



**HAL**  
open science

# Grandes valeurs propres de graphes aléatoires dilués

Simon Coste

► **To cite this version:**

Simon Coste. Grandes valeurs propres de graphes aléatoires dilués. Mathématiques [math]. Université Paul Sabatier (Toulouse 3), 2019. Français. NNT: . tel-02303646

**HAL Id: tel-02303646**

**<https://hal.science/tel-02303646>**

Submitted on 2 Oct 2019

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



# THÈSE

## En vue de l'obtention du DOCTORAT DE L'UNIVERSITÉ DE TOULOUSE

Délivré par l'Université Toulouse 3 - Paul Sabatier

---

Présentée et soutenue par  
**Simon COSTE**

Le 30 septembre 2019

**Grandes valeurs propres de graphes aléatoires dilués**

---

Ecole doctorale : **EDMITT - Ecole Doctorale Mathématiques, Informatique et Télécommunications de Toulouse**

Spécialité : **Mathématiques et Applications**

Unité de recherche :  
**IMT : Institut de Mathématiques de Toulouse**

Thèse dirigée par  
**Charles BORDENAVE et Justin SALEZ**

Jury

**Mme Ioanna DUMITRIU**, Rapporteur  
**M. Nati LINIAL**, Rapporteur  
**M. Laurent Massoulié**, Examineur  
**Mme Alice Guionnet**, Examinatrice  
**M. Antti Knowles**, Examineur  
**M. Djalil Chafaï**, Examineur  
**M. Charles BORDENAVE**, Directeur de thèse  
**M. Justin SALEZ**, Co-directeur de thèse



« Oh ! reprit-il, je donnerais bien cent sous au mathématicien qui me démontrerait par une équation algébrique l'existence de l'enfer. Il jeta une pièce en l'air en criant :

— Face pour Dieu !

— Ne regarde pas, dit Raphaël en saisissant la pièce, que sait-on ? Le hasard est si plaisant. »

BALZAC, *La peau de chagrin* (1830)



Je remercie en premier lieu mes deux directeurs de thèse, Charles Bordenave et Justin Salez, qui ont su me mettre en confiance et m'aider durant ces trois années. Je mesure mal la chance que j'ai eue de pouvoir travailler aux côtés de personnes si exceptionnelles ; ils resteront pour moi des exemples à suivre.

Je suis particulièrement honoré que Ioana Dumitriu et Nati Linial aient accepté d'être rapporteurs de cette thèse, et je les remercie chaleureusement pour leur relecture attentive du manuscrit. Je remercie également Antti Knowles, Alice Guionnet, Djalil Chafaï et Laurent Massoulié pour avoir accepté de participer à mon jury, ainsi que pour les discussions que nous avons pu avoir durant ces trois années.

Je remercie dans le désordre Selim Sukhtaiev, Nick Cook, Yizhe Zhu, Raj Rao Nadakuditi, Guilhem Semerjian, Andrea Montanari, Lorenzo Dall'Amico, Simone Warzel, Yan Fyodorov, Pierre Youssef, Andris Ambainis pour les simulantes discussions que nous avons eues lors de séminaires passés. Je remercie aussi les doctorants ou post-docs avec lesquels j'ai partagé une fraction de cette thèse ou son intégralité :

Anna	Côme	Henri	Mina	Sothea
Alejandro	Cyril	Hiroshi	Michel	Sylvain
Arturo	Dimas	Houzhi	Nathan	William
Assaf	Enzo	Junchao	Othmane	Xiaoli
Barbara	Fabio	Laure	Paul	Yann
Benjamin	Fanny	Luca	Pierre	Yiyang
Bohdan	Fernando	Lucas	Reda	Ziad
Clément	Guillaume	Marc	Rémy	
Clément	Guillaume	Mi-Song	Shangjie	

C'est grâce à eux que règne une bonne ambiance dans tous les bureaux de doctorants du monde entier, et du LPSM en particulier.

Je remercie Valérie et Nathalie (de Paris) et Agnès et Martine (de Toulouse), sans lesquelles la situation administrative *compliquée* dans laquelle je me suis parfois trouvé m'aurait certainement tenu en échec. Cette thèse est aussi la leur.

Mes derniers remerciements vont naturellement à ma famille, mon épouse et mes amis, qui m'ont toujours entouré de leur bienveillance et de leur amour.

---

# Table des matières

<b>1</b>	<b>Introduction</b>	<b>1</b>
1.1	La convergence du spectre . . . . .	2
1.2	Valeurs propres des graphes réguliers . . . . .	6
1.3	Digression : la suite des degrés . . . . .	13
1.4	Spectre des arbres . . . . .	17
1.5	Valeurs propres de matrices aléatoires diluées . . . . .	25
1.6	Deux remarques pour finir . . . . .	32
<b>2</b>	<b>The spectral gap of sparse random digraphs</b>	<b>37</b>
2.1	Introduction and statement of the results . . . . .	37
2.2	Proof of the main theorem . . . . .	43
2.3	Proof of Proposition 2.2.8 . . . . .	48
2.4	Expectation of a product of centered random variables . . . . .	52
2.5	General strategy and definitions for the proof of Proposition 2.2.6 . . . . .	53
2.6	Combinatorics of $\mathcal{C}_m$ . . . . .	56
2.7	Upper bound for $f$ . . . . .	61
2.8	Asymptotic analysis . . . . .	63
2.9	Proof of Proposition 2.2.7 . . . . .	66
2.10	Algebraic tools . . . . .	72
2.11	Proof of Theorem 3 . . . . .	74
<b>3</b>	<b>An Erdős-Gallai generalization</b>	<b>81</b>
3.1	Reconstruction of a graph with its universal covering . . . . .	81
3.2	Related work . . . . .	83
3.3	Definitions and statement of the main result . . . . .	84
3.4	Proof of Theorem 7 . . . . .	85
<b>4</b>	<b>Extended states in the spectrum of random trees</b>	<b>89</b>
4.1	Introduction . . . . .	89
4.2	Main ingredients . . . . .	94
4.3	Unimodular Galton-Watson trees . . . . .	98
4.4	The skeleton tree . . . . .	103
<b>5</b>	<b>Matrix reconstruction from sparse observations</b>	<b>109</b>
5.1	Introduction and statement of the results. . . . .	109
5.2	Related work . . . . .	114
5.3	An algebraic perturbation lemma . . . . .	117
5.4	Proof of Theorem 15 . . . . .	120
5.5	Consequences of algebraic incoherence . . . . .	125
5.6	Coupling graphs and trees . . . . .	128
5.7	Graph functionals . . . . .	133



5.8	Near eigenvectors: proofs . . . . .	137
5.9	Eigenwaves on Galton-Watson trees . . . . .	144
5.10	Proof of Proposition 5.8.5 . . . . .	147
<b>A</b>	<b>Introduction: english version</b>	<b>155</b>
A.1	Convergence of the spectrum . . . . .	156
A.2	Eigenvalues of regular graphs . . . . .	160
A.3	Digression : the degree sequence . . . . .	166
A.4	Spectra of trees . . . . .	171
A.5	Eigenvalues of random diluted matrices . . . . .	178
A.6	Two concluding remarks . . . . .	185

# Chapitre 1

## Introduction

*The english version of this introduction is in the appendix.*

Dans cette introduction, je m'attacherai à présenter les travaux effectués durant ces trois années de thèse. Ils portent principalement sur le spectre des graphes (ou des matrices) aléatoires. Une idée importante qui sous-tend de nombreux chantiers de la recherche moderne est que les plus grandes valeurs propres d'une matrice sont celles qui apportent le plus d'informations sur cette matrice. Il est donc d'un grand intérêt de disposer de descriptions fines de ces valeurs propres. J'essaierai de présenter autant que possible cette importance à travers les travaux présentés.

Le premier travail de cette thèse est inspiré de la prépublication *The spectral gap of sparse random digraphs* ([62]). On y étudie la deuxième valeur propre de la matrice de transition sur les graphes de configurations dirigés : dans ce modèle, on spécifie le degré entrant  $d_i^-$  et sortant  $d_i^+$  de chaque sommet et on prend un multi-graphe dirigé  $G$  uniformément au hasard parmi les graphes qui ont ces degrés. La deuxième valeur propre de la matrice de transition possède une borne elle-même fonction des degrés ; au passage, cela nous permet de résoudre la conjecture *dirigée* d'Alon sur les graphes réguliers dirigés.

De façon générale, une question importante de la théorie des graphes consiste à relier un graphe à sa suite des degrés ; on peut en premier lieu se demander si, étant donnés des entiers  $(d_1, \dots, d_n)$ , il est possible de construire un graphe  $G$  ayant ces degrés. Ce n'est pas toujours le cas : la question fut entièrement résolue par Erdős et Gallai dans les années 1960, et a généré tout un champ de recherches sur la possibilité de reconstruire des graphes à partir d'informations locales comme les degrés. Dans le deuxième travail présenté ici, inspiré d'un article en collaboration avec Charles Bordenave ([41]), on s'intéresse à l'existence de graphes ayant des *voisinages* donnés, la question ayant récemment émergé dans le cadre de travaux sur la convergence locale des graphes.

Ces deux premières parties sont relativement proches par leur objet, les graphes et leurs degrés. La troisième partie, fruit d'une collaboration avec Justin Salez ([63]), diffère un peu des deux premières ; l'objet principal reste les grands graphes aléatoires avec peu d'arêtes, mais on y étudie la nature du spectre, et en particulier l'existence d'une partie continue, qui correspond à la notion d'*états étendus* venue de la mécanique quantique. On démontre notamment un critère permettant de déterminer l'existence ou l'absence d'états étendus dans le spectre d'un modèle très général, les arbres de Galton-Watson unimodulaires. Cela nous a permis de répondre entre autres à une question posée par les physiciens Bauer et Golinelli en 2001.

Enfin, dans la dernière partie, issue d'un travail en collaboration avec Charles Bordenave et Raj Rao Nadakuditi, je présente un résultat sur les grandes valeurs propres des matrices. On considère un graphe d'Erdős-Rényi dirigé sur  $n$  sommets, dans le régime dilué où le degré moyen est  $d$ , indépendant de  $n$ . On pondère chaque arête  $(i, j)$  du graphe par un poids  $P_{i,j}$ . L'objectif est d'obtenir des

informations sur la matrice  $P$  à partir de l'observation de la matrice d'adjacence pondérée  $A$  ; c'est le problème de la complétion de matrice, extrêmement étudié depuis une quinzaine d'années. Nous démontrons une transition de phase spectaculaire : sous des hypothèses naturelles sur la matrice  $P$ , il existe un seuil  $\vartheta = \vartheta(P, d)$  tel que les plus grandes valeurs propres de  $(n/d)A$  convergent vers les valeurs propres de  $P$  plus grandes que  $\vartheta$ . Toutes les autres valeurs propres de  $(n/d)A$  sont contenues dans  $D(0, \vartheta)$ . Nous démontrons également que les vecteurs propres de  $A$  associés à ces valeurs propres sont corrélés avec les vecteurs propres correspondants de  $P$ . Ce résultat est inattendu, car il montre que la reconstruction (faible) de  $P$  ou d'une partie de  $P$  est possible même dans le régime où  $d$  est fixé.

Cette introduction donne une description complète de ces résultats, ainsi que des éléments de contexte sur chacun des domaines abordés. En premier lieu (§1.1), je donne quelques généralités sur le spectre des matrices et des graphes, avant de décrire la théorie de la convergence locale qui est sous-jacente à tous les travaux présentés ici. Les autres chapitres (§1.2-1.5) sont chacun consacrés aux travaux mentionnés ci-dessus et sont relativement indépendants.

Enfin, je donne en conclusion (§1.6, page 32) deux idées qui émanent de cette thèse : le régime dilué (où est  $d$  vraiment petit) est le régime intéressant du point de vue des applications, et parfois il vaut mieux dé-symétriser des problèmes qui sont naturellement symétriques, un programme qui semble assez prometteur au vu de certains résultats récents de la littérature.

## 1.1 La convergence du spectre

On rappelle quelques notions classiques d'algèbre linéaire et de théorie des graphes, puis on décrit la convergence des graphes au sens de Benjamini-Schramm.

### Les valeurs propres

Toute matrice symétrique  $A \in \mathcal{M}_{n,n}(\mathbb{R})$  possède une décomposition spectrale, au sens où elle s'écrit

$$A = \lambda_1 \varphi_1 \varphi_1^* + \cdots + \lambda_n \varphi_n \varphi_n^*, \quad (1.1.1)$$

avec  $\lambda_1 \geq \cdots \geq \lambda_n$  ses valeurs propres, et  $\varphi_i$  la famille orthonormée des vecteurs propres associés. Les valeurs propres capturent une grande quantité d'informations sur la matrice  $A$ , et disposent de caractérisations classiques très utiles, par exemple

$$\lambda_i = \max_{V \in \mathcal{V}_{i,n}} \min_{\substack{x \in V \\ \|x\|=1}} \langle x, Ax \rangle \quad (1.1.2)$$

où  $\mathcal{V}_{i,n}$  est l'ensemble des sous-espaces vectoriels de  $\mathbb{R}^n$  de dimension  $i$ . L'étude des valeurs propres de matrices symétriques *aléatoires*, initiée avec les travaux de Wigner, s'est révélée extrêmement fertile en développements venus de la combinatoire, des statistiques, de la physique ou même de la théorie des nombres.

Lorsque la matrice  $A$  n'est plus symétrique, elle possède tout de même  $n$  valeurs propres complexes, mais il n'y a pas forcément de décomposition de la forme (1.1.1). Une variante populaire et utile est la décomposition en valeurs singulières : pour toute matrice  $A \in \mathcal{M}_{m,n}(\mathbb{R})$ , il existe des  $\sigma_1 \geq \cdots \geq \sigma_r > 0$ , avec  $r = \text{rang}(A)$ , et deux familles orthonormales  $(\phi_1, \dots, \phi_r)$  dans  $\mathbb{R}^m$  et  $(\psi_1, \dots, \psi_r)$  dans  $\mathbb{R}^n$ , telles que

$$A = \sigma_1 \phi_1 \psi_1^* + \cdots + \sigma_r \phi_r \psi_r^*.$$

Des caractérisations variationnelles similaires à (1.1.2) existent pour les  $\sigma_i$ . Dans beaucoup d'applications, les valeurs singulières des matrices se sont révélées extrêmement utiles, jouant *a priori* le rôle des valeurs propres pour les matrices symétriques. Ce n'est pourtant pas exactement le cas, et dans plusieurs problèmes naturellement non-symétriques, les valeurs singulières se sont révélées moins informatives que les valeurs propres. L'un des objectifs de cette thèse est d'étudier les valeurs propres de certaines matrices non-hermitiennes.

## Spectre des graphes

Commençons par rappeler un peu de terminologie des graphes : un graphe  $G = (V, E)$  est la donnée d'un ensemble de sommets  $V$ , toujours dénombrable dans cette thèse, et d'une partie  $E$  de  $V \times V$ . Les éléments  $(u, v)$  dans  $E$  sont les arêtes du graphe. Les graphes simples sont les graphes sans boucles  $(v, v)$  et dont les arêtes sont symétriques, au sens où  $(u, v) \in E$  si et seulement si  $(v, u) \in E$  ; on dit aussi graphe non-orienté ou non-dirigé, par opposition au cas où des arêtes  $(u, v)$  peuvent exister sans que l'arête  $(v, u)$  ne soit dans  $E$ . Les travaux présentés dans cette thèse se rapportent aux graphes *dilués*, c'est-à-dire les graphes dont le nombre d'arêtes est comparable au nombre de sommets.

Tout graphe  $G = (V, E)$  est entièrement caractérisé par sa *matrice d'adjacence*  $A$ . Cette matrice est indexée par  $V$  et se définit par

$$A_{u,v} = \mathbf{1}_{(u,v) \in E}.$$

La matrice  $A$  est hermitienne si et seulement si le graphe correspondant est non-orienté. La théorie algébrique des graphes consiste à étudier  $G$  via le spectre de sa matrice d'adjacence — pour des manuels de référence, on renvoie à [51, 64]. L'un des avantages de cette vaste idée est qu'il existe de nombreuses procédures numériques pour approcher numériquement le spectre, en temps  $O(n^3)$  dans le pire des cas, et que la connaissance du spectre permet d'obtenir des bonnes approximations de quantités très difficiles à calculer, voire souvent NP-complètes, comme la constante isopérimétrique ou le nombre chromatique — nous y reviendrons dans la section 1.2. De nombreux problèmes statistiques ou informatiques qui se posent en termes de graphes peuvent ainsi être résolus par des algorithmes spectraux ; voir par exemple [134, 90, 10].

## Graphes aléatoires

La théorie classique des graphes s'est attachée à étudier les propriétés de graphes particuliers : étant donné un graphe, quelles relations peut-on donner entre, par exemple, son nombre chromatique, son nombre d'arête, sa planarité, etc. C'est souvent nécessaire dans les applications, où l'on dispose d'un graphe précis (par exemple *le* graphe de Facebook) duquel on veut tirer des informations.

Pourtant, dans beaucoup d'autres applications, on ne dispose pas d'un seul graphe, mais de plusieurs graphes qui vérifient certaines propriétés communes (comme *les* graphes des liens entre utilisateurs de divers réseaux sociaux). Plutôt que d'étudier ces graphes individuellement, on étudie donc des classes entières de graphes, et on essaie de déterminer des propriétés globalement vraies sur ces classes. Mathématiquement, il s'agit évidemment d'identifier une classe particulière, par exemple l'ensemble  $\mathcal{G}(n)$  des graphes simples sur  $n$  sommets, et d'étudier les propriétés d'une variable aléatoire (uniforme ou non)  $G$  sur cet ensemble. Dans cette thèse, on étudie plusieurs modèles classiques de graphes aléatoires : les graphes d'Erdős-Rényi, les arbres uniformes et les graphes avec degrés prescrits.

La plupart des résultats sont énoncés sous le régime asymptotique dans lequel la taille  $n$  des graphes tend vers l'infini. On étudiera donc des suites de graphes aléatoires de plus en plus grands. Dans ce régime, les modèles que nous étudions peuvent être vus comme des approximations finies d'objets-limites infinis, à l'aide de la *convergence locale faible*. Ce point de vue est à la fois une heuristique fertile pour comprendre certains problèmes (en particulier, ceux exposés dans la première), et une théorie puissante pour en résoudre d'autres (comme dans la troisième partie).

## Convergence de Benjamini-Schramm

L'idée principale de la convergence locale faible des graphes est d'étudier l'aspect *typique* du graphe autour d'un sommet quelconque. On présente dans cette section les grandes lignes et résultats de cette théorie ; on renvoie à [8, 34] pour des introductions complètes, ainsi qu'à [27].

Pour décrire la théorie, on a besoin de définir quelques termes d'usage courant : signalons pour commencer que tous les graphes étudiés auront pour sommets un même ensemble  $V$  (ou une partie

de  $V$ ), supposé dénombrable, et que tous les graphes considérés auront tous leurs degrés finis (on dit aussi qu'ils sont localement finis).

- Un graphe enraciné est un couple  $(G, v)$  où  $G$  est un graphe connexe et  $v \in V$  est un sommet particulier de  $G$ , appelé racine.
- Lorsque  $(G, v)$  est un graphe enraciné, on notera  $(G, v)_t$  l'ensemble des sommets de  $G$  à distance inférieure ou égale à  $t$  de la racine  $v$ .
- Un isomorphisme entre deux graphes enracinés  $(G_1, v_1)$  et  $(G_2, v_2)$  est un isomorphisme de graphes  $\varphi : G_1 \rightarrow G_2$  qui conserve la racine, c'est-à-dire tel que  $\varphi(v_1) = v_2$ .
- Lorsqu'il existe un tel isomorphisme entre deux graphes enracinés  $(G_1, v_1)$  et  $(G_2, v_2)$ , on écrit  $(G_1, v_1) \simeq (G_2, v_2)$ .
- On note enfin  $\mathcal{G}_*$  l'ensemble des (classes d'isomorphismes) des graphes connexes enracinés localement finis.

On peut munir  $\mathcal{G}_*$  d'une distance, dite « locale », en posant  $d((G_1, v_1), (G_2, v_2)) = (1 + T)^{-1}$ , où  $T = \sup\{t \geq 0 : (G_1, v_1)_t \simeq (G_2, v_2)_t\}$ . L'espace métrique ainsi obtenu est polonais. Il est donc possible de munir  $\mathcal{P}(\mathcal{G}_*)$ , l'ensemble des mesures de probabilités sur  $\mathcal{G}_*$ , de la topologie de la convergence faible. Concrètement, cela veut dire que  $\mu_n \rightarrow \mu$  si et seulement si pour toute fonction  $f : \mathcal{G}_* \rightarrow \mathbb{R}$  continue bornée, on a  $\int f d\mu_n \rightarrow \int f d\mu$ . On dira qu'une suite de graphes enracinés aléatoires  $(G_n, v_n)$  converge vers un graphe  $(G_\infty, v_\infty)$  si les lois de  $(G_n, v_n)$  convergent vers la loi de  $(G_\infty, v_\infty)$ . En pratique, on a  $G_n \rightarrow G_\infty$  si et seulement si pour tout graphe enraciné fixé  $(g, v)$ , on a

$$\forall t \in \mathbb{N}, \quad \mathbf{P}((G_n, v_n)_t \simeq (g, v)_t) \rightarrow \mathbf{P}((G_\infty, v_\infty)_t \simeq (g, v)_t).$$

L'idée centrale de la convergence de Benjamini-Schramm est que n'importe quelle suite de graphes finis  $(G_n)$  donne naissance à une suite de graphes connexes enracinés **aléatoires** par enracinement uniforme : il suffit de prendre la racine  $v_n$  uniformément au hasard sur  $V_n$ , et de considérer le graphe  $(G_n(v_n), v_n)$ , où la notation  $G(v)$  désigne la composante connexe de  $G$  contenant  $v$ .

Cette idée très puissante permet d'étudier des suites de graphes finis d'un point de vue purement local : la convergence d'une suite de graphes a lieu si les voisinages de ses points ressemblent (en loi) aux voisinages de la racine du graphe limite.

De nombreux modèles de graphes convergent presque sûrement au sens de Benjamini-Schramm. C'est le cas des modèles étudiés dans cette thèse, que nous présentons (entre autres) maintenant. La littérature sur chacun de ces modèles occupe des pages entières de bibliographies ; les trois premiers modèles sont extensivement décrits dans [32].

- (1) **Erdős-Rényi.** On note  $ER(n, p)$  la loi d'Erdős-Rényi de paramètres  $d/n$  ; l'ensemble des sommets est de taille  $n$  et chacune des  $n(n-1)/2$  arêtes possibles est indépendamment mise dans le graphe  $G$  avec probabilité  $p$ . Si  $p = d/n$  avec  $d$  un réel positif fixé, alors  $G_n \rightarrow PGW(d)$ , l'arbre de Galton-Watson avec loi de reproduction poissonnienne de paramètre  $\lambda$ . On renvoie à [34] pour une démonstration.
- (2) **Graphes avec degrés prescrits.** Soit  $\mathbf{d}^{(n)} = (d_1, \dots, d_n)$  une suite graphique<sup>1</sup>, et soit  $G_n$  pris uniformément au hasard sur  $\mathcal{G}(n, \mathbf{d}^{(n)})$ . On suppose que la loi empirique des degrés converge vers une mesure  $\pi$  sur  $\mathbb{N}$ , en loi et dans  $L^2$ . Si  $\pi(\{0\}) < 1$ , alors  $G_n \rightarrow UGW(\pi)$ , l'arbre de Galton-Watson unimodulaire avec loi des degrés  $\pi$ . Cet objet sera défini rigoureusement dans la dernière partie. Pour les démonstrations, voir [34, Chapitre 3].

---

<sup>1</sup>Une suite  $(d_1, \dots, d_n)$  est graphique s'il existe effectivement un graphe simple à  $n$  sommets dans lequel le degré du sommet  $i$  est  $d_i$ . Toutes les suites ne sont pas graphiques : cela fait l'objet de toute la section 1.3.

- (3) **Graphes réguliers.** C'est un cas particulier du précédent : en effet, dans ce cas on a  $\mathbf{d}^{(n)} = (d, \dots, d)$ , le graphe  $G_n$  est uniforme parmi tous les graphes  $d$ -réguliers à  $n$  sommets ; dans ce cas,  $\pi = \delta_d$  et l'arbre limite UGW( $\delta_d$ ) n'est autre que  $\mathbb{T}_d$ , l'arbre infini  $d$ -régulier<sup>2</sup> (en particulier, la limite est déterministe).
- (4) **Arbres uniformes.** Soit  $T_n$  un arbre choisi uniformément parmi l'ensemble des  $n^{n-2}$  arbres étiquetés sur  $n$  sommets. La suite  $T_n$  converge au sens de Benjamini-Schramm vers l'arbre squelette d'Aldous, noté  $T_{\text{squel}}$ , défini de la manière suivante : on commence par le graphe de  $\mathbb{N}$ , enraciné en zéro, appelé *colonne vertébrale* ; puis, à chaque noeud  $i$  de cette colonne vertébrale, on colle un arbre  $t_i$  qui est aléatoire et de loi PGW(1) — les arbres  $t_0, t_1, t_2, \dots$  sont indépendants ; voir figure 1.1b. Pour la démonstration de la convergence  $T_n \rightarrow T_{\text{squel}}$ , voir [85].
- (5) **Arbre canopée.** On note  $\mathbb{T}_{d,n} = (\mathbb{T}_n, o)_d$  l'arbre infini  $d$ -régulier dont a gardé uniquement les  $n$  premières générations. Contrairement au cas des graphes réguliers, cet arbre possède un bord : une proportion asymptotiquement égale à  $(d-2)/(d-1)$  des sommets sont des feuilles. La limite n'est donc pas  $\mathbb{T}_d$ , mais un bel objet récursif,  $(\mathbb{C}_d, o)$ , appelé arbre canopée : voir la figure 1.1a. L'identification du bon modèle limite entre  $\mathbb{T}_d$  et  $\mathbb{C}_d$  est une question intéressante en physique du modèle d'Anderson, où elle reflète la prise en compte des effets de bord ([5], [122]).

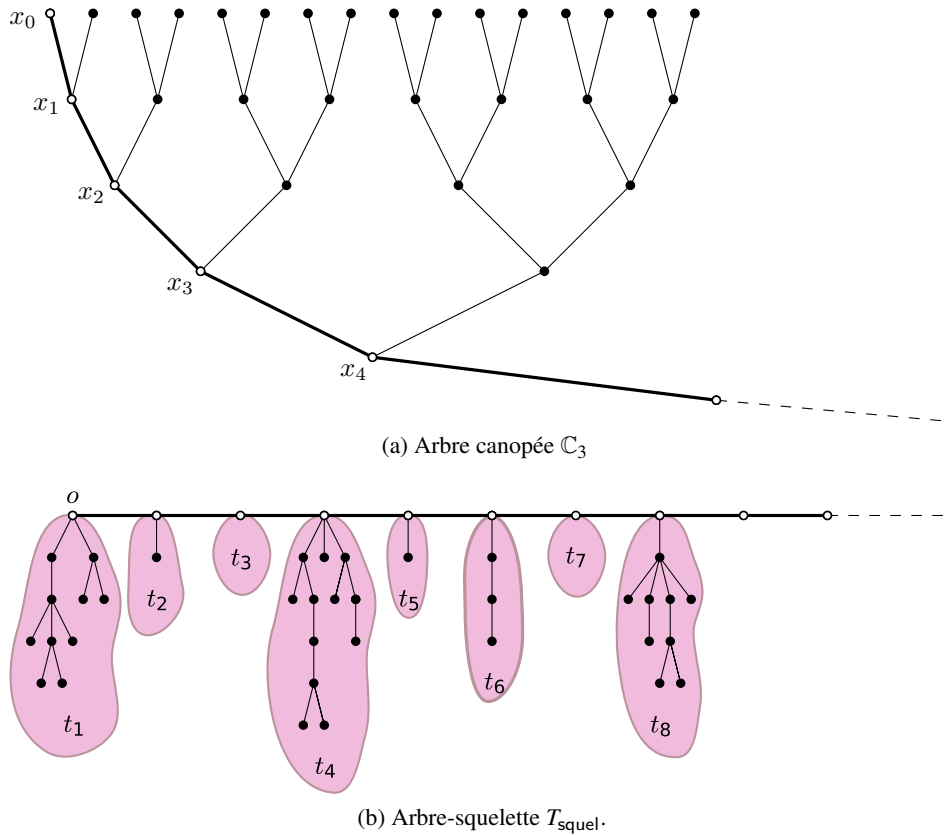


FIGURE 1.1 – Deux exemples d'arbres unimodulaires. Dans l'arbre-squelette, les  $t_i$  sont iid de loi PGW(1).

<sup>2</sup>Aussi appelé *réseau de Bethe*.

### Continuité du spectre

Soit  $G_n$  une suite de graphes finis, qui converge au sens de Benjamini-Schramm vers un graphe  $(G, \nu)$ . La matrice d'adjacence  $A_n$  de  $G_n$  a pour spectre  $\lambda_1(A_n) \geq \dots \geq \lambda_n(A_n)$  et on notera

$$\mu_{A_n} = \frac{1}{n} \sum_{i=1}^n \delta_{\lambda_i(A_n)}$$

la mesure spectrale empirique de  $G_n$ . Un des grands intérêts de la convergence de Benjamini-Schramm est de garantir que  $\mu_{A_n}$  converge vers une mesure limite, définie à partir de  $(G, \nu)$  (voir [1, 126] ou encore [34, Proposition 2.5]). Plus précisément, si  $G_n$  converge au sens de Benjamini-Schramm vers un graphe limite  $(G, \nu)$  de loi  $\rho$ , alors

$$\sup_{t \in \mathbb{R}} |F_{A_n}(t) - F_\rho(t)| \rightarrow 0 \tag{1.1.3}$$

où  $F_{A_n}$  est la fonction de répartition de  $\mu_{A_n}$ , et  $F_\rho$  est la fonction de répartition d'une mesure  $\mu_\rho$  sur  $\mathcal{G}_*$  qui est définie à partir de  $\rho$ . La convergence (1.1.3) équivaut en fait à la convergence de tous les atomes (i.e.  $\mu_{A_n}(\{\lambda\}) \rightarrow \mu_\rho(\{\lambda\})$  pour tout réel  $\lambda$ ) plus la convergence faible de  $\mu_{A_n}$  vers  $\mu_\rho$ . Toute la difficulté est évidemment d'identifier la mesure limite  $\mu_\rho$ , et ici intervient une propriété importante : l'unimodularité. Nous donnerons une définition explicite de  $\mu_\rho$  à la page 19, ainsi qu'une description de l'unimodularité.

Telle quelle, la convergence (1.1.3) ne permet pas d'élucider le comportement exact de certaines valeurs propres particulières, comme  $\lambda_1(A_n)$  par exemple, car le comportement d'un nombre fini ou négligeable devant  $n$  de valeurs propres est perdu dans la convergence locale faible. L'étude du comportement asymptotique des plus grandes valeurs propres ne peut donc pas entièrement se fonder sur la continuité du spectre (même si cette dernière permet d'obtenir quelques informations), et nécessite des méthodes ad hoc comme la méthode de la trace. C'est ce genre de méthodes qui gouverne l'étude de la deuxième valeur propre de certains graphes avec des degrés prescrits, que je présente maintenant.

## 1.2 Valeurs propres des graphes réguliers

On s'intéresse dans cette partie aux graphes avec une suite de degrés fixés, et à la deuxième valeur propre de leur matrice de transition ; les résultats et les démonstrations sont rassemblés dans le chapitre 1 de cette thèse, qui reprend la prépublication [62].

Pour commencer, on donne un bref aperçu des motivations autour de cette question, notamment à travers le théorème d'Alon. Pour fixer les idées, je commence par présenter ce dernier dans le cadre des graphes simples (non orientés, sans boucles) et  $d$ -réguliers.

### Spectre des graphes réguliers et expansion

Donnons-nous un graphe  $d$ -régulier  $G$ . La marche aléatoire sur ce graphe a pour matrice de transition  $P_{u,v} = \frac{1}{d} \mathbf{1}_{(u,v) \in E}$ . Il est clair que  $P = \frac{1}{d}A$ , où  $A$  est la matrice d'adjacence ; dans ce cadre, le spectre de  $P$  et le spectre de  $A$  sont identiques à une dilatation près. Les résultats de cette section sont formulés en termes de la matrice de transition  $P$ .

Il n'est pas difficile de vérifier que la plus grande valeur propre de  $P$  est 1, et la multiplicité de 1 est égale au nombre de composantes connexes de  $G$  ; d'autre part,  $-1$  est valeur propre si et seulement si le graphe est biparti (voir par exemple [31, Section VIII.2]). Le comportement de la plus grande valeur propre est donc finalement assez peu informatif. La question de connaître le comportement de la *seconde plus grande valeur propre* est autrement plus riche et difficile. Notons

$$\lambda_* = \max\{|\lambda| : \lambda \text{ valeur propre de } P \text{ avec } |\lambda| < 1\}$$

le module de la plus grande valeur propre non triviale. Il existe de nombreux résultats qui relient cette quantité à des propriétés importantes du graphe liées à son *expansion* : l'idée générale est que si  $\lambda_*$  est petite, le graphe est un bon «*expanseur*», au sens où ses arêtes sont très bien distribuées dans le graphe et où la suppression de quelques-unes d'entre elles ne provoque pas de changement majeur dans sa géométrie. Plusieurs outils permettent de quantifier cette intuition.

**Inégalité de Cheeger.** Pour tout graphe  $d$ -régulier  $G$ , on a

$$\frac{d(1 - \lambda_*)}{2} \leq \Phi(G) \leq d\sqrt{1 - \lambda_*} \quad (1.2.1)$$

où  $\Phi(G) = \min_{A \subset V, |A| \leq |V|/2} \frac{|\partial A|}{|A|}$  est la constante isopérimétrique<sup>3</sup> du graphe. Lorsque  $\lambda_*$  est petite, cet encadrement entraîne que  $\Phi(G)$  est grande : pour chaque partie  $A$ , il y a beaucoup d'arêtes qui sortent de  $A$ .

**Expander-mixing lemma.** Un autre lien entre  $\lambda_*$  et les propriétés de  $G$  est le célèbre *expander mixing lemma* : si  $G$  est un graphe  $d$ -régulier, alors pour tous  $X, Y \subset V$ , on a

$$\left| E(X, Y) - \frac{d|X||Y|}{n} \right| \leq d\lambda_* \sqrt{|X||Y|} \quad (1.2.2)$$

où  $E(X, Y) = \{(u, v) \in E : u \in X, v \in Y\}$  est le nombre d'arêtes entre  $X$  et  $Y$ . Le nombre d'arêtes possibles entre  $X$  et  $Y$  est  $|X||Y|/2$  et la densité des arêtes est  $d/(n-1)$ , approximativement  $d/n$  si  $n$  est grand. Par conséquent, si les arêtes de  $G$  étaient mises au hasard dans  $G$  avec cette densité (ce qui revient à prendre un graphe d'Erdős-Rényi de paramètres  $(n, d/n)$ ), le nombre moyen d'arêtes entre  $X$  et  $Y$  devrait être  $d|X||Y|/n$ . L'expander-mixing lemma donne donc l'écart entre ce nombre moyen et le nombre effectif d'arêtes entre  $X$  et  $Y$ .

Ces deux inégalités sont des outils extrêmement efficaces pour obtenir des informations sur  $G$  à partir d'une simple borne sur  $\lambda_*$  ou  $\lambda_2$ . Citons par exemple des bornes sur le nombre chromatique ou le diamètre ; on renvoie à [87, Chapitre 4] pour de nombreux développements autour de ce thème.

Les deux inégalités (1.2.1) et (1.2.2) ne sont vraies *stricto sensu* que pour les graphes  $d$ -réguliers, mais elles possèdent une infinité de variantes dans le cadre de graphes orientés, ou encore lorsque le graphe n'est pas régulier.

**Chaînes de Markov.** Dans un autre cadre,  $\lambda_*$  donne la vitesse de convergence en temps long des chaînes de Markov. En règle générale, la matrice de transition  $P$  d'une chaîne de Markov n'est ni symétrique, ni même diagonalisable. Ses valeurs propres ne sont donc pas nécessairement réelles, et on les ordonne par module décroissant :  $\lambda_1 = 1 \geq |\lambda_2| \geq \dots \geq |\lambda_n|$ . Lorsque la chaîne est irréductible aperiodique, la théorème de Perron-Frobenius entraîne que 1 est valeur propre simple, donc on a  $|\lambda_2| = \lambda_*$ .

**PROPOSITION 1.2.1** ([104], [117]). *Soit  $P$  la matrice de transition d'une chaîne de Markov irréductible aperiodique sur un espace d'états finis  $S = \{1, \dots, n\}$ , de loi invariante  $\pi_*$ . Soient  $1 = |\lambda_1| \geq |\lambda_2| \geq \dots \geq |\lambda_n|$  les valeurs propres de  $P$  ordonnées par taille décroissante, et soit  $d(t)$  la distance à l'équilibre au temps  $t$ , définie par  $d(t) = \max_{x \in S} \|P^t(x, \cdot) - \pi_*\|_{TV}$ , où  $\|\cdot\|_{TV}$  est la distance en variation totale. Alors,*

$$\lim_{t \rightarrow \infty} d(t)^{\frac{1}{t}} = |\lambda_2|. \quad (1.2.3)$$

Là encore, si  $|\lambda_2| = \lambda_*$  est petite, la chaîne de Markov sur le graphe  $G$  va rapidement converger vers son équilibre. Le calcul de  $|\lambda_2|$  en fonction des propriétés de  $G$  est donc crucial pour étudier des propriétés fines de la marche aléatoire sur le graphe. C'est ce qui a motivé de nombreux travaux sur l'étude de  $\lambda_*$  pour divers modèles de graphes, le plus simple — mais déjà difficile — étant celui des graphes réguliers.

<sup>3</sup>Aussi appelée constante de Cheeger, conductivité, constante d'expansion, etc.



### Borne d'Alon-Boppana

La célèbre borne d'Alon-Boppana exprime que  $\lambda_*$  est plus grande que  $2\sqrt{d-1}/d$ ; la formulation la plus précise, à ma connaissance, est la suivante : si  $P$  est la matrice de transition d'un graphe  $d$ -régulier  $G$ , alors

$$\lambda_2(P) \geq 2 \frac{\sqrt{d-1}}{d} \cos\left(\frac{\pi}{\text{diam}(G)+2}\right) \quad (1.2.4)$$

et en particulier  $\lambda_* \geq \lambda_2(P)$  est également plus grand que le terme de droite. En substance, si  $r = \text{diam}(G)$  est grand, on a

$$\lambda_*(P) \geq \frac{2\sqrt{d-1}}{d} \times \left(1 - O\left(\frac{1}{r^2}\right)\right).$$

La première démonstration est due à Alon ([120]), et cette version là est due à Mohar ([116]); toutes deux utilisent les formulations variationnelles (1.1.2) propres aux matrices symétriques.

Dans le cadre de la matrice d'adjacence des graphes simples, de nombreuses généralisations existent (voir notamment [125] et ses références). Si  $G_n$  est une suite de graphes  $d$ -réguliers dont le diamètre tend vers l'infini, on a donc

$$\liminf \lambda_*(P_n) \geq \frac{2\sqrt{d-1}}{d}. \quad (1.2.5)$$

Toute suite de graphes  $d$ -réguliers dont le diamètre tend vers l'infini a un diamètre qui tend aussi vers l'infini, et vérifie donc l'inégalité (1.2.5).

### Le nombre $2\sqrt{d-1}/d$

L'apparition du nombre  $2\sqrt{d-1}/d$  n'est pas fortuite, et résulte notamment de l'approximation de la mesure spectrale empirique de  $G$  par la loi de Kesten-McKay (voir figure 1.2). Plus précisément, soient  $G_n = (V_n, E_n)$  des graphes réguliers dont la taille  $|V_n|$  tend vers l'infini, et dont le diamètre tend vers l'infini.

On a vu plus tôt que si  $G_n$  est pris uniformément au hasard sur  $\mathcal{G}(n, d)$  alors  $G_n$  converge vers l'arbre infini  $d$ -régulier  $\mathbb{T}_d$  au sens de Benjamini-Schramm, et donc la mesure spectrale<sup>4</sup>  $\mu_{P_n}$  converge ([113]) vers le spectre de l'opérateur de transition sur  $\mathbb{T}_d$ , connue sous le nom de loi de Kesten-McKay. C'est une mesure absolument continue par rapport à la mesure de Lebesgue, de support  $[-2\sqrt{d-1}/d, 2\sqrt{d-1}/d]$ , et dont la densité est donnée par

$$\rho_d(x) = d \mathbf{1}_{\left[-\frac{2\sqrt{d-1}}{d}, \frac{2\sqrt{d-1}}{d}\right]} \frac{d\sqrt{4(d-1) - (dx)^2}}{2\pi(d^2 - (dx)^2)}. \quad (1.2.6)$$

---

<sup>4</sup>Stricto sensu, la convergence (1.1.3) s'applique à la mesure spectrale de la matrice d'adjacence de  $G_n$ , mais ici la renormalisation  $P = d^{-1}A$  est triviale, donc  $\mu_{P_n}$  converge vers la mesure  $\mu_{\mathbb{T}_d} \circ d$ , qui est bien entendu le spectre de l'opérateur de la marche aléatoire sur  $\mathbb{T}_d$ .

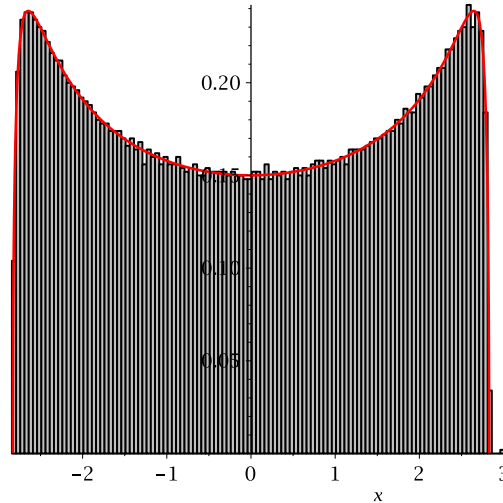


FIGURE 1.2 – Histogramme des valeurs propres de la matrice d’adjacence  $P$  d’un graphe 3-régulier avec 10000 sommets ; la mesure limite de Kesten-McKay  $\rho_3$  est en rouge.

Comme nous avons déjà mentionné, la convergence faible  $\mu_{G_n} \rightarrow \rho_d(x)dx$  ne donne aucune information précise sur le comportement de certaines valeurs propres particulières comme  $\lambda_2$ . Cependant, le lemme du porte-manteau entraîne que pour tout ouvert  $O \subset \mathbb{R}$  fixé, on a  $\liminf \mu_{G_n}(O) \geq \int_O \rho_d$ . En prenant  $O = ]2\sqrt{d-1}/d - \varepsilon, +\infty[$ , on a donc

$$\frac{|\{\text{valeurs propres de } P_n \text{ telles que } \lambda > \frac{2\sqrt{d-1}}{d} - \varepsilon\}|}{n} \rightarrow \int_{2\sqrt{d-1}/d - \varepsilon}^{\infty} \rho_d(t)dt := c(d, \varepsilon) > 0.$$

La proportion de valeurs propres de  $G_n$  qui sont plus grandes que  $2\sqrt{d-1}/d - \varepsilon$  est donc strictement positive (un résultat dû à Jean-Pierre Serre, [130]), ce qui est beaucoup plus fort que  $\lambda_* > 2\sqrt{d-1}/d - \varepsilon$ . Pour autant, la convergence (1.1.3) ne permet pas d’en dire plus : on pourrait très bien avoir  $\lambda_*(P_n) \rightarrow 1$ , par exemple. On verra que ce n’est pas le cas.

### Graphes de Ramanujan et théorème d’Alon-Friedman

Étant donnée l’importance de la deuxième valeur propre sur la structure des graphes et la borne inférieure d’Alon-Boppana, il est naturel de s’intéresser aux graphes qui sont extrémaux pour  $\lambda_2$ , ceux qui ont une deuxième valeur propre plus petite que  $2\sqrt{d-1}/d$ . Un graphe  $d$ -régulier connexe est appelé *graphe de Ramanujan* lorsque les valeurs propres de  $P$  sont soit de module 1, soit de module inférieur à  $2\sqrt{d-1}/d$ , comme dans la figure suivante.

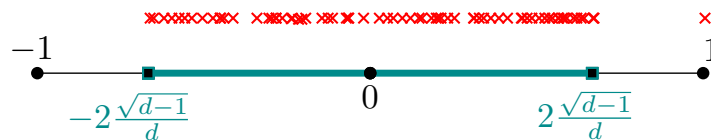


FIGURE 1.3 – En rouge, le spectre d’un graphe de Ramanujan. Lorsque  $-1$  est valeur propre, on parle souvent de graphe de Ramanujan *biparti*.

Ces graphes constituent des expandeurs optimaux au vu de la section précédente, et on ne s’étendra pas sur l’ampleur incroyable de leurs applications, pour laquelle on renvoie aux excellentes synthèses [66] et [87], ou plus récemment [124]. Signalons cependant que la question de leur existence s’est révélée hautement non-triviale.

La construction de familles de graphes de Ramanujan est un problème difficile, qui fut d'abord résolu pour certaines valeurs de  $d$  ([111, 107], pour les cas où  $d - 1$  est premier) : il est possible de construire des graphes de Ramanujan comme graphes de Cayley de  $\mathrm{PSL}_2(\mathbb{F}_q)$  à partir d'un ensemble bien choisi de générateurs<sup>5</sup> ; voir [66] pour une présentation accessible. L'existence de familles de graphes de Ramanujan  $d$ -réguliers a finalement été montrée pour tous les  $d > 2$  ([108, 109, 110], constructions de graphes de Ramanujan bipartis) ; signalons cependant que ces derniers travaux ont une nature purement existentielle et ne permettent pas de construire explicitement ces graphes de Ramanujan.

Ces "constructions" représentent un tour de force, mais ici la méthode probabiliste a permis de court-circuiter le problème de façon spectaculaire. En effet, vers la fin des années 1980, Alon conjecture que *presque tous les graphes réguliers sont presque des graphes de Ramanujan*. Plus précisément, il conjecture ([9]) que si  $G_n$  est choisi uniformément au hasard sur  $\mathcal{G}(n, d)$ , alors

$$\lambda_* \xrightarrow[n \rightarrow \infty]{\mathbf{P}} 2 \frac{\sqrt{d-1}}{d}. \quad (1.2.7)$$

Évidemment, (1.2.7) ne résout pas la question de la construction des graphes de Ramanujan, dans la mesure où on peut avoir des suites  $G_n$  qui satisfont  $\lambda_*(G_n) \rightarrow 2\sqrt{d-1}/d$  tout en n'étant pas de Ramanujan. Cependant, du point de vue des applications, il n'y a pas tant de différence, puisqu'une telle suite est asymptotiquement optimale au sens où elle atteint la borne inférieure d'Alon-Boppana.

La première démonstration de (1.2.7) date de 2004 ([79]), dans un article désormais célèbre ; une seconde démonstration, due à Bordenave [33], a introduit une méthode puissante pour étudier les grandes valeurs propres de certaines matrices. C'est sur cette méthode que se fonde le premier chapitre de cette thèse.

### Graphes dirigés réguliers

Un graphe dirigé est dit *régulier* lorsque tous les sommets ont les mêmes degrés, entrants et sortants :  $d_i^+ = d_i^- = d$ .

Comme mentionné au début de cette introduction, il serait illusoire de croire que les graphes dirigés sont simplement une variante des graphes simples. Toute la section précédente se fonde sur deux outils : d'abord, la borne d'Alon-Boppana, et ensuite la convergence de  $\mu_{P_n}$ . Or, à ce jour, on n'a pas d'équivalent satisfaisant ni de l'une, ni de l'autre, dans le cadre des graphes dirigés. Une des raisons est que la matrice de transition  $P_n$  n'est plus symétrique : ses valeurs propres sont des nombres complexes, et on n'a plus de caractérisations variationnelles. La théorie des *expandeurs dirigés* attire pourtant une attention croissante dans divers domaines, comme en témoigne la récente synthèse [124].

Commençons par examiner un éventuel analogue dirigé de la borne d'Alon-Boppana. La théorie de la convergence de Benjamini-Schramm est encore valable pour les graphes dirigés, et il n'est pas difficile de se convaincre que si  $G_n$  est un graphe  $d$ -régulier dirigé uniforme, alors  $G_n \rightarrow \vec{\mathbb{T}}_d$ , l'arbre  $d$ -régulier dirigé : chaque sommet possède  $d$  sommets entrants et  $d$  sommets entrants. Le problème réside dans la définition de la mesure spectrale sur  $\vec{\mathbb{T}}_d$ , puisque l'opérateur de transition sur cet arbre n'est pas auto-adjoint. On sait malgré tout que le spectre de cet opérateur, calculé dans [67], est égal à  $\{|z| \leq 1/\sqrt{d}\}$ . En fait, la forme de la mesure spectrale limite pour  $\mu_{A_n}$  est encore à l'état de conjecture ; on suppose que la mesure limite, supportée sur  $D(0, 1/\sqrt{d})$ , est donnée par ([40, page 70]) :

$$\frac{1}{\pi} \frac{d^2(d-1)}{(d^2 - |z|^2)^2} \mathbf{1}_{|z| < \sqrt{d}} dz. \quad (1.2.8)$$

Cette distribution est radiale mais n'est pas la radialisée de la mesure de Kesten-Mckay (1.4).

<sup>5</sup>La démonstration du fait que ces graphes sont de Ramanujan utilise des outils de théorie des nombres dus notamment à Ramanujan, d'où leur nom.

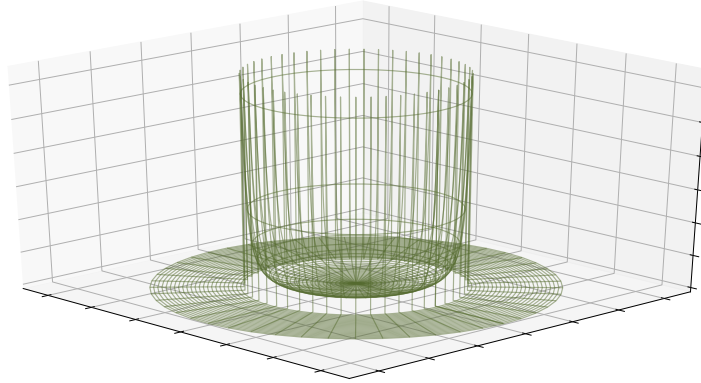


FIGURE 1.4 – Un aperçu de la densité de la « loi orientée de Kesten-McKay » pour  $d = 13$ . Lorsque  $d \rightarrow \infty$ , cette loi converge (après renormalisation par  $\sqrt{d}$ ) vers la loi du cercle, c'est-à-dire la loi uniforme sur  $D(0, 1)$ . Lorsque  $G_n$  est uniforme sur  $\mathcal{G}(n, d_n)$  et que  $d_n \rightarrow \infty$ , la convergence de  $\mu_{A_n}$  vers la loi du cercle est démontrée ([58]).

Une telle convergence impliquerait immédiatement un analogue dirigé de la borne (1.2.5) (elle-même plus faible qu'Alon-Boppana), mais il y a peu de chances que cette méthode soit généralisable à d'autres modèles — le calcul de la mesure spectrale d'opérateurs sur des arbres de Galton-Watson orientés n'est pas évident (la bonne notion de spectre pour les opérateurs non auto-adjoint est alors la mesure de Brown), et de même la convergence du spectre de graphes dirigés vers la mesure de Brown de leur limite au sens de Benjamini-Schramm semble pour l'instant hors de portée.

### Digraphes de Ramanujan

Un graphe  $d$ -régulier dirigé est appelé *digraphe de Ramanujan* si sa matrice de transition  $P$  vérifie  $\lambda_* \leq 1/\sqrt{d}$ . L'analogue du théorème de Friedman et de la convergence (1.2.7) s'énonce de la manière suivante (eq. 5.4 dans [124]) : pour toute suite  $(G_n)$  de graphes uniformes sur  $\mathcal{G}(n, d)$  de matrice de transition  $P_n$ , on a

$$\lambda_*(P_n) \rightarrow \frac{1}{\sqrt{d}} \quad (1.2.9)$$

en probabilité lorsque  $n \rightarrow \infty$ . Noter qu'avec grande probabilité,  $G_n$  est connexe, donc  $\lambda_* = |\lambda_2|$ . Le premier chapitre de cette thèse démontre la borne supérieure, analogue du théorème d'Alon-Friedman dans le cadre dirigé.

**Théorème 1.** *Soit  $d \geq 3$ , et soit  $G$  un graphe uniformément distribué sur  $\mathcal{G}(n, d)$ , de matrice de transition  $P$ . Alors, pour tout  $\varepsilon > 0$  on a*

$$\lambda_*(P_n) \leq \frac{1}{\sqrt{d}} + \varepsilon \quad (1.2.10)$$

avec une probabilité qui tend vers 1 lorsque  $n \rightarrow \infty$ .

La borne inférieure, analogue dirigé d'Alon-Boppana, reste à ce jour non démontrée. Comme expliqué dans le paragraphe précédent, un analogue probabiliste semble vrai ; en revanche, un analogue strict de (1.2.7) ne peut pas être vrai : certains digraphes  $d$ -réguliers non triviaux, comme le graphe de De Bruijn, ont toutes leurs valeurs propres non nulles de module 1 ([68]).

Le résultat que j'ai démontré dans ce premier article est en réalité considérablement plus général que le théorème 1, car il s'applique à *tous les graphes dirigés ayant une suite de degrés prescrits*.

### Graphes avec degrés prescrits

Les graphes réguliers ont une forme d'homogénéité très forte, au sens où chaque sommet a un comportement identique au vu des degrés. Dans de très nombreuses applications, notamment celles qui étudient des graphes issus de réseaux réels, les sommets ont des degrés différents, et souvent ces degrés sont accessibles : il est plus facile de disposer de la liste des degrés d'un réseau, que de toutes les connections dans le réseau. Pour modéliser de tels réseaux, on utilise les modèles de graphes à *degrés prescrits* ; on renvoie à [119, 60] et à leur bibliographie pour de nombreux exemples issus des mathématiques appliquées.

Dans un tel modèle, les degrés (entrants et sortants) de chaque sommet sont donnés : on connaît la liste  $\mathbf{d} = (d_1^+, d_1^-, \dots, d_n^+, d_n^-)$  des degrés, et on choisit  $G$  uniformément au hasard sur l'ensemble  $\mathcal{G}(n, \mathbf{d})$  des digraphes qui ont cette suite de degrés. La question de savoir si  $\mathcal{G}(n, \mathbf{d})$  n'est pas vide, c'est-à-dire de savoir s'il existe vraiment des graphes avec cette suite de degrés<sup>6</sup>, est une question ancienne et intéressante, sur laquelle nous reviendrons dans la deuxième partie de ce travail. Cette question est en réalité secondaire en raison d'un artifice célèbre, le *modèle de configurations* (voir [32], pour une présentation dans le cadre des graphes non dirigés).

Le problème du modèle uniforme sur  $\mathcal{G}(n, \mathbf{d})$  (mais aussi sur  $\mathcal{G}(n, d)$ ) est qu'on ne dispose pas d'une méthode facile et maniable pour générer une variable uniforme sur ces ensembles<sup>7</sup>. Plutôt que de faire cela, on préfère prendre non plus des graphes, mais des multi-graphes, au sens où les boucles  $(i, i)$  et les arêtes multiples sont autorisés. Pour n'importe quelle suite d'entiers  $\mathbf{d}$  vérifiant  $d_1^+ + \dots + d_n^+ = d_1^- + \dots + d_n^- := m$ , un tel multigraphe existe, et il y a une façon simple de la générer : on colle  $d_i^-$  demi-arêtes entrantes au sommet  $i$  et  $d_i^+$  demi-arêtes sortantes, puis pour chacune des  $m$  demi-arêtes sortantes, on la recolle au hasard à une des demi-arêtes entrantes disponibles. Le digraphe  $G$  est alors obtenu en identifiant chaque couple de demi-arêtes collées à une arête dirigée.

Or, le modèle uniforme sur  $\mathcal{M}(n, \mathbf{d})$  vérifie une forme d'absolue continuité par rapport au modèle uniforme  $\mathcal{G}(n, \mathbf{d})$ . Plus précisément, il est possible de montrer deux choses :

- i) il existe une constante  $c > 0$  dépendant uniquement de  $\max \mathbf{d}$  (et pas de  $n$ ) telle que si  $n$  est suffisamment grand,  $\mathbf{P}(G \in \mathcal{G}(n, \mathbf{d})) \geq c$ ,
- ii) la loi de  $G$  conditionné à être dans  $\mathcal{G}(n, \mathbf{d})$  est la loi uniforme sur  $\mathcal{G}(n, \mathbf{d})$ .

Par conséquent, si l'on se donne une suite  $(\mathbf{d}^{(n)})$  de listes de degrés, avec  $\mathbf{d}^{(n)}$  de longueur  $n$ , et  $\max \mathbf{d}^{(n)} \leq \Delta$  pour un certain entier  $\Delta$  ne dépendant pas de  $n$ , alors on dispose de la propriété suivante : pour tout événement  $E$ , on a

$$\lim_{n \rightarrow \infty} \mathbf{P}_{\mathcal{M}(n, \mathbf{d})}(G \in E) = 0 \implies \lim_{n \rightarrow \infty} \mathbf{P}_{\mathcal{G}(n, \mathbf{d})}(G \in E) = 0.$$

Autrement dit, toutes les propriétés asymptotiquement vraies pour le modèle de multigraphes seront vraies pour le modèle de digraphes correspondant. Le modèle sur les multigraphes est donc beaucoup plus général que le modèle sur les graphes, au moins au vu du comportement asymptotique.

### Forme générale du théorème d'Alon

Le résultat principal du premier chapitre de cette thèse est le théorème suivant, qui est l'analogue dirigé de (1.2.7) lorsque  $P_n$  est la matrice de transition d'un digraphe avec une suite de degrés fixée.

<sup>6</sup>Dans ce cas, on dit que  $\mathbf{d}$  est une suite digraphique.

<sup>7</sup>On ne connaît même pas leur cardinal en général, et il y a peu de chances pour que celui-ci s'exprime sous une forme fermée ou explicite.

**Théorème 2** ([62]). Soit  $\mathbf{d}^{(n)} = (d_1^+, \dots, d_n^+, d_1^-, \dots, d_n^-)$  une suite d'entiers avec  $d_1^- + \dots + d_n^- = d_1^+ + \dots + d_n^+ := m$ , et soit  $G_n$  le multigraphe obtenu via le modèle de configuration sur  $\mathcal{M}(n, \mathbf{d}^{(n)})$ . On suppose qu'il existe deux constantes  $3 \leq \delta \leq \Delta < \infty$  telles que pour tout  $n$ ,

$$\delta \leq \min \mathbf{d}^{(n)} \leq \max \mathbf{d}^{(n)} \leq \Delta. \quad (1.2.11)$$

Alors, pour tout  $\varepsilon > 0$ ,

$$|\lambda_\star(P_n)| \leq \max \left\{ \frac{1}{\delta}, \sqrt{\frac{1}{m} \sum_{i=1}^n \frac{d_i^-}{d_i^+}} \right\} + \varepsilon \quad (1.2.12)$$

avec une probabilité qui tend vers 1 lorsque  $n \rightarrow \infty$ .

En vertu des remarques précédentes, le résultat est le même lorsque  $\mathbf{d}$  est digraphique et  $G$  est pris uniformément sur  $\mathcal{G}(n, \mathbf{d})$ . Lorsque la suite des degrés est constante, on retrouve bien le théorème 1, puisqu'alors  $d_i^\pm = d$  pour tout  $i$ . Notons que si  $d_i^+ = d_i^-$  pour tout  $i$  (le graphe est alors eulérien), la borne obtenue est  $\sqrt{n/m} = 1/\sqrt{d_{\text{moy}}}$ , la racine inverse du degré moyen, un résultat à mettre en relation avec une borne similaire dans le modèle  $\text{ER}(n, d/n)$  démontrée dans [43].

La présence d'une éventuelle valeur propre de module proche de  $\delta^{-1}$  reste incertaine ; la démonstration ne permet pas de se passer du maximum dans (1.2.12) et les simulations numériques semblent indiquer que dans le régime où ce maximum est  $1/\delta$ , il existe effectivement une valeur propre de module proche de  $1/\delta$  avec probabilité *a priori* non nulle, comme dans la figure 1.5b. La question de savoir si la mesure limite possède effectivement un atome en  $\delta^{-1}$  reste ouverte.

### 1.3 Digression : la suite des degrés

Dans la partie précédente, nous avons étudié les propriétés spectrales de certains graphes ayant une suite de degré fixée, disons  $d_1, \dots, d_n$ . La question de savoir s'il existe effectivement un graphe avec cette suite de degré était secondaire, dans la mesure où le modèle de configuration permettait de se passer de cette hypothèse. Cependant, la question en elle-même est intéressante : quelles sont les conditions pour qu'une suite d'entiers soit la suite des degrés d'un graphe simple ? De telles suites sont dites *graphiques*.

#### La caractérisation d'Erdős-Gallai

La question fut résolue par Erdős et Gallai dans un célèbre article de 1960 :

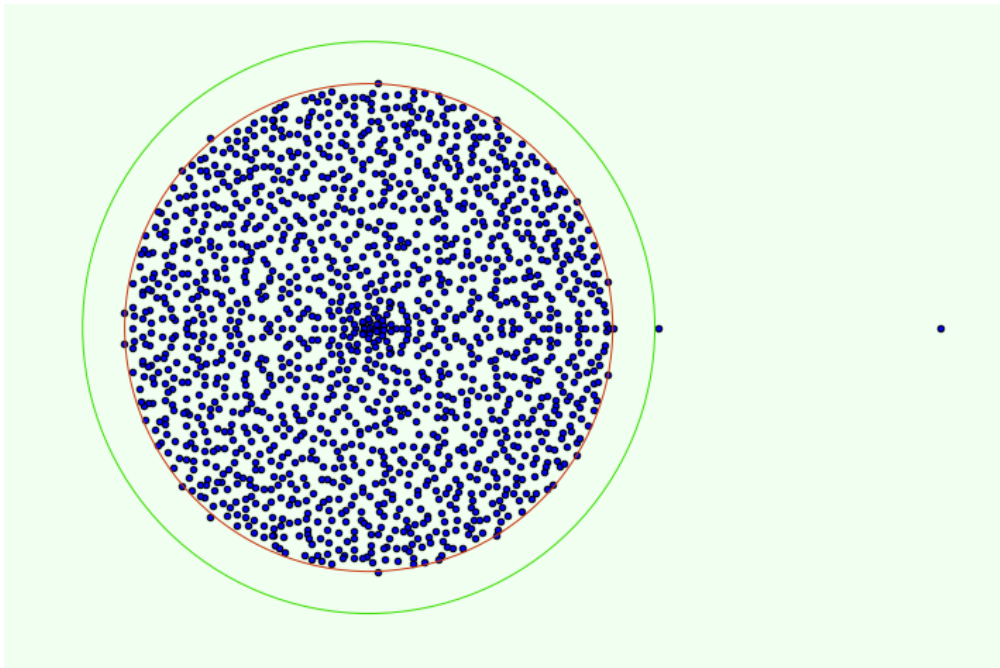
**Théorème 3** (Erdős-Gallai, [73]). Soit  $\mathbf{d} = (d_1, \dots, d_n)$  une suite d'entiers, que l'on supposera triés par ordre décroissant. Alors, il existe un graphe simple  $G$  dont les degrés sont donnés par cette suite si et seulement si les deux conditions suivantes sont réunies :

$$d_1 + \dots + d_n \text{ est pair}, \quad (1.3.1)$$

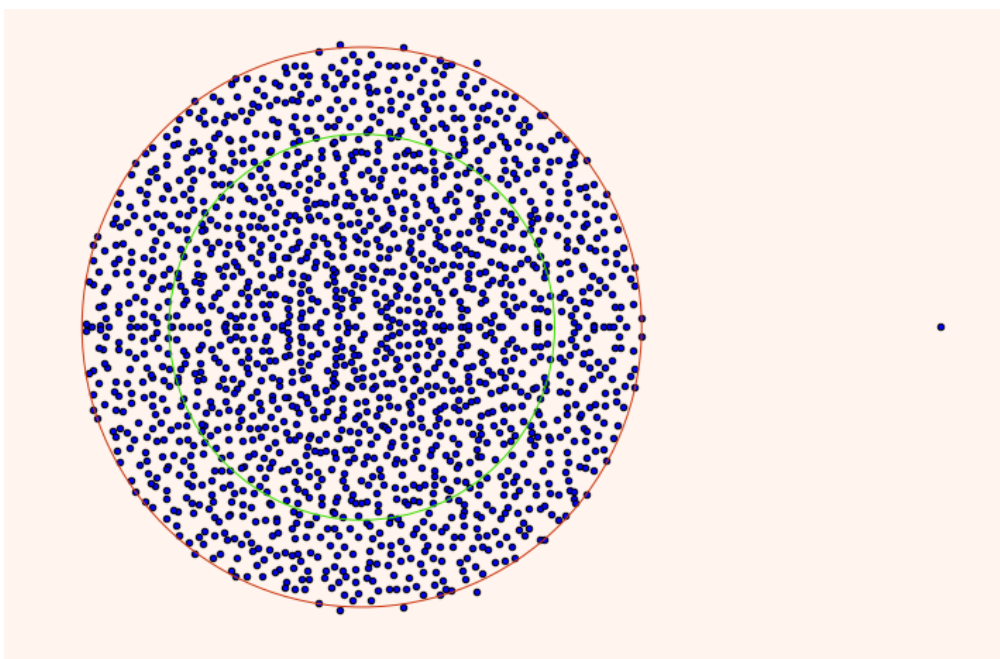
et si pour tout  $k \in \{1, \dots, n\}$  on a

$$d_1 + \dots + d_k \leq k(k-1) + \sum_{i=k+1}^n \min(d_i, k). \quad (1.3.2)$$

Il existe de nombreuses démonstrations de ce théorème, la plus courte étant due à Tripathi et Vijay ([138]) ; il existe aussi d'autres caractérisations des suites graphiques, toutes équivalentes à celle-ci, dont on trouvera une liste dans [131]. Dans le cadre des graphes orientés, il existe également des critères semblables, que l'on trouvera dans la synthèse [28]. Le deuxième résultat présenté dans cette thèse est une généralisation de ces résultats, et une extension de la notion de 'suite des degrés' à des voisinages entiers.



(a) Un exemple du cas où  $\tilde{\rho} = \delta^{-1}$ .



(b) Un exemple du cas où  $\tilde{\rho} = \rho$ .

FIGURE 1.5 – Deux réalisations du spectre de  $P$  pour un graphe  $G$  issu du modèle de configurations dirigées. Le cercle rouge a pour rayon  $\rho$  et le cercle vert a pour rayon  $1/\delta$ . Le point isolé à droite est 1, la valeur propre de Perron-Frobenius.

- Dans la figure (a) il y a  $n = 1600$  sommets, dont 700 de degré (2, 2) et 800 de degré (9, 9), de sorte que dans ce cas  $\tilde{\rho} = \delta^{-1} = 1/2$ .
- Dans (b), on a  $n = 1800$  sommets, dont 600 de degré (5, 6), 600 de degré (3, 7) et 600 de degré (9, 4). Dans ce cas, on a  $\tilde{\rho} = \rho$ .

### La couverture universelle d'un graphe

On peut voir le degré d'un sommet  $x$  comme une caractérisation de son voisinage immédiat : plus précisément, le degré de  $x$  est  $d$  si et seulement si  $x$  possède  $d$  voisins. Or, dans le cadre de la convergence de Benjamini-Schramm, on a besoin d'étudier des voisinages d'une profondeur supérieure à 1 ([35]). Pour cela, il est nécessaire d'introduire la notion de couverture universelle.

Soit  $G = (V, E)$  un graphe fini. Un graphe  $G' = (V', E')$  est un recouvrement<sup>8</sup> de  $G$  s'il existe une surjection  $p : V' \rightarrow V$  qui est un isomorphisme local, au sens où pour tout sommet  $x$ , l'application  $p$  est une bijection entre les arêtes de  $G$  incidentes à  $x$  et les arêtes de  $G'$  incidentes à  $p(x)$ . Il s'agit de l'analogie de la notion de *relèvement* venue de la topologie.

Un graphe connexe possède un relèvement particulier, qui est en un sens canonique : le recouvrement universel. Il s'agit de l'unique recouvrement de  $G$  qui est un arbre ; on le note  $T_G$ . Ce recouvrement possède deux propriétés essentielles :

- i) il est unique à isomorphisme près,
- ii) c'est un recouvrement de tous les recouvrements de  $G$ .

Le recouvrement universel d'un graphe peut être vu comme le graphe  $G$  *déplié*, c'est-à-dire dans lequel chaque cycle est oublié tout en conservant l'adjacence des sommets. Donnons quelques exemples pour éclaircir les idées.

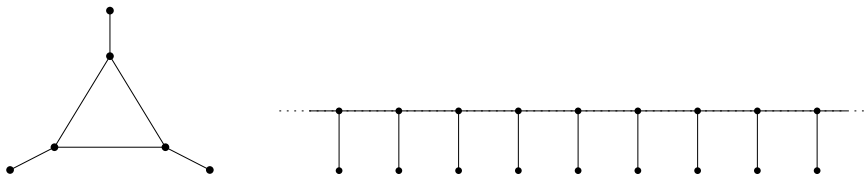


FIGURE 1.6 – Un exemple de recouvrement universel : le peigne infini de droite est le recouvrement universel du graphe de gauche.

- Le recouvrement universel de n'importe quel graphe  $d$ -régulier (avec  $d > 1$ ) est  $\mathbb{T}_d$  ;
- le recouvrement universel d'un cycle est une ligne bi-infinie ;
- le recouvrement universel du graphe à gauche de la figure 1.6 est le « peigne infini » à gauche de 1.6.

Le recouvrement universel d'un graphe porte en lui de nombreuses informations sur ce graphe, car il en donne une bonne approximation locale — d'autant meilleure que le graphe  $G$  ne possède pas beaucoup de petits cycles. En informatique théorique, le recouvrement universel a donné lieu à de nombreuses avancées sur le calcul parallèle ([13], [14]). En probabilités, il est d'une importance capitale pour étudier des processus sur des graphes qui sont localement des arbres ([84], [12]).

### Les $h$ -voisinages

Soit donc  $G$  un graphe connexe et  $T$  son recouvrement universel. Soit  $x$  un sommet de  $G$ . Son  $h$ -voisinage universel est la (classe d'isomorphisme de) la boule  $B_T(y, h)$  où  $y$  est n'importe quel antécédent de  $x$  par  $p$  ; on vérifiera facilement qu'elle ne dépend pas de l'antécédent choisi. Ce  $h$ -voisinage doit être vu comme le (vrai) voisinage de  $x$  dans  $G$ , dans lequel les cycles ont été dépliés. La donnée de  $T$  permet donc de connaître la liste  $(t_1, \dots, t_n)$  des  $h$ -voisinages universels du graphe  $G$ . L'inverse est-il possible ? Autrement dit, si l'on dispose d'une suite  $\mathbf{t} = (t_1, \dots, t_n)$  d'arbres enracinés non étiquetés de

<sup>8</sup>On trouve parfois les termes *relèvement* ou *couverture*.



profondeur  $\leq h$ , comment décider si cette suite correspond effectivement à la suite des  $h$ -voisinages d'un vrai graphe  $G$  ?

Lorsque  $h = 1$ , cette question est parfaitement équivalente au problème d'Erdős-Gallai. En effet, la donnée d'un arbre enraciné de profondeur 1 est équivalente à la donnée du nombre de ses feuilles. Si l'on note  $t(d)$  l'arbre à  $d + 1$  sommets ayant  $d$  feuilles, alors la donnée d'un graphe  $G$  dont les 1-voisinages sont  $(t(d_1), \dots, t(d_n))$  est parfaitement équivalente à la donnée d'un graphe dont la suite des degrés est  $(d_1, \dots, d_n)$ , comme dans la figure 1.7.



FIGURE 1.7 – Cette collection d'arbres de profondeur 1 est-elle la collection des 1-voisinages d'un graphe simple à 8 sommets ? Cela revient à chercher si la suite  $\mathbf{d} = (3, 1, 2, 3, 5, 2, 3, 1)$  est une suite graphique.

La donnée des  $h$ -voisinages universels d'un graphe permet de générer des graphes aléatoires ayant ces voisinages. Par exemple, supposons que  $(t_1, \dots, t_n)$  soit une suite d'arbres. Dans [35], les auteurs créent une variante du modèle de configurations, permettant de construire des multi-graphes ayant cette suite comme suite des  $h$ -voisinages universels ; leur construction est similaire à celle des graphes de configurations, avec des demi-arêtes qui sont recollées ensemble. Nous avons généralisé cette construction, afin de répondre à la question de l'existence de vrais graphes ayant cette suite de voisinages universels.

### Une caractérisation

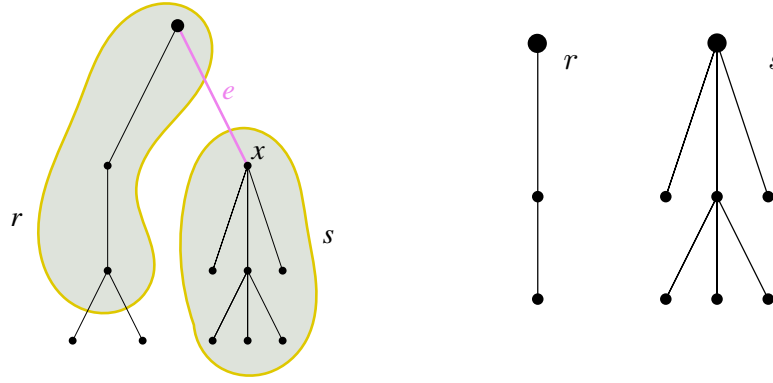
La question qui se pose est donc la suivante. On note  $\mathcal{T}_h$  l'ensemble des arbres enracinés non-étiquetés de profondeur maximale  $h$ .

**QUESTION 1.3.1.** Soit  $\mathbf{t} = (t_1, \dots, t_n)$  un  $n$ -uplet d'éléments de  $\mathcal{T}_h$ . Est-ce le  $n$ -uplet des  $h$ -voisinages universels d'un graphe simple  $G$  ?

Avant de répondre à la question, nous avons besoin de quelques définitions — la chose sera plus claire si l'on se réfère à la figure 1.8.

Soit  $t$  un arbre de racine  $\bullet$  et soit  $e = (\bullet, x)$  une arête adjacente à la racine. La suppression de  $e$  sépare  $t$  en deux composantes connexes, l'une contenant la racine notée  $r'$ , et l'autre ne la contenant pas, notée  $s$ . On supprime de  $r'$  les sommets à profondeur  $h$ . L'arbre obtenu, enraciné en  $\bullet$ , est un élément de  $\mathcal{T}_{h-1}$  que l'on notera  $r$ . L'autre arbre,  $s$ , enraciné en  $x$ , est également un élément de  $\mathcal{T}_{h-1}$ . On dira alors que le *type* de l'arête  $e$  est le couple  $\tau = (r, s)$ , et le type opposé de l'arête  $e$  est défini comme étant  $(s, r) = \tau^{-1}$ . Il est possible de décomposer l'ensemble des types des arêtes de  $t$  adjacentes à la racine en trois ensembles disjoints,  $\Delta, A, B$  :

- $\Delta$  est l'ensemble des types diagonaux, c'est-à-dire de la forme  $(r, r)$  ;
- $A \cup B$  est l'ensemble des types non diagonaux, et les ensembles  $A$  et  $B$  sont choisis de sorte que si le type  $\tau$  est dans  $A$ , alors  $\tau^{-1}$  est dans  $B$ .

FIGURE 1.8 – Construction de  $\tau(e) = (r, s)$  pour l'arête  $e$  dans l'arbre  $t$ .

Si  $\tau$  est le type d'une des arêtes d'un des arbres de  $\mathbf{t}$ , on note  $d_i^\tau$  le nombre d'arêtes  $e$  incidentes à la racine dans  $t_i$  et telles que  $\tau(e) = \tau$ , c'est-à-dire le « degré en  $\tau$  de  $t_i$  ». On pose enfin  $N_\tau = d_1^\tau + \dots + d_n^\tau$ . Le théorème suivant est le principal résultat de [41], travail en collaboration avec Charles Bordenave. Les indices entre parenthèses signifient que la double suite  $(d_i^\tau, d_i^{\tau^{-1}})$  a été ordonnée dans l'ordre lexicographique décroissant.

**Théorème 4.** Soit  $\mathbf{t} = (t_1, \dots, t_n)$  un  $n$ -uplet d'éléments de  $\mathcal{T}_h$ . Il est graphique si et seulement si

(1) pour tout  $\tau \in \Delta$ , le nombre  $N_\tau$  est pair et si pour tout  $k$  on a

$$\sum_{i=1}^k d_{(i)}^\tau \leq k(k-1) + \sum_{i=k+1}^n \min(d_{(i)}^\tau, k) \quad (1.3.3)$$

(2) pour tout  $\tau \in A$ , on a  $N_\tau = N_{\tau^{-1}}$ , et pour tout  $k$  on a

$$\sum_{i=1}^k d_{(i)}^\tau \leq \sum_{i=1}^k \min(d_{(i)}^{\tau^{-1}}, k). \quad (1.3.4)$$

### Problèmes de reconstruction

Le théorème 4 a été démontré dans le cas particulier  $h = 2$  par [19, 16]. Une des motivations des auteurs est le problème général de la reconstruction des graphes : soit  $G$  un graphe possédant une propriété  $\mathcal{P}$ . Est-il possible de reconstruire  $G$  à partir de la simple donnée de  $\mathcal{P}$  ? Plus généralement, est-il possible de déterminer s'il existe un graphe vérifiant  $\mathcal{P}$ , et si oui, un tel graphe est-il unique ?

Lorsque  $\mathcal{P}$  est la propriété d'avoir une suite de degrés donnée, le théorème d'Erdős-Gallai résout la question de l'existence ; la question de l'unicité a été résolue plus tard (voir [16], théorème 4.2 et références).

Si la question 1.3.1 peut être résolue, c'est notamment parce que l'on considère uniquement des voisinages dans le recouvrement universel, c'est-à-dire que l'on oublie les cycles. La même question dans laquelle on remplace ces voisinages universels par les vrais voisinages, où les cycles sont conservés, semble réellement hors de portée. Citons, à titre d'exemple, la célèbre conjecture de la reconstruction de Kelly-Ulam : partant d'un graphe  $G$  à  $n$  sommets, on note  $g_i$  la classe d'isomorphisme du graphe  $G$  privé de son sommet  $i$ . Est-il possible de retrouver  $G$  à partir de la donnée de  $(g_1, \dots, g_n)$  ? Cette question simple d'apparence a été posée par Ulam dans les années 1940 ; elle reste non résolue.

## 1.4 Spectre des arbres

Nous revenons dans cette dernière partie au modèle le plus simple de graphes aléatoires : le graphe d'Erdős-Rényi non-orienté et sa matrice d'adjacence. Comme mentionné plus haut, si  $G_n \sim \text{ER}(n, d/n)$ ,

alors  $G_n$  converge au sens de Benjamini-Schramm vers  $\text{PGW}(d)$  et de plus la mesure spectrale empirique de la matrice d'adjacence  $A_n$  de  $G_n$ , notée  $\mu_{A_n}$ , converge au sens de (1.1.3) vers une mesure limite  $\mu_{\text{PGW}(d)}$ . L'objectif de cette section est d'étudier la nature du spectre (partie continue, atomes) de cette mesure, pour laquelle on énoncera deux résultats nouveaux et quelques questions.

Comme promis, on commence par donner une définition précise de  $\mu_{\text{PGW}(d)}$ , et plus généralement  $\mu_\rho$  lorsque  $\rho$  est la mesure de certains arbres enracinés. Une notion importante pour cela est l'*unimodularité*.

### Unimodularité

Dans la convergence au sens de Benjamini-Schramm, l'enracinement uniforme des graphes finis est une idée profonde et a des conséquences importantes. En prenant une racine uniforme, on étudie les graphes finis du point de vue de leur sommet *typique*, en négligeant d'éventuels sommets peu nombreux au comportement déviant. Si une suite de graphes finis  $(G_n)$  converge vers  $(G, \nu)$ , sa limite devra donc donner une forme d'homogénéité au comportement des sommets : en particulier, il faut s'attendre à ce que la loi du graphe enraciné  $(G, \nu)$  ne change pas par réenracinement du graphe. Ce phénomène est appelé *unimodularité* ([8, 7, 27, 34]).

Introduisons d'abord l'ensemble  $\mathcal{G}_{**}$  des (classes d'isomorphismes de) graphes bi-enracinés, c'est-à-dire des triplets  $(G, \nu, w)$  où  $\nu, w$  sont deux sommets de  $G$ . On le munit de la topologie induite par la distance locale, comme pour  $\mathcal{G}_*$ . On dit qu'une loi de probabilité sur  $\mathcal{G}_*$  est unimodulaire si pour toute fonction  $f : \mathcal{G}_{**} \rightarrow \mathbb{R}$  mesurable,

$$\mathbf{E}_\rho \left[ \sum_{w \in V} f(G, \nu, w) \right] = \mathbf{E}_\rho \left[ \sum_{w \in V} f(G, w, \nu) \right], \quad (1.4.1)$$

où la notation  $\mathbf{E}_\rho$  signifie que la variable aléatoire  $(G, \nu)$  a pour loi  $\rho$  sous  $\mathbf{E}_\rho$ . Lorsque  $G = (V, E)$  est un graphe fini, la variable aléatoire  $(G(\nu), \nu)$  avec  $\nu$  uniforme sur  $V$ , est unimodulaire, et les limites de Benjamini-Schramm de lois unimodulaires sont encore unimodulaires, par passage à la limite dans (1.4.1). Ainsi, toute loi  $\rho$  sur  $\mathcal{G}_*$  qui est limite au sens de Benjamini-Schramm d'une suite de graphes finis<sup>9</sup> est unimodulaire. L'inverse est une question non résolue à ce jour (voir [7, 27] pour des développements sur cette question et ses conséquences).

Un **arbre de Galton-Watson unimodulaire** de loi de reproduction  $\pi = (\pi_n)_{n \geq 0}$  est un arbre aléatoire enraciné, dans lequel le nombre d'enfants de la racine a pour loi  $\pi$ , et le nombre d'enfants de tout autre sommet a pour loi  $\hat{\pi}$ , qui est la loi  $\pi$  biaisée par sa taille :

$$\hat{\pi}_n := \frac{(n+1)\pi_{n+1}}{\sum_{k=0}^{\infty} k\pi_k}.$$

Les seules lois telle que  $\hat{\pi} = \pi$  sont les lois de Poisson, et donnent naissance au prototype classique des arbres de Galton-Watson, les arbres de Galton-Watson poissonniens  $\text{PGW}(d)$ , qui apparaissent naturellement comme limites des graphes d'Erdős-Rényi  $\text{ER}(n, d/n)$ .

Lorsque  $\pi = \delta_d$ , on a  $\hat{\pi} = \delta_{d-1}$  et dans ce cas, l'arbre  $\text{UGW}(\delta_d)$  n'est autre que l'arbre infini  $d$ -régulier  $\mathbb{T}_d$ , limite des graphes uniformes  $d$ -réguliers.

### Mesure spectrale des arbres unimodulaires

Soit  $G = (V, E)$  un graphe localement fini sur un ensemble de sommets  $V$  que l'on prendra toujours dénombrable. Son opérateur d'adjacence  $A$  est l'opérateur linéaire sur l'espace de Hilbert

$$\mathcal{H} = \ell^2(V) = \left\{ f : V \rightarrow \mathbb{C}, \sum_{v \in V} |f(v)|^2 < \infty \right\}$$

<sup>9</sup>De telles lois sont dites *sophiques*.

défini par

$$\langle e_x, A e_y \rangle = \mathbf{1}_{(x,y) \in E} \quad (1.4.2)$$

où  $(e_x : x \in V)$  est la base canonique de  $\mathcal{H}$ . Le domaine  $D_A$  de  $A$  est l'ensemble dense des vecteurs à support fini :  $D_A = \text{vect}(e_x : x \in V)$ . Lorsque le graphe  $G$  est fini,  $A$  est une matrice symétrique, et le spectre ne pose aucun problème de définitions. En revanche, lorsque  $G$  est infini, des difficultés apparaissent. Si les degrés de  $G$  sont uniformément bornés par un nombre même  $d$ , alors  $A$  est un opérateur borné et auto-adjoint, et peut donc s'étendre en un opérateur borné et symétrique sur  $\mathcal{H}$  tout entier : il n'y a pas de difficultés à définir son spectre.

Cependant, beaucoup des graphes que nous étudions (en particulier les arbres de Galton-Watson poissoniens) n'ont pas des degrés bornés ; signalons également qu'il existe des arbres qui ne sont pas auto-adjoints ([118]).

Fort heureusement, les mesures unimodulaires sont concentrées sur les arbres qui sont essentiellement auto-adjoints ([34, Prop. 2.2]), pour lesquels l'analyse fonctionnelle classique permet de définir efficacement le spectre. Plus précisément, le théorème spectral montre que pour tout graphe enraciné  $(G, o)$  dont l'opérateur d'adjacence est essentiellement auto-adjoint, il existe une unique mesure borélienne de probabilité  $\mu_{(G,o)}$  sur  $\mathbb{R}$  telle que

$$\forall z \in \mathbb{C} \setminus \mathbb{R} \quad \langle e_o, (A - z)^{-1} e_o \rangle = \int_{\mathbb{R}} \frac{1}{z - t} d\mu_{(G,o)}(t). \quad (1.4.3)$$

Le membre de droite,  $s(z)$ , est la transformée de Stieltjes de la mesure  $\mu_{(G,o)}$  et elle caractérise cette dernière. Il est donc possible de retrouver  $\mu_{(G,o)}$  par la formule d'inversion de Stieltjes ; même lorsque ce calcul inverse est impossible, on peut tout de même obtenir des informations supplémentaires. Par exemple, les atomes sont donnés par

$$\mu_{(G,o)}(\{\lambda\}) = \lim_{t \rightarrow 0} \frac{1}{\pi} \Im t s(\lambda + it).$$

Si  $(G, o)$  est un arbre enraciné aléatoire de loi  $\rho$  unimodulaire, la mesure  $\mu_\rho$  est alors définie par

$$\mu_\rho = \mathbf{E}_\rho[\mu_{(G,o)}]. \quad (1.4.4)$$

On renvoie à [54] pour vérifier que tous les objets se comportent bien, notamment vis-à-vis de la mesurabilité.

Nous avons déjà vu que  $\text{UGW}(\delta_d) = \delta_{\mathbb{T}_d}$  : dans ce cas,  $(G, o)$  n'est pas aléatoire, il est toujours égal à l'arbre infini  $d$ -régulier  $\mathbb{T}_d$ . Le membre de gauche dans (1.4.3) peut alors se développer en série de Laurent dont les coefficients sont les nombres de marches fermées à la racine de  $\mathbb{T}_d$ , lesquels sont calculables explicitement par des arguments combinatoires ([113]). Cela donne une expression explicite pour la transformée de Stieltjes de  $\mu_\rho$ , et le calcul de la transformée inverse permet de voir que la mesure  $\mu_\rho$  est précisément la loi de Kesten-McKay déjà évoquée précédemment.

On trouvera dans [29] d'autres exemples de limites de spectres d'arbres, comme l'arbre récursif.

### La mesure limite des graphes d'Erdős-Rényi dilués

On se donne un nombre réel  $d > 0$ , et on va étudier le spectre d'un graphe  $G$  de distribution  $\text{ER}(n, d/n)$ . On a vu dans la section 1.1 que  $G_n$  converge au sens de Benjamini-Schramm vers  $\text{PGW}(d)$ . En vertu de la continuité du spectre (1.1.3), la loi empirique des valeurs propres de la matrice d'adjacence  $A_n$  de  $G_n$ ,

$$\mu_{A_n} := \frac{1}{n} \sum_{i=1}^n \delta_{\lambda_i},$$

converge vers une mesure de probabilité  $\mu_{\text{PGW}(d)}$  que l'on notera simplement  $\mu_d$ , mesure spectrale moyenne de l'arbre de Galton-Watson comme définie en (1.4.4). Cette convergence peut être établie

sans l'aide de la convergence de Benjamini-Schramm ([143, 94]), mais c'est cette dernière qui permet de disposer de la représentation de  $\mu_{\text{PGW}(d)}$  comme en (1.4.3)-(1.4.4). Si l'on note  $F_n, F_{\text{PGW}(d)}$  les fonctions de répartition de  $\mu_{A_n}, \mu_d$ , alors

$$\sup_{t \in \mathbb{R}} |F_n(t) - F_{\text{PGW}(d)}(t)| \rightarrow 0. \quad (1.4.5)$$

Comme déjà observé, cette convergence implique la convergence des atomes, au sens où pour tout nombre réel  $\lambda$ , on a  $\lim \mu_{G_n}(\{\lambda\}) = \mu_d(\{\lambda\})$ . En particulier, comme

$$\mu_{G_n}(\{0\}) = \dim \ker(A_n)/n,$$

on connaît le comportement asymptotique du noyau de  $A_n$ , à condition d'être capable de calculer  $\mu_{\text{PGW}(d)}(\{0\})$ .

Les physiciens Bauer et Golinelli, en s'intéressant au noyau de  $G$ , ont formulé plusieurs conjectures frappantes sur le comportement de  $\mu_d$  autour de zéro ([23, 21]). Grâce à la méthode des répliques symétriques, ils ont pu formuler la valeur exacte de l'atome en zéro, à savoir

$$\mu_{\text{PGW}(d)}(\{0\}) = e^{-dx} + dx e^{-dx} + x - 1 \quad (1.4.6)$$

où  $x$  est la plus petite solution dans  $]0, 1[$  de l'équation  $x = e^{-de^{-dx}}$ . Cette formule fut rigoureusement démontrée plus tard dans [44], notamment à l'aide de formules de récursion sur la transformée de Stieltjes de  $\mu_{\text{PGW}(d)}$ .

Bauer et Golinelli ont finement remarqué que le membre de droite dans (1.4.6) est analytique en  $d$  lorsque  $d < e$ , et possède une singularité en  $d = e$ ; ils ont interprété ce phénomène comme une transition dans la nature de la mesure spectrale, remarquant sur leurs simulations numériques l'apparition d'une partie continue non nulle autour de zéro lorsque  $d > e$ .

### Décomposition spectrale et modèle d'Anderson

Toute mesure borélienne  $\mu$  sur  $\mathbb{R}$  se décompose en

$$\mu = \mu_{\text{ac}} + \mu_{\text{pp}} + \mu_{\text{sc}} \quad (1.4.7)$$

où  $\mu_{\text{ac}}$  est absolument continue par rapport à la mesure de Lebesgue,  $\mu_{\text{pp}}$  est purement atomique et  $\mu_{\text{sc}}$  est singulière par rapport à la mesure de Lebesgue, mais n'a aucun atome.

Cette décomposition est centrale dans la théorie physique des opérateurs : si  $H$  est un opérateur auto-adjoint sur un espace de Hilbert, cette décomposition porte en elle des propriétés essentielles de  $H$  liées au comportement de la dynamique quantique  $t \mapsto e^{-itH} \phi$ , notamment via le théorème RAGE (voir ([54, 135] ou plus spécifiquement [101]).

Les physiciens ont porté une attention considérable à ces propriétés spectrales dans le modèle d'Anderson, pour lequel l'hamiltonien est donné  $H = A + \lambda V$  avec  $A$  est un opérateur laplacien et  $V$  un opérateur diagonal ; lorsque  $A$  est l'opérateur d'adjacence sur  $\mathbb{Z}^d$ , on a le modèle d'Anderson discret classique. L'existence d'une partie ponctuelle dans le spectre d'un tel opérateur est liée au phénomène de localisation, et c'est la raison pour laquelle la décomposition (1.4.7) est tant étudiée, y compris lorsque  $A$  est l'opérateur d'adjacence d'un arbre ([4], [95], [2]). Dans le langage des physiciens, on dit qu'une mesure  $\mu$  ne possède pas *d'états étendus* en un réel  $\lambda$  lorsque

$$\lim_{\varepsilon \rightarrow 0} \frac{\mu(] \lambda - \varepsilon, \lambda + \varepsilon[) - \mu(\{\lambda\})}{2\varepsilon} = 0. \quad (1.4.8)$$

Dans le cas contraire, on dit qu'elle possède des états étendus en  $\lambda$  ; par exemple, si  $\mu = f(x)dx$  avec  $f$  mesurable, la limite ci-dessus est égale à  $f(\lambda)$  pour presque tout  $\lambda$ .

On a déjà vu que la mesure spectrale de  $\mathbb{T}_d$  était absolument continue, avec pour mesure la loi de Kesten-McKay ; il existe cependant des arbres infinis dont la mesure spectrale est

- purement atomique : l'arbre canopée ([5, Proposition A.2]) en particulier, sur lequel on reviendra plus loin ;
- ou au contraire, purement singulière continue ; on trouvera dans [49, 48] pour une construction explicite de tels arbres à partir de  $\mathbb{T}_d$  et dans [5, section 6] des exemples d'arbres semi-infinis. Dans le cadre du modèle d'Anderson, on renvoie aussi à [133, 132].

Lorsque l'arbre en lui-même est aléatoire (Galton-Watson par exemple), la question est plus délicate. Une condition nécessaire et suffisante pour que  $\mu_{\text{PGW}(d)}$  ait une partie continue, et donc des états étendus, est simplement  $d > 1$ , et plus généralement ([46]) la mesure  $\mu_{\text{UGW}(\pi)}$  possède une partie continue si et seulement si  $\pi$  est sur-critique. La raison pour cela est le fait qu'avec probabilité 1, des sous-graphes isomorphes à  $\mathbb{Z}$  apparaissent dans  $\text{UGW}(\pi)$ . Cependant, ce résultat ne donne aucune information sur le support de la partie continue lorsqu'elle existe, et encore moins sur sa forme.

### La transition de Bauer-Golinelli

Les simulations numériques précises de Bauer et Golinelli [23, Section 5.1] les ont conduit à supposer que la partie continue de  $\text{PGW}(d)$  vérifie une transition en  $d = e$  :

**CONJECTURE 1.** *La mesure  $\mu_{\text{PGW}(d)}$  possède des états étendus en 0 si et seulement si  $d > e$ .*

On pourra constater ce phénomène sur les figures 1.9a-1.9b-1.9c.

Une telle transition de phase n'est pas en soi étonnante ; on sait déjà que si  $d$  est suffisamment petit (inférieur à 1), la mesure  $\mu_{\text{PGW}(d)}$  est purement atomique et qu'il n'y a donc pas d'états étendus en zéro. Or, lorsque  $d \rightarrow \infty$ , les mesures  $\mu_{\text{PGW}(d)}$  convergent vers la célèbre loi du semi-cercle de Wigner ([89, 137], voir figure 1.10), qui est absolument continue par rapport à la mesure de Lebesgue et dont la densité  $\rho(t) = \mathbf{1}_{|t| \leq 2} \sqrt{4 - t^2}$  est non nulle sur tout son support. Cela ne permet pas de prouver rigoureusement l'apparition d'états étendus en zéro, mais cela donne tout de même une intuition sur le résultat. Ce qui est étonnant est plutôt la *localisation exacte* de cette transition au point  $e \approx 2,718$ .

Dans la prépublication [63] en collaboration avec Justin Salez, nous avons démontré la conjecture 1. Ce résultat n'est pas propre à la mesure spectrale des arbres de Galton-Watson poissoniens, mais il s'énonce dans le cadre général des arbres de Galton-Watson unimodulaires  $\text{UGW}(\pi)$  (avec  $\pi$  sur-critique). Pour formuler les résultats, on a besoin de la fonction génératrice de  $\pi$ , que l'on notera

$$\varphi(z) := \sum_{n=0}^{\infty} \pi_n z^n.$$

De façon similaire, on note  $\hat{\varphi}$  la fonction génératrice de  $\hat{\pi}$ . Posons  $M(t) = \varphi(t) + (1-t)\varphi'(t) + \varphi(1 - \hat{\varphi}(t)) - 1$ , pour  $t \in [0, 1]$ . Un élément clé dans la démonstration de (1.4.6) dans [44] était l'identité  $\mu_{\text{PGW}(d)}(\{0\}) = \max M$  et le fait qu'il existe un unique nombre  $z_*$  dans  $]0, 1[$  tel que  $z_* = 1 - \hat{\varphi}(z_*)$ . Les deux hypothèses sur  $M$  qui interviennent dans notre théorème principal sont les suivantes :

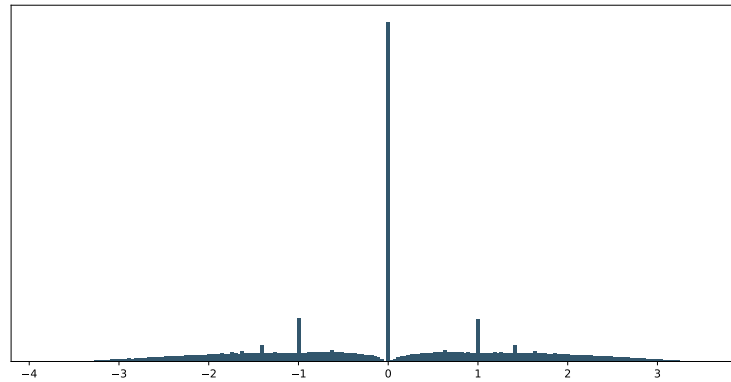
- (1) La fonction  $M$  possède un unique maximum en  $z_*$ .
- (2)  $M''(z_*) \neq 0$ .

Les conditions d'apparition d'états étendus en zéro sont décrites par le théorème suivant.

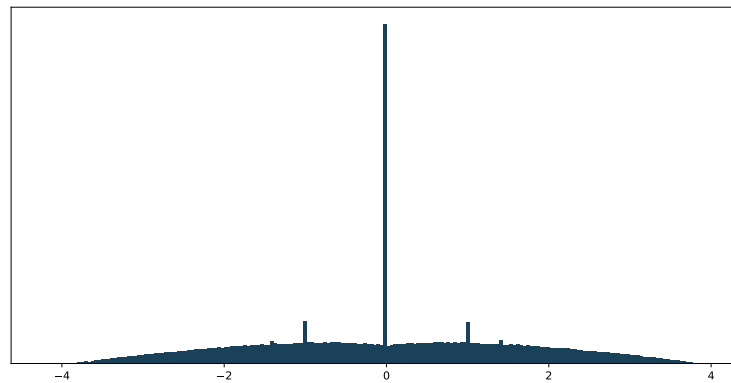
**Théorème 5.** *Soit  $\pi$  une mesure sur  $\mathbb{N}$ , avec  $\pi_0 < 1$ .*

- Si  $M$  ne vérifie pas la condition 1, alors  $\mu_{\text{UGW}(\pi)}$  a des états étendus en zéro.
- Si  $M$  vérifie 1 et 2, alors  $\mu_{\text{UGW}(\pi)}$  n'a pas d'états étendus en zéro.

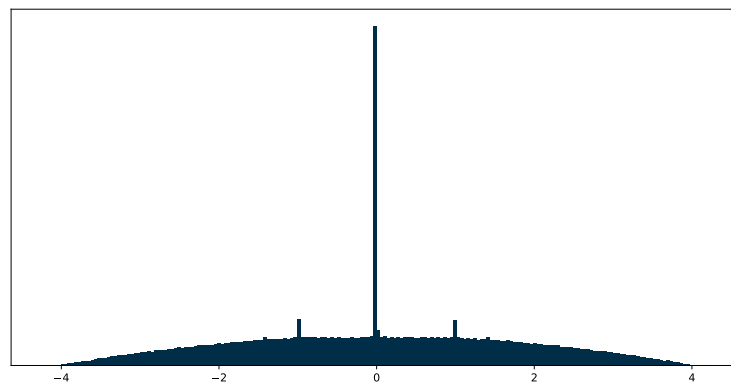
L'existence d'une partie continue ou d'états étendus en zéro est donc réglée, au moins dans le cas des arbres de Galton-Watson poissoniens. La question de l'existence de tels résultats à d'autres arbres unimodulaires se pose.



(a)  $d = 2$



(b)  $d = e$



(c)  $d = 3$

FIGURE 1.9 – Histogrammes des valeurs propres de  $G \sim \text{ER}(n, d/n)$  avec  $n = 1000$  (sur 100 échantillons). La partie ‘continue’ est nulle en zéro lorsque  $d < e$ , conformément à la prédiction de Bauer-Golinelli.

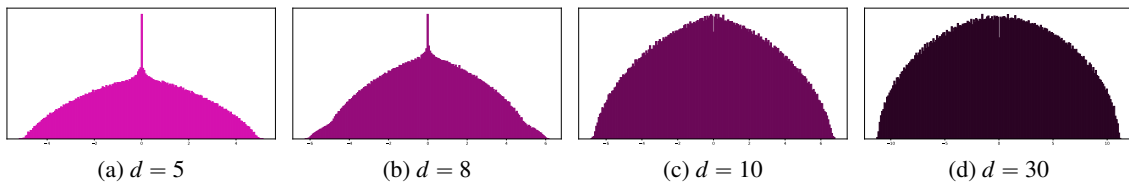


FIGURE 1.10 – Histogrammes des valeurs propres de  $G \sim \text{ER}(n, d/n)$  avec  $n = 1000$  (100 échantillons). Lorsque  $d \rightarrow +\infty$  on a  $\mu_d \rightarrow \mu_{\text{Wigner}}$ .

### La nature du spectre dans les arbres semi-infinis

La démonstration de l'existence d'une partie continue dans le spectre des arbres de Galton-Watson ([46]) reposait sur le fait que, lorsque la loi de reproduction  $\pi$  est sur-critique, l'arbre contient un bi-rayon<sup>10</sup>, c'est-à-dire un sous-graphe isomorphe à  $\mathbb{Z}$ . Ce n'est pas le cas pour d'autres arbres unimodulaires comme l'arbre squelette  $T_{\text{squel}}$ , qui sont infinis presque sûrement, mais qui ne contiennent qu'un rayon, c'est-à-dire un sous-graphe isomorphe à  $\mathbb{N}$  : de tels arbres sont appelés *arbres semi-infinis* (ou *single-infinite trees* par Aldous ([6, Section 4])). On peut identifier n'importe quel arbre semi-infini avec une suite  $(T_n)$  d'arbres enracinés finis, la racine de  $t_n$  étant reliée à celle de  $t_{n+1}$ , comme dans la figure 1.1b.

On peut vérifier facilement que la mesure spectrale de  $\mathbb{N}$  est précisément la loi du semi-cercle ; on pourrait donc s'attendre à ce que la mesure spectrale des arbres semi-infinis possède toujours une partie continue. Ce n'est pas le cas, comme le montre l'exemple de l'arbre canopée  $(\mathbb{C}_d, \circ)$ ,  $d > 3$  entier.

**LEMMA 1.4.1** (Aizenman, Warzel, [5]). *La mesure spectrale  $\mathbf{E}\mu_{(\mathbb{C}_d, \circ)}$  est purement atomique. Ses atomes sont toutes les valeurs propres des arbres finis  $\mathbb{T}_{n,d} := (\mathbb{T}_d, \circ)_n$ .*

Les arbres semi-infinis peuvent être vus comme des généralisations des matrices de Jacobi. Dans le modèle d'Anderson sur  $\mathbb{N}$ , chaque sommet  $n$  de  $\mathbb{N}$  est doté d'un poids  $V(n)$ , et on dispose de nombreux critères pour relier la nature du spectre de  $H = A + V$  avec les propriétés d'ergodicité de la suite  $(V(n))$  ([54], [102]). Dans le cas des arbres semi-infinis, on peut donc se poser la question suivante :

**QUESTION 1.4.2.** *Quelle relation y a-t-il entre la nature du spectre des arbres semi-infinis et les propriétés de croissance ou d'ergodicité de leur suite  $(T_n)$  ?*

### L'arbre squelette

L'arbre squelette, que l'on notera  $T_{\text{squel}}$ , est la limite au sens de Benjamini-Schramm des arbres uniformes. Plus précisément, si  $T_n$  est uniformément distribué parmi les  $n^{n-2}$  arbres étiquetés à  $n$  sommets, alors  $T_n \rightarrow T_{\text{squel}}$ , une convergence établie dans [85].

<sup>10</sup>*Invariant line ensemble*, dans la terminologie de [46].



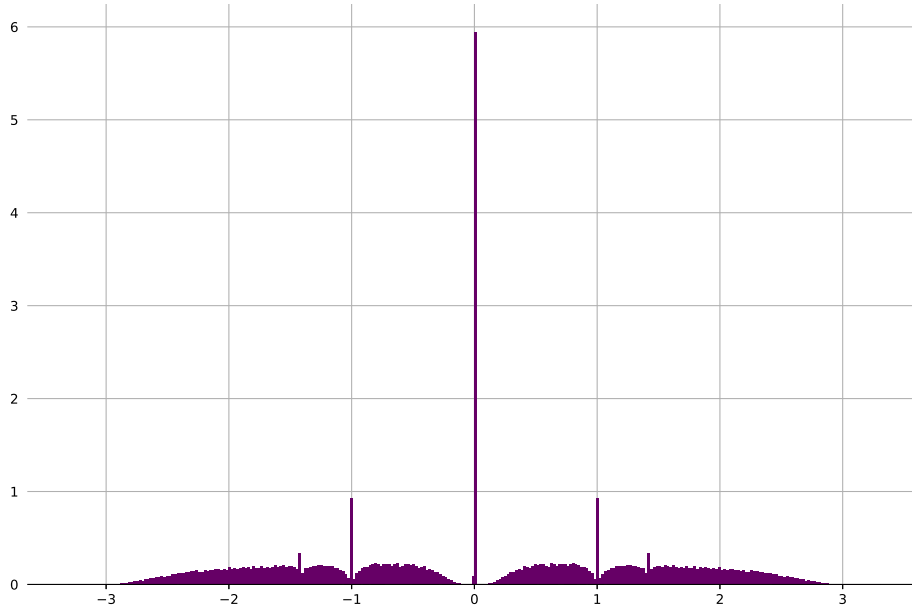


FIGURE 1.11 – Histogramme des valeurs propres du spectre d’un arbre aléatoire sur 1000 sommets, qui donne une idée approchée de l’allure de  $\mu_{\text{squel}}$ .

En vertu de la continuité du spectre pour la convergence locale faible, on a bien la convergence de la mesure spectrale empirique  $\mu_{T_n}$  vers la mesure  $\mu_{\text{squel}}$ . Cette mesure est aussi mal connue que  $\mu_{\text{UGW}(\pi)}$  ; en fait, les seules informations intéressantes sur  $\mu_{\text{squel}}$  sont obtenues en passant à la limite dans  $\mu_{T_n}$ . Donnons deux exemples.

**Atomes.** Par des arguments élémentaires, il est possible de voir que *toutes les valeurs propres d’arbres finis sont des atomes de  $\mu_{\text{squel}}$*  ; cet ensemble est une partie dense de  $\mathbb{R}$  et il est égal à l’ensemble  $\mathbb{A}$  des nombres algébriques totalement réels<sup>11</sup>, un résultat de [127]. Dans les simulations (voir figure 1.11), on voit la présence des atomes en 0,  $\pm 1$  (le spectre de l’arbre avec une seule arête),  $\pm\sqrt{2}$  (le spectre de l’étoile à deux branches).

La valeur précise de  $\mu_{\text{squel}}(\{0\})$  est connue ; ce calcul résulte d’une analyse purement combinatoire de  $T_n$ . Le lien entre la dimension du noyau d’un arbre  $T$  et les feuilles de cet arbre est connu depuis longtemps : pour tout arbre fini  $T$ , on a  $\dim \ker T = \min\{k : M_{n-2k} \neq 0\}$  où  $M_j$  est le nombre de couplages parfaits dans  $T$  qui contiennent  $j$  arêtes. Bauer et Golinelli ont exploité ce lien pour obtenir une formule *exacte* de l’espérance du noyau de  $T_n$ . Plus précisément, ils démontrent dans [21] que

$$\mathbf{E}[\dim \ker(T_n)] = n \left( 1 - 2 \sum_{m=2}^n \frac{(-1)^m}{m} \left(\frac{m}{n}\right)^m \binom{n}{m} \right). \quad (1.4.9)$$

En utilisant des outils de combinatoire analytique, ils ont également calculé la fonction génératrice de  $\dim \ker(T_n)$  et ont obtenu l’asymptotique de  $\mathbf{E}[\dim \ker(T_n)]$  sous la forme :

$$\lim_{n \rightarrow \infty} \mu_{T_n}(\{0\}) = \mu_{\text{squel}}(\{0\}) = 2x_\star - 1$$

<sup>11</sup>Un nombre est *algébrique totalement réel* s’il est racine d’un polynôme à coefficients entiers, dont toutes les racines sont réelles.

où  $x_* \approx 0,56714\dots$  est l'unique solution de l'équation  $x = e^{-x}$ . Cela réglait donc le problème du calcul de l'atome en zéro pour l'arbre squelette, comme (1.4.6) réglait ce problème pour les arbres de Galton-Watson poissonniens. Les simulations indiquent l'absence d'états étendus en zéro pour  $\mu_{\text{squel}}$  (voir figure 1.11. Dans une courte section du chapitre 3, non publiée, j'ai démontré ce résultat :

**Théorème 6.** *La mesure spectrale  $\mu_{\text{squel}}$  ne possède pas d'états étendus en zéro.*

Les arguments sont une adaptation des techniques déjà utilisées pour démontrer la conjecture 1. Il reste à résoudre la question de l'existence d'une partie continue ; les simulations semblent indiquer que  $\mu_{\text{squel}}$  possède effectivement une partie continue.

## 1.5 Valeurs propres de matrices aléatoires diluées

L'objet de cette section est l'étude de certains graphes pondérés, c'est-à-dire de graphes dont les arêtes sont munies de poids.

Formellement, on se donne une matrice de poids  $W \in \mathcal{M}_{n,n}(\mathbb{R})$ , et la matrice d'adjacence  $M$  d'un graphe quelconque, et on veut étudier la matrice  $P = M \odot W$  où  $\odot$  est le produit de Hadamard terme à terme :  $P_{i,j} = M_{i,j}W_{i,j}$ . De tels modèles ont été très étudiés lorsque le graphe sous-jacent est relativement peu dense (typiquement lorsque le nombre d'arêtes de  $G$  est d'ordre  $n \ln(n)$  — on verra la raison pour laquelle ce cadre est plus étudié) ou encore lorsque tout le problème est hermitien, c'est-à-dire lorsque  $W$  est hermitienne et le graphe sous-jacent  $G$  est un graphe simple non dirigé ; un très récent article de Tikhomirov et Youssef ([136]) décrit entièrement le comportement des grandes valeurs propres de  $P$  lorsque  $W$  est issue du GOE et  $G$  est un graphe d'Erdős-Rényi de paramètre  $d/n$  avec  $d \rightarrow \infty$ .

Les auteurs posent notamment les questions suivantes, que nous reprenons *verbatim* de [136, page 3] :

- 1) Y a-t-il un seuil (en fonction de  $d$ ) pour l'apparition de valeurs propres détachées du spectre de  $W$  (outliers) ?
- 2) Y a-t-il une expression explicite pour ce seuil ?
- 3) A-t-on une explication pour l'apparition de ces outliers ?
- 4) Quelle est l'asymptotique exacte de ces outliers ?

Dans le chapitre IV de cette thèse, fondé sur une collaboration avec Charles Bordenave et Raj Rao Nadakuditi, nous répondons à ces quatre questions dans le cadre difficile des graphes **dilués dirigés**, la principale motivation pratique venant du problème statistique de la complétion des matrices, que nous exposerons ci-dessous. Le résultat principal (Théorème 7 en page 28) semble étonnant, parce qu'il dit en substance que les valeurs propres de matrices non-symétriques sont en réalité plus intéressantes que les valeurs singulières des matrices symétriques — la sagesse populaire<sup>12</sup> voulant plutôt que les valeurs propres de matrices non-symétriques soient très instables. Nous reviendrons dans la dernière section sur cette idée récente de la recherche sur les matrices aléatoires.

### Reconstruction

Les problèmes de reconstruction ont pour objectif de reconstruire un objet en général complexe (un graphe, une matrice, un tenseur, une fonction périodique) à partir d'une petite quantité d'informations sur cet objet, typiquement quelques entrées de la matrice, quelques coefficients de Fourier. Ces problèmes sont devenus extrêmement populaires en mathématiques appliquées notamment via l'essor des systèmes de recommandation.

<sup>12</sup>... dans le monde des mathématiques numériques.

Nous nous intéresserons plus particulièrement aux problèmes de reconstruction des matrices. Le problème est le suivant : une matrice  $W$  est cachée. On sait qu'elle possède certaines propriétés structurales (sa taille  $n \times m$  est connue, parfois des informations sur son rang ou ses entrées). Cependant, l'observateur n'a accès qu'à l'observation de certaines entrées  $W_{i,j}$  pour un petit ensemble d'indices  $(i, j) \in E \subset [n] \times [m]$  et cherche à utiliser au mieux cette information pour retrouver (partiellement, avec une faible erreur, ou même exactement) cette matrice  $W$ .

La littérature sur le sujet est gigantesque et on renvoie à [65] et ses références pour une synthèse assez complète ; citons tout de même les articles de Candès, Tao et Candès, Recht [53, 52], l'article de Keshavan, Montanari et Oh [92] et celui de Chatterjee [55].

### Obstructions au problème et hypothèses nécessaires

Il n'est pas toujours possible de reconstruire une matrice  $W$  à partir d'un faible nombre d'observation de ses entrées. Les deux obstacles principaux sont le nombre trop faible de données, et la complexité intrinsèque de  $W$ .

**Manque de données.** Prenons un exemple de reconstruction très simple, dans lequel la matrice à retrouver est de rang 1 et de norme 1, c'est-à-dire de la forme  $W = xy^*$  avec  $x, y \in \mathbb{R}^n$  deux vecteurs unitaires. Retrouver  $W$  revient à retrouver  $x$  et  $y$ , à un signe ou une phase près ; or, supposons que la ligne  $i$  de  $W$  ne soit pas du tout observée. Il n'y alors aucune manière d'accéder à  $x_i$ . Autrement dit, il est nécessaire que l'observateur dispose au moins d'une observation par ligne et d'une observation par colonne s'il veut espérer reconstruire sa matrice originale  $W$  aussi précisément que possible. Supposons que les entrées révélées le soient au hasard, uniformément parmi toutes les entrées possibles de la matrice ; par le principe du collectionneur, pour avoir au moins une entrée révélée par ligne, il est nécessaire de disposer d'au moins  $\ln n$  observations sur chaque ligne et donc  $n \ln n$  observations en tout, autrement dit toute densité  $d \ll \ln n$  laissera nécessairement des lignes ou des colonnes entières inobservées.

La littérature sur la complétion de matrices s'est ainsi concentrée sur ce régime  $d \geq \ln n$ . L'un des apports de ce travail est de démontrer que tout n'est pas perdu lorsque  $d \leq \ln n$  ou même lorsque  $d$  est fixé, et qu'au contraire il est possible d'obtenir un estimateur de  $W$  bien corrélé avec  $W$ .

**Difficulté intrinsèque du problème.** Toutes les matrices ne peuvent pas être facilement reconstruites à partir d'une observation aléatoire de leurs entrées. Par exemple, supposons que la matrice  $W$  ait toutes ses entrées non nulles, sauf éventuellement une ou deux, comme dans l'exemple

$$W = e_1 e_1^* + e_1 e_2^* = \begin{pmatrix} 1 & 1 & 0 & \dots & 0 & 0 \\ 0 & 0 & 0 & \dots & 0 & 0 \\ \vdots & & & & & \vdots \\ 0 & & \dots & & & 0 \end{pmatrix}.$$

Si les entrées révélées sont peu nombreuses et si leur localisation est suffisamment aléatoire, il n'y a aucune chance pour que les deux seules entrées non nulles de la matrice ci-dessus soient observées. On ne verra que des zéros. Il faut donc que les entrées de  $W$  soient suffisamment étalées dans la matrice. Mathématiquement, cela revient par exemple à demander à ce que la plus grande entrée de  $W$  soit d'ordre  $O(1/n)$ , de sorte que la masse  $L^2$  (norme de Frobenius) soit équitablement répartie entre les entrées de la matrice. Ceci est parfaitement équivalent au fait de demander à ce que la masse des vecteurs singuliers soit suffisamment répartie entre les entrées, une condition fréquemment appelée *condition d'incohérence*.

Dans toute la suite, nous adopterons donc des hypothèses sur  $W$  qui rendent au moins le problème accessible. On va pour cela définir une large classe de matrices dont les paramètres de complexité

sont contrôlés. Signalons que ces hypothèses sont standard dans la littérature ; voir par exemple les conditions A0-A1 dans [52] ou encore A1-A2 dans [92]. On renvoie à la synthèse [65] sur ce sujet.

On commencera par s'intéresser exclusivement à des matrices hermitiennes positives ; ceci n'est une restriction qu'en apparence, nous y reviendrons plus tard. Notons

$$W = \sum_{i=1}^r \mu_i \varphi_i \varphi_i^* \quad (1.5.1)$$

où  $\mu_1 \geq \dots \geq \mu_r$  sont les valeurs propres classées dans l'ordre décroissant, l'entier  $r$  est le rang, et les  $\varphi_i$  sont des vecteurs propres unitaires orthonormaux.

(i) Le rang de  $W$  est inférieur à  $r$ .

(ii) Les valeurs propres de  $W$  sont positives.

(iii) Il existe un réel  $b \geq 1$  tel que

$$\max_{i \in [r]} |\varphi_i|_\infty = \frac{b}{\sqrt{n}}. \quad (1.5.2)$$

Enfin, conformément au thème général de cette thèse, nos résultats sont valables à  $d$  fixé, sans aucune restriction ; il s'agit donc des premiers et seuls résultats sur les valeurs propres de matrices *réellement* diluées. Toutes nos démonstrations s'étendent au cas où  $d$  croît lentement avec  $n$ , typiquement  $d = n^{o(1)}$ .

### Méthodes spectrales

En règle générale, la matrice observée  $P$  est de rang élevé avec grande probabilité ; cependant, une idée fructueuse pour reconstruire  $W$  à partir de  $P$  consiste à calculer sa décomposition en valeurs singulières, disons

$$P = \sum_{i=1}^{\text{rang}(P)} \sigma_i(P) x_i y_i^*$$

avec les  $\sigma_i$  décroissants, et à couper toutes les valeurs singulières et vecteurs singuliers au-delà du véritable rang  $r$  de  $W$ , c'est-à-dire prendre

$$\hat{W} = \sum_{i=1}^r \sigma_i(W) x_i y_i^*.$$

Le théorème d'Eckhart-Young dit que  $\hat{W}$  est précisément la meilleure approximation de  $P$  par une matrice de rang  $r$  (au sens de la norme de Frobenius) ; le problème est qu'en règle générale, les valeurs singulières des matrices non-hermitiennes et les valeurs propres des matrices hermitiennes sont « polluées » par les plus grands degrés ([136, 24, 25, 96, 139]).

Une idée étonnante déjà exploitée par Feige et Ofek ([77]) et popularisée par Keshavan, Montanari et Oh dans leur célèbre article [92] consiste à régulariser le spectre en supprimant les hauts degrés, c'est-à-dire les lignes de  $P$  avec trop d'entrées révélées, obtenant ainsi une matrice  $\tilde{P}$ , et ensuite seulement à effectuer la troncation au rang  $r$  de la nouvelle matrice, disons  $T_r(\tilde{P})$ . Un des résultats importants de [92] est que cette opération permet de retrouver les valeurs singulières de la matrice originale ; plus précisément, sous des hypothèses simples<sup>13</sup>, les auteurs démontrent que

$$\frac{1}{n} \|W - T_r(\tilde{P})\| = O(\sqrt{r/d}). \quad (1.5.3)$$

<sup>13</sup>Cf infra.

Les inégalités de Weyl permettent alors de voir que si  $d \rightarrow \infty$ , on retrouve la matrice  $W$ , et que les  $r$  plus grandes valeurs singulières de  $\tilde{P}$  sont alignées avec celles de  $W$  ; l'équation (1.5.3) est valable à  $d$  fixé, mais dans ce cas elle ne permet pas de déterminer l'asymptotique en  $n \rightarrow \infty$  des grandes valeurs singulières de  $\tilde{P}$ .

L'aspect le plus intéressant de cette méthode est l'étape d'émondage, où l'on supprime certaines données pour régulariser le spectre de la matrice observée et faire apparaître certains paramètres de  $W$ . L'idée qui sous-tend les sections suivantes est qu'**il ne fallait pas regarder les valeurs singulières de la matrice non hermitienne  $P$ , mais justement ses valeurs propres, qui sont directement alignées avec les grandes valeurs propres de  $W$** . Cette idée neuve, selon laquelle la non-symétrie conserve plus d'informations que la symétrie, semble prometteuse et a récemment émergé dans un article de Chen et al ([56]). Nous reviendrons sur ce point dans la conclusion.

### Asymptotique des grandes valeurs propres et détection des vecteurs propres

Le résultat principal que nous avons obtenu est une description exacte du comportement des valeurs propres de la matrice  $P = (n/d)M \odot W$ , avec  $M$  la matrice d'adjacence d'un graphe d'Erdős-Rényi dirigé de paramètres  $n$  et  $d/n$  — la normalisation par  $(n/d)$  est prise de sorte que  $\mathbf{E}[P] = W$ .

Nous démontrons une transition de phase spectaculaire : il existe un seuil  $\vartheta$  tel que toutes les valeurs propres de  $P$  plus grandes que ce seuil sont asymptotiquement égales aux valeurs propres de  $W$  plus grandes que ce seuil.

Pour expliciter ce résultat, nous introduisons la matrice  $X = nW \odot W$ , ou encore

$$X_{x,y} = nW_{x,y}^2. \quad (1.5.4)$$

Nous notons  $\rho = \|X\|$  sa norme d'opérateurs, qui est du même ordre de grandeur que celle de  $W$  sous l'hypothèse d'incohérence 1.5.2. On pose également  $L = n \max_{x,y} |W_{x,y}|$ . Sous les hypothèses précédentes sur  $W$ , le réel  $L$  est d'ordre  $O(1)$ . Le seuil dont il est question plus tôt est défini par

$$\vartheta = \max \left\{ \sqrt{\frac{\rho}{d}}, \frac{L}{d} \right\}.$$

Notons  $s$  le nombre de valeurs propres de  $W$  plus grandes que  $\vartheta$  :

$$\mu_1 \geq \dots \geq \mu_s > \vartheta \geq \mu_{s+1} \geq \dots \geq \mu_r. \quad (1.5.5)$$

On note  $|\lambda_1| \geq \dots \geq |\lambda_n|$  les valeurs propres de la matrice  $A$ . Le premier résultat est une asymptotique exacte des grandes valeurs propres de  $P$ .

**Théorème 7.** *Soit  $d > 1$  un réel et  $W$  une matrice réelle de taille  $n \times n$  vérifiant les hypothèses mentionnées plus haut. Si  $n$  est assez grand, pour tout  $i \in \{1, \dots, s\}$  on a*

$$|\mu_i - \lambda_i| = o(1). \quad (1.5.6)$$

Par ailleurs, pour tout  $i > s$  on a

$$|\lambda_i| \leq \vartheta + o(1). \quad (1.5.7)$$

Signalons que le seuil  $\vartheta$  est défini comme un maximum : c'est le même phénomène que celui déjà noté dans la première partie de cette thèse, où la seconde valeur propre était asymptotiquement inférieure à  $\rho \vee \delta^{-1}$ . Le nombre  $\delta^{-1}$  était d'ailleurs égal au maximum des entrées de la matrice de transition, comme ici  $L/d$  est le maximum des entrées de  $P$  avec grande probabilité.

Lorsque  $d$  est suffisamment grand (plus précisément lorsque  $d > L^2/\rho$ ) c'est le premier terme qui l'emporte dans la définition de  $\vartheta$  ; on peut vérifier facilement que  $L^2/\rho$  est une mesure de la délocalisation de la matrice  $W$ , au sens où pour une matrice très délocalisée (typiquement, avec toutes

les entrées égales) on a  $L^2/\rho = 1$ . Il y a donc un phénomène de compétition entre la délocalisation de la matrice et la densité  $d$ .

Notre second résultat est une description du comportement des vecteurs propres de  $P$ . Le résultat n'est pour l'instant démontré que pour les matrices de rang 1 (un cadre déjà très intéressant, voir [56]) mais il sera étendu à tous les faibles rangs dans de prochains travaux.

**Théorème 8.** *Supposons que  $W = \mu\varphi\varphi^*$  avec  $\varphi$  un vecteur localisé. Si  $d > n|\varphi|_4^4$ , alors  $\lambda_1 \rightarrow \mu_1$ , et si  $\psi$  est le vecteur propre de  $P$  associé à  $\lambda_1$ , alors avec probabilité  $1 - o(1)$ ,*

$$|\langle \psi, \varphi \rangle| = (1 + o(1)) \sqrt{1 - \frac{n|\varphi|_4^4}{d}}. \quad (1.5.8)$$

Le seuil  $n|\varphi|_4^4$  provient simplement du calcul de  $\rho$  dans ce cas précis. Noter que plus  $\varphi$  est délocalisé, plus ce nombre est petit, autre exemple de la compétition évoquée plus haut entre délocalisation de  $W$  et densité  $d$ .

Il est donc possible d'obtenir une corrélation strictement positive avec le véritable vecteur propre  $\varphi$ ; en l'absence totale d'informations, le meilleur estimateur possible pour  $\varphi$  consiste simplement à prendre un vecteur uniforme  $u$  sur  $\mathbb{S}^{n-1}$  et la corrélation  $\langle u, \varphi \rangle$  tend alors vers 0.

### Non-orthogonalité

Nos travaux ont mis en évidence un phénomène étonnant : les vecteurs propres de la matrice  $P$  associés aux valeurs propres au-dessus du seuil  $\vartheta$  possèdent une structure de covariance non-triviale. Plus précisément, il existe une matrice  $\Gamma$  de taille  $s$ , dépendant uniquement de  $W$  et de  $d$ , telle que si  $\psi_i, \psi_j$  sont les vecteurs propres de  $P$  associés à  $\lambda_i \sim \mu_i$  et  $\lambda_j \sim \mu_j$ , alors  $\langle \psi_i, \psi_j \rangle \approx \Gamma_{i,j}$  (pour un énoncé précis, on renvoie au dernier chapitre du manuscrit). En règle générale, la matrice  $\Gamma$  n'est pas diagonale :

$$\Gamma_{i,j} = \sum_{k=0}^{\infty} \frac{\langle \mathbf{1}, X^k \varphi^{i,j} \rangle}{(\mu_i \mu_j d)^k} \quad (1.5.9)$$

où  $X$  a été définie en (1.5.4) et  $\varphi^{i,j} := \varphi_i \odot \varphi_j$  est le produit terme à terme des vrais vecteurs propres. Dans certains cas (par exemple, si la masse  $\ell^2$  de chaque ligne de  $W$  est constante), on peut facilement voir que  $\Gamma$  est diagonale, mais il est possible de construire des exemples simples pour lesquels ce n'est pas le cas. Une des difficultés dans la démonstration de ces résultats provient de cette non-orthogonalité (qui disparaît lorsque  $d \rightarrow \infty$ ).

### Erdős-Rényi dirigé

Une conséquence immédiate des théorèmes précédents concerne le spectre du graphe d'Erdős-Rényi dilué, obtenu en prenant tout simplement  $W_{i,j} = 1/n$ , dont le spectre est 1 et 0 avec multiplicité  $n - 1$ . Comme  $nW \odot W = W$ , on obtient  $\rho = 1$  et  $\vartheta = 1/\sqrt{d}$ .

Ces graphes sont pour l'instant relativement peu étudiés du point de vue spectral, surtout comparés à leurs homologues non-dirigés : pour ces derniers, très bien connus maintenant, les résultats de Komlos et Füredi [81] et Krievlevitch et Sudakov [96] complétés par les récents travaux de Benaych-Georges, Bordenave et Knowles ([24, 25], Tikhomirov et Youssef [136] et Alt, Ducatez et Knowles [11] élucident entièrement le comportement des grandes valeurs propres pour tous les régimes de  $d$ .

Pour les graphes d'Erdős-Rényi dirigés, notre résultat est le suivant.

**CONSÉQUENCE 1.5.1.** *Soit  $d$  un nombre réel fixé et soit  $A$  la matrice d'adjacence d'un graphe d'Erdős-Rényi  $ER(n, d/n)$  dirigé. Les assertions suivantes sont vraies avec probabilité  $1 - o(1)$  lorsque  $n \rightarrow \infty$  :*

1. Si  $d \leq 1$ , toutes les valeurs propres de  $A$  sont de module plus petit que  $1 + o(1)$ .

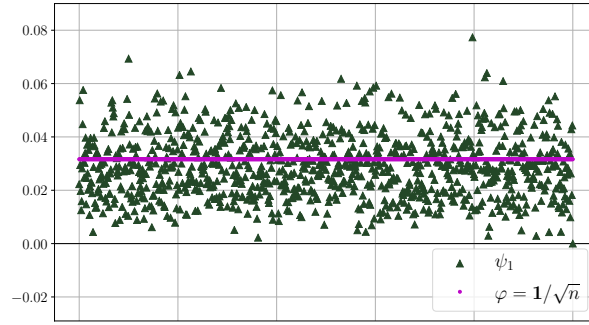


FIGURE 1.12 – Représentation (en vert) du vecteur propre  $\psi$  associé à la plus grande valeur propre de  $A$  pour un graphe d’Erdős-Rényi sur 1000 sommets avec  $d = 7$ . Toutes les entrées de  $\psi$  sont positives (théorème de Perron) et la corrélation entre  $\psi$  et le véritable vecteur propre  $\mathbf{1}/\sqrt{n}$  (en violet) est proche de  $\sqrt{1 - 1/7} \approx 0.9258$ .

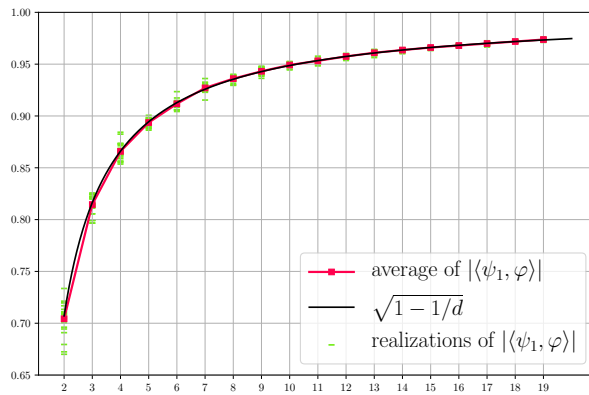


FIGURE 1.13 – Représentation de  $|\langle \psi, \mathbf{1}/\sqrt{n} \rangle|$  pour plusieurs valeurs de  $d$ , avec  $n = 1000$ . Pour chaque  $d = 2, \dots, 20$ , on a généré 20 réalisations de  $\text{ER}(1000, d)$  et calculé  $|\langle \psi, \mathbf{1}/\sqrt{n} \rangle|$  (petits tirets verts). La moyenne est en rouge et la limite théorique  $\sqrt{1 - 1/d}$  est en noir.

2. Sinon,  $d > 1$  ; dans ce cas,  $\lambda_1(A) \rightarrow d$  et toutes les autres valeurs propres sont de module plus petit que  $\sqrt{d} + o(1)$ .

Si  $\psi$  désigne le vecteur propre de  $A$  associé à  $\lambda_1(A)$ , alors

$$|\langle \psi, \varphi \rangle| \rightarrow \sqrt{1 - \frac{1}{d}}. \quad (1.5.10)$$

Ces points sont illustrés dans les figures 1.12-1.13.

L’énoncé précédent démontre ainsi une observation bien connue et souvent mentionnée par les physiciens, à savoir que la mesure empirique spectrale a son support dans  $D(0, \sqrt{d})$  y compris dans le cas dilué ; on renvoie à l’excellente synthèse [114] (venue de la physique), et en particulier aux sections 4.2 et 4.3, qui mentionnent ces faits sans les démontrer rigoureusement toutefois.

Notons  $\bar{\mu}_{A_n}$  la mesure empirique spectrale de la matrice d’adjacence d’un graphe d’Erdős-Rényi dirigé de paramètre  $d/n$  ; à ce jour et à ma connaissance, il n’existe strictement aucun résultat rigoureux sur ces mesures dans le cadre dilué. Seul un article de Basak et Rudelson [18] démontre que si  $d = \Omega(\ln(n)^2)$ , alors la mesure empirique spectrale converge vers la loi du cercle. Il n’est à ce jour pas démontré que  $\bar{\mu}_{A_n}$  converge vers une mesure sur  $\mathbb{C}$ , et une éventuelle forme limite n’est même pas conjecturée, au contraire de la conjecture de Kesten-McKay orientée déjà mentionnée (Figure 1.4).

Reprenons enfin une remarque déjà faite dans la première section de cette introduction : nous ne démontrons qu’une borne supérieure. La borne inférieure correspondante, à savoir  $|\lambda_2| \geq \sqrt{d} -$

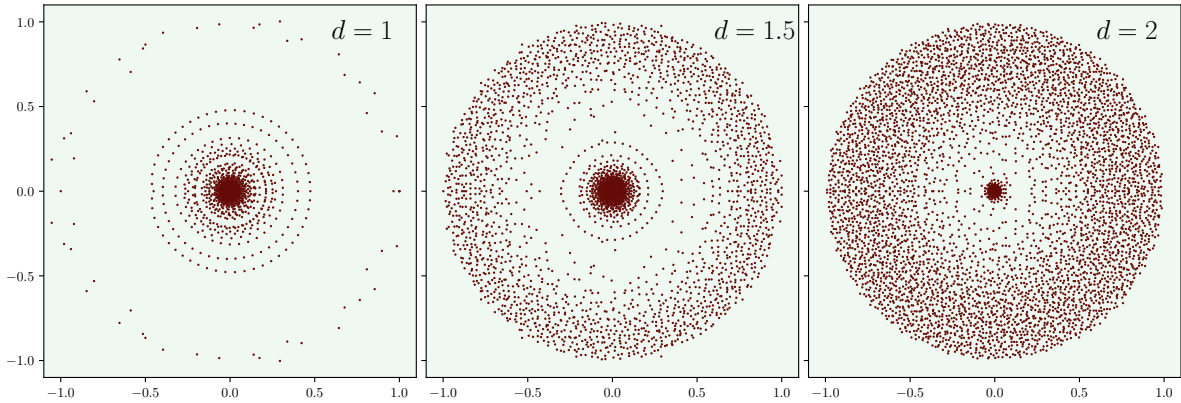


FIGURE 1.14 – Trois réalisations de spectres de graphes d’Erdős-Rényi dirigés à  $n = 10000$  sommets, pour trois valeurs de  $p = d/n$ . On a coupé l’outlier et renormalisé par  $1/\sqrt{d}$ . Lorsque  $d = 1$ , le graphe limite est fini presque sûrement et le spectre se trouve concentré sur les valeurs propres des graphes dirigés finis.

$o(1)$ , n’est pas démontrée, mais semble vraie (voir en particulier la figure 1.14). Les démonstrations classiques de telles bornes inférieures (comme la démonstration de la borne d’Alon-Boppana) ne se transposent pas au cas dirigé.

### Cas typique : vecteurs propres uniformes.

Nos résultats sont valables pour n’importe quelle matrice initiale fixée  $W$  vérifiant les hypothèses mentionnées. On pourrait penser que ces hypothèses sont restrictives, mais ce n’est pas le cas. En un sens, *presque toutes les matrices* vérifient ces hypothèses. En effet, étant donné un spectre  $\Sigma = \text{diag}(\mu_1, \dots, \mu_r)$ , on peut choisir la matrice des vecteurs propres  $U \in \mathcal{M}_{n,r}(\mathbb{R})$  distribuée selon la mesure de Haar sur l’ensemble des matrices orthonormales de taille  $n \times r$ ; la matrice

$$W = U^* \Sigma U$$

est alors uniforme parmi toutes les matrices ayant  $\Sigma$  comme spectre, et il est bien connu que  $U$  vérifie l’hypothèse de délocalisation au sens où  $\|U\|_\infty = O(\sqrt{\ln n/n})$  avec grande probabilité. Un tel modèle est populaire dans la littérature de la complétion de matrices, et déjà utilisé dans [53, 92].

Les théorèmes énoncés plus hauts seront donc valables pour « la plupart » des matrices de rang  $r$  ayant un spectre fixé. On trouvera dans les figures 1.15 et 1.16 des illustrations du phénomène décrit dans le théorème 8 pour une matrice de rang 1.

Ces résultats complètent bien ceux obtenus dans [56], qui se placent toutefois dans le régime  $d \rightarrow \infty$  (voir en particulier [56, Page 10]) et n’obtiennent pas la transition en  $\vartheta$  ou les asymptotiques exactes ; en effet, notre méthode n’utilise pas de théorème de perturbation des vecteurs propres comme Davis-Kahan ou l’identité de Neumann (voir [142, 71, 56]), qui sont souvent optimaux dans le pire des cas mais assez mauvais dans les cas typiques.

Nos démonstrations intègrent directement des approximations de vecteurs propres, et permettent ainsi d’obtenir des informations fines sur ces derniers : un élément clé dans nos démonstrations est que si  $\ell = c \ln(n)$ , où  $c$  est une constante judicieusement choisie, alors le vecteur  $A^\ell \varphi_i$  est asymptotiquement aligné avec le  $i$ -ème vecteur propre de  $A$  lorsque  $i \in [r_0]$ .



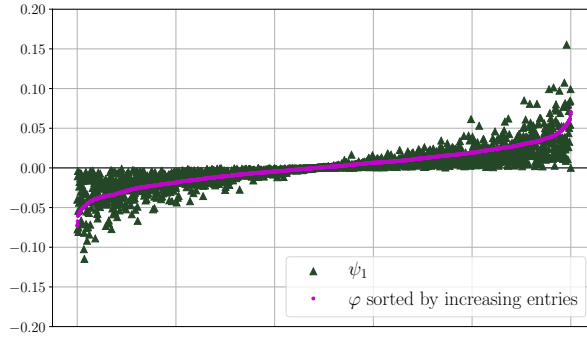


FIGURE 1.15 – On a pris  $P = \varphi\varphi^*$  avec  $\varphi \sim \text{Unif}(\mathbb{S}^{n-1})$ ,  $n = 1000$ . Les entrées de  $\varphi$  ont été triées par ordre croissant. Les entrées du vecteur  $\psi$  sont en vert. La corrélation théorique est proche de  $\sqrt{1 - n|\varphi|_4^4/d} \approx \sqrt{1 - 3/d}$  avec grande probabilité.

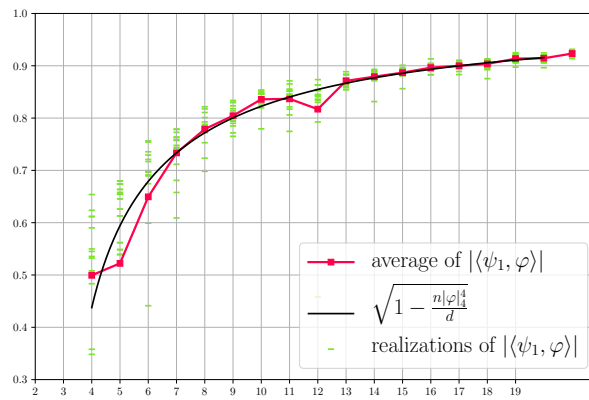


FIGURE 1.16 – Représentation de  $|\langle \psi_1, \varphi \rangle|$  pour différents  $d$ . On a d’abord généré  $W = \varphi\varphi^*$  avec  $\varphi$  uniforme sur  $\mathbb{S}^{n-1}$ . Ensuite, pour chaque  $d = 2, \dots, 20$ , on a effectué 20 simulations de  $A$  avec  $W$  comme matrice sous-jacente, puis on a calculé  $\langle \psi_1, \varphi \rangle$ . La moyenne est en rouge et la valeur limite  $\sqrt{1 - n|\varphi|_4^4/d}$  est en noir.

## 1.6 Deux remarques pour finir

Cette dernière section présente deux idées nées des résultats de cette thèse.

### ‘Diluted’ is the new ‘sparse’

Dans la littérature sur les réseaux ou les matrices peu denses, on observe plusieurs interprétations sur ce que signifie « peu dense ». Le consensus général est que le degré moyen des connections d’un agent, disons  $d$ , doit être négligeable devant la taille  $n$  du système :  $d = o(n)$ . Cependant, deux régimes sont peu denses en ce sens : le régime que nous appellerons **dilué**, où  $d = O(1)$  est borné indépendamment de  $n$ , et le régime **épars** où  $d = o(n)$  mais tend vers l’infini, typiquement  $d = (\ln n)^c$  ou  $d = n^\alpha$ , avec  $\alpha < 1$ . Signalons que cette terminologie n’est pas fermement établie, et que les deux termes ont été rencontrés l’un pour l’autre dans la littérature.

L’étude « moderne » des réseaux peu denses a commencé dans les années 1960 avec les travaux sur la transition de phase dans le modèle d’Erdős-Rényi, puis elle a véritablement explosé depuis les années 1990 avec l’émergence de réseaux informatiques ou sociaux, dans lesquels les agents possèdent un nombre relativement faible de liens avec les autres agents. Or, du point de vue des applications, les deux régimes dilué et épars sont identiques. Les ingénieurs, statisticiens ou physiciens ne seront jamais

confrontés à des réseaux de taille supérieure à, mettons,  $10^{80}$  (le nombre d'atomes dans l'univers observable), et en pratique les régimes dilués ou épars seront strictement indiscernables (on a  $\ln 10^{80} \approx 184$ ). Pour ces applications, on peut donc *a priori* se contenter de résultats sur les cas épars, où  $d \rightarrow \infty$  lentement, car ces derniers se sont révélés plus simples à traiter.

La différence entre les régimes *dilué* et *épars* n'est pourtant pas artificielle, bien au contraire, et révèle une transition de phase profonde et à mon sens sous-estimée dans la littérature : le régime épars se situe naturellement dans la classe d'universalité spectrale « dense », celle du semi-cercle pour les modèles hermitiens, et celle du cercle pour les modèles non-hermitiens. Dans les principaux modèles de graphes ( $d$ -réguliers, Erdős-Rényi  $d/n$ ) avec densité  $d \rightarrow \infty$ , la mesure spectrale converge systématiquement vers  $\rho_{sc}$  ou vers  $\rho_{cercle}$ , et ce y compris aux échelles microscopiques ([18, 57, 17, 59, 75, 47, 75, 74]). Or, une partie des réseaux réels étudiés par les ingénieurs ou les statisticiens ne fait pas partie de la classe d'universalité de Wigner ([76, 129]), comme les graphes des réseaux sociaux, dont le spectre ne ressemble visiblement pas à un semi-cercle (figure 1.17).

Le bon critère pour les modèles de réseaux réels ne semble donc pas être uniquement le faible nombre de liens entre les agents du système, mais aussi la structure spectrale assez éloignée des lois de référence comme la loi du semi-cercle ou du cercle : présence d'atomes, supports non bornés, états étendus et non étendus. Le lecteur pourra s'en convaincre en observant les quelques exemples de spectres de réseaux réels dans la figure 1.17 page 34 (les réseaux en question sont décrits à la fin du paragraphe). Ce sont ces aspects là des modèles réels que l'on devrait retrouver dans les modèles des mathématiques appliquées. De ce point de vue, même des modèles de graphes dilués ne conviennent pas forcément : les résultats obtenus pour un graphe d'Erdős-Rényi avec paramètre  $d$  fixé mais grand, par exemple  $d = 10^{45}$ , seront certainement plus éloignés de la réalité que ceux obtenus pour tout  $n$  et pour tout  $d \leq \ln n$ .

Les données des réseaux de la figure 1.17, page 34, sont en accès libre, voir la base de donnée [99] et l'article [140] et les références ci-dessous. Les quatre jeux de données dont j'ai représenté le spectre sont les suivants.

- (a) La Bible du roi Jacques (*King James Bible*, 1611) est une version anglaise de la Bible, encore aujourd'hui référence dans l'Église Anglicane. Elle contient de nombreux noms (de villes ou de personnes), qui sont les 1773 sommets du graphe. Deux noms sont reliés par une arête s'ils apparaissent dans le même verset ; chaque arête est pondérée par le nombre de versets dans lequel apparaissent les deux noms (ce sont donc des multi-arêtes). Il y a en tout 9131 arêtes et 16401 multi-arêtes ; le degré moyen est 18,5 et le degré maximal est 364 — il s'agit du nom Israël. Ces données ont été compilées par Chris Harrison.
- (b) PGP (PrettyGoodPrivacy) est un algorithme populaire de cryptage, utilisé en particulier pour les échanges de courriers. Les noeuds du réseau sont les utilisateurs (10680) dans la composante connexe géante, et deux utilisateurs sont connectés s'ils ont échangé des informations sécurisées par PGP. Il y a 24316 arêtes, le degré moyen est 4,55 et le degré maximal est 205 (voir [30]).
- (c) Le réseau électrique des États-Unis (un des réseaux étudiés dans le célèbre article [140]) est composé de 4941 unités (en règle générale, des transformateurs ou des centrales) ; les arêtes représentent les lignes électriques et il y en a 6594. Le degré moyen est 2,66 et le degré maximal est 19.
- (d) En 2003, le réseau de courrier électronique de l'université Rovira i Virgile à Tarragone ([86]) contenait 1133 utilisateurs (composante géante). Deux utilisateurs sont reliés s'ils se sont envoyés un mail. Il y a 5451 arêtes, le degré moyen est 9,6 et le degré maximal est 71.

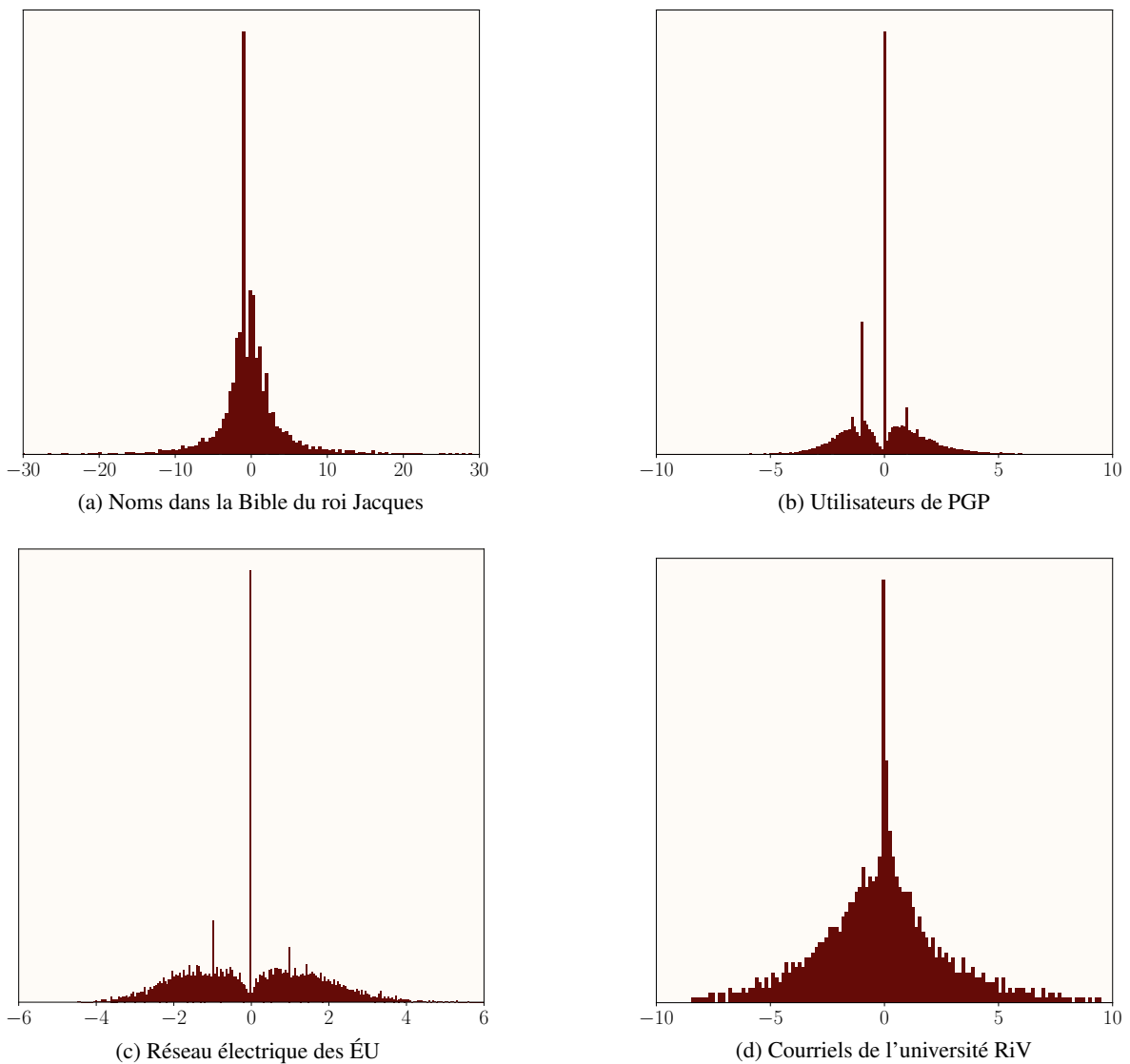


FIGURE 1.17 – Quelques spectres de réseaux réels dilués.

### Moins de symétrie, plus de valeurs propres

On trouvera dans la plupart des manuels d'algèbre des descriptions de l'instabilité des valeurs propres des matrices non symétriques. Par exemple, il est facile de construire deux matrices non symétriques très proches dans plusieurs normes, mais dont les valeurs propres sont parfaitement dissemblables. Pourtant, la remarquable et omniprésente loi du cercle semble indiquer que les valeurs propres de certaines matrices aléatoires non symétriques sont en fait très stables (voir [40, Figure 2 et commentaire]), comme si l'aléa dans la matrice régularisait suffisamment le spectre.

Des résultats récents semblent en effet montrer que les matrices non-symétriques possèdent des propriétés spectrales bien plus fines qu'on aurait pu l'imaginer, et reflètent même les statistiques des réseaux sous-jacents *mieux que leurs équivalents symétriques*. Par exemple, dans le problème de la détection de communautés, ce n'est pas la matrice d'adjacence qu'il fallait étudier, mais une de ses variantes non-symétriques, la matrice non-backtracking ([97, 43]), parce que les grandes valeurs propres de cette dernière reflètent mieux la structure interne du graphe.

Récemment, un article de Chen et al ([56]) a franchi le pas, en suggérant qu'il serait parfois bénéfique de dé-symétriser des problèmes symétriques. Les auteurs remarquent et démontrent un phé-

nomène tout à fait similaire à celui exposé dans la dernière partie de cette thèse : si  $M$  est une matrice symétrique de faible rang perturbée par un bruit  $H$  non-symétrique, il est plus efficace et plus facile de retrouver la structure de  $M$  à partir des valeurs propres de  $M' = M + H$  qu'à partir des valeurs singulières de  $M'$  ; de plus, même si la matrice sous-jacente  $M$  est symétrique, il est sous-optimal d'utiliser cette information en symétrisant  $M'$ .

C'est également la conclusion de nos travaux : y compris lorsque la matrice sous-jacente est symétrique, dé-symétriser les observations revient à les réorganiser pour mieux faire apparaître leur structure, et ce

- sans supprimer d'informations, contrairement à d'autres procédures déjà utilisées, par exemple l'émondage des hauts degrés dans [77] ou [92],
- sans perdre en dimension, comme c'était le cas avec la matrice non-backtracking dont le passage de la dimension  $n$  à la dimension  $dn$  pouvait s'avérer pénalisant en pratique, même lorsque  $d = 10$ .

Une telle dé-symétrisation peut ne pas être optimale ; les auteurs de [56] notent par exemple que leurs procédures de dé-symétrisation entraînent une augmentation de la variance du bruit. Deux perspectives semblent prometteuses : premièrement, les résultats théoriques que nous avons exposés dans la dernière partie de cette thèse ouvrent la voie à des calculs de seuils très précis, qui permettront souvent de résoudre entièrement des problèmes statistiques malgré l'augmentation du bruit ; et deuxièmement, d'autres procédures de dé-symétrisation moins élémentaires permettront certainement de mieux tirer parti de cette nouvelle philosophie.



## Chapter 2

# The spectral gap of sparse random digraphs

This chapter is drawn from the prepublication [62].

### 2.1 Introduction and statement of the results

#### 2.1.1 Directed configurations

Given two  $n$ -tuples of positive integers, say  $(d_1^+, \dots, d_n^+)$  and  $(d_1^-, \dots, d_n^-)$ , we build a sequence of directed multigraphs  $G_1, G_2, \dots$  using the configuration model: at each of the  $n$  vertices (labeled from 1 to  $n$ ), we glue tails and heads. The vertex  $i$  has  $d_i^+$  heads and  $d_i^-$  tails. For consistency we ask the total number of tails to be equal to the total number of heads:

$$\sum_{i=1}^n d_i^+ = \sum_{i=1}^n d_i^- := M. \quad (2.1.1)$$

We then choose uniformly at random a matching of the tails into the heads, that is a random permutation  $\sigma_n \in \mathfrak{S}_M$ . If  $\mathbf{e}$  is a head attached to vertex  $x$ , we glue it to the tail  $\sigma_n(\mathbf{e}) = \mathbf{f}$ . If  $\mathbf{f}$  is attached to vertex  $y$ , this gives rise to an oriented edge from  $x$  to  $y$ . The whole construction leads to a directed multigraph  $G_n$  (we will often say *digraph*) on  $n$  vertices called the *directed configuration graph* associated with the so-called *degree sequence*  $d_1^+, d_1^-, \dots, d_n^+, d_n^-$ . The permutation  $\sigma_n$  will sometimes be called the *environment*.

The random graph  $G_n$  will simply be noted  $G$ , the  $n$ -dependence being implicit through all this paper. We are interested in properties of  $G$  in the asymptotic regime  $n \rightarrow \infty$ : we say that an event depending on  $n$  holds *with high probability* if its probability tends to 1 as  $n \rightarrow \infty$ .

If  $u$  is a vertex, we will adopt the following notations:  $E^+(u)$  is the set of all heads attached to  $u$ , and  $E^-(u)$  is the set of all tails attached to  $u$ . Therefore,  $\#E^+(u) = d_u^+$  and  $\#E^-(u) = d_u^-$ . Through all this paper, and unless specified otherwise, heads will be denoted by the bold letter  $\mathbf{e}$  and tails by  $\mathbf{f}$ .

#### 2.1.2 Statement of the theorem and illustrations

The transition probability matrix  $P$  on the graph  $G$  is defined as follows:

$$P(u, v) = \frac{\#\{\mathbf{e} \in E^+(u) : \sigma(\mathbf{e}) \in E^-(v)\}}{d_u^+}. \quad (2.1.2)$$

The matrix  $P$  is thus a random stochastic matrix. The *eigenvalues* of  $P$  are the  $n$  complex roots (counted with multiplicity) of its characteristic polynomial  $\det(P - zI)$ . We order them by decreasing modulus:

$$|\lambda_n| \leq |\lambda_{n-1}| \leq \dots \leq |\lambda_2| \leq \lambda_1 = 1.$$

Recall that all those eigenvalues are random variables depending implicitly on  $n$  and on the degree sequence  $(d_i^+, d_i^-)_{i \leq n}$ . We will impose that all the degrees are bounded independently on  $n$ , meaning that there are two constants  $\delta \geq 2$  and  $\Delta \geq \delta$  such that for every  $n$ ,

$$\delta \leq \min\{d_1^+, d_1^-, \dots, d_n^+, d_n^-\} \quad \text{and} \quad \max\{d_1^+, d_1^-, \dots, d_n^+, d_n^-\} \leq \Delta. \quad (\text{H1})$$

Under the first assumption, the minimal degree is greater than two (which means there are no dead-ends) and the graph  $G$  is strongly connected with high probability as shown in [61]. Let us introduce a central parameter of this model:

$$\rho := \sqrt{\frac{1}{M} \sum_{i=1}^n \frac{d_i^-}{d_i^+}}. \quad (2.1.3)$$

Our goal is to link the modulus of the second eigenvalue with  $\rho$ . The main result is the following theorem.

**Theorem 1.** *Let  $P$  be the transition matrix (2.1.2) of the random digraph associated with the degree sequence  $(d_1^+, d_n^-, \dots, d_n^+, d_n^-)$  satisfying (H1). Let  $\rho$  be as in (2.1.3) and define  $\tilde{\rho} = \rho \vee \delta^{-1}$ . Then, as  $n$  goes to infinity, we have for every  $\varepsilon > 0$ :*

$$\lim_{n \rightarrow \infty} \mathbf{P}(|\lambda_2| > \tilde{\rho} + \varepsilon) = 0. \quad (2.1.4)$$

Hence, for every  $\varepsilon > 0$ , with high probability as  $n$  goes to infinity, the second eigenvalue satisfies

$$|\lambda_2| \leq \max \left\{ \frac{1}{\delta}, \sqrt{\frac{1}{M} \sum_{i=1}^n \frac{d_i^-}{d_i^+}} \right\} + \varepsilon.$$

This theorem only provides an upper bound for  $|\lambda_2|$ ; knowing if the bound is optimal and having a symmetric lower bound are questions not addressed in this paper. The following figure shows an illustration of (2.1.4).

**REMARK 2.1.1.** *When  $\delta^{-1}$  is smaller than  $\rho$ , the bound of theorem 1 is equal to  $\rho$ . This happens when*

$$\delta \rho > 1 \quad (2.1.5)$$

and this is not always verified as shown in the following example:

$$\begin{cases} d_i^+ = d_i^- = 2 & \forall i \in \{1, \dots, 100\} \\ d_i^+ = d_i^- = 8 & \forall i \in \{101, \dots, 200\}. \end{cases}$$

This degree sequence satisfies  $\rho = \sqrt{n/M} = \sqrt{200/1000} \simeq 0.45$  and in this case we have  $\delta \rho < 1$ . In fact, using Jensen's inequality, one can give a slightly stronger form of (2.1.5). Let  $\pi^-$  be the so-called in-degree distribution on vertices  $\{1, \dots, n\}$ , that is  $\pi^-(i) = d_i^-/M$ . Let  $U$  be a random variable with probability distribution  $\pi^-$ : we have

$$\rho^2 = \mathbf{E} \left[ \frac{1}{d_U^+} \right].$$

Using Jensen's inequality for the convex function  $x \mapsto 1/x$ , we get  $\mathbf{E}[d_U^+]^{-1} \leq \rho^2$ . A direct consequence of hypothesis (H1) is  $\delta \leq \mathbf{E}[d_U^+] \leq \Delta$ , so (2.1.5) is fulfilled when  $\mathbf{E}[d_U^+] < \delta^2$ . This hypothesis can be interpreted as a concentration hypothesis in the sense that the out-degree of a  $\pi^-$ -distributed random vertex has an expectation not far from the minimum out-degree.

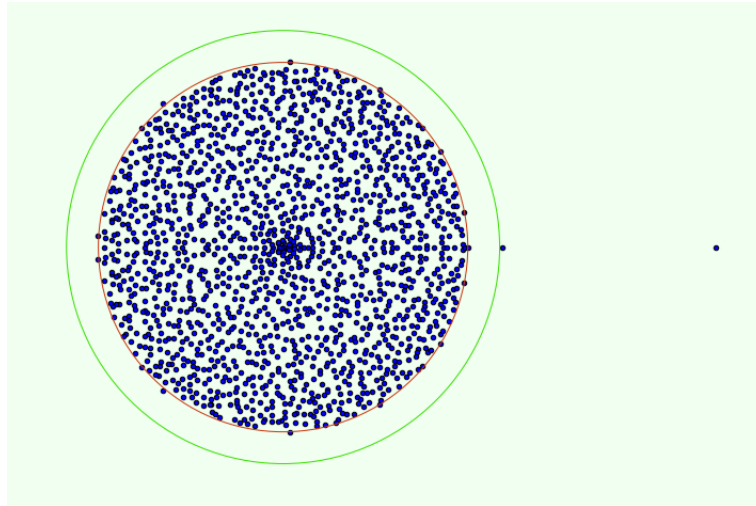
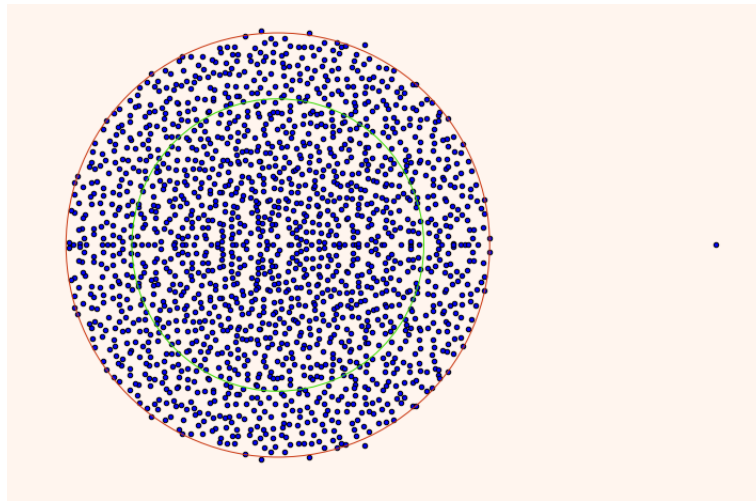
(a) Case with  $\tilde{\rho} = \delta^{-1}$ .(b) Case with  $\tilde{\rho} = \rho$ .

Figure 2.1 – Two spectra of the transition matrix on a random configuration digraph. We drew in red the circle with radius  $\rho$ ; in green, the circle with radius  $\delta^{-1}$ . The rightmost outlier is the Perron eigenvalue  $\lambda_1 = 1$ .

- In figure (a) there are  $n = 1600$  vertices: 700 of them have type  $(2, 2)$  and 800 have type  $(9, 9)$ . In this case we have  $\tilde{\rho} = \delta^{-1} = 1/2$ . Notice that there are very few outliers outside the circle of radius  $\rho$ : only one in this case.
- In figure (b), there are  $n = 1800$  vertices, 600 of them have type  $(5, 6)$ , 600 of type  $(3, 7)$  and 600 of type  $(9, 4)$ . Here we have  $\tilde{\rho} = \rho$ .

### 2.1.3 Ramanujan digraphs and the Alon conjecture

A  $d$ -regular undirected graph is said to be *Ramanujan* if every eigenvalue  $\lambda$  of its transition matrix has  $|\lambda| = 1$  or  $|\lambda| \leq 2\sqrt{d-1}/d$ . Those graphs have been very well studied, notably for their optimal expansion properties ([66, 87]). The reason why the value  $2\sqrt{d-1}/d$  appears here is because the universal cover of every  $d$ -regular graph is the infinite  $d$ -regular tree  $\mathbb{T}_d$ , and its transition operator has spectrum  $[-2\sqrt{d-1}/d, 2\sqrt{d-1}/d]$ , a classical result of Kesten [93]; Ramanujan graphs are the regular graphs whose non-trivial eigenvalues are included in the spectrum of their universal cover.



A recent line of research generalized this to digraphs, as recently<sup>1</sup> surveyed in [124]: the universal cover of a  $d$ -regular digraph is the infinite  $d$ -regular tree  $\vec{\mathbb{T}}_d$  obtained from the infinite  $2d$ -regular tree  $\mathbb{T}_{2d}$  by assigning a direction for  $d$  edges at every vertex and the other direction for the  $d$  other edges at this vertex. The spectrum of the transition operator  $\vec{\mathbb{T}}_d$  is precisely  $\{z \in \mathbb{C} : |z| \leq 1/\sqrt{d}\}$  as proven in [67]. By analogy, a  $d$ -regular digraph is called Ramanujan if every eigenvalue  $\lambda$  of its adjacency matrix has  $|\lambda| = 1$  or  $|\lambda| \leq 1/\sqrt{d}$ .

Explicit constructions of Ramanujan graphs have been a challenging problem with a rich history, but one of the most striking phenomenon in the domain is that *most* regular graphs are *nearly* Ramanujan. More precisely, Alon conjectured in [9] that for every  $d, \varepsilon$ , the second eigenvalue  $\lambda_2$  of the transition matrix of a uniform  $d$ -regular graph on  $n$  vertices is smaller than  $2\sqrt{d-1}/d + \varepsilon$  with high probability when  $n \rightarrow \infty$ . The question remained open for two decades and was solved by Friedman in his celebrated 2004 paper [79]. In fact, the bound was optimal due to a simple inequality already shown by Alon, sometimes referred to as the Alon-Boppana inequality ([120]). This is now called *Friedman's second eigenvalue theorem*:

**Theorem 2** ([79, 33]). *Fix an integer  $d > 2$ . For every  $\varepsilon > 0$ , as  $n \rightarrow \infty$  we have*

$$\mathbf{P}\left(\left|\lambda_2 - \frac{2\sqrt{d-1}}{d}\right| > \varepsilon\right) \rightarrow 0. \quad (2.1.6)$$

This solved the first-order asymptotic behaviour of the second eigenvalue for regular graphs; we refer the reader to the introductions of [9, 33, 66, 87] for further reference. When it comes to regular digraphs, our main theorem settles the Alon conjecture for *digraphs* (see [124, section 5.5]). In fact, in a  $d$ -regular digraph, we have  $d_i^+ = d_i^- = d$ , hence  $\bar{p}$  is equal to  $\frac{1}{d} \vee \frac{1}{\sqrt{d}} = \frac{1}{\sqrt{d}}$ . We state this as a corollary.

**COROLLARY 2.1.2.** *Let  $d \geq 2$  be a fixed integer and  $P$  be the transition matrix of a random  $d$ -regular digraph. Note  $|\lambda_n| \leq \dots \leq |\lambda_2| \leq \lambda_1 = 1$  the eigenvalues of  $P$ , ordered by decreasing modulus. Fix  $\varepsilon > 0$ . Then, as  $n$  goes to infinity, the following holds with high probability:*

$$|\lambda_2| \leq \frac{1}{\sqrt{d}} + \varepsilon. \quad (2.1.7)$$

## 2.1.4 Motivation, background and related work

### Random digraphs

In this paper, we consider random directed (multi)graphs with a specified sequence of in-degrees and out-degrees; when all the degrees are equal to  $d$ , this model reduces to the directed  $d$ -regular case. Our construction with half-edges is a directed variant of the classical configuration model (see [32]). When the degrees are bounded independently of the size of the graph, such multigraphs are sparse, meaning they have few edges. Even if digraphs are much more difficult to handle than undirected graphs, they are also one step closer to reality when modelling real-life situations: see [119, 60] and references for (many) examples of graph-modelling that go beyond the Internet graph.

### Eigenvalues of Markov chains

Many strong connections exist between the second eigenvalue of a transition matrix and the convergence properties of the corresponding Markov chain. The following proposition is the most known result:

---

<sup>1</sup>The survey [124] appeared on the ArXiv after the first version of this paper.

**PROPOSITION 2.1.3** ([104], [117]). *Let  $P$  be the transition matrix of an irreducible, aperiodic Markov chain on the finite state space  $S = \{1, \dots, n\}$  with stationary distribution  $\pi_\star$ . Let  $1 = |\lambda_1| \geq |\lambda_2| \geq \dots \geq |\lambda_n|$  be the eigenvalues of  $P$  ordered by decreasing modulus and  $d(n)$  be the distance to equilibrium at time  $n$ , defined as  $d(n) = \max_{x \in S} \|P^n(x, \cdot) - \pi_\star\|_{\text{TV}}$ , with  $\|\cdot\|_{\text{TV}}$  the usual total variation distance. Then,*

$$\lim_{n \rightarrow \infty} d(n)^{\frac{1}{n}} = |\lambda_2|. \tag{2.1.8}$$

In other words, large values of the *spectral gap*  $\gamma_\star := 1 - |\lambda_2|$  are linked with fast convergence. For random walks on graphs,  $\lambda_2$  is also known to be strongly linked with expansion properties of the underlying graph (see [87] for an excellent survey). It is thus of special interest to study the spectrum of transition matrices; however, instead of focusing on a fixed chain  $P$ , researchers now study “generic” models of transition matrices. Most of the time, the transition matrix is chosen at random among a certain type of matrices and its properties are studied in a probabilistic setting. In this line, random walks on random graphs have attracted an extraordinary attention during the last decades.

Another very important aspect of Markov chains linked with  $|\lambda_2|$  is *mixing*, and especially the cutoff phenomenon ([69, 104]). Proving cutoffs for large classes of random walks is an active line of research. In the context of random graphs, cutoff had been proven with high probability in the  $d$ -regular model ([106]), but it was recently shown by Lubetzky and Peres in their influential paper [105] that *every* Ramanujan graph exhibits cutoff, suggesting that optimality of the second eigenvalue is linked with optimal mixing. Our paper gives the first upper bound for the second eigenvalue for a non-reversible model of Markov chains. The cutoff phenomenon for our model has been established whp in the inspiring paper [39], with a logarithmic mixing time (see Theorems 1 and 2 in [39]). Note that our main result (Theorem 1) immediately implies Theorem 3 in [39], as a consequence of Proposition 2.1.3.

### Random transition matrices

While we are interested in the spectral gap of a special kind of those matrices, some serious advances on global asymptotics of the spectrum have recently been made. In a series of papers [57, 58, 17], Nicholas Cook and coauthors established convergence towards the circular law of the empirical spectrum of matrices related to the adjacency matrix of  $d$ -regular directed graphs, when  $d$  grows to infinity with  $n$ . In another series of papers ([37, 36, 38]), Bordenave, Caputo and Chafaï considered the spectra of a transition matrix  $P$  constructed by row-normalizing a random matrix with nonnegative iid entries  $X_{i,j}$ , that is  $P(i, j) := X_{i,j} \rho(i)^{-1}$  where  $\rho(i) := X_{i,1} + \dots + X_{i,n}$ . A key result is formulated in [38] where the authors prove the convergence towards the circular law in the sparse case where the  $X_{i,j}$  are heavy-tailed with index  $\alpha \in ]0, 1[$ . They also conjecture ([38, Remark 1.3]) that in this case, whp the second eigenvalue  $|\lambda_2|$  will be smaller than  $\sqrt{1 - \alpha}$ . We believe that our method could be adapted to tackle this conjecture.

### Non-reversible chains

A key feature of random walks on random unoriented graphs is *reversibility* of the Markov chain. When the walk is reversible, the transition matrix  $P$  has a known stationary distribution  $\pi_\star$  and is self-adjoint relatively to the hilbert product  $\langle \cdot, \cdot \rangle_\star$  defined by

$$\langle x, y \rangle_\star = \sum_{x \in V} x_i y_i \pi_\star(i) \quad (x, y \in \mathbb{R}^n).$$

In this reversible case, all the classical tools from hermitian algebra can be used to study the spectrum of  $P$ . When  $P$  is not reversible but when its stationary distribution  $\pi_\star$  is known, we can still use the *reversibilization trick* introduced by Fill ([78]; see also [117]): if  $P^\star$  denotes the time-reversibilization of  $P$ , defined as  $P^\star(i, j) = P(j, i) \pi_\star(j) \pi_\star(i)^{-1}$ , then  $PP^\star$  is self-adjoint for  $\langle \cdot, \cdot \rangle_\star$ . All

the eigenvalues  $1 = \mu_1 \geq \mu_2 \geq \dots \geq \mu_n \geq 0$  of  $PP^*$  are real and positive, and  $\mu_2 \geq |\lambda_2|^2$ , thus giving informations about  $|\lambda_2|$ . However, in any model where  $\pi_\star$  is not explicitly known, those techniques are useless.

Our method is the first one to efficiently deal with the top eigenvalue of non-hermitian matrices with no information on the eigenvectors; we strongly believe this method could prove extremely useful in other problems within the random matrix theory, especially in the non-hermitian setting.

In fact, after the first version of this paper was put on the ArXiv, other results on the spectral gap of random matrix models have been proven with this method, such as the spectral gap for random biregular bipartite graphs [50], and for sparse bistochastic matrices [45].

We finally mention some related questions and conjectures.

1. What is the link between  $|\lambda_2|$  and the cutoff phenomenon for the Markov chain ? Do *all* graphs in our model having  $|\lambda_2| \leq \rho$  exhibit cutoff ?
2. Is the upper bound (2.1.4) optimal ? In the Friedman theorem, the difficult part was to prove the upper bound while the lower bound had been proven very early ([120]) using the full strength of the symmetric nature of  $P$ . We have proven an upper bound for our model, but no lower bound is known yet.
3. This paper deals with the second eigenvalue of random digraphs in general. In the specific case of  $d$ -regular digraphs, it is conjectured in [40, Section 7] that the whole empirical spectral measure of the adjacency matrix of a  $d$ -regular digraph converges almost surely in distribution to  $\mu_{\text{OKMC}}$ , a complex version of the Kesten-McKay distribution, namely

$$\mu_{\text{OKMC}}(dz) = \pi^{-1} \frac{d^2(d-1)}{(d^2 - |z|^2)^2} \mathbf{1}_{|z| \leq \sqrt{d}} dz.$$

### 2.1.5 Conventions and notations

The operator norm of a real square matrix  $A \in \mathcal{M}_n(\mathbb{R})$  is

$$\|A\| = \sup_{x \neq 0} \frac{\|Ax\|}{\|x\|}$$

where  $\|x\| = (x_1^2 + \dots + x_n^2)^{\frac{1}{2}}$  is the standard euclidean norm. If  $M$  is any matrix,  $A^\top$  is its usual transpose. We will also note  $\mathbf{1}$  the column vector  $\mathbf{1} = (1, \dots, 1)^\top$ . If  $(a_n)$  and  $(b_n)$  are two real sequences, we use the classical Landau notations  $a_n \sim b_n$ ,  $a_n = o(b_n)$  and  $a_n = O(b_n)$ .

We will also adopt the following notations for half-edges in our model. Formally, a half-edge will be coded by a triple  $(u, i, \varepsilon)$ , where

- $u$  is a vertex,
- $\varepsilon \in \{-, +\}$  is a sign indicating the nature of the half-edge: a  $+$  symbol denotes a head, a  $-$  denotes a tail,
- $i$  is an integer in  $\{1, \dots, d_u^\varepsilon\}$ .

With this notation, we have  $E^+(u) = \{(u, i, +) : i = 1, \dots, d_u^+\}$  and also  $E^-(u) = \{(u, i, -) : i = 1, \dots, d_u^-\}$ . These notations will specifically be used in the combinatorial section 2.6. In general, it will be more convenient to adopt the following conventions, much easier to read: heads will be denoted by the bold letter  $\mathbf{e}$  and tails will be denoted by the bold letter  $\mathbf{f}$ . If a half-edge  $\mathbf{e}$  is attached to vertex  $u$ , we will write  $d_u^\pm$  instead of  $d_u^\pm$ .

For example, a 2-step path in the graph between vertices  $a$  and  $b$  is a sequence of the form  $(\mathbf{e}_1, \mathbf{f}_1, \mathbf{e}_2, \mathbf{f}_2)$  with  $\mathbf{e}_1$  attached to  $a$ ,  $\mathbf{f}_2$  attached to  $b$ ,  $\mathbf{e}_2$  and  $\mathbf{f}_1$  attached to the same vertex and  $\sigma(\mathbf{e}_1) = \mathbf{f}_1, \sigma(\mathbf{e}_2) = \mathbf{f}_2$ . We will give a complete and precise definition of *paths* further in the paper.

In the rest of the paper, we will denote all universal constants by  $C > 0$ .

### 2.1.6 Acknowledgements

The author is grateful to his advisors Charles Bordenave and Justin Salez for their valuable help and advice during the writing of this paper, from preliminary discussions about the problem and the understanding of [33] to the final remarks on the manuscript.

## 2.2 Proof of the main theorem

### 2.2.1 Outline

We give a motivated sketch of the main difficulties in the proof of our theorem and the core ideas to overcome them.

As mentioned in the beginning of [79] or [33], the standard trace method for bounding  $|\lambda_2|$  is doomed to fail: the main obstruction comes from the fact that with small probability, some very small graphs with many cycles (“tangled graphs”) are present in the graph, and they drastically perturb the expectation of the trace of  $P^t$ . To tackle the problem, a powerful idea is to use a *selective trace*.

Recall that the coefficient  $(i, j)$  of  $P^t$  is the sum over all paths of length  $t$  from  $i$  to  $j$  of the probability that the simple random walk follows this path. Instead of taking all those paths, we are going to select only those that are not “too much tangled” and replace the matrix  $P^t$  with a “tangle-free” matrix  $P^{(t)}$  — all proper definitions will be stated in Section 2.2.2 — and use the fact that with high probability, when  $t$  is not too large, there are no tangles in the original graph (Proposition 2.2.3). This idea was introduced in [79] for the proof of the Friedman theorem and was refined in [33] and [43].

In the models studied in these papers, it was easier to study paths that are non-backtracking, i.e. that do not take the same edge twice in a row. In our own model of directed graphs, no edge can be crossed twice in a row except self-loops — which are rare — hence we can concentrate on the transition matrix  $P^t$  or its tangle-free analog  $P^{(t)}$  instead of resorting to non-backtracking matrices.

The next step will be to relate the second eigenvalue of  $P^{(t)}$  with the matrix norm of different other related matrices, namely  $\underline{P}^{(t)}$  and  $R^{t, \ell}$ , defined in 2.2.2. Those matrices are easier to study, because their components are nearly centered. Their norms are given in Propositions 2.2.6 and 2.2.7.

The key difficulty of our model, compared to the regular case studied in [33], lies in the fact that the stationary distribution is unknown. In the regular case, the stationary distribution — i.e., the top left-eigenvector — is known to be  $(1/n, \dots, 1/n)$ , which could be used in Lemma 3 of [33] for deriving a Courant-Fisher-like variational formulation of  $|\lambda_2|$ . This is no longer the case here and we had to perform different algebraic manipulations and to approximate the stationary distribution; this will be done in the proof of Proposition 2.2.8 (Section 2.3).

### 2.2.2 Definitions: tangles and variants of $P$

This subsection introduces the main tools for our proof of Theorem 1.

#### Paths

Even though the graph  $G$  is a multigraph, its construction with half-edges described in Section 2.1.1 is extremely useful and will be of paramount importance in the paper. This is why we do not define

paths as a usual path in a graph (or multigraph), but as a sequence of half-edges that could be paired through  $\sigma$ . Through all the sequel,  $t > 0$  is an integer.

**DEFINITION 2.2.1.** A path of length  $t$  between two vertices  $i$  and  $j$  is a sequence of half-edges  $(\mathbf{e}_1, \mathbf{f}_1, \dots, \mathbf{e}_t, \mathbf{f}_t)$  such that

1. for every  $s \leq t$ ,  $\mathbf{e}_s$  is a head and  $\mathbf{f}_s$  is a tail,
2. for every  $s < t$ ,  $\mathbf{f}_s$  and  $\mathbf{e}_{s+1}$  are attached to the same vertex,
3.  $\mathbf{e}_1$  is attached to  $i$  and  $\mathbf{f}_t$  is attached to  $j$ .

We note  $\mathcal{P}^t(i, j)$  the set of paths of length  $t$  connecting  $i$  to  $j$ . Usually, we will denote paths by the bold letter  $\mathbf{p}$ , meaning  $\mathbf{p} = (\mathbf{e}_1, \mathbf{f}_1, \dots, \mathbf{e}_t, \mathbf{f}_t)$ .

Keep in mind that our definition of a *path* does not depend on  $\sigma$  or  $G$ : it is a *potential* path in  $G$ . The path itself is a purely combinatorial object and is not random; it will become a true path in the random graph  $G$  if in addition,  $\sigma(\mathbf{e}_s) = \mathbf{f}_s$  for every  $s \in \{1, \dots, t\}$ . In this setting we have the following useful expression for powers of the matrix  $P$ :

$$P^t(i, j) = \sum_{\mathbf{p} \in \mathcal{P}^t(i, j)} \prod_{s=1}^t \frac{\mathbf{1}_{\sigma(\mathbf{e}_s) = \mathbf{f}_s}}{d_{\mathbf{e}_s}^+} \quad (2.2.1)$$

where  $d_{\mathbf{e}}^+$  is in fact  $d_u^+$  if the half-edge  $\mathbf{e}$  is attached to the vertex  $u$  (see notation 2.1.5). When  $t = 1$ , this expression reduces to

$$P(i, j) = \sum_{\mathbf{e} \in E^+(i)} \sum_{\mathbf{f} \in E^-(j)} \frac{\mathbf{1}_{\sigma(\mathbf{e}) = \mathbf{f}}}{d_i^+}.$$

Taking expectations on both sides yields the following identity:

$$\mathbf{E}[P(i, j)] = \frac{d_j^-}{M} := \pi^-(j). \quad (2.2.2)$$

The probability distribution  $\pi^-$  is also called the *out-degree distribution*.

### Tangles and cycles

In an oriented multigraph, we say that two vertices  $u$  and  $v$  are adjacent if there is an edge between them, regardless of its orientation. A *cycle* is a sequence of vertices  $(x_1, \dots, x_n)$  such that for every  $i \neq n$ ,  $x_i$  and  $x_{i+1}$  are adjacent and  $x_n$  is adjacent with  $x_1$ . Loops and multi-edges count as cycles.

If  $G$  is an oriented multigraph and  $x, y$  are two vertices, a digraph-path from  $x$  to  $y$  is a sequence  $(x_1, \dots, x_n)$  such that  $x_1 = x, x_n = y$ , and for every  $i$  the vertex  $x_i$  leads to the vertex  $x_{i+1}$ . Its length is  $n - 1$ . We denote by  $d(x, y)$  the length of the shortest digraph-path from  $x$  to  $y$ . Let  $x$  be a vertex and  $r$  a positive integer. The forward ball of center  $x$  and radius  $r$ , noted  $B^+(x, r)$ , is the oriented multigraph induced by  $G$  on the vertices  $y$  such that  $d(x, y) \leq r$ .

We now give our first definition of *tangles*, in the context of digraphs:

- Let  $G$  be an oriented multigraph. We say that it is **tangled** if it has at least two cycles. If  $G$  is not tangled, it is tangle-free.
- Let  $d$  be a positive integer. If, for every vertex  $x$ , the oriented multigraph  $B^+(x, d)$  is tangle-free, we say that  $G$  is **d-tangle free**. Otherwise, it is  $d$ -tangled.

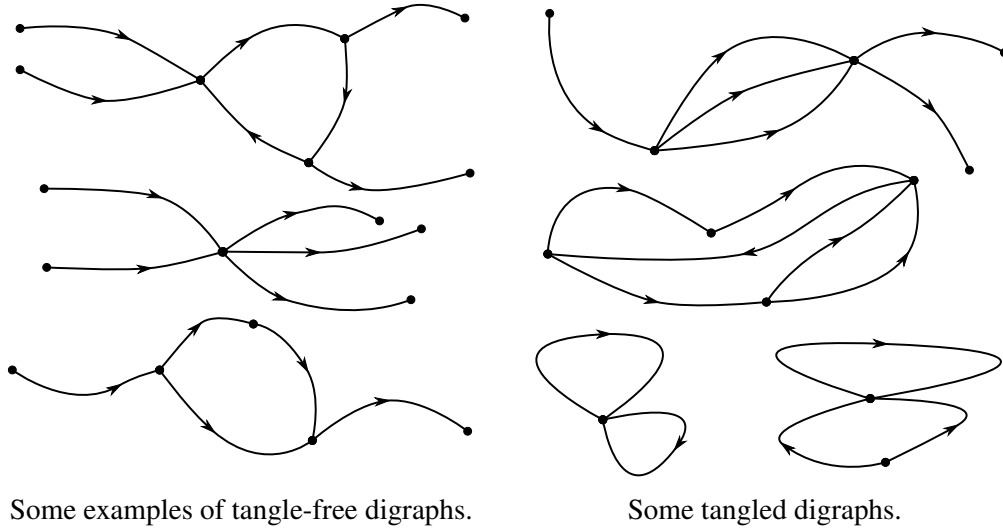


Figure 2.2 – Examples.

We now extend this to *paths*, as defined in definition 2.2.1. Fix a path  $\mathbf{p}$ . It induces an oriented multigraph  $G(\mathbf{p})$  with the following construction:

- the vertices of  $G(\mathbf{p})$  are the vertices having an half-edge appearing in  $\mathbf{p}$ ,
- the number of edges going from vertex  $x$  to vertex  $y$  is the number of *distinct* couples  $(\mathbf{e}, \mathbf{f})$  appearing in  $\mathbf{p}$ , such that  $\mathbf{e}$  is a head attached to  $x$  and  $\mathbf{f}$  is a tail attached to  $y$ .

If  $(\mathbf{e}, \mathbf{f})$  appears more than once in the path  $\mathbf{p}$ , then it will only account for one edge in  $G(\mathbf{p})$ . The definition of tangles naturally extends to paths  $\mathbf{p}$ :

**DEFINITION 2.2.2** (tangle-free paths). *Let  $\mathbf{p}$  be a path. It is tangle-free if  $G(\mathbf{p})$  is tangle free. The set of all paths of length  $t$  going from  $i$  to  $j$  that are tangle-free will be noted  $\mathcal{T}^t(i, j)$ .*

Note that a path  $\mathbf{p}$  can be tangle-free and have a cycle crossed many times. For example, fix a head  $\mathbf{e}$  and a tail  $\mathbf{f}$  attached to the same vertex  $x$ . Define the path

$$\mathbf{p} = (\mathbf{e}, \mathbf{f}, \mathbf{e}, \mathbf{f}, \mathbf{e}, \mathbf{f}).$$

The corresponding graph  $G(\mathbf{p})$  is the simple loop based at  $x$ , which has only one cycle, thus  $\mathbf{p}$  is tangle-free. However, the loop is explored three times by the path  $\mathbf{p}$ .

Now take another tail attached to  $x$ , say  $\mathbf{f}'$ . Consider the path

$$\mathbf{q} = (\mathbf{e}, \mathbf{f}, \mathbf{e}, \mathbf{f}').$$

Then  $G(\mathbf{q})$  is simply the multigraph with one vertex and two distinct loops based at  $x$ , thus  $\mathbf{q}$  is tangled.

### Variants of $P$ .

We now define:

- the **centered** analogue of  $P^t$ , which is  $\underline{P}^t$  defined by

$$\underline{P}^t(i, j) = \sum_{\mathbf{p} \in \mathcal{T}^t(i, j)} \prod_{s=1}^t \frac{\mathbf{1}_{\sigma(\mathbf{e}_s) = \mathbf{f}_s} - 1/M}{d_{\mathbf{e}_s}^+}. \quad (2.2.3)$$

Using (2.2.2), we see that the matrix  $\underline{P}^1$  is centered. This is not true for  $\underline{P}^t$ , but an important step in this work will be to prove that  $\underline{P}^t$  is *nearly* centered.

- the **tangle-free** analogue of  $P$ , defined by

$$P^{(t)}(i, j) = \sum_{\mathbf{p} \in \mathcal{T}^t(i, j)} \prod_{s=1}^t \frac{\mathbf{1}_{\sigma(\mathbf{e}_s) = \mathbf{f}_s}}{d_{\mathbf{e}_s}^+}. \quad (2.2.4)$$

Here, we just got rid of all the tangled paths. When the underlying graph is  $t$ -tangle free, we obviously have  $P^t = P^{(t)}$ .

- and finally the **centered tangle-free** analogue of  $P$ , defined by

$$\underline{P}^{(t)}(i, j) = \sum_{\mathbf{p} \in \mathcal{T}^t(i, j)} \prod_{s=1}^t \frac{\mathbf{1}_{\sigma(\mathbf{e}_s) = \mathbf{f}_s} - 1/M}{d_{\mathbf{e}_s}^+}. \quad (2.2.5)$$

The matrix  $\underline{P}^{(t)}$  is the main tool of the forthcoming analysis, because it is “nearly centered” and the sum runs over tangle-free paths. A key step in this paper will be to check if the perturbation  $P^t - \underline{P}^{(t)}$  is small: to this end, first remark that the sparsity of the graph  $G$  implies that tangles are not frequent if we choose the right scale for the path length  $t$ :

**PROPOSITION 2.2.3.** *Let  $G$  be the random graph associated with the degree sequence  $(d_i^+, d_i^-)$  satisfying hypothesis (H1). Define  $t = \lceil \alpha \log_{\Delta}(n) \rceil$ . Then, as  $n$  goes to infinity, we have*

$$\lim_{n \rightarrow \infty} \mathbf{P}(G \text{ is } t\text{-tangled}) = 0. \quad (2.2.6)$$

The proof relies on a classical breadth-first-search exploration argument and can be found in section 3.2 of [39]. In particular, under assumption (H1),  $\mathcal{T}^s = \mathcal{P}^s$  with high probability for every  $s \leq t$ , so  $P^s = P^{(s)}$ . Some related work on cycles in those random digraphs can be found in [61].

For the rest of the paper, we fix  $t$  as in the preceding proposition with  $\alpha < 1/4$ , that is

$$t = \lceil \alpha \log_{\Delta}(n) \rceil. \quad (2.2.7)$$

The parameter  $\alpha$  can be chosen arbitrarily small, as long as it is strictly smaller than  $1/4$ . This freedom will be used in Section 2.2.3.

### Tangled remainders

We finally define our last ingredient: tangles. We first need a notation for the concatenation of two paths.

**NOTATION 2.2.4** (concatenation). *If  $\mathbf{p} = (\mathbf{e}_s, \mathbf{f}_s)_{1 \leq s \leq k}$  is a path of length  $k$  and if  $\mathbf{p}' = (\mathbf{e}'_s, \mathbf{f}'_s)_{1 \leq s \leq k'}$  is a path of length  $k'$ , with  $\mathbf{f}_k$  attached to the same vertex as  $\mathbf{e}'_1$ , then the concatenation  $(\mathbf{p}, \mathbf{p}')$  will be the path of length  $k + k'$  defined by*

$$(\mathbf{e}_1, \mathbf{f}_1, \dots, \mathbf{e}_k, \mathbf{f}_k, \mathbf{e}'_1, \mathbf{f}'_1, \dots, \mathbf{e}'_{k'}, \mathbf{f}'_{k'}).$$

*This definition obviously extends to the concatenation of three or more paths, provided that the final tail of each path is attached to the same vertex as the beginning head of the next path.*

**DEFINITION 2.2.5.**  $\mathcal{R}^{t, \ell}(i, j)$  is the set of all tangled paths  $\mathbf{p}$  going from  $i$  to  $j$ , but which can be written in the form  $\mathbf{p} = (\mathbf{p}_1, \mathbf{p}_2, \mathbf{p}_3)$  where

- the path  $\mathbf{p}_1$  belongs to  $\mathcal{T}^{\ell-1}(i, g)$  where  $g$  is a vertex of the graph,
- $\mathbf{p}_2 = (\mathbf{e}, \mathbf{f})$  is a path which goes from  $g$  to  $h$  in only one step, with  $h$  a vertex of the graph,

- the path  $\mathbf{p}_3$  belongs to  $\mathcal{T}^{t-\ell}(h, j)$ .

We also define the tangled rest by

$$R^{t,\ell}(i, j) = \sum_{\mathbf{p} \in \mathcal{R}^{t,\ell}(i, j)} \prod_{s=1}^{\ell-1} A(\mathbf{e}_s, \mathbf{f}_s) \frac{1}{d_{\mathbf{e}_\ell}^+} \prod_{s=\ell+1}^t \underline{A}(\mathbf{e}_s, \mathbf{f}_s). \quad (2.2.8)$$

In other words, the set  $\mathcal{R}^{t,\ell}$  is the set of all paths that can be obtained by gluing two tangle-free paths with a bridge, but which in the end are tangled.

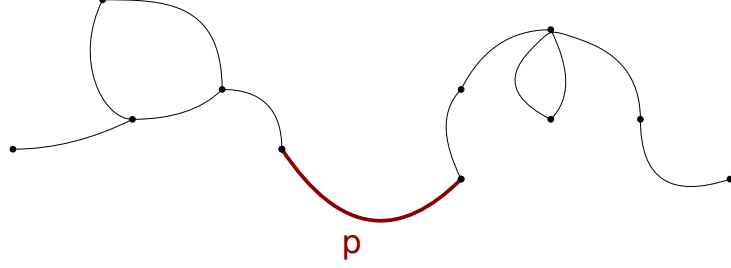


Figure 2.3 – An element in  $\mathcal{R}^{t,\ell}$ . The two black paths are tangle-free, but when we glue them together with the “bridge”  $\mathbf{p}$ , we create a tangle.

### 2.2.3 Proof of the main theorem

The main algebraic idea of the proof relies on the fact that one can bound  $|\lambda_2|$  using the operator norm of matrices  $\underline{P}^{(t)}$  and  $R^{t,\ell}$  for  $\ell \leq t$ . The core of the paper will consist in bounds for  $\|\underline{P}^{(t)}\|$  and  $\|R^{t,\ell}\|$ . Recall that  $\tilde{\rho} = \rho \vee \delta^{-1}$ .

**PROPOSITION 2.2.6.** *Let  $t$  be as in (2.2.7). For any  $c > 1$ , with high probability, we have*

$$\|\underline{P}^{(t)}\| \leq \ln(n)^D (c\tilde{\rho})^t, \quad (2.2.9)$$

where  $D$  is a positive constant.

**PROPOSITION 2.2.7.** *Let  $t$  be as in (2.2.7) and let  $\ell$  be in  $\{1, \dots, t\}$ . With high probability, we have*

$$\|R^{t,\ell}\| \leq n \ln(n)^D (c\tilde{\rho})^{t+\ell} \quad (2.2.10)$$

where  $D$  is a positive constant.

The proof of those two propositions is an application of the classical trace method and is quite technical. It will be postponed at Sections 2.4 - 2.9. We now state the central proposition for bounding the second eigenvalue of  $P$ . Its proof is exposed in Section 2.3.

**PROPOSITION 2.2.8.** *With high probability, the second eigenvalue  $\lambda_2$  of the matrix  $P$  satisfies the following inequality:*

$$|\lambda_2|^t \leq 2 \ln(n)^3 \left( \|\underline{P}^{(t)}\| + \frac{1}{M} \sum_{\ell=1}^t \|R^{t,\ell}\| \right). \quad (2.2.11)$$

We now conclude the proof of Theorem 1 from Propositions 2.2.8, 2.2.6 and 2.2.7. For simplicity, note

$$K_t = \|\underline{P}^{(t)}\| + \frac{1}{M} \sum_{\ell=1}^t \|R^{t,\ell}\|. \quad (2.2.12)$$



As a direct consequence of the two preceding theorems and the fact  $M \geq \delta n \geq n$ , it is clear that with high probability,  $K_t \leq \ln(n)^D (c\tilde{\rho})^t + \delta^{-1} (c\tilde{\rho})^t \ln(n)^D \sum_{\ell=1}^t (c\tilde{\rho})^\ell$  which is equal to  $\ln(n)^D (c\tilde{\rho})^t (1 + \delta^{-1} c\tilde{\rho} \frac{(c\tilde{\rho})^t - 1}{c\tilde{\rho} - 1})$ . If  $c$  is close enough to 1 to ensure that  $c\tilde{\rho} < 1$ , then as  $n$  goes to infinity the term  $1 + \delta^{-1} c\tilde{\rho} \frac{(c\tilde{\rho})^t - 1}{c\tilde{\rho} - 1}$  is bounded by some absolute constant  $C$ . We have proven that, with high probability,

$$K_t \leq \ln(n)^D (c\tilde{\rho})^t C. \quad (2.2.13)$$

We now use Proposition 2.2.8 which states that  $|\lambda_2|^t \leq 2 \ln(n)^3 K_t$ , hence

$$|\lambda_2|^t \leq 2C \ln(n)^{D+3} (c\tilde{\rho})^t. \quad (2.2.14)$$

Take powers  $1/t$  on both sides and use  $t = \Theta(\ln(n))$ :

$$|\lambda_2| \leq (2C \ln(n)^{D+3})^{\frac{1}{t}} c\tilde{\rho} = (1 + o(1)) c\tilde{\rho} \quad (2.2.15)$$

which finally ends the proof of (2.1.4) and Theorem 1.

### 2.2.4 Organisation of the rest of the paper

The rest of this paper is mainly devoted to the proof of Propositions 2.2.6-2.2.7. Both are inspired from [33].

1. Section 2.3 gives the proof of Proposition 2.2.8.
2. In Section 2.4, we state a lemma on correlation functions in the multigraph  $G$  that will be used in the proof of Propositions 2.2.6 and 2.2.7. This section is essentially technical and the proof of (2.4.1) is postponed to Appendix 2.11.
3. In Section 2.5, we develop the general strategy used to prove Proposition 2.2.6 which is an adaptation of the trace method. This leads to two subproblems, one purely combinatorial and one purely probabilistic. The combinatorial part (counting paths) is treated in Section 2.6 and the probabilistic one (bounding expectations) in Section 2.7.
4. Finally, the asymptotic analysis is done in Section 2.8, thus concluding the proof of Proposition 2.2.6.
5. The exact same steps are adapted to the proof of Proposition 2.2.7 in the last section.

## 2.3 Proof of Proposition 2.2.8

The method for the bound (2.2.11) is inspired from [112] and was developed in [43] and [33]. The main steps are as follows:

1. express  $P^t$  as a weighted sum of matrix products involving the tangle-free centered matrices  $\underline{P}^{(t)}$  and the tangled rest  $R^{t,\ell}$ ,
2. use this expression to make  $P^t$  appear as a perturbation of a rank 1 matrix,
3. and finally use classical results from linear algebra to link the eigenvalues of  $P^t$  with those of this perturbed matrix.

**Notation.** If  $\mathbf{e}$  is a head and  $\mathbf{f}$  is a tail, then we will adopt the following notations:

$$A(\mathbf{e}, \mathbf{f}) = \frac{\mathbf{1}_{\sigma(\mathbf{e})=\mathbf{f}}}{d_{\mathbf{e}}^+} \quad \text{and} \quad \underline{A}(\mathbf{e}, \mathbf{f}) = \frac{\mathbf{1}_{\sigma(\mathbf{e})=\mathbf{f}} - 1/M}{d_{\mathbf{e}}^+}. \quad (2.3.1)$$

With these notations, the matrix  $P^t$  has the following expression:

$$P^t(i, j) = \sum_{\mathbf{p} \in \mathcal{P}_{i,j}^t} \prod_{s=1}^t A(\mathbf{e}_s, \mathbf{f}_s)$$

### 2.3.1 Telescoping products of real numbers

If  $x_1, \dots, x_t, y_1, \dots, y_t$  are arbitrary complex numbers, we have the following ‘‘telescoping product-sum’’ :

$$\prod_{s=1}^t y_s = \prod_{s=1}^t x_s - \sum_{\ell=1}^t \prod_{s=1}^{\ell-1} y_s (x_\ell - y_\ell) \prod_{s=\ell+1}^t x_s. \quad (2.3.2)$$

Recall Definitions 2.2.3 of  $\underline{P}^t$  and Definition 2.2.5 of  $\underline{P}^{(t)}$  on page 46. We apply (2.3.2) to the matrix  $P^{(t)}$ , with  $y_s = A(\mathbf{e}_s, \mathbf{f}_s)$  and  $x_s = y_s - (Md_{\mathbf{e}_s}^+)^{-1}$ . Note that the choice (2.2.7) for  $t$  implies that  $P^t = P^{(t)}$  with high probability due to Proposition 2.2.3. Hence, with high probability,

$$P^t = P^{(t)} = \sum_{\mathbf{p} \in \mathcal{T}^t(i,j)} \prod_{s=1}^t A(\mathbf{e}_s, \mathbf{f}_s) \quad (2.3.3)$$

$$= \sum_{\mathbf{p} \in \mathcal{T}^t(i,j)} \prod_{s=1}^t \underline{A}(\mathbf{e}_s, \mathbf{f}_s) - \sum_{\mathbf{p} \in \mathcal{T}^t(i,j)} \sum_{\ell=1}^t \prod_{s=1}^{\ell-1} A(\mathbf{e}_s, \mathbf{f}_s) \left( \underline{A}(\mathbf{e}_\ell, \mathbf{f}_\ell) - A(\mathbf{e}_\ell, \mathbf{f}_\ell) \right) \prod_{s=\ell+1}^t \underline{A}(\mathbf{e}_s, \mathbf{f}_s). \quad (2.3.4)$$

By definition (see (2.3.1)), we have  $\underline{A}(\mathbf{e}_\ell, \mathbf{f}_\ell) - A(\mathbf{e}_\ell, \mathbf{f}_\ell) = -(Md_{\mathbf{e}_\ell}^+)^{-1}$ , so finally

$$P^t = \underline{P}^{(t)} - \sum_{\ell=1}^t \frac{1}{M} \sum_{\mathbf{p} \in \mathcal{T}^t(i,j)} \prod_{s=1}^{\ell-1} A(\mathbf{e}_s, \mathbf{f}_s) \frac{1}{d_{\mathbf{e}_\ell}^+} \prod_{s=\ell+1}^t \underline{A}(\mathbf{e}_s, \mathbf{f}_s). \quad (2.3.5)$$

### 2.3.2 Gluing paths and gathering the remainders

We now decompose the set  $\mathcal{T}^t(i, j)$  appearing in the sum in the right hand side of (2.3.5). Recall that the out-degree distribution  $\pi^-$  was defined in (2.2.2) on page 44.

**LEMMA 2.3.1.** *With high probability,*

$$P^t = \underline{P}^{(t)} - \sum_{\ell=1}^t P^{\ell-1} \mathbf{1}(\pi^-)^\top \underline{P}^{t-\ell} + \frac{1}{M} \sum_{\ell=1}^t R^{t,\ell}. \quad (2.3.6)$$

*Proof.* We start from (2.3.5): our main task will be to reorganize the sum

$$\frac{1}{M} \sum_{\mathbf{p} \in \mathcal{T}_{i,j}^t} \prod_{s=1}^{\ell-1} A(\mathbf{e}_s, \mathbf{f}_s) \frac{1}{d_{\mathbf{e}_\ell}^+} \prod_{s=\ell+1}^t \underline{A}(\mathbf{e}_s, \mathbf{f}_s). \quad (2.3.7)$$

We have the following decomposition when  $\ell < t$  (remind that the union over  $g, h$  is taken over all pairs of vertices):

$$\mathcal{T}^t(i, j) = \bigcup_{g,h} \{(\mathbf{p}_1, \mathbf{p}_2, \mathbf{p}_3) : \mathbf{p}_1 \in \mathcal{T}^{\ell-1}(i, g), \mathbf{p}_2 \in \mathcal{T}^1(g, h), \mathbf{p}_3 \in \mathcal{T}^{t-\ell}(h, j)\} \setminus \mathcal{R}^{t,\ell}(i, j). \quad (2.3.8)$$

Therefore, we have the following symbolic identity between sums:

$$\sum_{\mathcal{T}^t(i,j)} = \sum_g \sum_h \sum_{\mathcal{T}^{\ell-1}(i,g)} \sum_{\mathcal{T}^1(g,h)} \sum_{\mathcal{T}^{t-\ell}(h,j)} - \sum_{\mathcal{R}^{t,\ell}(i,j)}. \quad (2.3.9)$$

In the RHS, the sum over  $\mathcal{R}^{t,\ell}$  will be exactly the  $(i, j)$  entry of the matrix  $R^{t,\ell}$  (see (2.2.8)). Note that, if the path  $\mathbf{p} = (\mathbf{e}_s, \mathbf{f}_s)_{s \leq t}$  can be written in the form  $(\mathbf{p}_1, \mathbf{p}_2, \mathbf{p}_3)$  with  $\mathbf{p}_1$  in  $\mathcal{T}^{\ell-1}$  and so on as in (2.3.8), then

$$\prod_{s=1}^{\ell-1} A(\mathbf{e}_s, \mathbf{f}_s) \frac{1}{d_{\mathbf{e}_\ell}^+} \prod_{s=\ell+1}^t \underline{A}(\mathbf{e}_s, \mathbf{f}_s) = \left( \prod_{s=1}^{\ell-1} A(\mathbf{e}_s^1, \mathbf{f}_s^1) \right) \left( \frac{1}{d_{\mathbf{e}_\ell^+}^2} \right) \left( \prod_{s=1}^{t-\ell} A(\mathbf{e}_s^3, \mathbf{f}_s^3) \right) \quad (2.3.10)$$

where we noted  $\mathbf{p}_1 = (\mathbf{e}_1^1, \mathbf{f}_1^1, \dots, \mathbf{e}_{\ell-1}^1, \mathbf{f}_{\ell-1}^1)$  and so on. With the same notations, we plug this into the five sums found above:

$$\begin{aligned} \frac{1}{M} R^{t,\ell}(i,j) + \frac{1}{M} \sum_{\mathbf{p} \in \mathcal{I}^t(i,j)} \prod_{s=1}^{\ell-1} A(\mathbf{e}_s, \mathbf{f}_s) \frac{1}{d_{\mathbf{e}}^+} \prod_{s=\ell+1}^t \underline{A}(\mathbf{e}_s, \mathbf{f}_s) = \\ \sum_{g,h} \left( \sum_{\mathbf{p}_1 \in \mathcal{I}^{\ell-1}(i,g)} \prod_{s=1}^{\ell-1} A(\mathbf{e}_s^1, \mathbf{f}_s^1) \right) \left( \sum_{\mathbf{p}_2 \in \mathcal{I}^1(g,h)} \frac{1}{M d_{\mathbf{e}}^+} \right) \left( \sum_{\mathbf{p}_3 \in \mathcal{I}^{t-\ell}(h,j)} \prod_{s=1}^{t-\ell} \underline{A}(\mathbf{e}_s^3, \mathbf{f}_s^3) \right) \end{aligned} \quad (2.3.11)$$

This is a matrix product : the first and third parentheses are  $P^{(\ell-1)}(i,h)$  and  $\underline{P}^{(t-\ell)}(h,j)$ . The term in the middle is equal to  $\sum_{\mathbf{e} \in E^+(g), \mathbf{f} \in E^-(h)} \frac{1}{M d_g^+}$  which simplifies to  $d_h^- / M = \pi^-(h)$ . We define  $X(g,h) = \pi^-(h)$  — note the useful identity  $X = \mathbf{1}(\pi^-)^\top$ . The RHS of (2.3.11) then becomes

$$\sum_{g,h} P^{(\ell-1)}(i,h) X(g,h) \underline{P}^{(t-\ell)}(h,j) = (P^{(\ell-1)} X \underline{P}^{(t-\ell)})(i,j) \quad (2.3.12)$$

and the whole expression (2.3.7) becomes equal to  $(P^{(\ell-1)} X \underline{P}^{(t-\ell)})(i,j) - M^{-1} R^{t,\ell}(i,j)$ . Putting it back in (2.3.5), we get

$$P^t = \underline{P}^{(t)} - \sum_{\ell=1}^t P^{(\ell-1)} X \underline{P}^{(t-\ell)} + \frac{1}{M} \sum_{\ell=1}^t R^{t,\ell}$$

which is exactly the claim in the lemma because because (due to Proposition 2.2.3), with high probability we have  $\underline{P}^{(t-\ell)} = \underline{P}^{t-\ell}$  and  $P^\ell = P^{(\ell)}$ .  $\square$

### 2.3.3 Expressing $P$ as a perturbation of a rank 1 matrix

We first define two real vectors  $x, y \in \mathbb{R}^n$  by

$$x = \mathbf{1}, \quad y = \frac{1}{n} (P^t)^\top x \quad (2.3.13)$$

and we recall the definition of  $K_t$  given in (2.2.12):

$$K_t = \|\underline{P}^{(t)}\| + \frac{1}{M} \sum_{\ell=1}^t \|R^{t,\ell}\|.$$

Note the presence of the important  $M^{-1}$  factor in the right. The following lemma is crucial: it quantifies the distance between the matrix  $P^t$  and a rank-1 matrix, namely  $xy^\top$ .

**LEMMA 2.3.2.** *With high probability,*

$$\|P^t - xy^\top\| \leq K_t. \quad (2.3.14)$$

*Proof.* Let  $f$  be a vector such that  $\langle f, \mathbf{1} \rangle = 0$ ; multiply (2.3.6) to the left by  $f^\top$  to get

$$f^\top P^t = f^\top \underline{P}^{(t)} - \sum_{\ell=1}^t f^\top P^{\ell-1} \mathbf{1} (\pi^-)^\top \underline{P}^{t-\ell} + \frac{1}{M} \sum_{\ell=1}^t f^\top R^{t,\ell}. \quad (2.3.15)$$

The matrix  $P^{\ell-1}$  is a Markov matrix, therefore  $P^{\ell-1} \mathbf{1} = \mathbf{1}$  and the product  $f^\top P^{\ell-1} \mathbf{1} (\pi^-)^\top \underline{P}^{t-\ell}$  vanishes. We get the fundamental inequality

$$\|(P^t)^\top f\| = \|f^\top P^t\| \leq \left( \|\underline{P}^{(t)}\| + \frac{1}{M} \sum_{\ell=1}^t \|R^{t,\ell}\| \right) \cdot \|f\| = K_t \|f\|. \quad (2.3.16)$$

Let us momentarily note  $Q = P^t - xy^\top$  so that

$$P^t = xy^\top + Q. \quad (2.3.17)$$

These choices imply the crucial following observation:  $x^\top P^t = x^\top xy^\top + x^\top Q = ny^\top + x^\top Q$ . But as  $x^\top P^t = ny^\top$  we get  $x^\top Q = 0$ . Hence,  $Q$  vanishes when multiplied on the left by  $\mathbf{1}$ . Let  $v$  be any unit vector: there is a real number  $\alpha$  and a vector  $f$  with  $\langle f, x \rangle = 0$  such that  $v = f + \alpha x$ . The triangle inequality implies  $\|v^\top Q\| \leq \|f^\top Q\| + \alpha \|x^\top Q\| = \|f^\top Q\|$ , hence  $\|Q\| \leq \sup \|f^\top Q\| / \|Q\|$ , where the supremum is taken over all nonzero vectors  $f$  such that  $\langle f, x \rangle = 0$ . Moreover, as  $\langle f, x \rangle = 0$  we have  $f^\top P^t = f^\top xy^\top + f^\top Q = f^\top Q$ . Putting all these observations together with (2.3.16) yields the following:

$$\begin{aligned} \|Q\| &\leq \sup_{\langle f, \mathbf{1} \rangle = 0} \frac{\|f^\top Q\|}{\|f\|} \\ &\leq \sup_{\langle f, \mathbf{1} \rangle = 0} \frac{\|f^\top P^t\|}{\|f\|} \\ &\leq K_t \end{aligned}$$

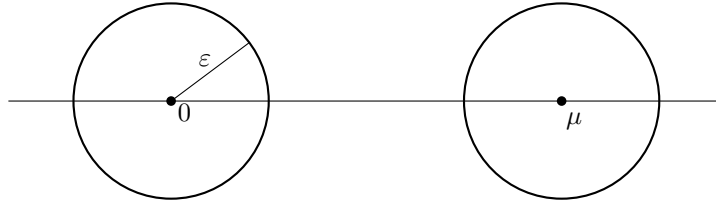
which is exactly the claim in the lemma.  $\square$

### 2.3.4 Classical algebra to link the eigenvalues of $P$ with those of $xy^\top$

The main ingredient for the proof of Proposition 2.2.8 will be the following basic algebraic lemma (see Appendix 2.10 for a complete proof of this result).

**LEMMA 2.3.3** (eigenvalue perturbation for rank 1 matrices). *Let  $H, M$  be two real  $n \times n$  matrices, with  $M$  diagonalizable with rank 1. Let  $x, y$  be two vectors such that  $M = xy^\top$ . Define  $\mu = \langle x, y \rangle$ .*

1. *The eigenvalues of  $M + H$  lie in the union of the two balls  $B(0, \varepsilon)$  and  $B(\mu, \varepsilon)$ , with  $\varepsilon = 2\|x\|^2\|y\|^2\mu^{-2}\|H\|$ .*



2. *If  $B(0, \varepsilon) \cap B(\mu, \varepsilon) = \emptyset$ , then there is exactly one eigenvalue of  $M + H$  inside  $B(\mu, \varepsilon)$  and all the other eigenvalues of  $M + H$  are contained in  $B(0, \varepsilon)$ .*

*Proof of Proposition 2.2.8.* Let  $x, y$  be as in (2.3.13). We apply Lemma 2.3.3 to the matrix  $P^t = xy^\top + Q$ . First of all, note that  $\mu = \langle x, y \rangle = \langle P^t x, x/n \rangle = \langle x, x/n \rangle = 1$ . All the eigenvalues of  $P^t$  lie in the union of the two balls  $B(0, \varepsilon)$  and  $B(\langle x, y \rangle, \varepsilon)$  where  $\varepsilon$  is smaller than

$$\frac{2\|x\|^2\|y\|^2}{\langle x, y \rangle^2} K_t = 2\|x\|^2\|y\|^2 K_t.$$

We clearly have  $\|x\| = \sqrt{n}$ . We should now have a control over the norm of  $y$ . Note that  $\|y\|^2 = \sum_{i=1}^n (\pi_0^\top P^t)_i^2$  where  $\pi_0$  is the uniform measure over the vertices of the graph (i.e.  $\pi_0(v) = 1/n$ ); hence,  $\pi_0^\top P^t$  can be interpreted as the distribution of the Markov chain after  $t$  steps on the directed graph  $G$  when started from a uniform vertex. In particular, for every  $i$  the term  $(\pi_0^\top P^t)_i^2$  is equal to  $\mathbf{P}(X_t = Y_t = i)$  when  $X, Y$  are two independent Markov chains, each one being independently started from a

uniform vertex. We will note  $\mathbf{P}', \mathbf{E}'$  the probability and expectation of the Markov chain conditionally on  $G$ . The overall term  $\|y\|^2$  is thus equal to  $\mathbf{P}'(X_t = Y_t)$ . An elegant argument from [39, Section 4] shows that

$$\mathbb{P}(X_t = Y_t) = O\left(\frac{\ln(n)^2}{n}\right) = o\left(\frac{\ln(n)^3}{n}\right) \quad (2.3.18)$$

where  $\mathbb{P}$  denotes the so-called *annealed probability*, that is the probability according to both the environment and the walk:  $\mathbb{P}(X_t = Y_t) = \mathbf{E}[\mathbf{P}'(X_t = Y_t)]$ . Using the Markov inequality with  $\mathbf{P}$ , (2.3.18) yields that with high probability,

$$\|y\| = \sqrt{\mathbf{P}(X_t = Y_t)} \leq \sqrt{\frac{\ln(n)^3}{n}}.$$

Finally, with high probability we have  $\|x\|^2 \|y\|^2 \leq \ln(n)^3$ , hence  $\varepsilon \leq 2\ln(n)^3 K_t$ .

We now use the second part of Lemma 2.3.3. To this end, we have to check that the two balls  $B(0, \varepsilon)$  and  $B(1, \varepsilon)$  are disjoint, at least when  $n$  is big. It is easy to see that

$$\varepsilon = O(\ln(n)^{D+3} (c\tilde{\rho})^t), \quad (2.3.19)$$

see for instance the short computations on page 48 leading to (2.2.13). As a consequence of (H1), we also get  $\tilde{\rho} < 1$  so if  $c$  is close enough to 1, then  $c\tilde{\rho} < 1$  and  $\varepsilon$  goes to 0 as  $n$  goes to infinity. The two balls  $B(0, \varepsilon)$  and  $B(1, \varepsilon)$  are thus disjoint. Using the second point of Lemma 2.3.3, exactly one eigenvalue of  $P^t$  is inside the ball  $B(1, \varepsilon)$  and this eigenvalue is obviously 1 because  $P^t$  is a transition matrix. All the other eigenvalues, and in particular  $\lambda_2$ , are in  $B(0, \varepsilon)$ .  $\square$

## 2.4 Expectation of a product of centered random variables

In this technical section, we present a method for obtaining upper bounds on the expectations of a product having the form  $\prod_{s \in I} (\mathbf{1}_{E_s} - \mathbf{P}(E_s))$  when the events  $E_s$  are nearly independant for most of them, and strongly dependent for a few ones. The general setting is the same as before. Such expectations will appear in the proofs of Propositions 2.2.6 and 2.2.7.

For the sake of clarity in the following sections, we need a definition of “potential paths”, i.e. collections of half-edges that are not paths, but who could give rise to real paths in the graph. Those are called *proto-paths*:

**DEFINITION 2.4.1.** A proto-path is a sequence  $\mathfrak{p} = (\mathbf{e}_1, \mathbf{f}_1, \dots, \mathbf{e}_N, \mathbf{f}_N)$  with  $N$  an integer, such that for every  $s$  in  $\{1, \dots, N\}$ ,  $\mathbf{e}_s$  is a head and  $\mathbf{f}_s$  is a tail.

There is no restriction whatsoever on the half-edges of a proto-path. Indeed, a proto-path is meant to be a path in the graph  $G$ , but it is not necessarily a path: some half-edge could appear twice or more in  $\mathfrak{p}$ , there is no vertex-consistency statement.

We are interested in computing different probabilistic quantities depending on  $\mathfrak{p}$ , the simplest of them being the probability of the event “for all  $s$ , the head  $\mathbf{e}_s$  is matched with the tail  $\mathbf{f}_s$ ”.

Fix some integer  $p$  smaller than  $N$ . Recall that  $\underline{A}$  and  $A$  had been defined in (2.3.1). We define a function  $F_p$  by

$$F_p(\mathfrak{p}) = \mathbf{E} \left[ \prod_{s=1}^p \underline{A}(\mathbf{e}_s, \mathbf{f}_s) \prod_{s=p+1}^N A(\mathbf{e}_s, \mathbf{f}_s) \right]$$

Most of the times, the index  $p$  will be dropped and we will just note  $F$ . We introduce several useful definitions and notations.

- We will note  $B(\mathbf{e}, \mathbf{f}) = \mathbf{1}_{\sigma(\mathbf{e})=\mathbf{f}} - 1/M$  and  $B'(\mathbf{e}, \mathbf{f}) = \mathbf{1}_{\sigma(\mathbf{e})=\mathbf{f}}$ . This implies  $\underline{A}(\mathbf{e}, \mathbf{f}) = B(\mathbf{e}, \mathbf{f})/d_{\mathbf{e}}^+$ .

- An edge of  $\mathfrak{p}$  is a couple  $(\mathbf{e}_s, \mathbf{f}_s)$  appearing in  $\mathfrak{p}$ .
- $a$  is the number of distinct edges appearing in the proto-path  $\mathfrak{p}$ :

$$a = \#\{(\mathbf{e}_s, \mathbf{f}_s) : 1 \leq s \leq N\}.$$

We will denote those edges by  $y_1, \dots, y_a$ .

- For each  $i \in \{1, \dots, a\}$ , the weight  $w_i$  of edge  $y_i$  is the number of times edge  $y_i$  is visited by the proto-path before  $p$  and  $w'_i$  is the number of times edge  $y_i$  is visited after  $p$ :

$$w_i = \#\{s \leq p : (\mathbf{e}_s, \mathbf{f}_s) = y_i\} \quad w'_i = \#\{s > p : (\mathbf{e}_s, \mathbf{f}_s) = y_i\}.$$

- If  $y_i = (\mathbf{e}, \mathbf{f})$ , we will note  $B(y_i)$  or  $\underline{A}(y_i)$  instead of  $B(\mathbf{e}, \mathbf{f})$  or  $\underline{A}(\mathbf{e}, \mathbf{f})$ .
- The *weight* of the proto-path  $\mathfrak{p}$  is

$$\omega(\mathfrak{p}) = \prod_{s=1}^N \frac{1}{d_{\mathbf{e}_s}^+}.$$

- Call an edge  $y_i$  *consistent* if both of its end-half-edges appear only once in the proto-path  $\mathfrak{p}$ . Call an edge *simple* if its weight is 1. If an edge is not consistent, it is *inconsistent*. If the edge  $(\mathbf{e}, \mathbf{f})$  is inconsistent, there is another edge  $(\mathbf{e}', \mathbf{f}')$  in the proto-path such that  $\{\mathbf{e}, \mathbf{f}\} \cap \{\mathbf{e}', \mathbf{f}'\} \neq \emptyset$ .

The main result of this section is the following theorem.

**Theorem 3.** *Let  $\mathfrak{p}$  be any proto-path of length  $N \leq \sqrt{M}$ ,  $p$  an integer smaller than  $N$ , and let  $a_1$  be the number of simple, consistent edges of  $\mathfrak{p}$ , before  $p$ . Also, let  $b$  be the number of inconsistent edges of  $\mathfrak{p}$ . Then, for every  $c > 1$ , there is an integer  $n_0$  such that if  $n$  is larger than  $n_0$ , we have*

$$|F(\mathfrak{p})| \leq 24 \cdot \omega(\mathfrak{p}) 3^b \left(\frac{c}{M}\right)^a \left(\frac{N}{\sqrt{M}}\right)^{a_1}. \quad (2.4.1)$$

The proof of Theorem 3 is essentially technical and is a mere adaptation of [33]. The complete proof can be found in Appendix 2.11.

## 2.5 General strategy and definitions for the proof of Proposition 2.2.6

In this section, we study the quantity  $\|\underline{P}^{(t)}\|$  for the choice of  $t = \lfloor \alpha \log_{\Delta}(n) \rfloor$  as in (2.2.7). For the rest of the paper, we set

$$m = \left\lfloor \frac{\ln(n)}{50 \ln \ln(n)} \right\rfloor. \quad (2.5.1)$$

### 2.5.1 A simplified version of Proposition 2.2.6

In order to prove Proposition 2.2.6, we are going to prove the following lemma.

**LEMMA 2.5.1.** *Fix  $t$  as in (2.2.7) and  $m$  as in (2.5.1). Fix  $c$  close to 1 and  $\tilde{\rho} = \rho \vee \delta^{-1}$ . When  $n$  is large enough, we have*

$$\mathbf{E}[\|\underline{P}^{(t)}\|^{2m}] = o(1)n^3(c\tilde{\rho})^{2tm}. \quad (2.5.2)$$

*Proof of Proposition 2.2.6 using (2.5.2).* For any constant  $D$ ,

$$\mathbf{P}(\|\underline{P}^{(t)}\| > \ln(n)^D (c\tilde{\rho})^t) \leq \frac{\mathbf{E}[\|\underline{P}^{(t)}\|^{2m}]}{(\ln(n)^D)^{2m} (c\tilde{\rho})^{2tm}} \quad (2.5.3)$$

$$\leq \frac{o(1)n^3}{(\ln(n)^D)^{2m}}. \quad (2.5.4)$$

$$(2.5.5)$$

Now, the choice of  $D = 50 \times 3/2$  yields  $\ln(n)^{2Dm} \sim n^3$ , and  $\mathbf{P}(\|\underline{P}^{(t)}\| > \ln(n)^D (c\tilde{\rho})^t) = o(1)$ .  $\square$

Before going further in the application of the trace method, we gather here some basic consequences of the choice  $m = \Theta(\ln(n)/\ln\ln(n))$  as in (2.5.1). They will be used several times in the forthcoming analysis without necessary reference.

**LEMMA 2.5.2.** *For any  $m = \Theta\left(\frac{\ln(n)}{\ln\ln(n)}\right)$  and any  $c_n > 0$  such that  $\ln(c_n) = o(\ln\ln(n))$  we have  $(c_n)^m = n^{o(1)}$ . In particular, for any constant  $c > 0$  we have  $c^m = n^{o(1)}$ .*

*For any  $A > 0$  and  $m = \frac{A\ln(n)}{\ln\ln(n)}$  and any  $t_n = O(\ln(n))$  we have  $(t_n)^m \leq n^{A+o(1)}$ .*

*For any  $A > 0$  and  $m = \frac{A\ln(n)}{\ln\ln(n)}$  and any  $t_n = O(\ln(n)^B)$  we have  $(t_n)^m \leq n^{AB+o(1)}$ .*

## 2.5.2 Using the classical trace method

The proof of (2.5.2) relies on the trace method. To somewhat lighten the notations, we will note  $X = \underline{P}^{(t)}$  in this paragraph. From now on we will choose an even integer  $r = 2m$ , so that  $\|X\|^{2m} = \|X^*X\|^m$ . As  $X^*X$  is symmetric, we have

$$\|X\|^{2m} \leq \text{tr}((X^*X)^m) = \sum_{i_1, \dots, i_m} \prod_{s=1}^m (X^*X)_{i_s, i_{s+1}} \quad (2.5.6)$$

$$= \sum_{i_1, i_2, \dots, i_{2m}} \prod_{s=1}^m X_{i_{2s-1}, i_{2s}} X_{i_{2s+1}, i_{2s}} \quad (2.5.7)$$

where we adopted the cyclic notation  $i_{m+1} = i_1$  in the first line and  $i_{2m+1} = i_1$  in the second line. With  $\underline{P}^{(t)}$  this becomes

$$\|\underline{P}^{(t)}\|^{2m} \leq \sum_{i_1, \dots, i_{2m}} \prod_{s=1}^m \underline{P}^{(t)}(i_{2s-1}, i_{2s}) \underline{P}^{(t)}(i_{2s+1}, i_{2s}). \quad (2.5.8)$$

Developping according to the definition of  $\underline{P}^{(t)}$ , we get

$$\prod_{s=1}^m X_{i_{2s-1}, i_{2s}} X_{i_{2s+1}, i_{2s}}^* = \sum_{\mathbf{p}_1 \in \mathcal{F}^t(i_1, i_2)} \sum_{\mathbf{p}_2 \in \mathcal{F}^t(i_3, i_2)} \dots \sum_{\mathbf{p}_{2m} \in \mathcal{F}^t(i_1, i_{2m})} \prod_{i=1}^{2m} \prod_{s=1}^t \underline{A}(\mathbf{e}_{i,s}, \mathbf{f}_{i,s}) \quad (2.5.9)$$

where we noted  $\mathbf{p}_i = (\mathbf{e}_{i,s}, \mathbf{f}_{i,s})_{s \leq t}$  the  $i$ -th path in the ‘‘path of paths’’  $\mathbf{p} = (\mathbf{p}_1, \dots, \mathbf{p}_{2m})$  (remember the concatenation notation 2.2.4). We define  $\mathcal{C}_m$  as the set of ‘‘paths of paths’’ corresponding to the sum, that is  $2m$ -tuples  $(\mathbf{p}_1, \dots, \mathbf{p}_{2m})$  such that  $\mathbf{p}_1$  and  $\mathbf{p}_2$  have the same endpoint,  $\mathbf{p}_2$  and  $\mathbf{p}_3$  have the same beginning point, and so on. For the following analysis, it will be easier to ‘‘reverse’’ all odd paths in  $\mathbf{p}$ , leading to the following central definition:

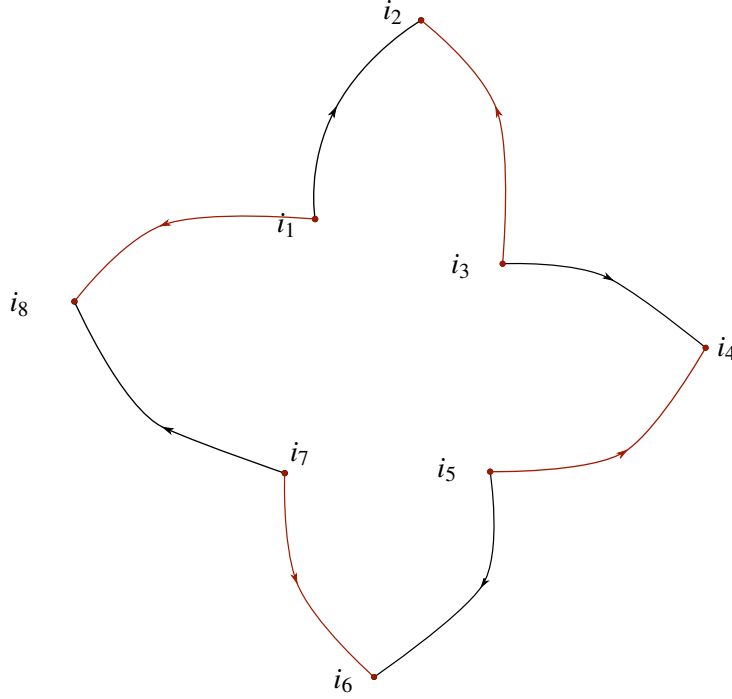


Figure 2.4 – A path in  $\mathcal{C}_4$ . The red paths are the "odd" paths, corresponding to "reversed tangle-free paths". The black ones are "even" paths.

**DEFINITION 2.5.3.**  $\mathcal{C}_m$  is the set of  $2m$ -tuples  $\mathbf{p} = (\mathbf{p}_1, \dots, \mathbf{p}_{2m})$  such that

- for every  $i$ , the path  $\mathbf{p}_{2i-1}$  is in  $\mathcal{T}^t$  and the "reversed path"

$$\bar{\mathbf{p}}_{2i} = (\mathbf{f}_{2i,t}, \mathbf{e}_{2i,t}, \dots, \mathbf{f}_{2i,1}, \mathbf{e}_{2i,1})$$

is in  $\mathcal{T}^t$ .

- For every  $i$ , the last half-edge of  $\mathbf{p}_i$  and the first half-edge of  $\mathbf{p}_{i+1}$  are attached to the same vertex (boundary condition).

Note that there is a little lack of consistency with our convention that  $\mathbf{e}$  denotes heads and  $\mathbf{f}$  denotes tails, for in this case  $\mathbf{e}_{2i,s}$  denotes a tail and  $\mathbf{f}_{2i,s}$  denotes a head. For every element  $\mathbf{p} \in \mathcal{C}_m$ , we note

$$f(\mathbf{p}) = \mathbf{E} \left[ \prod_{i=1}^m \prod_{s=1}^t \underline{A}(\mathbf{e}_{2i-1,s}, \mathbf{f}_{2i-1,s}) \prod_{s=1}^t \underline{A}(\mathbf{f}_{2i,s}, \mathbf{e}_{2i,s}) \right]. \quad (2.5.10)$$

We have obtained the following fundamental inequality:

$$\mathbf{E} \left[ \|\underline{A}^{(t)}\|^{2m} \right] \leq \sum_{\mathbf{p} \in \mathcal{C}_m} |f(\mathbf{p})|. \quad (2.5.11)$$

In the last expression, the probabilistic part, which is contained in the function  $f$ , is entirely decoupled from the combinatoric part, which is contained in the set  $\mathcal{C}_m$ . Both parts will be separately treated in the forthcoming analysis.

### 2.5.3 Geometry of paths in $\mathcal{C}_m$

We now introduce some definitions that will be commonly used in the sequel. Let  $\mathbf{p}$  be any element in  $\mathcal{C}_m$ . It induces a walk on the vertices of the graph  $G$ . We will note  $V(\mathbf{p})$  (or generally  $V$  if there is no ambiguity) the set of all visited vertices, and  $v = v(\mathbf{p}) = \#V(\mathbf{p})$ . Any  $\mathbf{p} \in \mathcal{C}_m$  is composed of  $2m$  path of length  $t$ , hence we have  $v \leq 2tm$ .



**DEFINITION 2.5.4.** We had already defined an edge of  $\mathbf{p}$  as any pair of a head followed by a tail appearing in one of the  $\mathbf{p}_i$ 's (for example  $(\mathbf{e}_{1,s}, \mathbf{f}_{1,s})$  or  $(\mathbf{f}_{2,s}, \mathbf{e}_{2,s})$ ) A graph edge is the corresponding (oriented) edge between vertices.

**EXAMPLE 2.5.5.** Let  $(\mathbf{e}, \mathbf{f})$  be an edge of  $\mathbf{p}$ , with  $\mathbf{e}$  a head and  $\mathbf{f}$  a tail. If  $\mathbf{e}$  is attached to vertex  $u$  and  $\mathbf{f}$  to vertex  $u'$ , then the corresponding graph edge will be  $(u, u')$ . Thus, each graph-edge  $(u, v)$  corresponds to at most  $d_u^+ d_{u'}^-$  distinct edges.

We will note  $E(\mathbf{p})$  the set of edges. The total number of distinct edges will be noted  $a = a(\mathbf{p}) = \#E(\mathbf{p})$ . Any  $\mathbf{p} \in \mathcal{C}_m$  induces an oriented multigraph on the set of vertices  $V(\mathbf{p})$ : its edges are just the graph edges of  $\mathbf{p}$ , counted with multiplicities. Let us call  $\vec{G}(\mathbf{p})$  this oriented multigraph; the corresponding *unoriented multigraph*  $G(\mathbf{p})$  is connected. We will note  $\chi = \chi(\mathbf{p}) = a - v + 1$  the **tree excess** of  $G(\mathbf{p})$ . This quantity will be used many times in the sequel.

## 2.6 Combinatorics of $\mathcal{C}_m$

We split  $\mathcal{C}_m$  in various disjoint subsets, taking into account the number of visited vertices and also the number of edges. The counting argument is inspired from [33] which itself stems from the seminal paper [81].

**DEFINITION 2.6.1.** Let  $a, v$  be integers and let  $\mathbf{i} = (i_1, \dots, i_v)$  be a  $v$ -tuple of vertices. We define

$$X_m^{v,a}(\mathbf{i}) = X_m^{v,a}(i_1, \dots, i_v)$$

as the set of all the elements in  $\mathcal{C}_m$  whose vertex set is precisely  $(i_1, \dots, i_v)$  (visited in this order) and who have  $a$  edges.

The aim of this section is to prove the following result on the number of elements in  $X_m^{v,a}(\mathbf{i})$ .

**PROPOSITION 2.6.2.** Fix  $v, \mathbf{i}$  and  $a$ . Recall that  $\chi = a - v + 1$ . Then, there is a constant  $C > 0$  and an integer  $n_1$  such that for every  $n \geq n_1$ , we have

$$\#X_m^{v,a}(\mathbf{i}) \leq \left( \prod_{i \in \mathbf{i}} d_i^+ d_i^- \right) C^\chi n^{\frac{25}{50} + \frac{17}{50}\chi}. \quad (2.6.1)$$

The core tool for the proof of (2.6.1) will be a simple partition of the elements of  $\#X_m^{v,a}(\mathbf{i})$  with the following notion of equivalence:

**DEFINITION 2.6.3.** Let  $\mathbf{p}$  and  $\mathbf{p}'$  be two elements in  $\mathcal{C}_m$ ; we note  $\mathbf{e}_{i,s}, \mathbf{f}_{i,s}$  the half-edges of  $\mathbf{p}$  and  $\mathbf{e}'_{i,s}, \mathbf{f}'_{i,s}$  those of  $\mathbf{p}'$ . They are said **equivalent** if

- they both belong to  $X_m^{a,v}(\mathbf{i})$  and they visit the same vertices at the same time,
- for every vertex  $u \in \mathbf{i}$ , there are two permutations  $\sigma_u \in \mathfrak{S}_{d_u^+}$  and  $\tau_u \in \mathfrak{S}_{d_u^-}$  such that for every  $i$  and  $s$ , if  $\mathbf{e}_{i,s}$  is a head attached to  $u$  and  $\mathbf{f}_{i,s}$  a tail attached to  $u$ , then

$$\mathbf{e}_{i,s} = \sigma_u(\mathbf{e}'_{i,s}) \quad \text{and} \quad \mathbf{f}_{i,s} = \tau_u(\mathbf{f}'_{i,s}).$$

In other words, two elements of  $\mathcal{C}_m$  are equivalent if they only differ by a permutation of their half-edges.

The proof is organized as follows:

- In 2.6.1, we prove an upper bound for the number of elements within each equivalence class.
- In 2.6.2, we prove an upper bound for the number of equivalence classes.
- In 2.6.3 we prove Proposition 2.6.2.

### 2.6.1 Cardinal of equivalence classes

Let  $\mathbf{p}$  be an element of  $X_m^{v,a}(\mathbf{i})$ . How many elements of  $\mathcal{C}_m$  are equivalent to  $\mathbf{p}$ ? The vertices are fixed so there is no choice from this part. We have to choose the half-edges. If there is exactly one tail and one head attached to each of these vertices, we would have  $d_{i_1}^+$  choices for the first head, then  $d_{i_2}^-$  for the first tail, and so on until the last head with  $d_{i_v}^+$  choices and the last tail with  $d_{i_1}^-$  choices. Thus, we have at most  $\prod_{i \in \mathbf{i}} d_i^+ d_i^-$  paths equivalent with  $\mathbf{p}$  in this case. In the general case, there are some vertices with *more* than one half-edge visited by  $\mathbf{p}$  attached to these vertices.

**LEMMA 2.6.4.** *Let  $\mathbf{p}$  be in  $X_m^{v,a}(\mathbf{i})$ . Note  $\alpha_s$  the number of heads visited by  $\mathbf{p}$  attached to the vertex  $i_s$ , and let  $\beta_s$  be the same with tails. Then, we have at most*

$$C^\chi \prod_{i \in \mathbf{i}} d_i^+ d_i^- \quad (2.6.2)$$

elements in  $\mathcal{C}_m$  equivalent to  $\mathbf{p}$ , where  $C > 0$  is a constant.

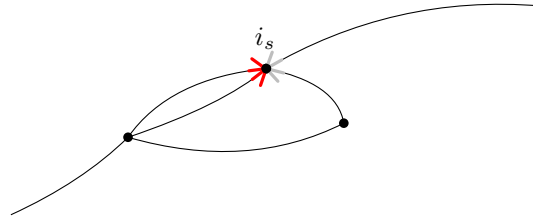


Figure 2.5 – Here, we have  $d_i = 4$ , but  $\alpha_i = 2$  and  $\beta_i = 2$ .

In the proof we will make use of the Pochhammer symbol: if  $a$  is a real number and  $k$  and integer, then  $(a)_k = a(a-1)\dots(a-k+1)$ .

*Proof.* Fix  $\mathbf{p}$ . When choosing equivalent elements to  $\mathbf{p}$ , we have *at most*

$$\prod_{i \in \mathbf{i}} (d_i^+)^{\alpha_i} (d_i^-)^{\beta_i} = \prod_{i \in \mathbf{i}} d_i^+ d_i^- \prod_{i \in \mathbf{i}} (d_i^+ - 1)^{\alpha_i - 1} (d_i^- - 1)^{\beta_i - 1}$$

choices, with the convention that a product over an empty set is equal to 1. We also have  $(d_i^+ - 1)^{\alpha_i - 1} \leq (\Delta - 1)^{\alpha_i - 1}$  and  $(d_i^- - 1)^{\beta_i - 1} \leq (\Delta - 1)^{\beta_i - 1}$ , so if we set  $K_t^+ = \#\{i \in \mathbf{i} : \alpha_i = t\}$  and  $K_t^- = \#\{i \in \mathbf{i} : \beta_i = t\}$  we have

$$\begin{aligned} \prod_{i \in \mathbf{i}} (d_i^+ - 1)^{\alpha_i - 1} (d_i^- - 1)^{\beta_i - 1} &= \prod_{t \geq 1} \prod_{i \in K_t^+} (\Delta - 1)^{t-1} \prod_{i \in K_t^-} (\Delta - 1)^{t-1} \\ &\leq \prod_{t \geq 1} (\Delta - 1)^{(t-1)(K_t^+ + K_t^-)} \\ &\leq (\Delta - 1)^{\sum_{t \geq 1} (t-1)K_t^+ + (t-1)K_t^-}. \end{aligned}$$

Counting edges going out of every vertex yields  $\sum_t t K_t^+ = a$  and counting vertices according to the number of edges going out this vertex gives  $\sum_t K_t^+ = v$  (the same holds for  $K_t^-$ ), so we get  $\sum_{t \geq 1} (t-1)K_t^+ = \sum_{t \geq 1} (t-1)K_t^- = a - v$ , and

$$\prod_{i \in \mathbf{i}} (d_i^+ - 1)^{\alpha_i - 1} (d_i^- - 1)^{\beta_i - 1} \leq (\Delta - 1)^{2(a-v)} \leq C^\chi$$

where  $C = (\Delta - 1)^2$ , thus closing the proof of (2.6.2).  $\square$

### 2.6.2 Number of equivalence classes

Now, we count the number of equivalence classes in  $X_m^{a,v}(\mathbf{i})$ . The result of this paragraph is:

**LEMMA 2.6.5.** *There is an integer  $n_1$  such that for every  $n \geq n_1$ , the total number of equivalence classes of paths in  $\mathcal{C}_m$  visiting vertices  $(i_1, \dots, i_v)$  and having  $a$  edges is bounded by*

$$n^{\frac{25}{50} + \frac{17}{50}\chi}. \quad (2.6.3)$$

We now prove this lemma. The explored vertices are  $\mathbf{i} = (i_1, \dots, i_v)$ , in this order. Recall Notation 2.1.5: half-edges are noted  $(u, i, \varepsilon)$  with  $\varepsilon \in \{-, +\}$  and  $i \leq d_u^\varepsilon$ . We first describe a coding pattern for the equivalence classes (in Paragraphs 2.6.2-2.6.2) and then prove (2.6.3) in Paragraph 2.6.2.

#### Choice of the path.

In any equivalence class, we choose a  $\mathbf{p}$  visiting heads and tails in the ‘‘alternating lexicographic order’’, that is

- vertex  $u$  before vertex  $v > u$ ,
- head  $(u, s, +)$  before head  $(u, s', +)$  with  $s' > s$  and the same for tails,
- and such that
  - if  $i$  is even,  $\mathbf{e}_{i,s}$  is a head and  $\mathbf{f}_{i,s}$  is a tail,
  - if  $i$  is odd,  $\mathbf{e}_{i,s}$  is a tail and  $\mathbf{f}_{i,s}$  is a head.

The chosen  $\mathbf{p}$  will be called the *representative path* of the class  $X_m^{a,v}(\mathbf{i})$ . We will note  $\mathbf{p} = (\mathbf{e}_{i,s}, \mathbf{f}_{i,s})_{i,s}$ . The edge  $(\mathbf{e}_{i,s}, \mathbf{f}_{i,s})$  will be noted  $y_{i,s}$ . We see  $\mathbf{p}$  as a walk on the vertices  $\mathbf{i}$ . The index  $(i, s)$  in  $\mathbf{p}$  is seen as a time parameter. At time  $(i, s)$ , the walk is located on the vertex  $u$  attached to  $\mathbf{e}_{i,s}$ , and then moves along the edge  $y_{i,s}$  to go to the vertex  $v$  to which is attached  $\mathbf{f}_{i,s}$ .

#### Creating the spanning tree.

We build a marked graph  $T$  on the vertex-set  $\mathbf{i}$  by adding the graph-edge<sup>2</sup>  $(u, v)$  with mark  $y_{i,s}$  when vertex  $v$  is explored for the first time at time  $(i, s)$ . The edge  $y_{i,s}$  is called a tree edge. The (unmarked) graph  $T$  is clearly a tree on the vertex set  $\mathbf{i}$ . The mark over every edge of  $T$  keeps track of the half-edges used to discover for the first time the endvertex of this edge.

Suppose that we are at time  $(i, t)$  and the edge we are currently exploring is  $y_{i,t} = (\mathbf{e}_{i,t}, \mathbf{f}_{i,t})$  and leads to vertex  $u$ . If the vertex  $u$  is already part of the tree  $T$  then the edge  $y_{i,t}$  is called an excess edge and time  $(i, t)$  is called a *cycling time* for obvious reasons.

Due to the very specific structure of  $\mathbf{p}$  (a sequence of tangle-free paths with boundary conditions), such times can easily be understood: either they count as cycling times inside a tangle-free path  $\mathbf{p}_i$  (which can happen only once for every  $i \leq 2m$ ), or they are cycling times between different  $\mathbf{p}_i$ .

We are now going to give an encoding of  $\mathbf{p}$ : the idea is roughly that if there were no cycling times,  $\mathbf{p}$  would perfectly be uncoded without needing anything, due to the choice of lexicographic ordering of half-edges. Therefore, by noting the different cycling times and giving them a minimal amount of information on how to decode them, we will be able to explore the non-cycling times as usual and create the tree  $T$  in the process, and when stepping on a cycling time we will use all the previous information (mainly,  $T$ ) and the mark to determine where to go.

<sup>2</sup>Recall notations from section 2.5. *Edges* are pairs of half-edges seen in  $\mathbf{p}$  while *graph-edges* are pairs of vertices corresponding to some edge.

### Short cycling times.

Each sub-path  $\mathbf{p}_i$  is tangle-free. Let  $r_i$  denotes the first time when  $\mathbf{f}_{i,r_i}$  is attached to a vertex *already visited by  $\mathbf{p}_i$* : this time is called a *short cycling time*. If this cycling time does not exist, we artificially set it to be the symbol  $\otimes$ ; thus,  $r_i = \otimes$  means that  $\mathbf{p}_i$  has no cycles. Also, let  $\sigma_i$  be the first time when the path left this vertex after its first visit in  $\mathbf{p}_i$ . If  $r_i = \otimes$ , we set  $\sigma_i = 0$ . If  $r_i \neq \otimes$ , the cycle  $\mathcal{C}_i$  in  $\mathbf{p}_i$  is precisely given by the edges  $\mathcal{C}_i = \{y_{i,\sigma_i}, y_{i,\sigma_i+1}, \dots, y_{i,r_i}\}$  and it might be visited more than once. Note  $\ell_i$  the “total time spent in the loop”, that is the number of times  $(i,t)$  such that  $y_{i,t}$  is in  $\mathcal{C}_i$ . Then, the knowledge of

1. the cycling time  $(i, r_i)$
2. the half-edges  $\mathbf{e}_{i,r_i}$  and  $\mathbf{f}_{i,r_i}$
3. the total time spent “in the loop”  $\ell_i$  and the half-edge  $\mathbf{e}_{i,\tau_i}$  where we’re leaving the cycle,
4. the next vertex  $u_i$  where we will leave the edges of the tree  $T$ ,

are sufficient to reconstruct the path  $\mathbf{p}_i$  up to the visit of vertex  $u_i$ . Note that in the second step, if  $\mathbf{e}_{i,r_i} = (v_{r_i}, j_{r_i}, \pm)$ , the vertex  $v_{r_i}$  is already known, and whether  $\mathbf{e}_{i,r_i}$  is a head or a tail is also known according to the parity of  $i$ , so we only need to know  $j_{r_i}$ . Thus, if  $r_i \neq \otimes$ , the mark for the  $i$ -th short cycling time  $(i, r_i)$  will be

$$(j_{i,r_i}, \mathbf{f}_{i,r_i}, \ell_i, \mathbf{e}_{i,\tau_i}, u_i) \quad (2.6.4)$$

and if  $r_i = \otimes$  this mark is set to be  $\emptyset$ .

We have at most one short cycling time per  $\mathbf{p}_i$  which is a path of length  $t$ . Fix  $i$ : if there is no cycling time,  $r_i = \emptyset$  (one possibility). If there is a cycling time, there are  $t$  choices for its location. Once this time has been chosen, there are at most  $\Delta(\Delta v)t(\Delta v)v = \Delta^3 v^3 t$  possible marks as (2.6.4) for the short cycling time. This bound is extremely crude but will be sufficient for our purpose. Thus, the total number of possible marks for the short cycling time of  $\mathbf{p}_i$  is  $1 + \Delta^3 v^3 t$ .

**REMARK 2.6.6.** *Suppose we are decoding a short cycling time. The last part of the mark is  $u_i$ ; as  $T$  is a tree, this means that the path to follow is perfectly known up to  $u_i$ . Arriving at  $u_i$  at a certain time, say  $(i', t')$ , we know that we are going to leave the tree  $T$  constructed so far, and this can lead to two situations.*

- *The time  $(i', t')$  can be another cycling time. In this case, the procedure defined on this paragraph (if the cycling time is short) or the next paragraph (if it is long) will tell us where to go next.*
- *The time  $(i', t')$  is not a cycling time. If we note  $v$  the next vertex after  $u_i$ , this means that the edge  $(u_i, v)$  is not in the tree  $T$  constructed so far, and that  $v$  is not already discovered. Therefore, the path is just going to explore this new vertex  $v$  and we are going to add the edge  $(u_i, v)$  to  $T$ . Note that the use of the lexicographic order clearly tells us which half-edges to use.*

### Long cycling times.

There are also cycling times that are not “short cycling times”: basically, it is when a path  $\mathbf{p}_i$  collides with another path  $\mathbf{p}_j$  with  $j < i$ . More precisely, let  $(i, t)$  be a cycling time leading to the (already known) vertex  $u$ . If  $u$  is not one of the vertices discovered by  $\mathbf{p}_i$ , then  $(i, t)$  is called a *long cycling time*: in this case,  $u$  had already been visited by some  $\mathbf{p}_j$  with  $j < i$ . Here again, we are going to mark long cycling times with different items, so they could be easily deduced from the marks. When arriving at a long cycling time, we need to know:

1. the head  $\mathbf{e}_{i,t}$  and the tail  $\mathbf{f}_{i,t}$ ,

2. the next vertex  $u_i$  where we will leave the edges of the tree  $T$  (no extra information is needed: see Remark 2.6.6).

The mark obtained has the form

$$(j_{i,t}, \mathbf{f}_{i,t}, u_i). \tag{2.6.5}$$

For every long cycling time, there are at most  $\Delta^2 v^2$  marks like (2.6.5).

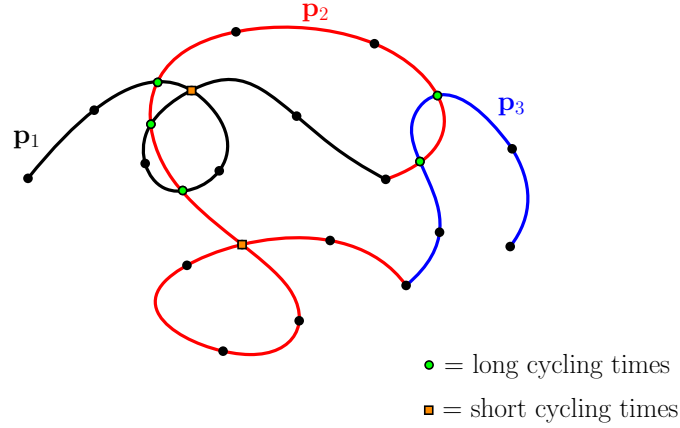
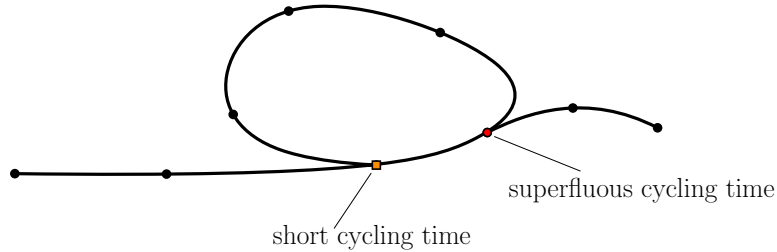


Figure 2.6 – Some examples of vertices generating long and short cycling times.

### Superfluous times.

There is another kind of cycling times we have not yet coded: those times are the cycling times “embedded in the loop” of a short cycling time, that is all the times *except for the first one* when  $(i, t)$  when  $\mathbf{f}_{i,t}$  is attached to a vertex already visited by  $\mathbf{p}_i$ . Those times need no special treatment as they are decoded with the mark of the short cycling time associated with  $i$ . For this reason, they will be called *superfluous cycling times* and play no role in the coding procedure.



### Proof of Lemma 2.6.5.

We now gather the number of different types of marks to get a bound on the number of equivalence classes in  $\mathcal{C}_m$ . Recall the definitions given in Subsection 2.5.3 (page 55) and the difference between *edges of  $\mathbf{p}$*  and *graph-edges of  $\mathbf{p}$* . Consider the undirected multi-graph spanned by the *unoriented* graph-edges of  $\mathbf{p}$  on vertices  $\mathbf{i} = (i_1, \dots, i_v)$ . This graph is connected. Its total number of edges is at most  $a$  (if no edge is visited two times in opposite directions<sup>3</sup>). Therefore, there are at most  $\chi := a - v + 1$  excess edges. For each  $i \leq 2m$ , there are at most  $\chi$  cycling times, *a fortiori* there are at most  $\chi$  long cycling times. Therefore, we have at most  $t^{2m\chi}$  choices for the positions for the long cycling times and we have already seen that we have  $t^{2m}$  choices for the positions of the short cycling times. Now the total count amounts to  $t^{2m(\chi+1)}((\Delta v)^2)^{2m\chi}((\Delta v)^3 t)^{2m}$  possible codings. Organizing termes leads to  $t^{2m\chi+4m}(\Delta v)^{4m\chi+6m}$  which (using  $v \leq 2tm$ ) is bounded by

$$(2\Delta tm)^{8m\chi+12m}.$$

<sup>3</sup>Observe that it is also at least  $a/2$  if all edges are visited twice, in opposite directions. This will not be used in the proof.

Using the asymptotic properties exposed in Lemma 2.5.2, this expression can be simplified. Note for example that there is an integer  $n_1$  only depending on  $\Delta$  such that for every  $n \geq n_1$ , we have  $(2\Delta tm)^{8m} \leq n^{\frac{17}{50}}$ , and the same argument gives  $(2\Delta tm)^{12m} \leq n^{\frac{24}{50}}$ . Hence, when  $n$  is larger than  $n_1$ , we have

$$(2\Delta tm)^{8m\chi+12m} \leq n^{\frac{25}{50} + \frac{17}{50}\chi}$$

which ends the proof of (2.6.3)

### 2.6.3 Proof of Proposition 2.6.2

Let us note  $\mathcal{N}(a, v, \mathbf{i})$  the set of equivalence classes  $\mathcal{E}$  inside  $X_m^{v,a}(\mathbf{i})$ . We have

$$\#X_m^{v,a}(\mathbf{i}) = \sum_{\mathcal{E} \in \mathcal{N}(a, v, \mathbf{i})} \#\mathcal{E}.$$

Using Lemmas 2.6.4 and 2.6.5, when  $n$  is larger than  $n_1$  we get

$$\#X_m^{v,a}(\mathbf{i}) \leq \sum_{\mathcal{E} \in \mathcal{N}(a, v, \mathbf{i})} C^\chi \prod_{i \in \mathbf{i}} d_i^+ d_i^- \leq n^{\frac{25}{50} + \frac{17}{50}\chi} C^\chi \prod_{i \in \mathbf{i}} d_i^+ d_i^-$$

which is the conclusion of Proposition 2.6.2.

## 2.7 Upper bound for $f$

Our aim in the next paragraphs will be to bound  $f(\mathbf{p})$  (which was defined in (2.5.10)) with an expression that depends on the variables  $a, v, m, t, \mathbf{i}$ . We recall a definition from Section 2.4: if  $\mathbf{p}$  is a proto-path of length  $N$ , then

$$\omega(\mathbf{p}) = \prod_{s=1}^N \frac{1}{d_{\mathbf{e}_s}^+}.$$

Every path is itself a proto-path, so we can extend the definition of the weight  $\omega$  in a natural way to  $\mathbf{p} \in \mathcal{C}_m$ :

$$\omega(\mathbf{p}) = \prod_{i=1}^m \omega(\mathbf{p}_i) \omega(\bar{\mathbf{p}}_i).$$

The result of this section is the following proposition which gives upper bounds for  $|f(\mathbf{p})|$  depending on  $a, \chi$ .

**PROPOSITION 2.7.1.** *Let  $\mathbf{p}$  be any path with  $v$  vertices and  $a$  edges. Note  $\chi = a - v + 1$ . Then, there is a constant  $C > 0$  and an integer  $n_2$  such that for every  $n \geq n_2$ , we have the following inequalities:*

- If  $\chi \geq v - tm - 1$ , then

$$|f(\mathbf{p})| \leq \frac{n^{o(1)}}{\delta^{2(tm-v)}} \prod_{i \in \mathbf{i}} \left( \frac{1}{d_i^+} \right)^2 \left( \frac{C}{M} \right)^\chi \left( \frac{c}{M} \right)^{v-1}.$$

- Else  $\chi \leq v - tm - 1$  and we have

$$|f(\mathbf{p})| \leq \frac{n^{o(1)}}{\delta^{2(tm-v)}} \prod_{i \in \mathbf{i}} \left( \frac{1}{d_i^+} \right)^2 \left( \frac{C}{M} \right)^\chi \left( \frac{c}{M} \right)^{v-1} \left( \frac{6tm}{\sqrt{M}} \right)^{2(v-tm-1-\chi)}.$$

The rest of the section is devoted to the proof of this proposition.

### 2.7.1 Expressing the weight $\omega(\mathbf{p})$ with graph-dependant variables

Fix  $\mathbf{p}$  in  $X_m^{a,v}(\mathbf{i})$ . For every  $s > 0$ , let  $V_s$  be the set of vertices that are visited by  $\mathbf{p}$  exactly  $s$  times and note  $v_s = \#V_s$ , so that  $\sum_{s>0} v_s = v$  and  $\sum_{s>0} s v_s = 2tm$ . A vertex is called a *boundary vertex* if it is the endpoint or beginning point of a sub-path of  $\mathbf{p}$ : if  $\mathbf{p} = (\mathbf{p}_i)_{i \leq 2m}$  (with each of the  $\mathbf{p}_i$ 's being tangle-free paths of length  $t$ ) then boundary vertices are those attached to half-edges  $\mathbf{e}_{i,0}$  or  $\mathbf{f}_{i,t}$ . We also recall that  $a_1$  is the number of consistent edges of  $\mathbf{p}$  visited exactly once: this quantity was introduced in Section 2.4 and appears in the statement of Theorem 3. Also, recall that  $b$  is the number of inconsistent edges.

**LEMMA 2.7.2.** *There is a constant  $C > 0$  such that for every  $\mathbf{p} \in X_m^{a,v}(\mathbf{i})$  we have*

$$\omega(\mathbf{p}) \leq n^{o(1)} \prod_{i \in \mathbf{i}} \left( \frac{1}{d_i^+} \right)^2 \frac{C^{\chi+a_1}}{\delta^{2(tm-v)}}. \quad (2.7.1)$$

*Proof.* As a consequence of the definition of the sets  $V_s$ , we have

$$\omega(\mathbf{p}) = \prod_{s>0} \prod_{i \in V_s} \left( \frac{1}{d_i^+} \right)^s,$$

with the usual convention that a product over an empty set is equal to 1. All the products are in fact finite. Isolating  $(d_i^+)^2$  for each  $i$ , we get the following:

$$\omega(\mathbf{p}) = \prod_{i \in \mathbf{i}} \left( \frac{1}{d_i^+} \right)^2 \prod_{i \in V_1} d_i^+ \prod_{s>2} \prod_{i \in V_s} \left( \frac{1}{d_i^+} \right)^{s-2}. \quad (2.7.2)$$

Using hypothesis (H1), this can be bounded by

$$\prod_{i \in \mathbf{i}} \left( \frac{1}{d_i^+} \right)^2 \Delta^{v_1} \left( \frac{1}{\delta} \right)^{\sum_{s>2} \sum_{i \in V_s} (s-2)}. \quad (2.7.3)$$

We also have

$$\begin{aligned} \sum_{s>2} \sum_{i \in V_s} (s-2) &= \sum_{s>2} s v_s - 2 \sum_{s>2} v_s \\ &= 2tm - v_1 - 2v_2 - 2v + 2v_1 + 2v_2 \\ &= 2(tm - v) + v_1. \end{aligned}$$

Thus, we have  $\omega(\mathbf{p}) \leq \prod_{i \in \mathbf{i}} (d_i^+)^{-2} \Delta^{v_1} \delta^{-2(tm-v)} \delta^{-v_1}$ .

We are now going to give a bound on  $v_1$ , the number of vertices visited once. At most  $2m$  of them belong to the boundary vertices of  $\mathbf{p}$ . If  $i$  is in  $V_1$  but is not a boundary vertex, there are exactly two simple edges adjacent with  $i$ , one entering in  $i$  and one going out of  $i$ . One simple edge is adjacent to at most two vertices, so two distinct vertices in  $V_1$  can be adjacent to at most one common simple edge, and we have an injection from the set of non-boundary vertices in  $V_1$  into the set of simple edges, whose cardinal will be denoted by  $a'_1$ : as there are no more than  $2m$  boundary vertices, we have  $v_1 \leq 2m + a'_1$ . Those  $a'_1$  edges might however be inconsistent: if  $a'_1 = a_1 + z'$  with  $z'$  the simple and inconsistent edges, we have  $z' \leq b$ .

**LEMMA 2.7.3.** *With the preceding notations,  $b \leq 4\chi$ .*

This yields  $v_1 \leq 2m + 4\chi + a_1$ . As  $\Delta/\delta \geq 1$ , we have  $(\Delta/\delta)^{v_1} \leq (\Delta/\delta)^{2m+4\chi+a_1}$  and finally

$$\omega(\mathbf{p}) \leq \prod_{i \in \mathbf{i}} (d_i^+)^{-2} (\Delta/\delta)^{2m+4\chi+a_1} \frac{1}{\delta^{2(tm-v)}}.$$

Asymptotics 2.5.2 give  $\Delta^{2m} = n^{o(1)}$ . Taking  $C = (\Delta/\delta)^4$  ends the proof of (2.7.1).  $\square$

*Proof of lemma 2.7.3.* Fix some inconsistent edge  $y = (\mathbf{e}, \mathbf{f})$ . Without loss of generality we can suppose that there is another edge with  $\mathbf{e}$  as its beginning half-edge (say,  $(\mathbf{e}, \mathbf{f}')$  with  $\mathbf{f}' \neq \mathbf{f}$ ) in  $\mathbf{p}$ . If  $\mathbf{e}$  is attached to vertex  $v$ , then there are at most 4 excess edges caused by the fact that  $y$  is not consistent. Therefore, the total number of inconsistent edges is at most  $4\chi$ .  $\square$

### 2.7.2 Expressing $f$ with graph-dependant variables

Let  $\mathbf{p}$  be in  $X_m^{a,v}(\mathbf{i})$ . In order to apply Theorem 3 to  $\mathbf{p}$ , we need a finer knowledge on the number of consistent or simple edges depending on  $a$  and  $v$ . The general idea is the following: the more excess edges, the lesser simple and consistent edges. To apply Theorem 3, we define  $\mathfrak{p}$  to be the proto-path naturally given by  $\mathbf{p}$ . All the quantities  $a, a_1$  and  $b$  appearing in (2.4.1) depend on  $\mathfrak{p}$ . A plain application of Theorem 3 and (2.7.1) with any  $n$  greater than  $n_0$ ,  $N = 2tm$  and  $p = 2tm$  yields the following inequality:

$$|f(\mathbf{p})| \leq 24n^{o(1)} \prod_{i \in \mathbf{i}} \left( \frac{1}{d_i^+} \right)^2 \frac{C_1^\chi 3^b}{\delta^{2(tm-v)}} \left( \frac{c}{M} \right)^a \left( \frac{6tm}{\sqrt{M}} \right)^{a_1}. \quad (2.7.4)$$

We now simplify this expression. The term  $24n^{o(1)}$  is still of order  $n^{o(1)}$ . Let  $a'_1$  be the number of simple edges (not necessarily consistent) and  $a'_2$  be the number of other edges. It is clear that

$$\begin{cases} a'_1 + a'_2 = a \\ a'_1 + 2a'_2 \leq 2mt \end{cases}$$

so  $a'_1 \geq 2(a - mt)$ . If  $b$  is the number of inconsistent edges we have  $a_1 \geq a'_1 - b$  so  $a_1 \geq (2(a - tm) - b)_+$ . Using Lemma 2.7.3, we get  $a_1 \geq (2(a - tm) - 4\chi)_+$ . We use again Lemma 2.7.3:

$$|f(\mathbf{p})| \leq n^{o(1)} \prod_{i \in \mathbf{i}} \left( \frac{1}{d_i^+} \right)^2 \frac{(3^4 C)^\chi}{\delta^{2(tm-v)}} \left( \frac{c}{M} \right)^a \left( \frac{6tm}{\sqrt{M}} \right)^{(2(a-tm)-4\chi)_+}. \quad (2.7.5)$$

Proposition 2.7.1 now follows from (2.7.5) by noting that  $(2(a - tm) - 4\chi)_+ = 0$  if and only if  $\chi \geq v - tm - 1$ .

## 2.8 Asymptotic analysis

We finally gather all the results from Sections 2.6-2.7 and study their limit as  $n$  grows to infinity. More precisely, we will pick only integers  $n$  greater than  $\max\{n_0, n_1\}$ . We first decompose the sum (2.5.11) according to  $v, \chi$  and  $\mathbf{i}$ :

$$\mathbf{E} \left[ \|\underline{P}^{(t)}\|^{2m} \right] \leq \sum_{v=2}^{2mt} \sum_{\mathbf{i}=(i_1, \dots, i_v)} \sum_{\chi=0}^{2tm-v+1} \left( \sum_{\mathbf{p} \in X_m^{a,v}(\mathbf{i})} |f(\mathbf{p})| \right) = \mathcal{H}_1 + \mathcal{H}_2 + \mathcal{L} \quad (2.8.1)$$

where



$$\mathcal{H}_1 = \sum_{v=2}^{mt+1} \sum_{i_1, \dots, i_v} \sum_{\chi=0}^{2tm-v+1} \left( \sum_{\mathbf{p} \in X_m^{a,v}(\mathbf{i})} |f(\mathbf{p})| \right) \quad (2.8.2)$$

$$\mathcal{H}_2 = \sum_{v=mt+2}^{2mt} \sum_{i_1, \dots, i_v} \sum_{\chi=v-tm-1}^{2tm-v+1} \left( \sum_{\mathbf{p} \in X_m^{a,v}(\mathbf{i})} |f(\mathbf{p})| \right) \quad (2.8.3)$$

$$\mathcal{L} = \sum_{v=mt+2}^{2mt} \sum_{i_1, \dots, i_v} \sum_{\chi=0}^{v-tm-2} \left( \sum_{\mathbf{p} \in X_m^{a,v}(\mathbf{i})} |f(\mathbf{p})| \right). \quad (2.8.4)$$

Each term will be separately bounded by  $o(1)n^3(c\tilde{\rho})^{2tm}$  as claimed in (2.5.2).

### 2.8.1 Bound for $\mathcal{H}_1$

In this sum we sum over  $v \leq tm + 1$ . We use Proposition 2.7.1 and (2.6.1) with  $n$  greater than  $n_0$ .

$$\sum_{\mathbf{p} \in X_m^{a,v}(\mathbf{i})} |f(\mathbf{p})| \leq \sum_{\mathbf{p} \in X_m^{a,v}(\mathbf{i})} \frac{n^{o(1)}}{\delta^{2(tm-v)}} \prod_{i \in \mathbf{i}} \left( \frac{1}{d_i^+} \right)^2 \left( \frac{C}{M} \right)^\chi \left( \frac{c}{M} \right)^{v-1} \quad (2.8.5)$$

$$\leq \left( \prod_{i \in \mathbf{i}} d_i^+ d_i^- \right) C^\chi n^{\frac{25}{50} + \frac{17}{50}\chi} \frac{n^{o(1)}}{\delta^{2(tm-v)}} \prod_{i \in \mathbf{i}} \left( \frac{1}{d_i^+} \right)^2 \left( \frac{C}{M} \right)^\chi \left( \frac{c}{M} \right)^{v-1} \quad (2.8.6)$$

$$\leq \left( \prod_{i \in \mathbf{i}} \frac{d_i^-}{d_i^+} \right) \frac{n^{\frac{25}{50} + o(1)}}{\delta^{2(tm-v)}} \left( \frac{Cn^{\frac{17}{50}}}{\delta n} \right)^\chi \left( \frac{c}{M} \right)^{v-1} \quad (2.8.7)$$

$$\leq \left( \prod_{i \in \mathbf{i}} \frac{d_i^-}{d_i^+} \right) \frac{n^3}{\delta^{2(tm-v)}} (Cn^{-\gamma})^\chi (cM^{-1})^v \quad (2.8.8)$$

where we noted  $\gamma = 1 - 17/50 \in ]0, 1[$  and we chose  $n$  large enough to ensure that the term  $o(1)$  is smaller than  $1/50$ . Putting (2.8.8) into (2.8.2) yields

$$\mathcal{H}_1 \leq \sum_{v=2}^{mt+1} \frac{n^3}{\delta^{2(tm-v)}} \sum_{\mathbf{i}} (cM^{-1})^v \left( \prod_{i \in \mathbf{i}} \frac{d_i^-}{d_i^+} \right) \left\{ \sum_{\chi=0}^{2tm-v+1} (Cn^{-\gamma})^\chi \right\}.$$

The sum in  $\chi$  (between braces) is a geometric sum started at 2 and the ratio goes to 0 as  $n$  goes to infinity, so the whole term in braces is of order  $o(1)$ . Recall the definition of  $\rho$ : we have

$$\sum_{\mathbf{i}} (cM^{-1})^v \left( \prod_{i \in \mathbf{i}} \frac{d_i^-}{d_i^+} \right) \leq \left( cM^{-1} \sum_{i=1}^n \frac{d_i^-}{d_i^+} \right)^v \leq (c\rho)^{2v}. \quad (2.8.9)$$

Now

$$\mathcal{H}_1 \leq \frac{o(1)n^3}{\delta^{2tm}} \sum_{v=2}^{mt+1} (c\delta\rho)^{2v}. \quad (2.8.10)$$

Here again, the sum is indeed geometric with ratio  $c\delta\rho \leq c\delta\tilde{\rho}$  where we recall that  $\tilde{\rho} = \rho \vee \delta^{-1}$ . As  $\delta\tilde{\rho} \geq \delta^{-1}$ , we have  $c\delta\tilde{\rho} \geq 1$ , and

$$\sum_{v=2}^{mt+1} (c\delta\tilde{\rho})^{2v} \leq (c\delta\tilde{\rho})^{2mt+2}. \quad (2.8.11)$$

After simplifications, we get  $\mathcal{H}_1 \leq o(1)n^3(c\tilde{\rho})^{2mt}$  which is the desired bound.

### 2.8.2 Bound for $\mathcal{H}_2$

In this sum,  $v > tm + 1$  and  $\chi \geq v - tm - 1$ . The computations are extremely similar to what was done in the preceding section, so we omit the details. As in the preceding section we have

$$\sum_{\mathbf{p} \in X_m^{a,v}(\mathbf{i})} |f(\mathbf{p})| \leq \left( \prod_{i \in \mathbf{i}} \frac{d_i^-}{d_i^+} \right) \frac{n^{\frac{25}{50}+1}}{\delta^{2(tm-v)}} (Cn^{-\gamma})^\chi (cM^{-1})^v. \quad (2.8.12)$$

The sum in  $\chi$  is now started at  $v - tm - 1$ . We have

$$\mathcal{H}_2 \leq \sum_{v=mt+2}^{2mt} \frac{Mn^{\frac{25}{50}}}{c\delta^{2(tm-v)}} \sum_{\mathbf{i}} \left(\frac{c}{M}\right)^v \left( \prod_{i \in \mathbf{i}} \frac{d_i^-}{d_i^+} \right) \left\{ \sum_{\chi=v-tm-1}^{2tm-v+1} \left(\frac{C}{n^\gamma}\right)^\chi \right\} \quad (2.8.13)$$

The sum between braces is geometric and the ratio is  $o(1)$ , hence it is bounded by the first term times some constant close to 1. The first term is  $(Cn^{-\gamma})^{v-tm-1}$ . We also have (2.8.9) and the fact  $Mn^{25/50}/c \leq n^2$  when  $n$  is large enough. Putting it all together, we get

$$\mathfrak{H}_2 \leq n^2 \sum_{v=mt+2}^{2mt} \frac{(c\rho)^{2v} (Cn^{-\gamma})^{v-tm-1}}{\delta^{2(tm-v)}}. \quad (2.8.14)$$

This is indeed a geometric sum and the ratio is of order  $O(n^{-\gamma})$ . After quick simplifications left to the reader, we get  $\mathfrak{H}_2 \leq n^2 (c\tilde{\rho})^{2tm} Cn^{-\gamma}$  which is also generously bounded by  $o(1)n^3 (c\tilde{\rho})^{2tm}$  when  $n$  is large.

### 2.8.3 Bound for $\mathcal{L}$

In this sum,  $v > tm + 1$  and  $\chi \leq v - tm - 1$ . The main difference with the two other regions is the extra term in the bound for  $f(\mathbf{p})$ . We use Proposition 2.7.1 and (2.6.1).

$$\sum_{\mathbf{p} \in X_m^{a,v}(\mathbf{i})} |f(\mathbf{p})| \leq \left( \prod_{i \in \mathbf{i}} d_i^+ d_i^- \right) C^\chi n^{\frac{25}{50} + \frac{17}{50}\chi} \frac{n^{o(1)}}{\delta^{2(tm-v)}} \prod_{i \in \mathbf{i}} \left(\frac{1}{d_i^+}\right)^2 \left(\frac{C}{M}\right)^\chi \left(\frac{c}{M}\right)^{v-1} \left(\frac{6tm}{\sqrt{M}}\right)^{2(v-tm-1-\chi)}$$

This can be simplified when  $n$  is large enough to

$$\left(\frac{c}{M}\right)^v \prod_{i \in \mathbf{i}} \frac{d_i^-}{d_i^+} \frac{n^{\frac{27}{50}}}{\delta^{2(tm-v)}} (Cn^{1-\gamma})^\chi \left(\frac{6tm}{\sqrt{M}}\right)^{2(v-tm-1)} \quad (2.8.15)$$

We plug (2.8.15) into the definition of  $\mathcal{L}$  and we use (2.8.9):

$$\mathcal{L} \leq \sum_{v=mt+2}^{2mt} \sum_{\mathbf{i}} \sum_{\chi=0}^{v-tm-2} \left(\frac{c}{M}\right)^v \prod_{i \in \mathbf{i}} \frac{d_i^-}{d_i^+} \frac{n^{\frac{27}{50}}}{\delta^{2(tm-v)}} (Cn^{1-\gamma})^\chi \left(\frac{6tm}{\sqrt{M}}\right)^{2(v-tm-1)} \quad (2.8.16)$$

$$\leq \sum_{v=mt+2}^{2mt} \frac{(c\tilde{\rho})^{2v} n^{\frac{27}{50}}}{\delta^{2(tm-v)}} \left(\frac{6tm}{\sqrt{M}}\right)^{2(v-tm-1)} \left\{ \sum_{\chi=0}^{v-tm-2} (Cn^{1-\gamma})^\chi \right\}. \quad (2.8.17)$$

As for other regions, the term between braces is a geometric with ratio greater than 1 so it is bounded by  $(Cn^{1-\gamma})^{v-tm-1}$ . We are now left with a sum in  $v$

$$\mathcal{L} \leq \sum_{v=mt+2}^{2mt} \frac{(c\tilde{\rho})^{2v} n^{\frac{27}{50}}}{\delta^{2(tm-v)}} \left(\frac{6tm}{\sqrt{M}}\right)^{2(v-tm-1)} (Cn^{1-\gamma})^{v-tm-1} \quad (2.8.18)$$

and this is generously bounded by  $o(1)n^2 (c\tilde{\rho})^{2tm}$ ; note that  $\mathcal{L}$  is negligible in front of  $\mathcal{H}_1, \mathcal{H}_2$ .

## 2.9 Proof of Proposition 2.2.7

We now prove Proposition 2.2.7. The strategy is exactly the same as for Proposition 2.2.6 and runs along the lines of its proof. We omit the details. First, we recall (2.2.8):

$$R^{t,\ell}(i, j) = \sum_{\mathbf{p} \in \mathcal{R}^{t,\ell}(i, j)} \prod_{s=1}^{\ell-1} A(\mathbf{e}_s, \mathbf{f}_s) \frac{1}{d_{\mathbf{e}_\ell}^+} \prod_{s=\ell+1}^t \underline{A}(\mathbf{e}_s, \mathbf{f}_s)$$

where  $\mathcal{R}^{t,\ell}(i, j)$  had been defined in Definition 2.2.5 on page 46.

### 2.9.1 Trace method

We note  $Y(\mathbf{p}) = \prod_{s=1}^{\ell-1} A(\mathbf{e}_s, \mathbf{f}_s) (d_{\mathbf{e}_\ell}^+)^{-1} \prod_{s=\ell+1}^t \underline{A}(\mathbf{e}_s, \mathbf{f}_s)$  when  $\mathbf{p}$  is in  $\mathcal{R}^{t,\ell}$ . Using the classical trace method as in Subsection 2.5.2, we find

$$\|R^{t,\ell}\|^{2m} \leq \sum_{i_1, \dots, i_{2m}} \prod_{s=1}^m R^{t,\ell}(i_{2s-1}, i_{2s}) R^{t,\ell}(i_{2s+1}, i_{2s}). \quad (2.9.1)$$

$$\leq \sum_{i_1, \dots, i_{2m}} \prod_{s=1}^m \left( \sum_{\mathbf{p} \in \mathcal{R}^{t,\ell}(i_{2s-1}, i_{2s})} Y(\mathbf{p}) \right) \left( \sum_{\mathbf{p} \in \mathcal{R}^{t,\ell}(i_{2s+1}, i_{2s})} Y(\mathbf{p}) \right) \quad (2.9.2)$$

$$\leq \sum_{i_1, \dots, i_{2m}} \sum_{(\mathbf{p}_1, \dots, \mathbf{p}_{2m})} \prod_{s=1}^{2m} Y(\mathbf{p}_i) \quad (2.9.3)$$

where the sum is over all  $2m$ -tuples  $(\mathbf{p}_1, \dots, \mathbf{p}_{2m})$  such that  $\mathbf{p}_{2s}$  is in  $\mathcal{R}^{t,\ell}(i_{2s-1}, i_{2s})$  and  $\mathbf{p}_{2s+1}$  is in  $\mathcal{R}^{t,\ell}(i_{2s+1}, i_{2s})$ ; note that we used the cyclic convention  $i_{2m+1} = i_1$ . Now, going back to the definition of  $R^{t,\ell}$ , we have

$$\mathbf{E}[\|R^{t,\ell}\|^{2m}] \leq \sum_{\mathbf{p} \in \mathcal{C}'_{m,\ell}} |g(\mathbf{p})| \quad (2.9.4)$$

where  $\mathcal{C}'_{m,\ell}$  and  $g$  are now defined in the same fashion as  $\mathcal{C}_m$  and  $f$  in Section 2.5.

**DEFINITION 2.9.1.**  $\mathcal{C}'_{m,\ell}$  is the set of  $2m$ -tuples  $(\mathbf{p}_1, \dots, \mathbf{p}_{2m})$  such that

- $\mathbf{p}_s$  is in  $R^{t,\ell}$  for every  $s$  odd,
- $\bar{\mathbf{p}}_s$  is in  $R^{t,\ell}$  for every  $s$  even, where  $\bar{\mathbf{p}}_s$  denotes path  $\mathbf{p}_s$  “reversed”,
- the beginning vertex of  $\bar{\mathbf{p}}_{2s}$  is the beginning vertex of  $\mathbf{p}_{2s+1}$
- the endvertex vertex of  $\mathbf{p}_{2s-1}$  is the endvertex of  $\bar{\mathbf{p}}_{2s}$ .

Finally, for every  $\mathbf{p} = (\mathbf{p}_1, \dots, \mathbf{p}_{2m})$  in  $\mathcal{C}'_{m,\ell}$ , we set

$$g(\mathbf{p}) = \prod_{s=1}^m Y(\mathbf{p}_{2i-1}) Y(\bar{\mathbf{p}}_{2i}). \quad (2.9.5)$$

**REMARK 2.9.2.** If  $\mathbf{p}_i$  is in  $\mathcal{R}^{t,\ell}$ , then it has at least two cycles. This fact has two consequences: the number of vertices visited by  $\mathbf{p}_i$  is smaller than  $t - 2$ , and the tree excess  $\chi(\mathbf{p}_i)$  is greater than 2. When this is applied to  $\mathbf{p}$ , we get the following facts:

- $\mathbf{p}$  visits no more than  $2tm - 2m = 2m(t - 1)$  vertices.
- $\chi(\mathbf{p})$  is greater than  $4m$ .

Our task is to prove Proposition 2.2.7. To this end, we are going to prove the following lemma:

**LEMMA 2.9.3.** *If  $n$  is large enough, then*

$$\mathbf{E}[\|R^{t,\ell}\|^{2m}] = o(1)n^{2m+3}(c\tilde{\rho})^{2m(t+\ell)} \quad (2.9.6)$$

with  $D > 0$  a constant (we can take  $D = 100$ ).

*Proof of Proposition 2.2.7 using (2.9.6).* By the Markov inequality, we have

$$\begin{aligned} \mathbf{P}(\|R^{t,\ell}\| > n \ln(n)^D (c\tilde{\rho})^{t+\ell}) &\leq \frac{\mathbf{E}[\|R^{t,\ell}\|^{2m}]}{n^{2m} \ln(n)^{2Dm} (c\tilde{\rho})^{2m(t+\ell)}} \\ &\leq \frac{o(1)n^3 n^{2m} (c\tilde{\rho})^{2m(t+\ell)}}{n^{2m} \ln(n)^{2Dm} (c\tilde{\rho})^{2m(t+\ell)}} \leq o(1)n^{3-\frac{2D}{50}}. \end{aligned}$$

If  $D$  is chosen great enough ( $D = 100$  is sufficient), then the last term goes to zero and we get  $\mathbf{P}(\|R^{t,\ell}\| > n \ln(n)^D (c\tilde{\rho})^{t+\ell}) = o(1)$ , which is the desired result.  $\square$

We are now going to prove (2.9.6), first studying the combinatorics of  $\mathcal{C}'_{m,\ell}$ , then bounding  $g(\mathbf{p})$  and finally doing the asymptotic analysis.

## 2.9.2 Combinatorics of $\mathcal{C}'_{m,\ell}$

We split  $\mathcal{C}'_{m,\ell}$  into disjoint subsets.

**DEFINITION 2.9.4.** *Let  $a, v$  be integers and let  $\mathbf{i} = (i_1, \dots, i_v)$  a  $v$ -tuple of vertices. We define*

$$X_{m,\ell}^{v,a}(\mathbf{i}) = X_{m,\ell}^{v,a}(i_1, \dots, i_v)$$

as the set of all the paths in  $\mathcal{C}'_{m,\ell}$  whose vertex set is precisely  $(i_1, \dots, i_v)$  (in this order) and who have  $a$  edges.

Let  $\mathbf{p}$  and  $\mathbf{p}'$  be two paths in  $\mathcal{C}'_{m,\ell}$ ; we note  $\mathbf{e}_{i,s}, \mathbf{f}_{i,s}$  the half-edges of  $\mathbf{p}$  and  $\mathbf{e}'_{i,s}, \mathbf{f}'_{i,s}$  those of  $\mathbf{p}'$ . Those paths are said equivalent if

- they both belong to  $X_{m,\ell}^{a,v}(\mathbf{i})$  and they visit the same vertices at the same time,
- for every vertex  $u \in \mathbf{i}$ , there are two permutations  $\sigma_u \in \mathfrak{S}_{d_u^+}$  and  $\tau_u \in \mathfrak{S}_{d_u^-}$  such that for every  $i$  and  $s$ , if  $\mathbf{e}_{i,s}$  is a head attached to  $u$  and  $\mathbf{f}_{i,s}$  a tail attached to  $u$ , then

$$\mathbf{e}_{i,s} = \sigma_u(\mathbf{e}'_{i,s}) \quad \text{and} \quad \mathbf{f}_{i,s} = \tau_u(\mathbf{f}'_{i,s}).$$

Indeed, two paths are equivalent if they only differ by a permutation of their half-edges. We state again Lemma 2.6.4. Its proof remains unchanged, and Lemma 2.7.3 is also true in this case.

**LEMMA 2.9.5.** *Let  $\mathbf{p}$  be a path in  $X_{m,\ell}^{v,a}(\mathbf{i})$ . Then, we have at most*

$$C^\chi \prod_{i \in \mathbf{i}} d_i^+ d_i^- \quad (2.9.7)$$

paths equivalent to  $\mathbf{p}$ , where  $C$  is a constant.

### 2.9.3 Number of equivalence classes

Now, we count the number of equivalence classes in  $X_{m,\ell}^{a,v}(\mathbf{i})$ . The explored vertices are  $\mathbf{i} = (i_1, \dots, i_v)$ , in this order.

In any equivalence class, we choose a path  $\mathbf{p}$  visiting heads and tails in the ‘‘alternating lexicographic order’’ in the same fashion as in 2.6.2. The chosen path  $\mathbf{p}$  will be called the *representative path* of the class  $X_{m,\ell}^{a,v}(\mathbf{i})$ . We build the tree  $T$  in the exact same way.

Cycling times are defined in the same way, but now another phenomenon can occur: there can be more than one cycle inside one subpath  $\mathbf{p}_i$ . However, the path  $\mathbf{p}_i$  is composed of two subpaths, say  $\mathbf{p}'_i$  and  $\mathbf{p}''_i$ , linked by a single edge<sup>4</sup>, and inside one of the two paths  $\mathbf{p}'_i, \mathbf{p}''_i$ , there can be no more than one cycle. Thus, a small variation of the code for  $\mathcal{C}_m$  will be sufficient for our purpose. To this end, define the *bridging time*

$$\ell_i := \begin{cases} \ell - 1 & \text{if } i \text{ is even} \\ t - \ell & \text{else.} \end{cases}$$

#### Short cycling times.

Each sub-path  $\mathbf{p}'_i, \mathbf{p}''_i$  is tangle-free. Let  $r'_i$  denotes the first time when  $\mathbf{f}_{i,r'_i}$  is attached to a vertex *already visited* by  $\mathbf{p}'_i$ , and similarly  $r''_i$  for  $\mathbf{p}''_i$ . Those are *short cycling times*.

If these cycling times does not exist, we artificially set them to be the symbol  $\otimes$ . Let  $\sigma'_i, \sigma''_i$  be the first time when the path  $\mathbf{p}'_i, \mathbf{p}''_i$  left this vertex after its first visit. Finally, note  $h'_i, h''_i$  the ‘‘total time spent in the loop’’.

We mark the cycling times  $r'_i, r''_i$  as follows:

$$(j'_{i,r'_i}, \mathbf{f}_{i,r'_i}, h'_i, \mathbf{e}_{i,\tau'_i}, u'_i) \quad \text{and} \quad (j''_{i,r''_i}, \mathbf{f}_{i,r''_i}, h''_i, \mathbf{e}_{i,\tau''_i}, u''_i) \quad (2.9.8)$$

and if  $r'_i, r''_i = \otimes$  this mark is set to be  $\emptyset$ .

We also have to deal with what happens at the bridge between  $\mathbf{p}'_i$  and  $\mathbf{p}''_i$ . To this end, we simply mark the bridging time  $\ell_i$  with the whole bridge, that is we set

$$\beta_i = (\mathbf{e}_{i,\ell_i}, \mathbf{f}_{\ell_i+1}).$$

All those informations are enough to reconstruct the short cycling times and the bridge. Note that we did not fully exploit the  $R^{t,\ell}$ -structure of the paths  $\mathbf{p}_i$ : in particular, we did not use the fact that in the end,  $\mathbf{p}_i$  is tangled. This will be used further.

Let us count those codes. We have at most two short cycling time per  $\mathbf{p}'_i$  or  $\mathbf{p}''_i$ . There are  $\ell_i$  choices for the first short cycling time and at most  $\Delta(v\Delta)t(v\Delta)v = t\Delta^2v^3$  choices for its mark, then there are at most  $\Delta v$  choices for the bridge, and finally there are at most  $t - \ell_i$  choices for the second short cycling time and  $t\Delta^2v^3$  choices for its mark.

#### Long cycling times.

Let  $(i, t)$  be a cycling time leading to the (already known) vertex  $u$ . If  $(i, t)$  is not a short cycling time, then

1. either  $u$  belongs to the verties discovered by some  $\mathbf{p}_j$  with  $j < i$ ,
2. either  $t > \ell_i$  and  $u$  belongs to the vertices discovered by  $\mathbf{p}'_i$ .

<sup>4</sup>Which can also be considered as a tangle-free path of length 1.

In either cases, we say  $(i, t)$  is a long cycling time. We mark long cycling times with a triple

$$(j_{i,t}, \mathbf{f}_{i,t}, u_i). \quad (2.9.9)$$

where  $j_{i,t}$  is the index of the head<sup>5</sup> by which we're leaving the current vertex,  $\mathbf{f}_{i,t}$  is the tail we are going to, and  $u_i$  is the next vertex when we will be leaving the tree  $T$ . For every long cycling time, there are at most  $\Delta^2 v^2$  marks like (2.9.9).

### Superfluous times.

Superfluous cycling times are defined as in 2.6.2 and play no role in the sequel.

### Total number.

We now gather the number of different types of marks to get a bound on the number of equivalence classes in  $\mathcal{C}'_{m,\ell}$ .

**PROPOSITION 2.9.6.** *The total number of equivalence classes of paths in  $\mathcal{C}'_{m,\ell}$  visiting vertices  $\mathbf{i} = (i_1, \dots, i_v)$  and having  $a$  edges is at most*

$$4^{-m} (2\Delta t m)^{4m\chi + 22m}. \quad (2.9.10)$$

*Proof.* Recall the definitions of section 2.5 and the difference between *edges* of  $\mathbf{p}$  and *graph-edges* of  $\mathbf{p}$ . Consider the undirected multi-graph spanned by the *unoriented* graph-edges of  $\mathbf{p}$  on vertices  $\mathbf{i} = (i_1, \dots, i_v)$ . This graph is connected. Its total number of edges is at most  $a$  (if no edge is visited two times in opposite directions<sup>6</sup>). Therefore, there are at most  $\chi := a - v + 1$  excess edges. For each  $i \leq 2m$ , there are at most  $\chi$  cycling times, *a fortiori* there are at most  $\chi$  long cycling times. Therefore, we have at most  $t^{2m\chi}$  choices for the positions of the long cycling times. For each  $i$ , there are at most two cycling times, one before  $\ell_i$  and one after. The total number of choices for these short cycling times is thus  $\prod_{i=1}^{2m} \ell_i (k - \ell_i) = \ell^{2m} (t - \ell)^{2m} \leq 4^{-m} t^{2m}$ .

For each one of these choices, we have the following number of possibilities for the marks:  $(t\Delta^2 v^3)^{2 \times 2m}$  for short cyclings,  $(\Delta v)^{4m}$  for bridges,  $(\Delta^2 v^2)^{2m\chi}$  for long cyclings. The total number of codings is at most  $4^{-m} t^{6m} \Delta^{12m+4m\chi} v^{16m+4m\chi}$  which (using  $v \leq 2tm$ ) is largely bounded by (2.9.10).  $\square$

Using the asymptotic properties exposed in lemma 2.5.2, the reader can check that (2.9.10) is bounded by  $n^{\frac{45}{50} + \frac{17}{50}\chi}$  when  $n$  is large enough. Using Lemma 2.9.5, we get the following variant of Proposition (2.6.2):

**PROPOSITION 2.9.7.** *Fix  $\ell, v, \mathbf{i}$  and  $a$ . Then, when  $n$  is big enough we have*

$$\#X_{m,\ell}^{v,a}(\mathbf{i}) \leq \left( \prod_{i \in \mathbf{i}} d_i^+ d_i^- \right) C^\chi n^{\frac{45}{50} + \frac{17}{50}\chi}. \quad (2.9.11)$$

### 2.9.4 Analysis of $g$

We now bound  $g(\mathbf{p})$  when  $\mathbf{p}$  is in  $\mathcal{C}'_{m,\ell}$ , following the ideas in Section 2.7. Recall the definition of  $g$  as in (2.9.5). When developping the terms in  $Y$ , if we note  $\mathbf{p}_i = (\mathbf{e}_{i,s}, \mathbf{f}_{i,s})_{s \leq \ell}$  for  $i$  odd and  $\bar{\mathbf{p}}_i = (\mathbf{e}_{i,s}, \mathbf{f}_{i,s})_{s \leq \ell}$  for  $i$  even, then we have

$$g(\mathbf{p}) = \prod_{i=1}^{2m} \frac{1}{d_{\mathbf{e}_{i,\ell}}^+} \times \prod_{i=1}^{2m} \prod_{s < \ell} A(\mathbf{e}_{i,s}, \mathbf{f}_{i,s}) \prod_{s > \ell} \underline{A}(\mathbf{e}_{i,s}, \mathbf{f}_{i,s}). \quad (2.9.12)$$

<sup>5</sup>Or the head, depending on the parity of  $i$ .

<sup>6</sup>Observe that it is also at least  $a/2$  if all edges are visited twice, in opposite directions. This will not be used in the proof.

Fix a path  $\mathbf{p}$  in  $X_{m,\ell}^{a,v}(\mathbf{i})$ . Lemma 2.7.2 remains exactly the same.

**LEMMA 2.9.8.** *There is a constant  $C$  such that for every  $\mathbf{p} \in X_{m,\ell}^{a,v}(\mathbf{i})$  we have*

$$|\omega(\mathbf{p})| \leq n^{o(1)} \prod_{i \in \mathbf{i}} \left( \frac{1}{d_i^+} \right)^2 \frac{C^{\chi+a_1}}{\delta^{2(tm-v)}}. \quad (2.9.13)$$

Now comes the application of Theorem 3 to the second factor in the right of (2.9.12). Let  $\mathbf{p}$  be a path in  $X_{m,\ell}^{a,v}(\mathbf{i})$ . In order to apply Theorem 3, we need to define an auxiliary path, say  $\hat{\mathbf{p}}$ , by deleting each  $\ell$ -th edge in a subpath  $\mathbf{p}_i$ . We plug (2.9.13) into the bound given by Theorem 3 to get

$$|g(\mathbf{p})| \leq 24n^{o(1)} \prod_{i \in \mathbf{i}} \left( \frac{1}{d_i^+} \right)^2 \frac{C^{\chi} 3^b}{\delta^{2(tm-v)}} \left( \frac{c}{M} \right)^{\hat{a}} \left( \frac{Cm(t-1)}{\sqrt{M}} \right)^{a_1}.$$

where  $\hat{a}$  is the total number of edges of  $\hat{\mathbf{p}}$ , so  $\hat{a} \geq a - 2m$  with  $a$  the total number of edges of  $\mathbf{p}$ . Also,  $a_1$  is now the number of simple, consistent edges that appear in the path  $\hat{\mathbf{p}}$ :

- in  $\mathbf{p}_i$ , after  $\ell$  if  $i$  is odd,
- in  $\mathbf{p}_i$ , before  $t - \ell$  if  $i$  is even.

Such edges will be called *good edges* just for this paragraph. Note  $\bar{a}_1$  the total number of simple, consistent edges in  $\mathbf{p}$ ; as there are no more than  $2m\ell$  edges that are not good, we have  $a_1 \geq (\bar{a}_1 - 2m\ell)_+$ .

Let  $\bar{a}'_1$  be the number of simple edges (not necessarily consistent) of  $\mathbf{p}$  and  $\bar{a}'_2$  be the number of other edges. It is clear that  $\bar{a}'_1 + \bar{a}'_2 = a$  and  $\bar{a}'_1 + 2\bar{a}'_2 \leq 2mt$  so  $\bar{a}'_1 \geq 2(a - mt)$ . If  $b$  is the number of inconsistent edges we have  $\bar{a}_1 \geq \bar{a}'_1 - b$  so  $\bar{a}_1 \geq 2(a - tm) - b$ , and using Lemma 2.7.3, we get  $\bar{a}_1 \geq 2(a - tm) - 4\chi$  and finally  $a_1 \geq (2(a - tm) - 4\chi - 2\ell m)_+$ . We also have  $24n^{o(1)} = n^{o(1)}$ . Note that  $2(a - tm) - 4\chi - 2\ell m = 2((v-1) - (t+\ell)m - \chi)$ . Using once again Lemma 2.7.3, we get

$$|g(\mathbf{p})| \leq n^{o(1)} \prod_{i \in \mathbf{i}} \left( \frac{1}{d_i^+} \right)^2 \frac{C^{\chi}}{\delta^{2(tm-v)}} \left( \frac{c}{M} \right)^{a-2m} \left( \frac{Ctm}{\sqrt{M}} \right)^{2((v-1)-(t+\ell)m-\chi)_+}.$$

The  $2((v-1) - (t+\ell)m - \chi)_+$  term is zero if and only if  $\chi \geq v - tm - t\ell - 1$ , hence the following result.

**PROPOSITION 2.9.9.** *Let  $\mathbf{p}$  be any path in  $\mathcal{C}'_{m,\ell}$  with  $v$  vertices and  $a$  edges. Note  $\chi = a - v + 1$ . There is a constant  $C$  such that when  $n$  is large enough, we have*

- If  $\chi \geq v - (t+\ell)m - 1$ , then

$$|g(\mathbf{p})| \leq \frac{n^{o(1)}}{\delta^{2(tm-v)}} \prod_{i \in \mathbf{i}} \left( \frac{1}{d_i^+} \right)^2 \left( \frac{C}{M} \right)^{\chi} \left( \frac{c}{M} \right)^{v-1-2m}.$$

- Else,  $\chi \leq v - (t+\ell)m - 1$  and in this case,

$$|g(\mathbf{p})| \leq \frac{n^{o(1)}}{\delta^{2(tm-v)}} \prod_{i \in \mathbf{i}} \left( \frac{1}{d_i^+} \right)^2 \left( \frac{C}{M} \right)^{\chi} \left( \frac{c}{M} \right)^{v-1} \left( \frac{C2tm}{\sqrt{M}} \right)^{2(v-tm-\ell m-1-\chi)}.$$

### 2.9.5 Asymptotic analysis

All the computations in this section have already been done in Section 2.8, so we do not write the details. Go back to (2.9.4) and decompose the sum according to  $a, v, \mathbf{i}$ :

$$\mathbf{E} [\|R^{t,\ell}\|^{2m}] \leq \sum_{v=2}^{2mt} \sum_{\mathbf{i}=(i_1, \dots, i_v)} \sum_{\chi=4m}^{2tm-v+1} \left( \sum_{\mathbf{p} \in X_{m,\ell}^{a,v}(\mathbf{i})} |g(\mathbf{p})| \right) = \mathcal{H}'_1 + \mathcal{H}'_2 + \mathcal{L}'$$

where

$$\mathcal{H}'_1 = \sum_{v=2}^{m(t+\ell)+1} \sum_{i_1, \dots, i_v} \sum_{\chi=4m}^{2tm-v+1} \left( \sum_{\mathbf{p} \in X_{m,\ell}^{a,v}(\mathbf{i})} |g(\mathbf{p})| \right) \quad (2.9.14)$$

$$\mathcal{H}'_2 = \sum_{v=m(t+\ell)+2}^{2m(t-2)} \sum_{i_1, \dots, i_v} \sum_{\chi=v-(t+\ell)m-1}^{2tm-v+1} \left( \sum_{\mathbf{p} \in X_{m,\ell}^{a,v}(\mathbf{i})} |g(\mathbf{p})| \right) \quad (2.9.15)$$

$$\mathcal{L}' = \sum_{v=m(t+\ell)+2}^{2mt} \sum_{i_1, \dots, i_v} \sum_{\chi=4m}^{v-tm-\ell m-2} \left( \sum_{\mathbf{p} \in X_{m,\ell}^{a,v}(\mathbf{i})} |g(\mathbf{p})| \right). \quad (2.9.16)$$

Each one of those terms can be bounded by the appropriate quantity as requested in Proposition 2.2.7, that is  $o(1)n^{2m+3}(c\tilde{\rho})^{2m(t+\ell)}$ .

For example, in  $\mathcal{H}'_1$ , we sum over indices such that  $v \leq (t+\ell)m+1$ . We then have

$$\begin{aligned} \sum_{\mathbf{p} \in X_{m,\ell}^{a,v}(\mathbf{i})} |g(\mathbf{p})| &\leq \sum_{\mathbf{p} \in X_{m,\ell}^{a,v}(\mathbf{i})} \frac{n^{o(1)}}{\delta^{2(tm-v)}} \prod_{i \in \mathbf{i}} \left( \frac{1}{d_i^+} \right)^2 \left( \frac{C}{M} \right)^\chi \left( \frac{c}{M} \right)^{v-1-2m} \\ &\leq \left( \prod_{i \in \mathbf{i}} d_i^+ d_i^- \right) C^\chi n^{\frac{45}{50} + \frac{17}{50}\chi} \frac{n^{o(1)}}{\delta^{2(tm-v)}} \prod_{i \in \mathbf{i}} \left( \frac{1}{d_i^+} \right)^2 \left( \frac{C}{M} \right)^\chi \left( \frac{c}{M} \right)^{v-1-2m} \\ &\leq \left( \prod_{i \in \mathbf{i}} \frac{d_i^-}{d_i^+} \right) \frac{n^{\frac{45}{50}}}{\delta^{2(tm-v)}} (Cn^{-\gamma})^\chi (cM^{-1})^{v-1-2m} \end{aligned}$$

with  $\gamma = 1 - 17/50 \in ]0, 1[$ . As noted in Remark 2.9.2, if  $\mathbf{p}$  is in  $\mathcal{E}'_{m,\ell}$ , then  $\chi$  cannot be less than  $4m$ . We thus have

$$\begin{aligned} \mathcal{H}'_1 &\leq \sum_{v=2}^{mt+m\ell+1} \left( \frac{M}{c} \right)^{2m+1} \frac{n^{\frac{45}{50}}}{\delta^{2(tm-v)}} \sum_{\mathbf{i}} (cM^{-1})^v \left( \prod_{i \in \mathbf{i}} \frac{d_i^-}{d_i^+} \right) \left\{ \sum_{\chi=4m}^{2tm-v+1} (Cn^{-\gamma})^\chi \right\} \\ &\leq \sum_{v=2}^{mt+m\ell+1} n^{2m+1+o(1)} \frac{n^{\frac{45}{50}}}{\delta^{2(tm-v)}} \sum_{\mathbf{i}} (cM^{-1})^v \left( \prod_{i \in \mathbf{i}} \frac{d_i^-}{d_i^+} \right) (Cn^{-\gamma})^{4m} \left\{ \sum_{\chi=0}^{2m(t-1)-v+1} (Cn^{-\gamma})^\chi \right\} \end{aligned}$$

The sum in  $\chi$  (between braces) is a bounded by 2 if  $n$  is large enough and the sum in  $\mathbf{i}$  is bounded by  $(c\rho)^{2v}$ , hence

$$\begin{aligned} \mathcal{H}'_1 &\leq 2n^{2+2m+o(1)-4m\gamma} \sum_{v=2}^{mt+m\ell+1} \frac{(c\rho)^{2v}}{\delta^{2(tm-v)}} \leq n^{2m+2-4m\gamma} \delta^{-2tm} \frac{(c\delta\rho)^{2mt+2m\ell-2} - 1}{(c\delta\rho) - 1} (c\delta\rho)^2 \\ &\leq n^{2m+3-4m\gamma} \delta^{2\ell m} (c\tilde{\rho})^{2mt+2m\ell}. \end{aligned}$$



To conclude, note that  $\delta^{2tm} = n^{2m\alpha/\ln(\Delta)}$ ; when  $\alpha$  is chosen to be strictly smaller than  $2\ln(\Delta)\gamma$ , the term  $n^{-4m\gamma}\delta^{2m\ell}$  becomes  $o(1)$ .

We bound  $\mathcal{H}'_2$  and  $\mathcal{L}'$  in the same way, adapting the computations already done in the preceding sections.

\*\*\*

## 2.10 Algebraic tools

In this section, we prove Lemma 2.3.3. We begin with a classical theorem ([20]) connecting the eigenvalues of any diagonalizable matrix  $A$  with the eigenvalues of any perturbation of  $A$ . If  $M$  is a matrix, we note  $\sigma(M)$  the set of its eigenvalues.

**Theorem 4** (Bauer-Fike). *Let  $A$  be a diagonalizable matrix,  $A = PDP^{-1}$  with  $P$  invertible and  $D$  diagonal, and let  $H$  be any matrix.*

1. Define  $\varepsilon = \|P\| \cdot \|P^{-1}\| \cdot \|H\|$ . Then,

$$\sigma(A + H) \subset \bigcup_{\lambda \in \sigma(A)} B(\lambda, \varepsilon). \quad (2.10.1)$$

2. If  $I$  is a subset of  $\{1, \dots, n\}$  such that

$$\bigcup_{i \in I} B(\lambda_i, \varepsilon) \cap \bigcup_{i \notin I} B(\lambda_i, \varepsilon) = \emptyset$$

then the number of eigenvalues of  $A + H$  lying in  $\bigcup_{i \in I} B(\lambda_i, \varepsilon)$  is exactly  $\#I$ .

Hence, the spectrum of the perturbed matrix  $A + H$  is entirely contained in the  $\varepsilon$ -blowup around the spectrum of  $A$  (see also Figure 2.7). Note that whenever  $A$  is hermitian, the matrix  $P$  is unitary and  $\|P\| = \|P^{-1}\| = 1$ . Therefore, the ‘‘eigenvalue maximal perturbation’’, namely  $\varepsilon$ , depends on the amplitude of the perturbation matrix (i.e. the term  $\|H\|$ ) and on the ‘‘lack of hermitian-ness’’ of the matrix  $A$  (since we always have  $\|P\| \cdot \|P^{-1}\| \geq 1$ ).

Here is the entertaining proof of the Bauer-Fike theorem.

*Proof of the first point.* Let  $\mu$  be an eigenvalue of the perturbed matrix  $A + H$ ; then  $A + H - \mu\text{Id}$  is singular. Suppose that  $\mu \notin \sigma(A)$ ; in this case,  $D - \mu\text{I}$  is nonsingular, and we have

$$A + H - \mu\text{I} = P(D - \mu\text{I})(\text{I} + (D - \mu\text{I})^{-1}P^{-1}HP)P^{-1}.$$

This shows that  $\text{I} + (D - \mu\text{I})^{-1}P^{-1}HP$  is singular, so  $-1$  is an eigenvalue of  $M := (D - \mu\text{I})^{-1}P^{-1}HP$ ; in particular,  $1 \leq \|M\| \leq \|(D - \mu\text{I})^{-1}\| \cdot \|P^{-1}\| \cdot \|H\| \cdot \|P\|$ . It is easy to see that the norm of the diagonal matrix  $(D - \mu\text{I})^{-1}$  is  $|\lambda_k - \mu|^{-1}$ , where  $k$  is such that  $|\lambda_k - \mu| = \min |\lambda_i - \mu|$ . This proves the inequality  $|\lambda_k - \mu| \leq \|P^{-1}\| \cdot \|H\| \cdot \|P\|$  which is the claim (2.10.1).  $\square$

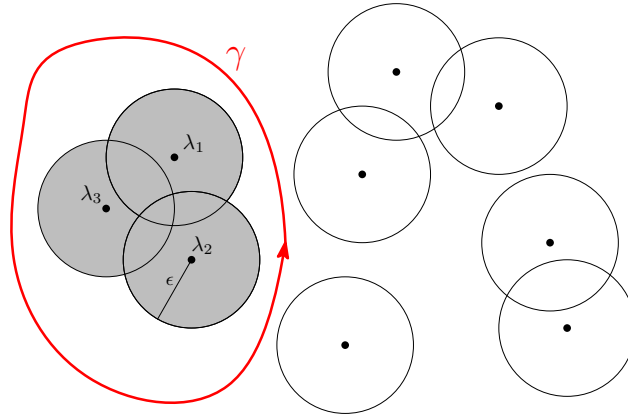


Figure 2.7 – Black dots denote the spectrum of  $A = A_0$ . All the eigenvalues of  $A + H$  are inside the circles and the number of eigenvalues of  $A + H$  inside the grey zone is equal to exactly 3.

*Proof of the second point.* Let  $s$  be in  $[0, 1]$ . Note  $A_s = A + sH$  and  $p_s(z) = \det(A_s - z\text{Id})$ . The eigenvalues of  $A$  are the roots of  $p_0$  and those of  $A + H$  are the roots of  $p_1$ . Let  $\gamma$  be a simple Jordan curve in the complex plane and let  $U$  be the bounded connected component of  $\mathbb{C} \setminus \gamma$  and  $V$  the other component; suppose that  $\cup_{i \in I} B(\lambda_i, \varepsilon) \subset U$  and  $\cup_{i \notin I} B(\lambda_i, \varepsilon) \subset V$  (see figure 2.7). Then, the argument principle yields that the number  $n(s)$  of roots of  $p_s$  in  $U$  is equal to

$$\frac{1}{2i\pi} \oint_{\gamma} \frac{p'_s(\zeta)}{p_s(\zeta)} d\zeta.$$

The polynomial  $p_s$  depends continuously on the coefficients of  $A_s$ , so the application  $s \mapsto n(s)$  is continuous from  $[0, 1]$  into  $\mathbb{N}$ , so by connectedness it is constant. We thus have  $n(0) = n(1)$  and it is clear that  $n(0) = \#I$ . □

In order to use the Bauer-Fike theorem, we need a control on the *condition number* of  $P$ , that is  $c(P) = \|P\| \cdot \|P^{-1}\|$ . When  $A$  has rank 1 this can be easily done; note that every rank 1 matrix can be written  $xy^\top$  with  $x, y$  two nonzero vectors.

**PROPOSITION 2.10.1.** *Let  $A$  be a diagonalizable matrix with rank 1,  $A = PDP^{-1}$  with  $P$  invertible and  $D$  diagonal, say  $D = \text{diag}(\mu, 0, \dots, 0)$  with  $\mu$  the unique non-zero eigenvalue of  $A$ . Let  $x, y$  be two vectors such that  $A = xy^\top$ . Then,  $\mu$  is equal to  $\langle x, y \rangle$  and*

$$c(P) \leq \frac{2\|x\|^2\|y\|^2}{\mu^2}. \quad (2.10.2)$$

*Proof.* First, note that if  $A = xy^\top$ , then by Sylvester's determinant formula, for every  $z$  we have  $\det(zI - xy^\top) = z^n(1 - z^{-1}y^\top x) = z^{n-1}(z - \langle x, y \rangle)$ , so the eigenvalues of  $A$  are 0 and  $\langle x, y \rangle$ ; indeed, if  $A$  is diagonalizable and has rank 1, then necessarily  $\langle x, y \rangle \neq 0$  and  $\mu = \langle x, y \rangle$ .

We first suppose that  $\|x\| = \|y\| = 1$ . The right-eigenvector associated with  $\mu$  is  $x$ , the left-eigenvector is  $y^\top$ . Every basis of  $\text{vect}(y)^\perp$  provides a family of right-eigenvectors for the eigenvalue 0 and every basis of  $\text{vect}(x)^\perp$  provides a family of left-eigenvectors for the eigenvalue 0. For every orthonormal basis of  $\text{vect}(y)^\perp$ , say  $(e_2, \dots, e_n)$ , define a matrix by  $P = (x, e_2, \dots, e_n)$ . Then  $P$  is a diagonalization matrix for  $A$ . Now, define  $X = (y, e_2, \dots, e_n)$ : this matrix is unitary and we can check

that

$$X^*P = \begin{pmatrix} \mu & 0 & 0 & \dots & 0 \\ \langle x, e_2 \rangle & 1 & 0 & \dots & 0 \\ \langle x, e_3 \rangle & 0 & 1 & \dots & 0 \\ \vdots & & & \ddots & \vdots \\ \langle x, e_n \rangle & 0 & \dots & & 1 & 0 \\ & & & & 0 & 1 \end{pmatrix}$$

We can also choose the basis  $(e_i)$  so that  $x$  belongs to  $\text{vect}(y, e_2)$ . Let  $b$  be a real number such that  $x = \mu y + be_2$ . As  $\|x\| = 1$  and  $\mu \neq 0$ , we have  $b \in ]-1, 1[$  [we must have  $b^2 = 1 - \mu^2$  and  $b \in ]-1, 1[$ . Then,

$$X^*P = \begin{pmatrix} \mu & 0 \\ b & 1 \\ & & \mathbf{I}_{n-2} \end{pmatrix}.$$

We thus have proven that if

$$R = \begin{pmatrix} \mu & 0 \\ b & 1 \end{pmatrix}$$

then  $c(P) = c(X^*P) = c(R)$ , and the condition number  $c(R)$  can be computed; indeed, we find  $c(R) = \sqrt{(1+|b|)/(1-|b|)}$ . Remember that  $|b| = \sqrt{1-\mu^2} \leq 1 - \mu^2/2$ . Let  $f$  be the increasing function defined on  $[0, 1[$  by  $f : t \mapsto \sqrt{(1+t)/(1-t)}$ . Then  $c(R) = f(|b|) \leq f(1 - \mu^2/2)$  and it can be quickly checked, using  $\sqrt{s-1} \leq s/2$ , that  $f(1 - \mu^2/2) \leq 2/\mu^2$ . We thus have proven that  $c(P) \leq 2/\mu^2$ .

Now, suppose that  $\|x\|$  or  $\|y\|$  are not equal to 1 and define  $\tilde{x} = x/\|x\|$  and  $\tilde{y} = y/\|y\|$ . Set  $\tilde{A} = \tilde{x}\tilde{y}^\top$  so that  $\|x\|\|y\|\tilde{A} = A$ . Note  $r = \|x\|\|y\|$ . We have  $\tilde{A} = P\tilde{D}P^{-1}$  with  $\tilde{D} = \text{diag}(\mu/r, 0, \dots, 0)$  and  $c(P) \leq 2r^2/\mu^2$  by the preceding arguments. As we also have  $A = PDP^{-1}$ , this yields the final conclusion

$$c(P) \leq \frac{2\|x\|^2\|y\|^2}{\mu^2}.$$

□

We now conclude the proof of Lemma 2.3.3 on Theorem 4 and Proposition 2.10.1.

*Proof of Lemma 2.3.3.* Apply the first point of the Bauer-Fike theorem to the matrix  $M + H$ : all the eigenvalues of  $M + H$  lie in the union of the balls  $B(\lambda, \varepsilon)$  with  $\varepsilon = c(P)\|H\|$ . As  $M$  has rank 1, apply Proposition 2.10.1:  $c(P) \leq 2\|x\|^2\|y\|^2\mu^{-2}$ . Now, apply the second part of the Bauer-Fike theorem and suppose that  $B(\mu, \varepsilon)$  and  $B(0, \varepsilon)$  are disjoint. There is exactly one eigenvalue of  $M$  in  $B(\mu, \varepsilon)$  which is  $\mu$ , so there is exactly one eigenvalue of  $M + H$  in  $B(\mu, \varepsilon)$  and all other eigenvalues are in  $B(0, \varepsilon)$ . □

## 2.11 Proof of Theorem 3

In this appendix, we prove Theorem 3 by adapting the arguments of the proof of Proposition 8 of [33] to our setting. All the required definitions and notations have already been introduced in Section 2.4 (page 52). The proof begins with the simple case where all edges of  $\mathfrak{p}$  are consistent and then goes on to the general case. We start with a preliminary remark.

**REMARK 2.11.1.** Remember that  $c$  is a constant arbitrarily close to 1. As  $a \leq N \leq \sqrt{M}$ , we have  $(M-a)^{-1} \leq (M-N)^{-1} \leq (M-\sqrt{M})^{-1}$  and when  $n$  is large, this is smaller than  $cM^{-1}$ . This inequality will be used multiple times in the proof of Theorem 3.

### 2.11.1 Proof, part I: all edges are consistent

#### Definitions of some useful sets

This section deals with the general case, where some edges might not be simple in the proto-path. However, we suppose for the moment that no edge is inconsistent. First, define sets  $T, T_q$  as follows:

- $T$  is the set of all edges such that  $w'_i > 0$ . Those edges appear after  $p$ ; they can appear both before and after  $p$ . We note  $d = \#T$ .
- $T_q$  is the set of all edges such that  $w_i = q$  (with  $q > 0$ ).

The sets  $T_q$  are distinct, but  $T$  and  $T_q$  might have a nonempty intersection. However, we still have

$$F(\mathbf{p}) = \omega(\mathbf{p}) \mathbf{E} \left[ \prod_{q>0} \prod_{i \in T_q} B(y_i)^q \prod_{i \in T} B'(y_i) \right].$$

We note  $\mathbf{1}_\Omega = \prod_{i \in T} B'(y_i)$ . Some of the edges  $y_i$  with  $i \in T$  might also appear in the proto-path before  $p$ , and in this case  $B(y_i)B'(y_i) = (1 - 1/M)B'(y_i)$ ; we must keep track of these edges. We define:

- $T'_q = \{i : w_i = q, w'_i > 0\}$  and  $d'_q = \#T'_q$ ,
- $T_q^* = \{i : w_i = q, w'_i = 0\}$  so that  $T'_q \cup T_q^* = T_q$  and  $T'_q, T_q^*$  are disjoint.

Through the definition of  $T_q^*$ , we see that  $|T_1^*| = a_1$ , the number of simple (and consistent) edges of the proto-path, before  $p$ . Noting  $\zeta = \sum_{q>0} qd'_q$ , this yields

$$F(\mathbf{p}) = \omega(\mathbf{p})(1 - 1/M)^\zeta \mathbf{E} \left[ \mathbf{1}_\Omega \prod_{q>0} \prod_{i \in T_q^*} B(y_i)^q \right]. \quad (2.11.1)$$

The greatest contribution to the expectation (2.11.1) is due to the  $q = 1$  factor, so we are going to split the edges into two parts, those matched with another edge in some  $T_q^*$  and those who are not.

- $\hat{T}_1$  is the set of all  $i \in T_1^*$  such that there is a  $j$  in  $T_q^*$  for some  $q > 0$ , such that if  $y_i = (\mathbf{e}, \mathbf{f})$  and  $y_j = (\mathbf{e}', \mathbf{f}')$ , then either  $\sigma(\mathbf{e}) = \mathbf{f}'$  or  $\sigma(\mathbf{e}') = \mathbf{f}$  (or maybe both).
- For every  $q > 1$ ,  $\hat{T}_q$  is the set of all  $i \in T_q^*$  such that there is a  $j$  in  $T_1^*$ , such that if  $y_i = (\mathbf{e}, \mathbf{f})$  and  $y_j = (\mathbf{e}', \mathbf{f}')$ , then either  $\sigma(\mathbf{e}) = \mathbf{f}'$  or  $\sigma(\mathbf{e}') = \mathbf{f}$  (or maybe both).
- Finally, note  $S_q = T_q^* \setminus \hat{T}_q$ . If  $i$  is in  $S_1$  and  $y_i = (\mathbf{e}, \mathbf{f})$ , then either  $\sigma(\mathbf{e}) = \mathbf{f}$ , or  $\sigma(\mathbf{e})$  is some tail  $\mathbf{f}$  which does not belong to any other edge of the proto-path  $\mathbf{p}$ .

Those sets are random as they depend on the environment  $\sigma$ . Finally, note  $X_q = \prod_{i \in S_q} B(y_i)^q$  and  $\zeta' = \sum_{q \geq 1} q|\hat{T}_q|$ . Then, we have

$$\mathbf{1}_\Omega \prod_{q>0} \prod_{i \in T_q^*} B(y_i)^q = \left( \frac{-1}{M} \right)^{\zeta'} \mathbf{1}_\Omega \prod_{q>0} X_q. \quad (2.11.2)$$

**First conditioning**

Let  $\mathcal{F}$  be the sigma-algebra generated by

- the event  $\Omega$ ,
- the matchings  $\sigma(\mathbf{e})$  and  $\sigma^{-1}(\mathbf{f})$  for every  $y_i = (\mathbf{e}, \mathbf{f})$  with  $i$  not in  $S_1$ .

**LEMMA 2.11.2.** *With the notations given above, if  $n$  is large enough we have*

$$|\mathbf{E}[X_1 | \mathcal{F}]| \leq 8 \left( \frac{3cN}{M\sqrt{M}} \right)^{|S_1|}. \quad (2.11.3)$$

The proof of this lemma relies on the following remark:  $|S_1|$  is measurable with respect to  $\mathcal{F}$ , so if  $H$  is the number of  $i \in S_1$  such that  $\sigma(\mathbf{e}) \neq \mathbf{f}$ , then

$$\mathbf{E}[X_1 | \mathcal{F}] = \mathbf{E} \left[ \left( 1 - \frac{1}{M} \right)^{|S_1| - H} \left( \frac{1}{M} \right)^H \right]. \quad (2.11.4)$$

We first give the law of  $H$  conditionally on  $\mathcal{F}$ . For simplicity we note  $r = |S_1|$ .

**LEMMA 2.11.3.** *Given  $\mathcal{F}$ , for every  $k$ , we have*

$$\mathbf{P}(H = k | \mathcal{F}) = \frac{\binom{r}{k} (M - a)_k}{\sum_{k=0}^r \binom{r}{k} (M - a)_k}. \quad (2.11.5)$$

*Proof.* Let us count the favorable cases for the event  $\{H = k\}$  (again, reasoning conditionally on  $\mathcal{F}$ ). We have to choose those  $k$  edges among the  $r$  that haven't been matched yet. Once they have been chosen, all the  $r - k$  remaining ones have to be matched with one tail not belonging to any edge in the proto-path  $\mathfrak{p}$ , and those edges are exactly  $M - a$ . Thus there are  $\binom{r}{k} (M - a)_k$  favorable cases. The sum in the denominator is the sum of all cases.  $\square$

The reader can check that if  $a \leq \sqrt{M}$ , then if  $n$  is large enough, for every  $k \leq a$ , we have  $(M - a)_k \geq (M - a)^k / 2$ , so if we note  $Z = Z(a, r, M) = \sum_{k=0}^r \binom{r}{k} (M - a)_k$  then we have

$$Z \geq \frac{1}{2} \sum_{k=0}^r \binom{r}{k} (M - a)^k = \frac{1}{2} (M - a + 1)^r \geq \frac{1}{2} (M - a)^r. \quad (2.11.6)$$

On the other hand,

$$\mathbf{E}[X | \mathcal{F}] = \frac{1}{Z} \sum_{k=0}^r \binom{r}{k} (M - a)_k \left( 1 - \frac{1}{M} \right)^k \left( \frac{-1}{M} \right)^{r-k}. \quad (2.11.7)$$

$$= \frac{(-1)^r}{Z} \mathbf{E}[(M - a)_Q (-1)^Q] \quad (2.11.8)$$

where  $Q$  is a random variable with law  $\mathcal{B}(r, 1/M)$ . Note that

$$\begin{aligned} (M - a)_Q (-1)^Q &= \prod_{n=0}^{Q-1} (M - a - n) \times (-1) \\ &= \prod_{n=0}^{Q-1} (n - (M - a)). \end{aligned}$$

We now use ([33], Lemma 9):

**LEMMA 2.11.4.** *Let  $z \geq 1, r \in \mathbb{N}^*$  and  $0 < p \leq q < 1$ . Let  $B$  a binomial random variable with parameters  $r, p$ . If  $8(1 - p/q)^2 \leq 2zqr^2 \leq 1$ , then*

$$\left| \mathbf{E} \left[ \prod_{n=1}^B \left( zn - \frac{1}{q} \right) \right] \right| \leq 4(r\sqrt{8zq})^r. \quad (2.11.9)$$

We apply the lemma with

- $q = 1/(M - a)$  and  $p = 1/M$  (they satisfy  $p \leq q$ ),
- the random variable  $Q$  as  $B$ ,
- $z = 1$  (we can check that the condition of the lemma is verified).

Then, the lemma yields

$$|\mathbf{E}[(-1)^{Q(M-a)Q}]| \leq 4 \left( r\sqrt{\frac{8}{M-a}} \right)^r. \quad (2.11.10)$$

We now plug this into  $|\mathbf{E}[X_1 | \mathcal{F}]|$ . Using this and the preliminary remark on  $n$  large and using inequality (2.11.6), we get

$$|\mathbf{E}[X_1 | \mathcal{F}]| \leq \frac{4}{Z} \left( \frac{3r}{\sqrt{M-a}} \right)^r \quad (2.11.11)$$

$$\leq \frac{8}{(M-a)^r} \left( \frac{3r}{\sqrt{M-a}} \right)^r \quad (2.11.12)$$

$$\leq 8 \left( \frac{3cN}{M\sqrt{M}} \right)^r \quad (2.11.13)$$

This ends the proof of Lemma 2.11.2. As a consequence, we get

$$\mathbf{E} \left[ \mathbf{1}_\Omega \prod_{q>0} \prod_{i \in T_q^*} B(y_i)^q \right] \leq 8 \mathbf{E} \left[ \left( \frac{3cN}{M\sqrt{M}} \right)^{|S_1|} \left( \frac{1}{M} \right)^{\zeta'} \mathbf{1}_\Omega \prod_{q>1} |X_q| \right]. \quad (2.11.14)$$

### Second conditioning

Let  $\mathcal{G}_i$  be the  $\sigma$ -algebra generated by

- the event  $\Omega$ ,
- the matchings  $\sigma(\mathbf{e})$  and  $\sigma^{-1}(\mathbf{f})$  for every  $y_j = (\mathbf{e}, \mathbf{f})$  with  $i \neq j$ .

The random variables  $\zeta', |S_q|$  are  $\mathcal{G}_i$ -measurable. Fix  $i$  in some  $S_q$ . Then, as  $q > 1$  we have

$$\begin{aligned} \mathbf{E}[|B(y_i)|^q | \mathcal{G}_i] &\leq \mathbf{E}[|B(y_i)|^2 | \mathcal{G}_i] \\ &= \left( 1 - \frac{1}{M} \right)^2 \mathbf{P}(\sigma(\mathbf{e}) = \mathbf{f} | \mathcal{G}_i) + \frac{1}{M^2} \mathbf{P}(\sigma(\mathbf{e}) \neq \mathbf{f} | \mathcal{G}_i) \end{aligned}$$

Conditionnally on  $\mathcal{G}_i$ , the head  $\mathbf{e}$  cannot be matched with a tail belonging to  $y_j$  for  $j \neq i$  (recall the definition of  $S_p$ ). Hence, if  $M_i$  is the total number of unmatched tails after the matching of all the

heads belonging to some  $y_j$ , we have  $\mathbf{P}(\sigma(\mathbf{e}) = \mathbf{f} | \mathcal{G}_i) = 1/M_i$ . Remember that if  $n$  is large enough, we have  $1/(M-a) \leq c/M$  (see the preliminary remark). Hence, we have

$$\begin{aligned} \mathbf{E}[|B(y_i)|^q | \mathcal{G}_i] &\leq \mathbf{E}[|B(y_i)|^2 | \mathcal{G}_i] \\ &\leq c \left(1 - \frac{1}{M}\right)^2 \frac{1}{M} + c \frac{1}{M^2} \left(1 - \frac{1}{M}\right) \\ &\leq c \frac{1}{M} \left(1 - \frac{1}{M}\right) \leq \frac{c}{M}. \end{aligned}$$

By conditioning repeatedly on all the  $\mathcal{G}_i$  for every  $i$  in some  $S_q$ , for  $q > 1$ , we get

$$\mathbf{E} \left[ \mathbf{1}_\Omega \prod_{p>0} \prod_{i \in T_p^*} B(y_i)^p \right] \leq 8 \mathbf{E} \left[ \left( \frac{3cN}{M\sqrt{M}} \right)^{|S_1|} \left( \frac{c}{M} \right)^{\zeta' + \sum_{q>1} q|S_q|} \mathbf{1}_\Omega \right].$$

As  $\zeta' = \sum_{q>0} q|\hat{T}_q|$ , we have  $\zeta' + \sum_{q>1} q|S_q| = |\hat{T}_1| + \sum_{q>1} q(|\hat{T}_q| + |S_q|) = |\hat{T}_1| + \sum_{q>1} q|T_q^*|$ .

### Third conditioning

We now condition on the sigma-algebra  $\mathcal{G}$  generated by the matchings  $\sigma(\mathbf{e})$  and  $\sigma^{-1}(\mathbf{f})$  for every  $y_j = (\mathbf{e}, \mathbf{f})$  with  $i \notin T$ . Note  $\tilde{\Omega}$  the event "no half-edge belonging to  $y_i$  for some  $i \notin T$  has been matched with a half-edge  $y_j$  with  $j$  in  $T$ ". This event is  $\mathcal{G}$ -measurable, and when  $n$  is large enough,

$$\mathbf{E}[\mathbf{1}_\Omega | \mathcal{G}] \leq \mathbf{1}_{\tilde{\Omega}} \left( \frac{c}{M} \right)^{|T|} \leq \left( \frac{c}{M} \right)^{|T|}.$$

Hence,

$$\mathbf{E} \left[ \mathbf{1}_\Omega \prod_{q>0} \prod_{i \in T_q^*} B(y_i)^q \right] \leq 8 \mathbf{E} \left[ \left( \frac{3cN}{M\sqrt{M}} \right)^{|S_1|} \left( \frac{c}{M} \right)^{|\hat{T}_1| + \sum_{q>1} q|T_q^*| + |T|} \right]. \quad (2.11.15)$$

### Endstep

Recall (2.11.1); we have

$$F(\mathbf{p}) = \omega(\mathbf{p})(1 - 1/M)^\zeta \mathbf{E} \left[ \mathbf{1}_\Omega \prod_{q>0} \prod_{i \in T_q^*} B(y_i)^q \right] \quad (2.11.16)$$

$$\leq 8 \omega(\mathbf{p})(1 - 1/M)^\zeta \mathbf{E} \left[ \left( \frac{3cN}{M\sqrt{M}} \right)^{|S_1|} \left( \frac{c}{M} \right)^{|\hat{T}_1| + \sum_{q>1} q|T_q^*| + |T|} \right] \quad (2.11.17)$$

$$\leq 8 \omega(\mathbf{p}) \left( \frac{3cN}{M\sqrt{M}} \right)^{a_1} \mathbf{E} \left[ \left( \frac{3cN}{M\sqrt{M}} \right)^{-|\hat{T}_1|} \left( \frac{c}{M} \right)^{|\hat{T}_1| + \sum_{q>1} q|T_q^*| + |T|} \right] \quad (2.11.18)$$

$$\leq 8 \omega(\mathbf{p}) \left( \frac{3cN}{M\sqrt{M}} \right)^{a_1} \mathbf{E} \left[ \left( \frac{3N}{\sqrt{M}} \right)^{-|\hat{T}_1|} \left( \frac{c}{M} \right)^{\sum_{q>1} q|T_q^*| + |T|} \right]. \quad (2.11.19)$$

where in the third line we used  $a_1 = |T_1^*| = |S_1| + |\hat{T}_1|$ . By construction, we have  $\sum_{q>0} |T_q^*| + |T| = a$ , therefore

$$\sum_{q>1} q|T_q^*| + |T| = a - |T_1^*| + \sum_{q>1} (q-1)|T_q^*| \geq a - a_1 \quad (2.11.20)$$

and we have  $(c/M)^{\sum_{q>1} q|T_q^*|+|T|} \leq (c/M)^{a-a_1}$ . This finally yields

$$F(\mathfrak{p}) \leq 8\omega(\mathfrak{p}) \left(\frac{c}{M}\right)^{a-a_1} \left(\frac{3cN}{M\sqrt{M}}\right)^{a_1} \mathbf{E} \left[ \left(\frac{3N}{\sqrt{M}}\right)^{-|\hat{T}_1|} \right].$$

In the next lemma, we bound the expectation on the right side.

**LEMMA 2.11.5.** *If  $n$  is large enough,*

$$\mathbf{E} \left[ \left(\frac{3N}{\sqrt{M}}\right)^{-|\hat{T}_1|} \right] \leq 3. \quad (2.11.21)$$

*Proof.* We have

$$\mathbf{E} \left[ \left(\frac{3N}{\sqrt{M}}\right)^{-|\hat{T}_1|} \right] = \sum_{\ell=0}^{\infty} \mathbf{P}(|\hat{T}_1| = \ell) \left(\frac{\sqrt{M}}{3N}\right)^{\ell}.$$

Using the pigeonhole principle, on the event  $\{|\hat{T}_1| = \ell\}$ , at least  $\lfloor \ell/2 \rfloor$  couples of edges  $(y, y')$  are “mismatched”, which means that  $\sigma(\mathbf{e}) = \mathbf{f}'$  or  $\sigma(\mathbf{e}') = \mathbf{f}$ . A (very) crude bound for the choice of those  $\lfloor \ell/2 \rfloor$  couples is  $(a^2)^{\lfloor \ell/2 \rfloor}$ . For each choice of those  $\lfloor \ell/2 \rfloor$  couples, the probability that they are indeed mismatched is at most  $(1/(M-a))^{\lfloor \ell/2 \rfloor}$  which is smaller than  $(2/\sqrt{M})^{\ell}$  if  $n$  is large enough. In the end, we get

$$\mathbf{P}(|\hat{T}_1| = \ell) \leq a^{\ell} \left(\frac{2}{\sqrt{M}}\right)^{\ell}$$

Finally, as  $a \leq N$ , we have

$$\mathbf{E} \left[ \left(\frac{3t}{\sqrt{M}}\right)^{-|\hat{T}_1|} \right] \leq \sum_{\ell=0}^{\infty} \left(\frac{2a}{3N}\right)^{\ell} \leq \sum_{\ell=0}^{\infty} \left(\frac{2}{3}\right)^{\ell}$$

which ends the proof of the lemma.  $\square$

We finally get the desired bound, that is

$$|F(\mathfrak{p})| \leq 24 \cdot \omega(\mathfrak{p}) \left(\frac{c}{M}\right)^a \left(\frac{N}{\sqrt{M}}\right)^{a_1} \quad (2.11.22)$$

### 2.11.2 Proof, part II: some edges are not consistent

We now suppose some edges are not consistent: for example, there might be in  $\mathfrak{p}$  two edges having the form  $y = (\mathbf{e}, \mathbf{f})$  and  $y' = (\mathbf{e}, \mathbf{f}')$  with  $\mathbf{f} \neq \mathbf{f}'$ . Without loss of generality we can suppose  $y = y_1$  and  $y' = y_a$ . The contributions of those two edges in the product has the form  $B(y)^w B'(y)^z B(y')^{w'} B'(y')^{z'}$ . Note that  $B'(y)B'(y')$  is always zero. From this, we see that we can't have  $z$  and  $z'$  be both non zero. Without loss of generality, we suppose that  $z' = 0$ .

**First case:**  $z \neq 0$ .

Here, we immediately have

$$B(y)^w B'(y)^z B(y')^{w'} = B'(y) B(y)^w \left(-\frac{1}{M}\right)^{w'}. \quad (2.11.23)$$

This expression does not longer rely upon  $y'$ . Hence, in this case, we have

$$F(\mathfrak{p}) = \frac{1}{(d_{\mathbf{e}}^+)^{w'}} \left(-\frac{1}{M}\right)^{w'} F(\mathfrak{q}) \quad (2.11.24)$$



where the proto-path  $q$  is the proto-path  $p$  without all the  $w'$  instances of the  $y'$  edge. This new proto-path  $q$  has length  $N - w'$ , has  $d' = a - 1$  distinct edges before  $p$ , and its number of simple, consistent edges before  $p$  is greater than  $a_1$ .

**Second case:**  $z = 0$ .

The product is now reduced to  $B(y)^w B(y')^{w'}$ . After a short development we find that

$$B(y)^w B(y')^{w'} = B(y)^w \left(-\frac{1}{M}\right)^{w'} + \left(-\frac{1}{M}\right)^{w'} B(y)^w - \left(-\frac{1}{M}\right)^{w'+w} \quad (2.11.25)$$

Hence,  $F(p)$  splits into three parts:

$$F(p) = \left(-\frac{1}{d_e^+ M}\right)^{w'} F(q) + \left(-\frac{1}{d_e^+ M}\right)^w F(q') - \left(-\frac{1}{d_e^+ M}\right)^{w'+w} F(q''). \quad (2.11.26)$$

All the three new proto-paths  $q, q', q''$  now have

- length  $N - w', t - w$  and  $N - w' - w$ ,
- at most  $a - 1$  distinct edges,
- less inconsistent edges than  $p$ .

### Iteration of the procedure.

We repeat the procedure as many times as needed to get rid of every inconsistent edge. Each step gives rise to at most 3 terms having the form

$$\pm \left(\frac{1}{M}\right)^\alpha \omega(p) \mathbf{E} \left[ \prod_{i=1}^{a-1} B(y_i)^{w_i} B(y_i) \right]$$

or

$$\pm \left(\frac{1}{M}\right)^\alpha \omega(p) \mathbf{E} \left[ \prod_{i=2}^a B(y_i)^{w_i} B(y_i) \right]$$

or

$$\pm \left(\frac{1}{M}\right)^\alpha \omega(p) \mathbf{E} \left[ \prod_{i=2}^{a-1} B(y_i)^{w_i} B(y_i) \right]$$

where  $\alpha$  is either  $w_a, w_1$  or  $w_a + w_1$ .

Now, we repeat the procedure for each term. Each step removes one inconsistent edge, so there are no more than  $3^b$  steps, and in the end we get at most  $3^b$  terms. In each one of the final  $3^b$  terms, all edges are consistent so we can apply (2.11.22). The number of simple, consistent edges of those new proto-paths is greater than  $a_1$  but still smaller than  $N$ . Hence, applying (2.11.22) to each term, we can bound  $|F(p)|$  with at most  $3^b$  terms having the form

$$24 \cdot \omega(p) \left(\frac{c}{M}\right)^a \left(\frac{N}{\sqrt{M}}\right)^{a_1}$$

which yields the final desired result (2.4.1).

## Chapter 3

# An Erdős-Gallai generalization

*This chapter is drawn from the published paper [41], a joint work with Charles Bordenave.*

### 3.1 Reconstruction of a graph with its universal covering

Let  $G = (V, E)$  be a finite graph. A graph  $G' = (V', E')$  is a *covering* of  $G$  if there is a surjective map  $p : V' \rightarrow V$  which is a local isomorphism: for every  $x \in V'$ , the map  $p$  induces a bijection between the edges incident to  $x$  and the edges incident to  $p(x)$ . This is the graph-theoretic analog of the notion of covering space in topology; note that some authors tend to use the words *lift* instead of *covering*. A comprehensive presentation on graph coverings can be found in [100].

#### 3.1.1 The universal covering of a graph

A *universal covering* of  $G$  is a covering of  $G$  which is a tree. When the graph  $G$  is a tree, then it is a covering of itself. However, if the graph  $G$  contains a cycle, then any universal covering of  $G$  is infinite: starting from a cycle  $(x_1, \dots, x_k)$  in  $G$ , one can easily find an infinite path in  $G'$  by taking  $y_1 \in p^{-1}(x_1)$ , then  $y_2 \in p^{-1}(x_1)$  the only antecedent of  $x_1$  adjacent to  $y_1$ , and so on until one reaches  $y_k \in p^{-1}(x_k)$ ; then pick  $y_{k+1} \in p^{-1}(x_1)$  which is adjacent to  $y_k$ . This last  $y_{k+1}$  is distinct from  $y_1$  because the covering  $G'$  is a tree. We can continue this construction forever and we end with an infinite path in  $G'$ .

If  $G_1, G_2$  are two universal coverings of  $G$  with maps  $p_1, p_2$ , then one might pick a vertex  $x$  of  $G_1$  and a vertex  $y$  of  $G_2$  such that  $p_1(x) = p_2(y)$  and define a morphism  $\phi : G_1 \rightarrow G_2$  by setting  $\phi(x) = y$ , and then mapping every neighbor  $x'$  of  $x$  to the unique vertex  $y'$  in  $G_2$  which is adjacent to  $y$  and has  $p_2(y') = p_1(x')$ , then do the same thing for the neighbors of  $x'$ , and so on. If  $G$  is connected, then  $G_1, G_2$  are also connected, and the resulting map  $\phi$  can easily be seen to be an isomorphism between  $G_1$  and  $G_2$ : up to isomorphism, there is only one universal cover of any connected graph, and it is denoted by  $T_G$ .

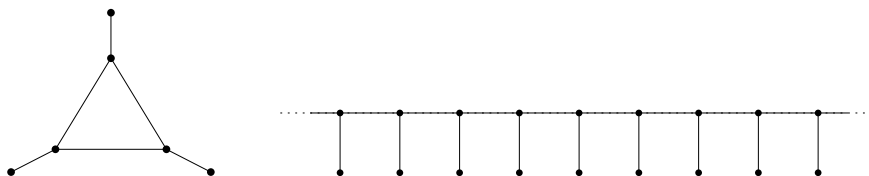


Figure 3.1 – On the right, the rooted product  $K_3 \circ K_2$  and on the left, its universal covering.

Let us give three examples:

- the universal cover of any  $d$ -regular graph with  $d \geq 2$  is the  $d$ -regular tree  $\mathbb{T}_d$ ;

- the universal cover of the complete bipartite graph  $K_{m,n}$  is the biregular tree  $\mathbb{T}_{m,n}$  where each vertex of degree  $m$  has only neighbors of degree  $n$ , and each vertex of degree  $n$  has only neighbors of degree  $m$ ;
- the universal covering of the graph depicted on the left of Figure 3.1 is the infinite comb in the right of Figure 3.1.

Universal coverings of graphs can be seen as trees which are good local approximations of  $G$ . They have widely been used in the context of distributed computing ([13], [14]). In the context of random graphs, where many models of graphs are locally tree-like, the universal cover is a good approximation of the local structure of the graph, which conveys information on crucial quantities such as the spectrum ([84], [12]).

### 3.1.2 Neighborhoods in the universal covering and the reconstruction problem

Let  $h$  be a positive integer. Given any vertex  $x$  of  $G$ , its  $h$ -depth universal covering neighborhood is the unlabeled ball of radius  $h$  in  $T_G$  around any antecedent of  $x$  by  $p$ . One can easily see that this ball does not depend — up to isomorphism — on the chosen antecedent  $x$ . Informally, this is the neighborhood of  $x$  in  $G$ , but where cycles have been forgotten.

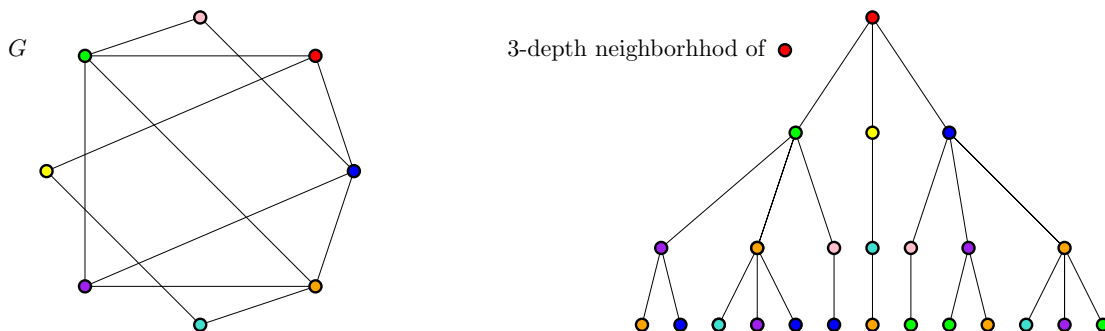


Figure 3.2 – The graph  $G$  is on the left ; the 3-depth neighborhood of the red vertex is drawn on the right. Two vertex having the same color in  $T_G$  are antecedents of the vertex with the corresponding color in  $G$ .

Our main goal in this paper will be to give a characterization of the collections of trees which correspond to  $h$ -depth neighborhoods of vertices in the universal covering of the graph.

**QUESTION 3.1.1.** Let  $\mathbf{t} = (t_1, \dots, t_n)$  be a collection of  $n$  unlabeled rooted trees with maximal depth  $h$ . Is it the collection of  $h$ -depth universal covering neighborhoods of some graph  $G$  ?

If this is the case, we say that the  $n$ -tuple  $\mathbf{t}$  is *graphical* and we say that  $G$  is a *realization* of  $\mathbf{t}$ . In the sequel, we will sometimes say “a  $n$ -tuple  $\mathbf{t}$ ” instead of “a  $n$ -tuple of unlabeled rooted trees with maximal depth  $h$ ”; the height  $h$  is fixed once for all.

### 3.1.3 Notation

From now on, we will adopt the term “tree” instead of “unlabeled, rooted tree”, unless explicitly stated otherwise. The set of all trees with depth  $h$  will be noted  $\mathcal{T}_h$ .

If  $x$  is a vertex of a graph  $G$ , we will note  $\deg_G(x)$  the number of neighbors of  $x$  in  $G$ .

We will frequently refer to directed graphs as *digraphs*. In a digraph  $G$ , we will note  $\deg_G^\pm$  the in and out degree of the vertices.

Generally, the root of a (rooted) graph will be noted  $\bullet$  and if  $k$  is an integer and  $g$  a rooted graph,  $(g)_k$  denotes the ball  $B_g(\bullet, k)$  of radius  $k$  around the root of  $g$ .

We will frequently use the notation  $x \wedge y$  to denote the minimum of two numbers  $x$  and  $y$ .

## 3.2 Related work

*Graph reconstruction problems* ask the following question : given any property  $\mathcal{P}$  about graphs, how can we ensure that there is a graph (or digraph, or multigraph) having this property  $\mathcal{P}$  ? What is the number of graphs that have this property ? Can we determine the properties  $\mathcal{P}$  that have a single graph realization ?

Reconstructing a graph (or bipartite graph, or digraph) only by the list of its degree has been a well-known and studied problem since the seminal works of Erdős, Gallai and many others. In fact, question 3.1.1 had been settled long time ago for  $h = 1$  by the celebrated Erdős-Gallai theorem. Suppose that  $\mathbf{t} = (t_1, \dots, t_n)$  is an  $n$ -tuple of 1-depth trees. A 1-depth tree  $t_i$  is just a root with some leaves, say  $d_i$  leaves ; thus,  $\mathbf{t}$  can be identified with a  $n$ -tuple of integers  $(d_1, \dots, d_n)$ .

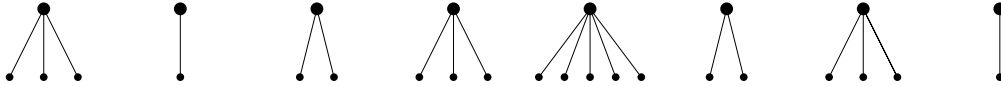


Figure 3.3 – Is there a graph on 8 vertices with this 1-depth neighborhood ? The associated degree sequence if  $\mathbf{d} = (3, 1, 2, 3, 5, 2, 3, 1)$ . According to Theorem 5, yes.

Finding a graph  $G$  with  $\mathbf{t}$  as 1-depth neighborhood boils down to finding a graph  $G$  with degree sequence  $\mathbf{d}$  — such sequences are called *graphical*. Not all integer sequences are graphical; the Erdős-Gallai theorem gives two conditions which are altogether necessary and sufficient for a sequence to be graphical.

**Theorem 5** (Erdős, Gallai, [73]). *Let  $\mathbf{d} = (d_1, \dots, d_n)$  be a  $n$ -tuple of integers. Rearrange them in decreasing order  $d_{(1)} \geq \dots \geq d_{(n)}$ . Then,  $\mathbf{d}$  is graphical if and only if it satisfies the two following conditions :*

$$d_1 + \dots + d_n \text{ is even,} \quad (3.2.1)$$

and the “Erdős-Gallai condition”

$$\forall k \in [n], \quad d_{(1)} + \dots + d_{(k)} \leq k(k-1) + \sum_{i=k+1}^n d_{(i)} \wedge k. \quad (3.2.2)$$

A short and constructive proof is available at [138]. In fact, the Erdős-Gallai condition is not the only sufficient and necessary condition for an integer sequence to be graphical ; there are some other equivalent conditions, notably listed in [131]. The corresponding realization problem for digraphs had also been solved quite early; see the interesting note [28] for a complete history and presentation of the many variants. When a double sequence  $\mathbf{d}^\pm = (d_i^+, d_i^-)_{i \in [n]}$  is the sequence of in and out degrees of a digraph, we say that  $\mathbf{d}^\pm$  is *digraphical*.

**Theorem 6.** *Let  $\mathbf{d}^\pm = (d_i^+, d_i^-)_{i \in [n]}$  be a  $2n$ -tuple of integers. We order the first component by decreasing lexicographic order:  $d_{(1)}^+ \geq \dots \geq d_{(n)}^+$  and if  $d_{(i)}^+ = d_{(j)}^+$  with  $i < j$ , then  $d_{(i)}^- \geq d_{(j)}^-$ . Then,  $\mathbf{d}^\pm$  is digraphical if and only if it satisfies the two following conditions :*

$$\sum_{i=1}^n d_i^+ = \sum_{i=1}^n d_i^-, \quad (3.2.3)$$

and the “directed Erdős-Gallai condition” :

$$\forall k \in [n], \quad \sum_{i=1}^k d_{(i)}^+ \leq \sum_{i=1}^k d_{(i)}^- \wedge (k-1) + \sum_{i=k+1}^n k \wedge d_{(i)}^-. \quad (3.2.4)$$

This settled our question for  $h = 1$ . The case  $h = 2$  had recently been solved by [19, 16]; in those papers, a 2-depth neighborhood is called a *neighborhood degree list* (NDL). In [16], the authors not only settle Question 3.1.1 and give a sufficient and necessary condition for a NDL to be graphical, but they also characterize those NDL that are “unigraphical”, meaning that they have a unique graphical realization — we do not address this problem, but we solve Question 3.1.1 for arbitrary depths  $h$ .

For  $h = 1$ , the number of labeled graphs with a given degree sequence is asymptotically known in many asymptotic regimes, see notably [88], [32, Theorem 2.16] and references therein. For general  $h$ , this question has been recently addressed in [35] in the regime where the maximal degree is uniformly bounded. The motivation came from the Benjamini-Schramm topology of rooted graphs.

In this paper, we only deal with neighborhoods *in the universal covering*, thus ignoring the eventual cycles in the  $h$ -neighborhood of a vertex. If some  $n$ -tuple  $\mathbf{t}$  is graphical, then it might as well have very different realizations, for instance ones that are  $h$ -locally tree-like, or others having many short cycles. When the same question is addressed with the  $h$ -neighborhoods *in the graph*, Question 3.1.1 becomes much more arduous; a similar problem in graph reconstruction, the famous Kelly-Ulam reconstruction problem, was asked during the 1940s and still remains opened.

### 3.3 Definitions and statement of the main result

Fix some  $n$ -tuple  $\mathbf{t} = (t_1, \dots, t_n)$  where  $t_i \in \mathcal{T}_h$ . The associated degree sequence  $\mathbf{d} = (d_1, \dots, d_n)$  is the sequence of degrees of the root  $\bullet$  of the tree  $t_i$ , that is  $d_i = \deg_{t_i}(\bullet)$ . An obvious necessary condition for  $\mathbf{t}$  to be graphical is that  $\mathbf{d}$  is itself a graphical sequence, hence satisfying (3.2.1)-(3.2.2). From now on, we will assume that  $d_1 + \dots + d_n = 2m$  where  $m$  is an integer.

Let  $t$  be a tree with depth at most  $h$  and root  $\bullet$ . Let  $e$  be an edge incident with the root, say  $e = (\bullet, x)$ . The tree  $t \setminus e$  has exactly two connected components. The connected component containing the root is  $r'$  and the other one is  $s$ ; we root  $s$  at  $x$ . We erase from  $r'$  all the vertices at depth exactly  $h$  in  $t$ , and we keep the same root; this yields a new rooted tree  $r$  — see Figure 3.4. The *type* of the edge  $e$  is defined as the couple of rooted trees  $(r, s)$  and we will denote it by  $\tau(e)$ .

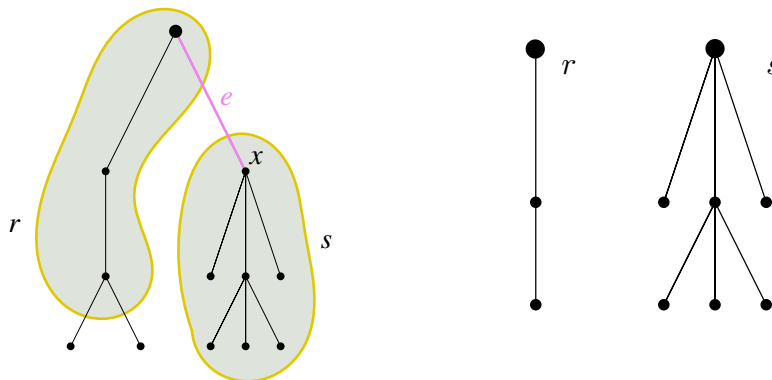


Figure 3.4 – Construction of the type  $\tau(e) = (r, s)$  of edge  $e$  in some tree  $t$ .

If  $\tau = (r, s)$  is a type, its *opposite type*  $\tau^{-1}$  is defined as  $(s, r)$ . A type is an element of  $\mathcal{T}_{h-1} \times \mathcal{T}_{h-1}$ . The set of all types induced by the edges in the elements of  $\mathbf{t}$  is noted  $\text{types}(\mathbf{t})$ . It can be decomposed into the disjoint union of three sets

$$\text{types}(\mathbf{t}) = \Delta \cup A \cup B$$

where

- $\Delta$  is the set of “diagonal” types  $\tau = (r, r)$  for some  $r \in \mathcal{T}_{h-1}$ ;
- $A \cup B$  is the set of types  $\tau = (r, s)$  with  $r \neq s$ , and the sets  $A, B$  are chosen such that if  $\tau \in A$ , then  $\tau^{-1} \in B$ .

If  $\tau \in \text{types}(\mathbf{t})$ , we define

- the  $\tau$ -degree of any index  $i \in [n]$  as the number of edges in  $t_i$  incident to the root and whose type is  $\tau$ . We will denote it by  $d_i^\tau$ ;
- the  $\tau$ -number  $N_\tau$  as the total number of edges in elements of  $\mathbf{t}$  with type  $\tau$ , that is

$$N_\tau = \sum_{i \in [n]} d_i^\tau.$$

It should be clear that for every  $i \in [n]$ , we have  $\sum_{\tau \in \text{types}(\mathbf{t})} d_i^\tau = d_i$ .

**Theorem 7.** *Let  $\mathbf{t} = (t_1, \dots, t_n)$  be a  $n$ -tuple of elements of  $\mathcal{T}_h$ ; it is graphical if and only if it satisfies the following conditions:*

- for every  $\tau \in \Delta$ , the integer sequence  $(d_i^\tau)_{i \in [n]}$  is graphical;
- for every  $\tau \in A$ , the integer double sequence  $(d_i^\tau, d_i^{\tau^{-1}})_{i \in [n]}$  is digraphical.

Using classical characterizations of graphical and digraphical sequences given earlier in Theorems 5 and 6, this result can be detailed:

**Theorem 8.** *Let  $\mathbf{t} = (t_1, \dots, t_n)$  be a  $n$ -tuple of elements of  $\mathcal{T}_h$ ; it is graphical if and only if it satisfies the following conditions:*

- for every  $\tau \in \Delta$ , the integer  $N_\tau$  is even and for every  $k \in [n]$  we have

$$\sum_{i=1}^k d_{(i)}^\tau \leq k(k-1) + \sum_{i=k+1}^n d_{(i)} \wedge k, \quad (3.3.1)$$

- for every  $\tau \in A$ , we have  $N_\tau = N_{\tau^{-1}}$  and for every  $k \in [n]$ , we have

$$\sum_{i=1}^k d_{(i)}^\tau \leq \sum_{i=1}^k d_{(i)}^{\tau^{-1}} + \sum_{i=k+1}^n d_{(i)}^{\tau^{-1}} \wedge k \quad (3.3.2)$$

where indices correspond to decreasing lexicographic ordering of the sequence  $(d_i^\tau, d_i^{\tau^{-1}})_{i \in [n]}$  as in Theorem 6.

Note that those conditions together imply that  $(d_1, \dots, d_n)$  is itself a graphical sequence (sum over all the types  $\tau$ ), which is a necessary, but clearly non sufficient condition.

### 3.4 Proof of Theorem 7

We assume without loss of generality  $h \geq 2$ . The conditions are easily seen to be necessary, for if  $\mathbf{t}$  is graphical and  $\tau$  is a type, then

- either  $\tau \in \Delta$  and the graph induced in  $G$  by keeping only the edges  $e$  such that  $\tau(e) = \tau$  has  $(d_i^\tau)_{i \in [n]}$  as its degree sequence,
- either  $\tau \notin \Delta$ ; in this case either  $\tau \in A$  or  $\tau^{-1} \in A$ , so without loss of generality we can assume that  $\tau \in A$ . The graph induced by edges such that  $\tau(e) = \tau$  can be oriented: if  $e = (i, j) \in G_\tau$ , then one vertex  $k \in \{i, j\}$  satisfies  $\tau(e) = \tau$  in  $t_k$ . We orient the edge  $(i, j)$  from  $k$  to the other vertex. This yields a digraph  $\vec{G}_\tau$  with oriented bi-degree sequence  $(d_i^\tau, d_i^{\tau^{-1}})_{i \in [n]}$ , so the second condition of Theorem 7 is met.

We now prove the sufficiency. We suppose that  $\mathbf{t} = (t_1, \dots, t_n)$  is a  $n$ -tuple satisfying the assumptions in Theorem 7 and we build a graph  $G$  which is a realization of  $\mathbf{t}$ . We first fix some type  $\tau$ .

- We suppose in the first time that  $\tau \in A$ , in particular  $\tau = (r, s)$  with  $r \neq s$ . As  $(d_i^\tau, d_i^{\tau^{-1}})_{i \in [n]}$  is digraphical, there is some digraph  $\vec{G}_\tau$  on  $n$  vertices such that  $\deg_{\vec{G}_\tau}^+(i) = d_i^\tau$  and  $\deg_{\vec{G}_\tau}^-(i) = d_i^{\tau^{-1}}$  for every vertex  $i \in [n]$ . We now define a (non-directed) multigraph  $G_\tau$  by simply forgetting the directions of edges in  $\vec{G}_\tau$  — indeed, this multigraph will be proven to be simple in Lemma 3.4.1.
- Else, if  $\tau \in \Delta$ , then by assumption  $(d_i^\tau)_{i \in [n]}$  is graphical and there is a simple graph  $G_\tau$  such that  $\deg_{G_\tau}(i) = d_i^\tau$ .

We now “glue together” the graphs  $G_\tau$  to get our realization of  $\mathbf{t}$ , namely  $G$ . Formally, if  $E(G_\tau)$  denotes the set of edges in  $G_\tau$ , then  $G = ([n], E)$  with the edge set  $E$  being defined as

$$E := \bigcup_{\tau \in \Delta \cup A} E(G_\tau). \quad (3.4.1)$$

The following lemma is the crucial ingredient of the proof of Theorem 7.

**LEMMA 3.4.1.**  *$G$  is a simple graph.*

*Proof.* Suppose that  $G_\tau$  contains a double edge, for instance  $(x, y)$ . We are going to prove the two following facts :

1. first, this double edge can not arise from two distinct  $G_\tau$ . In other words, if  $(x, y) \in G_\tau$ , then  $(x, y) \notin G_{\tau'}$  for every  $\tau' \neq \tau$  ;
2. then we check that for every  $\tau \in A$ , the multi-graph  $G_\tau$  contains no double edge.

Together, those two facts imply that  $G$  is simple : indeed, if there is a double edge, then it can only belong to a single  $G_\tau$  ; but if  $\tau \in A$ ,  $G_\tau$  cannot contain any double edge, and if  $\tau \in \Delta$  then  $G_\tau$  is simple by construction, hence the conclusion.

Suppose that there is a double edge between vertices  $i$  and  $j$ , one belonging to  $G(\tau)$  and the other to  $G(\tau')$  for two types  $\tau = (r, s)$  and  $\tau' = (r', s')$ . We prove that  $\tau = \tau'$ . As manipulating unlabeled rooted trees is quite inconvenient, we will work with two labeled rooted trees  $T_i, T_j$  in the equivalence classes of  $t_i, t_j$ , and the same with  $R, R', S, S'$  which are representatives of the equivalence classes of  $r, r', s, s'$ . We are going to prove that  $R \simeq R'$  and  $S \simeq S'$  (as rooted labeled trees) , hence proving  $r = r'$  and  $s = s'$  as needed. The following arguments are illustrated in Figure 3.5.

- The presence of an edge between  $i$  and  $j$  in  $G_\tau$  has the following consequence : there is an edge  $e$  in  $T_i$ , adjacent with the root, such that  $T_i \setminus e$  has two connected components, one isomorphic with  $S$  and the other having its ball of radius  $h$  isomorphic with  $R$ . On the other hand, as  $(i, j) \in G(\tau')$ , there is an edge  $e'$  such that  $T_i \setminus e'$  has one component isomorphic with  $S'$  and the ball of radius  $h - 1$  of the other is isomorphic with  $R'$ .
- The same holds with  $T_j$ .

It is clear that  $\deg_R(\bullet) + 1 = \deg_{T_i}(\bullet) = d_i$  and also  $\deg_{R'}(\bullet) + 1 = d_i$ , hence  $\deg_R(\bullet) = \deg_{R'}(\bullet)$ . The same is true with  $S, S'$  ; we have just proven that  $(R)_1 \simeq (R')_1$  and  $(S)_1 \simeq (S')_1$ . We are now going to prove that if  $(S)_k \simeq (S')_k$  and  $(R)_k \simeq (R')_k$  for some  $k < h - 1$ , then this is also true with  $k + 1$ .

First, the ball  $(T_i)_{k+1}$  can be decomposed in two ways shown in Figure 3.5 :

$$(T_i)_{k+1} = e \cup (S)_k \cup (R)_{k+1} \quad \text{and} \quad (T_i)_{k+1} = e' \cup (S')_k \cup (R')_{k+1}$$

but as  $(S)_k \simeq (S')_k$ , we can erase both branches pending at  $e$  and  $e'$ , to get  $(R)_{k+1} \simeq (R')_{k+1}$ . The same idea applies to  $T_j$ , to show that  $(S)_{k+1} \simeq (S')_{k+1}$ , hence closing the recurrence. We have proven that  $(S)_{h-1} \simeq (S')_{h-1}$  and  $(R)_{h-1} \simeq (R')_{h-1}$ , thus  $r = r'$  and  $s = s'$  as needed. We thus have proven the first point exposed earlier.

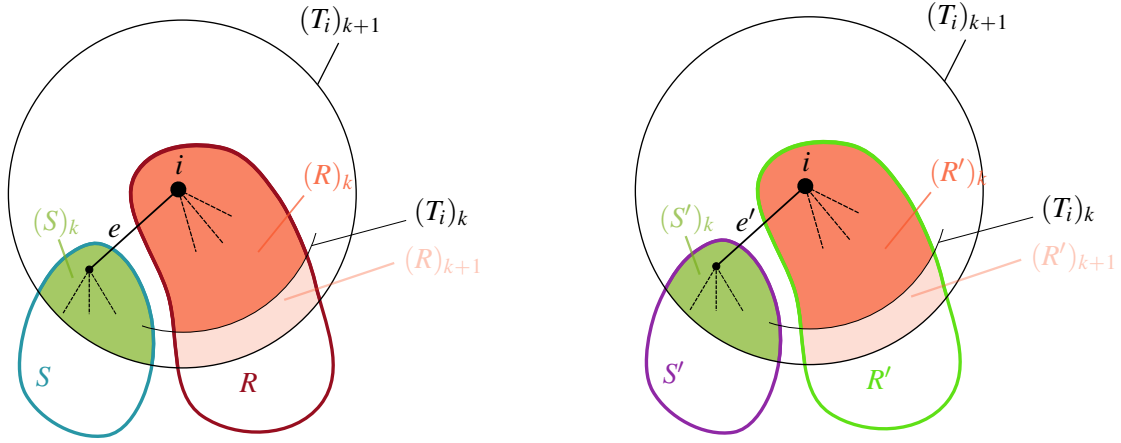


Figure 3.5 – An illustration of the proof of Lemma 3.4.1. The green parts represent  $(S)_k$  and  $(S')_k$  which are isomorphic (as recurrence hypothesis) and the dark red parts are representing  $(R)_k$  and  $(R')_k$ , which are isomorphic too ; hence, the light pink parts are also isomorphic, thus proving  $(R)_{k+1} \simeq (R')_{k+1}$ . A similar procedure applies to  $T_j$ .

We now check the second point, i.e. that for every  $\tau \in A$ , the multi-graph  $G_\tau$  is indeed a simple graph. The proof runs along the same lines : suppose that there is a double directed edge between  $i$  and  $j$  in  $\vec{G}_\tau$ . This can only happen if  $(i, j)$  and  $(j, i)$  are both directed edges in  $\vec{G}_\tau$ . We suppose that  $\tau = (r, s)$ , and with a recurrence we prove that  $r = s$ , hence  $\tau \in \Delta$  which had been discarded since  $\Delta \cap A = \emptyset$ .

To do this, first check that  $\deg_r(\bullet) = \deg_s(\bullet)$ , then suppose that for some  $k < h$ , we have  $(r)_k = (s)_k$  and prove that  $(r)_{k+1} = (s)_{k+1}$ . This step uses the exact same procedure as before.  $\square$

We now check that  $G$  solves our problem.

**LEMMA 3.4.2.**  $G$  is a realization of  $\mathbf{t}$ .

*Proof.* We want to show that the  $h$ -neighborhood of any vertex  $i$  in the universal cover of  $G$  matches  $t_i$ . We show by strong recurrence that for  $k \leq h$ , if  $\bar{t}_i$  denotes the  $h$ -neighborhood of  $i$  in the universal cover, then for every  $i \in [n]$  we have  $(t_i)_k = (\bar{t}_i)_k$ . It is clear by our construction of  $G$  that  $\deg_G(i) = \sum_\tau d_i^\tau = d_i$ , hence  $(t_i)_1 = (\bar{t}_i)_1$ . Now suppose that  $(t_i)_k = (\bar{t}_i)_k$  for some  $k < h$ . If  $N_G(i)$  is the set of neighbors of  $i$  in  $G$ , then for every  $j \in N_G(i)$  we have  $(t_j)_k = (\bar{t}_j)_k$  by the recurrence hypothesis. This readily implies that  $(t_i)_{k+1} = (\bar{t}_i)_{k+1}$ , hence the lemma is proven.  $\square$





## Chapter 4

# Extended states in the spectrum of random trees

*This chapter is drawn from the prepublication [63], a joint work with Justin Salez.*

### 4.1 Introduction

This paper deals with the general question of existence of a non-trivial absolutely continuous part at zero in the adjacency spectrum of unimodular Galton-Watson trees. To motivate our work, let us first briefly describe its implications for the Erdős-Renyi random graph.

#### 4.1.1 The Bauer-Golinelli prediction

Let  $G_n$  be an Erdős-Renyi random graph with size  $n$  and density  $p \in (0, 1)$ . Its adjacency matrix  $A_n$  is a random symmetric  $n \times n$  matrix with zero entries along the diagonal and independent Bernoulli( $p$ ) entries above the diagonal. The associated empirical eigenvalue distribution is

$$\mu_{G_n} := \frac{1}{n} \sum_{k=1}^n \delta_{\lambda_k}, \quad (4.1.1)$$

where  $\lambda_1 \geq \dots \geq \lambda_n$  are the eigenvalues of  $A_n$ . When  $n \rightarrow \infty$  while  $p$  is kept fixed, a celebrated result of Wigner [141] asserts that a suitably rescaled version of  $\mu_{G_n}$  converges weakly in probability to the semi-circle law. This remains true if  $p = p_n$  tends to 0 as  $n \rightarrow \infty$ , as long as  $np_n \rightarrow \infty$  (see, e.g. [137]). The situation changes significantly, however, when instead,

$$np_n \xrightarrow[n \rightarrow \infty]{} c \in (0, \infty). \quad (4.1.2)$$

In this *sparse* regime, the semi-circle law gives place to a non-explicit, densely-discontinuous measure  $\mu_c$ , discovered in [143, 94] and later identified in [42] as the expected spectral measure of the Poisson-Galton-Watson tree with mean offspring  $c$  (see below). The latter has attracted a considerable attention [44, 46, 127, 72, 89], as it captures the asymptotics of many properties of  $G_n$ . One emblematic example is the nullity  $\dim \ker(A_n)$ , which is known to satisfy

$$\frac{1}{n} \dim \ker(A_n) \xrightarrow[n \rightarrow \infty]{\mathbf{P}} \mu_c(\{0\}). \quad (4.1.3)$$

In a remarkable work [22], physicists Bauer and Golinelli used the so-called *replica-symmetric ansatz* to predict the following intriguing formula for the limit in (4.1.3).

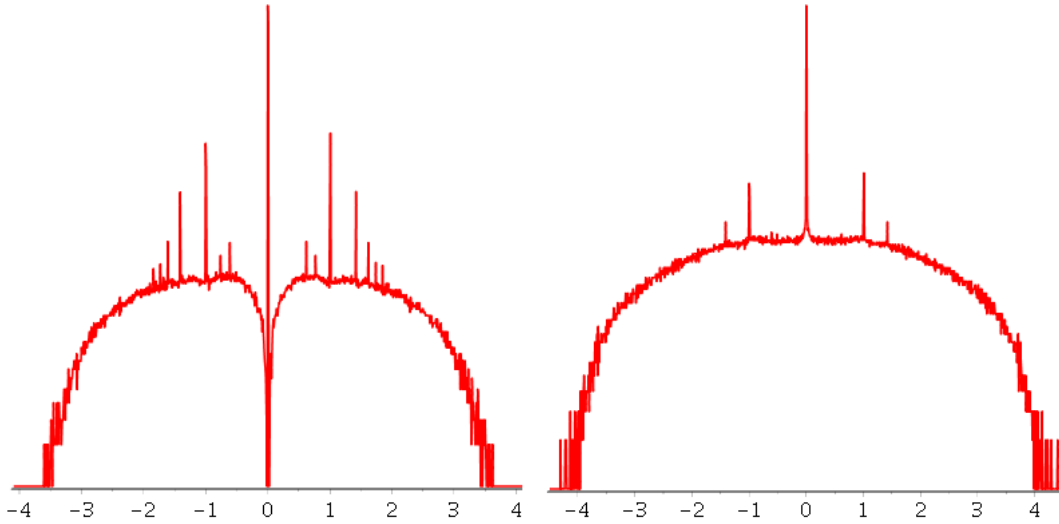


Figure 4.1 – Logarithmic plots of the adjacency spectrum of an Erdős-Renyi random graph of size  $n = 10000$  and average degree  $c = 2$  (left) or  $c = 3$  (right). The presence or absence of an absolutely continuous part at zero in the  $n \rightarrow \infty$  limit is already manifest on these finitary plots.

**CONJECTURE 2** (Atomic mass at zero). *For any  $c \in (0, \infty)$ ,*

$$\mu_c(\{0\}) = q(c) + e^{-cq(c)} + cq(c)e^{-cq(c)} - 1, \quad (4.1.4)$$

where  $q(c)$  denotes the smallest point  $q \in (0, 1)$  satisfying the fixed-point equation

$$q = e^{-ce^{-cq}}. \quad (4.1.5)$$

A quick analysis of (4.1.5) – or an even quicker look at Figure 4.2 – reveals that the right-hand side of (4.1.4) undergoes a rupture of analyticity as  $c$  reaches the value  $e \approx 2.718$ . Bauer and Golinelli proposed an interpretation of this anomaly as a phase transition in the asymptotic structure of the kernel of  $A_n$ . Guided by numerical simulations, they further predicted the point  $c = e$  to be the threshold for the emergence of a continuous part at zero in the limiting measure  $\mu_c$  [22, 23]. To be more precise, we will say that a measure  $\mu$  has *no extended states* at a location  $E \in \mathbb{R}$  if

$$\frac{\mu([E - \varepsilon, E + \varepsilon]) - \mu(\{E\})}{\varepsilon} \xrightarrow{\varepsilon \rightarrow 0+} 0, \quad (4.1.6)$$

and has *extended states* at  $E$  otherwise. This terminology is borrowed from the theory of random Schrödinger operators (see, e.g., [4] for a recent treatment).

**CONJECTURE 3** (Emergence of extended states at zero). *The following phase transition occurs:*

1. *If  $c < e$ , then  $\mu_c$  has no extended states at 0.*
2. *If  $c > e$ , then  $\mu_c$  has extended states at 0.*

Conjecture 2 was established almost a decade ago [44] by a detailed *first-order* analysis of the random operator  $(A - z)^{-1}$  near the singular point  $z = 0$ , where  $A$  is the adjacency operator of the Poisson-Galton-Watson tree with mean offspring  $c$ . To the best of our knowledge however, Conjecture 3 – reiterated in [44] – had so far remained open. In the present work, we establish this long-predicted phase transition, illustrated on Figure 4.1. This is achieved by investigating the *second-order* behavior of the random operator  $(A - z)^{-1}$  near  $z = 0$ . As already mentioned, our result is not limited to the Erdős-Renyi model: we provide general, explicit criteria for the presence or absence of extended states at zero in the limiting spectral measure of any graph sequence whose local weak limit is a unimodular Galton-Watson tree, as defined next.

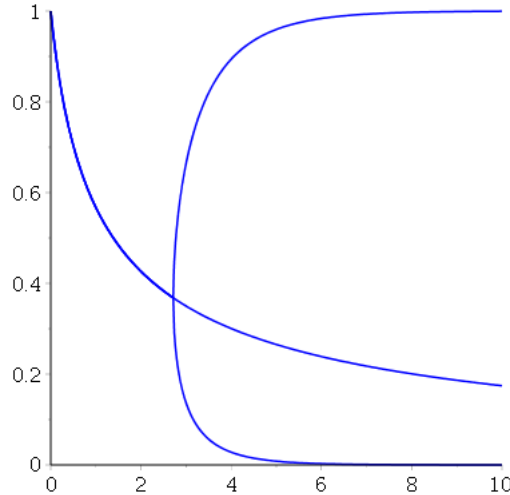


Figure 4.2 – The set of pairs  $(c, q)$  satisfying the equation (4.1.5). The branch point  $(e, e^{-1})$  causes a rupture of analyticity in the spectral mass at zero  $\mu_c(\{0\})$ , as  $c$  reaches  $e$ .

#### 4.1.2 General framework

The purpose of this section is to introduce our main objects of study, namely spectral measures of unimodular Galton-Watson trees. We only recall the necessary notions, and refer to the comprehensive survey [34] for more details on graph limits and their spectral theory.

**Spectral measures.** Let  $G = (V, E)$  be a countable, locally finite graph. Its adjacency operator  $A$  is a symmetric linear operator on the Hilbert space  $\ell^2_{\mathbb{C}}(V)$ . The domain of  $A$  consists of all finitely-supported vectors, and the action of  $A$  on the canonical basis  $(\mathbf{e}_x : x \in V)$  is given by

$$\langle \mathbf{e}_x | A \mathbf{e}_y \rangle = \begin{cases} 1 & \text{if } \{x, y\} \in E \\ 0 & \text{otherwise.} \end{cases} \quad (4.1.7)$$

As long as  $A$  is essentially self-adjoint, the Spectral Theorem applies: the resolvent  $(A - z)^{-1}$  is a well-defined bounded operator for all  $z \in \mathbb{C} \setminus \mathbb{R}$ , and for every  $\mathfrak{o} \in V$ , we have the representation

$$\forall z \in \mathbb{C} \setminus \mathbb{R}, \quad \langle \mathbf{e}_{\mathfrak{o}} | (A - z)^{-1} \mathbf{e}_{\mathfrak{o}} \rangle = \int_{\mathbb{R}} \frac{1}{\lambda - z} \mu_{(G, \mathfrak{o})}(\mathrm{d}\lambda), \quad (4.1.8)$$

for a unique probability measure  $\mu_{(G, \mathfrak{o})}$  on  $\mathbb{R}$ , called the spectral measure of the rooted graph  $(G, \mathfrak{o})$ . This fundamental object will be central to our work. It may be thought of as the local contribution of  $\mathfrak{o}$  to the spectrum of  $G$ . Indeed, when  $G$  is finite, there is an orthonormal basis of  $n = |V|$  eigenfunctions  $\phi_1, \dots, \phi_n$  of  $A$  with respective eigenvalues  $\lambda_1, \dots, \lambda_n$ , and we have the expression

$$\mu_{(G, \mathfrak{o})} = \sum_{k=1}^n |\phi_k(\mathfrak{o})|^2 \delta_{\lambda_k}. \quad (4.1.9)$$

In particular, the empirical eigenvalue distribution  $\mu_G := \frac{1}{n} \sum_{k=1}^n \delta_{\lambda_k}$  can be recovered from the spectral measures  $(\mu_{(G, \mathfrak{o})} : \mathfrak{o} \in V)$  by averaging over the choice of the root:

$$\mu_G = \frac{1}{|V|} \sum_{\mathfrak{o} \in V} \mu_{(G, \mathfrak{o})}. \quad (4.1.10)$$

Of course, neither side of this identity makes sense when  $G$  is infinite. However, the framework of *local weak convergence* enables us to pass to the “infinite-volume limit”, in an appropriate sense.

**Local weak convergence.** Write  $\mathcal{G}_*$  for the space of locally finite, connected rooted graphs, considered up to root-preserving isomorphism. Make it complete and separable by letting the distance between two rooted graphs be  $1/(1+r)$ , where  $r$  is the largest integer such that the balls of radius  $r$  around the root in the two graphs are isomorphic. Any finite graph naturally induces a probability measure on  $\mathcal{G}_*$ , via choosing a root uniformly at random and restricting to its connected component. If  $(G_n)_{n \geq 1}$  is a sequence of finite graphs, and if the sequence of probability measures thus induced admits a weak limit  $\mathcal{L}$ , then  $\mathcal{L}$  is called the *local weak limit* of  $(G_n)_{n \geq 1}$  [8]. In words,  $\mathcal{L}$  is the law of a random rooted graph  $(G, o)$  that describes what  $G_n$  asymptotically looks like when seen from a uniformly chosen vertex.

This limiting object has been shown to capture a number of asymptotic properties of  $(G_n)_{n \geq 1}$ , including the empirical eigenvalue distribution itself. More precisely, we always have

$$\sup_{\lambda \in \mathbb{R}} |\mu_{G_n}((-\infty, \lambda]) - \mu_{\mathcal{L}}((-\infty, \lambda])| \xrightarrow{n \rightarrow \infty} 0, \quad (4.1.11)$$

where  $\mu_{\mathcal{L}}(\cdot) := \mathbf{E}[\mu_{(G, o)}(\cdot)]$  denotes the expected spectral measure under  $\mathcal{L}$ <sup>1</sup>. This remarkable continuity principle has a long history [42, 1, 34]. In short, it allows one to replace the spectral analysis of sparse graphs by that of their local weak limits. Luckily, the latter turn out to be much more convenient to work with than the finite graphs that they approximate. For example, although they have many cycles, most sparse random graphs admit a local weak limit that is supported on trees. Moreover, in many cases of interest, including the Erdős-Renyi and configuration models, the limit has a particularly simple recursive structure, which we now describe.

**Unimodular Galton-Watson trees.** Let  $\pi = (\pi_k)_{k \geq 0}$  be a probability distribution on  $\mathbb{N}$  with finite, non-zero mean. A unimodular Galton-Watson tree with degree distribution  $\pi$  is a random rooted tree obtained by a Galton-Watson branching process in which the root has offspring distribution  $\pi$  and all descendants have the size-biased offspring distribution  $\hat{\pi} = (\hat{\pi}_k)_{k \geq 0}$  given by

$$\hat{\pi}_k := \frac{(k+1)\pi_{k+1}}{\sum_i i\pi_i}. \quad (4.1.12)$$

The law of this random rooted tree plays a distinguished role in the theory and will be denoted by  $\text{UGW}(\pi)$ . It arises as the local weak limit of uniform random graphs with prescribed degrees, when the number of vertices tends to infinity while the empirical degree distribution tends to  $\pi$ .

A simple example is random  $d$ -regular graphs, for which  $\pi$  is just a Dirac mass at  $d$ : the resulting tree is then the infinite  $d$ -regular rooted tree, whose spectral measure is the well-known Kesten-McKay distribution [113], see Figure 4.3. Another important example is the Erdős-Renyi model with parameters as in (4.1.2), for which  $\pi$  is the Poisson distribution with mean  $c$ . In that case, we have  $\hat{\pi} = \pi$ , so that  $\text{UGW}(\pi)$  is the law of the standard Poisson-Galton-Watson tree with mean offspring  $c$ . Its expected spectral measure  $\mu_{\text{UGW}(\pi)}$  is precisely the limit  $\mu_c$  mentioned in Section 4.1.1. The striking difference in the spectra of these two models (see Figures 4.1 and 4.3) motivates the following research program, to which the present paper is intended to contribute.

**PROBLEM 9.** *Understand the regularity of  $\mu_{\text{UGW}(\pi)}$  – in particular, the supports of its pure-point, absolutely-continuous, and singular-continuous parts – as a function of the degree distribution  $\pi$ .*

**State of the art.** This relatively young line of research has already witnessed notable progress. A comprehensive account, as well as a list of exciting conjectures, can be found in the introductory survey [34]. The pure-point part of the spectrum is now reasonably well understood. In particular, the work [44] provides an explicit formula for the mass at zero, while [127, 128] investigate the

<sup>1</sup>This definition implicitly relies on the (non-trivial) fact that the adjacency operator of a unimodular random graph is essentially self-adjoint with probability 1, see [34, Proposition 2.2] for a proof.

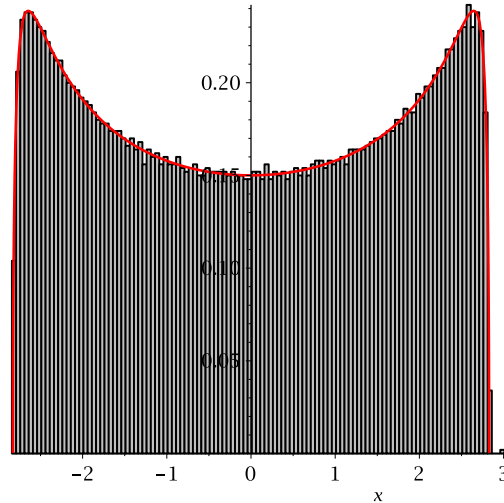


Figure 4.3 – Histogram of the eigenvalues of a uniform 3-regular random graph on  $10^4$  vertices (in gray), and the limiting Kesten McKay density (in red).

locations of other atoms. Rigorous results on the support of the continuous part are more limited. A remarkably general criterion by Bordenave, Sen and Virág [46] guarantees the presence of a non-trivial continuous part as soon as the Galton-Watson tree is super-critical. Unfortunately, the result is existential in nature and can not be used to ensure the presence of extended states at a given location. More precise information is available when  $\pi$  is sufficiently close to a Dirac mass, thanks to Keller [91]. However, the method used there is intrinsically perturbative and does not yield information for explicit choices of  $\pi$  such as the Poisson distribution involved in Conjecture 3. In the present paper, we provide explicit criteria for the presence or absence of extended states at zero in the spectral measure  $\mu_{\text{UGW}(\pi)}$ , for a general degree distribution  $\pi$ .

### 4.1.3 Results

Throughout this section, we fix a probability measure  $\pi$  on  $\mathbb{N}$  with finite, non-zero mean, and we let  $\mu = \mu_{\text{UGW}(\pi)}$  denote the expected spectral measure of the unimodular Galton-Watson tree with degree distribution  $\pi$ . In the degenerate case where  $\pi_0 + \pi_1 = 1$ , our random tree is just an isolated vertex with probability  $\pi_0$  and an isolated edge with probability  $\pi_1$ , so its expected spectral measure is  $\mu_{\text{UGW}(\pi)} = \pi_0 \delta_0 + \frac{\pi_1}{2} \delta_{-1} + \frac{\pi_1}{2} \delta_1$ , which trivially has no extended states anywhere. To avoid degeneracies, we will henceforth always assume that

$$\pi_0 + \pi_1 < 1. \quad (4.1.13)$$

All our results will be expressed in terms of the degree generating series

$$\varphi(z) := \sum_{k \geq 0} \pi_k z^k, \quad \widehat{\varphi}(z) := \sum_{k \geq 0} \widehat{\pi}_k z^k = \frac{\varphi'(z)}{\varphi'(1)}. \quad (4.1.14)$$

It was shown in [44, Theorem 2] that  $\mu(\{0\}) = \max M$ , where the function  $M: [0, 1] \rightarrow \mathbb{R}$  is given by

$$M(z) := \varphi(z) + (1-z)\varphi'(z) + \varphi(1 - \widehat{\varphi}(z)) - 1. \quad (4.1.15)$$

Our main finding is that the presence or absence of extended states at zero depends on the set

$$\operatorname{argmax} M := \{z \in [0, 1] : M(z) = \max M\}. \quad (4.1.16)$$

A quick differentiation shows that any  $z \in \operatorname{argmax} M$  must satisfy

$$z = 1 - \widehat{\varphi}(1 - \widehat{\varphi}(z)). \quad (4.1.17)$$

Among the (possibly many) solutions to this fixed-point equation, the following one will play a crucial role: we let  $z_\star \in (0, 1)$  denote the unique point satisfying

$$z_\star = 1 - \widehat{\varphi}(z_\star). \quad (4.1.18)$$

It is easily checked that  $M'(z_\star) = 0$ , and that  $M''(z_\star)$  has the same sign as  $\widehat{\varphi}'(z_\star) - 1$ . The presence or absence of extended states at zero in  $\mu$  turns out to be dictated by the following two conditions:

1.  $M$  achieves its maximum uniquely at  $z_\star$ , i.e.  $\operatorname{argmax} M = \{z_\star\}$ .
2.  $M''(z_\star) \neq 0$  (or equivalently,  $\widehat{\varphi}'(z_\star) \neq 1$ ).

More precisely, our first main result states that (i) and (ii) characterize a strong square-integrability property which, in particular, implies the absence of extended states at zero.

**Theorem 10** (No extended states at zero). *The square-integrability property*

$$\int_{\mathbb{R} \setminus \{0\}} \frac{1}{\lambda^2} \mu(d\lambda) < \infty \quad (4.1.19)$$

holds if and only if (i) and (ii) are both satisfied. In particular, when this is the case,  $\mu$  satisfies

$$\mu([- \varepsilon, \varepsilon]) = \mu(\{0\}) + o(\varepsilon^2), \quad (4.1.20)$$

as  $\varepsilon \rightarrow 0$ , which is much stronger than the absence of extended states at zero.

Conversely, our second main result guarantees the existence of extended states at zero as soon as (i) fails. Note that this only leaves aside the critical situation where (i) holds but (ii) fails, in which case we do not know whether the measure  $\mu$  has extended states at zero. We emphasize that this situation is not generic, as it forces  $z \mapsto \widehat{\varphi}(z) + z$  and  $z \mapsto \widehat{\varphi}'(z)$  to reach 1 at the same point.

**Theorem 11** (Extended states at zero). *If condition (i) fails, then  $\mu$  has extended states at zero.*

We end this section by applying our results to the special case where  $\pi$  is the Poisson distribution with mean  $c$ , i.e.  $\varphi(z) = \widehat{\varphi}(z) = e^{c(z-1)}$ . Under the change of variable  $q = 1 - z$ , the fixed-point equation (4.1.17) reduces to (4.1.5), whose solutions were represented on Figure 4.2. When  $c \leq e$ , the solution is unique, so condition (i) trivially holds. When  $c > e$ , there are three solutions, and condition (i) must fail because  $M''(z_\star) > 0$ . In fact, the double equality  $\widehat{\varphi}'(z_\star) = ce^{c(z_\star-1)} = c(1 - z_\star)$  shows that  $M''(z_\star)$  is negative, null or positive according to whether  $c$  is less than, equal to, or more than  $e$ . Thus, Theorem 10 applies if and only if  $c < e$ , and Theorem 11 applies if and only if  $c > e$ . This establishes Conjecture 3 and leaves aside the critical case  $c = e$ , which remains open.

## 4.2 Main ingredients

In this section, we introduce the main ingredients of our proof. We start by reformulating the problem of extended states at zero in terms of Stieltjes transforms, and then recall the well-known recursion satisfied by the latter on rooted trees. We then combine this recursion with the Mass Transport Principle to establish a new identity that will be crucial to our proof. We emphasize that all results in this section apply to general trees. The special structure of unimodular Galton-Watson trees will only enter the play in Section 4.3 below.

### 4.2.1 Stieltjes transform

The integral appearing in the definition (4.1.8) is known as the *Stieltjes transform* of the measure  $\mu_{(G,0)}$ . Here we will focus on the imaginary part of its restriction to the imaginary axis. More precisely, given a finite Borel measure  $\mu$  on  $\mathbb{R}$ , we consider the observable  $\mathfrak{s}: (0, \infty) \rightarrow \mathbb{R}$  defined by

$$\mathfrak{s}(t) := \int_{\mathbb{R}} \frac{t}{\lambda^2 + t^2} \mu(d\lambda). \quad (4.2.1)$$

The leading-order asymptotics of  $\mathfrak{s}(t)$  as  $t \rightarrow 0$  are directly related to the behavior of  $\mu$  around 0. In particular, the following two limits emerge naturally:

$$\alpha := \lim_{t \rightarrow 0} \downarrow t\mathfrak{s}(t) = \mu(\{0\}), \quad \beta := \lim_{t \rightarrow 0} \uparrow \frac{\mathfrak{s}(t)}{t} = \int_{\mathbb{R}} \frac{1}{\lambda^2} \mu(d\lambda). \quad (4.2.2)$$

Note that we can not simultaneously have  $\alpha > 0$  and  $\beta < \infty$ . To investigate the presence or absence of extended states at zero, we first need to subtract the atom at zero from  $\mu$ , i.e. consider the measure  $\mu^* := \mu - \alpha\delta_0$  and its associated transform,

$$\mathfrak{s}^*(t) = \mathfrak{s}(t) - \frac{\alpha}{t} = \int_{\mathbb{R} \setminus \{0\}} \frac{t}{\lambda^2 + t^2} \mu(d\lambda). \quad (4.2.3)$$

We then have the following exact characterization of the absence of extended states at zero.

**LEMMA 4.2.1** (Characterization).  *$\mu$  has no extended states at zero if and only if  $\mathfrak{s}^*(t) \rightarrow 0$  as  $t \rightarrow 0$ .*

*Proof.* Fix  $t > 0$  and set  $I_t := [-t, t] \setminus \{0\}$ . Since  $\frac{t}{\lambda^2 + t^2} \geq \frac{1}{2t}$  for all  $\lambda \in I_t$ , we have

$$\mathfrak{s}^*(t) \geq \frac{\mu(I_t)}{2t}.$$

Thus,  $\mu$  has no extended states at zero whenever  $\mathfrak{s}^*(t) \rightarrow 0$ . Conversely, observe that for any  $\varepsilon, t > 0$ , we have by Fubini's Theorem

$$\int_{I_\varepsilon} \frac{t}{\lambda^2 + t^2} \mu(d\lambda) = \int_0^\infty \frac{2tu}{(t^2 + u^2)^2} \mu(I_\varepsilon \wedge u) du.$$

On the other hand, the same identity with the measure  $\mu$  replaced by Lebesgue's measure gives

$$\arctan\left(\frac{\varepsilon}{t}\right) = \int_0^\infty \frac{2tu}{(t^2 + u^2)^2} (\varepsilon \wedge u) du.$$

Comparing these two lines, we deduce that

$$\int_{I_\varepsilon} \frac{t}{\lambda^2 + t^2} \mu(d\lambda) \leq \sup_{u \in (0, \varepsilon)} \left\{ \frac{\mu(I_u)}{u} \right\} \arctan\left(\frac{\varepsilon}{t}\right).$$

Since  $\arctan(\cdot) \leq \frac{\pi}{2}$ , and since  $\frac{t}{\lambda^2 + t^2} \leq \frac{t}{\varepsilon^2}$  for all  $\lambda \notin I_\varepsilon$ , we conclude that

$$\mathfrak{s}^*(t) \leq \frac{t}{\varepsilon^2} + \frac{\pi}{2} \sup_{u \in (0, \varepsilon)} \left\{ \frac{\mu(I_u)}{u} \right\}.$$

Sending  $t \rightarrow 0$  and then  $\varepsilon \rightarrow 0$  shows that  $\mathfrak{s}^*(t) \rightarrow 0$  whenever  $\mu$  has no extended states at zero.  $\square$

This lemma reduces the absence of extended states at zero to the condition  $\mathfrak{s}^*(t) = o(1)$  as  $t \rightarrow 0$ . Moreover, the square-integrability property (4.1.19) can be rephrased as  $\beta^* < \infty$ , where

$$\beta^* := \lim_{t \rightarrow 0} \uparrow \frac{\mathfrak{s}^*(t)}{t} = \int_{\mathbb{R} \setminus \{0\}} \frac{1}{\lambda^2} \mu(d\lambda). \quad (4.2.4)$$

Thus, our two main theorems will follow from a careful analysis of  $\mathfrak{s}^*(t)$  as  $t \rightarrow 0$ , when  $\mu$  is the expected spectral measure of a unimodular Galton-Watson tree. The starting point of this analysis is a well-known local recursion satisfied by spectral measures of rooted trees.



### 4.2.2 Local recursion

As many graph-theoretical quantities, spectral measures admit a recursive structure when evaluated on trees. Fix a tree  $T = (V, E)$  whose adjacency operator is self-adjoint, and let  $o \in V$  be an arbitrary vertex. We write  $\partial o = \{x \in V : \{x, o\} \in E\}$  for the set of its neighbours, and  $\deg(o) = |\partial o|$  for its degree. Deleting  $o$  splits  $T$  into  $\deg(o)$  disjoint subtrees which will naturally be denoted by  $(T_{x \rightarrow o} : x \in \partial o)$ . We let  $\mathfrak{s}_o, \mathfrak{s}_o^*, \alpha_o, \beta_o, \beta_o^*$  be the objects  $\mathfrak{s}, \mathfrak{s}^*, \alpha, \beta, \beta^*$  defined above when the general measure  $\mu$  is taken to be the spectral measure  $\mu_{(T, o)}$ . Similarly, we let  $\mathfrak{s}_{x \rightarrow o}, \mathfrak{s}_{x \rightarrow o}^*, \alpha_{x \rightarrow o}, \beta_{x \rightarrow o}, \beta_{x \rightarrow o}^*$  correspond to the choice  $\mu = \mu_{(T_{x \rightarrow o}, x)}$ . We then have the following elementary but fundamental relation (see, e.g., [44] for a proof): for all  $t \in (0, \infty)$ ,

$$\mathfrak{s}_o(t) = \frac{1}{t + \sum_{x \in \partial o} \mathfrak{s}_{x \rightarrow o}(t)}. \quad (4.2.5)$$

In particular, multiplying or dividing both sides by  $t$  and sending  $t \rightarrow 0$  yields

$$\alpha_o = \frac{1}{1 + \sum_{x \in \partial o} \beta_{x \rightarrow o}}; \quad (4.2.6)$$

$$\beta_o = \frac{1}{\sum_{x \in \partial o} \alpha_{x \rightarrow o}}. \quad (4.2.7)$$

In view of these identities, it is natural to decompose the degree as  $\deg(o) = \mathcal{N}_o^+ + \mathcal{N}_o^- + \mathcal{N}_o^*$  where

$$\mathcal{N}_o^+ := \sum_{x \in \partial o} \mathbf{1}_{(\alpha_{x \rightarrow o} > 0)} \quad (4.2.8)$$

$$\mathcal{N}_o^- := \sum_{x \in \partial o} \mathbf{1}_{(\beta_{x \rightarrow o} < \infty)} \quad (4.2.9)$$

$$\mathcal{N}_o^* := \sum_{x \in \partial o} \mathbf{1}_{(\alpha_{x \rightarrow o} = 0, \beta_{x \rightarrow o} = \infty)}. \quad (4.2.10)$$

It then readily follows from (4.2.6) and (4.2.7) that

$$\alpha_o > 0 \iff \mathcal{N}_o^+ = \mathcal{N}_o^* = 0 \quad (4.2.11)$$

$$\beta_o < \infty \iff \mathcal{N}_o^+ \geq 1 \quad (4.2.12)$$

$$(\alpha_o = 0, \beta_o = \infty) \iff (\mathcal{N}_o^+ = 0, \mathcal{N}_o^* \geq 1). \quad (4.2.13)$$

Of course, the recursion (4.2.5) also applies to the tree  $T_{o \rightarrow y}$  (for any  $y \in \partial o$ ), yielding

$$\mathfrak{s}_{o \rightarrow y}(t) = \frac{1}{t + \sum_{x \in \partial o \setminus \{y\}} \mathfrak{s}_{x \rightarrow o}(t)}; \quad (4.2.14)$$

$$\alpha_{o \rightarrow y} = \frac{1}{1 + \sum_{x \in \partial o \setminus \{y\}} \beta_{x \rightarrow o}}; \quad (4.2.15)$$

$$\beta_{o \rightarrow y} = \frac{1}{\sum_{x \in \partial o \setminus \{y\}} \alpha_{x \rightarrow o}}. \quad (4.2.16)$$

These recursions will play a crucial role in our analysis.

### 4.2.3 Mass Transport Principle

The second-order quantity  $\beta_o^*$  is *a priori* much harder to analyze than its first-order counterpart  $\beta_o$ , as we have to remove the singularity caused by the atom at zero. To overcome this difficulty, we will exploit a powerful identity known as the *Mass Transport Principle* (see, e.g., [7]): any random rooted

graph  $(G, \mathfrak{o})$  whose law is the local weak limit of some sequence of finite graphs is *unimodular*, in the sense that it satisfies the distributional symmetry

$$\mathbf{E} \left[ \sum_{x \in V(G)} f(G, \mathfrak{o}, x) \right] = \mathbf{E} \left[ \sum_{x \in V(G)} f(G, x, \mathfrak{o}) \right], \quad (4.2.17)$$

for any Borel-measurable function  $f: \mathcal{G}_{**} \rightarrow [0, \infty]$ , where  $\mathcal{G}_{**}$  denotes the natural analogue of  $\mathcal{G}_*$  for *doubly-rooted* graphs. At an intuitive level, this identity expresses the fact that the root is “equally likely” to be any vertex (even though the underlying graph is possibly infinite). Here we use this spatial stationarity to prove the following key formula, which expresses  $\mathbf{E}[\beta_{\mathfrak{o}}^*]$  in terms of  $\beta_{\mathfrak{o}}$  only.

**PROPOSITION 4.2.2** (Getting rid of the atom at zero). *For any unimodular random tree  $(T, \mathfrak{o})$ ,*

$$\mathbf{E}[\beta_{\mathfrak{o}}^*] = \mathbf{E}[\mathbf{1}_{(\alpha_{\mathfrak{o}}=0)}\beta_{\mathfrak{o}}] + \mathbf{E}[\mathbf{1}_{(\mathcal{N}_{\mathfrak{o}}^+ \geq 2)}\beta_{\mathfrak{o}}] + \mathbf{E} \left[ \mathbf{1}_{(\mathcal{N}_{\mathfrak{o}}^+ \geq 2)} \frac{\sum_{x \in \partial \mathfrak{o}} \beta_{x \rightarrow \mathfrak{o}} \mathbf{1}_{(\alpha_{x \rightarrow \mathfrak{o}}=0)}}{\sum_{x \in \partial \mathfrak{o}} \alpha_{x \rightarrow \mathfrak{o}} \mathbf{1}_{(\alpha_{x \rightarrow \mathfrak{o}} > 0)}} \right]. \quad (4.2.18)$$

*Proof.* Fix  $t \in (0, \infty)$  and  $y \in \partial \mathfrak{o}$ . Combining (4.2.5) and (4.2.14), we have

$$\mathfrak{s}_{\mathfrak{o}}(t) = \left( \frac{1}{\mathfrak{s}_{\mathfrak{o} \rightarrow y}(t)} + \mathfrak{s}_{y \rightarrow \mathfrak{o}}(t) \right)^{-1}. \quad (4.2.19)$$

Multiplying by  $\mathfrak{s}_{y \rightarrow \mathfrak{o}}(t)$  clearly makes the right-hand side symmetric in  $\mathfrak{o}$  and  $y$ , and hence

$$\mathfrak{s}_{\mathfrak{o}}(t) \mathfrak{s}_{y \rightarrow \mathfrak{o}}(t) = \mathfrak{s}_y(t) \mathfrak{s}_{\mathfrak{o} \rightarrow y}(t). \quad (4.2.20)$$

Summing over all  $y \in \partial \mathfrak{o}$  and using again (4.2.5), we obtain

$$1 - t \mathfrak{s}_{\mathfrak{o}}(t) = \sum_{y \in \partial \mathfrak{o}} \mathfrak{s}_y(t) \mathfrak{s}_{\mathfrak{o} \rightarrow y}(t). \quad (4.2.21)$$

On the other hand, it easily follows from (4.2.6), (4.2.15) and (4.2.16) that for any  $y \in \partial \mathfrak{o}$ , we have

$$\alpha_{\mathfrak{o}} > 0 \iff (\alpha_{\mathfrak{o} \rightarrow y} > 0 \text{ and } \mathcal{N}_y^+ \geq 2). \quad (4.2.22)$$

Combining this with (4.2.21), we deduce that

$$(1 - t \mathfrak{s}_{\mathfrak{o}}(t)) \mathbf{1}_{(\alpha_{\mathfrak{o}} > 0)} = \sum_{y \in \partial \mathfrak{o}} \mathfrak{s}_y(t) \mathfrak{s}_{\mathfrak{o} \rightarrow y}(t) \mathbf{1}_{(\alpha_{\mathfrak{o} \rightarrow y} > 0)} \mathbf{1}_{(\mathcal{N}_y^+ \geq 2)}. \quad (4.2.23)$$

We may now take expectation and use unimodularity to obtain

$$\mathbf{E}[(1 - t \mathfrak{s}_{\mathfrak{o}}(t)) \mathbf{1}_{(\alpha_{\mathfrak{o}} > 0)}] = \mathbf{E} \left[ \mathbf{1}_{(\mathcal{N}_{\mathfrak{o}}^+ \geq 2)} \mathfrak{s}_{\mathfrak{o}}(t) \sum_{y \in \partial \mathfrak{o}} \mathfrak{s}_{y \rightarrow \mathfrak{o}}(t) \mathbf{1}_{(\alpha_{y \rightarrow \mathfrak{o}} > 0)} \right]. \quad (4.2.24)$$

Letting  $t \rightarrow 0$  and using  $\{\mathcal{N}_{\mathfrak{o}}^+ \geq 2\} \subseteq \{\beta_{\mathfrak{o}} < \infty\}$  for the right-hand side, we obtain

$$\begin{aligned} \mathbf{E}[(1 - \alpha_{\mathfrak{o}}) \mathbf{1}_{(\alpha_{\mathfrak{o}} > 0)}] &= \mathbf{E} \left[ \mathbf{1}_{(\mathcal{N}_{\mathfrak{o}}^+ \geq 2)} \beta_{\mathfrak{o}} \sum_{y \in \partial \mathfrak{o}} \alpha_{y \rightarrow \mathfrak{o}} \right] \\ &= \mathbf{P}(\mathcal{N}_{\mathfrak{o}}^+ \geq 2) \\ &= \mathbf{E} \left[ \mathbf{1}_{(\mathcal{N}_{\mathfrak{o}}^+ \geq 2)} \mathfrak{s}_{\mathfrak{o}}(0) \left( t + \sum_{y \in \partial \mathfrak{o}} \mathfrak{s}_{y \rightarrow \mathfrak{o}}(t) \right) \right], \end{aligned}$$

where the second line follows from (4.2.7) and the third from (4.2.5). Subtracting (4.2.24), we arrive at

$$\mathbf{E} \left[ (t\mathfrak{s}_o(t) - \alpha_o) \mathbf{1}_{(\alpha_o > 0)} \right] = \mathbf{E} \left[ \mathbf{1}_{(\mathcal{N}_o^+ \geq 2)} \mathfrak{s}_o(t) \left( t + \sum_{y \in \partial o} \mathfrak{s}_{y \rightarrow o}(t) \mathbf{1}_{(\alpha_{y \rightarrow o} = 0)} \right) \right]. \quad (4.2.25)$$

Dividing through by  $t^2$  and sending  $t \rightarrow 0$  yields

$$\mathbf{E}[\beta_o^* \mathbf{1}_{(\alpha_o > 0)}] = \mathbf{E} \left[ \mathbf{1}_{(\mathcal{N}_o^+ \geq 2)} \beta_o \left( 1 + \sum_{x \in \partial o} \beta_{x \rightarrow o} \mathbf{1}_{(\alpha_{x \rightarrow o} = 0)} \right) \right]. \quad (4.2.26)$$

On the other hand, on the event  $\{\alpha_o = 0\}$ , we have  $\beta_o^* = \beta_o$ , which concludes the proof.  $\square$

### 4.3 Unimodular Galton-Watson trees

The above results were valid for any unimodular random tree. We now consider the special case of unimodular Galton-Watson trees, and exploit their self-similar nature to turn the above recursions into distributional fixed-point equations that will be amenable to analysis.

#### 4.3.1 Distributional fixed-point equations

From now on, we fix a degree distribution  $\pi$  as in Section 4.1.3, and we equip the space of rooted trees  $(T, o)$  with two different probability measures: we reserve the letter  $\mathbf{P}$  for the unimodular law  $\text{UGW}(\pi)$ , and use  $\widehat{\mathbf{P}}$  to denote the homogeneous Galton-Watson law with offspring distribution  $\widehat{\pi}$ . We naturally use  $\mathbf{E}$  and  $\widehat{\mathbf{E}}$  to denote the corresponding expectations. Thus, the distribution of the root-degree  $\text{deg}(o)$  is  $\pi$  under  $\mathbf{P}$  and  $\widehat{\pi}$  under  $\widehat{\mathbf{P}}$  and in both cases, conditionally on  $\text{deg}(o)$ , the subtrees  $(T_{x \rightarrow o}, x \in \partial o)$  are i.i.d. homogeneous Galton-Watson trees with offspring distribution  $\widehat{\pi}$ . In particular, the recursion (4.2.5) takes the following simple distributional form.

**COROLLARY 4.3.1** (Distributional structure). *Under both  $\mathbf{P}$  and  $\widehat{\mathbf{P}}$ , the conditional distribution of  $(\mathcal{N}_o^+, \mathcal{N}_o^-, \mathcal{N}_o^*)$  given  $\text{deg}(o)$  is Multinomial with size parameter  $\text{deg}(o)$  and probability parameters  $(\widehat{\mathbf{P}}(\alpha_o > 0), \widehat{\mathbf{P}}(\beta_o < \infty), \widehat{\mathbf{P}}(\alpha_o = 0, \beta_o = \infty))$ . Moreover, conditionally on  $(\mathcal{N}_o^+, \mathcal{N}_o^-, \mathcal{N}_o^*)$ , the random sums*

$$\begin{aligned} & \sum_{x \in \partial o} \mathfrak{s}_{x \rightarrow o}(t) \mathbf{1}_{(\alpha_{x \rightarrow o} > 0)} \\ & \sum_{x \in \partial o} \mathfrak{s}_{x \rightarrow o}(t) \mathbf{1}_{(\alpha_{x \rightarrow o} < \infty)} \\ & \sum_{x \in \partial o} \mathfrak{s}_{x \rightarrow o}(t) \mathbf{1}_{(\alpha_{x \rightarrow o} = 0, \beta_{x \rightarrow o} = \infty)} \end{aligned}$$

are independent, the first (resp. second, resp. third) being distributed as a sum of  $\mathcal{N}_o^+$  (resp.  $\mathcal{N}_o^-$ , resp.  $\mathcal{N}_o^*$ ) i.i.d. random variables with law  $\widehat{\mathbf{P}}(\mathfrak{s}_o(t) \in \cdot | \alpha_o > 0)$  (resp.  $\widehat{\mathbf{P}}(\mathfrak{s}_o(t) \in \cdot | \beta_o < \infty)$ , resp.  $\widehat{\mathbf{P}}(\mathfrak{s}_o(t) \in \cdot | \alpha_o = 0, \beta_o = \infty)$ ).

We shall use this fact (and its  $t \rightarrow 0$  counterparts) repeatedly below, without notice. For example, an immediate consequence of this and (4.2.11)-(4.2.12) is that

$$\mathbf{P}(\alpha_o > 0) = \varphi \left( \widehat{\mathbf{P}}(\beta_o < \infty) \right), \quad \mathbf{P}(\beta_o = \infty) = \varphi \left( \widehat{\mathbf{P}}(\alpha_o = 0) \right), \quad (4.3.1)$$

$$\widehat{\mathbf{P}}(\alpha_o > 0) = \widehat{\varphi} \left( \widehat{\mathbf{P}}(\beta_o < \infty) \right), \quad \widehat{\mathbf{P}}(\beta_o = \infty) = \widehat{\varphi} \left( \widehat{\mathbf{P}}(\alpha_o = 0) \right), \quad (4.3.2)$$

where we recall that  $\varphi$  and  $\hat{\varphi}$  are the generating series of  $\pi$  and  $\hat{\pi}$  respectively. In particular, the number  $z = \hat{\mathbf{P}}(\alpha_0 = 0)$  must solve the fixed-point equation (4.1.17). In fact,  $\hat{\mathbf{P}}(\alpha_0 = 0)$  was shown in [44] to coincide with the last point at which the function  $M$  achieves its maximum, i.e.

$$\hat{\mathbf{P}}(\alpha_0 = 0) = \max(\operatorname{argmax} M). \quad (4.3.3)$$

With this characterization in hands, we may reformulate our main assumption (i) as follows.

**LEMMA 4.3.2** (Reformulation of assumption (i)). *The following conditions are equivalent.*

1.  $\operatorname{argmax} M = \{z_\star\}$ ;
2.  $\hat{\mathbf{P}}(\alpha_0 = 0) = z_\star$ ;
3.  $\mathbf{P}(\alpha_0 = 0, \beta_0 = \infty) = 0$ ;
4.  $\hat{\mathbf{P}}(\alpha_0 = 0, \beta_0 = \infty) = 0$ .

*Proof.* Set  $z := \hat{\mathbf{P}}(\alpha_0 = 0)$ . Since  $\{\alpha_0 > 0\} \subseteq \{\beta_0 = \infty\}$ , we always have

$$\begin{aligned} \mathbf{P}(\alpha_0 = 0, \beta_0 = \infty) &= \mathbf{P}(\beta_0 = \infty) - \mathbf{P}(\alpha_0 > 0) \\ &= \varphi(z) - \varphi(1 - \hat{\varphi}(z)). \end{aligned}$$

where the second line follows from (4.3.1). Similarly,

$$\hat{\mathbf{P}}(\alpha_0 = 0, \beta_0 = \infty) = \hat{\varphi}(z) - \hat{\varphi}(1 - \hat{\varphi}(z)).$$

From these equalities and the fact that  $\varphi, \hat{\varphi}$  are increasing, we immediately deduce that the conditions (2),(3) and (4) are equivalent. Moreover, it is clear from (4.3.3) that (1) implies (2). To see that (2) implies (1), recall that any point  $z \in \operatorname{argmax} M$  must satisfy the fixed-point equation (4.1.17), and that the latter implies  $M(1 - \hat{\varphi}(z)) = M(z)$ . Thus, the set  $\operatorname{argmax} M$  is stable under the map  $z \mapsto 1 - \hat{\varphi}(z)$ , and so it can not intersect  $(z_\star, 1]$  without also intersecting  $[0, z_\star)$ .  $\square$

### 4.3.2 Proof of Theorem 10

In this section, we prove Theorem 10, namely, that (i) and (ii) are necessary and sufficient for  $\mathbf{E}[\beta_0^\star] < \infty$ . The necessity of (i) is easy: if (i) fails, then  $\mathbf{P}(\alpha_0 = 0, \beta_0 = \infty) > 0$  by Lemma 4.3.2, and so the first term on the right-hand side of (4.2.18) is already infinite. We will thus henceforth assume that (i) holds. By Lemma 4.3.2, this ensures that

$$\hat{\mathbf{P}}(\alpha_0 = 0) = \hat{\mathbf{P}}(\beta_0 < \infty) = z_\star. \quad (4.3.4)$$

Let us note here for future use that, in view of Corollary 4.3.1, we now have

$$\mathbf{P}(\mathcal{N}_0^+ = 1) = \sum_{n \geq 1} n \pi_n z_\star^{n-1} (1 - z_\star) = (1 - z_\star) \varphi'(z_\star) = (1 - z_\star)^2 \varphi'(1) \quad (4.3.5)$$

$$\hat{\mathbf{P}}(\mathcal{N}_0^+ = 1) = \sum_{n \geq 1} n \hat{\pi}_n z_\star^{n-1} (1 - z_\star) = (1 - z_\star) \hat{\varphi}'(z_\star) \quad (4.3.6)$$

$$\hat{\mathbf{E}}[\mathcal{N}_0^- | \mathcal{N}_0^+ = \mathcal{N}_0^\star = 0] = \frac{1}{1 - z_\star} \sum_{n=0}^{\infty} n \hat{\pi}_n z_\star^n = \frac{z_\star \hat{\varphi}'(z_\star)}{1 - z_\star}. \quad (4.3.7)$$

Our first task consists in reducing the finiteness of  $\mathbf{E}[\beta_0^\star]$  to that of  $\hat{\mathbf{E}}\left[\frac{1}{\alpha_0} | \alpha_0 > 0\right]$ .

**LEMMA 4.3.3** (Reduction). *Under (i), we have  $\mathbf{E}[\beta_0^\star] < \infty$  if and only if  $\hat{\mathbf{E}}\left[\frac{1}{\alpha_0} | \alpha_0 > 0\right] < \infty$ .*

*Proof of the “only if” part.* Since  $\{\mathcal{N}_o^+ = 1\} \subseteq \{\alpha_o = 0\} \subseteq \{\beta_o^* = \beta_o\}$ , we have

$$\begin{aligned} \mathbf{E}[\beta_o^*] &\geq \mathbf{E}\left[\beta_o \mathbf{1}_{(\mathcal{N}_o^+ = 1)}\right] = \mathbf{E}\left[\frac{1}{\sum_{x \in \partial o} \alpha_{x \rightarrow o}} \mathbf{1}_{(\mathcal{N}_o^+ = 1)}\right] \\ &= \mathbf{E}\left[\frac{1}{\sum_{x \in \partial o} \alpha_{x \rightarrow o} \mathbf{1}_{(\alpha_{x \rightarrow o} > 0)}} \mathbf{1}_{(\mathcal{N}_o^+ = 1)}\right] \\ &= \mathbf{P}(\mathcal{N}_o^+ = 1) \widehat{\mathbf{E}}\left[\frac{1}{\alpha_o} \mid \alpha_o > 0\right]. \end{aligned}$$

This is enough to conclude, since  $\mathbf{P}(\mathcal{N}_o^+ = 1) > 0$ , by (4.3.5).  $\square$

*Proof of the “if” part.* Let us now assume that  $\widehat{\mathbf{E}}\left[\frac{1}{\alpha_o} \mid \alpha_o > 0\right] < \infty$ , and verify that each term on the right-hand side of Formula (4.2.18) is finite. For the first term, we write

$$\begin{aligned} \mathbf{E}\left[\mathbf{1}_{(\alpha_o = 0)} \beta_o\right] &= \mathbf{E}\left[\mathbf{1}_{(\beta_o < \infty)} \beta_o\right] \\ &= \mathbf{E}\left[\mathbf{1}_{(\mathcal{N}_o^+ \geq 1)} \frac{1}{\sum_{x \in \partial o} \alpha_{x \rightarrow o}}\right] \end{aligned}$$

where the first line follows from Lemma 4.3.2, and the second from (4.2.7) and (4.2.12). Now, conditionally on  $\mathcal{N}_o^+$ , the random variable  $\sum_{x \in \partial o} \alpha_{x \rightarrow o} = \sum_{x \in \partial o} \alpha_{x \rightarrow o} \mathbf{1}_{(\alpha_{x \rightarrow o} > 0)}$  is distributed as the sum of  $\mathcal{N}_o^+$  i.i.d. random variables with law  $\widehat{\mathbf{P}}(\alpha_o \in \cdot \mid \alpha_o > 0)$ . Keeping only one of them yields

$$\mathbf{E}\left[\mathbf{1}_{(\alpha_o = 0)} \beta_o\right] \leq \mathbf{P}(\mathcal{N}_o^+ \geq 1) \widehat{\mathbf{E}}\left[\frac{1}{\alpha_o} \mid \alpha_o > 0\right], \quad (4.3.8)$$

which is finite. The second term is less than the first because  $\{\mathcal{N}_o^+ \geq 2\} \subseteq \{\alpha_o = 0\}$ . For the third one, we observe the following: conditionally on  $\mathcal{N}_o^+$  and  $\mathcal{N}_o^- + \mathcal{N}_o^*$ , the two random variables

$$\sum_{x \in \partial o} \alpha_{x \rightarrow o} \mathbf{1}_{(\alpha_{x \rightarrow o} > 0)} \quad \text{and} \quad \sum_{x \in \partial o} \beta_{x \rightarrow o} \mathbf{1}_{(\alpha_{x \rightarrow o} = 0)},$$

are independent, the first being distributed as a sum of  $\mathcal{N}_o^+$  i.i.d. random variables with distribution  $\widehat{\mathbf{P}}(\alpha_o \in \cdot \mid \alpha_o > 0)$ , and the second as a sum of  $\mathcal{N}_o^* + \mathcal{N}_o^-$  i.i.d. random variables with law  $\widehat{\mathbf{P}}(\beta_o \in \cdot \mid \alpha_o = 0)$ . Keeping only one of the  $\mathcal{N}_o^+$  i.i.d. random variables in the first sum, we obtain

$$\begin{aligned} \mathbf{E}\left[\mathbf{1}_{(\mathcal{N}_o^+ \geq 2)} \frac{\sum_{x \in \partial o} \beta_{x \rightarrow o} \mathbf{1}_{(\alpha_{x \rightarrow o} = 0)}}{\sum_{x \in \partial o} \alpha_{x \rightarrow o} \mathbf{1}_{(\alpha_{x \rightarrow o} > 0)}}\right] &\leq \\ &\widehat{\mathbf{E}}\left[\frac{1}{\alpha_o} \mid \alpha_o > 0\right] \mathbf{E}\left[\mathbf{1}_{(\mathcal{N}_o^+ \geq 2)} (\mathcal{N}_o^- + \mathcal{N}_o^*)\right] \widehat{\mathbf{E}}[\beta_o \mid \alpha_o = 0]. \end{aligned}$$

The product on the right-hand side consists of three terms. The first is finite by assumption. The second is less than the expected degree at the root of our unimodular Galton-Watson tree, which is also finite. Finally, the inequality (4.3.8) with  $\mathbf{P}$  replaced by  $\widehat{\mathbf{P}}$  shows that the third term is finite.  $\square$

Since our running assumption (i) forces  $M''(z_*) \leq 0$ , the condition (ii) becomes  $M''(z_*) < 0$  or equivalently,  $\widehat{\varphi}'(z_*) < 1$ . To complete the proof of Theorem 10, it therefore only remains to show that  $\widehat{\mathbf{E}}\left[\frac{1}{\alpha_o} \mid \alpha_o > 0\right] < \infty$  if and only if  $\widehat{\varphi}'(z_*) < 1$ , which we now do.

**LEMMA 4.3.4.** *Under assumption (i),  $\widehat{\mathbf{E}}\left[\frac{1}{\alpha_o} \mid \alpha_o > 0\right] < \infty$  if and only if  $\widehat{\varphi}'(z_*) < 1$ .*

*Proof of the “only if” part.* On the one hand, using (4.2.6) and (4.2.11), we have

$$\begin{aligned}
 \widehat{\mathbf{E}} \left[ \frac{1}{\alpha_o} \mid \alpha_o > 0 \right] &= 1 + \widehat{\mathbf{E}} \left[ \sum_{x \in \partial o} \beta_{x \rightarrow o} \mid \alpha_o > 0 \right] \\
 &= 1 + \widehat{\mathbf{E}} \left[ \sum_{x \in \partial o} \beta_{x \rightarrow o} \mathbf{1}_{(\beta_{x \rightarrow o} < \infty)} \mid \mathcal{N}_o^+ = \mathcal{N}_o^* = 0 \right] \\
 &= 1 + \widehat{\mathbf{E}} [\mathcal{N}_o^- \mid \mathcal{N}_o^+ = \mathcal{N}_o^* = 0] \widehat{\mathbf{E}} [\beta_o \mid \beta_o < \infty] \\
 &= 1 + \frac{\widehat{\varphi}'(z_*)}{1 - z_*} \widehat{\mathbf{E}} [\beta_o \mathbf{1}_{(\beta_o < \infty)}],
 \end{aligned}$$

where the last line uses (4.3.7) and (4.3.4). On the other hand, using (4.2.7), we have

$$\begin{aligned}
 \widehat{\mathbf{E}} [\beta_o \mathbf{1}_{(\mathcal{N}_o^+ = 1)}] &= \widehat{\mathbf{E}} \left[ \frac{1}{\sum_{x \in \partial o} \alpha_{x \rightarrow o}} \mathbf{1}_{(\mathcal{N}_o^+ = 1)} \right] \\
 &= \widehat{\mathbf{E}} \left[ \frac{1}{\sum_{x \in \partial o} \alpha_{x \rightarrow o} \mathbf{1}_{(\alpha_{x \rightarrow o} > 0)}} \mathbf{1}_{(\mathcal{N}_o^+ = 1)} \right] \\
 &= \widehat{\mathbf{P}} (\mathcal{N}_o^+ = 1) \widehat{\mathbf{E}} \left[ \frac{1}{\alpha_o} \mid \alpha_o > 0 \right] \\
 &= (1 - z_*) \widehat{\varphi}'(z_*) \widehat{\mathbf{E}} \left[ \frac{1}{\alpha_o} \mid \alpha_o > 0 \right],
 \end{aligned}$$

where the third line uses Corollary 4.3.1 and the last line uses (4.3.6). Since  $\{\mathcal{N}_o^+ = 1\} \subseteq \{\beta_o < \infty\}$ , we deduce from these two facts that

$$\widehat{\mathbf{E}} \left[ \frac{1}{\alpha_o} \mid \alpha_o > 0 \right] \geq 1 + (\widehat{\varphi}'(z_*))^2 \widehat{\mathbf{E}} \left[ \frac{1}{\alpha_o} \mid \alpha_o > 0 \right].$$

The desired conclusion now clearly follows.  $\square$

*Proof of the “if” part.* Fix  $t > 0$ , and observe that by (4.2.5) and (4.2.11),

$$\widehat{\mathbf{E}} \left[ \frac{1}{t \mathfrak{s}_o(t)} \mid \alpha_o > 0 \right] = \widehat{\mathbf{E}} \left[ 1 + \sum_{x \in \partial o} \frac{\mathfrak{s}_{x \rightarrow o}(t)}{t} \mid \alpha_o > 0 \right] \tag{4.3.9}$$

$$= 1 + \widehat{\mathbf{E}} \left[ \sum_{x \in \partial o} \frac{\mathfrak{s}_{x \rightarrow o}(t)}{t} \mathbf{1}_{(\beta_{x \rightarrow o} < \infty)} \mid \mathcal{N}_o^+ = \mathcal{N}_o^* = 0 \right] \tag{4.3.10}$$

$$= 1 + \widehat{\mathbf{E}} [\mathcal{N}_o^- \mid \mathcal{N}_o^+ = \mathcal{N}_o^* = 0] \widehat{\mathbf{E}} \left[ \frac{\mathfrak{s}_o(t)}{t} \mid \beta_o < \infty \right] \tag{4.3.11}$$

$$= 1 + \frac{z_* \widehat{\varphi}'(z_*)}{1 - z_*} \widehat{\mathbf{E}} \left[ \frac{\mathfrak{s}_o(t)}{t} \mid \beta_o < \infty \right] \tag{4.3.12}$$

where the last line uses (4.3.7). On the other hand, using (4.2.5) and (4.2.12), we have

$$\begin{aligned}
 \widehat{\mathbf{E}} \left[ \frac{\mathfrak{s}_o(t)}{t} \mid \beta_o < \infty \right] &= \widehat{\mathbf{E}} \left[ \left( \sum_{x \in \partial o} t \mathfrak{s}_{x \rightarrow o}(t) \right)^{-1} \mid \mathcal{N}_o^+ \geq 1 \right] \\
 &\leq \widehat{\mathbf{E}} \left[ \left( \sum_{x \in \partial o} t \mathfrak{s}_{x \rightarrow o}(t) \mathbf{1}_{(\alpha_{x \rightarrow o} > 0)} \right)^{-1} \mid \mathcal{N}_o^+ \geq 1 \right].
 \end{aligned} \tag{4.3.13}$$

Conditionally on  $\mathcal{N}_0^+$ , the integrand on the right-hand side is distributed as the reciprocal of the sum of  $\mathcal{N}_0^+$  i.i.d. random variables with law  $\widehat{\mathbf{P}}(t\mathfrak{s}_0(t) \in \cdot | \alpha_0 > 0)$ . To exploit this i.i.d. structure, we transform the reciprocal  $(\cdot)^{-1}$  into a power via the trivial identity

$$r^{-1} = \int_0^1 z^{r-1} dz,$$

valid for any  $r > 0$ . With  $r = \sum_{x \in \partial_0} t\mathfrak{s}_{x \rightarrow 0}(t) \mathbf{1}_{(\alpha_{x \rightarrow 0} > 0)}$ , we obtain

$$\begin{aligned} \widehat{\mathbf{E}} \left[ \left( \sum_{x \in \partial_0} t\mathfrak{s}_{x \rightarrow 0}(t) \mathbf{1}_{(\alpha_{x \rightarrow 0} > 0)} \right)^{-1} \middle| \mathcal{N}_0^+ \geq 1 \right] &= \int_0^1 \widehat{\mathbf{E}} \left[ z^{\sum_{x \in \partial_0} t\mathfrak{s}_{x \rightarrow 0}(t) \mathbf{1}_{(\alpha_{x \rightarrow 0} > 0)}} \middle| \mathcal{N}_0^+ \geq 1 \right] \frac{dz}{z} \\ &= \sum_{n=1}^{\infty} \widehat{\mathbf{P}}(\mathcal{N}_0^+ = n | \mathcal{N}_0^+ \geq 1) \int_0^1 \widehat{\mathbf{E}} \left[ z^{t\mathfrak{s}_0(t)} \middle| \alpha_0 > 0 \right]^n \frac{dz}{z}. \end{aligned}$$

We now fix some  $\varepsilon \in (0, 1)$  and  $n \geq 1$ , and estimate the integral on the right-hand side by splitting it into two parts: for  $z \in (\varepsilon, 1)$ , we use the crude bound  $\widehat{\mathbf{E}} \left[ z^{t\mathfrak{s}_0(t)} \middle| \alpha_0 > 0 \right] \leq 1$  to obtain

$$\int_{\varepsilon}^1 \left( \widehat{\mathbf{E}} \left[ z^{t\mathfrak{s}_0(t)} \middle| \alpha_0 > 0 \right] \right)^n \frac{dz}{z} \leq \ln \left( \frac{1}{\varepsilon} \right).$$

For  $z \in (0, \varepsilon)$ , we use the observation that  $z^{t\mathfrak{s}_0(t)} \leq \varepsilon^{\alpha_0}$  to write

$$\begin{aligned} \int_0^{\varepsilon} \left( \widehat{\mathbf{E}} \left[ z^{t\mathfrak{s}_0(t)} \middle| \alpha_0 > 0 \right] \right)^n \frac{dz}{z} &\leq \widehat{\mathbf{E}}[\varepsilon^{\alpha_0} | \alpha_0 > 0]^{n-1} \int_0^{\varepsilon} \widehat{\mathbf{E}} \left[ z^{t\mathfrak{s}_0(t)} \middle| \alpha_0 > 0 \right] \frac{dz}{z} \\ &\leq \widehat{\mathbf{E}}[\varepsilon^{\alpha_0} | \alpha_0 > 0]^{n-1} \widehat{\mathbf{E}} \left[ \frac{1}{t\mathfrak{s}_0(t)} \middle| \alpha_0 > 0 \right]. \end{aligned}$$

Inserting these estimates into the above series and recalling (4.3.9)-(4.3.13), we arrive at

$$\widehat{\mathbf{E}} \left[ \frac{1}{t\mathfrak{s}_0(t)} \middle| \alpha_0 > 0 \right] \leq 1 + \frac{z_{\star} \widehat{\varphi}'(z_{\star})}{1 - z_{\star}} \left( \log \left( \frac{1}{\varepsilon} \right) + \Phi \left( \widehat{\mathbf{E}}[\varepsilon^{\alpha_0} | \alpha_0 > 0] \right) \widehat{\mathbf{E}} \left[ \frac{1}{t\mathfrak{s}_0(t)} \middle| \alpha_0 > 0 \right] \right), \quad (4.3.14)$$

where we have introduced the short-hand

$$\Phi(u) = \widehat{\mathbf{E}} \left[ u^{\mathcal{N}_0^+ - 1} \middle| \mathcal{N}_0^+ \geq 1 \right] = \sum_{n=0}^{\infty} \widehat{\mathbf{P}}(\mathcal{N}_0^+ = n + 1 | \mathcal{N}_0^+ \geq 1) u^n.$$

Now, observe that

$$\Phi(0) = \frac{\widehat{\mathbf{P}}(\mathcal{N}_0^+ = 1)}{\widehat{\mathbf{P}}(\mathcal{N}_0^+ \geq 1)} = \frac{1 - z_{\star}}{z_{\star}} \widehat{\varphi}'(z_{\star}),$$

where we have used (4.2.12), (4.3.4) and (4.3.6). By continuity of  $\Phi$ , we deduce that

$$\frac{z_{\star} \widehat{\varphi}'(z_{\star})}{1 - z_{\star}} \Phi \left( \widehat{\mathbf{E}}[\varepsilon^{\alpha_0} | \alpha_0 > 0] \right) \xrightarrow{\varepsilon \rightarrow 0} (\widehat{\varphi}'(z_{\star}))^2.$$

If  $\widehat{\varphi}'(z_{\star}) < 1$ , we can choose  $\varepsilon > 0$  so that  $\frac{z_{\star} \widehat{\varphi}'(z_{\star})}{1 - z_{\star}} \Phi \left( \widehat{\mathbf{E}}[\varepsilon^{\alpha_0} | \alpha_0 > 0] \right) < 1$  and rewrite (4.3.14) as

$$\widehat{\mathbf{E}} \left[ \frac{1}{t\mathfrak{s}_0(t)} \middle| \alpha_0 > 0 \right] \leq \frac{1 + \frac{z_{\star} \widehat{\varphi}'(z_{\star})}{1 - z_{\star}} \log \left( \frac{1}{\varepsilon} \right)}{1 - \frac{z_{\star} \widehat{\varphi}'(z_{\star})}{1 - z_{\star}} \Phi \left( \widehat{\mathbf{E}}[\varepsilon^{\alpha_0} | \alpha_0 > 0] \right)}.$$

The right-hand side is finite and independent of  $t$ , so letting  $t \rightarrow 0$  concludes the proof.  $\square$

### 4.3.3 Proof of Theorem 11

We use the following Lemma, whose proof is trivial once we observe that a sequence of non-negative random variables  $(X_n)_{n \geq 1}$  tends to  $\infty$  in probability if and only if  $\mathbf{E}[\exp(-X_n)] \xrightarrow[n \rightarrow \infty]{} 0$ .

**LEMMA 4.3.5.** *Let  $k \geq 1$  be a fixed integer, and let  $(X_n^{(1)})_{n \geq 1}, \dots, (X_n^{(k)})_{n \geq 1}$  be  $k$  i.i.d. copies of an arbitrary sequence  $(X_n)_{n \geq 1}$  of non-negative random variables. Then the sequence  $(Y_n)_{n \geq 1}$  defined by  $Y_n := X_n^{(1)} + \dots + X_n^{(k)}$  tends to  $\infty$  in probability if and only if  $(X_n)_{n \geq 1}$  does.*

*Proof of Theorem 11.* Assume that condition (i) fails. By Lemma 4.3.2, this means that the event

$$\mathcal{E} := \{\alpha_0 = 0, \beta_0 = \infty\} = \{\mathcal{N}_0^+ = 0, \mathcal{N}_0^* \geq 1\}$$

has positive probability under  $\mathbf{P}$  and  $\widehat{\mathbf{P}}$ . On this event, the recursion (4.2.5) can be rewritten as

$$\mathfrak{s}_0^*(t) = \left( t + \sum_{x \in \partial_0} \mathfrak{s}_{x \rightarrow 0}(t) \mathbf{1}_{(\beta_{x \rightarrow 0} < \infty)} + \sum_{x \in \partial_0} \mathfrak{s}_{x \rightarrow 0}^*(t) \mathbf{1}_{(\alpha_{x \rightarrow 0} = 0, \beta_{x \rightarrow 0} = \infty)} \right)^{-1}.$$

The first sum on the right-hand side tends to 0 as  $t \rightarrow 0$  by definition of  $\beta_{x \rightarrow 0}$ . On the other hand, by Corollary 4.3.1, conditionally on  $\mathcal{N}_0^*$ , the second sum is distributed as the sum of  $\mathcal{N}_0^*$  i.i.d. variables with law  $\widehat{\mathbf{P}}(\mathfrak{s}_0^* \in \cdot | \mathcal{E})$ . We emphasize that this statement is valid under both  $\mathbf{P}$  and  $\widehat{\mathbf{P}}$  (only the distribution of  $\mathcal{N}_0^*$  differ). Applying Lemma 4.3.5 to both situations, we deduce that along any deterministic sequence  $(t_n)_{n \geq 1}$  of positive numbers with  $t_n \rightarrow 0$  as  $n \rightarrow \infty$ , the following conditions are equivalent:

- (a)  $\mathfrak{s}_0^*(t_n) \xrightarrow[n \rightarrow \infty]{} 0$  in probability under  $\mathbf{P}(\cdot | \mathcal{E})$ ;
- (b)  $\mathfrak{s}_0^*(t_n) \xrightarrow[n \rightarrow \infty]{} \infty$  in probability under  $\widehat{\mathbf{P}}(\cdot | \mathcal{E})$ ;
- (c)  $\mathfrak{s}_0^*(t_n) \xrightarrow[n \rightarrow \infty]{} 0$  in probability under  $\widehat{\mathbf{P}}(\cdot | \mathcal{E})$ .

Of course, (b) and (c) are incompatible, and so (a) can never hold. In particular, this rules out the possibility that  $\mathbf{E}[\mathfrak{s}_0^*(t_n)] \rightarrow 0$  as  $n \rightarrow \infty$ , and since  $(t_n)_{n \geq 1}$  is arbitrary, we conclude that

$$\liminf_{t \rightarrow 0^+} \mathbf{E}[\mathfrak{s}_0^*(t)] > 0. \tag{4.3.15}$$

By Lemma 4.2.1, this is more than enough to ensure that  $\mu_{\text{UGW}(\pi)}$  has extended states at zero.  $\square$

## 4.4 The skeleton tree

We end this chapter by an extension of the preceding result to another infinite unimodular tree, namely the skeleton tree. This semi-infinite tree naturally arises as the Benjamini-Schramm limit of uniform rooted trees ([85]). It is defined as follows: let  $(T_n)$  be an iid sequence of random rooted trees, with distribution  $\text{PGW}(1)$ , and let  $P$  be the ‘semi-infinite spine’, whose vertex set is  $\mathbb{N} = \{0, 1, 2, \dots\}$  and where the edges are  $(x, x + 1)$ . The Skeleton tree  $T_{\text{squel}}$  is obtained by gluing the root of the tree  $T_n$  on the  $n$ -th vertex of  $P$ . The obtained tree  $T_{\text{squel}}$  is then rooted at the first vertex of the spine.



#### 4.4. The skeleton tree

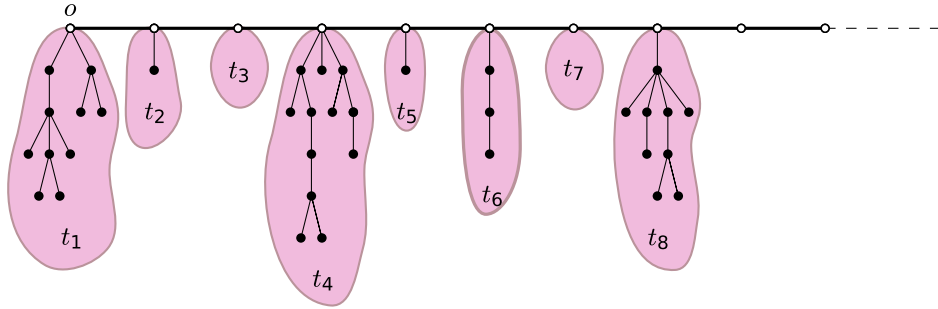


Figure 4.4 – Illustration for one construction of the skeleton tree  $T_{\text{squel}}$ . The trees  $(t_i)$  are realization of an iid sequence PGW(1) trees.

This tree is unimodular as a limite of finite connected graphs, hence its adjacency operator is as self-adjoint, and we can define its spectral measure at the root  $\mu_{\text{squel}}$  through its Stieltjes transform (4.1.8).

Let  $G_n$  be a uniform tree over the  $n^{n-2}$  trees on  $n$  vertices. As shown in [85], this sequence convergences to  $T_{\text{squel}}$  in the Benjamini-Schramm topology, hence the spectral continuity (4.1.11) directly applies; if  $\mu_{G_n}$  is the spectral measure of the adjacency matrix of  $G_n$ , then

$$\mu_{G_n} \xrightarrow{d_{\text{KS}}} \mathbf{E}[\mu_{\text{squel}}] \quad (4.4.1)$$

where  $d_{\text{KS}}$  denotes the Kolmogorov-Smirnoff distance on finite measures, as in (4.1.11). Numerical simulations seem to show the absence of extended states in the origin of this measure, as in Figure 4.5.

Using the same tools as for unimodular Galton-Watson trees, we prove this observation.

**Theorem 12.** *The expected spectral measure  $\mathbf{E}\mu_{\text{squel}}$  has no extended states at zero.*

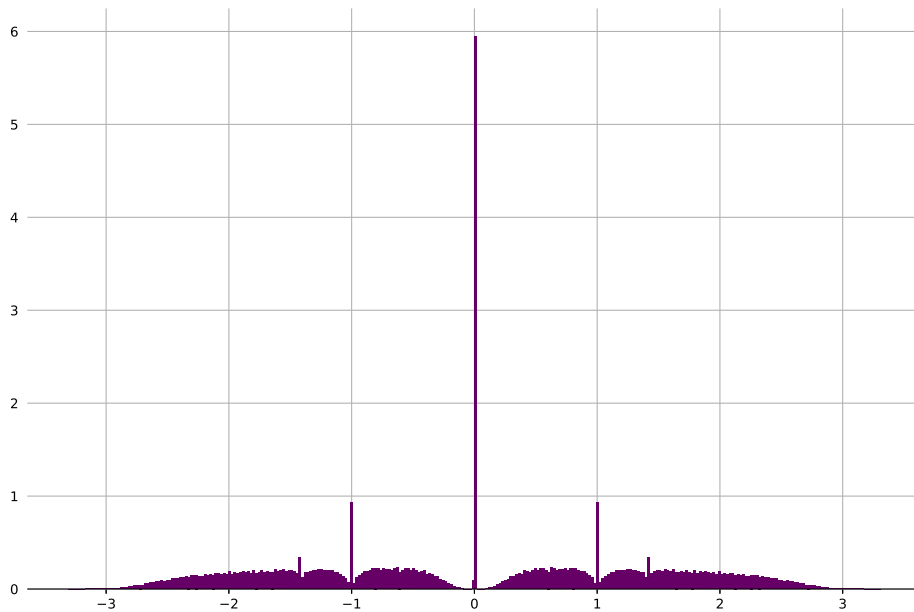


Figure 4.5 – Histogram of the eigenvalues of large uniform random trees on  $n = 1000$  vertices.

Our proof strategy uses the Stieltjes transform and especially the characterization lemma 4.2.1. Let us recall some notations. If  $(T, o)$  is a rooted essentially self-adjoint tree, with spectral measure at the root denoted by  $\mu_{(T,o)}$ , we consider the observable

$$\mathfrak{s}(t) = \int_{\mathbb{R}} \frac{t}{\lambda^2 + t^2} \mu_{(T,o)}(d\lambda)$$

and the two corresponding quantities

$$\alpha := \lim_{t \rightarrow 0} \downarrow t \mathfrak{s}(t) = \mu_{(T,o)}(\{0\}), \quad \beta := \lim_{t \rightarrow 0} \uparrow \frac{\mathfrak{s}(t)}{t} = \int_{\mathbb{R}} \frac{1}{\lambda^2} \mu_{(T,o)}(d\lambda) \quad (4.4.2)$$

and finally, we set

$$\mathfrak{s}^*(t) = \mathfrak{s}(t) - \frac{\alpha}{t}.$$

Following Lemma 4.2.1, our goal is to prove that when  $(T, o)$  is distributed as  $T_{\text{squel}}$ , then

$$\lim_{t \rightarrow 0} \mathbf{E}[\mathfrak{s}^*(t)] = 0 \quad (4.4.3)$$

thus showing that  $\mathbf{E}[\mu_{\text{squel}}]$  has no extended states at the origin.

#### 4.4.1 Proof of Theorem 11

**Tree notations and cavity equations.** Let us introduce some notations. The vertices of the spine of  $T_{\text{squel}}$  are noted with integers; the root is 0. If  $n$  is one of those spine vertices, we note  $V_n$  the set of its children *which are not on the spine*; they correspond to the children of the root in  $T_n$ .

If  $n$  is an integer, we will denote by  $(T_{\text{squel}})_n$  the tree which is formally defined as  $(T_{\text{squel}})_{n \rightarrow n-1}$ ; it is the tree obtained from  $T_{\text{squel}}$  by deleting the vertices  $0, 1, \dots, n-1$  and their attached trees  $T_0, \dots, T_{n-1}$ , and setting the root at  $n$ . It is clear that  $(T_{\text{squel}})_n$  has the same distribution as  $T_{\text{squel}}$ . Quantities such as  $\alpha_n$  or  $\beta_{0 \rightarrow n}$ , etc., are defined similarly as in Section 4.2.2.

The cavity equations applied at the root of the skeleton tree now take the form:

$$\alpha_0 = \frac{1}{1 + \sum_{x \in V_0} \beta_{x \rightarrow 0} + \beta_1} \quad (4.4.4)$$

and

$$\beta_0 = \frac{1}{\sum_{x \in V_0} \alpha_{x \rightarrow 0} + \alpha_1} \quad (4.4.5)$$

**The alternative holds true on the skeleton tree.** We will now note  $p = \mathbf{P}_{\text{PGW}(1)}(\alpha > 0)$  and  $q = \mathbf{P}_{\text{PGW}(1)}(\beta < \infty)$ . As noted in the end of Section 4.1.3, on a Poisson-GW tree with parameter  $c = 1 < e$ , condition (i) trivially holds, and from the equivalences listed in Lemma 4.3.2, this is equivalent to  $\mathbf{P}_{\text{PGW}(1)}(\alpha = 0, \beta = \infty) = 0$ , or more simply  $p + q = 1$ ; in this paragraph, we show that the same alternative holds for  $T_{\text{squel}}$ , namely that  $\mathbf{P}(\alpha_0 > 0) + \mathbf{P}(\beta_0 < \infty) = 1$ .

From the preceding equations we have  $\{\alpha_0 > 0\} = \{\beta_1 < \infty\} \cap \{\forall x \in V_0, \beta_{x \rightarrow 0} < \infty\}$ , hence by independence,

$$\mathbf{P}(\alpha_0 > 0) = \mathbf{P}(\beta_0 < \infty)p. \quad (4.4.6)$$

On the other hand,  $\{\beta_0 = \infty\} = \{\alpha_1 = 0\} \cap \{\forall x \in V_0 : \alpha_{x \rightarrow 0} = 0\}$ , hence by the same argument

$$\mathbf{P}(\beta_0 = \infty) = \mathbf{P}(\alpha_1 = 0)p \quad (4.4.7)$$

and so we get

$$\mathbf{P}(\beta_0 < \infty) = 1 - \mathbf{P}(\alpha_1 = 0)p = q + p\mathbf{P}(\alpha_0 > 0). \quad (4.4.8)$$

and this finally gives  $\mathbf{P}(\alpha_0 > 0) = pq/(1 - p^2)$ , and  $\mathbf{P}(\beta_0 < \infty) = q/(1 - p^2)$ . Adding both, we get

$$\mathbf{P}(\alpha_0 > 0) + \mathbf{P}(\beta_0 < \infty) = \frac{pq}{1 - p^2} + \frac{q}{1 - p^2} = \frac{pq + q}{1 - p^2} = 1 \quad (4.4.9)$$

because  $pq + q = p(1 - p) + 1 - p = p - p^2 + 1 - p = 1 - p^2$ .

**No extended states: the proof.** We are now going to show that  $\mathbf{E}[\beta_0^*] < \infty$ . The Skeleton tree is a unimodular random graph, hence it satisfies the identity (4.2.18) from Proposition 4.2.2, namely

$$\mathbf{E}[\beta_0^*] = \mathbf{E}[\beta_0 \mathbf{1}_{\alpha_0=0}] + \mathbf{E}[\beta_0 \mathbf{1}_{\mathcal{N}_0^+ \geq 2}] + \mathbf{E} \left[ \mathbf{1}_{\mathcal{N}_0^+ \geq 2} \frac{\sum_{x \in V_0} \beta_{x \rightarrow 0} \mathbf{1}_{\alpha_{x \rightarrow 0}=0} + \beta_1 \mathbf{1}_{\alpha_1=0}}{\sum_{x \in V_0} \alpha_{x \rightarrow 0} \mathbf{1}_{\alpha_{x \rightarrow 0} > 0} + \alpha_1 \mathbf{1}_{\alpha_1 > 0}} \right] \quad (4.4.10)$$

where  $\mathcal{N}_0^+ = \mathbf{1}_{\alpha_1 > 0} + \sum_{x \in V_0} \mathbf{1}_{\alpha_{x \rightarrow 0} > 0}$  is the number of children of the root having  $\alpha > 0$ ; similarly, we define  $\mathcal{N}_0^-$  as the number of children of the root having  $\beta < \infty$ .

In (4.4.10), there are three terms. As  $\{\mathcal{N}_0^+ \geq 2\} \subset \{\alpha = 0\}$ , we have  $\mathbf{E}[\beta_0 \mathbf{1}_{\alpha_0=0}] \geq \mathbf{E}[\beta_0 \mathbf{1}_{\mathcal{N}_0^+ \geq 2}]$ , hence we only have to prove that  $\mathbf{E}[\beta_0 \mathbf{1}_{\alpha_0=0}] < \infty$  (*first claim*) and  $\mathbf{I} < \infty$ , where  $\mathbf{I}$  is the last expectation in (4.4.10) (*second claim*).

*Proof of the first claim.* We have  $\{\beta_0 < \infty\} = \{\alpha_1 > 0\} \cup \{\beta_{0 \rightarrow 1} < \infty\}$ , hence we have  $\mathbf{1}_{\beta_0 < \infty} \leq \mathbf{1}_{\alpha_1 > 0} + \mathbf{1}_{\beta_{0 \rightarrow 1} < \infty}$ , and

$$\begin{aligned} \mathbf{E}[\beta_0 \mathbf{1}_{\beta_0 < \infty}] &= \mathbf{E} \left[ \frac{1}{\alpha_1 + \sum_{x \in V_0} \alpha_{x \rightarrow 0}} \mathbf{1}_{\beta_0 < \infty} \right] \\ &\leq \mathbf{E} \left[ \frac{\mathbf{1}_{\alpha_1 > 0}}{\alpha_1 + \sum_{x \in V_0} \alpha_{x \rightarrow 0}} \right] + \mathbf{E} \left[ \frac{\mathbf{1}_{\beta_{0 \rightarrow 1} < \infty}}{\alpha_1 + \sum_{x \in V_0} \alpha_{x \rightarrow 0}} \right] \\ &\leq \mathbf{E} \left[ \frac{\mathbf{1}_{\alpha_1 > 0}}{\alpha_1} \right] + \mathbf{E} \left[ \frac{\mathbf{1}_{\beta_{0 \rightarrow 1} < \infty}}{\sum_{x \in V_0} \alpha_{x \rightarrow 0}} \right] \\ &\leq \mathbf{E} \left[ \frac{\mathbf{1}_{\alpha_0 > 0}}{\alpha_0} \right] + \mathbf{E} [\mathbf{1}_{\beta_{0 \rightarrow 1} < \infty} \beta_{0 \rightarrow 1}] \end{aligned}$$

The second term is precisely the quantity  $\beta$  applied to the dangling tree  $T_0$ , which is distributed as a PGW(1) tree, and it has been already proven that  $\mathbf{E}_{\text{PGW}(1)}[\mathbf{1}_{\beta < \infty} \beta]$  is finite; more precisely, the combination of (4.3.8), assumptions (i)-(ii) (which are true for PGW(1)) and Lemma 4.3.4 clearly shows that this expectation is finite.

We now repeat the argument for  $\alpha_0$ :

$$\begin{aligned} \mathbf{E} \left[ \frac{\mathbf{1}_{\alpha_0 > 0}}{\alpha_0} \right] &= \mathbf{E} \left[ \mathbf{1}_{\alpha_0 > 0} \left( 1 + \beta_1 + \sum_{x \in V_0} \beta_{x \rightarrow 0} \right) \right] \\ &= \mathbf{E}[\beta_1 \mathbf{1}_{\alpha_0 > 0}] + \mathbf{E} \left[ \frac{1}{\alpha_{0 \rightarrow 1}} \mathbf{1}_{\alpha_0 > 0} \right] \\ &\leq \mathbf{E}[\beta_1 \mathbf{1}_{\beta_1 < \infty}] \mathbf{P}(\alpha_{0 \rightarrow 1} > 0) + \mathbf{E} \left[ \frac{\mathbf{1}_{\alpha_{0 \rightarrow 1} > 0}}{\alpha_{0 \rightarrow 1}} \right] \end{aligned}$$

where in the last line, we used  $\{\alpha_0 > 0\} = \{\beta_1 < \infty\} \cap \{\alpha_{0 \rightarrow 1} > 0\}$  and independence of these two events. Again, Lemma 4.3.4 yields that  $\mathbf{E} \left[ \frac{\mathbf{1}_{\alpha_{0 \rightarrow 1} > 0}}{\alpha_{0 \rightarrow 1}} \right]$  is finite. Gathering the preceding inequalities, we get

$$\mathbf{E}[\beta_0 \mathbf{1}_{\beta_0 < \infty}] \leq p \mathbf{E}[\beta_0 \mathbf{1}_{\beta_0 < \infty}] + c \quad (4.4.11)$$

where  $p = \mathbf{P}_{\text{PGW}(1)}(\alpha_0 > 0)$  had been defined earlier and  $c$  is the real constant

$$c := \mathbf{E}_{\text{PGW}(1)}[\beta_0 \mathbf{1}_{\beta_0 < \infty}] + \mathbf{E}_{\text{PGW}(1)} \left[ \frac{\mathbf{1}_{\alpha_0 > 0}}{\alpha_0} \right].$$

As  $p > 0$ , this automatically ensures that  $\mathbf{E}[\beta_0 \mathbf{1}_{\beta_0 < \infty}] \leq \frac{c}{1-p} < \infty$  as needed. Note that, in the course of the proof, the equation

$$\mathbf{E} \left[ \frac{\mathbf{1}_{\alpha_0 > 0}}{\alpha_0} \right] < \infty \quad (4.4.12)$$

has been shown to be true.  $\square$

*Proof of the second claim.* We now treat the last term in (4.4.10), dubbed I, and prove its finiteness. We can split the event  $\{\mathcal{N}_0^+ \geq 2\}$  in two disjoint events  $\{\alpha_1 > 0, \widetilde{\mathcal{N}}_0^+ \geq 1\} \cup \{\alpha_1 = 0, \widetilde{\mathcal{N}}_0^+ \geq 2\}$ , where

$$\widetilde{\mathcal{N}}_0^+ = \sum_{x \in V_0} \mathbf{1}_{\alpha_{x \rightarrow 0} > 0}.$$

Splitting I on these two events yields

$$\mathbf{I} \leq \mathbf{E} \left[ \mathbf{1}_{\widetilde{\mathcal{N}}_0^+ \geq 1, \alpha_1 > 0} \frac{\sum_{x \in V_0} \beta_{x \rightarrow 0} \mathbf{1}_{\alpha_{x \rightarrow 0} = 0}}{\alpha_1 \mathbf{1}_{\alpha_1 > 0}} \right] + \mathbf{E} \left[ \mathbf{1}_{\widetilde{\mathcal{N}}_0^+ \geq 2, \alpha_1 = 0} \frac{\sum_{x \in V_0} \beta_{x \rightarrow 0} \mathbf{1}_{\alpha_{x \rightarrow 0} = 0} + \beta_1 \mathbf{1}_{\alpha_1 = 0}}{\sum_{x \in V_0} \alpha_{x \rightarrow 0} \mathbf{1}_{\alpha_{x \rightarrow 0} > 0}} \right]. \quad (4.4.13)$$

By independence, the first term is equal to

$$\mathbf{E} \left[ \frac{\mathbf{1}_{\alpha_1 > 0}}{\alpha_1} \right] \mathbf{E} \left[ \mathbf{1}_{\widetilde{\mathcal{N}}_0^+ \geq 1} \sum_{x \in V_0} \beta_{x \rightarrow 0} \mathbf{1}_{\alpha_{x \rightarrow 0} = 0} \right]$$

The first term has already been proven finite in (4.4.12), and the second one is smaller than

$$\mathbf{E} \left[ \sum_{x \in V_0} \beta_{x \rightarrow 0} \mathbf{1}_{\alpha_{x \rightarrow 0} = 0} \right] = \sum_{n=0}^{\infty} n \pi_n \mathbf{E}_{\text{PGW}(1)}[\beta \mathbf{1}_{\beta < \infty}] = \mathbf{E}_{\text{PGW}(1)}[\beta \mathbf{1}_{\beta < \infty}] < \infty.$$

We finally treat the second expectation in (4.4.13). Let us note  $S_\beta$  the sum on the denominator and  $S_\alpha$  the sum on the numerator. The term is equal to

$$\mathbf{P}(\alpha_1 = 0) \mathbf{E} \left[ \mathbf{1}_{\widetilde{\mathcal{N}}_0^+ \geq 2} \frac{S_\beta}{S_\alpha} \right] + \mathbf{E} \left[ \mathbf{1}_{\widetilde{\mathcal{N}}_0^+ \geq 2} \frac{1}{S_\alpha} \right] \mathbf{E}[\beta_1 \mathbf{1}_{\alpha_1 = 0}]$$

and every term in this expression is finite by the preceding results.  $\square$

To conclude, every term in the RHS of (4.4.10) is finite, hence  $\mathbf{E}[\beta_0^*]$  is itself finite, which directly proves by Lemma 4.2.1 that  $\mathbf{E}[\mu_{\text{sequel}}]$  has no extended states at the origin. In fact, it also proves that  $\beta_0^*$  is finite, hence on an event with probability one, the random measure  $\mu_{\text{sequel}}$  has no extended states at the origin — a much stronger result than Theorem 12.



## Chapter 5

# Matrix reconstruction from sparse observations

*This chapter is based on a joint work with Charles Bordenave and Raj Rao Nadakuditi; the prepublication is in preparation at the time of submission of this dissertation .*

### 5.1 Introduction and statement of the results.

#### 5.1.1 Setting.

Let  $n \geq 1$  be an integer and let  $P = (P_{i,j})_{i,j \in [n]} \in \mathcal{M}_n(\mathbb{R})$  be a real hermitian matrix with spectral decomposition

$$P = \sum_{k=1}^r \mu_k \varphi_k \varphi_k^* \quad (5.1.1)$$

where  $r$  is the rank of  $P$ , the  $\mu_k$  are the eigenvalues of  $P$ , and  $\varphi_1, \dots, \varphi_r$  is an orthonormal basis of eigenvectors. By simplicity, we will consider matrices with *positive, distinct eigenvalues* in decreasing order:  $\mu_1 > \dots > \mu_r > 0$ .

Each one of the  $n^2$  entries of the matrix  $P$  is observed with equal and independent probability  $d/n$ . The observation is thus a matrix  $A$  defined as

$$A_{x,y} = \begin{cases} \frac{n}{d} P_{x,y} & \text{if entry } (x,y) \text{ is observed} \\ 0 & \text{else.} \end{cases}$$

The normalization is chosen so that  $\mathbf{E}[A] = P$ . The main point of this paper is that  $A$  is *not hermitian*. The eigenvalues are complex numbers, ordered by decreasing magnitude  $|\lambda_1| \geq \dots \geq |\lambda_n|$ . Even though it seems tempting to symmetrize the matrix  $A$  because  $P$  is hermitian, our point is that this might not be the best thing to do, and that indeed  $A$  captures more information about  $P$  than its symmetrized version.

This paper answers the following question:

*What part of  $P$  can be recovered from the observation of  $A$  ?*

The literature around this problem — the so-called matrix completion problem — is gigantic, see Section 5.2. The general philosophy is that under natural assumptions on  $P$  (low-rank with no extravagant structure), we can recover  $P$  with arbitrary precision as soon as  $d$  has order  $\log n$ . This paper gives a much more detailed and complete description of the phenomenon, by describing the exact relation between  $d$  and the parts of  $P$  which can be recovered from  $A$ . More precisely, given any

fixed  $d$ , there is a threshold  $\vartheta$ , intrinsic to the matrix  $P$  and to  $d$ , such that all the eigenvalues above this threshold can be recovered, while the others cannot, being “lost in the bulk” of uninformative eigenvalues of  $A$ . The eigenvectors can also be recovered (weakly).

Those results are linked with the non-symmetry of the problem and are part of a new and promising philosophy, namely that *in many problems eigenvalues of non-symmetric matrices can perform better than eigenvalues of symmetric matrices*.

We hereby state this result without caring too much on hypothesis and definitions. We note  $L = n \max_{x,y} |P_{x,y}|$  and we define a matrix  $Q$  by  $Q_{x,y} = nP_{x,y}^2$ ; we note  $\rho$  its operator norm.

**Theorem 13.** *Let  $P$  be a matrix with rank  $r = O(\ln(n))$  and which is sufficiently incoherent. We note*

$$\vartheta = \sqrt{\rho/d} \quad \text{and} \quad \vartheta_0 = L/d. \quad (5.1.2)$$

Let  $r_0$  be the number of eigenvalues of  $P$  which are greater than

$$\max\{\vartheta, \vartheta_0\}.$$

Then, with high probability when  $n$  is large, the top  $r_0$  eigenvalues of  $A$  are asymptotically equal to the  $\mu_k$ , while all the other eigenvalues of  $A$  are asymptotically smaller than  $\max\{\vartheta, \vartheta_0\}$ .

By ‘incoherent’, we mean that the eigenvectors are delocalized, ie with entries  $O(\sqrt{n})$ , a universal assumption in the litterature. This notion is going to be precisely described in the next section, and a very detailed description of this result will be the object of Theorem 15.

In general,  $\vartheta$  is bigger than  $\vartheta_0$  but there is a regime where  $d$  is very small and as a consequence, the real threshold is  $\vartheta_0$ . This is the same phenomenon as what discovered in [62] — see the definition of  $\tilde{\rho}$  in Theorem 1.

### 5.1.2 Weak recovery is doable for $d$ fixed

One of our most striking result states that *weak recovery is feasible even on the diluted case where  $d$  is fixed*. This is in very sharp contrast with what would happen if the revealed entries were symmetric. As known in the litterature, the top eigenvalues would then be aligned with the high-degree vertices, but also the top eigenvectors would be localized on those vertices, losing all the signal information.

**Theorem 14** (phase transition for rank-one matrices). *Suppose that  $P = \mu\varphi\varphi^*$  with  $\varphi$  a delocalized unit vector:*

1. *If  $d < n|\varphi|_4^4$ , then all the eigenvalues of  $A$  have modulus smaller than  $\mu\sqrt{n|\varphi|_4^4/d}$ .*
2. *If  $d > n|\varphi|_4^4$ , then  $\lambda_1(A) \rightarrow \mu$  and all the other eigenvalues of  $A$  have modulus smaller than  $\mu\sqrt{\frac{n|\varphi|_4^4}{d}} + o(1)$ . Moreover, if  $\psi$  is the normalized eigenvector of  $A$  associated with  $\lambda_1$ , then*

$$|\langle \psi, \varphi \rangle| \sim \sqrt{1 - \frac{n|\varphi|_4^4}{d}}. \quad (5.1.3)$$

Here again, “sufficiently delocalized” will be defined later, but the reader can replace it by  $|\varphi|_\infty = O(1)$  or  $O(\ln(n))$ .

The proof is done in Subsection 5.4.3. We insist on the fact that this result is valid for any regime of  $d$  such that the  $\ell$  subsequently defined in the statement of Theorem 15 goes to infinity: this encompasses  $d = \text{polylog}(n)$ , a regime in which we can achieve full recovery because in this case, the threshold tends to zero and the vectors  $\psi, \varphi$  are aligned. This is a deep improvement of the novel results in [56].

**Illustration: Erdős-Rényi**

As an immediate consequence, in the Erdős-Rényi graph model with  $d > 1$ , we find that the eigenvector  $\psi$  associated with the outlier  $\lambda_1 \sim d$  satisfies  $|\langle \psi, \mathbf{1}/\sqrt{n} \rangle| \sim \sqrt{1 - 1/d}$ . Some experiments for the Erdős-Rényi graphs are depicted in Figures 5.1-5.2.

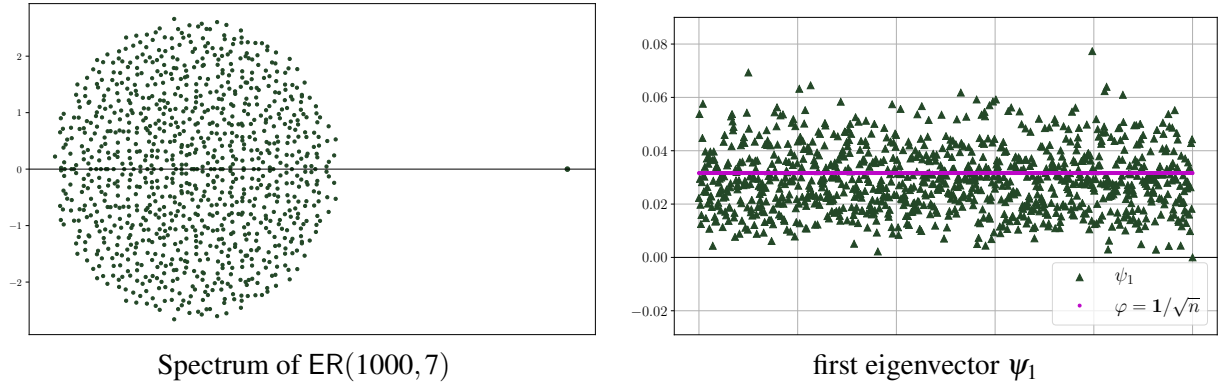


Figure 5.1 – In the left, the spectrum of a directed ER graph with  $d = 7$ . The outlier  $\lambda_1 \approx 7$  is clearly visible. On the right, a plot of the eigenvector of  $A$  associated with  $\lambda_1 \approx 7$ . All entries are positive (Perron’s theorem) and seem stacked close to the real eigenvector  $\varphi$  which is in violet; however, their scalar product is close to  $\sqrt{1 - 1/7} \approx 0.9258$  which is strictly smaller than 1.

This is quite a striking contrast with the (diluted) Erdős-Rényi graphs, where the high eigenvalues are aligned with the high-degree vertices, and the associated eigenvectors are very localized on those vertices.

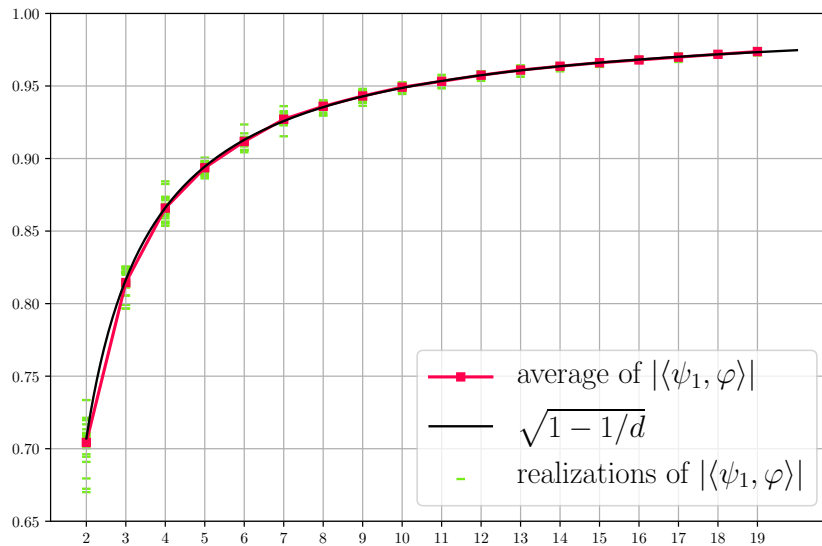


Figure 5.2 – A plot of  $|\langle \psi_1, \varphi \rangle|$  for different mean degrees  $d$ . For each  $d = 2, \dots, 20$ , we made 20 simulations of  $\text{ER}(1000, d)$  and computed  $\langle \psi_1, \varphi \rangle$  (little green dashes). The mean is plotted in red and the prediction  $\sqrt{1 - 1/d}$  is in black.

**Illustration: uniform rank-one matrix**

We also include illustrations of the phenomenon when  $\varphi$  is itself chosen at random on the unit sphere. It is well known in this case that with very high probability, the vector  $\varphi$  is delocalized in the sense that  $|\varphi|_\infty \leq C\sqrt{\ln(n)}/\sqrt{n}$ . We then fixed  $P = \mu\varphi\varphi^*$ . The fourth moment of a standard gaussian is 3, hence one can easily check that  $n|\varphi|_4^4 \approx 3$ .



The top eigenvector of  $A$ , for  $n = 1000, d = 10$ , is depicted in Figure 5.3. For more readability the entries of the signal  $\varphi$  have been sorted by increasing order.

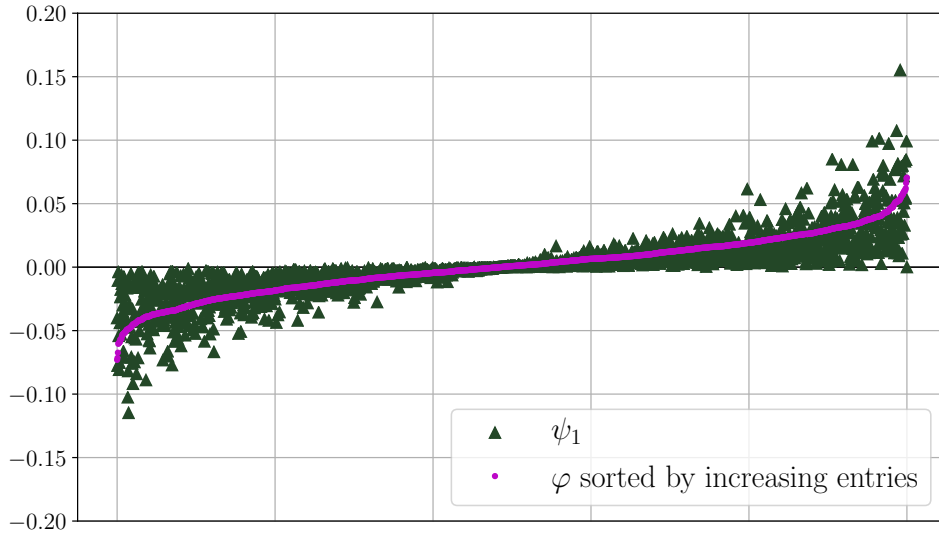


Figure 5.3 – The eigenvector associated with the outlier eigenvalue of the matrix  $A$ , where the matrix  $P$  is simply  $\mu\varphi\varphi^*$  with  $\varphi \sim \text{Haar}(\mathbb{S}^{n-1})$ . Here  $n = 1000$  and  $d = 10$ .

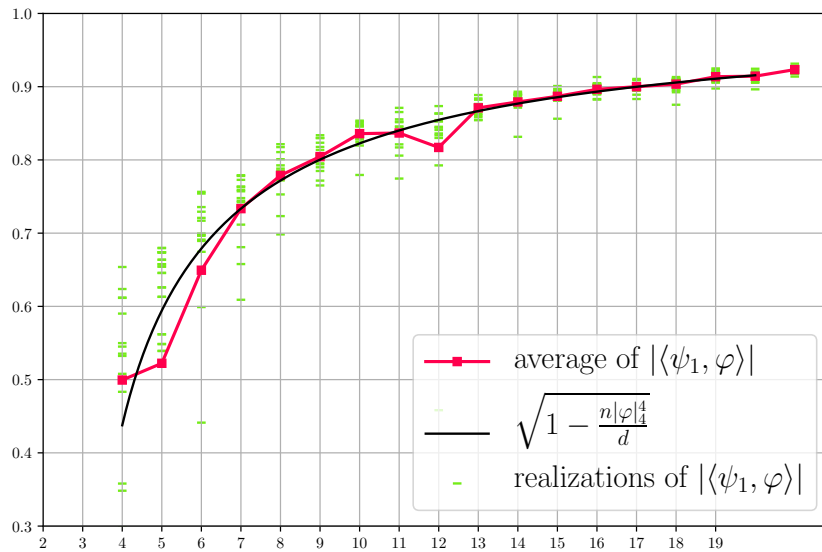


Figure 5.4 – A plot of  $|\langle \psi_1, \varphi \rangle|$  for different mean degrees  $d$ . We first generated one matrix  $P = \mu\varphi\varphi^*$  with  $\varphi$  uniform over  $\mathbb{S}^{n-1}$  (the vector  $\varphi$  is the one depicted in the preceding figure). Then, for each  $d = 2, \dots, 20$ , we made 20 simulations of  $A$  with  $P$  as common underlying matrix, and we computed  $\langle \psi_1, \varphi \rangle$  (little green dashes). The mean is plotted in red and the prediction  $\sqrt{1 - n|\varphi_4|^4/d}$  is in black.

**Notations**

- When  $n$  is an integer,  $[n]$  denotes the set  $\{1, \dots, n\}$ .
- We identify  $\mathbb{R}^n$  with the set  $\ell^2([n])$ . Elements in  $\mathbb{R}^n$  will be noted  $u = (u(x))_{x \in [n]}$ . We will note  $|\cdot|_\infty, |\cdot|_p$  the usual norms on  $\mathbb{R}^n$ , namely

$$|u|_\infty = \max_{x \in [n]} |u(x)| \quad |u|_p = \left| \sum_{x \in [n]} |u(x)|^p \right|^{1/p}.$$

The euclidean norm ( $p = 2$ ) will simply be noted  $|\cdot|$ .

- The operator norm of the matrix  $X$  is noted  $\|X\|$ ; it is the greatest singular value of the matrix. The frobenius norm is noted  $\|X\|_F$  and is defined by  $\|X\|_F = \sqrt{\text{tr}(X^*X)}$ . It is also the  $L^2$ -norm of the singular values.
- We will use the Landau notations.  $a_n = O(b_n)$  means that there is a constant  $c > 0$  such that  $a_n \leq cb_n$  and  $a_n = o(b_n)$  means  $a_n/b_n \rightarrow 0$ .
- When  $t$  is a real number,  $(t)_+$  is equal to  $\max\{t, 0\}$ .
- The letter  $c$  denotes any constant which does not depend on anything in this paper. It might be used from line to line to denote different constants.

### 5.1.3 Detailed statements

Let  $M \in \mathcal{M}_n(\mathbb{R})$  be a random matrix whose entries are iid Bernoulli with parameter  $d/n$ :

$$\mathbf{P}(M_{i,j} = 1) = 1 - \mathbf{P}(M_{i,j} = 0) = \frac{d}{n}.$$

The matrix  $M$  is the adjacency matrix of a directed Erdős-Rényi graph (with self-loops allowed) in the sparse regime, where each vertex has a mean of  $d$  out-neighbors and  $d$  in-neighbors. We will note  $G$  this random digraph. It corresponds to the entries of the matrix  $P$  that will be observed, the remaining ones will be hidden. The observed matrix is

$$A = \left(\frac{n}{d}\right) P \odot M,$$

where  $\odot$  denotes the Hadamard entrywise product of two matrices. Our goal is to infer information on  $P$  from the observation of  $A$ .

We define a real  $n \times n$  matrix  $Q$  and its norm  $\rho$  by

$$Q_{x,y} = nP_{x,y}^2 \quad \rho = \|Q\|. \quad (5.1.4)$$

The detection threshold  $\vartheta$  depends on  $P$  and  $d$  and is defined as

$$\vartheta = \sqrt{\frac{\rho}{d}}. \quad (5.1.5)$$

We also introduce  $\vartheta_0 = L/d$ .

We denote by  $r_0$  the number of eigenvalues of  $P$  which are larger than  $\vartheta \vee \vartheta_0$ :

$$\mu_1 \geq \dots > \mu_{r_0} > \max\left\{\sqrt{\frac{\rho}{d}}, \frac{L}{d}\right\} \geq \mu_{r_0+1} > \dots > \mu_r. \quad (5.1.6)$$

Our result will hold uniformly over a wide class of matrices that match the usual hypothesis from the litterature: low-rank with incoherence conditions. The goal is is not really to restrict the range of applications, but to track the dependence of the error terms with respect to the parameters at stake (such as rank, measure of incoherence or spectral separations).

First of all, we raise the fact that the whole problem is homogeneous, hence we can assume without loss of generality that  $\mu_1 = 1$ .

The class we are interested in is the set of all the real square matrices

$$\mathcal{H} = \mathcal{H}(b, r, \tau, d)$$

that have the following properties.

1. The rank of  $P$  is  $r$ .
2. The eigenvectors of  $P$  are  $b$ -delocalized, in the sense that for every  $k$  in  $[r]$ , we have

$$\sup_{x \in [n]} |\varphi_k(x)| \leq \frac{b}{\sqrt{n}}. \quad (\text{H:incoherent})$$

3. The eigenvalues of  $P$  are positive, smaller than 1, and the number  $\tau < 1$  satisfies

$$\tau > \left( \frac{\vartheta \vee \vartheta_0}{\mu_{r_0}} \right)^2. \quad (5.1.7)$$

We can now state our result in full generality.

**Theorem 15.** *Let  $b, r, \tau$  be positive parameters, the only restriction being that  $r$  is an integer. We define  $\ell = \lfloor 0.5\kappa \log_{2d}(n) \rfloor$ , where  $\kappa$  is the constant defined by*

$$\kappa = \frac{0.249}{3 + \frac{\ln(b)}{\ln(2d)}}. \quad (5.1.8)$$

*Then, there is a  $C_0 = C_0(b, r, \tau, \ln(n))$  and an integer  $n_0 = n_0(b, r, \tau, \ln(n))$  such that, with probability greater than  $1 - 4\ln(n)$ , the following holds true for any matrix  $P$  in the class  $\mathcal{K}(b, r, \tau, d)$  with size  $n > n_0$ :*

*For any  $k \in \{1, \dots, r_0\}$ ,*

$$\lambda_k = \mu_k(1 + \Upsilon_k) \quad (5.1.9)$$

*and for any  $k > r_0$ ,*

$$\lambda_k \leq [\vartheta \vee \vartheta_0](1 + \Upsilon) \quad (5.1.10)$$

*where the real numbers  $\Upsilon_k, \Upsilon$  satisfy*

$$|\Upsilon_k| \leq \frac{C_0}{\ell} \left( \frac{\vartheta \vee \vartheta_0}{\mu_k} \right)^\ell \quad \text{and} \quad |\Upsilon| \leq (C_0)^{\frac{1}{\ell}}. \quad (5.1.11)$$

*Moreover, the numbers  $C_0$  and  $n_0$  are smaller than*

$$c \left( \frac{br_0 \ln(n)}{1 - \tau} \right)^{40}$$

*with  $c$  a universal constant depending on nothing.*

We stated this theorem for any matrix  $P$  in a wide class  $\mathcal{K}$  with parameters  $b, r, \tau, d$  without mentioning any dependence on  $n$ . In fact, all those parameters can indeed depend on  $n$ ; for the result to be non-trivial (eg, error terms going to zero), one only needs the following conditions:

- $n > n_0$  if  $n$  is sufficiently large.
- $C_0^{\frac{1}{\ell}} \rightarrow 0$ , which implies that  $\ell \rightarrow \infty$ .

Those two conditions are met in a variety of cases. If the parameters are independent of  $n$ , then any regime of  $d$  such that  $\ell \rightarrow \infty$  is sufficient and this encompasses  $d = O(n^{o(1)})$ . We can also allow the parameters to go to infinity slowly, for instance  $b, r \sim n^{o(1)}$ . In this case we will need to take  $d$  growing with equivalent speed.

## 5.2 Related work

### Completion and sparsification

The problem of sparse completion consists in observing a very sparse sample of elements of a general object (a matrix, a subspace) carrying some structure (low-rank, delocalized), and trying to reconstruct it. The problem of *matrix* completion has attracted a gigantic amount of attention from researchers in

applied mathematics since the last 15 years; the general philosophy can be grasped by a handful of seminal papers from Candès and Tao [53] and Candès and Recht ([52]), Keshavan Montanari and Oh [92] and Chatterjee ([55]). The survey [65] gives a global view of the field.

The dual problem of completion is *sparsification*, where given a matrix  $P$ , one seeks a procedure to keep only a handful of entries of  $P$  without altering too much its properties ([3, 70, 98, 121]).

Those papers, although different in their methods, show that completing a matrix from the observation of  $nd$  of its entries can only be done if the underlying matrix  $P$  is not too complicated (ie low-rank and sufficiently incoherent), and in that case  $P$  can efficiently be recovered only if  $d$  is of order  $\ln(n)$  — the so-called *information-theoretic threshold* for completion. In [92], there are results for  $d$  fixed, but they are not sharp at all and do not allow any precise asymptotics on specific eigenvalues as we do. To our knowledge, the few works on completion from  $d = O(n)$  entries (see for instance Gamarnik, Li and Zhang [82] and references therein) is focused on  $\varepsilon$ -approximating the whole hidden matrix  $P$ , and never on exact estimation of a specific part of the matrix.

## Random matrices and Erdős-Rényi graphs

From the random matrix point of view, this is all about the spectrum of (sparse) random matrices, or on the eigenvalues of weighted (sparse) random graphs. Estimating the spectral properties of the simplest of random graphs, such as Erdős-Rényi, is already quite difficult ([96]). The complete description of the behavior of the greatest eigenvalues of Erdős-Rényi graphs have been totally explained, in the  $d = o(n)$  sparse setting, only recently by different works: Benaych-Georges, Bordenave, Knowles ([24, 25]) and Alt, Ducatez and Knowles ([11]). Recently, Tikhomirov and Youssef gave similar results for eigenvalues of Erdős-Rényi graphs with iid gaussian weights on the edges ([136]); here, the underlying matrix  $P$  is thus drawn from GOE, and does not meet the usual assumptions of matrix completion.

In those works, it turns out that the behaviour of the (suitably normalized) high eigenvalues of Erdős-Rényi graphs is governed by the high degrees of the graph when  $d < \ln(n)$ , and stick to the edge  $\pm 2$  of the limiting semi-circle law in hen  $d \rightarrow \infty$ . The exact threshold for the disappearance of outliers happens at  $d_* = \ln(4/e)^{-1} \ln(n)$  ([11, 136]). Those results only hold for undirected Erdős-Rényi graphs, and we are not aware of any similar results for directed ER graphs, and even less in the really sparse regime where  $d$  is fixed. Indeed, only the convergence of the global spectrum towards the circle law is now proven (when  $d > \ln(n)^2$ ) by Basak and Rudelson ([18]). Many questions and intuitions are given in the physicist survey [114]. Among them is listed (but not proved) our results on eigenvalues of Erdős-Rényi graphs proved in Paragraph 5.1.2 on page 111. Our results on eigenvectors completes the picture.

## Phase transitions

Our main result is a phase transition for the top eigenvalues of sparse non-hermitian matrices: the whole bulk is confined in a circle of radius  $O(1/\sqrt{d})$ , and depending on the strength of the noise  $d$ , a few outliers appear and they are aligned with the corresponding eigenvalues of the original matrix  $P$ , and their eigenvectors have a nontrivial correlation with the original eigenvector<sup>1</sup>.

This is of course similar to the celebrated BBP transition ([15]), and many similar transitions are already available in the litterature of PCA or low-rank matrix estimation ([26, 103] and references therein). Apart from [115], which has a very different setting than ours, there are no results for phase transitions in low-rank non-symmetric matrix estimations, or in sparse settings.

<sup>1</sup>At least in the rank-1 case; the general case is now under work.

### ‘Asymmetry helps’

One of the key features of this paper is that it deals with top eigenvalues of non-symmetric matrices. While the global behaviour of the spectrum of random matrices is now well understood (see the survey [40] on the circular law, or [83, 114] for physicist’s point of views), finer properties are less known.

Generally speaking, it is easier to deal with eigenvalues of hermitian matrices, notably thanks to the variational characterizations of the eigenvalues. However, in many problems from applied mathematics, it turns out that the spectrum of hermitian matrices can sometimes be less informative than the spectrum of other choices of non-hermitian matrices. A striking instance of this fact was the so-called ‘spectral redemption conjecture’ in community detection ([97] and [43]), where the interesting properties were not captured by the spectrum of the adjacency matrix, but of a non-hermitian matrix, the non-backtracking matrix.

In the setting of matrix perturbation, this insight was remarkably exposed in a very recent and inspiring paper by Chen, Cheng and Fan ([56]). Their setting is more or less the same as ours: an underlying hermitian matrix  $P$ , which is unsymmetrically perturbed into an observed non-hermitian matrix  $A = P + H$ , the entries of  $H$  being all iid. One might favor a singular value decomposition because of the conventional wisdom that SVD is more stable than eigendecomposition when it comes to non-hermitian matrices; but this in fact not true, as shown in their Figure 1, and indeed the eigenvalues are more accurate than the singular values; *verbatim*,

*“When it comes to spectral estimation for low-rank matrices, arranging the observed matrix samples in an asymmetric manner and invoking eign-decomposition properly (as opposed to SVD) could sometimes be quite beneficial.” [56, page 2]*

This is the philosophy we would like to convey here; however, their result hold only on the not-so-sparse regime where  $d > \ln(n)$ . We extend all their results to the fixed  $d$  regime, with an explicit threshold for the detection of  $P$  and exact asymptotics for perturbation of linear forms, at least when the rank is 1.

### Eigenvalues of perturbed matrices

Many works on completion or sparsification rely on a perturbation analysis of the eigenvalues/singular values of perturbed matrices.

For example, one of the key points in many papers is that the sparsification procedure (from  $P$  to  $A$ ) alters the spectral properties of  $P$ , but not too much; indeed the top singular values or eigenvalues do not differ too much, hence keeping only the ‘greater’ items in the SVD or the eigendecomposition of  $A$  is sufficient to weakly recover  $P$ ; that was the idea of [92, 55, 70] (and many of their heirs). The proofs usually rely on estimates on eigenvalues/singular values of the random matrix  $A$ , by combining concentration inequalities and eigenvalues inequalities (such as Weyl’s one). but no sharp asymptotics can be obtained with those methods, a limitation already visible in the seminal paper from Friedman, Kahn, Szemerédi ([80] and Feige and Ofek ([77]). This problem becomes unassailable when  $d$  is really smaller than  $\ln(n)$  or fixed, due to the fact that the underlying graphs are highly non-regular.

Our proof techniques rely on methods introduced by Massoulié and refined by Bordenave, Lelarge and Massoulié ([112, 43]). This powerful and versatile trace method has now been used in various problems for estimating high eigenvalues of sparse random matrices, such as random regular graphs ([33]), biregular bipartite graphs ([50]), digraphs with fixed degree sequence ([62]), bistochastic sparse matrices ([45]), multigraph stochastic blockmodels ([123]).

### Eigenvectors of perturbed matrices

Eigenvector perturbation has also attracted a lot of attention, mainly around variants of the Davis-Kahan theorem or the Neumann trick ([142, 71, 56]). As mentioned in [71], many algebraic bounds

(such as Weyl's inequality or the Davis-Kahan theorems) are tight in the worst case, but wasteful in typical cases. Our proof method does not rely on those general bounds, and naturally integrates the perturbation of eigenvectors; for the moment, the eigenvector perturbation results (Theorem 14) are formulated in the rank-one case, but subsequent work will extend it to the general case.

### 5.3 An algebraic perturbation lemma

We present an eigenvalue perturbation theorem, which extends some the results from [43, Section 4] by taking into account the lack of normality of the structures at stake. We formulate this tool in a separate section because it can be of independent interest.

Let us first give a simple description of the result: if  $u_i, v_i$  are vectors such that  $\langle u_i, v_j \rangle \approx \delta_{i,j}$ , then every matrix close to  $S = \sum \theta_i u_i v_i^*$  has eigenvalues close to the  $\theta_i$ , provided the  $u_i$  are sufficiently well-conditioned. Theorem 17 quantifies this. The novelty here is that **the vectors  $u_i$  need not form an orthonormal family** for the result to hold, and the same for the  $v_i$ 's.

Let us first recollect the Bauer-Fike theorem:

**Theorem 16** (Bauer-Fike, [20], [43] thm 6). *Let  $D$  be a diagonalizable matrix,  $D = P\Sigma P^{-1}$  with  $\Sigma = \text{diag}(\theta_1, \dots, \theta_n)$ . Let  $H$  be a matrix. Then, all the eigenvalues of  $D + H$  lie inside the union of the balls  $B(\theta_i, \varepsilon)$  where  $\varepsilon = \|H\| \|P\| \|P^{-1}\|$ . Moreover, if  $J \subset [n]$  is such that*

$$\left( \cup_{j \in J} B(\theta_j, \varepsilon) \right) \cap \left( \cup_{j \notin J} B(\theta_j, \varepsilon) \right) = \emptyset,$$

*then the number of eigenvalues of  $D + H$  inside  $\cup_{j \in J} B(\theta_j, \varepsilon)$  is exactly  $|J|$ .*

Let  $u_1, \dots, u_r, v_1, \dots, v_r$  be two families of nonzero vectors in  $\mathbb{R}^n$ . Let us note  $U = (u_1, \dots, u_r)$  and  $V = (v_1, \dots, v_r)$ ; those are real matrices with  $n$  lines and  $r$  columns. Our nearly diagonalizable' matrix will be  $S = U\Sigma V^*$  with  $\Sigma = \text{diag}(\theta_1, \dots, \theta_r)$ , the  $\theta_i$  being real positive numbers in decreasing order:

$$0 < \theta_r \leq \theta_{r-1} \leq \dots \leq \theta_1.$$

The center of our investigations will be some real square matrix  $A \in \mathcal{M}_{n,n}(\mathbb{R})$ , not necessarily diagonalizable, but close to  $S$  in operator norm. We make the following assumptions.

1. There is some small  $\eta$  such that

$$\|A - S\| \leq \eta. \quad (5.3.1)$$

2. The matrices  $U$  and  $V$  are well-conditioned, in the following sense:

- They have full rank.
- For some  $N > 1$  we have  $\|U\| \leq N$  and  $\|V\| \leq N$ .
- There is a constant  $h > 0$  such that

$$\|V^*V\| \geq h \quad \text{and} \quad \|U^*U\| \geq h. \quad (5.3.2)$$

- There is a small  $\delta > 0$  such that

$$\|U^*V - I_r\| \leq \delta. \quad (5.3.3)$$

For mainly aesthetic reasons we will assume that  $\delta$  is smaller than  $h$ .

3. The  $\theta_i$  are well-separated from zero, in the sense that

$$\theta_r > \frac{\varepsilon}{2} := 30N^7 \left( \eta + \frac{4\delta\theta_1 N^4}{h} \right). \quad (5.3.4)$$

Some light will be shed upon the definition of  $\varepsilon$  in (5.3.4) in the statement of Theorem 17; however, in  $\varepsilon$  one can see  $\delta/h$  as a measure of the ill-conditioned nature of the matrices  $U$  and  $V$ .

**Theorem 17.** *Under the preceding assumptions, the  $r$  eigenvalues of  $A$  with greater modulus, namely  $|\lambda_1| \geq \dots \geq |\lambda_r|$ , are close to the  $\theta_i$  in the following sense:*

$$|\lambda_i - \theta_i| < r \times \varepsilon := r \times 60N^7 \left( \eta + \frac{4\delta\theta_1 N^4}{h} \right). \quad (5.3.5)$$

Moreover, all the other  $n - r$  eigenvalues of  $A$  have modulus smaller than  $\varepsilon$ .

*Proof.* We begin by defining one alternative to  $U$ , named  $\bar{U}$ , which is ‘really orthonormal’ to  $V$ . To do this, we define the vector space

$$H_i = \text{vect}(v_j : j \neq i).$$

It was assumed earlier that  $V^*V$  was nonsingular, which is equivalent to  $V$  having full rank, hence  $H_i$  has dimension  $r - 1$ . The orthogonal projection on  $H_i$  is given by the formula

$$P_{H_i}(w) = V_i(V_i^*V_i)^{-1}V_i^*w$$

where  $V_i$  is  $V$  whose  $i$ -th column  $v_i$  has been deleted. Note that  $V_i^*V_i$  is a principal minor of  $V^*V$ , hence it is nonsingular itself; moreover, its eigenvalues interlace those of  $V^*V$  and in particular, its smallest eigenvalue is greater than  $c$  through (5.3.2); when taking the inverse, we get  $\|(V_i^*V_i)^{-1}\| \geq 1/h$ .

We now take a look at the vectors defined by  $\tilde{u}_i := u_i - P_{H_i}(u_i)$  and

$$\bar{u}_i := \frac{\tilde{u}_i}{\langle \tilde{u}_i, v_i \rangle} = \frac{u_i - P_{H_i}(u_i)}{\langle u_i - P_{H_i}(u_i), v_i \rangle}. \quad (5.3.6)$$

One feels that  $u_i$  is close to  $\tilde{u}_i$ , and more generally if  $\bar{U} = (\bar{u}_1, \dots, \bar{u}_r)$ , we want to prove that  $\bar{U}$  is close to  $U$ . This is achieved in (5.3.7). Here and after,  $e_j$  denotes the  $j$ -th element of the canonical basis.

By (5.3.3), we have  $|V^*u_i - e_i| = |V^*Ue_i - I_r e_i| \leq \|V^*U - I_r\| = \|U^*V - I_r\| \leq \delta$ , thus we also have  $|V_i^*u_i| = \sum_{j \neq i} |\langle v_j, u_i \rangle|^2 \leq |V^*u_i - e_i|^2 \leq \delta^2$ , and

$$|P_{H_i}(u_i)| = |V_i(V_i^*V_i)^{-1}V_i^*u_i| \leq \|V\| \|(V_i^*V_i)^{-1}\| \delta \leq \frac{\|V\| \delta}{h}$$

and by definition,  $|u_i - \tilde{u}_i| \leq \|V\| \delta/h$ . Moreover, we have  $|\langle u_i, v_i \rangle - 1| \leq \|U^*V - I_r\| \leq \delta$ , so

$$\begin{aligned} |\langle \tilde{u}_i, v_i \rangle - 1| &\leq |\langle u_i, v_i \rangle - 1| + |\langle u_i, v_i \rangle - \langle \tilde{u}_i, v_i \rangle| \\ &\leq \delta + |P_{H_i}(u_i)| \\ &\leq \delta(1 + \|V\|h^{-1}). \end{aligned}$$

When  $\xi$  is a real number smaller than  $1/2$  we have  $|(1 + \xi)^{-1} - 1| \leq 2\xi$ , thus when  $\delta$  is small enough, we have

$$\left| \frac{1}{\langle \tilde{u}_i, v_i \rangle} - 1 \right| \leq 2\delta(1 + \|V\|h^{-1}).$$

We now write

$$\begin{aligned} |\bar{u}_i - u_i| &= \left| \frac{\tilde{u}_i}{\langle \tilde{u}_i, v_i \rangle} - u_i \right| \\ &\leq |\tilde{u}_i - u_i| + \left| \tilde{u}_i \left( \frac{1}{\langle \tilde{u}_i, v_i \rangle} - 1 \right) \right| \\ &\leq \delta \|V\| h^{-1} + |\tilde{u}_i| \delta (1 + \|V\| h^{-1}) \\ &\leq \delta \|V\| h^{-1} + (|u_i| + \delta \|V\| h^{-1}) \delta (1 + \|V\| h^{-1}). \end{aligned}$$

When  $\delta$  is small, this is generously bounded by  $\delta h^{-1}(1 + \|U\|)(1 + \|V\|)$ , which is itself smaller than  $4\delta N^3/h$ , so we get

$$\|\bar{U} - U\| \leq \frac{4\delta N^3}{h} \quad (5.3.7)$$

Finally, we have  $(\bar{U}^*V)_{i,j} = \langle \bar{u}_i, v_j \rangle = 0$  if  $i \neq j$ , and  $\langle \bar{u}_i, v_i \rangle = 1$ , a crucial fact which can also be written as

$$\bar{U}^*V = V^*\bar{U} = \mathbf{I}_r. \quad (5.3.8)$$

We now set  $\Sigma = \text{diag}(\theta_i)$  and  $\bar{S} = \bar{U}\Sigma V^*$ ;

$$\begin{aligned} \|S - \bar{S}\| &\leq \|V\|\|\Sigma\|\|\bar{U} - U\| \\ &\leq \frac{4\delta\theta_1 N^4}{h} := \eta'. \end{aligned}$$

Moreover,  $\bar{S}$  is exactly diagonalizable with eigenvalues  $\theta_1, \dots, \theta_r$ . We note  $\bar{S} = P^{-1}\Sigma'P$  with  $\Sigma' = \text{diag}(\theta_1, \dots, \theta_r, 0, \dots, 0)$  and  $P$  a whole diagonalization matrix. The matrices  $A$  and  $\bar{S}$  are close:

$$\|A - \bar{S}\| \leq \|S - \bar{S}\| + \|A - S\| \leq \eta + \eta'.$$

By the Bauer-Fike theorem, the eigenvalues of  $A$  are entirely included in the union of the balls  $B(\theta_i, \varepsilon)$  and  $B(0, \varepsilon)$ , where  $\varepsilon = (\eta + \eta')\|P\|\|P^{-1}\|$ . We now compute  $P$  and  $\|P\|\|P^{-1}\|$ .

Let  $K = \text{span}(v_1, \dots, v_r)^\perp = \text{im}(V)^\perp = \ker(V^*)$ ; the dimension of  $K$  is  $n - r$ . Let us choose any orthonormal basis  $(w_{r+1}, \dots, w_n)$  of  $K$  and set up  $P = (\bar{U}, W)$  where  $W$  is the  $n \times (n - r)$  matrix whose columns are the  $w_k$ 's. Then, the family  $(\bar{u}_1, \dots, \bar{u}_r, w_{r+1}, \dots, w_n)$  is a diagonalization basis for the matrix  $S$ : more precisely, we have  $S\bar{u}_i = \bar{U}\Sigma V^*\bar{u}_i = \theta_i\bar{u}_i$ , and  $Sw_j = 0$ . We now claim that the inverse of  $P$  is given by

$$P^{-1} = \begin{pmatrix} V^* \\ -W^*\bar{U}V^* + W^* \end{pmatrix}. \quad (5.3.9)$$

We can directly check this using the relations (5.3.8), the orthonormality relation  $W^*W = \mathbf{I}_{n-r}$  and  $V^*W = 0$ , which stems from the choice of  $W$  as a basis for  $\ker(V^*)$ . Indeed,

$$\begin{aligned} \begin{pmatrix} V^* \\ -W^*\bar{U}V^* + W^* \end{pmatrix} P &= \begin{pmatrix} V^* \\ -W^*\bar{U}V^* + W^* \end{pmatrix} (\bar{U} \quad W) \\ &= \begin{pmatrix} V^*\bar{U} & V^*W \\ -W^*\bar{U}V^*\bar{U} + W^*\bar{U} & -W^*\bar{U}V^*W + W^*W \end{pmatrix} \\ &= \begin{pmatrix} \mathbf{I}_r & 0 \\ -W^*\bar{U} + W^*\bar{U} & W^*W \end{pmatrix} \\ &= \begin{pmatrix} \mathbf{I}_r & 0 \\ 0 & \mathbf{I}_{n-r} \end{pmatrix} \\ &= \mathbf{I}_n. \end{aligned}$$

To compute the condition number of  $P$  we use the elementary Lemma 5.3.1, stated hereafter. Clearly,  $\|W\| \leq 1$  and we also supposed that  $\|U\| \geq 1$  and  $\|V\| \geq 1$ , hence

$$\|P\| \leq \sqrt{2}\|\bar{U}\|.$$

For  $P^{-1}$  we note that  $-W^*\bar{U}V^* + W^* = W^*(\mathbf{I}_n - \bar{U}V^*)$ , hence  $\|-W^*\bar{U}V^* + W^*\| \leq \|W\|\|\mathbf{I}_n - \bar{U}V^*\|$ , hence

$$\|P^{-1}\| \leq \sqrt{2}(1 + \|\bar{U}\|\|V\|).$$

We thus get

$$\|P\|\|P^{-1}\| \leq 2\|\bar{U}\|(1 + \|\bar{U}\|\|V\|).$$



We had proven that  $\|\bar{U} - U\| \leq 4\delta N^3/h$ , hence  $\|\bar{U}\| \leq N^3(1 + 4\delta/h) \leq 5N^3$ , an extremely crude bound where we used  $\delta < h$ . Finally, we get

$$\|P\| \|P^{-1}\| \leq 60N^7.$$

To conclude, the spectrum of  $A$  is entirely included in

$$B(0, \varepsilon) \cup B(\theta_r, \varepsilon) \cup \dots \cup B(\theta_1, \varepsilon)$$

where  $\varepsilon = 60N^7(\eta + \eta') = 60N^7(\eta + 4\delta\theta_1 N^4/h)$ . Under (5.3.4), the ball  $B(0, \varepsilon)$  is separated from the  $r$  other balls, hence there are  $r$  eigenvalues of  $A$  such that  $\lambda_i$  is  $\varepsilon$ -close to some  $\theta_j$ . When the balls do not overlap (which happens as soon as  $|\theta_{i+1} - \theta_i| < \varepsilon/2$ ), we directly get that  $|\lambda_i - \theta_i| < \varepsilon$ . However, in general we could very well have multiple values in the  $\theta_j$  or overlapping balls. Anyways, the reader will quickly check on picture 5.5 that the distance between  $\lambda_i$  and  $\theta_i$  can never be greater than  $r$  times  $\varepsilon$ .  $\square$

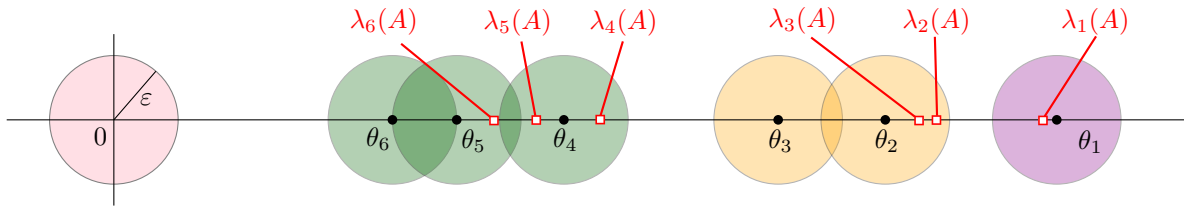


Figure 5.5 – The eigenvalue  $\lambda_i(A)$  need not be in  $B(\theta_i, \varepsilon)$  because there might be some overlap with the closest balls, such as for the green or yellow ones in the drawing. However,  $\lambda_i(A)$  will always be within distance  $r\varepsilon$  of  $\theta_i$ . Note that if  $A$  is real and if  $\lambda_i, \lambda_j$  are two eigenvalues outside  $B(0, \varepsilon)$  with distinct modulus, then they must be real, for the eigenvalues of real matrices are real or come into complex conjugate pairs.

As promised, here is a simple lemma used in the preceding proof.

**LEMMA 5.3.1.** *Let  $M_1 \in \mathcal{M}_{n,r}(\mathbb{R})$  and  $M_2 \in \mathcal{M}_{n,n-r}(\mathbb{R})$  be two matrices; we set  $M = (M_1, M_2) \in \mathcal{M}_{n,n}(\mathbb{R})$ . Then*

$$\|M\| \leq \sqrt{2} \max\{\|M_1\|, \|M_2\|\}.$$

*Proof.* For any  $x, y$  and  $z = (x, y)^*$  we have

$$\begin{aligned} |Mz| &= |M_1x + M_2y| \leq \|M_1\| |x| + \|M_2\| |y| \\ &\leq \max\{\|M_1\|, \|M_2\|\} \sqrt{2} \sqrt{|x|^2 + |y|^2} \\ &= \max\{\|M_1\|, \|M_2\|\} \sqrt{2} |z| \end{aligned}$$

which is valid for any  $z \in \mathbb{R}^n$ .  $\square$

## 5.4 Proof of Theorem 15

We now prove Theorem 15. We introduce an important parameter, namely the *depth* at which we will study the neighborhoods in the graph  $G$ . It is defined as

$$\ell = \left\lfloor \frac{1}{2} \kappa \log_{2d}(n) \right\rfloor \tag{5.4.1}$$

where  $\kappa$  is the positive constant defined in (5.1.8), namely

$$\kappa = \frac{0.249}{3 + \frac{\ln(b)}{\ln(2d)}}.$$

From now on, we pick a matrix  $P$  in the class  $\mathcal{K}(b, r, \tau, d)$  described before Theorem 15. We define  $\Phi = (\varphi_1, \dots, \varphi_{r_0})$  and  $D = \text{diag}(\mu_1, \dots, \mu_{r_0})$ . The columns of  $\Phi$  form an orthonormal family, hence  $\|\Phi\| \leq 1$ .

The ‘candidate eigenvectors’ are  $u_i = A^\ell \varphi_i / \mu_i^\ell$  and  $v_i = (A^*)^\ell \varphi_i / \mu_i^\ell$ , or to put it in matrix form they are the columns of

$$U = A^\ell \Phi D^{-\ell} \quad \text{and} \quad V = (A^*)^\ell \Phi D^{-\ell}. \quad (5.4.2)$$

We now set

$$S = UDV^*. \quad (5.4.3)$$

The matrix  $S$  can also be written as

$$\begin{aligned} S &= A^\ell \Phi D^{-\ell} D^\ell D^{-\ell} \Phi^* A^\ell \\ &= A^\ell \Phi D^{-\ell} \Phi^* A^\ell \\ &= U \Phi^* A^\ell. \end{aligned} \quad (5.4.4)$$

We finally introduce the vector space

$$H = \text{vect}(v_1, \dots, v_{r_0}) = \text{im}(V).$$

#### 5.4.1 Complete algebraic structure of $A, U, V$ with respect to $H$

The behaviour of the matrices  $U, V$  is dictated by a *theoretical covariance matrix*  $\Gamma^{(\ell)}$ , which is a good approximation of the Gram matrices of the columns of  $U$  and  $V$ . It is defined as follows: let  $i, j$  be in  $[r_0]$  and  $t$  be an integer. We will note  $\varphi^{i,j}$  for the Hadamard product between  $\varphi_i$  and  $\varphi_j$ :

$$\varphi^{i,j}(x) = \varphi_i(x) \varphi_j(x).$$

Then, we define the matrix  $\Gamma^{(t)} \in \mathcal{M}_{r_0, r_0}(\mathbb{R})$ :

$$\Gamma_{i,j}^{(t)} = \sum_{s=0}^t \frac{\langle \mathbf{1}, Q^s \varphi^{i,j} \rangle}{(\mu_i \mu_j d)^s}. \quad (5.4.5)$$

**LEMMA 5.4.1.** *For any  $t$ , the matrix  $\Gamma_{i,j}^{(t)}$  is a semi-definite positive matrix with eigenvalues greater than 1, and with*

$$1 \leq \|\Gamma^{(t)}\| \leq \frac{b^{10} r^3}{1 - \tau}$$

having noted that  $1 - \tau \geq 1 - (\vartheta \vee \vartheta_0 / \mu_{r_0})^2$ .

This lemma will be proved in Section 5.5. The main tool for the subsequent analysis of  $U$  and  $V$  is the following theorem, which could also be of independent interest.

**Theorem 18** (complete algebraic structure of  $U$  and  $V$ ). *There is an event with probability greater than  $1 - 4/\ln(n)$  and a universal constant  $c > 0$  such that the following holds for every matrix  $P$  in  $\mathcal{K}(b, r, \tau, d)$ :*

$$\|\Phi^* A^\ell \Phi - D\| \leq r_0 b^2 \times \nu \quad (5.4.6)$$

$$\|U^* V - \mathbf{I}_r\| \leq r_0 b^2 \times \nu \quad (5.4.7)$$

$$\|U^*U - \Gamma^{(\ell)}\| \leq r_0 b^2 \times \mathbf{v} \quad (5.4.8)$$

$$\|V^*V - \Gamma^{(\ell)}\| \leq r_0 b^2 \times \mathbf{v} \quad (5.4.9)$$

$$\|A^\ell U - UD\| \leq C[\vartheta \vee \vartheta_0]^\ell \quad (5.4.10)$$

$$\|A^\ell P_{H^\perp}\| \leq C[\vartheta \vee \vartheta_0]^\ell \quad (5.4.11)$$

where  $C = c(br_0 \ln(n))^{10}$ , and where  $\mathbf{v}$  satisfies

$$\mathbf{v} \leq \min \left\{ \frac{1}{n^{0.25}}, \vartheta^\ell, \vartheta_0^\ell \right\}. \quad (5.4.12)$$

On this event we also have the rough bound:

$$\|A^t\| \leq c \ln(n) n^{2\kappa} \vartheta_0^t \quad (5.4.13)$$

for any  $t \leq \ell$ .

The proof of this theorem occupies the next sections of this paper. We now use this theorem to prove all the results mentioned before.

### 5.4.2 Proof of the eigenvalue perturbation bounds

Our goal is to apply Theorem 17 to  $A^\ell$  and  $S$ .

Let us begin by checking that  $U, V$  satisfy the good behaviour in Condition 2:

$$\|U\| = \sqrt{\|U^*U\|} \leq \frac{b^{10} r_0^3}{1 - \tau} + \frac{r_0 b^2}{n^{0.25}} \quad (5.4.14)$$

and the same for  $\|V\|$ . Upon adjusting the constant  $c$ , the RHS is smaller than  $N := cb^{10} r_0^3 / (1 - \tau)$ . For the conditioning properties of  $U, V$ , one only has to note that by Weyl's perturbation principle, the smallest eigenvalue of  $\|U^*U\|$  or  $\|V^*V\|$  is greater than

$$\lambda_{\min}(\Gamma^{(\ell)}) - r_0 b^2 \mathbf{v} \geq 1 - \frac{cb^2}{n^{0.25}} \quad (5.4.15)$$

which automatically ensures that  $U^*U$  or  $V^*V$  are nonsingular, at least as soon as  $r_0 b^2 / n^{0.25} < 1$  — from now on we suppose that  $r_0 b^2 / n^{0.25} < 0.5$ , which happens as soon as  $n > 16r_0^4 b^8$ . As a consequence, with the  $h$  from Theorem 17 being  $h = 0.5$ , we get

$$\|U^*U\|, \|V^*V\| \geq h. \quad (5.4.16)$$

Moreover, we also have

$$\|U^*V - I_{r_0}\| \leq r_0 b^2 \mathbf{v} := \delta. \quad (5.4.17)$$

This settles the good-behaviour conditions for  $U, V$ . Note that (5.4.15) gives us for free that the norm of  $(V^*V)^{-1}$  or  $(U^*U)^{-1}$ , which is also the inverse of its smallest eigenvalue, is greater than  $1/h = 2$ , a bound to be evoked later:

$$\|(V^*V)^{-1}\| \leq 2. \quad (5.4.18)$$

We can now define the orthogonal projection  $P_H$  onto the subspace

$$H = \text{span}(v_1, \dots, v_{r_0}) = \text{im}(V),$$

and we have  $P_H = V(V^*V)^{-1}V^*$ . We also note  $P_{H^\perp} = I_n - P_H$  the projection matrix on  $H^\perp$ . Note that  $SP_{H^\perp} = 0$ , and that  $SP_H = S$ . As a consequence,

$$\|A^\ell - S\| \leq \|A^\ell P_H - SP_H\| + \|SP_H\| + \|A^\ell P_{H^\perp}\| \quad (5.4.19)$$

$$\begin{aligned} &\leq \|A^\ell P_H - S\| + \|A^\ell P_{H^\perp}\| \\ &\leq \|A^\ell V(V^*V)^{-1} - UD\| \|V^*\| + \|A^\ell P_{H^\perp}\|. \end{aligned} \quad (5.4.20)$$

We now perform a little trick: we first decompose  $U = P_H U + P_{H^\perp} U$ . We note  $W = P_{H^\perp} U$ . The rough idea here is that  $V^*U \approx I_{r_0}$ , hence  $P_H(U) \approx V(V^*V)^{-1}$ . More precisely,

$$\begin{aligned} P_H U &= V(V^*V)^{-1}V^*U \\ &= V(V^*V)^{-1} + H_1 \end{aligned}$$

where  $H_1 := V(V^*V)^{-1}(V^*U - I_{r_0})$  has very small norm. In fact, from the norm bounds in (5.4.14)-(5.4.17)-(5.4.18), we find that  $\|H_1\| \leq 2\delta N$ . We thus have

$$U = V(V^*V)^{-1} + H_1 + W.$$

But now, we get the following:

$$\begin{aligned} \|A^\ell V(V^*V)^{-1} - UD\| &\leq \|A^\ell U - UD\| + \|A^\ell H_1\| + \|A^\ell W\| \\ &\leq \|A^\ell U - UD\| + \|A^\ell\| \|H_1\| + \|A^\ell P_{H^\perp}\| \|U\| \end{aligned}$$

and everything here is either going to zero or is dominated (up to  $\log n$  terms and constants) by  $\vartheta^\ell$ . More precisely, going back to (5.4.20) yields (with  $C$  defined in the theorem):

$$\begin{aligned} \|A^\ell - S\| &\leq \|A^\ell U - UD\| + \|A^\ell\| \|H_1\| + \|A^\ell P_{H^\perp}\| \|U\| + \|A^\ell P_{H^\perp}\| \\ &\leq C\vartheta^\ell + \|A^\ell\| 2N\delta + C\vartheta^\ell 2N\delta + C\vartheta^\ell \\ &\leq C[\vartheta \vee \vartheta_0]^\ell + c \ln(n) n^{2\kappa} (L/d)^\ell 2N\delta + C[\vartheta \vee \vartheta_0]^\ell 2N\delta + C[\vartheta \vee \vartheta_0]^\ell \end{aligned}$$

and thanks to (5.4.12) everything here is of order  $[\vartheta \vee \vartheta_0]^\ell$ ; more precisely, there is a  $C'$  such that

$$\|A^\ell - S\| \leq C'[\vartheta \vee \vartheta_0]^\ell := \eta.$$

The number  $C'$  depends on  $b, r, (1 - \tau)^{-1}$  and  $\ln(n)$  polynomially and can indeed be taken to be as follows:

$$C' = \frac{c(br_0 \ln(n))^{20}}{1 - \tau} \quad (5.4.21)$$

with  $c$  some universal constant.

We are now ready to cast Theorem 17 to  $A^\ell$  and  $S$ . The two first conditions listed before the theorem are met. We still have to check the last one, (5.3.4). Remember that  $\mu_1 \leq 1$  (see the definition of the class  $\mathcal{K}$ ). The error term in Theorem 17 is going to be smaller than

$$\begin{aligned} \varepsilon &= 60N^7 (C'[\vartheta \vee \vartheta_0]^\ell N + 8\delta N^4) = [\vartheta \vee \vartheta_0]^\ell (60C'N^8 + c8b^2vN^{11}\delta / ([\vartheta \vee \vartheta_0]^\ell)) \\ &= C''[\vartheta \vee \vartheta_0]^\ell \end{aligned}$$

where  $C''$  depends on  $b, r, (1 - \tau)^{-1}$  and  $\ln(n)$  polynomially.

Due to the last item in the definition of  $\mathcal{K}$ , we know that  $\mu_{r_0}^\ell > \tau^{-\frac{\ell}{2}} [\vartheta \vee \vartheta_0]^\ell$  where  $\tau \in (0, 1)$  is a fixed number, so the separation condition (5.3.4) is always fulfilled when  $n$  is larger than some constant.

Theorem 17 now yields a  $C_0 = C_0(b, r, \tau, \ln(n))$  such that

$$|\lambda_i(A)^\ell - \mu_i^\ell| \leq r_0 \varepsilon = C_0[\vartheta \vee \vartheta_0]^\ell$$

for any  $i \in [r_0]$ , and  $|\lambda_i(A)|^\ell \leq \varepsilon$  for  $i > r_0$ .

In the first case, using  $|(1 + \xi)^{\frac{1}{\ell}} - 1| \leq \xi/\ell$ , we have  $\lambda_i(A) = \mu_i(1 + \Upsilon)$  where

$$\begin{aligned} \Upsilon &\leq \frac{C_0[\vartheta \vee \vartheta_0]^\ell}{\ell \mu_i^\ell} \\ &\leq \frac{C_0}{\ell} \left( \frac{\vartheta \vee \vartheta_0}{\mu_i} \right)^\ell. \end{aligned}$$

We can check that  $C_0$  is smaller than

$$c \left( \frac{br_0 \ln(n)}{1 - \tau} \right)^{40}$$

which is an extremely prodigal bound, but will nevertheless match all our needs. This closes the proof of Theorem 15.

### 5.4.3 Proof of Theorem 14

Suppose that the rank of  $P$  is  $r = 1$ , so that  $P = \mu \phi \phi^*$ . Let us note  $\phi = \varphi^2$ , so that  $Q = n\mu^2 \phi \phi^*$ . We have

$$\rho^2 = n^2 \mu^4 \max_{|w|=1} |\phi \phi^* w|^2 = n^2 \mu^4 |\phi|^2 \max_{|w|=1} |\langle \phi, w \rangle|^2 = n^2 \mu^4 |\phi|^4$$

hence we have  $\rho = n\mu^2 |\phi|_4^4$ .

In our case one can quickly check that  $Q^s \phi = (n\mu^2)^s |\phi|^{2s} \phi$  so  $\langle \mathbf{1}, Q^s \phi \rangle = (n\mu^2 |\phi|^2)^s$ . By definition, we have

$$\Gamma_{1,1}^{(\ell)} = \sum_{s=0}^{\ell} \frac{\langle \mathbf{1}, Q^s \phi \rangle}{(\mu^2 d)^s} = \sum_{s=0}^{\ell} \left( \frac{n\mu^2 |\phi|^2}{\mu^2 d} \right)^s = \frac{(n|\phi|^2/d)^{\ell+1} - 1}{n|\phi|^2/d - 1} = \frac{1 + o(1)}{1 - n|\phi|^2/d}. \quad (5.4.22)$$

It can easily be seen that we have

$$P_H(\psi) = \psi + o(1). \quad (5.4.23)$$

*Proof.* Suppose that  $\lambda$  is an eigenvalue of 1 with  $|\lambda| > \vartheta$  and note  $\psi$  its associated left-eigenvector. Then,  $\psi = P_H(\psi) + P_{H^\perp}(\psi)$ . But we also have

$$|P_{H^\perp}(\psi)| = \frac{1}{|\lambda|^\ell} |P_{H^\perp}(A^*)^\ell| \leq \frac{\|A^\ell P_{H^\perp}\|}{|\lambda|^\ell} \leq \frac{C\vartheta^\ell}{|\lambda|^\ell} \rightarrow 0$$

where we used (5.4.11). □

We suppose that we are above the threshold (ie  $d > n|\phi|_4^4$ ), and we note  $\psi$  the eigenvector associated with  $\lambda_1(A)$ . Equation (5.4.23) also means that if  $\bar{v} = v_1/|v_1|$  (where we recall that  $v_1 = (A^*)^\ell \phi/\mu^\ell$ ) then  $|\psi - \bar{v}| \rightarrow 0$  and  $|\langle \psi, \phi \rangle - \langle \bar{v}, \phi \rangle| \rightarrow 0$ . Finally, we have proven in (5.4.6) that  $|\langle v_1, \phi \rangle| \rightarrow 1$  and in (5.4.9) that  $|v_1| \sim \sqrt{\Gamma_{1,1}^{(\ell)}}$ . From all this and (5.4.22) it is clear that

$$|\langle \psi, \phi \rangle| \sim \frac{1}{\sqrt{\Gamma_{1,1}^{(\ell)}}} \sim \sqrt{1 - \frac{n|\phi|_4^4}{d}}.$$

## 5.5 Consequences of algebraic incoherence

Our goal in this section is to prove several useful estimates on  $Q$  and  $P$ , mainly linked with the incoherence property (H:incoherent). Through all the section,  $P$  is a matrix in  $\mathcal{K}(b, r, \tau, d)$  and in particular  $\mu_1 = 1$ .

### 5.5.1 The incoherence hypothesis for $P$

For any  $i$ , we supposed that  $|\varphi_i|_\infty \leq b/\sqrt{n}$ , therefore the entries of  $P$  are of order  $O(1/n)$ . More precisely,

$$\begin{aligned} |P_{x,y}| &= \left| \sum_{k=1}^r \mu_k \varphi_k(x) \varphi_k(y) \right| \\ &\leq \sqrt{\sum_{k=1}^r |\varphi_k(x)|^2} \sqrt{\sum_{k=1}^r |\varphi_k(y)|^2} \\ &\leq \frac{b^2}{n}. \end{aligned} \quad (5.5.1)$$

If we define

$$L = n \max_{x,y} |P_{x,y}|, \quad (5.5.2)$$

the preceding inequality ensures that  $L$  is bounded away infinity; more precisely,

$$L \leq b^2. \quad (5.5.3)$$

### 5.5.2 The threshold matrix

It is clear for its definition that the entries of  $Q$  are positive, hence by the Perron-Frobenius theorem its operator norm  $\rho$  is also an eigenvalue of  $Q$  with higher modulus.

In this section, we will note  $s$  the rank of  $Q$ . It is well known that the rank is submultiplicative for the Hadamard product, hence we have  $s = \text{rank}(P \odot P) \leq \text{rank}(P)^2 = r^2$ .

Now, let us briefly mention an elementary bound on the operator norm  $\rho = \|Q\|$ . By variational principles, we have  $\rho = \max_w \langle w, Qw \rangle / |w|^2$  for any nonzero  $w \in \mathbb{R}^n$ . Taking  $w = \mathbf{1}$  yields

$$\rho \geq \frac{\langle \mathbf{1}, Q\mathbf{1} \rangle}{n} = \sum_{x,y} P_{x,y}^2$$

and we hereby recognize the Frobenius norm of the matrix  $P$ , which is itself equal to  $\mu_1^2 + \dots + \mu_r^2$ . In addition, using (5.5.3), we get

$$\rho \geq \|P\|_F^2 \geq \frac{L^2}{b^4}. \quad (5.5.4)$$

The important fact here is that  $\rho$  does not go to zero, or equivalently  $Q$  still carries information, a normal consequence of (H:incoherent).

We now prove that  $Q$  is delocalized, in the sense that its eigenvectors have entries of order  $1/\sqrt{n}$ . We write the spectral decomposition of  $Q$  as

$$Q = v_1 \psi_1 \psi_1^* + \dots + v_s \psi_s \psi_s^* \quad (5.5.5)$$

with  $v_1 = \rho$  the Perron eigenvalue and  $v_2, \dots, v_s$  the other eigenvalues of  $Q$ , which are real, nonzero, and with absolute value smaller than  $\rho$ . We have

$$\begin{aligned} (Q^2)_{x,x} &= \sum_{y \in [n]} Q_{x,y}^2 = n^2 \sum_{y \in [n]} |P_{x,y}|^4 \\ &\leq n^2 \sum_{y \in [n]} \frac{L^4}{n^4} \\ &\leq \frac{L^4}{n}. \end{aligned}$$

However, we also have  $(Q^2)_{x,x} = v_1^2 \psi_1(x)^2 + \dots + v_s^2 \psi_s(x)^2$ . We thus have proven that the delocalization property for  $P$  implies delocalization for  $Q$ : for any  $x$  and  $k$ ,

$$|v_k \psi_k(x)| \leq \frac{L^2}{\sqrt{n}}. \quad (5.5.6)$$

For any vector  $w$ , we note  $w^2$  the vector defined as  $w^2(x) = |w(x)|^2$ . The following proposition will be crucial; the idea it conveys is that  $\langle \mathbf{1}, Q^t w^2 \rangle$  is essentially bounded by  $\rho^t$ , a result in the flavour of Perron-Frobenius theory.

**PROPOSITION 5.5.1.** *For any  $t \geq 2$  and any unit vector  $w \in \mathbb{R}^2$ ,*

$$\langle \mathbf{1}, Q^t w^2 \rangle \leq b^8 r^2 \rho^t \quad (5.5.7)$$

*Proof.* For any  $k$ , we have  $|\psi_k|_\infty \leq L^2/\sqrt{n}$ , hence  $\langle \mathbf{1}, \psi_k \rangle \leq L^2 \sqrt{n}$  and if  $w$  is a unit vector,  $\langle w^2, \psi_k \rangle \leq (L^2/\sqrt{n}) \sum_x w(x)^2 = L^2/\sqrt{n}$ , hence

$$\begin{aligned} \langle \mathbf{1}, Q^t w^2 \rangle &= \sum_{k=1}^s v_k^t \langle \mathbf{1}, \psi_k \rangle \langle w^2, \psi_k \rangle \\ &\leq \sum_{k=1}^s v_k^t \frac{L^2 \sqrt{n}}{v_k} \frac{L^2}{v_k \sqrt{n}} \\ &= \sum_{k=1}^s v_k^{t-2} L^4 \\ &= L^4 \rho^{t-2} (1 + (v_2/\rho)^{t-2} + \dots + (v_s/\rho)^{t-2}). \end{aligned}$$

When  $t$  is greater than 2, the last line is smaller than  $L^4 \rho^{t-2} s$ , and  $s$  is itself smaller than  $r^2$ . Finally, we use (5.5.4), which implies  $\rho^{-2} \leq b^8/L^4$ .  $\square$

Peeling through the details above, we could also get the bound

$$\langle \mathbf{1}, Q^t w^2 \rangle \leq L^4 \left( \frac{r}{\|P\|_F} \right)^2 \rho^t.$$

We will not use this.

Let us end this section by the proof of Lemma 5.4.1. We start by recalling the definition of the theoretical covariance  $\Gamma^{(t)} \in \mathcal{M}_{r_0, r_0}(\mathbb{R})$ :

$$\Gamma_{i,j}^{(t)} = \sum_{s=0}^t \frac{\langle \mathbf{1}, Q^s \varphi^{i,j} \rangle}{(\mu_i \mu_j d)^s}. \quad (5.5.8)$$

For the convenience of the reader we restate the lemma here.

**LEMMA 5.5.2.** For any  $t$ , the matrix  $\Gamma_{i,j}^{(t)}$  is a semi-definite positive matrix with eigenvalues greater than 1, and with

$$1 \leq \|\Gamma^{(t)}\| \leq \frac{b^{10}r^3}{1-\tau}$$

having noted  $1 - \tau = 1 - (\vartheta/\mu_{r_0})^2 > 0$ .

*Proof.* We use  $\|\Gamma^{(t)}\| \leq \|\Gamma^{(t)}\|_F \leq r\|\Gamma^{(t)}\|_\infty$ . We use the same bounds as for the preceding proof, except in the second line where we use

$$|\langle \varphi^{i,j}, \psi_k \rangle| \leq n|\varphi_i|_\infty|\varphi_j|_\infty|\psi_k|_\infty = n(b/\sqrt{n})^2(L^2/v_k\sqrt{n}) = b^2L^2/(v_k\sqrt{n}).$$

The same reasoning yields

$$\begin{aligned} \langle \mathbf{1}, Q^t \varphi^{i,j} \rangle &= \sum_{k=1}^s v_k^t \langle \mathbf{1}, \psi_k \rangle \langle \varphi^{i,j}, \psi_k \rangle \\ &\leq \sum_{k=1}^s v_k^t \frac{L^2\sqrt{n}}{v_k} \frac{L^2b^2}{v_k\sqrt{n}} \\ &= \sum_{k=1}^s v_k^{t-2} L^4 b^2 \\ &= L^4 b^2 \rho^{t-2} r^2 \\ &\leq b^{10} r^2 \rho^t. \end{aligned}$$

Going back to the sum defining  $\Gamma_{i,j}^{(t)}$ , we get

$$|\Gamma_{i,j}^{(t)}| \leq b^{10} r^2 \frac{1}{1 - (\rho/\mu_i \mu_j d)}$$

and as a consequence,

$$\|\Gamma^{(t)}\| \leq \frac{r^3 b^{10}}{1 - \frac{\rho}{\mu_0^2 d}} = \frac{b^{10} r^3}{1 - \tau}$$

On the other hand, if we note

$$C_{i,j}^{(s)} = \frac{\langle \mathbf{1}, Q^s \varphi^{i,j} \rangle}{(\mu_i \mu_j d)^s}$$

then it is not difficult to see that  $C^{(s)}$  is indeed a semi-definite positive matrix; more precisely, if we introduce  $\pi_s(x) = \sqrt{Q^s \mathbf{1}(x)} \geq 0$  and  $\Pi_s = \text{diag}(\pi_s)$ , then

$$C^{(s)} = d^{-s} \cdot D^{-s} \Phi^* \Pi_s \Phi D^{-s}$$

which is clearly SDP. The matrix  $\Gamma^{(t)}$  is thus a sum of  $C^{(0)} = I_{r_0}$  and  $t - 1$  SDP matrices, hence it is itself an SDP matrix and its eigenvalues are greater than the eigenvalues of  $I_{r_0}$ , hence the first statement.  $\square$

### 5.5.3 Bounds on the entries of $Q$

We will also need an analog of  $L$  but for the matrix  $Q$ . The inequalities proved in this section are more or less the same as in the preceding section.

We define  $K$  as

$$K = n \max_{x,y} Q_{xy} / \rho. \quad (5.5.9)$$



We have

$$L = n \max_{x,y} |P_{xy}| = \sqrt{K\rho}.$$

We have  $1 \leq K \leq r^2 b^4$ . The scalar  $K$  is a delocalization parameter: it is scale invariant.

For any  $x$ , we find

$$\sum_y Q_{xy} = n \sum_y (P_{xy})^2 = n(P^2)_{xx} \leq rb^2 \leq rb^2 \rho.$$

It follows that for any  $x, y$ ,

$$(Q^2)_{xy} = \sum_z Q_{xz} Q_{zy} \leq \frac{Krb^2}{n} \rho^2.$$

Let  $(\psi_k)$  be an ON basis of eigenvectors of  $Q$  with eigenvalues  $(v_k)$ . Let  $t \geq 2$ , we write for any  $x, y$ ,

$$(Q^t)_{xy} = \sum_k v_k^t \psi_k(x) \psi_k(y) \leq \rho^{t-2} \sum_k v_k^2 |\psi_k(x)| |\psi_k(y)| \leq \rho^{t-2} \sqrt{(Q^2)_{xx}} \sqrt{(Q^2)_{yy}},$$

where the last step follows from Cauchy-Schwarz inequality. In particular, for any  $t \geq 2$  and  $x, y$

$$(Q^t)_{xy} \leq \frac{Krb^2}{n} \rho^t. \quad (5.5.10)$$

It follows from (5.5.9) that Equation (5.5.10) also holds for  $t = 1$ .

## 5.6 Coupling graphs and trees

The basic ingredients for the proofs of Theorem 15 and related statements are directed Galton-Watson trees and martingales defined on them. We start to introduce the notations and vocabulary for this.

### 5.6.1 Galton-Watson trees and Erdős-Rényi graphs

#### Graph-theoretic definitions

A marked graph with mark space  $\mathbb{N}$  is a digraph  $(V, E)$ , loops allowed, endowed with a mark function  $\iota: V \rightarrow \mathbb{N}$  which is finitely supported. This definition might seem strange, but it really is a convenient setting for studying all the different kinds of graphs and trees that we are going to use. We are going to note  $\mathcal{G}_*$  the set of all the rooted directed graphs on a common countable set  $V$  and with mark space  $\mathbb{N}$ . Formally, the elements of  $\mathcal{G}_*$  are triples  $(g, o, \iota)$ , with  $o$  the root, but in general we will drop the mark function  $\iota$  and simply write  $(g, o)$ .

Let  $(g, o, \iota) \in \mathcal{G}_*$  and  $g = (V, E)$ . If  $W \subset V$  is a subset of  $V$  containing the root, then the **induced subgraph**  $(g, o, \iota)_W$  is defined as follows: the underlying graph is  $g_W := (V, E_W)$  where  $(i, j) \in E_W$  if and only if  $(i, j) \in E$  and both  $i$  and  $j$  are in  $W$ , and the mark function  $\iota_W$  is given by  $\iota_W(v) = \iota(v) \mathbf{1}_{v \in W}$ .

The elements in  $\mathcal{G}_*$  are digraphs, and therefore we need to make a distinction between directed paths and undirected paths. Let  $g = (V, E)$  be any directed graph.

- If  $(x, y) \in E$  we note  $x \rightarrow y$ ,
- if  $x \rightarrow y$  or  $y \rightarrow x$  or both, we note  $x \sim y$ .

Every directed graph  $g$  can be transformed into an undirected graph  $\hat{g} = (V, \hat{E})$  by simply forgetting the direction of the edges:  $(x, y) \in \hat{E}$  iff  $x \sim y$  in  $G$ .

If  $u, v \in V$ , a **directed path** or **dipath** from  $x$  to  $y$  is a sequence of vertices  $x_0 = x, x_1, \dots, x_k = y$  such that for every  $s$  we have  $u_s \rightarrow u_{s+1}$ . A **path** is the same except that we only ask  $x_s \sim x_{s+1}$ .

- The length of the shortest directed path between  $x$  and  $y$  is denoted by  $d^+(x, y)$ .

- The length of the shortest directed path between  $y$  and  $x$  is also denoted by  $d^-(x, y) = d^+(y, x)$ .
- The length of the shortest path is denoted by  $d(x, y)$ .
- When  $(g, o)$  is a rooted graph,  $\mathcal{P}_g(o, t)$  is the set of paths in  $g$  starting from the root  $o$  and having  $t$  steps.

The set of all  $y$  such that  $d^+(x, y) \leq t$  is the **forward ball**  $B_G^+(x, t)$  and the set of all  $y$  such that  $d(x, y) \leq t$  is the **ball**  $B_G(x, t)$ . When no confusion can arise, we write  $B^+$  or  $B$  instead of  $B_G, B_G^+$ .

A **cycle** in the graph  $g$  is a sequence of distinct vertices  $x_1, \dots, x_k$  such that  $x_s \sim x_{s+1}$  for every  $s < k$  and  $x_k \sim x_1$ . The number  $k$  is the length of the cycle.

A **tangle-free** subgraph of  $g$  is a subgraph of  $G$  that contains at most one cycle. The graph  $g$  is  $t$ -tangle free if for every vertex  $x$ , the ball  $B_g(x, t)$  is tangle-free.

If  $t$  is an integer and  $(g, x) \in \mathcal{G}_*$ , then  $(g, x)_t$  is the subgraph of  $(g, x)$  induced by  $B_g(x, t)$ , as defined several lines earlier, and similarly  $(g, x)_t^+$  is the subgraph of  $(g, x)$  induced by  $B^+(x, t)$ .

### Definition of the graph $G$ and the GW tree

We recall that  $G$  is the directed Erdős-Rényi graph (with loops) whose adjacency matrix is given by  $M$ . Let  $x$  be an arbitrary element of  $[n]$ . We root the graph  $G$  at  $x$ , and we mark every vertex with itself: the mark of vertex  $x \in [n]$  is simply the integer  $\iota(x) = x$  and the mark of every vertex in  $\mathbb{N} \setminus [n]$  is set to zero. The resulted marked graph  $(G, x)$  is an element of  $\mathcal{G}_*$ .

We now define the directed Galton-Watson tree  $T$  in the following way. Starting from its root  $o$ , every vertex has a  $\text{Poi}(2d)$  number of children. Every edge  $(u, v)$  is independently given a unique direction  $u \rightarrow v$  or  $v \rightarrow u$  with probability  $1/2$ . This yields a random directed tree. Equivalently, each vertex has a  $\text{Poi}(d)$  number of ‘out-children’ and a  $\text{Poi}(d)$  number of ‘in-children’.

Finally, every non-root vertex  $o'$  is independently given a random mark  $\iota(o')$  which is uniform on  $[n]$ . The root is given a special mark  $\iota(o) = \emptyset$ . The resulting element of  $\mathcal{G}_*$  will be noted  $(T, \emptyset)$ . We shall say that the tree  $(T, \emptyset)$  is *grown from the seed*  $\emptyset$ .

### 5.6.2 Growth properties: trees

Let us first state several properties on the growth of the tree  $T$  first, then on the graph  $G$ . They are directly drawn from [43], see Section 8 for the tree, and Sections 9.1-9.2 for the graph.

Clearly, the underlying undirected tree obtained from  $T$  by deleting the marks and orientations is simply a  $\text{Poi}(2d)$  Galton-Watson tree, which allows us to use without any further efforts all the growth properties of GW trees. For more readable statements, we note here and after

$$D := 2d.$$

**LEMMA 5.6.1.** *Let us note  $S_t$  the number of vertices at distance  $t$  from the root  $o$  of  $T$ . There are two constants  $c_0, c_1 > 0$  such that for all  $\lambda > 0$ ,*

$$\mathbf{P}(S_t \leq \lambda D^t \text{ for all } t) \geq 1 - c_0 e^{-c_1 \lambda}. \quad (5.6.1)$$

Moreover, for every  $p \geq 2$ , there is a constant  $c_p > 0$  such that

$$\mathbf{E} \left[ \max_{t \geq 1} \left( \frac{S_t}{D^t} \right)^p \right] \leq (c_p \ln(n))^2. \quad (5.6.2)$$

We take  $\lambda = c_0^{-1} \ln(c_1 n^2)$  in the first inequality of the lemma; for any  $n \geq 3$ , we get

$$\mathbf{P}(S_t \leq C \ln(n) D^t \text{ for all } t) \geq 1 - \frac{1}{n^2} \quad (5.6.3)$$

where  $C$  is the constant  $c_1^{-1}(2 + \ln(c_0))$ . On this event, we have

$$\begin{aligned} |(T, \mathcal{O})_t| &= 1 + S_1 + \dots + S_t \\ &\leq C \ln(n)(1 + D + \dots + D^t) \\ &\leq CD \ln(n) D^t. \end{aligned}$$

The most important consequence for us is the following: suppose that  $\{(T_x, x) : x \in [n]\}$  is a family of random rooted marked trees, the tree  $(T_x, x)$  having the distribution of  $(T, x)$ , but the whole family is not necessarily independent. Then, by sub-additivity, on an event with probability greater than  $1 - 1/n$ , the following holds for any integer  $t$  and any  $(T_x, x)$  in the family:

$$|(T_x, x)_t| \leq CD \ln(n) D^t. \quad (5.6.4)$$

Another easy consequence of (5.6.2) is the following: for any  $t$ ,

$$\begin{aligned} \mathbf{E}[|(T, \mathcal{O})_t|^2] &= \mathbf{E}[|1 + S_1 + \dots + S_t|^2] \\ &\leq (D^t)^2 \mathbf{E}[|1 + D^{-1}S_1 + \dots + D^{-t}S_t|^2] \\ &\leq D^{2t} t^2 \mathbf{E}[\max_{s \geq 0} (D^{-s} S_s)^2] \\ &\leq D^{2t} t^2 c_2^2 \ln(n)^2. \end{aligned}$$

As a consequence, we have

$$\sqrt{\mathbf{E}[|(T, \mathcal{O})_t|^2]} \leq c D^t t \ln(n) \leq c \ln(n)^2 n^\kappa \quad (5.6.5)$$

for some  $c > 0$ , where we used  $t \leq \ell \leq \ln(n)$  and  $D^t \leq n^\kappa$ .

### 5.6.3 Growth properties: graphs

We now establish the same properties as before, but for the directed graph  $G$  whose adjacency matrix is  $M$ . We start by proving that up to a depth of order  $\ln(n)$ , the graph  $G$  has few cycles. We recall the definition of  $\ell$

$$\ell := \frac{1}{2} \lfloor \kappa \log_{2d}(n) \rfloor \quad (5.6.6)$$

where  $\kappa \in ]0, 0.25[$  is the constant defined in (5.1.8). We will often go to depths such as  $2\ell$  so we'll use the notation

$$\ell' = 2\ell = \lfloor \kappa \log_{2d}(n) \rfloor.$$

All the results in this sections flow effortlessly from their undirected counterpart. In fact, note  $\hat{G}$  the graph  $G$  in which the directions have been erased. Pick any vertices  $x, y$ . Then,

$$\begin{aligned} \mathbf{P}(\{x, y\} \in E(\hat{G})) &= \mathbf{P}((x, y) \in E(G) \text{ or } (y, x) \in E(G)) \\ &= \mathbf{P}(x \rightarrow y, y \not\rightarrow x) + \mathbf{P}(y \rightarrow x, x \not\rightarrow y) + \mathbf{P}(x \leftrightarrow y) \\ &= 2 \frac{d}{n} \left(1 - \frac{d}{n}\right) + \frac{d^2}{n^2} \\ &= \frac{d}{n} \left(2 - \frac{d}{n}\right) \\ &\leq \frac{2d}{n}. \end{aligned}$$

This shows that the edge distribution of  $\hat{G}$  is stochastically dominated by the edge distribution of an undirected  $(n, 2d/n)$  Erdős-Rényi graph. Events which are monotone for the deletion of edges (such as having a few number of cycles) are thus of smaller probability in  $G$  than in  $\text{ER}(n, 2d/n)$ . As a consequence, Lemmas 29 and 30 in [43] directly transfer to our setting. We hereagain use the notation  $D = 2d$ .

**LEMMA 5.6.2** (sub-exponential growth). *Let us note  $S_t(x)$  be the number of vertices in  $G$  that are exactly at (unoriented) distance  $t$  from vertex  $x$ . There are two constants  $c_0, c_1 > 0$  such that for every positive  $\lambda$  and every vertex  $x \in [n]$ , we have*

$$\mathbf{P}(S_t(x) \leq \lambda D^t \text{ for all } t) \geq 1 - c_1 e^{-c_0 \lambda}. \quad (5.6.7)$$

Moreover, for every  $p \geq 2$ , there is a constant  $c_p$  such that

$$\mathbf{E} \left[ \max_{\substack{x \in [n] \\ t \geq 1}} \left( \frac{S_t(x)}{D^t} \right)^p \right] \leq (c_p \ln(n))^p \quad (5.6.8)$$

Let us apply this result to  $|(G, x)_t| = 1 + S_1(x) + \dots + S_t(x)$  with  $\lambda := c_0^{-1} \ln(c_1 n^2)$ . With probability greater than  $1 - 1/n$ , for any  $t$  and for any  $x$ ,

$$S_t(x) \leq C \ln(n) D^t$$

where  $C := c_1^{-1}(2 + \ln(c_0))$ . On this event, one also has

$$\begin{aligned} |(G, x)_t| &= 1 + S_1(x) + \dots + S_t(x) \\ &\leq C \ln(n)(1 + D + \dots + D^t) \\ &\leq CD \ln(n) D^t. \end{aligned}$$

Similarly, with  $p = 2$  or  $4$ , we have

$$\max_{x \in [n]} |(G, x)_t|^p \leq D^{tp} t^p \max_{\substack{x \in [n] \\ s \leq t}} |S_s(x) D^{-s}|^p$$

and from there we get the two following inequalities:

$$\mathbf{E} \left[ \max_{x \in [n]} |(G, x)_t|^2 \right]^{\frac{1}{2}} \leq c \ln(n)^2 n^\kappa \quad (5.6.9)$$

$$\mathbf{E} \left[ \max_{x \in [n]} |(G, x)_t|^4 \right]^{\frac{1}{4}} \leq c \ln(n)^2 n^\kappa \quad (5.6.10)$$

for some  $c > 0$ , where we used  $t \leq \ell \leq \ln(n)$  and  $D^t \leq n^\kappa$ .

By the same comparison trick between  $G$  and undirected Erdős-Rényi, the following holds:

**LEMMA 5.6.3** (tangle-free). *Let  $\kappa$  be a constant smaller than  $1/2$ . Set  $\ell' = 2\ell = \lfloor \kappa \log_{2d}(n) \rfloor$ .*

1. *With probability going to zero, the graph  $G$  is  $\ell'$ -tangle free.*
2. *For any vertex  $x$ , the graph  $(G, x)_{\ell'}$  has no cycles with probability greater than  $1 - cD^{\ell'}/n$ .*
3. *With probability greater than  $1 - 1/\ln(n)$ , there are no more than  $c \ln(n) D^{\ell'}$  vertices  $x$  such that  $(G, x)_{\ell'}$  has a cycle.*

#### 5.6.4 Total-variation distance between the graph and the tree

We quantify the distance between neighborhoods of  $G$  and  $T$  up to the logarithmic depth  $\ell' := 2\ell$ . Let us recall some definitions. If  $\mathbb{P}_1, \mathbb{P}_2$  are two probability measures on the space  $(\Omega, \mathcal{F})$ , their total variation distance is defined as

$$d_{\text{TV}}(\mathbb{P}_1, \mathbb{P}_2) = \min_{(X_1, X_2) \in \pi(\mathbb{P}_1, \mathbb{P}_2)} \mathbf{P}(X_1 \neq X_2)$$

where  $\pi(\mathbb{P}_1, \mathbb{P}_2)$  denotes the set of *couplings between*  $\mathbb{P}_1$  and  $\mathbb{P}_2$ : pairs of random variables  $(X_1, X_2)$  such that  $X_1$  is distributed as  $\mathbb{P}_1$  and  $X_2$  is distributed as  $\mathbb{P}_2$ . It is a well-known fact (see [104]) that the total variation distance is also given by

$$d_{\text{TV}}(\mathbb{P}_1, \mathbb{P}_2) = \max_{A \in \mathcal{F}} \mathbb{P}_1(A) - \mathbb{P}_2(A).$$

We note  $\mathcal{L}(X)$  the probability distribution of a random variable  $X$ .

**PROPOSITION 5.6.4 (GW-tree approximation).** *Let  $\ell' = \lfloor \kappa \log_{2d}(n) \rfloor = 2\ell$ , with  $\kappa < 1/2$ . There is a constant  $c > 0$  such that for every vertex  $x$ ,*

$$d_{\text{TV}}(\mathcal{L}((G, x)_{\ell'}), \mathcal{L}((T, x)_{\ell'})) \leq c \frac{\ln(n)n^\kappa}{n}. \quad (5.6.11)$$

The proof of this fact is classical; one can adapt the arguments in [43] to our setting. The difference is that our graphs are directed and now have  $[n]$  possible labels, but this only brings shallow difficulties. We sketch the main ideas.

### Coupling between graphs

Let us recall the following total variation distance:

$$d_{\text{TV}}(\text{Bin}(n, \lambda/n), \text{Poi}(\lambda)) \leq \frac{\lambda}{n}. \quad (5.6.12)$$

*Proof.* For the proof, we keep using the notation  $D = 2d$ . As a consequence of Lemma 5.6.3, with a probability than  $1 - cD^{\ell'}/n$ , the graph  $(G, x)_{\ell'}$  is a directed tree and contains no more than  $c \ln(n)D^{\ell'}$  vertices. Let us note  $E_{\ell'}$  this event and perform a breadth-first exploration starting from  $x$ . This explorations finishes at a time  $\tau \leq c \ln(n)D^{\ell'}$ . At each step, we reveal a set of  $\text{Poi}(d)$  out-vertices and  $\text{Poi}(d)$  in-vertices, thus making a total-variation error smaller than  $d/n + d/n$ . By repeatedly conditioning, the total variation error made on  $E_{\ell'}$  is not greater than  $\tau \times 2d/n \leq cD \ln(n)D^{\ell'}/n$ .

This gives a coupling between the unlabelled versions of  $(G, x)_{\ell'}$  and  $(T, x)_{\ell'}$  which fails with probability at most  $(c' + 2cd) \ln(n)n^\kappa/n = c'' \ln(n)n^\kappa/n$ . We now bring the labels in.

Suppose that you have drawn your coupling between the unlabelled versions of  $(G, x)_{\ell}$  and  $(T, x)_{\ell}$ . With probability greater than  $c'' \ln(n)n^\kappa/n$ , they agree and have size smaller than  $k := c \ln(n)n^\kappa$ . We then put the labels in the Erdős-Rényi graph by drawing a uniform ordered  $k$ -set from  $[n]$ , while we put the labels on the Galton-Watson tree by simply drawing  $k$  iid uniform samples from  $[n]$ . The total variation distance between these two random multi-sets is proven above to be smaller than  $k/n$ , hence the labels agree up to an extra error term of  $c \ln(n)n^\kappa/n$ .

In the end, the coupling created this way fails with probability at most  $(c + c' + 2dc) \ln(n)n^\kappa/n$  which is exactly what is needed, up to adjusting the constants.

### Sampling with and without replacement

Let  $m$  be an integer. We define two random multisets in the following way. Put  $m$  identical balls with labels from 1 to  $m$  in a big urn. Draw the first ball and set  $p_1$  and  $q_1$  to be its label. Put the ball back in the urn. Then, suppose that one has constructed  $(p_1, \dots, p_t)$  and  $(q_1, \dots, q_t)$ . Do the following :

- Draw a ball from the urn and set  $p_{t+1}$  to be the label of this ball.
- If this label is not already one of the  $q_s$ , set it onto  $q_{t+1}$ . Else, put the ball back in the urn and draw as many balls as need to get a label which is not already one of the  $q_s$ . Define  $q_{t+1}$  to be this label.

It is clear that for every  $k \leq m$ ,  $Q_k := (q_1, \dots, q_k)$  is a uniform ordered  $k$ -set from  $[m]$ , while  $P_k := (p_1, \dots, p_k)$  is distributed as  $k$  iid uniform elements in  $[m]$ . The random variable  $(P_k, Q_k)$  is thus a coupling between those two distributions. This coupling is successful if and only if  $P_k$  has exactly  $k$  distinct elements, which happens with probability

$$\frac{(m-1) \dots (m-k-1)}{m^k} \geq \left(1 - \frac{1}{m}\right)^k \geq 1 - \frac{k}{m}.$$

The coupling thus fails with probability smaller than  $k/m$ , an upper bound for the total-variation distance between  $P_k$  and  $Q_k$ . □

Proposition (5.6.4) tells us that for every fixed  $x$ , there exists a random rooted marked tree  $(T_x, x)$  defined on the same probabilistic space as  $(G, x)$  and such that

$$\mathbf{P}((T_x, x)_{\ell'} \neq (G, x)_{\ell'}) \leq \frac{c \ln(n) n^K}{n}.$$

We consider the family  $(T_x, x)$  for  $x \in [n]$ . Let us note  $\mathcal{E}(x)$  the event ‘the coupling between  $(G, x)_{\ell'}$  and  $(T_x, x)_{\ell'}$  fails’, so that  $\mathbf{P}(\mathcal{E}(x)) \leq c \ln(n) n^K / n$ . By the Markov inequality and the union bound, with probability greater than  $1 - 1/\ln(n)^3$  there are no more than  $c \ln(n)^4 n^K$  elements  $x \in [n]$  such that the coupling fails. We now gather all the high-probability results from the preceding sections in a general theorem.

**PROPOSITION 5.6.5.** *On an event with probability greater than  $1 - 1/\ln(n)$ , the following holds, for a constant  $c > 0$ .*

1. For every  $x \in [n]$  and for every  $t$  smaller than  $2\ell$ ,

$$|(G, x)_t| \leq c \ln(n) D^t \quad |(T_x, x)_t| \leq c \ln(n) D^t. \quad (5.6.13)$$

2. There are no more than  $c \ln(n)^4 n^K$  elements  $x$  in  $[n]$  such that the coupling between  $(T_x, x)_{\ell'}$  and  $(G, x)_{\ell'}$  fails.
3. The graph  $G$  is  $\ell'$ -tangle free.

Note that on this event, there is exactly zero or one cycle in  $(G, x)_t$ , hence for any  $y$  at distance  $t$  from  $x$ , there is at most two paths of length  $t$  from  $x$  to  $y$ ; to put it another way, for any  $x \in [n]$ , we have

$$|\mathcal{P}_G(x, t)| \leq 2|(G, x)_t| \quad (5.6.14)$$

where  $\mathcal{P}_G(x, t)$  denotes the set of directed paths  $x \rightarrow x_1 \rightarrow \dots \rightarrow x_t$  of length  $t$  started at  $x$ .

## 5.7 Graph functionals

### 5.7.1 Functionals on trees: computations

We now introduce a family of functionals on  $\mathcal{G}_*$  that will be used several times in the sequel. Remember that when  $(g, o)$  is a rooted marked graph, we note  $\mathcal{P}_g(o, t)$  the number of paths in  $g$  starting from the root  $o$  and having  $t$  steps, that is,  $(t+1)$ -uples  $x = (x_0, x_1, \dots, x_t)$  with  $x_0 = o$  and  $x_s \rightarrow x_{s+1}$ .

In this section,  $\psi, \phi$  represent two vectors in  $\mathbb{R}^n$  and  $t$  is an integer. We define

$$f_{\phi, \psi, t}(g, o) = \left(\frac{n}{d}\right)^t \phi(o) \sum_{\mathcal{P}_g(o, t)} P_{i(o), t(x_1)} \cdots P_{i(x_{t-1}), t(x_t)} \times \psi(t(x_t)). \quad (5.7.1)$$

We clearly have

$$f_{\phi, \psi, t}(G, x) = \phi(x)(A^t \psi)(x).$$

We will also need another functional:

$$F_{\mu, \psi, t}(g, \mathbf{o}) = f_{\mathbf{1}, \psi, t}(g, \mathbf{o}) - \mu^{-1} f_{\mathbf{1}, \psi, t+1}(g, \mathbf{o}). \quad (5.7.2)$$

We have

$$F_{\mu, \psi, t}(G, x) = A^t \psi(x) - \frac{1}{\mu} A^{t+1} \psi(x).$$

Before moving to several computations on those observables, we state general regularity facts. We say that a function is  $t$ -local if  $f(g, \mathbf{o})$  only depends on  $(g, \mathbf{o})_t$ .

**LEMMA 5.7.1.** *The function  $f_{\phi, \psi, t}$  is  $t$ -local and satisfies*

$$|f_{\phi, \psi, t}(g, \mathbf{o})| \leq |\phi|_{\infty} |\psi|_{\infty} |\mathcal{P}_g(\mathbf{o}, t)| \left(\frac{L}{d}\right)^t \quad (5.7.3)$$

The function  $F_{\psi, t}$  is  $(t+1)$ -local and satisfies

$$|F_{\mu, \psi, t}(g, \mathbf{o})| \leq 2 |\psi|_{\infty} |\mathcal{P}_g(\mathbf{o}, t+1)| \left(\frac{L}{d}\right)^{t+1}. \quad (5.7.4)$$

*Proof.* The locality property is obvious from the definition, while for the bound it suffices to write

$$\begin{aligned} |f_{\phi, \psi, t}(g, \mathbf{o})| &\leq \left(\frac{n}{d}\right)^t |\phi|_{\infty} \sum_{\mathcal{P}_g(\mathbf{o}, t)} \left(\frac{L}{d}\right)^t |\psi|_{\infty} \\ &\leq \left(\frac{L}{d}\right)^t |\phi|_{\infty} |\psi|_{\infty} |\mathcal{P}_g(\mathbf{o}, t)|. \end{aligned}$$

It is the same thing for  $F_{\psi, t}$ . □

The following crucial theorem gathers all the computations linked with expectations or variances of those functionals when specialized on a tree  $(T_x, x)$  with the distribution described before.

**Theorem 19.** *Let  $\psi$  be any vector in  $\mathbb{R}^n$  and  $t$  be an integer. For any  $i, j \in [r]$  and for any  $\phi \in \ker(P)$ , the following identities are true.*

$$\mathbf{E}[f_{\psi, \varphi_j, t}(T_x, x)] = \psi(x) \varphi_j(x) \mu_j^t \quad (5.7.5)$$

$$\mathbf{E}[f_{\psi, \phi, t}(T_x, x)] = 0 \quad (5.7.6)$$

$$\mathbf{E}[f_{\psi, \varphi_i, t}(T_x, x) f_{\psi, \varphi_j, t}(T_x, x)] = \mu_i^t \mu_j^t \psi(x)^2 \Gamma_{i,j}^{(t)} \quad (5.7.7)$$

$$\mathbf{E}[f_{\psi, \phi, t}(T_x, x)^2] = \psi(x)^2 \frac{Q^t \phi^2(x)}{d^t}. \quad (5.7.8)$$

$$\mathbf{E}[F_{\mu_i, \varphi_i, t}(T_x, x)^2] = \frac{Q^t \varphi_i^i(x)}{d^t}. \quad (5.7.9)$$

The proof consists in using the eigenvector equation to identify specific martingales and take advantage of their properties to compute those expectations and variances. It is postponed to Section 5.9.

### 5.7.2 Functionals on graphs: concentration

This section describes concentration of sum functionals on the graph  $G$ , having the form  $\sum_{o \in [n]} f(G, o)$  where  $f : \mathcal{G}_* \rightarrow \mathbb{R}$  is any measurable function. The tools and spirit of this section are identical to [43, Section 9], but slightly adapted to our needs.

The first proposition deeply exploits the fact that  $G$  is in fact a function of iid random variables  $M_{i,j}$ ; the Efron-Stein inequality is of great help here.

**PROPOSITION 5.7.2** (Variance bound for graph functionals). *Let  $f, \gamma : \mathcal{G}_* \rightarrow \mathbb{R}$  be two  $t$ -local functions for some  $t \leq 2\ell$ , and such that  $|f(g, o)| \leq \gamma(g, o)$  and  $\gamma$  is non-decreasing by the addition of edges. Then,*

$$\text{Var} \left( \sum_{o \in [n]} f(G, o) \right) \leq (c_4 \ln(n)^2 n^\kappa)^2 \sqrt{\mathbf{E} \left[ \max_{o \in [n]} |\gamma(G, o)|^4 \right]} \quad (5.7.10)$$

*Proof.* For any  $x$ , let  $E_x$  be the set of edges going out of  $x$ . The vector  $(E_x)_{x \in [n]}$  is an iid vector. Moreover, there is a measurable function  $F$  such that

$$\sum_{o \in [n]} f(G, o) = F(E_1, \dots, E_n).$$

Let us note  $Y = F(E_1, \dots, E_n)$  this sum, and for any  $x$  let us note  $Y_x$  the same sum where  $E_x$  has been emptied:

$$Y_x = F(E_1, \dots, E_{x-1}, \emptyset, E_{x+1}, \dots, E_n).$$

Equivalently, if  $G_x$  indicates the graph  $G$  where all the edges flowing out of  $x$  have been deleted, we have

$$Y_x = \sum_{o \in [n]} f(G_x, o).$$

The Efron-Stein inequality tells us that

$$\text{Var} \left( \sum_{o \in [n]} f(G, o) \right) \leq \sum_{x \in [n]} \mathbf{E}[|Y - Y_x|^2]. \quad (5.7.11)$$

Fix  $o$  and  $x$  in  $[n]$ . The function  $f$  is  $t$ -local, hence  $f(G, o) - f(G_x, o)$  is always zero, except maybe if  $x$  is  $(G, o)_t$ , or equivalently if  $o$  is in  $(G, x)_t$ . As a consequence, we have

$$\begin{aligned} |Y - Y_x| &\leq \sum_{o \in [n]} |f(G, o) - f(G_x, o)| \\ &\leq \sum_{o \in (G, x)_t} |\gamma(G, o)| + |\gamma(G_x, o)| \\ &\leq |(G, x)_t| 2 \max_{o \in [n]} |\gamma(G, o)| \end{aligned}$$

where in the last line we used the fact that  $\gamma$  is non-decreasing by the addition of edges. By the Cauchy-Schwarz inequality,

$$\mathbf{E}[|Y - Y_x|^2] \leq \sqrt{\mathbf{E}[|(G, x)_t|^4] \mathbf{E} \left[ 16 \max_{o \in [n]} |\gamma(G, o)|^4 \right]}.$$

Finally, by (5.6.10) we have a constant  $c_4$  such that

$$\mathbf{E}[|(G, x)_t|^4] \leq c_4^4 D^{4t} t^4 \ln(n)^4. \quad (5.7.12)$$



Consequently, using  $(2d)^t \leq n^\kappa$  and  $t \leq \ln(n)/2$ , we get

$$\mathbf{E}[|Y - Y_x|^2] \leq (c_4 \ln(n)^2 n^\kappa)^2 \sqrt{\mathbf{E} \left[ \max_{o \in [n]} |\gamma(G, o)|^4 \right]}$$

which closes the proof of the proposition.  $\square$

We now turn to a general comparison principle between graph functionals and the same functional specified on trees.

**Theorem 20.** *Let  $f, \gamma: \mathcal{G}_* \rightarrow \mathbb{R}$  be two  $t$ -local functions for some  $t \leq 2\ell$ , and such that  $|f(g, o)| \leq \gamma(g, o)$  and  $\gamma$  is non-decreasing by the addition of edges. Then, with probability greater than  $1 - 1/\ln(n)^2$ , the following inequality is valid:*

$$\left| \sum_{x \in [n]} f(G, x) - \mathbf{E} \left[ \sum_{x \in [n]} f(T_x, x) \right] \right| \leq c \ln(n)^3 n^\kappa \sqrt{n} \|\gamma\| \quad (5.7.13)$$

where  $\|\gamma\|$  is defined as

$$\|\gamma\| = \left( \mathbf{E} \left[ \max_{o \in [n]} \gamma(G, o)^4 \right] \right)^{\frac{1}{4}} \vee \max_{x \in [n]} (\mathbf{E}[|\gamma(T_x, x)|^2])^{\frac{1}{2}}. \quad (5.7.14)$$

*Proof.* By the Chebyshev inequality and the variance bound in the preceding proposition, we have

$$\left| \sum_{x \in [n]} f(G, x) - \mathbf{E} \left[ \sum_{x \in [n]} f(G, x) \right] \right| \leq c_4 \ln(n)^3 n^\kappa \|\gamma\| \quad (5.7.15)$$

with probability greater than  $1 - 1/\ln(n)^2$ . We now compare the expectation in the LHS with its counterpart on the trees.

Let  $\mathcal{E}(x)$  denote the event “the coupling between  $(G, x)_{2\ell}$  and  $(T_x, x)_{2\ell}$  fails”; as our functionals are  $t$ -local, we have  $f(G, x) = f(T_x, x)$  on  $\mathcal{E}(x)$ . We had proven in Proposition 5.6.4 that  $\mathbf{P}(\mathcal{E}(x)) \leq c \ln(n) n^\kappa / n$ . Consequently, by the Cauchy-Schwarz inequality,

$$\begin{aligned} \left| \mathbf{E} \sum_{x \in [n]} f(G, x) - f(T_x, x) \right| &\leq \sum_{x \in [n]} \mathbf{E}[|f(G, x)| \mathbf{1}_{\mathcal{E}(x)}] + \mathbf{E}[|f(T_x, x)| \mathbf{1}_{\mathcal{E}(x)}] \\ &\leq \sum_{x \in [n]} \sqrt{\mathbf{P}(\mathcal{E}(x))} \left( \sqrt{\mathbf{E}[|f(G, x)|^2]} + \sqrt{\mathbf{E}[|f(T_x, x)|^2]} \right) \\ &\leq \sqrt{c \frac{\ln(n) n^\kappa}{n}} \left( \sum_{x \in [n]} \sqrt{\mathbf{E}[|f(G, x)|^2]} + \sqrt{\mathbf{E}[|f(T_x, x)|^2]} \right) \\ &\leq \sqrt{cn \ln(n) n^\kappa} \max_{x \in [n]} \left( \mathbf{E}[|f(G, x)|^4]^{\frac{1}{4}} + \sqrt{\mathbf{E}[|f(T_x, x)|^2]} \right) \\ &\leq \sqrt{cn \ln(n) n^\kappa} \|\gamma\| \end{aligned}$$

which is generously bounded by the RHS in the claim, upon adjusting the constant.  $\square$

## 5.8 Near eigenvectors: proofs

In this Section we prove Theorem 18, using the tools introduced earlier. Here is the route taken: first, we prove different propositions related with precise bounds for the entries of the matrices  $U, V$  or  $\Phi^* A^\ell \Phi$ . Often, the error terms look like

$$c \frac{b^2 \ln(n)^6 n^{3\kappa}}{\sqrt{n}} \left(\frac{L}{d}\right)^\ell$$

or small variants; the reader might from now consider that all those terms are negligible in front of all the quantities of interest. In fact, we will make sure that this is precisely true in the last section, as a consequence of our choice for  $\kappa$ .

For functionals such as  $\langle \varphi, A^t \varphi \rangle$ , the plan is simple: we justify why those functionals can be well-approximated using the toolbox from the preceding section and then we use the computations done in Theorem 19.

Bounding  $\|A^\ell P_{H^\perp}\|$  is however much more difficult and will be done through a tangle-free decomposition, in Subsection 5.8.4. Performing the high-trace method is long, so we postponed this part to Section 5.10.

### 5.8.1 Entry-wise bounds for Theorem 18

**PROPOSITION 5.8.1.** *On an event with probability greater than  $1 - 1/\ln(n)$ , there is a constant  $c$  such that for any  $i, j \in [r]$  and  $t \leq 2\ell$ , the following holds.*

$$|\langle \varphi_i, A^t \varphi_j \rangle - \mu_j^t \delta_{i,j}| \leq c \frac{b^2 \ln(n)^4 n^{2\kappa}}{\sqrt{n}} \left(\frac{L}{d}\right)^t. \quad (5.8.1)$$

*Proof.* Fix  $i, j \in [r]$  and  $t \leq 2\ell$ . Using the notation already introduced in Definition 5.7.1, we define a function  $f$  by

$$f(g, \mathfrak{o}) = \mathbf{1}_{(g, \mathfrak{o})_t \text{ has no cycles}} f_{\varphi_i, \varphi_j, t}(g, \mathfrak{o}).$$

This function is clearly  $t$ -local — see Lemma 5.7.1 — and if  $(g, \mathfrak{o})_t$  has no cycles and is tangle-free, then from (5.7.3), (H:incoherent), and (5.6.14),

$$|f(g, \mathfrak{o})| \leq |\varphi_i|_\infty |\varphi_j|_\infty \left(\frac{L}{d}\right)^t |\mathcal{P}_g(\mathfrak{o}, t)| \quad (5.8.2)$$

$$\leq \frac{2b^2}{n} \left(\frac{L}{d}\right)^t |(g, \mathfrak{o})_t| := \gamma(g, \mathfrak{o}). \quad (5.8.3)$$

It is clear that this function  $\gamma$  is non-decreasing by the addition of edges.

Let us we note  $\mathcal{V}_n$  the set of vertices such that  $(G, x)_t$  is not tree like; in particular, if  $x$  is in  $\mathcal{V}_n$ , then the coupling between  $(G, x)_{\ell'}^t$  and  $(T_x, x)_{\ell'}^t$  described in Section 5.6.4 fails. On the event of Proposition 5.6.5, we have  $|\mathcal{V}_n| \leq c \ln(n)^4 n^\kappa$  and the graph is  $t$ -tangle free, hence

$$\left| \langle \varphi_i, A^t \varphi_j \rangle - \sum_{x \in [n]} f(G, x) \right| = \left| \sum_{x \in \mathcal{V}_n} \varphi_i(x) (A^t \varphi_j)(x) \right| \leq \frac{2b^2 c}{n} \left(\frac{L}{d}\right)^t \ln(n)^4 n^\kappa.$$

We now apply the concentration result in Theorem 20 to the function  $f$ . By the definition of  $\gamma$  and by the growth bounds (5.6.5)-(5.6.10) we have

$$\begin{aligned} \|\mathcal{V}\| &= \frac{2b^2}{n} \left(\frac{L}{d}\right)^t \left( \mathbf{E}[\max_{x \in [n]} |(G, x)_t|^4]^{\frac{1}{4}} \vee \max_{x \in [n]} \mathbf{E}[|(T_x, x)_t|^2]^{\frac{1}{2}} \right) \\ &\leq c \frac{b^2}{n} \left(\frac{L}{d}\right)^t \ln(n)^2 n^\kappa \end{aligned}$$

for some constant  $c$ . The error bound in Theorem 20 is thus

$$\frac{cb^2 \ln(n)^5 n^{2\kappa}}{\sqrt{n}} \left(\frac{L}{d}\right)^t.$$

Moreover, as computed in Theorem 19-(5.7.5), we have

$$\mathbf{E} \left[ \sum_{x \in [n]} f(T_x, x) \right] = \mathbf{E} \left[ \sum_{x \in [n]} f_{\varphi_i, \varphi_j, t}(T_x, x) \right] = \mu_i^t \delta_{i,j}.$$

When gathering all the preceding bounds and adjusting  $c$ , we get

$$|\langle \varphi_i, A^t \varphi_j \rangle - \mu_j^t \delta_{i,j}| \leq c \frac{b^2 \ln(n)^5 n^{2\kappa}}{\sqrt{n}} \left(\frac{L}{d}\right)^t.$$

□

**PROPOSITION 5.8.2.** *On an event with probability greater than  $1 - 2/\ln(n)$ , there is a constant  $c$  such that for any  $i, j \in [r]$  and  $t \leq 2\ell$ , the following holds.*

$$\left| \langle A^t \varphi_i, A^t \varphi_j \rangle - \mu_i^t \mu_j^t \Gamma_{i,j}^{(t)} \right| \leq \frac{cb^2 \ln(n)^5 n^{3\kappa}}{\sqrt{n}} \left(\frac{L}{d}\right)^{2t} \quad (5.8.4)$$

$$\left| \langle A^t \varphi_i, A^t \varphi_j \rangle - \mu_i^t \mu_j^t \Gamma_{i,j}^{(t)} \right| \leq \frac{cb^2 \ln(n)^5 n^{3\kappa}}{\sqrt{n}} \left(\frac{L}{d}\right)^{2t}. \quad (5.8.5)$$

Note that the second inequality is exactly the same as the first, except that we now look at  $A^*$ . However, since  $A^*$  and  $A$  are identical in distribution, one only has to prove the first one.

*Proof.* The proof is the same as for the proposition just before so we do not linger on details. Fix  $i, j \in [r]$  and  $t \leq 2\ell$ . The right function here is  $f$  defined by

$$f(g, \mathfrak{o}) = \mathbf{1}_{(g, \mathfrak{o})_t \text{ has no cycles}} f_{\mathbf{1}, \varphi_i, t}(g, \mathfrak{o}) f_{\mathbf{1}, \varphi_j, t}(g, \mathfrak{o}).$$

This function is clearly  $t$ -local and if  $(g, \mathfrak{o})_t$  has no cycles and is tangle-free,

$$|f(g, \mathfrak{o})| \leq |\varphi_i|_\infty |\varphi_j|_\infty \left(\frac{L}{d}\right)^t |\mathcal{P}_g(\mathfrak{o}, t)|^2 \quad (5.8.6)$$

$$\leq \frac{2b^2}{n} \left(\frac{L}{d}\right)^{2t} |(g, \mathfrak{o})_t|^2 := \gamma(g, \mathfrak{o}). \quad (5.8.7)$$

Let us we note  $\mathcal{V}_n$  the set of vertices such that  $(G, x)_t$  is not tree like. With probability greater than  $1 - 1/\ln(n)$  we have  $|\mathcal{V}_n| \leq c \ln(n)^4 n^\kappa$  and the graph is  $t$ -tangle free, hence

$$\left| \langle A^t \varphi_i, A^t \varphi_j \rangle - \sum_{x \in [n]} f(G, x) \right| = \left| \sum_{x \in \mathcal{V}_n} (A^t \varphi_i)(x) (A^t \varphi_j)(x) \right| \leq \frac{2b^2 c}{n} \left(\frac{L}{d}\right)^t \ln(n)^5 n^{2\kappa}.$$

We now apply Theorem 20. By the definition of  $\gamma$  and by (5.6.5)-(5.6.10) we have

$$\begin{aligned} \|\gamma\| &= \frac{2b^2}{n} \left(\frac{L}{d}\right)^{2t} \left( \mathbf{E}[\max_{x \in [n]} |(G, x)_t|^8]^{1/4} \vee \max_{x \in [n]} \mathbf{E}[|(T_x, x)_t|^4]^{1/2} \right) \\ &\leq c \frac{b^2}{n} \left(\frac{L}{d}\right)^{2t} \ln(n)^2 n^{2\kappa} \end{aligned}$$

for some  $c$ . The error bound in Theorem 20 is thus

$$\frac{cb^2 \ln(n)^5 n^{3\kappa}}{\sqrt{n}} \left(\frac{L}{d}\right)^{2t}.$$

Moreover, as computed in Theorem 19-(5.7.7), we have

$$\mathbf{E} \left[ \sum_{x \in [n]} f(T_x, x) \right] = \mathbf{E} \left[ \sum_{x \in [n]} f_{1, \varphi_i, t}(T_x, x) f_{1, \varphi_j, t}(T_x, x) \right] = \mu_i^t \mu_j^t \Gamma_{i,j}^{(t)}.$$

When gathering all the preceding bounds we get

$$\left| \langle \varphi_i, A^t \varphi_j \rangle - \mu_i^t \mu_j^t \Gamma_{i,j}^{(t)} \right| \leq \frac{cb^2 \ln(n)^5 n^{3\kappa}}{\sqrt{n}} \left(\frac{L}{d}\right)^{2t}.$$

□

### 5.8.2 Rough bound on $\|A^t\|$

We prove (5.4.13). We pick any unit vector  $w \in \mathbb{R}^n$  and we place ourselves on the good event described in Subsection 5.6.3 and Proposition 5.6.5. Using  $\|P\|_\infty \leq L/n$  and the Cauchy-Schwarz inequality, we get

$$\begin{aligned} |A^t w|^2 &= \left| \left(\frac{n}{d}\right)^{2t} \sum_{x \in [n]} \left( \sum_{x_0=x \rightarrow x_1 \rightarrow \dots \rightarrow x_t} \prod_{s=0}^{t-1} A_{x_s, x_{s+1}} w(x_t) \right)^2 \right| \\ &\leq \left(\frac{n}{d}\right)^{2t} \left(\frac{L}{n}\right)^{2t} \sum_{x \in [n]} \left( \sum_{x_t \in \partial B^+(x, t)} |w(x_t)| \right)^2 \\ &\leq \left(\frac{L}{d}\right)^{2t} \sum_{x \in [n]} |\partial B^+(x, t)|^2 \sum_{x_t \in \partial B^+(x, t)} |w(x_t)|^2 \end{aligned}$$

On the event described in Proposition 5.6.5,  $|\partial B^+(x, t)| \leq |G(x, t)|$  is smaller than  $c \ln(n) n^\kappa$  for every  $x \in [n]$ . As a consequence we get

$$\begin{aligned} |A^t w|^2 &\leq c \ln(n)^2 n^{2\kappa} \left(\frac{L}{d}\right)^{2t} \sum_{x \in [n]} \sum_{x_t \in \partial B^+(x, t)} |w(x_t)|^2 \\ &\leq c \ln(n)^2 n^{2\kappa} \left(\frac{L}{d}\right)^{2t} \sum_{x_t \in [n]} |w(x_t)|^2 \sum_{x \in \partial B^-(x_t, t)} 1 \\ &\leq c \ln(n)^4 n^{4\kappa} \left(\frac{L}{d}\right)^{2t} \end{aligned}$$

which clearly shows that  $\|A^t\| \leq c \ln(n)^2 n^{2\kappa} (L/d)^{2t}$  for some adjusted constant  $c$ , the requested bound in (5.4.13).

### 5.8.3 Control over the growth of a process

**PROPOSITION 5.8.3.** *On an event with probability greater than  $1 - 1/\ln(n)$ , one has for any  $t \leq \ell$ , for any  $w \in H^\perp$  and for any  $i \in [r_0]$  the following bound:*

$$\left| \langle (A^*)^t \varphi_i, w \rangle \right| \leq cb^4 r \ln(n) + c \frac{b^2 \ln(n)^6 n^{3\kappa}}{\sqrt{n}} \left(\frac{L}{d}\right)^\ell \vartheta^t \quad (5.8.8)$$

*Proof.* We follow the usual strategy. First, we note that by the mere definition of  $H$ , when  $w \in H^\perp$  we have  $\langle (A^*)^\ell \varphi_i, w \rangle = 0$ . Consequently,

$$\mu_i^{-t} \langle (A^*)^t \varphi_i, w \rangle = \mu_i^{-t} \langle (A^*)^t \varphi_i, w \rangle - \mu_i^{-\ell} \langle (A^*)^\ell \varphi_i, w \rangle$$

and from a telescopic sum we get

$$\begin{aligned} |\mu_i^{-t} \langle (A^*)^t \varphi_i, w \rangle| &= \left| \sum_{s=t}^{\ell-1} \mu_i^{-s} \langle (A^*)^s \varphi_i, w \rangle - \mu_i^{-(s+1)} \langle (A^*)^{s+1} \varphi_i, w \rangle \right| \\ &\leq \sum_{s=t}^{\ell-1} \mu_i^{-s} \left| \langle (A^*)^s \varphi_i, w \rangle - \frac{1}{\mu_i} \langle (A^*)^{s+1} \varphi_i, w \rangle \right| \\ &\leq \sum_{s=t}^{\ell-1} \mu_i^{-s} \left| (A^*)^s \varphi_i - \frac{1}{\mu_i} (A^*)^{s+1} \varphi_i \right| \end{aligned} \quad (5.8.9)$$

where in the last line we used the Cauchy-Schwarz inequality. Let us fix  $s \in [\ell]$ . We have

$$\left| (A^*)^s \varphi_i - \frac{1}{\mu_i} (A^*)^{s+1} \varphi_i \right|^2 = \sum_{x \in [n]} F_{\mu_i, \varphi_i, s}(G, x)^2$$

where  $F_{\mu_i, \varphi_i, s}$  was the functional defined in (5.7.2). By the computation (5.7.9) from Theorem 19, we have for any  $x \in [n]$

$$\mathbf{E}[F_{\mu_i, \varphi_i, s}(T_x, x)^2] = \frac{Q^s \varphi^{i,i}(x)}{d^s}$$

so summing over  $x$  yields

$$\mathbf{E} \left[ \sum_{x \in [n]} F_{\mu_i, \varphi_i, s}(T_x, x)^2 \right] = \frac{\langle \mathbf{1}, Q^s \varphi^{i,i} \rangle}{d^s}.$$

We are going to use the concentration bound from Theorem 20. In preparation, we need a majorant  $\gamma$ ; it is given by (5.7.4) from Lemma 5.7.1. More precisely, we have

$$F_{\mu_i, \varphi_i, s}(g, \circ) \leq |\varphi_i|_\infty |\mathcal{P}_g(\circ, s+1)| (L/d)^s := \gamma(g, \circ).$$

Applying Theorem 20 yields

$$\left| \left| (A^*)^s \varphi_i - \frac{1}{\mu_i} (A^*)^{s+1} \varphi_i \right|^2 - \frac{\langle \mathbf{1}, Q^s \varphi^{i,i} \rangle}{d^s} \right| \leq c \ln(n)^3 n^\kappa \sqrt{n} \|\gamma\|^2. \quad (5.8.10)$$

The upper bound on  $\|\gamma\|$  is done as usual from the growth bounds in (5.6.5)-(5.6.10), and they yield a constant  $c$  such that

$$\|\gamma\| \leq c \frac{b^2 \ln(n)^2 n^{2\kappa}}{n} \left( \frac{L}{d} \right)^{s+1}.$$

From this and (5.8.10) we find that on a high-probability event, for any  $s$ , we have

$$\left| \left| (A^*)^s \varphi_i - \frac{1}{\mu_i} (A^*)^{s+1} \varphi_i \right|^2 - \frac{\langle \mathbf{1}, Q^s \varphi^{i,i} \rangle}{d^s} \right| \leq c \frac{b^2 \ln(n)^5 n^{3\kappa}}{\sqrt{n}} \left( \frac{L}{d} \right)^{s+1}$$

and summing over all  $s$  between  $t$  and  $\ell$  as in (5.8.9) yields

$$\mu_i^{-t} \langle (A^*)^t \varphi_i, w \rangle \leq \sum_{s=t}^{\ell-1} \sqrt{\frac{\langle \mathbf{1}, Q^s \varphi^{i,i} \rangle}{(\mu_i^2 d)^s}} + c \frac{b^2 \ln(n)^6 n^{3\kappa}}{\sqrt{n}} \left( \frac{L}{d} \right)^\ell$$

where we used  $\sqrt{1+u} \leq 1 + 0.5u$  and we adjusted the constant  $c$ . Then, we use (5.5.7) which tells that  $\langle \mathbf{1}, Q^s \varphi^{i,i} \rangle \leq b^8 r^2 \rho^s$ . We get

$$\begin{aligned} |\langle (A^*)^t \varphi_i, w \rangle| &\leq \mu_i^t \sum_{s=t}^{\ell} \frac{b^4 r \vartheta^s}{\mu_i^s} + c \frac{b^2 \ln(n)^6 n^{3\kappa}}{\sqrt{n}} \left(\frac{L}{d}\right)^\ell \\ &\leq b^4 r \ln(n) \vartheta^t + c \frac{b^2 \ln(n)^6 n^{3\kappa}}{\sqrt{n}} \left(\frac{L}{d}\right)^\ell \end{aligned}$$

where we used the fact that  $\mu_i > \vartheta$ , hence every term in the sum is dominated by the first term, and there are less than  $\ell \leq \ln(n)$  terms inside the sum.  $\square$

#### 5.8.4 Norm of the remainder matrix: proof of (5.4.11)

Let us recall that  $H = \text{span}(v_1, \dots, v_{r_0}) = \text{im}(V)$ . Our goal in this section is to prove the last inequality (5.4.11) in Theorem 17, namely that with high probability,

$$\|A^\ell P_{H^\perp}\| = \max_{\substack{w \in H^\perp, \\ |w|=1}} |A^\ell w| \leq C \vartheta^\ell$$

with  $C$  some constant depending polynomially on the parameters.

#### The tangle-free decomposition

We notice that if the graph  $G$  is  $\ell$ -tangle free then  $A^\ell = A^{(\ell)}$  where

$$A_{x,y}^{(\ell)} = \left(\frac{n}{d}\right)^\ell \sum_{x \in F_{x,y}^\ell} \prod_{t=1}^{\ell} P_{x_{t-1}x_t} M_{x_{t-1}x_t},$$

and the sum runs over the set  $F_{x,y}^\ell$  of all paths  $(x_0, \dots, x_\ell)$  such that  $x_0 = x$ ,  $x_\ell = y$  and the graph of the path is tangle-free — we recall that tangle-free means that there are no more than one cycle, see the definitions in Subsection 5.6.1 on page 128. More generally,  $F^t$  denotes the set of all tangle-free paths of length  $t$ , whatever their endpoints. We also define the matrices  $\underline{M}$  and  $\underline{A}^{(\ell)}$  by  $A^{(0)} = \underline{A}^{(0)} = I_n$ , and

$$\begin{aligned} \underline{M}_{x,y} &= M_{x,y} - \frac{d}{n} \\ \underline{A}_{x,y}^{(\ell)} &= \left(\frac{n}{d}\right)^\ell \sum_{F_{x,y}^\ell} \prod_{t=1}^{\ell} P_{x_{t-1}x_t} \underline{M}_{x_{t-1}x_t}. \end{aligned} \quad (5.8.11)$$

We use the convention that the product over an empty set is 1. Then we may write for any  $a, b \in \mathbb{R}^\ell$ ,

$$\prod_{t=1}^{\ell} a_t = \prod_{t=1}^{\ell} b_t + \sum_{k=1}^{\ell} \left( \prod_{t=1}^{k-1} b_t \right) (a_k - b_k) \left( \prod_{t=k+1}^{\ell} a_t \right).$$

We thus get

$$A_{x,y}^{(\ell)} = \underline{A}_{x,y}^{(\ell)} + \sum_{k=1}^{\ell} \left(\frac{n}{d}\right)^\ell \sum_{F_{x,y}^\ell} \prod_{t=1}^{k-1} P_{x_{t-1}x_t} \underline{M}_{x_{t-1}x_t} \left(\frac{d}{n} P_{x_{k-1}x_k}\right) \prod_{t=k+1}^{\ell} P_{x_{t-1}x_t} M_{x_{t-1}x_t}.$$

This can then be rewritten as the following identity in  $\mathcal{M}_{n,n}(\mathbb{R})$ :

$$A^{(\ell)} = \underline{A}^{(\ell)} + \sum_{k=1}^{\ell} \underline{A}^{(k-1)} P A^{(\ell-k)} - \sum_{k=1}^{\ell} R_k^{(\ell)},$$

where

$$(R_k^{(\ell)})_{x,y} = \left(\frac{n}{d}\right)^{\ell-1} \sum_{(x_0, \dots, x_\ell) \in \mathcal{T}_{x,y}^{k,\ell}} \prod_{t=1}^{k-1} P_{x_{t-1}x_t} M_{x_{t-1}x_t} P_{x_{k-1}x_k} \prod_{t=k+1}^{\ell} P_{x_{t-1}x_t} M_{x_{t-1}x_t}$$

where the sum is over all ‘paths’  $(x_0, \dots, x_\ell)$  such that  $(x_0, \dots, x_{k-1}) \in F^{k-1}$ ,  $(x_k, \dots, x_\ell) \in F^{\ell-k}$  but  $(x_0, \dots, x_\ell)$  is not in  $F^\ell$ .

We now use the spectral decomposition  $P = \mu_1 \varphi_1 \varphi_1^* + \dots + \mu_r \varphi_r \varphi_r^*$ . For any unit vector  $w$ , we have

$$\underline{A}^{(k-1)} P A^{(\ell-k)} w = \underline{A}^{(k-1)} \sum_{j=1}^r \mu_j \varphi_j \langle \varphi_j, A^{(\ell-k)} w \rangle.$$

Hence, from the orthogonality of the  $\varphi_j$ ’s,

$$\begin{aligned} |\underline{A}^{(k-1)} P A^{(\ell-k)} w| &\leq \|\underline{A}^{(k-1)}\| \left| \sum_{j=1}^r \mu_j \varphi_j \langle \varphi_j, A^{(\ell-k)} w \rangle \right| \\ &= \|\underline{A}^{(k-1)}\| \sqrt{\sum_{j=1}^r \mu_j^2 \langle \varphi_j, A^{(\ell-k)} w \rangle^2}. \end{aligned} \quad (5.8.12)$$

We have proven in (5.8.8) that on a high-probability event, the following holds for any  $t$  smaller than  $\ell$  and  $i \in [r_0]$  and  $w \in H^\perp$ :

$$|\langle \varphi_i, A^t w \rangle| \leq c b^4 r \ln(n) \vartheta^t.$$

On the other hand, for  $i > r_0$ , we had proven in (5.8.5) that  $|(A^*)^t \varphi_i|^2 \leq \mu_i^{2t} \Gamma_{i,i}^{(t)}$ . However, when  $\mu_i > \vartheta$ , the sum defining  $\Gamma_{i,i}^{(t)}$  is smaller than  $t b^8 r^2 \rho^t / (\mu_i^2 d)^t$ , so we have  $|(A^*)^t \varphi_i|^2 \leq c b^8 r^2 \ln(n) \vartheta^t$ . As a consequence,

$$|\langle \varphi_i, A^t w \rangle| \leq |w| |(A^*)^t \varphi_i|^2 \leq c b^4 r \ln(n) \vartheta^t.$$

On the union of those events, the whole square root is thus bounded as follows (the constant  $c$  is adjusted):

$$\begin{aligned} \sqrt{\sum_{j=1}^r \mu_j^2 \langle \varphi_j, A^{(\ell-k)} w \rangle^2} &\leq \sqrt{c b^2 r \sum_{j=1}^{r_0} \mu_j^2 (\vartheta^{\ell-t})^2} \\ &\leq c \ln(n) b^2 r \vartheta^{t-\ell}. \end{aligned}$$

We can redefine the constant  $c$  and we get the following proposition.

**LEMMA 5.8.4.** *On an event with probability at least  $1 - 1/\ln(n)$ , one has*

$$\|A^\ell P_{H^\perp}\| \leq \|\underline{A}^{(\ell)}\| + c \ln(n) b^2 r \sum_{k=1}^{\ell} \|\underline{A}^{(k-1)}\| \vartheta^{\ell-t} + \sum_{k=1}^{\ell} \|R_k^{(\ell)}\|. \quad (5.8.13)$$

We now need bounds on  $\|\underline{A}^{(k-1)}\|, \|R_k^{(\ell)}\|$ .

**PROPOSITION 5.8.5.** *There exists a universal constant  $C$  such that if  $n \geq C \vee (K r b^2)^6$ , with probability at least  $1 - 1/n$ , the following holds for any  $k \in \{1, \dots, \ell\}$ :*

$$\|\underline{A}^{(k)}\| \leq 2 \ln(n)^{15} (r b^2 L^2 / \rho)^2 (\vartheta \vee \vartheta_0)^k. \quad (5.8.14)$$

$$\|R_k^{(\ell)}\| \leq 2 \frac{d}{n} \ln(n)^{24} L^\ell. \quad (5.8.15)$$

The proof of this proposition relies on a high-trace method. It is postponed to Section 5.10. We can now conclude: with high probability, for some constant  $c > 0$ , we have

$$\begin{aligned}
 \|A^\ell \mathbf{P}_{H^\perp}\| &\leq \|A^{(\ell)}\| + c \ln(n) b^2 r \sum_{k=1}^{\ell} \|A^{(k-1)}\| \vartheta^{\ell-k} + \sum_{k=1}^{\ell} \|R_k^{(\ell)}\| \\
 &\leq \ln(n)^{15} C \sum_{k=0}^{\ell} [\vartheta \vee \vartheta_0]^\ell + \frac{2d\ell \ln(n)^{24}}{n} L^\ell \\
 &\leq \ln(n)^{15} C_0 \ell [\vartheta \vee \vartheta_0]^\ell + \frac{2d\ell \ln(n)^{24}}{n} L^\ell \\
 &\leq C \ln(n)^{16} [\vartheta \vee \vartheta_0]^\ell
 \end{aligned} \tag{5.8.16}$$

with  $C_0 := cb^2 r (rb^2 L^2 / \rho)^2$  in the third line. To get the last line, all we have to do is to note that  $L^\ell / n = (L/d)^\ell d^\ell / n \leq \vartheta_0^\ell d^\ell / n \leq \vartheta_0^\ell$ . The constant  $C$  in this last line is taken to be

$$C = C(d, L, b, r) = cb^2 r (rb^2 L^2 / \rho)^2 \tag{5.8.17}$$

with  $c$  some absolute constant.

### 5.8.5 Proof of Theorem 18

We gather the events and bounds from the last propositions in a fashionable and easy-to-read summary, working out the error terms and presenting them in a way which keeps track of dependencies with the parameters in the class  $\mathcal{H}$ . The union of the three events in each of the three preceding propositions has probability greater than  $1 - 4/\ln(n)$ . The proof mainly relies on checking that the error terms are as claimed in the statement of Theorem 18 and this is more or less the same thing for all the entries, so we only do the first one.

For any  $\alpha > 0$  we have  $\ln(x) \leq x^\alpha$ , hence for every  $n$  we have the lavish domination  $\ln(n)^{20} \leq n^{0.001}$ .

We recall that  $\kappa$  had been defined in (5.1.8) by

$$\kappa = \frac{0.249}{3 + \frac{\ln(b)}{\ln(2d)}}. \tag{5.8.18}$$

The  $(i, j)$ -entry of the matrix  $\Phi^* A^\ell \Phi$  is precisely  $\langle \varphi_i, A^\ell \varphi_j \rangle$ , hence by (5.8.1) we have (with high probability):

$$\|\Phi^* A^\ell \Phi - D^\ell\| \leq r_0 \|\Phi^* A^\ell \Phi - D^\ell\|_\infty \leq c \frac{r_0 b^2 \ln(n)^5 n^{2\kappa}}{\sqrt{n}} \left(\frac{L}{d}\right)^\ell \leq \frac{c r_0 b^2 n^{2\kappa+0.001}}{\sqrt{n}} \vartheta_0^\ell.$$

We write this two error term in the form  $r_0 b^2 \delta$ . Our goal is to prove that  $\delta$  is smaller than  $n^{-0.25}$ ,  $\vartheta^\ell$  and  $\vartheta_0^\ell$ , as in (5.4.12). Note that by the choice of  $\kappa$  we have  $2\kappa + 0.001 - 0.5 < 0$ , so if  $n$  is large enough it is clear that  $cn^{2\kappa+0.001-0.5} < 1$  hence  $\delta < \vartheta_0^\ell$  as requested. On the other hand (we omit integer parts):

$$\vartheta^\ell = n^{\frac{\kappa}{4\ln(2d)} \ln(\rho/d)} \leq n^{\frac{\kappa}{4\ln(2d)} \ln(b^4/d)}$$

where we used  $\rho \leq L^2 \leq b^4$ . Simple rearrangements using our choice of  $\kappa$  now show that  $\delta \leq \vartheta^\ell$  and similar computations show that  $\delta$  is also smaller than  $n^{-0.25}$ , as requested.



## 5.9 Eigenwaves on Galton-Watson trees

This section carries out the details of the expectation and variance computation in Theorem 19. Let us recall the notation, and especially the definition (5.7.1) of the functionals: if  $\psi, \phi$  are two vectors in  $\mathbb{R}^n$  and  $t$  is an integer, then

$$f_{\phi, \psi, t}(g, \circ) = \left(\frac{n}{d}\right)^t \phi(\iota(\circ)) \sum_{\mathcal{P}_g(\circ, t)} P_{\iota(\circ), \iota(x_1)} \cdots P_{\iota(x_{t-1}), \iota(x_t)} \times \psi(\iota(x_t)).$$

### 5.9.1 An elementary computation on Poisson sums.

Let  $N$  be a  $\text{Poi}(d)$  random variable, and let  $(X_i), (Y_i)$  two iid sequences of random variables, both being independent from  $N$ ; we suppose that  $X_i$  is independent of  $Y_j$  for  $i \neq j$ , but there might be a nontrivial dependance between  $X_i$  and  $Y_i$ . Let us note

$$A = \sum_{i=1}^N X_i \quad B = \sum_{i=1}^N Y_i.$$

The following (classical) identity will be crucial in the next sections. For convenience, we provide a proof.

$$\text{Cov}(A, B) = d\mathbf{E}[XY]. \quad (5.9.1)$$

*Proof of (5.9.1).* Primary computations shows that  $\mathbf{E}[A] = d\mathbf{E}[X]$  and  $\mathbf{E}[B] = d\mathbf{E}[Y]$ , hence  $\text{Cov}(A, B) = \mathbf{E}[AB] - \mathbf{E}[A]\mathbf{E}[B] = \mathbf{E}[AB] - d^2\mathbf{E}[X]\mathbf{E}[Y]$ . The first term  $\mathbf{E}[AB]$  is thus equal to

$$\begin{aligned} \mathbf{E}[AB] &= \mathbf{E}\left[\sum_{i=1}^N \sum_{j=1}^N X_i Y_j\right] \\ &= \sum_{k=0}^{\infty} \frac{e^{-d} d^k}{k!} \mathbf{E}\left[\sum_{i=1}^k \sum_{j=1}^k X_i Y_j\right] \\ &= \sum_{k=0}^{\infty} \frac{e^{-d} d^k}{k!} (k\mathbf{E}[XY] + k(k-1)\mathbf{E}[X]\mathbf{E}[Y]) \\ &= \mathbf{E}[XY]\mathbf{E}[N] + \mathbf{E}[X]\mathbf{E}[Y]\mathbf{E}[N(N-1)]. \end{aligned}$$

We have  $\mathbf{E}[N(N-1)] = d^2$ , hence (5.9.1) holds true.  $\square$

Identity (5.9.1) will be used many times in the following context. Fix one vertex  $x \in [n]$  and suppose that  $X = P_{x,U}\varphi_i(U)$  and  $Y = P_{x,U}\varphi_j(U)$  with  $U \sim \text{Unif}[n]$ . By the eigenvector equation  $P\varphi_k = \mu_k\varphi_k$ , we have  $\mathbf{E}[X] = (\mu_i d \varphi_i(x))/n$  and  $\mathbf{E}[Y] = (\mu_j d \varphi_j(x))/n$ , hence in this case

$$\mathbf{E}[XY] = \frac{1}{n} \sum_{y \in [n]} P_{x,y}^2 \varphi_i(y) \varphi_j(y) = \frac{1}{n^2} (Q\varphi^{i,j})(x). \quad (5.9.2)$$

These identities will be used later in variance computations.

### 5.9.2 Proof of a martingale property

Let  $x$  be a fixed element in  $[n]$  and let  $(T_x, x)$  be the random rooted marked tree described in Section (5.6.1) and let  $\mathcal{F}_t$  be the sigma-algebra generated by  $(T_x, x)_t$ ; from now on we will use the filtration  $\mathcal{F} = (\mathcal{F}_t)_{t \geq 0}$ . The key observation for this whole section is that the process  $t \mapsto \mu_k^{-t} f_{\phi, \varphi_k, t}(T_x, x)$  is indeed an  $\mathcal{F}$ -martingale.

**LEMMA 5.9.1.** *Let  $\phi$  be any vector and let  $\varphi_i$  be an eigenvector of  $P$  associated with the nonzero eigenvalue  $\mu_i$ . Then, the discrete-time stochastic process*

$$Z_t := \frac{1}{\mu_i^t} f_{\phi, \varphi_i, t}(T_x, x)$$

is an  $\mathcal{F}$ -martingale.

From now on, the conditional expectation with respect to the sigma-algebra  $\mathcal{F}_t$  will be noted  $\mathbf{E}_t$  instead of  $\mathbf{E}[\cdot | \mathcal{F}_t]$ .

*Proof.* It is clear that  $f_{\phi, \varphi_i, t}(T_x, x) = \phi(x) f_{\mathbf{1}, \varphi_i, t}(T_x, x)$ , hence it is sufficient to prove the martingale property for  $Z_t = \mu_i^{-t} f_{\mathbf{1}, \varphi_i, t}(T_x, x)$ .

Let us fix an integer  $t$ . Then, upon factorizing up to depth  $t$  we have

$$Z_{t+1} - Z_t = \left( \frac{n}{d\mu_i} \right)^{t+1} \sum_{d^+(0, x) = t} \prod_{s=1}^t P_{t(x_{s-1}), t(x_s)} \left( \sum_{x_t \rightarrow y} P_{t(x_t), t(y)} \varphi_i(t(y)) - \frac{d\mu_i}{n} \varphi_i(t(x_t)) \right).$$

Let us note  $\Delta_t = Z_{t+1} - Z_t$  the martingale increment. Then,

$$\mathbf{E}_t[\Delta_t] = \left( \frac{n}{d\mu_i} \right)^{t+1} \sum_{d^+(0, x) = t} \prod_{s=1}^t P_{t(x_{s-1}), t(x_s)} \mathbf{E}_t \left[ \sum_{x_t \rightarrow y} P_{t(x_t), t(y)} \varphi_i(t(y)) - \frac{d\mu_i}{n} \varphi_i(t(x_t)) \right].$$

Let  $X_1, X_2, \dots$  be iid random variables with the following distribution (conditionally on  $\mathcal{F}_t$ ):

$$\mathbf{P}_t(X = P_{t(x_t), z} \varphi_i(z)) = \frac{1}{n} \quad \text{for each } z \in [n].$$

In other words, conditionally on  $\mathcal{F}_t$ , the rv's  $X_s$  are iid samples with distribution  $P_{t(x_t), U} \varphi_i(U)$  with  $U \sim \text{Unif}[n]$ , just as in the end of the preceding paragraph.

It is clear that for every children  $y$  of  $x_t$ , the random variable  $P_{t(x_t), t(y)} \varphi_i(t(y))$  has this distribution, and as already noted,

$$\mathbf{E}[X] = \frac{1}{n} \sum_{z \in [n]} P_{t(x_t), z} \varphi_i(z) = \frac{1}{n} (P\varphi_i)(t(x_t)) = \frac{\mu_i}{n} \varphi_i(t(x_t)).$$

The number  $N$  of children of  $x_t$  has a  $\text{Poi}(d)$  distribution, and is independent of  $\mathcal{F}_t$ , hence

$$\begin{aligned} \mathbf{E}_t \left[ \sum_{x_t \rightarrow y} P_{t(x_t), t(y)} \varphi_i(t(y)) - \frac{d\mu_i}{n} \varphi_i(t(x_t)) \right] &= \mathbf{E}_t \left[ \sum_{s=1}^N X_s \right] - \frac{d\mu_i}{n} \varphi_i(t(x_t)) \\ &= d\mathbf{E}[X] - \frac{d\mu_i}{n} \varphi_i(t(x_t)) \\ &= 0. \end{aligned}$$

We have  $\mathbf{E}_t[\Delta_t] = 0$  and the martingale property for  $Z_t$  is true. □

### 5.9.3 Proof of (5.7.5)-(5.7.6)

The proof of these two identities is straightforward. Indeed, the martingale property for  $Z_t = \mu_i^{-t} f_{\psi, \varphi_i, t}(T_x, x)$  shows that

$$\frac{\mathbf{E}[f_{\psi, \varphi_i, t}(T_x, x)]}{\mu_i^t} = \mathbf{E}[Z_t] = \mathbf{E}[Z_0] = \psi(x)\varphi_i(x)$$

which is exactly (5.7.5). For (5.7.6), one only has to note that the computations in the last section also show that if  $P\varphi = \mu\varphi$  for any  $\mu$  (possibly zero), then for any  $t > 0$  we have

$$\mathbf{E}[f_{\psi, \varphi, t}(T_x, x)] = \psi(x)\varphi(x)\mu^t$$

hence (5.7.6), when  $\mu$  is zero.

### 5.9.4 Proof of (5.7.7)-(5.7.9)

We fix  $i, j$  in  $[r]$  and  $x$  in  $[n]$  for the rest of the proof. Clearly, it is enough to do the computations with  $\psi = \mathbf{1}$ . We set

$$Z_t^i = \frac{f_{\mathbf{1}, \varphi_i, t}(T_x, x)}{\mu_i^t}$$

and

$$\Delta_t = \mathbf{E}_t[(Z_{t+1}^i - Z_t^i)(Z_{t+1}^j - Z_t^j)].$$

The  $Z^i$  are martingales, hence

$$\mathbf{E}[Z_t^i Z_t^j] = \mathbf{E}[\Delta_0 + \dots + \Delta_{t-1}] = \mathbf{E}[\Delta_0] + \dots + \mathbf{E}[\Delta_{t-1}].$$

Our goal is to compute those  $\Delta_s$ . First, we have

$$\Delta_t = \left( \frac{n^2}{\mu_i \mu_j d^2} \right)^{t+1} \sum_{\substack{d^+(o, x)=t \\ d^+(o, x')=t}} \prod_{s=0}^{t-1} P_{i(x_s), t(x_{s+1})} P_{i(x'_s), t(x'_{s+1})} \times E(x_t, x'_t) \quad (5.9.3)$$

where the sum runs over all the couples of paths of length  $t$  started at the root:  $o = x_0 \rightarrow x_1 \rightarrow \dots \rightarrow x_t$  and  $x'_0 = o \rightarrow x'_1 \rightarrow \dots \rightarrow x'_t$ , and where  $E(x_t, x'_t)$  is given by

$$\mathbf{E}_t \left[ \left( \sum_{x_t \rightarrow y} P_{i(x_t), t(y)} \varphi_j(t(y)) - \frac{d\mu_j \varphi_j(t(x_t))}{n} \right) \times \left( \sum_{x'_t \rightarrow y'} P_{i(x'_t), t(y')} \varphi_k(t(y')) - \frac{d\mu_k \varphi_k(t(x'_t))}{n} \right) \right].$$

We have already computed those expectations in (5.9.1). More precisely, when  $x_t \neq x'_t$ , the content of the two parentheses inside the expectation are totally independent and centered, hence the only contributions to (5.9.3) correspond to the summands where  $x_t = x'_t$ . In this case, (5.9.1) and (5.9.2) yields

$$E(x_t, x_t) = \frac{d}{n^2} Q\varphi^{i,j}(t(x_t)).$$

We thus have

$$\Delta_t = \frac{d}{n^2} \left( \frac{n^2}{\mu_i \mu_j d^2} \right)^{t+1} \sum_{d^+(o, x)=t} \prod_{s=0}^{t-1} P_{i(x_s), t(x_{s+1})}^2 Q\varphi^{i,j}(t(x_t)). \quad (5.9.4)$$

Since our goal is to compute  $\mathbf{E}[\Delta_s]$  for any  $s$ , we now apply repeated conditioning:  $\mathbf{E}[\Delta_t] = \mathbf{E}[\mathbf{E}_0[\mathbf{E}_1[\dots\mathbf{E}_s[\Delta_s]\dots]]]$ . By the computations done earlier, it is easy to see that

$$\mathbf{E}[\Delta_s] = \left(\frac{d}{n^2}\right)^{s+1} \left(\frac{n^2}{\mu_i \mu_j d^2}\right)^{s+1} Q^s \varphi^{i,j}(t(x)) = \frac{Q^s \varphi^{i,j}(t(x))}{(\mu_i \mu_j d)^s}.$$

This directly gives identity (5.7.9). Identity (5.7.7) also readily follows:

$$\begin{aligned} \mathbf{E}[f_{\phi, \varphi_i, t}(T_x, x) f_{\phi, \varphi_j, t}(T_x, x)] &= \mu_i^t \mu_j^t \phi(x)^2 \mathbf{E}[Z_i^i Z_j^j] \\ &= \mu_i^t \mu_j^t \phi(x)^2 \sum_{s=0}^t \frac{(Q^s \varphi^{i,j})(t(x))}{(\mu_i \mu_j d)^s} \\ &= \mu_i^t \mu_j^t \phi(x)^2 \Gamma_{i,j}^{(t)}. \end{aligned}$$

### 5.9.5 Proof of (5.7.8)

We do the same thing as before, but instead of looking at  $\mu_i^{-t} f_{1, \phi, t}(T_x, x)$  we directly do the computations on  $f_{1, \phi, t}(T_x, x)$ . The rest is identical.

## 5.10 Proof of Proposition 5.8.5

In this section, we prove Proposition 5.8.5. The proof relies on the expected high trace method introduced in random matrix theory by Füredi and Komlós [81] and on techniques developed in [43] for sparse random matrices.

### 5.10.1 Norm of $\underline{A}^{(k)}$

In this subsection, we prove the following lemma. We set

$$\vartheta_0 = \frac{L}{d} \quad \text{and} \quad \vartheta = \sqrt{\frac{\rho}{d}}.$$

**LEMMA 5.10.1.** *There exists a universal constant  $c_0 \geq 3$  such that for all integers  $1 \leq k \leq \ln(n)$  and  $n \geq c_0 \vee (Krb^2)^6$ ,*

$$\left(\mathbf{E}\left\{\|\underline{A}^{(k)}\|^{2m}\right\}\right)^{\frac{1}{2m}} \leq \ln(n)^{14} (Krb^2)^2 (\vartheta \vee \vartheta_0)^k,$$

where  $m = \ln(n/(Krb^2)^3)/(12 \ln(\ln(n)))$ .

From Markov inequality, Lemma 5.10.1 implies Equation (5.8.14). We start the proof of Lemma 5.10.1 by the norm identities

$$\|\underline{A}^{(k)}\|^{2m} = \|\underline{A}^{(k)} \underline{A}^{(k)*}\|^m = \left\| \left( \underline{A}^{(k)} \underline{A}^{(k)*} \right)^m \right\|.$$

From the trace formula, we get

$$\begin{aligned} \|\underline{A}^{(k)}\|^{2m} &\leq \text{tr} \left\{ \left( \underline{A}^{(k)} \underline{A}^{(k)*} \right)^m \right\} \\ &= \sum_{(x_1, \dots, x_{2m})} \prod_{t=1}^m (\underline{A}^{(k)})_{x_{2t-1} x_{2t}} (\underline{A}^{(k)})_{x_{2t+1} x_{2t}}, \end{aligned}$$

where the product is over all  $(x_1, \dots, x_{2m})$  in  $[n]^{2m}$  and we have set  $x_{2m+1} = x_1$ . From the definition of  $\underline{A}^{(k)}$  in (5.8.11), taking expectation, we get

$$\mathbf{E} \|\underline{A}^{(k)}\|^{2m} \leq \left(\frac{n}{d}\right)^{2km} \sum_{\gamma} \mathbf{E} \prod_{i=1}^{2m} \prod_{t=1}^k P_{\gamma_{i,t-1} \gamma_{i,t}} M_{\gamma_{i,t-1} \gamma_{i,t}}, \quad (5.10.1)$$

where the sum is over all  $\gamma = (\gamma_1, \dots, \gamma_{2m})$  with  $\gamma_i = (\gamma_{i,0}, \dots, \gamma_{i,k}) \in F^k$  and the boundary conditions: for all  $i \in [m]$ ,

$$\gamma_{2i,0} = \gamma_{2i+1,0} \quad \text{and} \quad \gamma_{2i-1,k} = \gamma_{2i,k}$$

with  $\gamma_{2m+1} = \gamma_1$ .

We associate to an element  $\gamma$  as above, a directed graph  $G_\gamma = (V_\gamma, E_\gamma)$  with vertices  $V_\gamma = \{\gamma_{i,t} : 1 \leq i \leq 2m, 0 \leq t \leq k\}$  and edge set  $E_\gamma = \{(\gamma_{i,t-1}, \gamma_{i,t}) : 1 \leq i \leq 2m, 1 \leq t \leq k\}$ . This graph may have loops (edges of  $E_\gamma$  of the form  $(x, x)$ ) and inverse edges (pair of edges  $(x, y)$  and  $(y, x)$  in  $E_\gamma$ ). From the above boundary conditions, the graph  $G_\gamma$  is simply connected. In particular, the genus of  $G_\gamma$  is non-negative:

$$|E_\gamma| - |V_\gamma| + 1 \geq 0. \quad (5.10.2)$$

Each oriented edge  $e \in E_\gamma$  has a multiplicity  $m_e$  defined as the number of times it is visited by  $\gamma$ :

$$m_e = \sum_{(i,t) \in [2m] \times [k-1]} \mathbb{1}_{(\gamma_{i,t}, \gamma_{i,t+1})=e}.$$

By construction,

$$\sum_{e \in E_\gamma} m_e = 2km. \quad (5.10.3)$$

We may now estimate the expectation on the right-hand side of (5.10.1). Recall that the random variables  $\underline{M}_{xy} = (M_{xy} - d/n)$ ,  $x, y$ , are iid, centered, bounded by 1 and with variance  $(d/n) - (d/n)^2$ . It follows that for any  $p \geq 1$ ,  $|\mathbf{E}[\underline{M}_{xy}^p]| \leq d/n$ . We deduce that

$$\mathbf{E} \prod_{i=1}^{2m} \prod_{t=1}^k P_{\gamma_{i,t-1}, \gamma_{i,t}} \underline{M}_{\gamma_{i,t-1}, \gamma_{i,t}} = \prod_{e \in E_\gamma} P_e^{m_e} \mathbf{E}[\underline{M}_{11}^{m_e}] \leq \left(\frac{d}{n}\right)^{|E_\gamma|} \prod_{e \in E_\gamma} |P_e|^{m_e}.$$

Moreover, the above expectation is zero unless all edges have multiplicity at least 2. From (5.10.1), we thus obtain that

$$\mathbf{E} \|\underline{A}^{(k)}\|^{2m} \leq \sum_{\gamma \in W_{k,m}} \left(\frac{n}{d}\right)^{2km - |E_\gamma|} \prod_{e \in E_\gamma} |P_e|^{m_e}, \quad (5.10.4)$$

where  $W_{k,m}$  is the set of paths  $\gamma$  as above such that each edge of  $E_\gamma$  is visited at least twice.

We now organize the sum (5.10.4) in terms of the topological properties of the paths. We introduce the equivalence class in  $W_{k,m}$ , we write  $\gamma \sim \gamma'$  if there exists a permutation  $\sigma \in S_n$  such that  $\gamma' = \sigma \circ \gamma$ , where  $\sigma$  acts on  $\gamma$  by mapping  $\gamma_{i,t}$  to  $\sigma(\gamma_{i,t})$ . We denote by  $\mathscr{W}_{k,m}$  the set of equivalence classes. Obviously,  $|V_\gamma|$  and  $|E_\gamma|$  are invariant in each equivalence class. For  $a, s$  integers, we denote by  $\mathscr{W}_{k,m}(s, a)$  the equivalence classes such that  $|V_\gamma| = s$  and  $|E_\gamma| = a$ . From (5.10.2),  $\mathscr{W}_{k,m}(s, a)$  is empty unless  $a - s + 1 \geq 0$ . Our first lemma is a rough estimate on  $\mathscr{W}_{k,m}(s, a)$ .

**LEMMA 5.10.2.** *Let  $a, s \geq 1$  be integers such that  $a - s + 1 \geq 0$ . We have*

$$|\mathscr{W}_{k,m}(s, a)| \leq (2km)^{6m(a-s+1)+2m}.$$

*Proof.* This lemma is contained in the proof of [43, Lemma 17]. We reproduce the proof for the reader convenience. Let  $\gamma = (\gamma_1, \dots, \gamma_{2m}) \in W_{k,m}$ . We order the set  $T = \{(i, t) : 1 \leq i \leq 2m, 0 \leq t \leq k-1\}$  with the lexicographic order. We think of  $T$  as time. For  $0 \leq t \leq k-1$  and  $i$  odd, we define  $e_{i,t} = (\gamma_{i,t}, \gamma_{i,t+1})$ ,  $y_{i,t} = \gamma_{i,t+1}$ , while for  $i$  even, we set  $e_{i,t} = (\gamma_{i,k-t-1}, \gamma_{i,k-t})$ ,  $y_{i,t} = \gamma_{i,k-t-1}$  (in words: we reverse  $\gamma_i$  for even  $i$ ). A vertex  $x \in V_\gamma \setminus \{\gamma_{1,1}\}$  is visited for the first time at  $\tau \in T$  if  $y_\tau = x$  and for all smaller  $\sigma \in T$ ,  $y_\sigma \neq x$ .

We pick a distinguished path in each equivalence class by saying that  $\gamma \in W_{k,m}$  is *canonical* if  $V_\gamma = \{1, \dots, |V_\gamma|\}$ ,  $\gamma_{1,1} = 1$  and vertices are first visited in order. There exactly one canonical path in

each equivalence class. We thus aim for an upper bound on the number of canonical paths in  $W_{k,m}$  with  $|V_\gamma| = s$  and  $|E_\gamma| = a$  by designing an injective map (or encoding) on such canonical paths.

Our goal is to retrieve unambiguously the values of  $y_\tau$ ,  $\tau \in T$ , from minimal information. For  $\tau \in T$ , we say that  $\tau$  is a *first time*, if  $y_\tau$  has not been seen before. If  $\tau$  is a first time the edge  $e_\tau$  is called a *tree edge*. By construction, the set of tree edges is a sub-graph of  $G_\gamma$  with no weak cycle (without orientation) and vertex set  $V_\gamma$ . We call the other edges of  $G_\gamma$  the *excess edges*. Any vertex different from 1 has its associated tree edge. It follows that the number of excess edges is

$$g = a - s + 1.$$

If  $e_\tau$  is an excess edge, we say that  $\tau$  is an *important time*. Other times are *tree times* (visit of a tree edge which has been seen before).

The set  $T_i = \{(i, t) : 0 \leq t \leq k-1\}$  is composed by the successive repetitions of (i) a sequence of the tree times (possibly empty), (ii) a sequence of first times (possibly empty), (iii) an important time.

We build a first encoding of canonical paths. We mark the important times  $(i, t)$  by the vector  $(y_{i,t}, y_{\tau-1})$ , where  $\tau \in T_i \cup \{(i, k)\}$  is the next time that  $e_\tau$  will not be a tree edge (by convention  $\tau = (i, k)$  if  $\gamma_i$  remains on the tree after  $(i, t)$ ). We can reconstruct a canonical path  $\gamma \in W_{k,m}$ , from the positions of the important times and their marks. Indeed, this follows from two observations (1) there is at most one path between two vertices in an oriented tree, and (2) if  $v$  vertices has been seen so far and  $\tau$  is a first time then  $y_\tau = v + 1$ . It is our first encoding.

We refine this encoding by using the assumption that for each  $i$ ,  $\gamma_i$  is tangle-free. We partition important times into three categories, *short cycling*, *long cycling* and *superfluous times* as follows. Assume that  $\gamma_i$  contains a cycle. Consider the smallest time  $(i, t_1)$  such that  $y_{i,t_1} \in \{y_{i,-1}, \dots, y_{i,t_1}\}$ , where  $y_{i,-1} = \gamma_{i,0}$  for odd  $i$  and  $y_{i,-1} = \gamma_{i,k}$  for even  $i$ . Let  $-1 \leq t_0 \leq t_1$  be such that  $y_{i,t_1} = y_{i,t_0}$ . By the assumption of  $\gamma_i$  being tangle-free,  $C = (y_{i,t_0}, \dots, y_{i,t_1})$  is the only directed cycle visited by  $\gamma_i$ . The last important time, say  $(i, t_i)$ , before  $(i, t_1)$  is called the short cycling time. We denote by  $t_2$  the next time after  $(i, t_1)$  that is not an edge of  $C$  ( $\gamma_i$  circles around  $C$  between times  $(i, t_0)$  and  $(i, t_2)$ ). We modify the mark of the short cycling time as  $(y_{i,t_i}, y_{i,t_i}, y_{\tau-1})$  where  $\tau \in T_i \cup \{(i, k)\}$  is the next time after  $(i, t_2)$  that  $e_\tau$  will not be a tree edge (by convention  $\tau = (i, k)$  if  $\gamma_i$  remains on the tree). Important times  $(i, t)$  with  $0 \leq t < t_i$  or  $t_2 \leq t \leq k-1$  are called long cycling times. The other important times are superfluous. The key observation is that for each  $1 \leq i \leq 2m$ , the number of long cycling times  $(i, t)$  is bounded by  $g-1$  (since there is at most one cycle, no edge of  $\gamma_i$  can be seen twice outside those of  $C$ ,  $-1$  coming from the fact that the short cycling time is an important time). Now consider the case where the  $i$ -th path does not contain a cycle, then all important times are called long cycling times and their number is bounded by  $g$ .

We can reconstruct a canonical path  $\gamma \in W_{k,m}$ , from the sole positions of the short and long cycling times and their marks. This our second encoding. For each  $i$ , there are at most  $(k+1)^g$  ways to position the short and long cycling times of  $T_i$ ,  $s^2$  possibilities for the mark of a long cycling time and  $s^3$  possibilities for the mark of a short cycling time. We deduce that

$$|\mathcal{W}_{k,m}(s, a)| \leq (k+1)^{2mg} (s^2)^{2m(g-1)} (s^3)^{2m}.$$

Since  $s \leq 2km$ , the conclusion follows.  $\square$

Our second lemma bounds the contributions of paths in each equivalence class. This lemma is the new main technical difference with [43].

**LEMMA 5.10.3.** *Let  $\gamma \in W_{k,m}$  such that  $|V_\gamma| = s$  and  $|E_\gamma| = a$ . We have*

$$\sum_{\gamma' : \gamma' \sim \gamma} \prod_{e \in E_{\gamma'}} |P_e|^{m_e} \leq n^{-2km+s} K^{km-a} (Krb^2)^{3(a-s)+4m} \rho^{km}.$$

*Proof.* We first express the product of entries of  $P$  in terms of the matrix  $Q$ :

$$\prod_{e \in E_\gamma} |P_e|^{m_e} = \prod_{e \in E_\gamma} |P_e|^{m_e-2} P_e^2 \leq \left( \frac{\sqrt{\rho K}}{n} \right)^{2km-2a} \prod_{e \in E_\gamma} \frac{Q_e}{n},$$

where we have used (5.10.3),  $m_e \geq 2$  and  $\max |P_{xy}| = \sqrt{\rho K}/n$ .

The statement of the lemma immediately follows from the claim:

$$\sum_{\gamma': \gamma' \sim \gamma} \prod_{e \in E_{\gamma'}} Q_e \leq \rho^a n^{s-a} (Krb^2)^{3(a-s)+4m} \quad (5.10.5)$$

Indeed, let us check (5.10.5). Let us define the degree of a vertex  $x$  in  $V_\gamma$  as the sum of in-degrees and out-degrees:  $\sum_{i,t} (\mathbb{1}_{\gamma_{i,t}=x} + \mathbb{1}_{\gamma_{t,i}=x})$ . Let  $s_k$  and  $s_{\geq k}$  be the set of vertices of degree  $k$  and at least  $k$ . We have

$$s_1 + s_2 + s_{\geq 3} = s \quad \text{and} \quad s_1 + 2s_2 + 3s_{\geq 3} \leq \sum_k ks_k = 2a.$$

Subtracting the right-hand side to twice the left-hand side, we find

$$s_{\geq 3} \leq 2(a-s) + s_1 \leq 2(a-s) + 2m.$$

The bound  $s_1 \leq 2m$  comes from the fact that only the vertices  $\gamma_{i,0}$  and  $\gamma_{i,k}$ , with  $i \in [2m]$ , can be of degree 1. Indeed, other vertices are of degree at least 2: for  $1 \leq t \leq k$ ,  $\gamma_{i,t}$  has in-degree at least 1 and out-degree at least 1.

Consider the set  $\hat{V}_\gamma$  of vertices  $V_\gamma$  which are of degree 1 or of degree at least 3. In other words,  $V_\gamma \setminus \hat{V}_\gamma$  is the set of vertices of degree 2. From what precedes

$$\hat{s} = |\hat{V}_\gamma| \leq 2(a-s) + 4m. \quad (5.10.6)$$

We may partition the edges of  $E_\gamma$  into  $\hat{a}$  sequences of edges of the form, for  $1 \leq j \leq \hat{a}$ ,  $\hat{e}_j = (e_{j,1}, \dots, e_{j,q_j})$ , with  $e_{j,t} = (x_{j,t-1}, x_{j,t}) \in E_\gamma$ ,  $x_{j,0}, x_{j,q_j}$  in  $\hat{V}_\gamma$  and  $x_{j,t} \notin \hat{V}_\gamma$  for  $1 \leq t \leq q_j - 1$ . By construction

$$\sum_{j=1}^{\hat{a}} q_j = a. \quad (5.10.7)$$

We consider the directed graph  $\hat{G}_\gamma$  on the vertex set  $\hat{V}_\gamma$  whose  $\hat{a}$  edges are, for  $1 \leq j \leq \hat{a}$ ,  $(x_{j,0}, x_{j,q_j})$  (this is a multi-graph: if two sequences  $\hat{e}_j$  and  $\hat{e}_i$ ,  $i \neq j$ , have the same extreme vertices, it creates two edges). It is straightforward to check that this operation preserves the genus:

$$a - s = \hat{a} - \hat{s}. \quad (5.10.8)$$

For ease of notation, let  $y_1, \dots, y_{\hat{s}}$  be the elements of  $\hat{V}_\gamma$ . Let  $a_j$  and  $b_j$  the indices such that  $x_{j,0} = y_{a_j}$  and  $x_{j,q_j} = y_{b_j}$ . Summing over all possible vertices, we get

$$\sum_{\gamma': \gamma' \sim \gamma} \prod_{e \in E_{\gamma'}} Q_e \leq \sum_{(y_1, \dots, y_{\hat{s}}) \in [n]^{\hat{s}}} \prod_{j=1}^{\hat{a}} Q_{y_{a_j} y_{b_j}}^{q_j},$$

where we have used that

$$\sum_{(x_{j,1}, \dots, x_{j,q_j-1})} \prod_{t=1}^{q_j} Q_{x_{j,t-1} x_{j,t}} = Q_{x_{j,0} x_{j,q_j}}^{q_j}.$$

We apply (5.5.10) and find

$$\sum_{\gamma': \gamma' \sim \gamma} \prod_{e \in E_{\gamma'}} Q_e \leq \sum_{(y_1, \dots, y_{\hat{s}}) \in [n]^{\hat{s}}} \prod_{j=1}^{\hat{a}} \left( \frac{Krb^2 \rho^{q_j}}{n} \right).$$

Using (5.10.6)-(5.10.8), we have  $\hat{a} \leq 3(a-s) + 4m$  and, from (5.10.7), Equation (5.10.5) follows.  $\square$

We are ready for the proof of Lemma 5.10.1.

*Proof of Lemma 5.10.1.* Note that  $\mathcal{W}_{k,m}(s,a)$  is empty unless  $0 \leq s-1 \leq a \leq km$  (since each edge has multiplicity at least 2, we have  $2|E_\gamma| \leq 2km$  from (5.10.3)) From (5.10.4), we get

$$\mathbf{E}\|\underline{A}^{(k)}\|^{2m} \leq \sum_{a=1}^{km} \sum_{s=1}^{a+1} \left(\frac{n}{d}\right)^{2km-a} |\mathcal{W}_{k,m}(s,a)| \max_{\gamma \in \mathcal{W}_{k,m}(s,a)} \sum_{\gamma': \gamma' \sim \gamma} \prod_{e \in E_{\gamma'}} |P_e|^{m_e}.$$

Using Lemma 5.10.2 and Lemma 5.10.3, we arrive at

$$\begin{aligned} \mathbf{E}\|\underline{A}^{(k)}\|^{2m} &\leq n \sum_{a=1}^{km} \sum_{g=0}^{\infty} d^{a-2km} (2km)^{6mg+2m} n^{-g} K^{km-a} (Krb^2)^{3g+4m-3} \rho^{km} \\ &= n \vartheta^{2km} (2km)^{2m} (Krb^2)^{4m-3} \sum_{a=1}^{km} \left(\frac{K}{d}\right)^{km-a} \sum_{g=0}^{\infty} \left(\frac{(Krb^2)^3 (2km)^{6m}}{n}\right)^g, \end{aligned}$$

where we have performed the change of  $s \rightarrow g = a - s + 1$  and used  $\vartheta = \sqrt{\rho/d}$ .

Recall  $k \leq \ln(n)$ . We take  $m = \lceil \ln(n/(Krb^2)^3)/(12 \ln(\ln(n))) \rceil$ . If  $n \geq C(Krb^2)^3$  for some universal constant  $C$ , we find that

$$\frac{(Krb^2)^3 (2km)^{6m}}{n} \leq \frac{1}{2}.$$

We deduce that

$$\mathbf{E}\|\underline{A}^{(k)}\|^{2m} \leq n \vartheta^{2km} \left(1 \vee \frac{K}{d}\right)^{km} (2km)^{2m} (Krb^2)^{4m} (2km).$$

For our choice of  $m$ ,  $2km \leq \ln(n)^2/\ln(\ln(n))$  and, if  $\sqrt{n} \geq (Krb^2)^3$ , then  $n^{1/(2m)} \leq \ln(n)^{12}$ . The conclusion follows easily.  $\square$

### 5.10.2 Norm of $R_k^{(\ell)}$

In this subsection, we prove (5.8.15).

**LEMMA 5.10.4.** *There exists a universal constant  $c_1 \geq 3$  such that for all integers  $1 \leq k \leq \ell \leq \ln(n)$  and  $n \geq c_1$ ,*

$$\left(\mathbf{E}\left\{\|R_k^{(\ell)}\|^{2m}\right\}\right)^{\frac{1}{2m}} \leq \frac{d}{n} \ln(n)^{23} L^\ell,$$

where  $m = \ln(n)/(24 \ln(\ln(n)))$ .

The proof follows from the same line than the proof of Lemma 5.10.1. It is also essentially contained in [43]. To avoid repetitions, we only focus on the main differences with the proof of Lemma 5.10.1. The computation leading to (5.10.1) gives

$$\mathbf{E}\|R_k^{(\ell)}\|^{2m} \leq \left(\frac{n}{d}\right)^{2(\ell-1)m} \sum_{\gamma \in W'_{\ell,m}} \mathbf{E} \prod_{i=1}^{2m} \prod_{t=1}^{\ell} P_{\gamma_{i,t-1}\gamma_{i,t}} \left( \mathbb{1}_{t < k} M_{\gamma_{i,t-1}\gamma_{i,t}} + \mathbb{1}_{t=k} + \mathbb{1}_{t > k} M_{\gamma_{i,t-1}\gamma_{i,t}} \right), \quad (5.10.9)$$

where  $W'_{\ell,m}$  is the set of  $\gamma = (\gamma_1, \dots, \gamma_{2m})$  with  $\gamma_i = (\gamma_{i,0}, \dots, \gamma_{i,\ell}) \notin F^\ell$ ,  $(\gamma_{i,0}, \dots, \gamma_{i,k-1}) \in F^{k-1}$ ,  $(\gamma_{i,k+1}, \dots, \gamma_{i,\ell}) \in F^{\ell-k}$  and the boundary conditions: for all  $i \in [m]$ ,

$$\gamma_{2i,0} = \gamma_{2i+1,0} \quad \text{and} \quad \gamma_{2i-1,\ell} = \gamma_{2i,\ell}$$

with  $\gamma_{2m+1} = \gamma_1$ .

We associate to an element  $\gamma \in W'_{\ell,m}$  the directed graph  $G'_\gamma = (V'_\gamma, E'_\gamma)$  with vertices  $V'_\gamma = \{\gamma_{i,t} : 1 \leq i \leq 2m, 0 \leq t \leq \ell\}$  and edge set  $E'_\gamma = \{(\gamma_{i,t-1}, \gamma_{i,t}) : 1 \leq i \leq 2m, 1 \leq t \leq \ell, t \neq k\}$ . The graph  $G'_\gamma$  is not necessarily weakly connected (since  $E'_\gamma$  does not contain the edges  $(\gamma_{i,k-1}, \gamma_{i,k})$ ). However, we have the following observation.



**LEMMA 5.10.5.** *If  $\gamma$  is as above then each connected component of  $G'_\gamma$  contains a cycle. In particular,  $|E'_\gamma| \geq |V'_\gamma|$ .*

*Proof.* By recursion, it is then enough to check that each connected component of  $G'_\gamma$  contains a cycle. By assumption,  $\gamma_i = (\gamma_{i,0}, \dots, \gamma_{i,\ell}) \notin F^\ell$  contains two distinct cycles. Up to recomposing a new cycle, we may assume without loss of generality that the edge  $(\gamma_{i,k-1}, \gamma_{i,k})$  is in zero or one of the two cycles. If it is in one of them, then the graph  $G'_\gamma$  is weakly connected and it contains the other cycle. Assume now that  $(\gamma_{i,k-1}, \gamma_{i,k})$  is in none of the two cycles. If  $G'_\gamma$  is weakly connected, there is nothing to prove. If  $G'_\gamma$  is not weakly connected, then the two connected components are the vertices of  $(\gamma_{i,0}, \dots, \gamma_{i,k-1})$  and  $(\gamma_{i,k}, \dots, \gamma_{i,\ell})$ . Since these two paths are tangle-free, each must contain exactly one of the two cycles and the statement follows.  $\square$

Using the independence of the entries of  $M$  and  $|P_{xy}| \leq L/n$ ,

$$\mathbf{E} \prod_{i=1}^{2m} \prod_{t=1}^{\ell} P_{\gamma_{i,t-1}\gamma_{i,t}} \left( \mathbb{1}_{t < k} M_{\gamma_{i,t-1}\gamma_{i,t}} + \mathbb{1}_{t=k} + \mathbb{1}_{t > k} M_{\gamma_{i,t-1}\gamma_{i,t}} \right) \leq \left( \frac{d}{n} \right)^{|E'_\gamma|} \left( \frac{L}{n} \right)^{2\ell m}.$$

We thus obtain that

$$\mathbf{E} \|R_k^{(\ell)}\|^{2m} \leq \left( \frac{d}{n} \right)^{2m} \left( \frac{L}{d} \right)^{2\ell m} \sum_{\gamma \in W'_{\ell,m}} \left( \frac{d}{n} \right)^{|E'_\gamma|}, \quad (5.10.10)$$

We introduce the equivalence class in  $W'_{\ell,m}$ , we write  $\gamma \sim \gamma'$  if there exists a permutation  $\sigma \in S_n$  such that  $\gamma' = \sigma \circ \gamma$ , where  $\sigma$  acts on  $\gamma$  by mapping  $\gamma_{i,t}$  to  $\sigma(\gamma_{i,t})$ . We denote by  $\mathscr{W}'_{\ell,m}(s, a)$  the set of equivalence classes such that  $|V'_\gamma| = s$  and  $|E'_\gamma| = a$ . From Lemma 5.10.5,  $\mathscr{W}'_{k,m}(s, a)$  is empty unless  $a \geq s$ . We have the following estimate on  $\mathscr{W}'_{\ell,m}(s, a)$ .

**LEMMA 5.10.6.** *Let  $a \geq s \geq 1$  be integers. We have*

$$|\mathscr{W}'_{\ell,m}(s, a)| \leq (2\ell m)^{12m(a-s)+22m}.$$

*Proof.* A proof is contained in [43, Lemma 18]. We give a proof for the reader convenience. We order the sets  $T' = \{(i, t) : 1 \leq i \leq 2m, 1 \leq t \leq \ell - 1, t \neq k\}$  and  $T = \{(i, t) : 1 \leq i \leq 2m, 1 \leq t \leq \ell - 1\}$  with the lexicographic order. We think of  $T$  and  $T'$  as times. If  $\tau \in T'$ , we denote  $\tau^-$  the largest element in  $T'$  smaller than  $\tau$ . By convention  $(1, 0)^- = (1, -1)$ . For  $\tau \in T$ , we define  $e_\tau$  and  $y_\tau$  as in Lemma 5.10.2 and we say  $\gamma \in W'_{\ell,m}$  is *canonical* if  $V'_\gamma = \{1, \dots, |V'_\gamma|\}$ ,  $\gamma_{1,1} = 1$  and vertices are first visited in order. We aim for an upper bound on the number of canonical paths in  $W'_{\ell,m}$  with  $|V'_\gamma| = s$  and  $|E'_\gamma| = a$  by designing an injective map.

We define a sequence of growing sub-forests  $(F_\tau)_{\tau \in T'}$  of  $G'_\gamma$  as follows. We start with  $F_{(1,-1)}$ , the trivial graph with no edge and a  $\gamma_{1,1} = 1$  as unique vertex. For  $\tau \in T'$ , we say that  $\tau$  is a *first time*, if adding  $e_\tau$  to  $F_{\tau^-}$  does not create a weak cycle. If  $\tau$  is a first time the edge  $e_\tau$  is called a *tree edge* and we define  $F_\tau$  as the union of  $e_\tau$  and  $F_{\tau^-}$ . Otherwise,  $F_\tau = F_{\tau^-}$ . We set  $F = F_{2m,\ell}$ . By construction, the set of tree edges is a sub-graph of  $G'_\gamma$  with no weak cycle and vertex set  $V'_\gamma$ . Moreover, the weak connected components of  $G'_\gamma$  and  $F$  are equal. We call the other edges of  $G'_\gamma$  the *excess edges*. In each weak connected component of  $G'_\gamma$  there are at most  $g = a - s + 1$  excess edges. Indeed, if  $a'$ ,  $s'$  are the numbers of directed edges and vertices of a connected component, then there are  $a' - s' + 1$  excess edges in this connected component. However by Lemma 5.10.5,  $a' - s' \leq a - s$ . If  $e_\tau$  is an excess edge, we say that  $\tau \in T'$  is an *important time*. Other times in  $T'$  are *tree times* (visit of a tree edge which has been seen before).

Let  $k_i = k$  for odd  $i$  and  $k_i = k - \ell + 1$  for even  $i$ . We define the sets  $T_i^1 = \{(i, t) : 0 \leq t \leq k_i - 1\}$  and  $T_i^2 = \{(i, t) : k_i \leq t \leq \ell\}$ . For each  $i$ , there could be a special first time  $(i, t) \in T_i^2$ , called the *merging time*, such that a connected component of  $F_{(i,t)^-}$  merges into a connected component of  $F_{i,k_i-1}$  by the addition of  $e_{i,t}$ .

The sets  $T_i^\varepsilon$  are composed by the successive repetitions of (i) a sequence of the tree times (possibly empty), (ii) a sequence of first times (possibly empty), (iii) an important time or the merging time. We mark the important and merging times (for  $\varepsilon = 2$ )  $(i, t) \in T_i^\varepsilon$  by the vector  $(y_{i,t}, y_{\tau-1})$ , where  $\tau \in T_i^\varepsilon \cup \{(i, k_i)\}$  is the next time that  $e_\tau$  will not be a tree edge (by convention  $\tau = (i, k_i)$  if  $T_i^\varepsilon$  only contains tree times after  $(i, t)$ ). We can reconstruct a canonical path  $\gamma \in W'_{\ell, m}$ , from the positions of the merging and important times and their marks.

We refine this encoding by partitioning important times into three categories, *short cycling*, *long cycling* and *superfluous times* exactly as done in Lemma 5.10.2, except that there are short and long cycling times for each  $i$  and  $\varepsilon \in \{1, 2\}$  in the sequence  $T_i^\varepsilon$ . There are either 0 short cycling times and at most  $g$  long cycling times, or 1 short cycling time and at most  $g - 1$  long cycling time (because in each connected component of  $G'_\gamma$  there are at most  $g$  excess edges).

We can reconstruct a canonical path  $\gamma \in W'_{\ell, m}$ , from the positions of the merging, short and long cycling times and their marks. There are at most  $\ell^{2m}$  ways to position the merging times. For each  $i, \varepsilon$ , there are at most  $\ell^g$  ways to position the short and long cycling times of  $T_i^\varepsilon$ ,  $s^2$  possibilities for the marks of a merging or long cycling time and  $s^3$  possibilities for the marks of a short cycling time. We deduce that

$$|\mathcal{W}'_{\ell, m}(s, a)| \leq \ell^{4mg+2m} (s^2)^{4m(g-1)+2m} (s^3)^{4m}.$$

Since  $s \leq 2\ell m$ , the conclusion follows.  $\square$

We are ready for the proof of Lemma 5.10.4.

*Proof of Lemma 5.10.4.* There are  $n(n-1)\cdots(n-s+1)$  elements of  $W'_{\ell, m}$  in an equivalence class in  $\mathcal{W}'_{\ell, m}(s, a)$ . From (5.10.10) and Lemma 5.10.6 we get

$$\mathbf{E}\|R_k^{(\ell)}\|^{2m} \leq \left(\frac{d}{n}\right)^{2m} \left(\frac{L}{d}\right)^{2\ell m} \sum_{a=1}^{2\ell m} \sum_{s=1}^{2\ell m} n^s (2\ell m)^{12m(a-s)+22m} \left(\frac{d}{n}\right)^a,$$

We perform the change of variable  $s \rightarrow p = a - s$ :

$$\mathbf{E}\|R_k^{(\ell)}\|^{2m} \leq \left(\frac{d}{n}\right)^{2m} \left(\frac{L}{d}\right)^{2\ell m} (2\ell m)^{22m} \sum_{a=1}^{2\ell m} d^a \sum_{p=0}^{\infty} \left(\frac{(2\ell m)^{12m}}{n}\right)^p.$$

Recall  $\ell \leq \ln(n)$  and  $d \geq 1$ . We take  $m = \lceil \ln(n)/(24 \ln(\ln(n))) \rceil$ . If  $n \geq c_1$  for some universal constant  $c_1$ , we find that

$$\frac{(2km)^{12m}}{n} \leq \frac{1}{2}.$$

We deduce that

$$\mathbf{E}\|R_k^{(\ell)}\|^{2m} \leq \left(\frac{d}{n}\right)^{2m} L^{2\ell m} (2\ell m)^{22m} (2\ell m).$$

For our choice of  $m$ ,  $2\ell m \leq \ln(n)^2$ . The conclusion follows easily.  $\square$



# Appendix A

## Introduction: english version

In this introduction, I will focus on presenting the works done during these three years of PhD. They mainly focus on the spectra of random graphs or matrices. An important idea underlying a consequential part of modern research in applied mathematics is that the greatest eigenvalues of a matrix are those that provide the best information about this matrix. It is therefore of great interest to have detailed descriptions of these eigenvalues. I will try to present this importance as much as possible through this dissertation.

The first work of this thesis is inspired by the prepublication *The spectral gap of sparse random digraphs* ([62]). I study the second eigenvalue of the transition matrix on directed configuration graphs: in this model, the incoming and outgoing degrees  $d_i^-$  and  $d_i^+$  of each vertex is specified, then one takes a directed multigraph  $G$  uniformly at random among the graphs that have these degrees. The second eigenvalue of the transition matrix is then upper bounded, the bound being explicit and depending only on the degrees; this allows us to solve Alon's *directed* conjecture, on regular directed graphs.

In general, an important question in graph theory is to link a graph to its degree sequence; one can first ask whether, given integers  $(d_1, \dots, d_n)$ , it is possible to construct a graph  $G$  with these degrees. This is not always the case: the question was entirely solved by Erdős and Gallai in the 1960s, and generated a whole field of research on the possibility of reconstructing graphs from local information such as degrees. In the second work presented here, inspired by a paper in collaboration with Charles Bordenave ([41]), we are interested in the existence of graphs with given *neighborhoods*, the question having recently emerged in the context of local graph convergence.

These first two parts are relatively akin by their object, graphs and degrees. The third part, the result of a collaboration with Justin Salez ([63]), slightly differs from them; the main object remains the large random graphs with few edges, but we study the *nature* of the spectrum, and in particular the existence of a continuous part, which corresponds to the notion of *extended states* from quantum mechanics. In particular, we prove a criterion for the existence or absence of extended states in the spectrum of a very general model, unimodular Galton-Watson trees. This allowed us to answer a question asked by physicists Bauer and Golinelli in 2001.

Finally, in the last part, resulting from a collaboration with Charles Bordenave and Raj Rao Nadakuditi, I present some results on the large eigenvalues of diluted matrices. We consider a directed Erdős-Rényi graph on  $n$  vertices, in the diluted regime where the average degree is  $d$ , independent of  $n$ . Each edge  $(i, j)$  of the graph is weighted by a weight  $P_{i,j}$ . The goal is to obtain information on the matrix  $P$  from the observation of the weighted adjacency matrix  $A$ ; this is the problem of matrix completion, which has been extensively studied for about fifteen years. We show a spectacular phase transition: under natural assumptions on  $P$ , there is a threshold  $\vartheta = \vartheta(P, d)$  such that the largest eigenvalues of  $(n/d)A$  converge as  $n \rightarrow \infty$  towards eigenvalues of  $P$  greater than  $\vartheta$ . All other eigenvalues

of  $(n/d)A$  are contained in  $D(0, \vartheta)$ . We also show that the eigenvectors of  $A$  associated with these eigenvalues are correlated with the corresponding eigenvectors of  $P$ . This result is unexpected, as it shows that (weak) reconstruction of  $P$  or part of  $P$  is feasible even in the regime where  $d$  is fixed.

This introduction provides a full description of these results, as well as background information on each of the covered areas. First (§A.1), I give some generalities about the spectra of matrices and graphs, before describing the theory of local convergence that underlies all the works presented here. The other chapters (§A.2-A.5) are devoted to the work mentioned above and are relatively independent.

Finally and most importantly, I give in conclusion (§A.6, page 185) two ideas that stem from this thesis: the diluted regime (where  $d$  is *really* small) is the interesting regime from the point of view of applications, and sometimes it is better to un-symmetrize problems that are naturally symmetrical, a promising program considering recent works in statistics.

## A.1 Convergence of the spectrum

We begin by recalling some classical notions of linear algebra and graph theory, then we describe the convergence of graphs in the sense of Benjamini-Schramm.

### Eigenvalues

Any symmetric matrix  $A \in \mathcal{M}_{n,n}(\mathbb{R})$  has a spectral decomposition, in the sense that it is written

$$A = \lambda_1 \varphi_1 \varphi_1^* + \cdots + \lambda_n \varphi_n \varphi_n^*, \quad (\text{A.1.1})$$

with  $\lambda_1 \geq \cdots \geq \lambda_n$  its eigenvalues, and  $\varphi_i$  an orthonormal family of associated eigenvectors. Eigenvalues capture a large amount of information about the matrix  $A$ , and have very useful classical characterizations, for example

$$\lambda_i = \max_{V \in \mathcal{V}_{i,n}} \min_{\substack{x \in V \\ |x|=1}} \langle x, Ax \rangle \quad (\text{A.1.2})$$

where  $\mathcal{V}_{i,n}$  is the set of subspaces of  $\mathbb{R}^n$  with dimension  $i$ . The study of the eigenvalues of *random* symmetric matrices, initiated with Wigner's work, has proved to be extremely fertile in developments from combinatorics, statistics, physics or even number theory.

When the matrix  $A$  is no longer hermitian, it still has  $n$  complex eigenvalues, but does not necessarily have a decomposition like (A.1.1). A popular and useful variant is the singular value decomposition: for any  $A \in \mathcal{M}_{m,n}(\mathbb{R})$ , there are  $\sigma_1 \geq \cdots \geq \sigma_r > 0$ , with  $r = \text{rank}(A)$ , and two orthonormal families  $(\phi_1, \dots, \phi_r)$  in  $\mathbb{R}^m$  and  $(\psi_1, \dots, \psi_r)$  in  $\mathbb{R}^n$ , such that

$$A = \sigma_1 \phi_1 \psi_1^* + \cdots + \sigma_r \phi_r \psi_r^*.$$

Variational characterizations similar to (A.1.2) exist for  $\sigma_i$ . In many applications, singular values of matrices have proven extremely useful, playing *a priori* the role of eigenvalues for symmetric matrices. However, this is not exactly the case, and in several naturally unsymmetrical problems, singular values have proved to be less informative than eigenvalues<sup>1</sup>. One of the goals of this thesis is to study the eigenvalues of certain non-Hermitian matrices, in particular the adjacency matrices of directed graphs.

### Spectra of graphs

Let's start by recalling some graph terminology: a graph  $G = (V, E)$  is a set of vertices  $V$ , always countable in this thesis, and a subset  $E$  of  $V \times V$ . The elements  $(u, v)$  in  $E$  are the edges of the graph.

---

<sup>1</sup>See in particular the concluding remarks of this introduction.

Simple graphs are graphs without loops  $(v, v)$  and whose edges are symmetrical, in the sense that  $(u, v) \in E$  if and only if  $(v, u) \in E$ ; such graphs are also called unoriented or undirected graphs, as opposed to the case where edges  $(u, v)$  can exist without the edge  $(v, u)$  being in  $E$ .

The works presented in this thesis refer to *diluted* graphs, i.e. graphs whose number of edges  $|E|$  is comparable to the number of vertices  $|V|$ .

Any graph  $G = (V, E)$  is entirely characterized by its *adjacency matrix*  $A$ . This matrix is indexed by  $V$  and defined by

$$A_{u,v} = \mathbf{1}_{(u,v) \in E}.$$

It is Hermitian if and only if the corresponding graph is undirected. The algebraic graph theory consists in studying  $G$  via the spectrum of its adjacency matrix — for classical books on the topic, we refer to [51, 64]. One of the advantages of this vast idea is that there are many numerical procedures for approaching the spectrum with arbitrary precision, in time  $O(n^3)$  in the worst case, and that knowledge of the spectrum makes it possible to obtain good approximations of quantities that are very difficult to calculate, or even often NP-complete, such as the isoperimetric constant or the chromatic number — we will return to this in Section A.2. Many problems in statistics or computer science that arise in terms of graphs can thus be solved by spectral algorithms; see for example [134, 90, 10].

### Random graphs

Classical graph theory has focused on studying the properties of particular graphs: given *one* graph, what relationships can be found between, for example, its chromatic number, its edge number, its planarity, etc. This is often necessary in applications, where you have a specific graph (for example, the Facebook graph) from which you want to get information.

However, in many other applications, there is not *one* single graph, but *several* graphs that share some common features (such as the graphs of links between the users of different social networks). Rather than studying these graphs individually, we study entire classes of graphs, and we try to determine properties that are generally true on these classes. Mathematically, we have to identify a particular class, for example the set  $\mathcal{G}(n)$  of simple graphs on  $n$  vertices, and study the properties of a random variable (uniform or not)  $G$  on this class. In this thesis, we study several classical models of random graphs: Erdős-Rényi graphs, uniform trees and graphs with prescribed degrees.

Most of the results are stated under the asymptotic regime in which the size  $n$  of the graphs tends to infinity. We will therefore study increasingly large random graph sequences. In this regime, the models we are studying can be seen as finite approximations of infinite limiting objects, using the *local weak convergence* theory. This point of view is both a fertile heuristic to understand some problems (in particular, those presented in the first part), and a powerful theory to solve others (as in the third and fourth parts).

### Benjamini-Shramm convergence

The main idea of local convergence is to study the *typical* aspect of a graph around any vertex. This section presents the main lines and results of this theory; we refer to [8, 34] for complete introductions, as well as [27].

To describe the theory, we need to define some commonly used terms: first of all, let us state that all the graphs we will study will have the same vertex set  $V$  or a subset of  $V$ , supposedly countable, and that all the graphs considered will have all their degrees finite (we also say that they are locally finite).

- A rooted graph is a pair  $(G, v)$  where  $G$  is a connected graph and  $v \in V$  is a particular vertex of  $G$ , called root.
- When  $(G, v)$  is a rooted graph, we will note  $(G, v)_t$  the set of vertices of  $G$  at a distance less than or equal than  $t$  from the root  $v$ .

- An isomorphism between two rooted graphs  $(G_1, v_1)$  and  $(G_2, v_2)$  is an isomorphism of graphs  $\varphi : G_1 \rightarrow G_2$  which preserves the root, i.e. such that  $\varphi(v_1) = v_2$ .
- When there is such an isomorphism between two rooted graphs  $(G_1, v_1)$  and  $(G_2, v_2)$ , we write  $(G_1, v_1) \simeq (G_2, v_2)$ .
- Finally, we note  $\mathcal{G}_*$  the set of (isomorphism classes of) locally finite connected rooted graphs.

We can endow  $\mathcal{G}_*$  with a distance, called the “local distance”, by putting

$$d((G_1, v_1)(G_2, v_2)) = (1 + T)^{-1},$$

where  $T = \sup\{t \geq 0 : (G_1, v_1)_t \simeq (G_2, v_2)_t\}$ . The metric space thus obtained is Polish. It is therefore possible to endow  $\mathcal{P}(\mathcal{G}_*)$ , the set of all probability measures on  $\mathcal{G}_*$ , with the weak topology. Concretely, this means that  $\mu_n \rightarrow \mu$  if and only if for any continuous bounded function  $f : \mathcal{G}_* \rightarrow \mathbb{R}$ , we have  $\int f d\mu_n \rightarrow \int f d\mu$ . We will say that a sequence of random rooted graphs  $(G_n, v_n)$  converges to a graph  $(G_\infty, v_\infty)$  if the distributions of  $(G_n, v_n)$  converge to the distribution of  $(G_\infty, v_\infty)$ . In practice, we have  $G_n \rightarrow G_\infty$  if and only if for any fixed rooted graph  $(g, v)$ , we have

$$\forall t \in \mathbb{N}, \quad \mathbf{P}((G_n, v_n)_t \simeq (g, v)_t) \rightarrow \mathbf{P}((G_\infty, v_\infty)_t \simeq (g, v)_t).$$

The central idea of the Benjamini-Schramm convergence is that any finite graph sequence  $(G_n)$  gives rise to a sequence of **random** connected rooted graphs by uniform rooting: just take the root  $v_n$  uniformly at random on  $V_n$ , and consider the graph  $(G_n(v_n), v_n)$ , where the notation  $G(v)$  refers to the connected component of  $G$  containing  $v$ .

This very powerful idea allows to study finite graph sequences from a purely local point of view: the convergence of a graph sequence happens if the neighbourhoods of its vertices resemble (in distribution) to the neighbourhood of the root of the limiting graph.

Many graph models converge almost surely in the Benjamini-Schramm sense. This is the case with the models studied in this thesis, which we now present (among others). The literature on each of these models spans entire pages of bibliographies; the first three models are extensively described in [32].

1. **Erdős-Rényi.** Note  $\text{ER}(n, p)$  the Erdős-Rényi distribution of parameter  $d/n$ ; the set of vertices is  $n$  and each of the possible  $n(n-1)/2$  edges is independently put in the graph with probability  $p$ . If  $p = d/n$  with  $d$  a fixed positive real, then  $G_n \rightarrow \text{PGW}(d)$ , the Galton-Watson tree with a Poisson progeny of parameter  $d$ . We refer to [34] for a proof.
2. **Graphs with prescribed degrees.** Let  $\mathbf{d}^{(n)} = (d_1, \dots, d_n)$  be a graphical sequence<sup>2</sup>, and  $G_n$  is uniformly chosen from  $\mathcal{G}(n, \mathbf{d}^{(n)})$ . It is assumed that the empirical distribution of degrees converges towards a measure  $\pi$  on  $\mathbb{N}$ , in distribution and in  $L^2$ . If  $\pi(\{0\}) < 1$ , then  $G_n \rightarrow \text{UGW}(\pi)$ , the unimodular Galton-Watson tree with degree distribution  $\pi$ . This object will be strictly defined in the last part. For proofs, see [34, Chapter 3].
3. **Regular graphs.** This is a special case of the previous one: when  $\mathbf{d}^{(n)} = (d, \dots, d)$ , the graph  $G_n$  is uniform among all  $d$ -regular graphs on  $n$  vertices; in this case,  $\pi = \delta_d$  and the limiting tree  $\text{UGW}(\delta_d)$  is precisely  $\mathbb{T}_d$ , the infinite  $d$ -regular tree<sup>3</sup> (in particular, the limit is deterministic).

---

<sup>2</sup>A sequence  $(d_1, \dots, d_n)$  is graphical if there is indeed a simple graph with  $n$  vertices in which the degree of the vertex  $i$  is  $d_i$ . Not all sequences are graphical: this is the subject of Section A.3.

<sup>3</sup> Also called *Bethe lattice*.

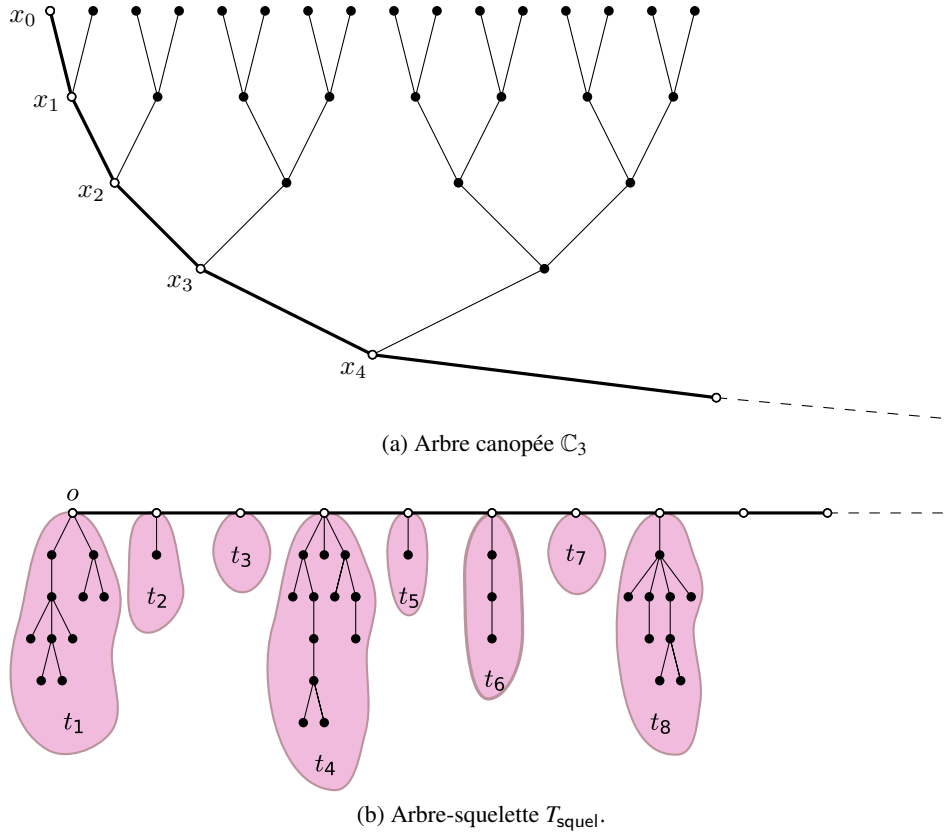


Figure A.1 – (explication)

4. **Uniform Trees.** Let  $T_n$  be a tree uniformly chosen from the set of the  $n^{n-2}$  labelled trees on  $n$  vertices. The sequence  $(T_n)$  converges in the Benjamini-Schramm sense towards the skeleton tree, noted  $T_{\text{squel}}$ , defined as follows: we start with the graph of  $\mathbb{N}$ , rooted at zero, called *spine*; then, at each node  $i$  of this spine, we glue a tree  $t_i$  which is random and distributed as  $\text{PGW}(1)$  — the trees  $t_0, t_1, t_2, \dots$  are independent; see Figure A.1b. For the proof of the convergence  $T_n \rightarrow T_{\text{squel}}$ , see [85].
5. **Canopy tree.** We note  $\mathbb{T}_{d,n} = (\mathbb{T}_n, o)_d$  the infinite  $d$ -regular tree from which we kept only the first  $n$  generations. Unlike regular graphs, this tree has a boundary: a proportion asymptotically equal to  $(d-2)/(d-1)$  of the vertices are leaves. The limit is therefore not  $\mathbb{T}_d$ , but a beautiful recursive object,  $(\mathbb{C}_d, o)$ , called *canopy tree*: see Figure A.1a. The identification of the right model between  $\mathbb{T}_d$  and  $\mathbb{C}_d$  is an interesting question in the physics of the Anderson model, where it reflects the treatment of boundary conditions ([5], [122]).

### Continuity of the spectrum

Let  $G_n$  be a sequence of finite graphs converging to  $(G, \nu)$  in the Benjamini-Schramm sense. The adjacency matrix  $A_n$  has spectrum  $\lambda_1(A_n) \geq \dots \geq \lambda_n(A_n)$  and we'll note

$$\mu_{A_n} = \frac{1}{n} \sum_{i=1}^n \delta_{\lambda_i(A_n)}$$

the empirical spectral measure of  $G_n$ . One of the strong features of the Benjamini-Schramm convergence is that it guarantees that  $\mu_{A_n}$  converges to some limit measure, defined from  $(G, \nu)$  (see [1, 126])



or [34, Proposition 2.5]). More precisely, if  $G_n$  converges in the BS sense to a random graph  $(G, \nu)$  with distribution  $\rho$ , then

$$\sup_{t \in \mathbb{R}} |F_{A_n}(t) - F_\rho(t)| \rightarrow 0 \tag{A.1.3}$$

where  $F_{A_n}$  is the cumulative distribution function of  $\mu_{A_n}$ , and  $F_\rho$  is the cumulative distribution function of some measure  $\mu_\rho$  over  $\mathcal{G}_*$  which depends only on  $\rho$ . The convergence (A.1.3) is in fact equivalent to the convergence of all the atoms (i.e.  $\mu_{A_n}(\{\lambda\}) \rightarrow \mu_\rho(\{\lambda\})$  for any real number  $\lambda$ ) and the weak convergence of  $\mu_{A_n}$  towards  $\mu_\rho$ . Obviously, identifying  $\mu_\rho$  is quite delicate and we will shun the problem for a moment; to do this, we will need an important concept, unimodularity. We will give an explicit description of  $\mu_\rho$  at page 172, as well as a definition of unimodularity.

As described, the convergence (A.1.3) does not give informations on the behaviour of some precise eigenvalues such as  $\lambda_1(A_n)$ , because the behaviour of a negligible number of eigenvalues vanishes in the weak convergence. We cannot entirely rely on the spectral continuity to study the extreme eigenvalues, even if this continuity still gives us some informations; in fact, we will need some *ad hoc* techniques such as trace methods. A consequential part of this thesis will be devoted to the use of trace methods to study the high eigenvalues of graphs.

## A.2 Eigenvalues of regular graphs

In this section, we are interested in graphs with a fixed degree sequence, and in the second eigenvalue of their transition matrix; the results and proofs are gathered in chapter 1 of this thesis, from the prepublication [62].

To begin, we give a brief overview of the motivations around this question, notably through Alon's theorem. To fix the ideas, I start by presenting it in the context of simple  $d$ -regular graphs.

### Spectra and expansion

Let us give ourselves a  $d$ -regular graph  $G$ . The simple random walk on this graph has transition matrix  $P_{x,y} = \frac{1}{d} \mathbf{1}_{(x,y) \in E}$ . It is clear that  $P = \frac{1}{d}A$ , where  $A$  is the adjacency matrix; in this framework, the spectrum of  $P$  and the spectrum of  $A$  are identical up to a homothety. The results of this section are formulated in terms of the transition matrix  $P$ .

It is not difficult to check that the largest eigenvalue of  $P$  is 1, and the multiplicity of 1 is equal to the number of connected components of  $G$ ; on the other hand,  $-1$  is an eigenvalue if and only if the graph is bipartite (see for example [31, Section VIII.2]). The behaviour of the highest eigenvalue is therefore not very informative. The question of knowing the behavior of the *second highest eigenvalue* is much richer and more difficult. We will note

$$\lambda_\star = \max\{|\lambda| : \lambda \text{ eigenvalue of } P \text{ with } |\lambda| < 1\}$$

the modulus of the largest non-trivial eigenvalue. There are many results that link this quantity to important properties of the graph related to its *expansion*: the general idea is that if  $\lambda_\star$  is small, the graph is a good "expander", in the sense that its edges are very well distributed in the graph and that the removal of some of them does not cause major harm in its geometry. Several tools can be used to quantify this intuition.

**Cheeger's Inequality.** For any  $d$ -regular graph  $G$ , we have

$$\frac{d(1 - \lambda_\star)}{2} \leq \Phi(G) \leq d\sqrt{1 - \lambda_\star} \tag{A.2.1}$$

where  $\Phi(G) = \min_{X \subset V, |X| \leq |V|/2} \frac{|\partial X|}{|X|}$  is the isoperimetric constant<sup>4</sup> of the graph. When  $\lambda_\star$  is small, this inequality means that  $\Phi(G)$  is large: for each subset  $X$ , there are many edges that come out of  $X$ .

---

<sup>4</sup> Also called Cheeger constant, conductivity, expansion constant, etc.

**Expander-mixing lemma.** Another link between  $\lambda_\star$  and  $G$  is the famous *expander mixing lemma*: if  $G$  is a  $d$ -regular graph, then for all  $X, Y \subset V$ , we have

$$\left| E(X, Y) - \frac{d|X||Y|}{n} \right| \leq d\lambda_\star \sqrt{|X||Y|} \quad (\text{A.2.2})$$

where  $E(X, Y) = \{(u, v) \in E : u \in X, v \in Y\}$  is the number of edges between  $X$  and  $Y$ . The number of possible edges between  $X$  and  $Y$  is  $|X||Y|/2$  and the density of the edges is  $d/(n-1)$ , approximately  $d/n$  if  $n$  is large. Therefore, if the edges of  $G$  were randomized in  $G$  with this density (which is equivalent to taking an Erdős-Rényi graph with parameters  $(n, d/n)$ ), the average number of edges between  $X$  and  $Y$  should be  $d|X||Y|/n$ . The expander-mixing lemma gives the difference between this average number of edges and the actual number of edges between  $X$  and  $Y$ .

These two inequalities are extremely effective for harvesting information on  $G$  from a simple bound on  $\lambda_\star$  or  $\lambda_2$ . For example, there are bounds on the chromatic number or diameter; we refer to [87, Chapter 4] for many developments around this theme.

The two inequalities (A.2.1) and (A.2.2) are only true *stricto sensu* for  $d$ -regular graphs, but they have an everlasting compilation of variants in the context of oriented graphs, or when the graph is not regular.

**Markov Chains.** In another context,  $\lambda_\star$  gives the long time speed of convergence of Markov chains. In general, the transition matrix of a Markov chain is neither symmetric nor even diagonalizable. Its eigenvalues are therefore not necessarily real, and are ordered by decreasing modulus:  $\lambda_1 = 1 \geq |\lambda_2| \geq \dots \geq |\lambda_n|$ . When the chain is irreducible and aperiodic, the Perron-Frobenius theorem results in 1 being a simple eigenvalue, so in this context we have  $|\lambda_2| = \lambda_\star$ .

**PROPOSITION A.2.1** ([104], [117]). *Let  $P$  be the transition matrix of an irreducible aperiodic Markov chain on a finite state space  $S = \{1, \dots, n\}$ , with invariant distribution  $\pi_\star$ . Let  $1 = |\lambda_1| \geq |\lambda_2| \geq \dots \geq |\lambda_n|$  be the eigenvalues of  $P$  ordered by decreasing modulus, and let  $d(t)$  be the distance to equilibrium at time  $t$ , defined by  $d(t) = \max_{x \in S} \|P^t(x, \cdot) - \pi_\star\|_{\text{TV}}$ , where  $\|\cdot\|_{\text{TV}}$  is the total variation distance. Then,*

$$\lim_{t \rightarrow \infty} d(t)^{\frac{1}{t}} = |\lambda_2|. \quad (\text{A.2.3})$$

Again, if  $|\lambda_2| = \lambda_\star$  is small, the Markov chain on  $G$  will quickly converge towards its equilibrium measure. The computation of  $|\lambda_2|$  is therefore crucial to study the fine properties of random walks on the graph. This is what motivated many studies around  $\lambda_\star$  for various graph models, the simplest — but already difficult — being that of regular graphs.

### The Alon-Boppana bound

The famous Alon-Boppana bound says that  $\lambda_\star$  for regular graphs is essentially larger than  $2\sqrt{d-1}/d$ ; the most precise formulation, to my knowledge, is as follows: if  $P$  is the transition matrix of a  $d$ -regular graph  $G$ , then

$$\lambda_2(P) \geq 2 \frac{\sqrt{d-1}}{d} \cos\left(\frac{\pi}{\text{diam}(G)+2}\right). \quad (\text{A.2.4})$$

In essence, if  $r = \text{diam}(G)$  is large, we have

$$\lambda_\star(P) \geq \lambda_2(P) \geq \frac{2\sqrt{d-1}}{d} \times \left(1 - O\left(\frac{1}{r^2}\right)\right).$$

The first proof is due to Alon ([120]), and this version is due to Mohar ([116]); both use the variational formulations (A.1.2). Within the setting of adjacency matrices of simple graphs, many generalizations

exist (see in particular [125] and its references). If  $G_n$  is a sequence of  $d$ -regular graphs whose diameter tends towards infinity, we have

$$\liminf \lambda_*(P_n) \geq \frac{2\sqrt{d-1}}{d}. \tag{A.2.5}$$

Any sequence of  $d$ -regular graphs whose diameter goes to infinity has a diameter that also goes to infinity, and therefore satisfies the inequality (A.2.5).

$$2\sqrt{d-1}/d$$

The appearance of the number  $2\sqrt{d-1}/d$  is not accidental, and results in particular from the approximation of the empirical spectral measure of  $G$  by the Kesten-McKay density (see Figure A.2). More precisely, let  $G_n = (V_n, E_n)$  be a regular graph whose size  $|V_n|$  goes to infinity, and whose diameter tends towards infinity.

We saw earlier that if  $G_n$  is taken uniformly at random on  $\mathcal{G}(n, d)$  then  $G_n$  converges towards the infinite  $d$ -regular tree  $\mathbb{T}_d$  in the Benjamini-Schramm sense, and therefore the spectral measure <sup>5</sup>  $\mu_{P_n}$  converges ([113]) to the spectrum of the transition operator on  $\mathbb{T}_d$ , known as the Kesten-McKay law. It is an absolutely continuous measure with respect to the Lebesgue measure, it has support  $[-2\sqrt{d-1}/d, 2\sqrt{d-1}/d]$ , and the density (see Figure A.2) is given by

$$\rho_d(x) = d \mathbf{1}_{\left[-\frac{2\sqrt{d-1}}{d}, \frac{2\sqrt{d-1}}{d}\right]} \frac{d\sqrt{4(d-1) - (dx)^2}}{2\pi(d^2 - (dx)^2)}. \tag{A.2.6}$$

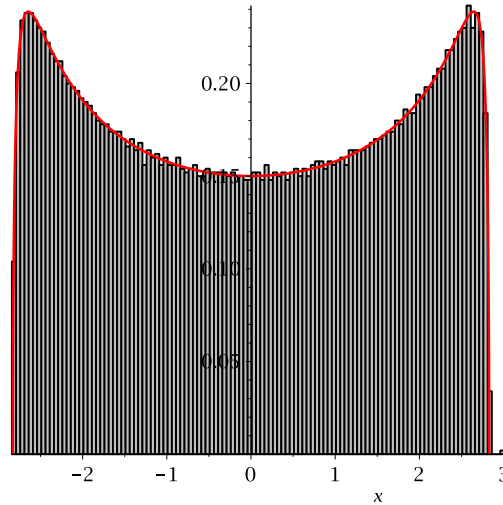


Figure A.2 – Histogram of the eigenvalues of the adjacency matrix  $P$  of a regular 3-graph with 10000 vertices; the Kesten-McKay limit measure is in red.

As we have already mentioned, the weak convergence  $\mu_{G_n} \rightarrow \rho_d(x)dx$  does not give any specific information on the behaviour of particular eigenvalues such as  $\lambda_2$ . However, the Portemanteau lemma implies that for every open subset  $O \subset \mathbb{R}$ , we have  $\liminf \mu_{G_n}(O) \geq \int_O \rho_d$ . By taking  $O = ]2\sqrt{d-1}/d - \varepsilon, +\infty[$ , we get

$$\frac{|\{\text{eigenvalues of } P_n \text{ such that } \lambda > \frac{2\sqrt{d-1}}{d} - \varepsilon\}|}{n} \rightarrow \int_{2\sqrt{d-1}/d - \varepsilon}^{\infty} \rho_d(t) dt := c(d, \varepsilon) > 0.$$

<sup>5</sup>Stricto sensu, the convergence (A.1.3) applies to the spectral measure of the adjacency matrix of  $G_n$ , but here the renormalization  $P = d^{-1}A$  is trivial, so  $\mu_{P_n}$  converges to the measure  $\mu_{\mathbb{T}_d} \circ d$ , which is of course the spectrum of the random walk operator on  $\mathbb{T}_d$ .

The proportion of eigenvalues of  $G_n$  that are greater than  $2\sqrt{d-1}/d - \varepsilon$  is therefore strictly positive (a result due to Jean-Pierre Serre, [130]), which is much stronger than  $\lambda_* > 2\sqrt{d-1}/d - \varepsilon$ . However, convergence (A.1.3) does not allow us to say more: we could very well have  $\lambda_*(P_n) \rightarrow 1$ , for example. We will see that this is not the case.

### Ramanujan graphs and the Alon-Friedman theorem

Given the importance of the second eigenvalue and the lower bound of Alon and Boppana, it is natural to focus on the extremal graphs for  $\lambda_2$ , those with a second eigenvalue smaller than  $2\sqrt{d-1}/d$ . A connected  $d$ -regular graph is called *Ramanujan* when the eigenvalues of  $P$  are either 1 or less than  $2\sqrt{d-1}/d$ , as in the following Figure.

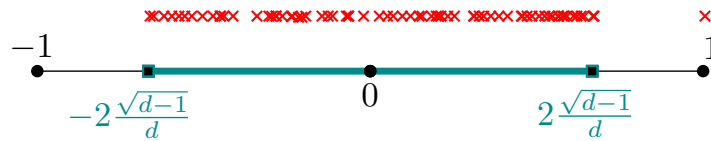


Figure A.3 – In red, the spectrum of a Ramanujan graph. When  $-1$  is eigenvalue, we often say that  $G$  is a bipartite Ramanujan graph.

These graphs are optimal expanders in view of the previous section, and we will not dwell on the incredible breadth of their applications, for which we refer to the influent syntheses [66] and [87], or more recently [124]. It should be noted, however, that the question of their existence has been proven to be highly non-trivial.

The construction of Ramanujan graph families is a difficult problem, which was first solved for some values of  $d$  ([111, 107], for cases where  $d - 1$  is a prime integer): one can construct Ramanujan graphs as Cayley graphs of  $\text{PSL}_2(\mathbb{F}_q)$  from a well chosen set of generators<sup>6</sup> see [66] for an accessible presentation. The existence of  $d$ -regular Ramanujan graphs has finally been shown for all  $d > 2$  ([108, 109, 110], bipartite Ramanujan graph constructions); however, it should be noted that the latter works are purely existential in nature and do not allow to explicitly construct these Ramanujan graphs.

These “constructions” represent a *tour de force*, but here the probabilistic method has made it possible to bypass the problem in a spectacular way. Indeed, in the late 1980s, Alon conjectured that *almost all regular graphs are almost Ramanujan*. More precisely, he conjectured ([9]) that if  $G_n$  is chosen uniformly at random on  $\mathcal{G}(n, d)$ , then

$$\lambda_* \xrightarrow[n \rightarrow \infty]{\mathbf{P}} 2 \frac{\sqrt{d-1}}{d}. \tag{A.2.7}$$

Obviously, (A.2.7) does not solve the question of the existence of Ramanujan graphs, insofar as we can have sequences  $G_n$  that satisfy  $\lambda_*(G_n) \rightarrow 2\sqrt{d-1}/d$  while not being Ramanujan. However, for applications, there is not so much difference, since such a sequence is asymptotically optimal in the sense that it reaches the lower bound of Alon-Boppana.

The first proof of (A.2.7) dates back to 2004 ([79]), in a now famous paper of Friedman; a second proof, due to Bordenave [33], introduced a powerful high-trace method to study the large eigenvalues of some matrices. The first and last chapter of this thesis are based on this method.

### Directed regular graphs

A digraph is said to be *regular* when all its vertices have the same degrees (in and out):  $d_i^+ = d_i^- = d$ . From now on, we will note  $\mathcal{G}_{\rightarrow}(n, d)$  the set of all directed  $d$ -regular graphs with  $n$  vertices.

<sup>6</sup>The proof that these graphs are Ramanujan uses number theory tools due in particular to Ramanujan, hence their name.

As mentioned in the beginning of this introduction, it would be misleading to believe that directed graphs are simply a variant of simple graphs. The entire previous section is based on two tools: first, the Alon-Boppana bound, and second, the convergence of  $\mu_{P_n}$ . However, to date, there is no satisfactory equivalent of either in the context of directed graphs. One of the reasons is that the transition matrix  $P_n$  is no longer hermitian: its eigenvalues are complex numbers, and there are no variational characterizations. However, the theory of *directed* expanders is attracting increasing attention in various fields, as evidenced by the recent synthesis [124].

Let us start by examining a possible directed analogue of the Alon-Boppana inequality. The Benjamini-Schramm convergence theory is still valid for directed graphs, and it is not difficult to convince oneself that if  $G_n$  is a uniform directed  $d$ -regular graph, then  $G_n \rightarrow \vec{\mathbb{T}}_d$ , the  $d$ -regular directed tree: each vertex has  $d$  in-neighbors and  $d$  out-neighbors. The limitation lies in the definition of the spectral measure of  $\vec{\mathbb{T}}_d$ , since the transition operator on this tree is not self-adjoint. However, we do know that the spectrum of this operator, calculated in [67], is equal to  $\{|z| \leq 1/\sqrt{d}\}$ . In fact, the shape of the limiting spectral measure for  $\mu_{A_n}$  is still conjectured; it is assumed but not proved that the limiting measure, supported on  $D(0, 1/\sqrt{d})$ , is given by ([40, page 70]) :

$$\frac{1}{\pi} \frac{d^2(d-1)}{(d^2 - |z|^2)^2} \mathbf{1}_{|z| < \sqrt{d}} dz. \tag{A.2.8}$$

This is *not* the radialization of the Kesten-McKay distribution (see Figure A.4).

Such convergence would immediately imply a directed analog of the bound (A.2.5) (which is itself weaker than Alon-Boppana), but it is unlikely that this method can be generalized to other models — the computation of the spectral measure of operators on directed Galton-Watson trees is not obvious (the right notion of spectra for non-self-adjoint operators is the Brown measure), and similarly the convergence of the spectrum of digraphs towards the Brown measure of their limit in the Benjamini-Schramm sense seems for the moment out of reach.

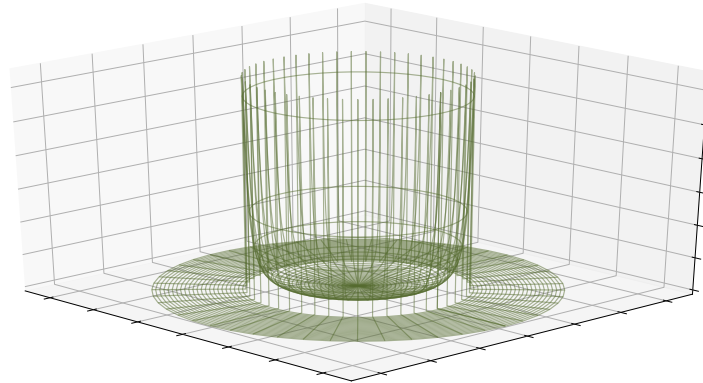


Figure A.4 – An overview of the oriented Kesten-McKay density for  $d = 13$ . When  $d \rightarrow \infty$ , this distribution converges (after renormalization by  $\sqrt{d}$ ) towards the circle, i.e. the uniform distribution on  $D(0, 1)$ . When  $G_n$  is uniform on  $\vec{\mathcal{G}}(n, d)$  and  $d_n \rightarrow \infty$ , the convergence of  $\mu_{A_n}$  towards the circle distribution is now proven ([58]).

### Ramanujan digraphs

A  $d$ -regular digraph is a *Ramanujan digraph* if its transition matrix  $P$  has  $\lambda_* \leq 1/\sqrt{d}$ . There is a conjectured analog of (A.2.7) and of Friedman’s theorem on those graphs (eqn. 5.4 in [124]) : for any

sequence of rv  $(G_n)$  where  $G_n$  is uniformly distributed over  $\mathcal{G}_{\rightarrow}(n, d)$  and with transition matrix  $P_n$ , we have

$$\lambda_{\star}(P_n) \rightarrow \frac{1}{\sqrt{d}} \quad (\text{A.2.9})$$

in probability as  $n \rightarrow \infty$ . Note that whp,  $G_n$  is connected, hence  $\lambda_{\star} = |\lambda_2|$ . The upper bound is proven in this thesis: it is the directed analog of Alon-Friedman's theorem.

**Theorem 21.** *Let  $d \geq 3$ , and let  $G$  be uniformly distributed over  $\mathcal{G}_{\rightarrow}(n, d)$ , with transition matrix  $P$ . Then, for any  $\varepsilon > 0$  we have*

$$\lambda_{\star}(P_n) \leq \frac{1}{\sqrt{d}} + \varepsilon \quad (\text{A.2.10})$$

with probability going to 1 as  $n \rightarrow \infty$ .

The lower bound, the directed analog of Alon-Boppana, is not proven up to this day. As explained in the preceding paragraph, a probabilistic analog seems to be true, but a deterministic analog of (A.2.7) cannot be true: some nontrivial  $d$ -regular digraphs such as De Bruijn graphs have all their nonzero eigenvalues with modulus 1 ([68]).

The result I proved in my first paper is indeed considerably stronger than Theorem 21, for it applies to *every graph with a fixed degree sequence*.

### Graphs with prescribed degrees

Regular graphs have a very strong form of homogeneity, in the sense that each vertex has an identical behaviour in terms of degrees. In many applications, especially those that study graphs from real networks, vertices have different degrees, and often these degrees are accessible: it is easier to have a list of the degrees of a network, than all the connections in the network. To model such networks, we use models of graphs *with prescribed degrees*; we refer to [119, 60] and their bibliography for many examples from applied mathematics.

In such a model, the degrees (incoming and outgoing) are given: we know the list  $\mathbf{d} = (d_1^+, d_1^-, \dots, d_n^+, d_n^-)$  of the degrees, and we choose  $G$  uniformly at random from the set  $\mathcal{G}_{\rightarrow}(n, \mathbf{d})$  of digraphs which have this sequence of degrees. The question of whether  $\mathcal{G}_{\rightarrow}(n, \mathbf{d})$  is not empty, i.e. whether there are graphs with this degree sequence<sup>7</sup> is an old and interesting question, to which we will return in the second part of this work. This question is actually secondary due to a famous technology, the *configuration model* (see [32], for a presentation in the context of non-directed graphs).

The problem with the uniform model on  $\mathcal{G}_{\rightarrow}(n, \mathbf{d})$  (but also on  $\mathcal{G}(n, d)$ ) is that there is no easy and manageable method to generate a uniform random variable on these sets<sup>8</sup>. Rather than doing this, we prefer to use multi-graphs instead of graphs, in the sense that  $(i, i)$  loops and multiple edges are allowed. For any sequence of integers  $\mathbf{d}$  satisfying  $d_1^+ + \dots + d_n^+ = d_1^- + \dots + d_n^- := m$ , such a multigraph exists, and there is a simple way to generate it: we paste  $d_i^-$  incoming half edges to vertex  $i$  and  $d_i^+$  outgoing half edges, and then for each of the  $m$  outgoing half edges, we randomly glue them to one of the available incoming half edges. The digraph  $G$  is then obtained by identifying each pair of half edges glued to a single directed edge. This yields a multi-graph which is uniform on the set  $\mathcal{M}_{\rightarrow}(n, d)$  of all directed multi-graphs with  $\mathbf{d}$  as degree sequence.

However, the uniform model on  $\mathcal{M}_{\rightarrow}(n, d)$  has a form of absolute continuity with respect to the uniform model  $\mathcal{G}_{\rightarrow}(n, \mathbf{d})$ . More precisely, one can show two things: if  $G$  is uniform on  $\mathcal{M}_{\rightarrow}(n, d)$ , then

1. there is a constant  $c > 0$  depending only on  $\max \mathbf{d}$  (and not  $n$ ) such that if  $n$  is large enough,  $\mathbf{P}(G \in \mathcal{G}_{\rightarrow}(n, \mathbf{d})) \geq c$ ,

<sup>7</sup>In this case, we say that  $\mathbf{d}$  is a digraphical sequence.

<sup>8</sup>No one even knows their cardinal in general, and it is unlikely that it will have any closed or explicit form. Only asymptotics are known.

2. the distribution of  $G$  conditioned on being in  $\mathcal{G}_{\rightarrow}(n, \mathbf{d})$  is the uniform distribution on  $\mathcal{G}_{\rightarrow}(n, \mathbf{d})$ .

Therefore, if we have a sequence of  $(\mathbf{d}^{(n)})$  of degrees, with  $\mathbf{d}^{(n)}$  of length  $n$ , and  $\max \mathbf{d}^{(n)} \leq \Delta$  for some integer  $\Delta$  not depending on  $n$ , then we have the following property: for any event  $E$ , we have

$$\lim_{n \rightarrow \infty} \mathbf{P}_{\mathcal{M}_{\rightarrow}(n, \mathbf{d})}(G \in E) = 0 \implies \lim_{n \rightarrow \infty} \mathbf{P}_{\mathcal{G}_{\rightarrow}(n, \mathbf{d})}(G \in E) = 0.$$

In other words, all asymptotically true properties for the multigraph model will be true for the corresponding digraph model. The model on multigraphs is therefore much more general than the model on graphs, at least regarding the asymptotic behaviour.

### Alon's theorem, general formulation

The main result of the first chapter of this thesis is the following theorem, which is the directed analog of (A.2.7) when  $P_n$  is the transition matrix of a digraph with a fixed degree sequence.

**Theorem 22** ([62]). *Let  $\mathbf{d}^{(n)} = (d_1^+, \dots, d_n^+, d_1^-, \dots, d_n^-)$  be a sequence of integers with  $d_1^- + \dots + d_n^- = d_1^+ + \dots + d_n^+ := m$ , and let  $G_n$  be the random multigraph obtained via the configuration model on  $\mathcal{M}_{\rightarrow}(n, \mathbf{d}^{(n)})$ . We assume that there are two constants  $3 \leq \delta \leq \Delta < \infty$  such that for all  $n$ ,*

$$\delta \leq \min \mathbf{d}^{(n)} \leq \max \mathbf{d}^{(n)} \leq \Delta. \tag{A.2.11}$$

Then, for every  $\varepsilon > 0$ ,

$$|\lambda_{\star}(P_n)| \leq \max \left\{ \frac{1}{\delta}, \sqrt{\frac{1}{m} \sum_{i=1}^n \frac{d_i^-}{d_i^+}} \right\} + \varepsilon \tag{A.2.12}$$

with a probability going to 1 when  $n \rightarrow \infty$ .

Thanks to the previous remarks, the result also holds when  $\mathbf{d}$  is digraphic and  $G$  is taken uniformly from  $\mathcal{G}_{\rightarrow}(n, \mathbf{d})$ . When the degree sequence is constant, we find back Theorem 21, since then  $d_i^{\pm} = d$  for all  $i$ . Note that if  $d_i^+ = d_i^-$  for any  $i$  (the graph is then Eulerian), the resulting bound is  $\sqrt{n/m} = 1/\sqrt{d_{\text{mean}}}$ , the inverse root of the average degree, a result to be related to a similar bound in the stochastic blockmodel shown in [43], and for Erdős-Rényi graphs in the last chapter of this thesis).

The presence of a possible eigenvalue with modulus close to  $1/\delta^{-1}$  remains uncertain; the proof does not allow to get rid of the maximum in (A.2.12) and numerical simulations seem to indicate that in the regime where this maximum is  $1/\delta$ , there is indeed an eigenvalue with modulus close to  $1/\delta$  with probability *a priori* not zero, as in Figure A.5b. The question of whether the limit measure actually has an atom in  $\delta^{-1}$  remains open.

## A.3 Digression : the degree sequence

In the preceding section, we studied the spectral properties of some graphs with a given degree sequence, say  $d_1, \dots, d_n$ . Knowing if there exists some graph with this specific degree sequence was secondary, because the configuration model allowed us to bypass this hypothesis. However, the question is interesting by itself: what are the conditions for an integer sequence to be the degree sequence of a simple graph? Such sequences are called *graphic sequences*.

### The Erdős-Gallai theorem

The question was solved by Erdős and Gallai in a famous 1960 paper:

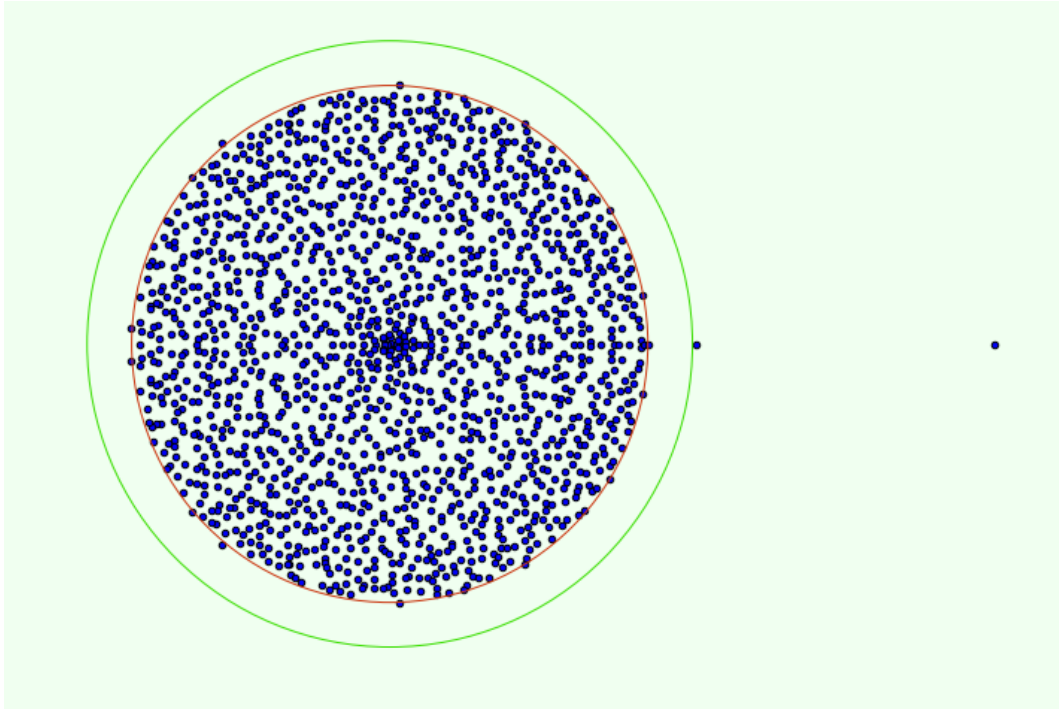
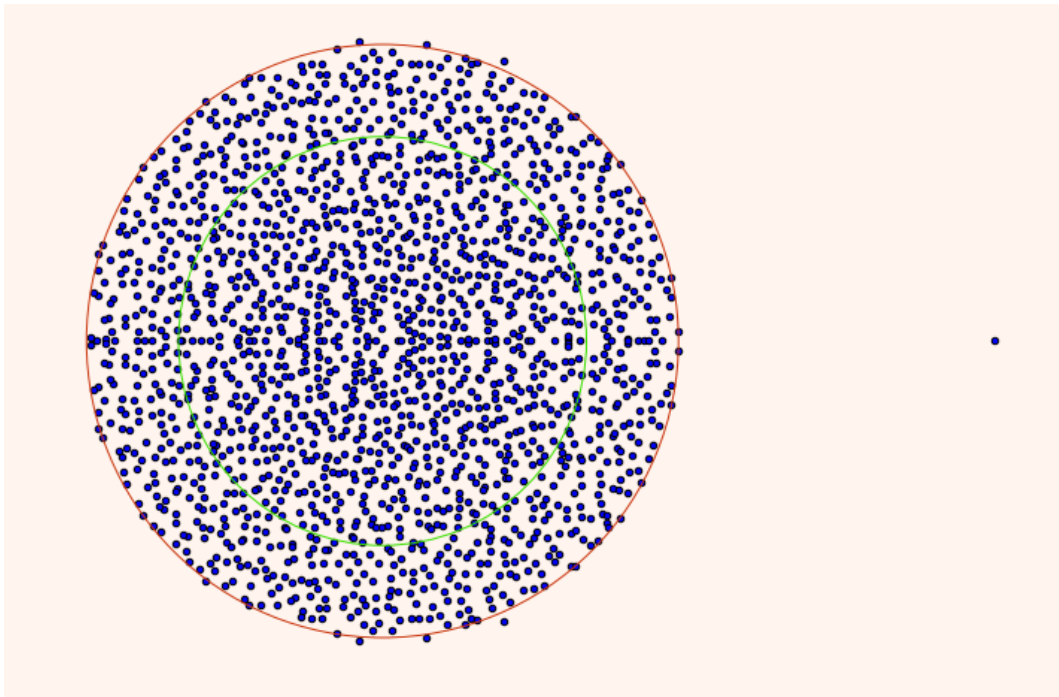
(a) Un exemple du cas où  $\tilde{\rho} = \delta^{-1}$ .(b) Un exemple du cas où  $\tilde{\rho} = \rho$ .

Figure A.5 – Two realizations of the spectrum of  $P$  for a graph  $G$  from the directed configuration model. The red circle has radius  $\rho$  and the green circle has radius  $1/\delta$ . The isolated point on the right is 1, the Perron-Frobenius eigenvalue.

- In the figure (a) there is  $n = 1600$  vertices, of which 700 of degree  $(2, 2)$  and 800 of degree  $(9, 9)$ , so that in this case  $\tilde{\rho} = \delta^{-1} = 1/2$ .
- In (b), we have  $n = 1800$  vertices, including 600 of degree  $(5, 6)$ , 600 of degree  $(3, 7)$  and 600 of degree  $(9, 4)$ . In this case, we have  $\tilde{\rho} = \rho$ .



**Theorem 23** (Erdős-Gallai, [73]). *Let  $\mathbf{d} = (d_1, \dots, d_n)$  be an integer sequence in decreasing order. There is a simple graph  $G$  with degrees given by this sequence if and only if those two conditions are met:*

$$d_1 + \dots + d_n \text{ is even,} \tag{A.3.1}$$

and if for any  $k \in \{1, \dots, n\}$  we have

$$d_1 + \dots + d_k \leq k(k-1) + \sum_{i=k+1}^n \min(d_i, k). \tag{A.3.2}$$

There are many proofs of this theorem, the shortest being the one of Tripathi and Vijay ([138]); there are also other characterizations of graphic sequences, all equivalent to this one; a list can be found in [131]. In the setting of directed graphs, there are also similar criteria, notably listed in [28]. The second result presented in this thesis is a generalization of those results and an extension of the “degree sequence” concept to whole neighborhoods.

### The universal covering of a graph

We can see the degree of a vertex  $x$  as a characterization of its immediate neighborhood: more precisely, the degree of  $x$  is  $d$  if and only if  $x$  has  $d$  neighbors. However, as part of the Benjamini-Schramm convergence, we need to study neighbourhoods deeper than 1 ([35]). To do this, it is necessary to introduce the concept of universal covering.

Let  $G = (V, E)$  be a finite graph. A graph  $G' = (V', E')$  is a covering of  $G$  if there is a surjection  $p : V' \rightarrow V$  which is a local isomorphism, in the sense that for any vertex  $x$ , the application  $p$  is a bijection between the edges in  $G$  incident to  $x$  and the edges in  $G'$  incident to  $p(x)$ . It is the analog of the notion of *lifting* from topology.

A connected graph has a particular covering, which is canonical in some sense: the universal covering. This is the only covering of  $G$  that is a tree; it is noted  $T_G$ . This covering has two essential properties:

- It is unique up to isomorphism,
- It is a covering of all the coverings of  $G$ .

The universal covering of a graph can be seen as the graph  $G$  *unfolded*, i.e. in which each cycle is forgotten while maintaining the adjacency of the vertices. Let us give some examples to clarify the ideas.

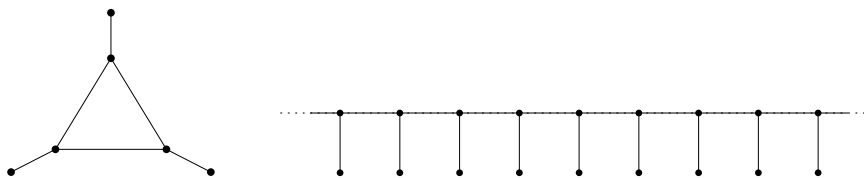


Figure A.6 – An example of universal covering: the infinite comb on the right is the universal covering of the graph in the left.

- The universal covering of any  $d$ -regular graph (with  $d > 1$ ) is  $\mathbb{T}_d$ ;
- the universal covering of a cycle is a bi-infinite line;
- the universal covering of the graph in the left of Figure A.6 is the infinite comb in the right of A.6.

The universal covering of a graph carries a lot of information about it, because it gives a good local approximation — especially when  $G$  does not have many small cycles. In theoretical computer science, universal coverings have led to many advances in parallel computing ([13], [14]). In probability, it is of paramount importance to study processes on graphs that are locally tree-like ([84], [12]).

**The  $h$ -neighborhoods**

Let  $G$  be a connected graph and  $T$  its universal covering. Let  $x$  be a vertex of  $G$ . Its  $h$ -universal covering neighborhood is the (isomorphism class of) the ball  $B_T(y, h)$  where  $y$  is any antecedent of  $x$  by  $p$ ; one will easily check that this does not depend on the chosen antecedent. This  $h$ -neighborhood can be seen as the true neighborhood of  $x$  in  $G$ , but in which the cycles have been “unfolded”. Given  $T$  we thus have access to the list  $(t_1, \dots, t_n)$  of the universal  $h$ -neighborhoods in  $G$ . Is the converse also true? To put it differently, if one takes a sequence  $\mathbf{t} = (t_1, \dots, t_n)$  of unlabelled rooted trees with depth  $\leq h$ , how can we decide if this sequence is the sequence of the  $h$ -neighborhoods in a real graph  $G$ ?

When  $h = 1$ , this is exactly the problem solved by Erdős and Gallai. Indeed, an unlabelled rooted tree with depth 1 is nothing more than the number of the leaves at the root. If we note  $t(d)$  the tree with  $d + 1$  vertices and  $d$  leaves, then any graph  $G$  whose 1-neighborhoods are  $(t(d_1), \dots, t(d_n))$  is in fact a graph whose degree sequence is  $(d_1, \dots, d_n)$ , as in Figure A.7.

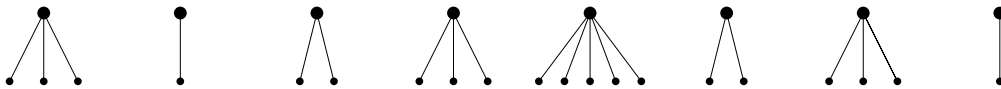


Figure A.7 – Is this collection of trees with depth 1 the collection of 1-neighborhoods in a simple graph with 8 vertices? This is exactly the same thing as telling if  $\mathbf{d} = (3, 1, 2, 3, 5, 2, 3, 1)$  is a graphical sequence.

The data of  $h$ -universal coverings in a graph can be used to generate random graphs with these neighborhoods. For instance, suppose that  $(t_1, \dots, t_n)$  is a sequence of trees. In [35], the authors use a variant of the configuration model to construct multi-graphs with this sequence as the sequence of universal  $h$ -neighborhoods; their construction is similar to the configuration model, with half-edges that are glued together. We generalized this construction and answered the question of existence of true graphs with a given sequence of universal neighborhoods.

**A characterization**

The question that arises is therefore the following. We note  $\mathcal{T}_h$  the set of unlabelled rooted trees with a maximum depth of  $h$ .

**QUESTION A.3.1.** *Let  $\mathbf{t} = (t_1, \dots, t_n)$  be a  $n$ -uplet of elements of  $\mathcal{T}_h$ . Is it the  $n$ -uplet of the universal  $h$ -neighborhoods of a simple graph  $G$ ?*

Before answering the question, we need some definitions — everything will be clearer if we refer to Figure A.8.

Let  $t$  be a rooted tree with root  $\bullet$  and  $e = (\bullet, x)$  be an edge adjacent to the root. The deletion of  $e$  separates  $t$  into two connected components, one containing the root noted  $r'$ , and the other not containing it, noted  $s$ . Vertices at depth  $h$  are removed from  $r'$ . The resulting tree, rooted at  $\bullet$ , is an element of  $\mathcal{T}_{h-1}$  that will be noted  $r$ . The other tree,  $s$ , rooted at  $x$ , is also an element of  $\mathcal{T}_{h-1}$ . We will then say that the *type* of  $e$  is the couple  $\tau = (r, s)$ , and the opposite type of  $e$  is defined as  $(s, r) = \tau^{-1}$ . It is possible to decompose all types of edges adjacent to the root into three separate sets,  $\Delta, A, B$ :

- $\Delta$  is the set of diagonal types, i.e. having the form  $(r, r)$ ;

- $A \cup B$  is the set of non-diagonal types, and the sets  $A$  and  $B$  are chosen so that if the type  $\tau$  is in  $A$ , then  $\tau^{-1}$  is in  $B$ .

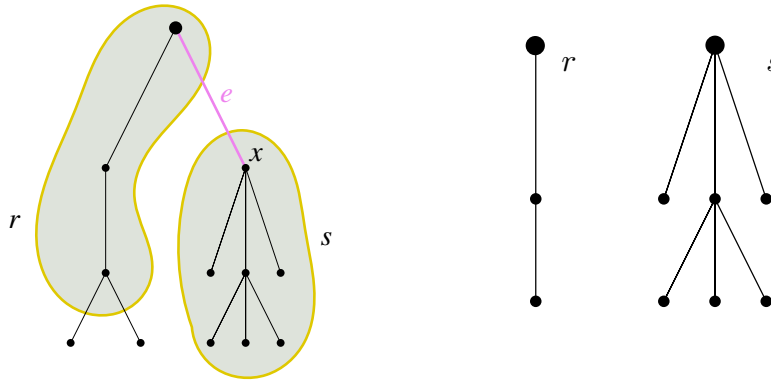


Figure A.8 – Construction of  $\tau(e) = (r, s)$  for the  $e$  ridge in the  $t$  tree.

If  $\tau$  is the type of one of the edges of one of the trees in  $\mathbf{t}$ , we note  $d_i^\tau$  the number of edges  $e$  incident at the root of  $t_i$  and such that  $\tau(e) = \tau$ , i.e. the “degree in  $\tau$  of  $t_i$ ”. Finally, we put  $N_\tau = d_1^\tau + \dots + d_n^\tau$ . The following theorem is the main result of [41], a work in collaboration with Charles Bordenave. The indices in parenthesis mean that the double sequence  $(d_i^\tau, d_i^{\tau^{-1}})$  has been ordered in decreasing lexicographical order.

**Theorem 24.** *Let be  $\mathbf{t} = (t_1, \dots, t_n)$  a  $n$ -tuple of elements of  $\mathcal{T}_h$ . It is graphical if and only if*

1. *for any  $\tau \in \Delta$ , the number  $N_\tau$  is even and if for any  $k$  we have*

$$\sum_{i=1}^k d_{(i)}^\tau \leq k(k-1) + \sum_{i=k+1}^n \min(d_{(i)}^\tau, k), \quad (\text{A.3.3})$$

2. *for any  $\tau \in A$ , we have  $N_\tau = N_{\tau^{-1}}$ , and for any  $k$  we have*

$$\sum_{i=1}^k d_{(i)}^\tau \leq \sum_{i=1}^k \min(d_{(i)}^{\tau^{-1}}, k). \quad (\text{A.3.4})$$

### Reconstruction problems

Theorem 24 was proven in the particular case  $h = 2$  by [19, 16]. One of the authors’ motivations is the general graph reconstruction problem: let  $G$  be a graph with some property  $\mathcal{P}$ . Is it possible to reconstruct  $G$  from the simple data of  $\mathcal{P}$ ? More generally, is it possible to determine if there is a graph verifying  $\mathcal{P}$ , and if so, is such a graph unique?

When  $\mathcal{P}$  is the property of having a given degree sequence, the Erdős-Gallai theorem answers the existence part of the question; the uniqueness part was solved later (see [16], theorem 4.2 and references therein).

One of the reasons why question A.3.1 can be solved is because we only considered neighbourhoods in the universal covering, i.e. we forget the cycles. The same question in which we replace these universal neighbourhoods with the real ones, where cycles are preserved, seems really out of reach. Let us quote, for example, the famous Kelly-Ulam reconstruction conjecture: starting from a graph  $G$  with  $n$  vertices, we note  $g_i$  the isomorphism class of the graph  $G$  deprived of its vertex  $i$ . Is it possible to find  $G$  from the knowledge of  $(g_1, \dots, g_n)$ ? This question seems quite easy; it was asked by Ulam in the 1940s, but it remains unresolved.

## A.4 Spectra of trees

We come back in this section to the simplest model of random graphs: the Erdős-Rényi model and its adjacency matrix. As mentioned earlier, if  $G_n \sim \text{ER}(n, d/n)$ , then  $G_n$  converges in the Benjamini-Schramm sense towards  $\text{PGW}(d)$  and the empirical spectral measure of the adjacency matrix  $A_n$ , noted  $\mu_{A_n}$ , converges as in (A.1.3) to a limit measure  $\mu_{\text{PGW}(d)}$ . The goal of this section is to describe the nature of the spectrum (continuous part, atoms); we will state two new results and a few questions.

As promised, we start by giving a precise definition of the measure  $\mu_{\text{PGW}(d)}$ , and more generally of  $\mu_\rho$  when  $\rho$  is the measure of certain rooted trees: the key notion here is *unimodularity*.

### Unimodularity

In the Benjamini-Schramm convergence, the uniform rooting of finite graphs is a powerful idea with deep consequences. By taking the root uniformly at random, one studies the graph from the viewpoint of a *typical* vertex, ruling out the possible extravagant behaviour of a few vertices. If a sequence of finite graphs  $(G_n)$  converges to  $(G, \circ)$ , its limit will exhibit some flavor of vertex uniformity: for instance, the distribution of the rooted graph  $(G, \circ)$  will be invariant by re-rooting. This is called *unimodularity* ([8, 7, 27, 34]).

Let us first introduce the set  $\mathcal{G}_{**}$  of (isomorphism classes of) bi-rooted graphs, i.e. triplets  $(G, \circ, w)$  where  $v, w$  are two vertices of  $G$ . We endow this space with the topology of local distance, as for  $\mathcal{G}_*$ . We say that a distribution over  $\mathcal{G}_*$  is unimodular if for any measurable function  $f : \mathcal{G}_{**} \rightarrow \mathbb{R}$ ,

$$\mathbf{E}_\rho \left[ \sum_{w \in V} f(G, \circ, w) \right] = \mathbf{E}_\rho \left[ \sum_{w \in V} f(G, w, \circ) \right], \tag{A.4.1}$$

where  $\mathbf{E}_\rho$  means that the random rooted graph  $(G, \circ)$  has distribution  $\rho$  under  $\mathbf{E}_\rho$ . When  $G = (V, E)$  is finite, the random variable  $(G(v), v)$  with uniform root  $v \in V$  is unimodular, and the Benjamini-Schramm limits of unimodular graphs are still unimodular, by taking the limit under (A.4.1). As a consequence, every distribution  $\rho$  on  $\mathcal{G}_*$  which is a Benjamini-Schramm limit of a sequence of finite graphs<sup>9</sup> is unimodular. The converse is not known to this day (see [7, 27] for many interesting developments on this question).

A **unimodular Galton-Watson tree** with progeny  $\pi = (\pi_n)_{n \geq 0}$  is a random rooted tree in which the number of children of the root has distribution  $\pi$ , and the number of children of the other vertices has distribution  $\hat{\pi}$ , which is the size-biased version of  $\pi$ :

$$\hat{\pi}_n := \frac{(n+1)\pi_{n+1}}{\sum_{k=0}^{\infty} k\pi_k}.$$

Distributions such that  $\hat{\pi} = \pi$  are the Poisson distributions, which generate the prototypical example of GW trees, namely Poisson Galton-Watson trees  $\text{PGW}(d)$ , who naturally appear as the limit of diluted Erdős-Rényi graphs  $\text{ER}(n, d/n)$ .

When  $\pi = \delta_d$ , we have  $\hat{\pi} = \delta_{d-1}$  and in this case  $\text{UGW}(\delta_d)$  is the infinite  $d$ -regular tree  $\mathbb{T}_d$ , the limit of uniform  $d$ -regular graphs.

### Spectral measure of unimodular graphs

Let  $G = (V, E)$  be a locally finite graph over a countable vertex set  $V$ . Its adjacency operator  $A$  is the operator defined on the Hilbert space

$$\mathcal{H} = \ell^2(V) = \left\{ f : V \rightarrow \mathbb{C}, \sum_{v \in V} |f(v)|^2 < \infty \right\}$$

<sup>9</sup>Such distributions are dubbed *sophic*.

by

$$\langle e_x, A e_y \rangle = \mathbf{1}_{(x,y) \in E} \quad (\text{A.4.2})$$

where  $(e_x : x \in V)$  is the canonical basis of  $\mathcal{H}$ . The domain  $D_A$  of  $A$  is the dense set of finitely supported vectors  $D_A = \text{vect}(e_x : x \in V)$ . When  $G$  is finite,  $A$  is a symmetric matrix, and there are no problems in defining its spectrum. When  $G$  is not finite, difficulties appear. If its degrees are uniformly bounded by a same integer  $\Delta$ , then  $A$  is a bounded self-adjoint operator, and can be extended to a bounded self-adjoint operator on the whole space  $\mathcal{H}$ , so here again the spectrum is well-defined through classical spectral theory.

However, many graphs of interest (such as Galton-Watson trees) do not have bounded degrees, and there are also trees with are not self-adjoint ([118]).

Luckily enough, unimodular measures are concentrated on trees that are essentially self-adjoint ([34, Prop. 2.2]), for which the notion of spectrum can easily be defined. More precisely, the spectral theorem shows that for any rooted graph  $(G, o)$  whose adjacency operator is essentially self-adjoint, there is a unique Borel probability measure  $\mu_{(G,o)}$  on  $\mathbb{R}$  such that

$$\forall z \in \mathbb{C} \setminus \mathbb{R}, \quad \langle e_o, (A - z)^{-1} e_o \rangle = \int_{\mathbb{R}} \frac{1}{z - t} d\mu_{(G,o)}(t) := \mathfrak{s}(z). \quad (\text{A.4.3})$$

The term  $\mathfrak{s}(z)$  is the Stieltjes transform of the measure  $\mu_{(G,o)}$ . One can thus retrieve the measure  $\mu_{(G,o)}$  using the Stieltjes inversion formula; even when this inversion cannot be performed explicitly, one can still obtain extra informations. For instance, the atoms are given by

$$\mu_{(G,o)}(\{\lambda\}) = \lim_{t \rightarrow 0} \frac{1}{\pi} \Im t \mathfrak{s}(\lambda + it).$$

If  $(G, o)$  is a random rooted tree with unimodular distribution  $\rho$ , the measure  $\mu_\rho$  is then defined as

$$\mu_\rho = \mathbf{E}_\rho[\mu_{(G,o)}]. \quad (\text{A.4.4})$$

We refer to [54] for the technical details.

We already saw that  $\text{UGW}(\delta_d) = \delta_{\mathbb{T}_d}$ : in this case,  $(G, o)$  is the deterministic  $d$ -regular tree  $\mathbb{T}_d$ . The LHS of (A.4.3) is a Laurent series whose coefficients are the numbers of closed walks at the root of  $\mathbb{T}_d$ , which can explicitly be computed through combinatorial arguments ([113]). This gives an explicit expression of the Stieltjes transform of  $\mu_\rho$ , and taking the inverse shows that  $\mu_\rho$  is the Kesten-McKay distribution already mentioned.

One can find in [29] many examples of converging sequences of finites trees and their spectra.

### The limiting spectral measure of diluted Erdős-Rényi graphs

We give ourselves a real number  $d > 0$ , and we study the spectrum of  $G$  with distribution  $\text{ER}(n, d/n)$ . We saw in Section A.1 that  $G_n$  converges in the Benjamini-Schramm sense towards  $\text{PGW}(d)$ . Thanks to the spectral continuity (A.1.3), the empirical distribution of the eigenvalues of the adjacency matrix  $A_n$ ,

$$\mu_{A_n} := \frac{1}{n} \sum_{i=1}^n \delta_{\lambda_i},$$

converges to a probability measure  $\mu_{\text{PGW}(d)}$  that will simply be noted  $\mu_d$ , and which is the averaged spectral measure of the Galton-Watson tree as defined in (A.4.4). This convergence can be established without the help of the Benjamini-Schramm machinery ([143, 94]), but it is the latter that provides the representation of  $\mu_{\text{PGW}(d)}$  as in (A.4.3)-(A.4.4). If we note  $F_n, F_{\text{PGW}(d)}$  the cumulative distribution functions of  $\mu_{A_n}, \mu_d$ , then

$$\sup_{t \in \mathbb{R}} |F_n(t) - F_{\text{PGW}(d)}(t)| \rightarrow 0. \quad (\text{A.4.5})$$

As already observed, this convergence implies the convergence of atoms, in the sense that  $\lim \mu_{G_n}(\{\lambda\}) = \mu_d(\{\lambda\})$  for any real number  $\lambda$ . In particular, as

$$\mu_{G_n}(\{0\}) = \dim \ker(A_n)/n,$$

we know the asymptotic behavior of the kernel of  $A_n$ , provided we are able to calculate  $\mu_{\text{PGW}(d)}(\{0\})$ .

Physicists Bauer and Golinelli, by focusing on the kernel of  $G$ , have made several striking conjectures about the behaviour of  $\mu_d$  around zero ([23, 21]). Thanks to the replica symmetry method, they were able to conjecture the exact value of the atom at zero, namely

$$\mu_{\text{PGW}(d)}(\{0\}) = e^{-dx} + dx e^{-dx} + x - 1 \quad (\text{A.4.6})$$

where  $x$  is the smallest solution in  $]0, 1[$  of the equation  $x = e^{-de^{-de^{-dx}}}$ . This formula was rigorously proved later in [44], notably by using recursion formulas on the Stieltjes transform of  $\mu_{\text{PGW}(d)}$ .

Bauer and Golinelli finely noted that the RHS in (A.4.6) is analytical in  $d$  when  $d < e$ , and has a singularity at the point  $d = e$ ; they interpreted this phenomenon as a transition in the nature of the spectral measure, noting on their numerical simulations the appearance of a continuous part around zero when  $d > e$ .

### The Anderson model and the nature of the spectrum

Any Borel measure  $\mu$  on  $\mathbb{R}$  can be split into

$$\mu = \mu_{\text{ac}} + \mu_{\text{pp}} + \mu_{\text{sc}} \quad (\text{A.4.7})$$

where  $\mu_{\text{ac}}$  is absolutely continuous with respect to the Lebesgue measure,  $\mu_{\text{pp}}$  is purely atomic and  $\mu_{\text{sc}}$  is singular with respect to the Lebesgue measure, but has no atoms.

This decomposition is central in the physical theory of operators: if  $H$  is a self-adjoint operator on a Hilbert space, this decomposition carries within it essential properties of  $H$  related to the behaviour of quantum dynamics  $t \mapsto e^{-itH}\phi$ , notably via the RAGE theorem (see ([54, 135] or more specifically [101]).

Physicists have paid considerable attention to these spectral properties in the Anderson model, for which the Hamiltonian is given  $H = A + \lambda V$  with  $A$  the Laplacian and  $V$  a diagonal operator; when  $A$  is the adjacency operator on  $\mathbb{Z}^d$ , we have the classical discrete Anderson model. The existence of a pure point spectrum for such an operator is related to the localization phenomenon, and this is the reason why the decomposition (A.4.7) has attracted so much attention, even when  $A$  is the adjacency operator on a tree ([4], [95], [2]). In the language of physicists, we say that a measure  $\mu$  does not have extended states at a real  $\lambda$  when

$$\lim_{\varepsilon \rightarrow 0} \frac{\mu(] \lambda - \varepsilon, \lambda + \varepsilon[) - \mu(\{\lambda\})}{2\varepsilon} = 0. \quad (\text{A.4.8})$$

Otherwise, it is said to have extended states at  $\lambda$ , which means that it has a continuous part (absolutely continuous or singular continuous) at  $\lambda$ ; for example, if  $\mu = f(x)dx$  with  $f$  measurable, the above limit is  $f(\lambda)$  for almost all  $\lambda$ .

We have already seen that the spectral measure of  $\mathbb{T}_d$  was absolutely continuous, with Kesten-McKay density; however, there are infinite trees whose spectral measure is

- purely atomic: the canopy tree ([5, Proposition A.2]) in particular, to which we will come back later;
- or to the contrary, purely singular continuous; we will find in [49, 48] an explicit construction of such trees from  $\mathbb{T}_d$  and in [5, Section 6] examples of such semi-infinite trees. In the Anderson model, we also refer to [133, 132].

When the tree itself is random (Galton-Watson for example), the question is more delicate. A necessary and sufficient condition for  $\mu_{\text{PGW}(d)}$  to have a continuous part, and therefore extended states, is simply  $d > 1$ , and more generally ([46]) the measure  $\mu_{\text{UGW}(\pi)}$  has a continuous part if and only if  $\pi$  is super-critical. The reason for this is the fact that almost surely, subgraphs which are isomorphic with  $\mathbb{Z}$  appear in  $\text{UGW}(\pi)$ . However, this result does not give any information on the support of the continuous part when it exists, and even less on its shape.

### The Bauer-Golinelli transition

Bauer and Golinelli's precise numerical simulations (see [23, Section 5.1]) led them to assume that the continuous part of  $\text{PGW}(d)$  exhibits a transition at  $d = e$ :

**CONJECTURE 4.** *The measure  $\mu_{\text{PGW}(d)}$  has extended states at 0 if and only if  $d > e$ .*

We can see this phenomenon on Figures A.9a-A.9b-A.9c.

Such a phase transition is not in itself surprising; we already knew that if  $d$  is small enough (less than 1), the measure  $\mu_{\text{PGW}(d)}$  is purely atomic and that there are no extended states at zero. However, when  $d \rightarrow \infty$ , the measures  $\mu_{\text{PGW}(d)}$  converge towards the famous Wigner's semicircle distribution ([89, 137], see Figure A.10), which is absolutely continuous with respect to the Lebesgue measure and whose density  $\rho(t) = \mathbf{1}_{|t| \leq 2} \sqrt{4 - t^2}$  is non-zero on all its support. This does not rigorously prove the appearance of extended states at zero, but it still gives an intuition of the result. What is surprising is rather the exact *location* of this transition at the point  $e \approx 2,718$ .

In the prepublication [63] in collaboration with Justin Salez, we demonstrated Conjecture 4. This result is not specific to the spectral measure of Poisson Galton-Watson trees, but is expressed within the general framework of unimodular Galton-Watson trees  $\text{UGW}(\pi)$  (with  $\pi$  supercritical). To formulate the results, we need the generating function of  $\pi$ , which we will note

$$\varphi(z) := \sum_{n=0}^{\infty} \pi_n z^n.$$

Similarly, we note  $\hat{\varphi}$  the generating function of  $\hat{\pi}$ . Set  $M(t) = \varphi(t) + (1-t)\varphi'(t) + \varphi(1 - \hat{\varphi}(t)) - 1$ , for  $t \in [0, 1]$ . A key element in the proof of (A.4.6) in [44] was the identity  $\mu_{\text{PGW}(d)}(\{0\}) = \max M$  and the fact that there is a unique number  $z_*$  in  $]0, 1[$  such as  $z_* = 1 - \hat{\varphi}(z_*)$ . The two assumptions on  $M$  that are involved in our main theorem are as follows:

1. The function  $M$  has a unique maximum in  $z_*$ .
2.  $M''(z_*) \neq 0$ .

The conditions for the emergence of extended states at zero are described by the following theorem.

**Theorem 25.** *Let  $\pi$  be a measure on  $\mathbb{N}$ , with  $\pi_0 < 1$ .*

- *If  $M$  does not meet condition 1, then  $\mu_{\text{UGW}(\pi)}$  has extended states at zero.*
- *If  $M$  meets conditions 1 and 2, then  $\mu_{\text{UGW}(\pi)}$  does not have extended states at zero.*

The existence of a continuous part at zero is therefore solved, at least in the case of Poisson Galton-Watson trees. The question arises as to whether such results extend to other unimodular trees.

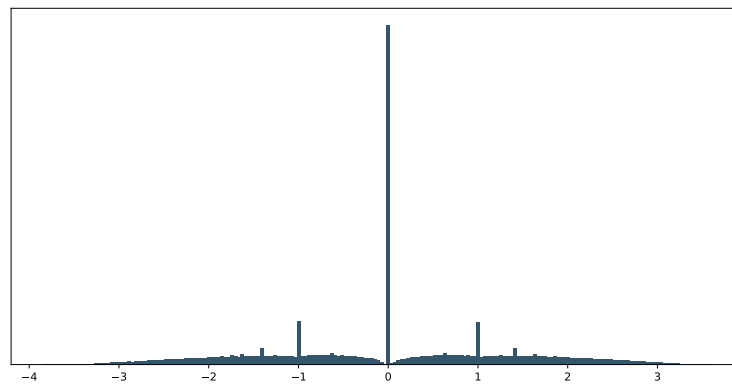
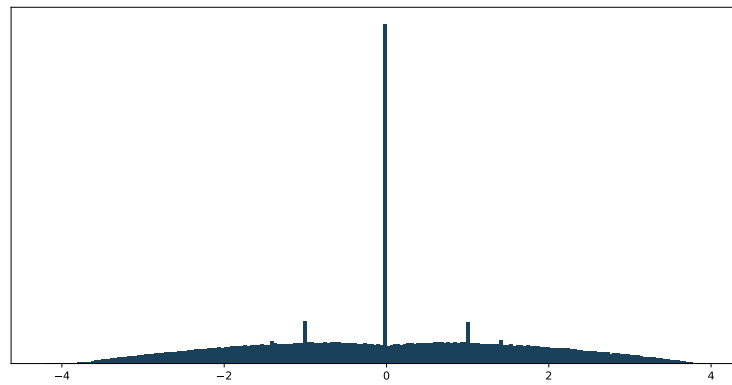
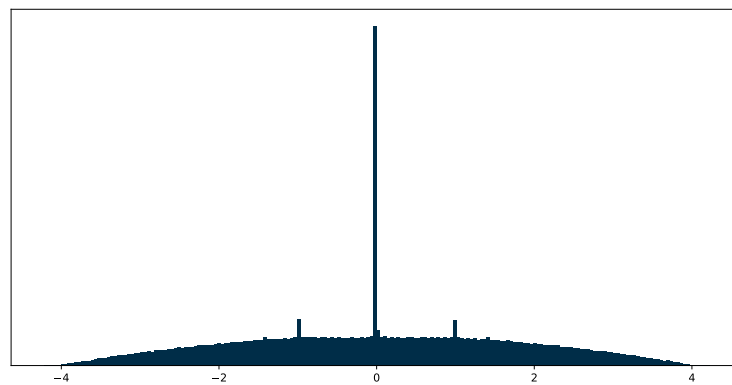
(a)  $d = 2$ (b)  $d = e$ (c)  $d = 3$ 

Figure A.9 – Histograms of eigenvalues of  $G \sim \text{ER}(n, d/n)$  with  $n = 1000$  (on 100 samples). The ‘continuous’ part is zero when  $d < e$ , as predicted by Bauer-Golinelli.



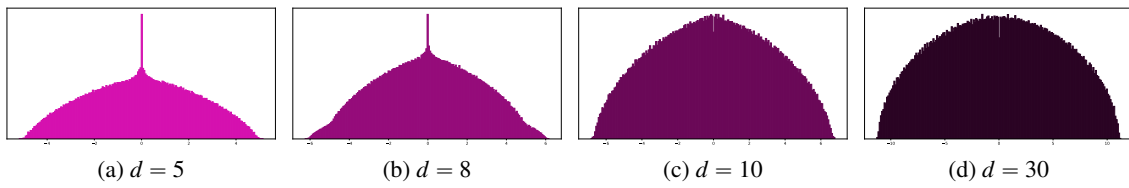


Figure A.10 – Histograms of eigenvalues of  $G \sim \text{ER}(n, d/n)$  with  $n = 1000$  (100 samples). When  $d \rightarrow +\infty$  you have  $\mu_d \rightarrow \mu_{\text{Wigner}}$ .

### The nature of the spectrum in semi-infinite trees

The proof of the existence of a continuous part in the spectrum of Galton-Watson trees ([46]) was based on the fact that, when the progeny  $\pi$  is supercritical, the tree contains a double ray<sup>10</sup> which is a sub-graph isomorphic to  $\mathbb{Z}$ . This is not the case for other unimodular trees such as the skeleton tree  $T_{\text{Squel}}$ , which are almost certainly infinite, but which contain only one ray, i.e. a subgraph isomorphic with  $\mathbb{N}$ : such trees are called *semi-infinite trees* (or single-infinite trees by Aldous ([6, Section 4])). Any semi-infinite tree can be identified with a sequence  $(T_n)$  of finite rooted trees, the root of  $t_n$  being connected to that of  $t_{n+1}$  with a single edge, as in Figure A.1b.

One can easily check that the spectral measure of  $\mathbb{N}$  is indeed the semi-circle distribution; one could therefore bargain for a continuous part in the spectrum of any semi-infinite tree. This is not the case, as shown by the canopy-tree  $(\mathbb{C}_d, o)$ .

**LEMMA A.4.1** (Aizenman, Warzel, [5]). *The spectral measure  $\mathbf{E}\mu_{(\mathbb{C}_d, o)}$  is purely atomic. Its atoms are the eigenvalues of the finite trees  $\mathbb{T}_{n,d} := (\mathbb{T}_d, o)_n$ .*

Semi-infinite trees can be seen as generalizations of Jacobi matrices. In the Anderson model over  $\mathbb{N}$ , each vertex  $n$  has a weight  $V(n)$ , and we have at our disposal numerous criteria for linking the nature of the spectrum of  $H = A + V$  with ergodic properties of  $V(n)$  ([54], [102]). In the case of semi-infinite tree, one can ask the following question:

**QUESTION A.4.2.** *What is the relation between the nature of the spectrum in semi-infinite trees and the growth or ergodicity properties of  $(T_n)$  ?*

### The skeleton tree

The skeleton tree, which will be noted as  $T_{\text{Squel}}$ , is the Benjamini-Schramm limit of uniform trees. More precisely, if  $T_n$  is uniformly distributed among the  $n^{n-2}$  labelled trees with  $n$  vertices, then  $T_n \rightarrow T_{\text{Squel}}$ , a convergence established in [85].

---

<sup>10</sup>*invariant line ensemble*, in the terminology of [46].

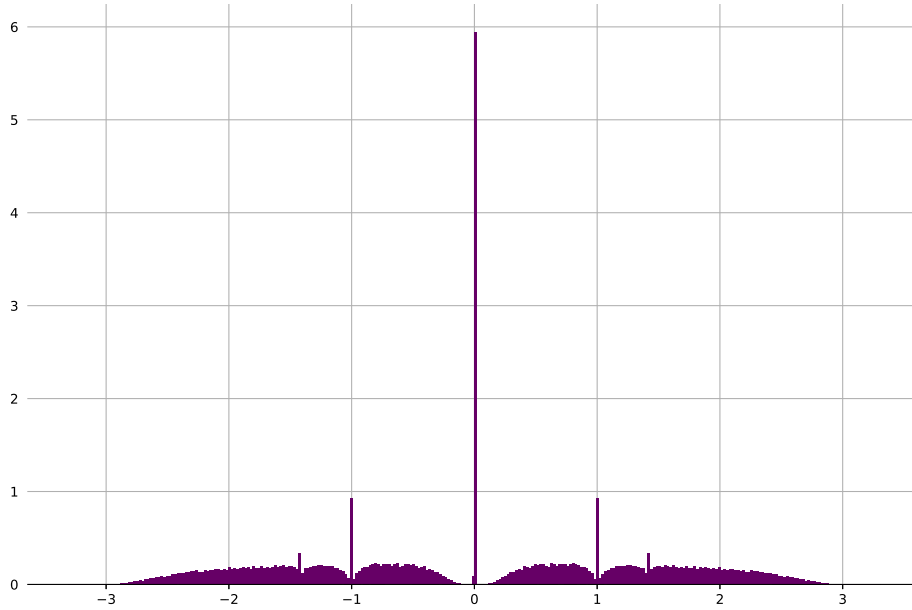


Figure A.11 – Histogram of the eigenvalues of the spectrum of a random tree on 1000 vertices, which gives an approximate idea of  $\mu_{\text{squel}}$ .

Thanks to the continuity of the spectrum for local weak convergence, we have the convergence of the empirical spectral measure  $\mu_{T_n}$  to  $\mu_{\text{squel}}$ . This measure is as poorly known as  $\mu_{\text{UGW}(\pi)}$ ; in fact, the only interesting informations about  $\mu_{\text{squel}}$  are obtained by passing through the limit in  $\mu_{T_n}$ . Let us give two examples.

**Atoms.** By elementary arguments, it is possible to see that *all the eigenvalues of finite trees are atoms of  $\mu_{\text{squel}}$* ; this set is a dense part of  $\mathbb{R}$  and it is equal to the set  $\mathbb{A}$  of the totally real algebraic numbers<sup>11</sup>, a result of [127]. In the simulations (see Figure A.11), we see the presence of atoms at 0,  $\pm 1$  (the spectrum of the tree with only one edge),  $\pm\sqrt{2}$  (the spectrum of the two-edges graph).

The exact value of  $\mu_{\text{squel}}(\{0\})$  is known; this calculation results from a purely combinatorial analysis of  $T_n$ . The link between the size of a tree's kernel and the leaves of this tree has long been known: for every finite tree  $T$ , we have  $\dim\ker T = \min\{k : M_{n-2k} \neq 0\}$  where  $M_j$  is the number of perfect couplings in  $T$  that contain  $j$  edges. Bauer and Golinelli used this link to obtain an *exact* formula for the expected rank nullity of  $T_n$ . More specifically, they prove in [21] that

$$\mathbf{E}[\dim\ker(T_n)] = n \left( 1 - 2 \sum_{m=2}^n \frac{(-1)^m}{m} m \binom{m}{n}^m \binom{n}{m} m \right). \quad (\text{A.4.9})$$

Using tools from analytic combinatorics, they also calculated the generating function of  $\dim\ker(T_n)$  and obtained the asymptotics of  $\mathbf{E}[\dim\ker(T_n)]$  in the form :

$$\lim_{n \rightarrow \infty} \mu_{T_n}(\{0\}) = \mu_{\text{squel}}(\{0\}) = 2x_\star - 1$$

where  $x_\star \approx 0.56714 \dots$  is the unique solution of the equation  $x = e^{-x}$ . This solved the problem of computing the atom at zero for the skeleton tree, just as (A.4.6) solved this problem for the Galton-

<sup>11</sup>A number is *algebraic totally real* if it is the root of a polynomial with integer coefficients, whose roots are all real.

Watson trees. The simulations indicate the absence of extended zero states for  $\mu_{\text{squel}}$  (see Figure A.11). In a short section of Chapter 3, unpublished, I proved this result:

**Theorem 26.** *The spectral measure  $\mu_{\text{squel}}$  does not have extended states at zero.*

The arguments are an adaptation of the techniques already used to prove Conjecture 4. The question of the existence of a continuous part remains unsolved; simulations suggest that  $\mu_{\text{squel}}$  does indeed have one.

## A.5 Eigenvalues of random diluted matrices

The purpose of this section is to study some weighted graphs, i.e. graphs with weights on the edges.

Formally, we give ourselves a weight matrix  $W \in \mathcal{M}_{n,n}(\mathbb{R})$ , and the adjacency matrix  $M$  of a graph, and we want to study the matrix  $P = M \odot W$  where  $\odot$  is the Hadamard product (termwise):  $P_{i,j} = M_{i,j}W_{i,j}$ . Such models have been extensively studied when the underlying graph is relatively sparse (typically when the number of edges of  $G$  is of the order  $n \log(n)$  — we will see why this framework is more studied) or when the whole problem is Hermitian, i.e. when  $W$  is Hermitian and the underlying graph  $G$  is a simple undirected graph; a very recent paper by Tikhomirov and Youssef ([136]) fully describes the behavior of the large eigenvalues of  $P$  when  $W$  is drawn from the GOE and  $G$  is an Erdős-Renyi graph with parameter  $d/n$ , in the regime  $d \rightarrow \infty$ .

In particular, the authors ask the following questions, which we extract *verbatim* from [136, page 3]:

- 1) Is there a threshold (depending on  $d$ ) for the appearance of outliers?
- 2) Is there an explicit expression for this threshold?
- 3) Do we have an explanation for the appearance of these outliers?
- 4) What is the exact asymptotic of these outliers?

In Chapter IV of this thesis, based on a collaboration with Charles Bordenave and Raj Rao Nadakuditi, we answer these four questions in the context of directed graphs, the main practical motivation coming from the statistical problem of matrix completion, which we will present below. The main results (Theorem 27 on page 181 *et seq.*) seem surprising, because they essentially say that the eigenvalues of non-symmetric matrices are actually more interesting than the singular values of non-symmetric matrices — common knowledge<sup>12</sup> usually states that eigenvalues of non-symmetric matrices are very unstable. In the last section, we will return to this recent and promising idea.

### Reconstruction

Reconstruction problems aim at reconstructing an object that is generally complex (a graph, a matrix, a tensor, a periodic function) from a small amount of information about this object, typically a few entries in the matrix, or some Fourier coefficients. These problems have become extremely popular in applied mathematics, particularly through the development of recommendation systems.

We will focus on the problem of matrix reconstruction. The problem is as follows: a matrix  $W$  is hidden. It is known to have certain structural properties (its size  $n \times m$  is known, sometimes information on its rank or entries). However, the observer only has access to the observation of certain entries  $W_{i,j}$  for a small set of indices  $(i, j) \in E \subset [n] \times [m]$  and tries to use this information to find (partially, with a small error, or even exactly) the matrix  $W$ .

The literature on the subject is gigantic and we refer to [65] and its references for a global survey; let us quote Candès, Tao and Candès, Recht [53, 52], Keshavan, Montanari and Oh [92] and Chatterjee [55] for a few landmark papers.

---

<sup>12</sup>... at least, in the world of numerical mathematics.

### Obstructions to the problem and necessary assumptions

It is not always possible to reconstruct a matrix  $W$  from a small number of observations of its entries. The two main obstacles are the sparsity of samples, and the intrinsic complexity of  $W$ .

**The lack of data.** Let's take an example of a very simple reconstruction problem, in which the matrix to be found has rank 1, i.e. it has the shape  $W = xy^*$  with  $x, y \in \mathbb{R}^n$  two unit vectors. Retrieving  $W$  is like retrieving  $x$  and  $y$ , up to a sign or a phase. Suppose that the  $i$ -th line of  $W$  is not observed at all. Then there is no way to access  $x_i$ . In other words, it is necessary that the observer has at least one observation per row and one observation per column if he wants to hope to reconstruct the original matrix  $W$  as accurately as possible. Suppose that the entries are revealed uniformly at random; by the coupon collector principle, to have at least one entry revealed on each line, it is necessary to have at least  $\log n$  observations on each line and therefore  $n \log n$  observations overall, in other words any density  $d \ll \log n$  will necessarily leave entire rows or columns unobserved.

The literature on matrix completion has thus focused on this  $d \geq \log n$  regime. One of the contributions of this work is to demonstrate that not everything is lost when  $d \leq \log n$  or even when  $d$  is fixed, and that it is even possible to obtain an estimator of  $W$  which is well correlated with  $W$ .

**The intrinsic difficulty of the problem.** Not all matrices can be easily reconstructed from a random observation of their inputs. For example, suppose that the matrix  $W$  is all zeroes, except possibly one or two entries, as in the example

$$W = e_1 e_1^* + e_1 e_2^* = \begin{pmatrix} 1 & 1 & 0 & \cdots & 0 & 0 \\ 0 & 0 & 0 & \cdots & 0 & 0 \\ \vdots & & & & & \vdots \\ 0 & & \cdots & & & 0 \end{pmatrix}.$$

If the number of revealed entries is small and their location is sufficiently random, there is no chance that the only two non-zero entries in the above matrix will be observed. We'll only see zeros. It is therefore necessary that the entries of  $W$  are sufficiently spread out in the matrix. Mathematically, this means, for example, that the largest entry of  $W$  has order  $O(1/n)$ , so that the  $L^2$  mass (Frobenius norm) is evenly distributed among the entries of the matrix. This is perfectly equivalent to requiring that the mass of singular vectors is sufficiently distributed among the entries, a condition frequently referred to as *inconsistency condition*.

In the sequel, we will adopt assumptions about  $W$  under which the problem is feasible. To do this, we will define a large class of matrices whose complexity parameters are controlled. Note that these assumptions are standard in the literature; see for example conditions A0-A1 in [52] or A1-A2 in [92]. We refer to the survey [65] on this subject.

We will start by focusing exclusively on positive Hermitian matrices; this is only a shallow restriction, we will come back to it later. We will note

$$W = \sum_{i=1}^r \mu_i \varphi_i \varphi_i^* \tag{A.5.1}$$

where  $\mu_1 \geq \cdots \geq \mu_r$  are the eigenvalues ordered in descending order, the integer  $r$  is the rank, and  $\varphi_i$  are orthonormal unit eigenvectors. The assumptions on  $W$  are as follows:

- (i) The  $W$  rank is less than  $r$ .
- (ii) The eigenvalues of  $W$  are positive.

(iii) There is a real number  $b \geq 1$  such that

$$\max_{i \in [r]} |\varphi_i|_\infty = \frac{b}{\sqrt{n}}. \quad (\text{A.5.2})$$

Finally, in accordance with the general theme of this thesis, our results are valid for any fixed  $d$ , without any restriction; they are therefore the first results on the top eigenvalues of diluted non-hermitian matrices. All our proofs extend to the case where  $d$  grows slowly with  $n$ , typically  $d = n^{o(1)}$ .

### Spectral methods

As a general rule, the observed matrix  $P$  has nearly full rank with high probability; however, a successful idea to reconstruct  $W$  from  $P$  is to calculate its singular value decomposition, say

$$P = \sum_{i=1}^{\text{rank}(P)} \sigma_i(P) x_i y_i^*$$

with decreasing  $\sigma_i$ , and to get rid of all singular values and singular vectors below the true rank  $r$  of  $W$ , i.e. to take

$$\hat{W} = \sum_{i=1}^{\text{rank}(W)} \sigma_i(W) x_i y_i^*.$$

The Eckhart-Young theorem says that  $\hat{W}$  is precisely the best rank- $r$  approximation of  $P$  (in Frobenius norm); the problem is that the singular values of non-Hermitian matrices and the eigenvalues of Hermitian matrices are ‘polluted’ by the highest degrees ([136, 24, 25, 96, 139]).

A deep idea already exploited by Feige and Ofek ([77]) and popularized by Keshavan, Montanari and Oh in their famous paper [92] is to regularize the spectrum by removing the high degrees, i.e. the lines of  $P$  with too many revealed entries, thus obtaining a matrix  $\tilde{P}$ , and only then to truncate at rank  $r$ , obtaining  $T_r(\tilde{P})$ . One of the important results of [92] is that this operation makes it possible to find the singular values of the original matrix; more precisely, under simple hypotheses<sup>13</sup>, the authors show that

$$\frac{1}{n} \|W - T_r(\tilde{P})\| = O(\sqrt{r/d}). \quad (\text{A.5.3})$$

Weyl’s inequalities then show that if  $d \rightarrow \infty$ , we find  $W$ , and that the  $r$  largest singular values of  $\tilde{P}$  are aligned with those of  $W$ ; the equation (A.5.3) is valid at  $d$  fixed, but in this case it does not allow us to determine the asymptotic in  $n \rightarrow \infty$  of the large singular values of  $\tilde{P}$ .

The most interesting aspect of this method is the trimming procedure, where some data are deleted to regularize the spectrum of the observed matrix. The idea behind the following sections is that **the good idea was not to look at the singular values of the non-Hermitian matrix  $P$ , but precisely its eigenvalues, which are directly aligned with the large eigenvalues of  $W$** . This new idea, according to which non-symmetry retains more information than symmetry, seems promising and has recently emerged in an article by Chen et al ([56]). We will come back to this point in the conclusion.

### Asymptotics of large eigenvalues and detection of eigenvectors

The main result we obtained is an exact description of the behavior of the eigenvalues of the matrix  $P = (n/d)M \odot W$ , with  $M$  the adjacency matrix of an Erdős-Rényi directed graph with parameters  $n$  and  $d/n$  — the normalization by  $(n/d)$  is taken so that  $\mathbf{E}[P] = W$ .

We show a spectacular phase transition: there is a threshold  $\vartheta$  such that all eigenvalues of  $P$  greater than this threshold are asymptotically equal to the eigenvalues of  $W$  greater than this threshold.

---

<sup>13</sup>Cf infra.

To explain this result, we introduce the matrix  $X = nW \odot W$ , or

$$X_{x,y} = nW_{x,y}^2. \quad (\text{A.5.4})$$

We note  $\rho = \|X\|$  its operator norm, which is the same order of magnitude as the operator norm of  $W$  under the incoherence assumption (A.5.2). We also put  $L = n \max_{x,y} |W_{x,y}|$ . Under the previous assumptions on  $W$ , this  $L$  is  $O(1)$ . The threshold referred to above is defined as

$$\vartheta = \max \left\{ \sqrt{\frac{\rho}{d}}, \frac{L}{d} \right\}.$$

Note  $s$  the number of eigenvalues of  $W$  greater than  $\vartheta$ :

$$\mu_1 \geq \dots \geq \mu_s > \vartheta \geq \mu_{s+1} \geq \dots \geq \mu_r. \quad (\text{A.5.5})$$

We note  $|\lambda_1| \geq \dots \geq |\lambda_n|$  the (complex) eigenvalues  $A$ . The first result is an exact asymptotic of the large eigenvalues of  $P$ .

**Theorem 27.** *Let  $d > 1$  be a real number and  $W$  a real matrix of size  $n \times n$  verifying the assumptions mentioned before. If  $n$  is large enough, then for any  $i \in \{1, \dots, s\}$  we have*

$$|\mu_i - \lambda_i| = o(1). \quad (\text{A.5.6})$$

In addition, for every  $i > s$  we have

$$|\lambda_i| \leq \vartheta + o(1). \quad (\text{A.5.7})$$

Note that the  $\vartheta$  threshold is defined as a maximum: it is the same phenomenon as already noted in the first part of this thesis, where the second eigenvalue was asymptotically lower than  $\rho \vee \delta^{-1}$ . The number  $\delta^{-1}$  was the maximum of the transition matrix entries, and here  $L/d$  is the maximum of the entries of  $P$  with high probability.

When  $d$  is large enough (more precisely when  $d > L^2/\rho$ ) it is the first term that prevails in the definition of  $\vartheta$ ; it is easy to verify that  $L^2/\rho$  is a good measurement of the delocalization of  $W$ , in the sense that for a highly delocalized matrix (typically, with all equal entries) we have  $L^2/\rho \approx 1$ . There is therefore a competition between the localization of the matrix and the density  $d$ .

Our second result is a description of the behavior of the eigenvectors of  $P$ . The result is only shown for rank-1 matrices (an already interesting framework, see [56]) but it will be extended to all low ranks in future work.

**Theorem 28.** *Suppose  $W = \mu\varphi\varphi^*$  with  $\varphi$  a delocalized unit vector. If  $d > n|\varphi|_4^4$ , then  $\lambda_1 \rightarrow \mu_1$ , and if  $\psi$  is the eigenvector of  $P$  associated with  $\lambda_1$ , then with probability  $1 - o(1)$ ,*

$$|\langle \psi, \varphi \rangle| = (1 + o(1)) \sqrt{1 - \frac{n|\varphi|_4^4}{d}}. \quad (\text{A.5.8})$$

The  $n|\varphi|_4^4$  threshold simply comes from the computation of  $\rho$  in this case. Note that the more delocalized  $\varphi$  is, the smaller  $n|\varphi|_4^4$  is, another example of the competition mentioned above between delocalization of  $W$  and  $d$ .

It is therefore possible to obtain a strictly positive correlation with the true eigenvector  $\varphi$ ; in the total absence of information, the best possible estimator for  $\varphi$  is simply to take a uniform vector  $u$  on  $\mathbb{S}^{n-1}$  and the correlation  $\langle u, \varphi \rangle$  then goes to 0.

### Non-orthogonality

Our work revealed an astonishing phenomenon: the eigenvectors of  $P$  associated with eigenvalues above  $\vartheta$  have a non-trivial covariance structure. More precisely, there is a matrix  $\Gamma$  of size  $s$ , depending only on  $W$  and  $d$ , such that if  $\psi_i, \psi_j$  are the eigenvectors of  $P$  associated with  $\lambda_i \sim \mu_i$  and  $\lambda_j \sim \mu_j$ , then  $\langle \psi_i, \psi_j \rangle \approx \Gamma_{i,j}$  (for a specific statement, we refer to the last chapter of the manuscript). As a general rule, the matrix  $\Gamma$  is not diagonal:

$$\Gamma_{i,j} = \sum_{k=0}^{\infty} \frac{\langle \mathbf{1}, X^k \varphi^{i,j} \rangle}{(\mu_i \mu_j d)^k} \quad (\text{A.5.9})$$

where  $X$  has been defined in (A.5.4) and  $\varphi^{i,j} := \varphi_i \odot \varphi_j$  is the termwise product of the real eigenvectors. In some cases (for example, if the  $\ell^2$  mass of each line of  $W$  is constant), it is easy to see that  $\Gamma$  is diagonal, but it is possible to construct simple examples for which this is not the case. One of the difficulties in proving these results comes from this non-orthogonality (which disappears when  $d \rightarrow \infty$ ).

### Erdős-Rényi directed graphs

An immediate consequence of the above theorems concerns the spectrum of diluted Erdős-Rényi graphs, obtained by simply taking  $W_{i,j} = 1/n$ , a matrix whose spectrum is 1 and 0 with multiplicity  $n - 1$ . As  $nW \odot W = W$ , we get  $\rho = 1$  and  $\vartheta = 1/\sqrt{d}$ .

From a spectral point of view, these graphs are for the moment relatively less studied than their un-directed siblings: for the latter, which are very well known now, the results of Komlos and Füredi [81] and Krievlevitch and Sudakov [96] completed by the recent works of Benaych-Georges, Bordenave and Knowles ([24, 25], Tikhomirov and Youssef [136] and Alt, Ducatez and Knowles [11] fully describe the behaviour of large eigenvalues in all regimes of  $d$ .

For the directed Erdős-Rényi graphs, our result is as follows.

**COROLLARY A.5.1.** *Let  $d$  be a fixed real number and  $A$  be the adjacency matrix of a directed Erdős-Rényi graph  $\text{ER}(n, d/n)$ . The following statements are true with probability  $1 - o(1)$  when  $n \rightarrow \infty$ :*

1. *If  $d \leq 1$ , all the eigenvalues of  $A$  have modulus smaller than  $1 + o(1)$ .*
2. *Otherwise,  $d > 1$ ; in this case,  $\lambda_1(A) \rightarrow d$  and all other eigenvalues have modulus smaller than  $\sqrt{d} + o(1)$ . If  $\psi$  is the eigenvector of  $A$  associated with  $\lambda_1(A)$ , then*

$$|\langle \psi, \varphi \rangle| \rightarrow \sqrt{1 - \frac{1}{d}}. \quad (\text{A.5.10})$$

These points are illustrated in the figures A.12-A.13.

The previous statement thus demonstrates a well-known observation often mentioned by physicists, namely that the empirical spectral measure is supported in  $D(0, \sqrt{d})$  including in the diluted case; we refer to the excellent survey [114] (from physics), and in particular to Sections 4.2 and 4.3, which mention these facts, but without rigorously proving them.

Note  $\vec{\mu}_{A_n}$  the spectral empirical measure of the adjacency matrix of a directed Erdős-Rényi graph with parameter  $d/n$ ; to my knowledge, there are strictly no rigorous results on these measures in the diluted framework. Only one paper by Basak and Rudelson [18] shows that if  $d = \Omega(\log(n)^2)$ , then the spectral empirical measure converges towards the circle distribution. Even the convergence of  $\vec{\mu}_{A_n}$  towards a measure on  $\mathbb{C}$  has not been proven, and a possible closed-form expression for the limit is not even conjectured, unlike the Kesten-McKay conjecture already mentioned (Figure A.4).

Finally, let us point something already encountered in the first section of this introduction: we only show an upper bound for  $|\lambda_2|$ . The corresponding lower bound,  $|\lambda_2| \geq \sqrt{d} - o(1)$ , is not demonstrated, but seems true (see in particular Figure A.14).

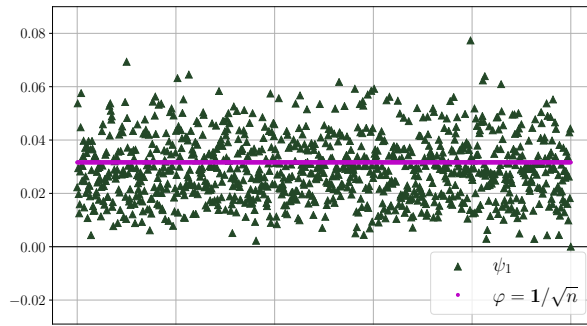


Figure A.12 – A plot of  $|\langle \psi_1, \varphi \rangle|$  for different mean degrees  $d$ . For each  $d = 2, \dots, 20$ , we made 20 simulations of  $\text{ER}(1000, d)$  and computed  $\langle \psi_1, \varphi \rangle$  (little green dashes). The mean is plotted in red and the prediction  $\sqrt{1 - 1/d}$  is in black.

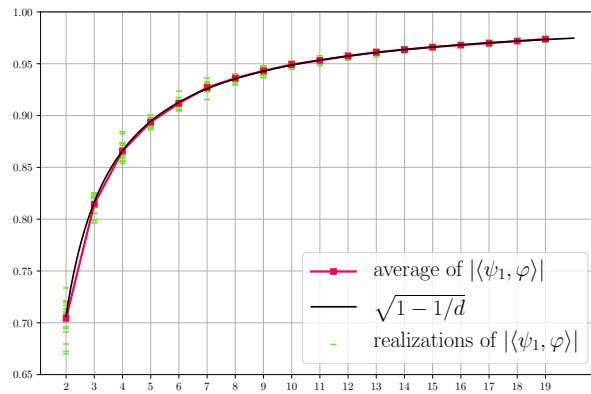


Figure A.13 – Representation of  $|\langle \psi, \mathbf{1}/\sqrt{n} \rangle|$  for several values of  $d$ , with  $n = 1000$ . For each  $d = 2, \dots, 20$ , we generated 20 realizations of  $\text{ER}(1000, d)$  and numerically computed  $|\langle \psi, \mathbf{1}/\sqrt{n} \rangle|$  (small green dashes). The average is in red and the theoretical limit  $\sqrt{1 - 1/d}$  is in black.

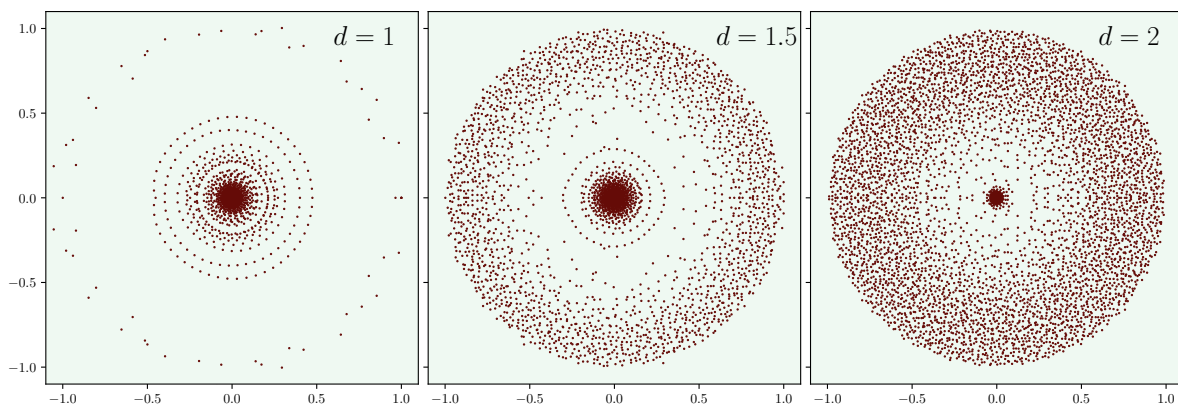


Figure A.14 – Three realizations of directed Erdős-Rényi graph spectra with  $n = 10000$  vertices, for three values of  $p = d/n$ . We cut the outlier and renormalized by  $1/\sqrt{d}$ .



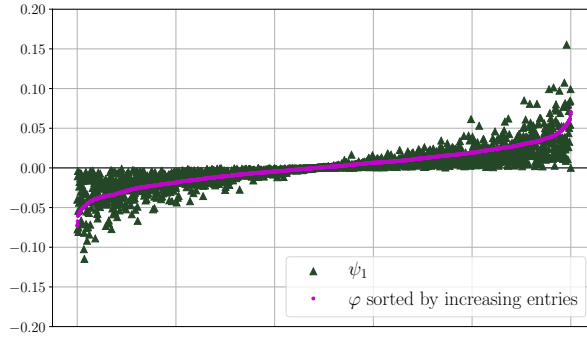


Figure A.15 – We took  $P = \varphi\varphi^*$  with  $\varphi \sim \text{Unif}(\mathbb{S}^{n-1})$ ,  $n = 1000$ . The entries in  $\varphi$  have been sorted in ascending order. The entries of  $\psi$  are in green. The theoretical correlation is close to  $\sqrt{1 - n|\varphi|_4^4/d} \approx \sqrt{1 - 3/d}$  with high probability.

### Typical case: uniform eigenvectors.

Our results are valid for any initial matrix  $W$  verifying the assumptions mentioned before. In some sense, *almost all matrices* verify those assumptions. Indeed, given a spectrum  $\Sigma = \text{diag}(\mu_1, \dots, \mu_r)$ , we can choose eigenvectors  $U$  distributed according to Haar's measure on all orthonormal matrices with size  $n \times r$ ; the matrix

$$W = U^*\Sigma U$$

is then uniform, and it is well known that  $U$  is delocalized in the sense that  $\|U\|_\infty = O(\sqrt{\log n/n})$  with high probability. Such a model is popular in the matrix completion literature, and it is already used in [53, 92].

The theorems stated above will therefore be valid for most of the rank- $r$  matrices. Figures A.15 and A.16 show illustrations of the phenomenon described in Theorem 28 for a rank-1 matrix.

These results complement those obtained in [56], which are formulated in the  $d \rightarrow \infty$  regime (see in particular [56, Page 10]); they do not obtain the transition in  $\vartheta$  nor the exact asymptotics; indeed, our method does not use any eigenvector perturbation theorem such as Davis-Kahan or Neumann's identity (see [142, 71, 56]), which are often optimal in the worst case but quite bad in typical cases. Our proof technique directly integrates the perturbation of eigenvectors, and thus make it possible to obtain detailed information on them.

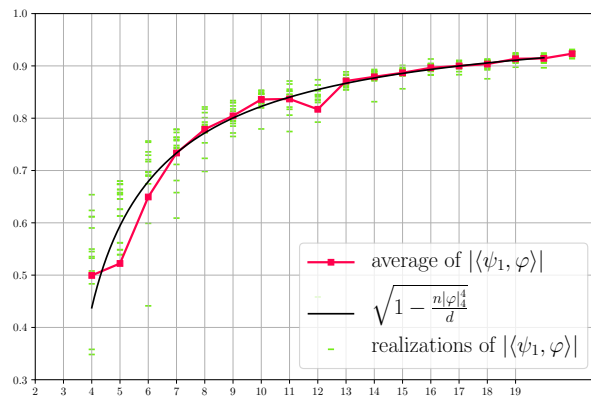


Figure A.16 – Representation of  $|\langle \psi_1, \varphi \rangle|$  for different  $d$ . First we generated  $W = \varphi\varphi^*$  with  $\varphi$  uniform on  $\mathbb{S}^{n-1}$ . Then, for each  $d = 2, \dots, 20$ , we performed 20 simulations of  $A$  with  $W$  as the underlying matrix, then we calculated  $\langle \psi_1, \varphi \rangle$ . The average is in red and the limit value  $\sqrt{1 - n|\varphi|_4^4/d}$  is in black.

## A.6 Two concluding remarks

This last section presents two ideas that emerged from the results of this thesis.

### ‘Dilute’ is the new ‘sparse’

In the literature on low-density networks or matrices, there are several interpretations on what ‘low-density’ or ‘sparse’ means. The general consensus is that the average number of connections of the agents, say  $d$ , must be negligible compared to the size of the network:  $d = o(n)$ . However, two regimes are sparse in this sense: the regime we will call **dilute**, where  $d = O(1)$  is bounded independently of  $n$ , and the regime we will call **sparse** where  $d = o(n)$  but goes to infinity, typically  $d = (\log n)^c$  or  $d = n^\alpha$ , with  $\alpha < 1$ . It should be noted that this terminology is not firmly established, and that the two terms have been used interchangeably in the literature.

The modern study of low-density networks began in the 1960s with the works on phase transitions in the Erdős-Rényi model, and has really exploded since the 1990s with the emergence of computer or social networks, in which agents have a relatively small number of links with other agents. However, from the point of view of applications, the diluted and sparse regimes are identical. Engineers, statisticians or physicists will never face networks with a size larger than, say,  $10^{80}$  (the number of atoms in the observable universe), and in practice diluted or sparse regimes will be strictly indistinguishable (we have  $\log 10^{80} \approx 184$ ). For these applications, the results obtained for the sparse regime, where  $d \rightarrow \infty$  slowly, should be largely satisfying, and indeed the sparse regime is much easier to handle than the diluted regime and the proof techniques are well-established now.

However, the difference between the *dilute* and *sparse* regimes is not artificial, and hides a deep phase transition which — in my opinion — is underestimated in the literature: when it comes to the spectrum, the sparse regime lies in the universality class of ‘dense’ regimes, the world of the semi-circle distribution for Hermitian models, and of the circle distribution for non-Hermitian models. In the main graph models ( $d$ -regular, Erdős-Rényi  $d/n$ ) with density  $d \rightarrow \infty$ , the spectral measure always converges to  $\rho_{\text{sc}}$  or to  $\rho_{\text{circle}}$ , and this including at microscopic scales ([18, 57, 17, 59, 75, 47, 75, 74]). However, some of the real networks studied by engineers or statisticians do not lie in Wigner’s universality class ([76, 129]), such as social network graphs, whose spectrum clearly does not look like a semi-circle (Figure A.17).

If we want the models to fit the reality, the good criterion therefore seems to be not only the small number of links between the agents of the system, but also the spectral structure which must be very far from the semi-circle of circle distributions: presence of atoms, unbounded supports, extended and localized states. The reader can see this by observing the few examples of real network spectra in Figure A.17 (the networks in question are described at the end of the paragraph). These are the features that should be found in applied mathematics models. From this point of view, even diluted graph models are not necessarily the best fitted to reality: the results obtained for an Erdős-Rényi graph with a fixed but large  $d$  parameter, for example  $d = 10^{45}$ , will certainly be further from reality than those obtained for any  $n$  and any  $d \leq \log n$ .

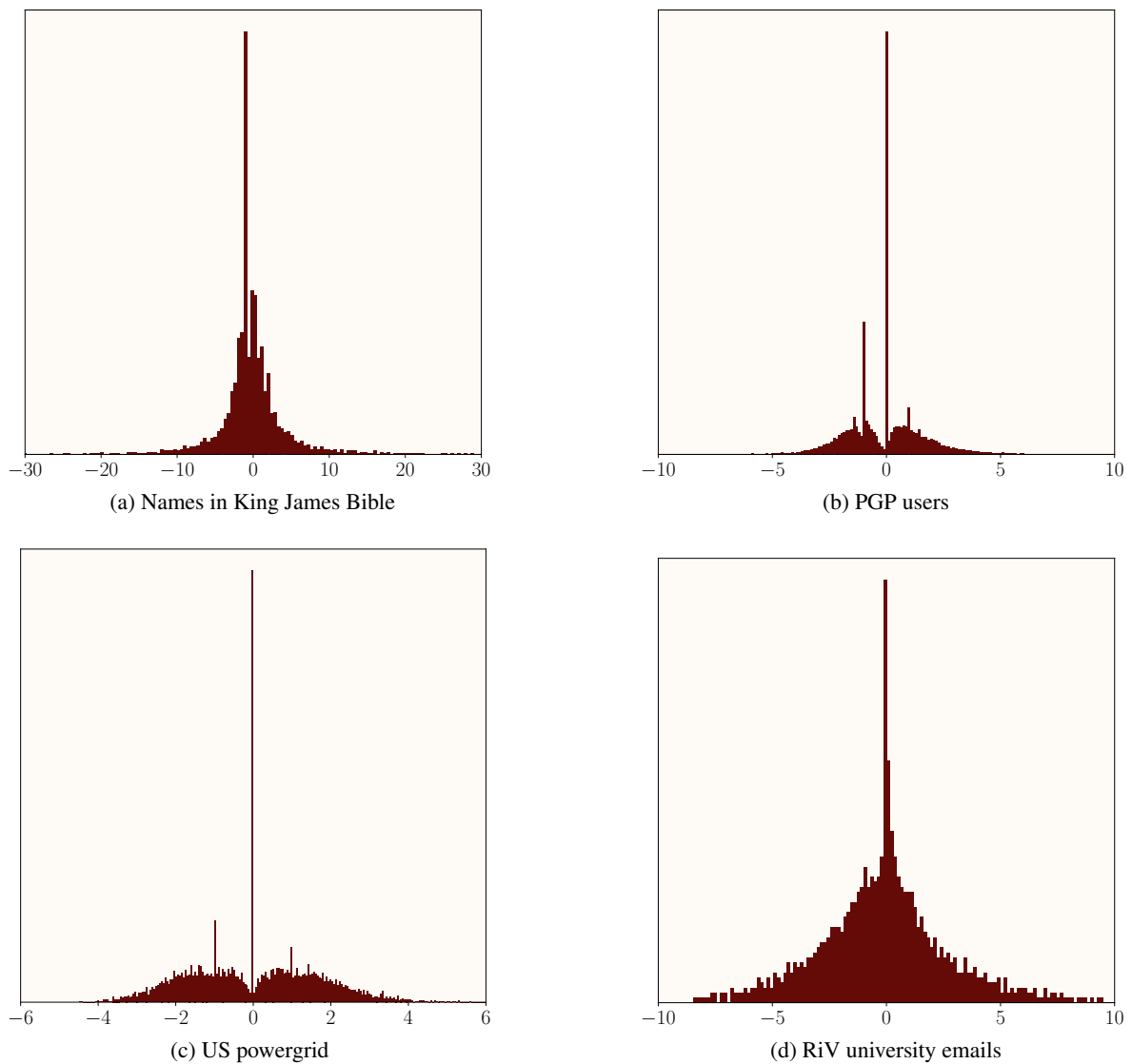


Figure A.17 – Some spectra of real diluted networks.

The network data in Figure A.17 are freely available, see the database [99] and the paper [140] and the references below.

- (a) King James' Bible (1611) is an English version of the Bible, and is still a reference for the Anglican Church. It contains many names (of places or people), which are the 1773 vertices of the graph. Two names are linked by an edge if they appear in the same verse; each edge is weighted by the number of verses in which the two names appear (so they are multi-edges). There are a total of 9131 edges and 16401 multi-edges; the average degree is 18.5 and the maximum degree is 364 — corresponding to the name Israel. These data were compiled by Chris Harrison and are available on his website.
- (b) PGP (PrettyGoodPrivacy) is a popular encryption algorithm, widely used in emails. The network nodes are the users (10680) in the giant connected component, and two users are connected if they have exchanged information secured by PGP. There are 24316 edges, the average degree is 4.55 and the maximum degree is 205 (see [30]).
- (c) The United States power grid (one of the networks studied in the famous paper [140]) consists of 4941 units (generally, transformers or power plants); the edges represent the power lines and

there are 6594 of them. The average degree is 2.66 and the maximum degree is 19.

- (d) In 2003, the email network of Rovira i Virgile University in Tarragona ([86]) contained 1133 users (giant component). Two users are connected if they have sent each other an email. There are 5451 edges, the average degree is 9.6 and the maximum degree is 71.

### Less symmetry, more eigenvalues

Most algebra textbooks describe the instability of the eigenvalues of non-symmetric matrices. For example, it is easy to construct two non-Hermitian matrices which are very close in norm, but whose eigenvalues are very different. Yet, the remarkable and ubiquitous circle law suggests that the eigenvalues of some non-symmetric random matrices are in fact very stable (see [40, Figure 2 and commentary]), as if the randomness in the matrix had the power to regularize the spectrum.

Recent results seem to show that non-symmetric matrices have much finer spectral properties than might have been expected, and that they capture the statistics of the underlying networks *better than their symmetric counterparts*. For example, in the community detection problem, it was not the adjacency matrix that needed to be studied, but one of its non-symmetric variants, the non-backtracking matrix ([97, 43]), because the large eigenvalues of the latter better reflect the graph's internal structure.

Recently, a paper by Chen et al ([56]) took a step forward, suggesting that it would sometimes be beneficial to un-symmetrize symmetric problems. The authors note and prove a phenomenon quite similar to that described in the last part of this thesis: if  $M$  is a low-rank symmetric matrix perturbed by unsymmetric noise  $H$ , it is more efficient to get back  $M$  from the eigenvalues of  $M' = M + H$  than from the singular values of  $M'$ ; moreover, even if the underlying matrix  $M$  is symmetric, it is sub-optimal to use this information by symmetrizing  $M'$ .

This is also the conclusion of our work: even when the underlying matrix is symmetrical, we can un-symmetrize it, which boils down to reorganizing the data in order to extract its essential features; this is done

- without deleting information, unlike other procedures already used, such as trimming the high degrees in [77] or [92],
- without losing dimension, as was the case with the non-backtracking matrix whose transition from dimension  $n$  to  $dn$  could be quite a problem in practice, even when  $d = 10$ .

Such un-symmetrization may not be optimal; the authors of [56] note, for example, that their un-symmetrization procedures lead to an increase in noise variance. Two perspectives seem promising: first, the theoretical results we have presented in the last part of this thesis pave the way for very precise threshold computations, which will often allow statistical problems to be fully solved despite the increase in noise; and secondly, some less elementary un-symmetrization procedures will certainly allow better use of this new philosophy.



# Bibliography

- [1] Miklos Abért, Andreas Thom, and Balint Virág. Benjamini-Schramm convergence and point-wise convergence of the spectral measure. *In preparation.*, 2016.
- [2] Ragi Abou-Chacra, David J. Thouless, and Philip W. Anderson. A selfconsistent theory of localization. *Journal of Physics C: Solid State Physics*, 6(10):1734, 1973.
- [3] Dimitris Achlioptas and Frank McSherry. Fast computation of low-rank matrix approximations. *Journal of the ACM (JACM)*, 54(2):9, 2007.
- [4] Michael Aizenman, Robert Sims, and Simone Warzel. Stability of the absolutely continuous spectrum of random Schrödinger operators on tree graphs. *Probab. Theory Related Fields*, 136(3):363–394, 2006.
- [5] Michael Aizenman and Simone Warzel. The canopy graph and level statistics for random operators on trees. *Mathematical Physics, Analysis and Geometry*, 9(4):291–333, 2006.
- [6] David Aldous. Asymptotic fringe distributions for general families of random trees. *The Annals of Applied Probability*, pages 228–266, 1991.
- [7] David Aldous and Russell Lyons. Processes on unimodular random networks. *Electron. J. Probab.*, 12:no. 54, 1454–1508, 2007.
- [8] David Aldous and J. Michael Steele. The objective method: probabilistic combinatorial optimization and local weak convergence. In *Probability on discrete structures*, volume 110 of *Encyclopaedia Math. Sci.*, pages 1–72. Springer, Berlin, 2004.
- [9] Noga Alon. Eigenvalues and expanders. *Combinatorica*, 6(2):83–96, 1986.
- [10] Noga Alon. Spectral techniques in graph algorithms (invited paper). In *LATIN’98: theoretical informatics (Campinas, 1998)*, volume 1380 of *Lecture Notes in Comput. Sci.*, pages 206–215. Springer, Berlin, 1998.
- [11] Johannes Alt, Raphaël Ducatez, and Antti Knowles. Extremal eigenvalues of critical erdős-rényi graphs. *arXiv preprint arXiv:1905.03243*, 2019.
- [12] Omer Angel, Joel Friedman, and Shlomo Hoory. The non-backtracking spectrum of the universal cover of a graph. *arXiv preprint arXiv:0712.0192*, 2007.
- [13] Dana Angluin. Local and global properties in networks of processors. In *Proceedings of the twelfth annual ACM symposium on Theory of computing*, pages 82–93. ACM, 1980.
- [14] Hagit Attiya, Marc Snir, and Manfred K. Warmuth. Computing on an anonymous ring. *J. ACM*, 35(4):845–875, October 1988.
- [15] Jinho Baik, Gérard Ben Arous, Sandrine Péché, et al. Phase transition of the largest eigenvalue for nonnull complex sample covariance matrices. *The Annals of Probability*, 33(5):1643–1697, 2005.

- [16] Michael D Barrus and Elizabeth A Donovan. Neighborhood degree lists of graphs. *Discrete Mathematics*, 2017.
- [17] Anirban Basak, Nicholas Cook, and Ofer Zeitouni. Circular law for the sum of random permutation matrices. *ArXiv e-prints*, May 2017.
- [18] Anirban Basak and Mark Rudelson. The circular law for sparse non-hermitian matrices. *arXiv preprint arXiv:1707.03675*, 2017.
- [19] Kevin E Bassler, Charo I Del Genio, Péter L Erdős, István Miklós, and Zoltán Toroczkai. Exact sampling of graphs with prescribed degree correlations. *New Journal of Physics*, 17(8):083052, 2015.
- [20] Friedrich L Bauer and Charles T Fike. Norms and exclusion theorems. *Numerische Mathematik*, 2(1):137–141, 1960.
- [21] Michel Bauer and Olivier Golinelli. On the kernel of tree incidence matrices. *J. Integer Seq.*, 3(1):Article 00.1.4, 1 HTML document, 2000.
- [22] Michel Bauer and Olivier Golinelli. Exactly solvable model with two conductor-insulator transitions driven by impurities. *Physical review letters*, 86(12):2621, 2001.
- [23] Michel Bauer and Olivier Golinelli. Random incidence matrices: moments of the spectral density. *J. Statist. Phys.*, 103(1-2):301–337, 2001.
- [24] Florent Benaych-Georges, Charles Bordenave, and Antti Knowles. Largest eigenvalues of sparse inhomogeneous Erdős-Rényi graphs. *arXiv preprint arXiv:1704.02953*, 2017.
- [25] Florent Benaych-Georges, Charles Bordenave, and Antti Knowles. Spectral radii of sparse random matrices. *arXiv e-prints*, page arXiv:1704.02945, Apr 2017.
- [26] Florent Benaych-Georges and Raj Rao Nadakuditi. The eigenvalues and eigenvectors of finite, low rank perturbations of large random matrices. *Advances in Mathematics*, 227(1):494–521, 2011.
- [27] Itai Benjamini, Russell Lyons, and Oded Schramm. Unimodular random trees. *Ergodic Theory Dynam. Systems*, 35(2):359–373, 2015.
- [28] Annabell Berger. A note on the characterization of digraphic sequences. *Discrete Mathematics*, 314:38–41, 2014.
- [29] Shankar Bhamidi, Steven N. Evans, and Arnab Sen. Spectra of large random trees. *J. Theoret. Probab.*, 25(3):613–654, 2012.
- [30] Marián Boguná, Romualdo Pastor-Satorras, Albert Díaz-Guilera, and Alex Arenas. Models of social networks based on social distance attachment. *Physical review E*, 70(5):056122, 2004.
- [31] Belá Bollobás. *Modern Graph Theory*. Springer, 1998.
- [32] Béla Bollobás. *Random graphs*, volume 73 of *Cambridge Studies in Advanced Mathematics*. Cambridge University Press, Cambridge, second edition, 2001.
- [33] Charles Bordenave. A new proof of Friedman’s second eigenvalue Theorem and its extension to random lifts. *ArXiv e-prints*, February 2015.
- [34] Charles Bordenave. *Spectrum of random graphs*. 2016. Lecture notes.

- 
- [35] Charles Bordenave and Pietro Caputo. Large deviations of empirical neighborhood distribution in sparse random graphs. *Probability Theory and Related Fields*, 163(1-2):149–222, 2015.
- [36] Charles Bordenave, Pietro Caputo, and Djalil Chafaï. Spectrum of large random reversible Markov chains: two examples. *ArXiv e-prints*, November 2008.
- [37] Charles Bordenave, Pietro Caputo, and Djalil Chafaï. Spectrum of large random reversible Markov chains: Heavy-tailed weights on the complete graph. *ArXiv e-prints*, March 2009.
- [38] Charles Bordenave, Pietro Caputo, Djalil Chafaï, and Daniele Piras. Spectrum of large random markov chains: heavy-tailed weights on the oriented complete graph. *Random Matrices: Theory and Applications*, 6(02):1750006, 2017.
- [39] Charles Bordenave, Pietro Caputo, and Justin Salez. Random walk on sparse random digraphs. *Probability Theory and Related Fields*, 170(3-4):933–960, 2018.
- [40] Charles Bordenave and Djalil Chafaï. Around the circular law. *Probability surveys*, 9, 2012.
- [41] Charles Bordenave and Simon Coste. Graphs with prescribed local neighborhoods of their universal coverings. *Journal of Combinatorial Theory, Series B*, 2019.
- [42] Charles Bordenave and Marc Lelarge. Resolvent of large random graphs. *Random Structures Algorithms*, 37(3):332–352, 2010.
- [43] Charles Bordenave, Marc Lelarge, and Laurent Massoulié. Non-backtracking spectrum of random graphs: community detection and non-regular Ramanujan graphs. *ArXiv e-prints*, January 2015.
- [44] Charles Bordenave, Marc Lelarge, and Justin Salez. The rank of diluted random graphs. *Ann. Probab.*, 39(3):1097–1121, 2011.
- [45] Charles Bordenave, Yanqi Qiu, and Yiwei Zhang. Spectral gap of sparse bistochastic matrices with exchangeable rows with application to shuffle-and-fold maps. *ArXiv e-prints*, May 2018.
- [46] Charles Bordenave, Arnab Sen, and Bálint Virág. Mean quantum percolation. *J. Eur. Math. Soc. (JEMS)*, 19(12):3679–3707, 2017.
- [47] Paul Bourgade, Jiaoyang Huang, Horng-Tzer Yau, et al. Eigenvector statistics of sparse random matrices. *Electronic Journal of Probability*, 22, 2017.
- [48] Jonathan Breuer. Singular continuous and dense point spectrum for sparse tree with finite dimensions. In *Probability and mathematical physics*, volume 42, pages 65–83. Amer. Math. Soc. Providence, RI, 2007.
- [49] Jonathan Breuer. Singular continuous spectrum for the laplacian on certain sparse trees. *Communications in mathematical physics*, 269(3):851–857, 2007.
- [50] Gerrandy Brito, Ioanna Dumitriu, and Kameron Decker Harris. Spectral gap in random bipartite biregular graphs and its applications. *ArXiv e-prints*, April 2018.
- [51] Andries E Brouwer and Willem H Haemers. *Spectra of graphs*. Springer Science & Business Media, 2011.
- [52] Emmanuel J Candès and Benjamin Recht. Exact matrix completion via convex optimization. *Found. Comput. Math.*, 9(6):717–772, 2009.
- [53] Emmanuel J Candès and Terence Tao. The power of convex relaxation: near-optimal matrix completion. *IEEE Trans. Inform. Theory*, 56(5):2053–2080, 2010.



- [54] René Carmona and Jean Lacroix. *Spectral theory of random Schrödinger operators*. Springer Science & Business Media, 2012.
- [55] Sourav Chatterjee. Matrix estimation by universal singular value thresholding. *Ann. Statist.*, 43(1):177–214, 2015.
- [56] Yuxin Chen, Chen Cheng, and Jianqing Fan. Asymmetry helps: Eigenvalue and eigenvector analyses of asymmetrically perturbed low-rank matrices. *arXiv preprint arXiv:1811.12804*, 2018.
- [57] Nicholas Cook. The circular law for signed random regular digraphs. *ArXiv e-prints*, August 2015.
- [58] Nicholas Cook. The circular law for random regular digraphs. *ArXiv e-prints*, March 2017.
- [59] Nicholas Cook, Larry Goldstein, and Tobias Johnson. Size biased couplings and the spectral gap for random regular graphs. *ArXiv e-prints*, October 2015.
- [60] Colin Cooper. Random walks, interacting particles, dynamic networks: Randomness can be helpful. In *International Colloquium on Structural Information and Communication Complexity*, pages 1–14. Springer, 2011.
- [61] Colin Cooper and Alan Frieze. The size of the largest strongly connected component of a random digraph with a given degree sequence. *Combinatorics, Probability and Computing*, 13(3):319 – –338, 5 2004.
- [62] Simon Coste. The Spectral Gap of Sparse Random Digraphs. *ArXiv e-prints*, August 2017.
- [63] Simon Coste and Justin Salez. Emergence of extended states at zero in the spectrum of sparse random graphs. *arXiv e-prints*, page arXiv:1809.07587, Sep 2018.
- [64] Dragoš Cvetkovic, Slobodan Simic, and Peter Rowlinson. *An introduction to the theory of graph spectra*. Cambridge University Press, 2009.
- [65] Mark A Davenport and Justin Romberg. An overview of low-rank matrix recovery from incomplete observations. *IEEE Journal of Selected Topics in Signal Processing*, 10(4):608–622, 2016.
- [66] Giuliana P. Davidoff, Peter Sarnak, and Alain Valette. *Elementary number theory, group theory, and Ramanujan graphs*. London Mathematical Society student texts. Cambridge University Press, Cambridge (UK), New York, 2003.
- [67] Pierre de la Harpe, A. Guyan Robertson, and Alain Valette. On the spectrum of the sum of generators for a finitely generated group. *Israel J. Math.*, 81(1-2):65–96, 1993.
- [68] Charles Delorme and Jean-Pierre Tillich. The spectrum of de Bruijn and Kautz graphs. *European J. Combin.*, 19(3):307–319, 1998.
- [69] Persi Diaconis. The cutoff phenomenon in finite markov chains. *Proceedings of the National Academy of Sciences*, 93(4):1659–1664, 1996.
- [70] Petros Drineas and Anastasios Zouzias. A note on element-wise matrix sparsification via a matrix-valued Bernstein inequality. *Inform. Process. Lett.*, 111(8):385–389, 2011.
- [71] Justin Eldridge, Mikhail Belkin, and Yusu Wang. Unperturbed: spectral analysis beyond Davis-Kahan. *arXiv e-prints*, page arXiv:1706.06516, Jun 2017.

- [72] Nathanaël Enriquez and Laurent Ménard. Spectra of large diluted but bushy random graphs. *Random Structures Algorithms*, 49(1):160–184, 2016.
- [73] Paul Erdős and Tibor Gallai. Graphs with prescribed degrees of vertices (hungarian). *Mat. Lapok*, 11:264–274, 1960.
- [74] László Erdős, Antti Knowles, Horng-Tzer Yau, and Jun Yin. Spectral statistics of erdős-rényi graphs ii: Eigenvalue spacing and the extreme eigenvalues. *Communications in Mathematical Physics*, 314(3):587–640, Sep 2012.
- [75] László Erdős, Antti Knowles, Horng-Tzer Yau, Jun Yin, et al. Spectral statistics of Erdős–Rényi graphs i: local semicircle law. *The Annals of Probability*, 41(3B):2279–2375, 2013.
- [76] Illes J Farkas, Imre Derényi, Albert-László Barabási, and Tamas Vicsek. Spectra of “real-world” graphs: Beyond the semicircle law. *Physical Review E*, 64(2):026704, 2001.
- [77] Uriel Feige and Eran Ofek. Spectral techniques applied to sparse random graphs. *Random Structures & Algorithms*, 27(2):251–275, 2005.
- [78] James Allen Fill et al. Eigenvalue bounds on convergence to stationarity for nonreversible markov chains, with an application to the exclusion process. *The annals of applied probability*, 1(1):62–87, 1991.
- [79] Joel Friedman. A proof of Alon’s second eigenvalue conjecture and related problems. *CoRR*, cs.DM/0405020, 2004.
- [80] Joel Friedman, Jeff Kahn, and Endre Szemerédi. On the second eigenvalue of random regular graphs. In *Proceedings of the twenty-first annual ACM symposium on Theory of computing*, pages 587–598. ACM, 1989.
- [81] Zoltan Füredi and Janosz Komlós. The eigenvalues of random symmetric matrices. *Combinatorica*, 1(3):233–241, 1981.
- [82] David Gamarnik, Quan Li, and Hongyi Zhang. Matrix completion from  $o(n)$  samples in linear time. *arXiv preprint arXiv:1702.02267*, 2017.
- [83] Luis Carlos García del Molino, Khashayar Pakdaman, and Jonathan Touboul. Real eigenvalues of non-symmetric random matrices: Transitions and Universality. *arXiv e-prints*, page arXiv:1605.00623, May 2016.
- [84] Rostislav Grigorchuk and Andrzej Zuk. On the asymptotic spectrum of random walks on infinite families of graphs.
- [85] Geoffrey Grimmett. Random labelled trees and their branching networks. *J. Austral. Math. Soc. Ser. A*, 30(2):229–237, 1980/81.
- [86] Roger Guimera, Leon Danon, Albert Diaz-Guilera, Francesc Giralt, and Alex Arenas. Self-similar community structure in a network of human interactions. *Physical review E*, 68(6):065103, 2003.
- [87] Shlomo Hoory, Nathan Linial, and Avi Wigderson. Expander graphs and their applications. *Bulletin of the American Mathematical Society*, 43(4):439–561, 2006.
- [88] Svante Janson. The probability that a random multigraph is simple. II. *J. Appl. Probab.*, 51A(Celebrating 50 Years of The Applied Probability Trust):123–137, 2014.

- [89] Paul Jung and Jaehun Lee. Delocalization and Limiting Spectral Distribution of Erdős-Rényi Graphs with Constant Expected Degree. *ArXiv e-prints*, October 2017.
- [90] Ravindran Kannan and Santosh Vempala. Spectral algorithms. *Found. Trends Theor. Comput. Sci.*, 4(3-4):front matter, 157–288 (2009), 2008.
- [91] Matthias Keller. Absolutely continuous spectrum for multi-type Galton Watson trees. *Ann. Henri Poincaré*, 13(8):1745–1766, 2012.
- [92] Raghunandan. Keshavan, Andrea Montanari, and Sewoong Oh. Matrix Completion from a Few Entries. *ArXiv e-prints*, January 2009.
- [93] Harry Kesten. Symmetric random walks on groups. *Trans. Amer. Math. Soc.*, 92:336–354, 1959.
- [94] Oleksiy Khorunzhy, Mariya Shcherbina, and Valentin Vengerovsky. Eigenvalue distribution of large weighted random graphs. *J. Math. Phys.*, 45(4):1648–1672, 2004.
- [95] Abel Klein. Extended states in the anderson model on the bethe lattice. *Advances in Mathematics*, 133(1):163–184, 1998.
- [96] Michael Krivelevich and Benny Sudakov. The largest eigenvalue of sparse random graphs. *Combinatorics, Probability and Computing*, 12(1):61–72, 2003.
- [97] Florent Krzakala, Cristopher Moore, Elchanan Mossel, Joe Neeman, Allan Sly, Lenka Zdeborová, and Pan Zhang. Spectral redemption in clustering sparse networks. *Proceedings of the National Academy of Sciences*, 110(52):20935–20940, 2013.
- [98] A. Kundu and P. Drineas. A Note on Randomized Element-wise Matrix Sparsification. *ArXiv e-prints*, April 2014.
- [99] Jérôme Kunegis. KONECT – The Koblenz Network Collection. In *Proc. Int. Conf. on World Wide Web Companion*, pages 1343–1350, 2013.
- [100] Jin Ho Kwak and Roman Nedela. Graphs and their coverings.
- [101] Yoram Last. Quantum dynamics and decompositions of singular continuous spectra. *Journal of Functional Analysis*, 142(2):406–445, 1996.
- [102] Yoram Last and Barry Simon. Eigenfunctions, transfer matrices, and absolutely continuous spectrum of one-dimensional Schrödinger operators. *Invent. Math.*, 135(2):329–367, 1999.
- [103] Marc Lelarge and Léo Miolane. Fundamental limits of symmetric low-rank matrix estimation. *Probability Theory and Related Fields*, pages 1–71, 2017.
- [104] David Asher Levin, Yuval Peres, and Elizabeth Lee Wilmer. *Markov chains and mixing times*. Providence, R.I. American Mathematical Society, 2009. With a chapter on coupling from the past by James G. Propp and David B. Wilson.
- [105] Eyal Lubetzky and Yuval Peres. Cutoff on all ramanujan graphs. *Geometric and Functional Analysis*, 26(4):1190–1216, 2016.
- [106] Eyal Lubetzky, Allan Sly, et al. Cutoff phenomena for random walks on random regular graphs. *Duke Mathematical Journal*, 153(3):475–510, 2010.
- [107] Alex Lubotzky, Ralph Phillips, and Peter Sarnak. Ramanujan graphs. *Combinatorica*, 8(3):261–277, 1988.

- 
- [108] Adam W Marcus, Daniel A Spielman, and Nikhil Srivastava. Ramanujan graphs and the solution of the Kadison-Singer problem. In *Proceedings of the International Congress of Mathematicians—Seoul 2014. Vol. III*, pages 363–386. Kyung Moon Sa, Seoul, 2014.
- [109] Adam W Marcus, Daniel A Spielman, and Nikhil Srivastava. Interlacing families I: Bipartite Ramanujan graphs of all degrees. *Ann. of Math. (2)*, 182(1):307–325, 2015.
- [110] Adam W Marcus, Daniel A Spielman, and Nikhil Srivastava. Interlacing Families IV: Bipartite Ramanujan Graphs of All Sizes. *SIAM J. Comput.*, 47(6):2488–2509, 2018.
- [111] Gregori Margulis. Explicit construction of concentrators. *PIT*, 9:325–332, 1973.
- [112] Laurent Massoulié. Community detection thresholds and the weak Ramanujan property. *ArXiv e-prints*, November 2013.
- [113] Brendan McKay. The expected eigenvalue distribution of a large regular graph. *Linear Algebra Appl.*, 40:203–216, 1981.
- [114] Fernando Lucas Metz, Izaak Neri, and Tim Rogers. Spectra of sparse non-hermitian random matrices. *arXiv preprint arXiv:1811.10416*, 2018.
- [115] Léo Miolane. Fundamental limits of low-rank matrix estimation: the non-symmetric case. *arXiv e-prints*, page arXiv:1702.00473, Feb 2017.
- [116] Bojan Mohar. A strengthening and a multipartite generalization of the Alon-Boppana-Serre theorem. *Proc. Amer. Math. Soc.*, 138(11):3899–3909, 2010.
- [117] Ravi Montenegro, Prasad Tetali, et al. Mathematical aspects of mixing times in markov chains. *Foundations and Trends® in Theoretical Computer Science*, 1(3):237–354, 2006.
- [118] Vladimír Müller. On the spectrum of an infinite graph. *Linear Algebra and its Applications*, 93:187–189, 1987.
- [119] Mark E Newman, Steven Strogatz, and Duncan J Watts. Random graphs with arbitrary degree distributions and their applications. *journal*, 64(2):026118, August 2001.
- [120] A. Nilli. On the second eigenvalue of a graph. *Discrete Mathematics*, 91(2):207 – 210, 1991.
- [121] Sean O’Rourke, Van H Vu, and Ke Wang. Random perturbation and matrix sparsification and completion. *ArXiv e-prints*, March 2018.
- [122] Massimo Ostilli. Cayley trees and bethe lattices: A concise analysis for mathematicians and physicists. *Physica A: Statistical Mechanics and its Applications*, 391(12):3417–3423, 2012.
- [123] Soumik Pal and Yizhe Zhu. Community Detection in the Sparse Hypergraph Stochastic Block Model. *arXiv e-prints*, page arXiv:1904.05981, Apr 2019.
- [124] Ori Parzanchevski. Ramanujan Graphs and Digraphs. *ArXiv e-prints*, April 2018.
- [125] Mustazee Rahman. A lower bound on the spectrum of unimodular networks. *arXiv e-prints*, page arXiv:1609.02209, Sep 2016.
- [126] Justin Salez. *Some implications of local weak convergence for sparse random graphs*. PhD thesis, Université Pierre et Marie Curie-Paris VI; Ecole Normale Supérieure de Paris, 2011.
- [127] Justin Salez. Every totally real algebraic integer is a tree eigenvalue. *Journal of Combinatorial Theory, Series B*, 111:249–256, 2015.

- [128] Justin Salez. Spectral atoms of unimodular random trees. *ArXiv e-prints*, September 2016.
- [129] Camellia Sarkar and Sarika Jalan. Spectral properties of complex networks. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, 28(10):102101, 2018.
- [130] Jean-Pierre Serre. Répartition asymptotique des valeurs propres de l'opérateur de hecke tp. *Journal of the American Mathematical Society*, pages 75–102, 1997.
- [131] Gerard Sierksma and Han Hoogeveen. Seven criteria for integer sequences being graphic. 15:223 – 231, 06 1991.
- [132] Barry Simon. Operators with singular continuous spectrum. VI. Graph Laplacians and Laplace-Beltrami operators. *Proc. Amer. Math. Soc.*, 124(4):1177–1182, 1996.
- [133] Barry Simon and Gunter Stolz. Operators with singular continuous spectrum. V. Sparse potentials. *Proc. Amer. Math. Soc.*, 124(7):2073–2080, 1996.
- [134] Daniel A Spielman. Spectral graph theory and its applications. In *Foundations of Computer Science, 2007. FOCS'07. 48th Annual IEEE Symposium on*, pages 29–38. IEEE, 2007.
- [135] Gerald Teschl. *Mathematical methods in quantum mechanics*, volume 157. American Mathematical Soc., 2014.
- [136] Konstantin Tikhomirov and Pierre Youssef. Outliers in spectrum of sparse Wigner matrices. *arXiv e-prints*, page arXiv:1904.07985, Apr 2019.
- [137] Linh V Tran, Van H Vu, and Ke Wang. Sparse random graphs: eigenvalues and eigenvectors. *Random Structures Algorithms*, 42(1):110–134, 2013.
- [138] Amitabha Tripathi and Sujith Vijay. A note on a theorem of erdős & gallai. *Discrete Mathematics*, 265(1):417–420, 2003.
- [139] Van H Vu. Spectral norm of random matrices. *Combinatorica*, 27(6):721–736, 2007.
- [140] Duncan J Watts and Steven H Strogatz. Collective dynamics of 'small-world' networks. *nature*, 393(6684):440, 1998.
- [141] Eugene Paul Wigner. *The collected works of Eugene Paul Wigner. Part A. The scientific papers. Vol. II. Nuclear physics*. Springer-Verlag, Berlin, 1996. Annotated by Herman Feshbach, Edited and with a preface by Arthur S. Wightman and Jagdish Mehra.
- [142] Yi Yu, Tengyao Wang, and Richard J Samworth. A useful variant of the davis–kahan theorem for statisticians. *Biometrika*, 102(2):315–323, 2014.
- [143] Inna Zakharevich. A generalization of Wigner's law. *Comm. Math. Phys.*, 268(2):403–414, 2006.

## RÉSUMÉ

Une matrice aléatoire  $n \times n$  est *diluée* lorsque le nombre d'entrées non nulles est d'ordre  $n$  ; les matrices d'adjacence de graphes  $d$ -réguliers ou les graphes d'Erdős-Rényi de degré moyen  $d$  fixé sont dilués.

Dans le premier chapitre, je démontre une borne supérieure sur la deuxième valeur propre de la matrice de transition sur certains graphes dilués, les graphes de configuration dirigés, dans lesquels on a spécifié le degré (entrant et sortant) de chaque sommet. On obtient aussi une généralisation importante du théorème de Friedman : la seconde valeur propre de la matrice d'adjacence d'un graphe  $d$ -régulier dirigé est inférieure à  $\sqrt{d} + o(1)$ .

Dans le second chapitre, issu d'une collaboration avec Charles Bordenave, on donne une généralisation du théorème d'Erdős-Gallai.

Le troisième chapitre, issu d'une collaboration avec Justin Salez, résout un problème posé en 2004 par Bauer et Golinelli : l'existence ou non d'états étendus dans le spectre limite des graphes d'Erdős-Rényi de paramètre  $d/n$ . On y démontre l'absence d'états étendus en zéro lorsque  $d < e$  et la présence d'états étendus lorsque  $d > e$ . Nos résultats s'étendent aux arbres de Galton-Watson unimodulaires. Je démontre également l'absence d'états étendus en zéro dans le spectre de l'arbre squelette d'Aldous.

Le dernier chapitre est issu d'une collaboration avec Charles Bordenave et Raj Rao Nadakuditi. On y étudie les valeurs propres de la matrice d'adjacence  $A$  d'un graphe d'Erdős-Rényi de paramètre  $d/n$ , dans lequel les arêtes sont pondérées par les entrées d'une matrice symétrique  $P$ . On montre une transition de phase spectaculaire : il existe un seuil  $\vartheta$  dépendant de  $P$  et de  $d$  tel que les plus grandes valeurs propres de  $(n/d)A$  convergent vers les valeurs propres de  $P$  plus grandes que  $\vartheta$ , et tel que les vecteurs propres de  $A$  associés sont alignés avec ceux de  $P$ .

## SUMMARY

A random  $n \times n$  matrix is *diluted* when the number of non-zero entries is of order  $n$ ; adjacency matrices of  $d$ -regular graphs or adjacency matrices of Erdős-Rényi graphs with fixed average degree  $d$  are diluted. This dissertation is about the spectrum of diluted random matrices.

In the first chapter I show an upper bound on the second eigenvalue of the transition matrix on a diluted directed graph model, the *directed configuration model*, in which the degree (in and out) of each vertex is specified. We also get an important generalization of Friedman's theorem: the second eigenvalue of the adjacency matrix of a directed  $d$ -regular graph is less than  $\sqrt{d} + o(1)$ .

A second short chapter, from a collaboration with Charles Bordenave, gives a generalization of the Erdős-Gallai theorem.

The third chapter, a collaboration with Justin Salez, solves a problem raised in 2004 by Bauer and Golinelli: the existence (or not) of extended states in the limiting spectrum of Erdős-Rényi graphs with parameter  $d/n$ . We show the absence of extended states at zero when  $d < e$  and the presence of extended states when  $d > e$ . Our results extend to the spectra of unimodular Galton-Watson tree. I also prove the absence of extended states at zero in the spectrum of the skeleton tree.

The last chapter is a collaboration with Charles Bordenave and Raj Rao Nadakuditi. We study the eigenvalues of the adjacency matrix  $A$  of a directed Erdős-Rényi graph with parameter  $d/n$ , in which the edges are weighted by the entries of a symmetric matrix  $P$ . We show a spectacular phase transition: there is a threshold  $\vartheta$  depending on  $P$  and  $d$  such that the largest eigenvalues of  $(n/d)A$  converge to the eigenvalues of  $P$  which are greater than  $\vartheta$ . The associated eigenvectors of  $A$  are aligned with those of  $P$ .