



HAL
open science

Structures biparties et communautés recouvrantes des graphes de terrains

Raphaël Tackx, Maximilien Danisch, Fabien Tarissan

► **To cite this version:**

Raphaël Tackx, Maximilien Danisch, Fabien Tarissan. Structures biparties et communautés recouvrantes des graphes de terrains. 5ème Conférence sur les Modèles et l'Analyse des Réseaux : Approches Mathématiques et Informatique (MARAMI'14), Oct 2014, Paris, France. pp.1-12. hal-01208339

HAL Id: hal-01208339

<https://hal.science/hal-01208339>

Submitted on 2 Oct 2015

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.

Structures biparties et communautés recouvrantes des graphes de terrains

Raphaël Tackx, Maximilien Danisch, Fabien Tarissan

Sorbonne Universités, UPMC Université Paris 6 and CNRS, UMR 7606, LIP6, Paris

ABSTRACT. De nombreux réseaux rencontrés en pratique se prêtent naturellement à la formalisation sous forme de graphes pour analyser et modéliser leur structure. Cette représentation plate des réseaux s'est montrée cependant peu efficace pour rendre compte de propriétés importantes et non triviales liées à la structure bipartie des réseaux. Des travaux récents ont montré notamment que des propriétés de recouvrements semblaient être présentes dans la plupart des réseaux réels et qu'elles permettaient de mieux expliquer des propriétés observées sur des graphes simples. Le présent travail entend poursuivre cette problématique en étudiant les propriétés liées aux recouvrements dans les structures communautaires des réseaux sociaux. Nous conduisons pour cela une étude basée sur le réseau des pages et catégories WIKIPEDIA et nous montrons notamment que parmi les métriques proposées récemment pour rendre compte de ces recouvrements complexes entre communautés, le coefficient de redondance était plus pertinent que le populaire coefficient de clustering biparti étudié généralement en pratique.

KEYWORDS: Réseaux, Graphes de terrains, Graphes bipartis, Communautés recouvrantes

DOI:10.3166/MARAMI.xx.1-12 © 2014 Lavoisier

1. Introduction

De nombreux réseaux rencontrés en pratique – aussi appelés *graphes de terrains* – se prêtent naturellement à la formalisation sous forme de graphes pour analyser et modéliser leur structure. Habituellement, les sommets du graphe représentent les nœuds du réseau et les arcs entre les sommets représentent les (possibles) interactions entre les nœuds du réseau. Cette approche s'est révélée utile pour identifier des propriétés non triviales dans la structure de réseaux issus de contextes pourtant différents, que ce soit en informatique (Internet, réseaux pair-à-pair, web), en biologie (réseaux d'interactions protéine-protéine, réseaux de régulation génétique), en sciences sociales (réseaux d'amitiés ou de collaboration), en linguistique, en économie, en droit, etc. (Watts, Strogatz, 1998; Cancho, Solé, 2001; Newman *et al.*, 2002; Battiston, Catanzaro, 2004; Le Fessant *et al.*, 2005; Prieur *et al.*, 2008; Tarissan *et al.*, 2009; Ahn *et al.*, 2011; Tarissan, Nollez-Goldbach, 2013).

Cependant, bien que pertinente, la représentation des réseaux sous forme de graphes ne permet pas de rendre compte de la complexité inhérente à la structure hiérarchique de la plupart des réseaux. Si on considère par exemple les réseaux d'acteurs qui relient les acteurs d'un même film (Newman *et al.*, 2001; Watts, Strogatz, 1998) ou les réseaux de co-publications qui relient les auteurs publiant ensemble (Newman *et al.*, 2001; 2002), il est bien plus réaliste de relier les acteurs aux films dans lesquels ils jouent, et les auteurs à leur articles. Cette remarque a amené la communauté scientifique à s'intéresser aux graphes bipartis dans lesquels l'ensemble des nœuds est scindé en deux ensembles disjoints, \top et \perp (par exemple les films et les acteurs), de façon à ce que les liens relient des nœuds du premier ensemble à des nœuds du second. Cet objet fondamental s'est révélé pertinent à la fois pour l'analyse (Ahn *et al.*, 2011; Latapy *et al.*, 2008; Tumminello *et al.*, 2011; Omodei *et al.*, 2013) et la modélisation (Guillaume, Latapy, 2006; Tarissan *et al.*, 2013) de la structures des graphes de terrains.

Dans un papier récent (Tarissan *et al.*, 2013), cette approche a été exploitée afin de proposer, pour la première fois un modèle biparti de la topologie de l'Internet. Il repose sur des techniques de mesures récentes (Mérindol *et al.*, 2009; 2010) qui permettent de révéler la structure à plusieurs niveaux d'Internet. Ce modèle reste simple en ce qu'il ne prend en paramètre que la séquence des degrés des deux niveaux du réseau et génère aléatoirement un graphe biparti respectant la séquence des degrés des deux niveaux. Le papier a montré que, malgré la simplicité du modèle, des propriétés réalistes du réseaux (comme la forte densité locale et des corrélations non triviales entre les nœuds du réseau) émergeaient naturellement. Mais ce travail a aussi montré que le modèle échouait dans la reproduction des recouvrements détectables dans la structure à deux niveaux des réseaux.

Le présent travail étend cette analyse à un autre type de réseaux, les réseaux sociaux et tente d'identifier quelles métriques liées à leur structure bipartie sont le mieux à même de rendre compte de ces propriétés. En particulier, nous étudions si deux métriques récentes, à savoir le *coefficient de clustering biparti* et le *coefficient de redondance*, sont pertinentes pour mettre en évidence les recouvrements observés.

Le reste du papier est organisé comme suit : Section 2 passe en revue les notions nécessaires pour la bonne compréhension du papier, Section 3 présente les résultats principaux obtenus et Section 4 expose les conclusions que l'on peut tirer de ce travail et ouvre vers de futurs travaux.

2. Jeux de données et définitions

Dans cette section, nous donnons les informations nécessaires pour la compréhension du reste du papier. Tout d'abord, nous présentons les jeux de données utilisés dans cette étude (Section 2.1). Puis nous rappelons les différentes définitions et métriques liées aux graphes bipartis (Section 2.2).

2.1. Jeux de données

Le jeu de données avec lequel nous allons travailler est extrait de l'encyclopédie en-ligne WIKIPEDIA, en anglais, de juillet 2012. Il est constitué des 4 millions de pages, de 200 millions de liens hypertextes entre elles ainsi que des 900 000 catégories auxquelles appartiennent ces pages. Les catégories ont une structure hiérarchique, c'est à dire que certaines catégories sont déclarées comme étant des sous-catégories de catégories plus générales.

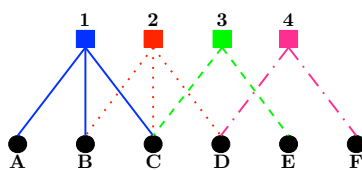
Étant donné la taille conséquente de ce jeu de donnée et afin d'avoir plusieurs jeux de données similaires, nous avons extrait 4 jeux de données plus petits en choisissant à chaque fois, à partir d'une catégories, l'ensemble de ses sous-catégories et des pages incluses. Le choix de ces 4 catégories, guidé par une taille permettant des calculs en temps raisonnable, s'est porté sur : *Anime*, *Linear algebra*, *Graph theory* et *Monument*.

Notons que ces jeux de données, liant articles et catégorie, est un excellent exemple de la structure communautaire des réseaux réels, les catégories pouvant s'apparenter en effet à des communautés de pages d'un point de vue sémantique. Pour cette raison, nous utiliserons indifféremment le terme de *communauté* et *catégorie* dans le contexte de cette étude.

2.2. Métriques

Un *graphe biparti* est un triplet $G = (\top, \perp, E)$, dans lequel \top est l'ensemble des nœuds du *haut* (dans notre cas, les catégories), \perp l'ensemble des nœuds du *bas* (ici les pages WIKIPEDIA), et $E \subseteq \top \times \perp$ l'ensemble des liens entre \top et \perp (qui relie donc les articles aux catégories).

Comparés aux graphes simples, les nœuds d'un graphe bipartis sont donc séparés en deux ensembles disjoints et les liens relient exclusivement les nœuds du haut à ceux du bas. Un exemple d'un tel graphe biparti est donné par la figure suivante, où \top est représenté par des carrés et \perp par des cercles.



Comme pour les graphes simples, les graphes bipartis sont munis de plusieurs métriques qui tentent de rendre compte de différents aspects de leur structure. On peut mentionner les extensions naturelles des caractéristiques des graphes simples comme le nombre de nœuds \top et \perp (notés respectivement n_{\top} et n_{\perp} dans ce papier), le nombre de liens (m), la densité ($\delta = \frac{m}{n_{\top} \cdot n_{\perp}}$), le degré moyen des nœuds haut et bas (resp. k_{\top}

et k_{\perp}), le degré max des nœuds haut et bas (resp. d_{\top}^{\pm} et d_{\perp}^{\pm}). Dans le cas de la figure précédente, on a ainsi $n_{\top} = 4, n_{\perp} = 6, m = 10, \delta = 0.42, k_{\top} = 2.5, k_{\perp} = 1.67$ et $d_{\top}^{\pm} = d_{\perp}^{\pm} = 3$.

Outre ces extensions naturelles, d'autres métriques plus spécifiques peuvent être proposées pour les graphes bipartis. En particulier, la notion classique de densité locale dans les graphes simples, souvent capturée par le *coefficient de clustering* (Watts, Strogatz, 1998), n'a pas d'équivalent bien établi pour les graphes bipartis. La notion de triangle sur laquelle ce dernier repose n'a en effet pas de sens dans ces structures à deux niveaux. Comme suggéré dans (Latapy *et al.*, 2008), on peut cependant s'appuyer sur le *coefficient de clustering biparti* (cc), certainement la métrique la plus communément utilisée, dont la définition est donnée pour toute paire de nœuds $u, v \in \top$ et tente de rendre compte des recouvrements existants entre ces deux nœuds :

$$cc_{\top}(u, v) = \frac{N_{\top}(u) \cap N_{\top}(v)}{N_{\top}(u) \cup N_{\top}(v)}$$

où $N_{\top}(u)$ désigne l'ensemble des voisins de u . Ce coefficient est intéressant en ce qu'il capture le recouvrement entre le voisinage de deux nœuds \top . Dès lors on peut dériver une valeur pour un nœud donné par

$$cc_{\top}(u) = \frac{\sum_{u \in N_{\perp} N_{\top}(v)} cc_{\top}(u, v)}{|N_{\perp} N_{\top}(v)|}$$

On peut enfin dériver une valeur synthétique pour l'ensemble du graphe par $cc_{\top}(G) = \frac{\sum_{u \in \top} cc_{\top}(u)}{n_{\top}}$. Ainsi, dans le graphe biparti donné en exemple, on a $cc(1, 2) = \frac{1}{2}$, $cc(1, 3) = \frac{1}{4}$ et donc $cc_{\top}(1) = \frac{3}{8}$. Le coefficient pour le graphe est lui de $cc_{\top}(G) = 0.24$.

Cependant, il a été montré dans (Latapy *et al.*, 2008) que ce coefficient n'était pas capable de rendre compte de certaines propriétés subtiles liées aux recouvrements. C'est pourquoi les auteurs ont suggéré d'utiliser un autre coefficient permettant également de mettre en évidence les recouvrements entre les voisinages de nœuds dans les graphes bipartis. Il s'agit du *coefficient de redondance* (Latapy *et al.*, 2008; Tarissan *et al.*, 2013; Tarissan, Coupechoux, 2013). Celui-ci est défini pour un nœud $v \in \top$ par :

$$rd_{\top}(v) = \frac{|\{\{u, w\} \in N_{\top}(v)^2 \text{ tel que } \exists v' \neq v, (v', u) \in E \text{ et } (v', w) \in E\}|}{\frac{|N_{\top}(v)|(|N_{\top}(v)|-1)}{2}}$$

Étant donnée une catégorie, ce coefficient calcule donc quelle est la probabilité (observée) que deux pages de cette catégorie appartiennent aussi à une autre catégorie commune.

3. Analyse de la structure communautaire recouvrante des réseaux

Le but de cette section est d'analyser les recouvrements présents dans la structure bipartie des quatre jeux de données issus de WIKIPEDIA et présentés en Section 2.1. Nous allons en particulier focaliser notre attention sur les deux métriques qui ont été proposées pour rendre compte de cette particularité topologique, à savoir le coefficient de clustering biparti et le coefficient de redondance (auxquels nous ferons désormais référence par clustering et redondance). Tout d'abord, nous présentons les caractéristiques globales et standard des réseaux considérés (Section 3.1). Puis, nous nous penchons sur les propriétés liées aux recouvrements, en comparant notamment les distributions et corrélations entre les différentes métriques (Section 3.2).

3.1. Propriétés globales des graphes

Les premières statistiques que nous regardons concernent les propriétés standard observées dans la plupart des graphes de terrains, présentées plus formellement dans la section précédente. La table 1 présente les résultats pour les quatre jeux de données de la section 2.1.

Table 1: Propriétés globales de la structure bipartie des réseaux

	Anime	Graph	Algebra	Monument
n_{\top}	888	122	796	1 295
n_{\perp}	12 009	2 225	10 646	19 345
m_b	32 086	3 267	15 513	21 596
$\delta_b (*10^{-2})$	0.30	1.20	1.81	0.87
k_{\top}	36.13	26.78	19.49	16.68
k_{\perp}	2.67	1.47	1.46	1.12
d_{\top}^+	1784	383	416	1 674
d_{\perp}^+	23	5	9	7

Comme attendu, les observations habituelles faites pour les réseaux réels sont aussi valables dans notre cas. En particulier, le réseau est dense (ordre de grandeur de 10^{-2} , 10^{-3}) et le degré maximal est quelques ordres de grandeurs plus grand que le degré moyen, ce qui indique en général un certain niveau d'hétérogénéité dans le degré des nœuds.

Ceci est confirmé par l'étude des distributions cumulatives inverses des degrés (à la fois pour les nœuds \top et \perp) présentées en figure 1 dans laquelle on voit clairement une distribution hétérogène, en queue lourde, pour les quatre jeux de données et les deux niveaux des réseaux.

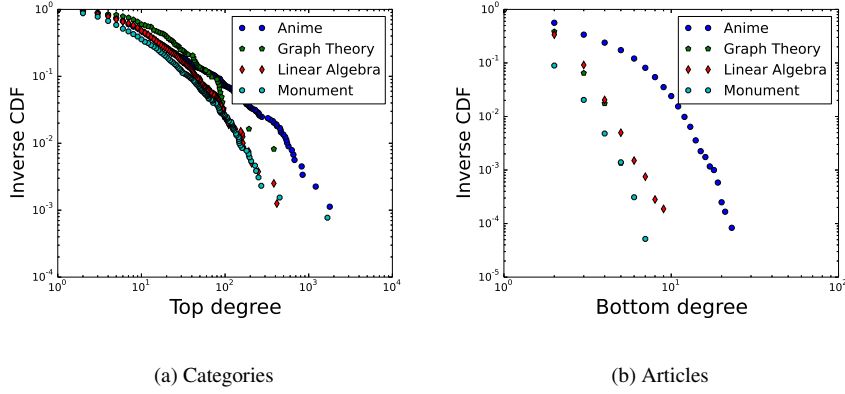


Figure 1: Distribution cumulative inverse des degrés

3.2. Analyse de la structure bipartie des graphes

On se concentre maintenant sur le cœur de l'analyse proposée dans le présent travail, à savoir l'étude des métriques liées aux recouvrement dans les graphes bipartis. Tout d'abord, la table 2 présente les valeurs globales synthétisant les valeurs observées en moyenne pour les nœuds des réseaux, pour les deux coefficients proposés et les quatre jeux de données.

Table 2: Propriétés liées aux recouvrements dans la structure bipartie des réseaux

	Anime	Graph	Algebra	Monument
cc_{bip}	0.03	0.05	0.07	0.06
rd_{bip}	0.33	0.13	0.21	0.09

Ce tableau montre tout de suite que, bien que proposés pour capturer la même propriété (les recouvrements), ils diffèrent nettement en fonction des jeux de données, sauf pour le cas de *Monument*. Pour les autres cas, ils diffèrent d'un ordre de grandeur, le cas le plus flagrant concernant *Anime* pour lequel le clustering global est très faible (0.03) alors que la redondance est plutôt élevée (0.33).

Ces statistiques moyennes et globales ne permettent pas de comprendre en détail le comportement des deux métriques étudiées. Il faut pour cela regarder les valeurs en chacun des nœuds, ce qui permet d'étudier plusieurs propriétés, comme la distribution de ces coefficients par exemple. La figure 2 présente la distribution cumulative inverse des distribution du coefficient de clustering (Figure 2a) et de redondance (Figure 2a).

On peut voir que les distributions sont extrêmement différentes. Pour le clustering, la courbe montre une décroissance brutale autour des faibles valeurs. La majorité

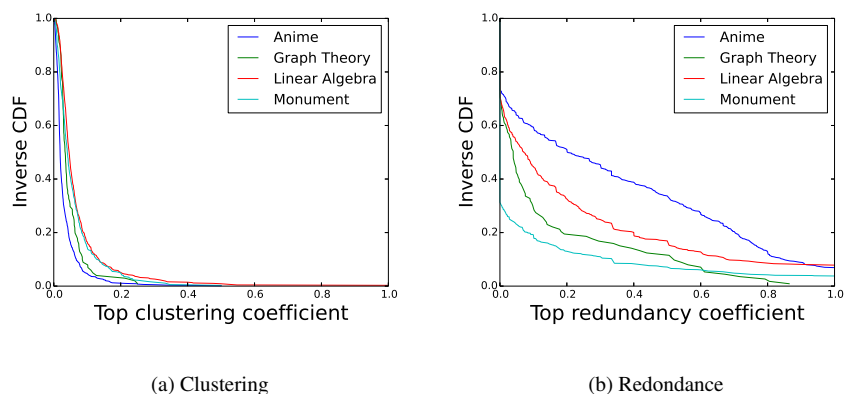


Figure 2: Distribution cumulative inverse des coefficients de clustering et de redondance

des catégories WIKIPEDIA ont ainsi un coefficient très faible. Ceci indique que les recouvrements, d'après cette métrique, ne sont pas très fréquents dans ces réseaux.

En ce qui concerne la redondance, le comportement est nettement différent. Mis à part pour le jeu de données *Monument* peut-être qui présente aussi une décroissance assez forte sur les petites valeurs, le coefficient semble pour le reste assez uniformément réparti sur le réseau, l'ensemble des valeurs possibles étant observable. À l'opposé du coefficient de clustering, celui-ci semble donc indiquer au contraire que les recouvrements sont présents dans ces réseaux. Notons en particulier qu'une fraction non négligeable des catégories ont une redondance observée égale à 1: 4% pour *Monument*, 7% pour *Anime* et 10% pour *Algebra*. Prenant ce dernier cas en exemple, cela signifie que pour 10% des catégories de ce jeu de données, tous les couples de pages ont *au moins* une autre catégorie en commun. Ceci indique un fort recouvrement entre communautés de pages dans le cas de ces catégories. Ceci n'est pas complètement surprenant dans le cas de WIKIPEDIA mais ce phénomène n'est apparemment pas capturé par le coefficient de clustering.

Les distributions présentées ci-dessus sont intéressantes en ce qu'elles permettent de discriminer entre l'utilité des deux métriques étudiées mais elles n'aident pas à comprendre *pourquoi* le clustering est particulièrement bas et pas la redondance. Afin de mieux comprendre la situation, nous avons regardé, figure 3, la corrélation entre la taille d'une catégorie et la valeur de son coefficient de clustering (figure 3a) et celle de son coefficient de redondance (figure 3b). Plus précisément, un point (x, y) dans ces figures signifie que la valeur moyenne du coefficient considéré pour une catégorie de taille x est y .

La figure 3a montre une légère corrélation qui est intéressante pour notre question: il semble que la valeur du clustering soit assez directement liée à celle de la taille

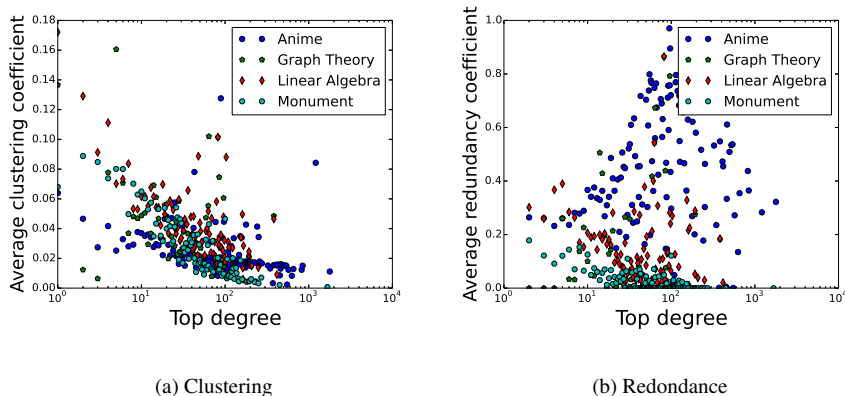


Figure 3: Corrélations entre le degré d'une catégorie et ces coefficients de recouvrement

de la communauté. Plus la communauté contient de pages, plus elle a tendance à avoir un clustering faible (en moyenne). Une telle corrélation n'est absolument pas observable pour la redondance, qui présente les mêmes valeurs pour toutes les tailles de communautés de pages (voir figure 3b). Il semble donc que l'information contenue dans la notion de *redondance* soit, de ce point de vue, plus riche que celle contenue dans la notion de *clustering*.

4. Conclusion

Dans ce papier, nous avons étudié les propriétés liées aux recouvrements observés dans la structure bipartite de différents jeux de données issus de WIKIPEDIA présentant une structure communautaire réelle (par opposition à détectée). La principale question abordée ici était d'être capable de discriminer entre deux métriques proposées pour rendre compte de ces recouvrements, à savoir le coefficient de clustering biparti et le coefficient de redondance.

À travers l'analyse de quatre jeux de données différents, nous avons montré que l'information capturée par la notion de clustering s'avère assez peu pertinente, en ce qu'elle ne détecte pas tous les schémas de recouvrements et semble avoir un lien fort avec la simple notion de taille d'une communauté. À l'inverse, le comportement du coefficient de redondance sur ces quatre jeux de données est plus intéressant. Il semble beaucoup moins prévisible au regard de simples propriétés locales.

Comprendre les caractéristiques de la structure bipartite des réseaux réels est fondamentale pour plusieurs raisons. Tout d'abord, comme montré à travers plusieurs études, cette structure permet de mieux comprendre des propriétés non triviales des graphes simples (cf (Tarissan *et al.*, 2013) par exemple). Mais, et c'est plus important

encore, il a été montré qu'elle constitue un meilleur support pour la définition de modèles permettant notamment de générer des structures semblables à celles observées dans les réseaux réels, mieux que les modèles classiques (Tarissan *et al.*, 2013).

De ce point de vue, le présent travail ouvre la voie à plusieurs améliorations des modèles récents. Il montre en particulier qu'on pourrait améliorer les modèles basés sur les graphes bipartis en intégrant dans de tels modèles les propriétés mises en évidence par le coefficient de redondance, mettant ainsi de côté celui de clustering qui semble moins pertinent. Une approche possible, comme suggérée dans (Tarissan, Coupechoux, 2013) pourrait consister à *coder* la redondance dans un troisième niveau afin de contrôler le coefficient par des permutations de liens dans la structure tripartite.

Mais ce travail amène également à d'autres approfondissements, suggérés ci-dessous.

4.1. Autres métriques.

Afin de mieux cerner le comportement du coefficient de redondance, qui semble après notre étude, le mieux adapté pour capturer les propriétés de recouvrement, on pourrait raffiner cette métrique en la mettant en relation avec d'autres. Par exemple, si la redondance d'une catégorie n'est pas liée à sa taille, elle est en revanche peut-être liée au degré des pages qu'elle contient. On pourrait ainsi mesurer la *dispersion* d'une catégorie en regardant le ratio entre le nombre de communautés voisines et le nombre de communautés voisines *qu'elle aurait dû avoir*, mesuré par exemple par la somme des degrés des pages qu'elle contient.

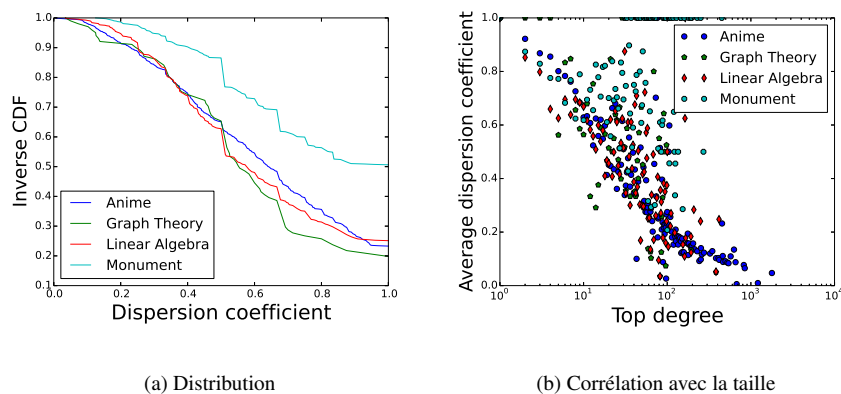


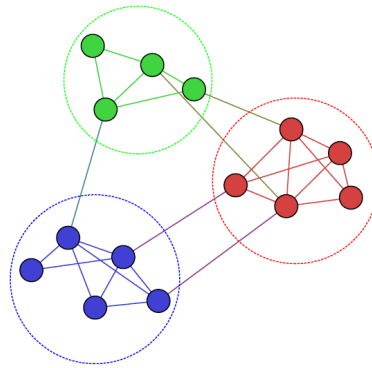
Figure 4: Coefficient de dispersion

La figure 4 montre ce que donne cette métrique sur les jeux de données du présent travail. Il semble que non seulement les valeurs sont réparties relativement uniformément dans le réseau (cf figure 4a) ce qui, comme la redondance, tend à rendre cette métrique pertinente mais, à la différence de la redondance, elle présente une corréla-

tion intéressante avec la taille de la communauté (cf figure 4b) qui demande à être étudiées plus avant.

4.2. Les graphes bipartis comme support pour la détection de communautés.

Comme remarqué plus haut, les *communautés* étudiées ici sont réelles, en ce qu'elles sont données *à priori* par les jeux de données. Ce sont simplement les catégories auxquelles sont liées les pages WIKIPEDIA. Mais cette notion de communauté existe aussi dans le cas des graphes simples, sans structure hiérarchique apparente. L'analyse des réseaux a en effet montré que l'une des caractéristiques importantes et non triviales communes à une grande majorité des réseaux est l'existence de communautés que l'on présente généralement comme des groupes de nœuds fortement liés entre eux et faiblement reliés à l'extérieur (Fortunato, 2010) comme dans la figure suivante :



Dans ce contexte, la problématique la plus répandue consiste à s'intéresser à la *détection automatique* de communautés, c'est-à-dire à la mise au point d'algorithmes efficaces capables de révéler la structure communautaire des réseaux réels. La pratique la plus courante étant alors de rechercher une partition des sommets du graphe qui fasse sens (Blondel *et al.*, 2008; Rosvall, Bergstrom, 2008). Mais on peut également voir la structure communautaire comme recouvrante, où chaque sommet peut appartenir à plusieurs communautés (Evans, Lambiotte, 2009; Lancichinetti *et al.*, 2009; Palla *et al.*, 2005). Dans cette vision, le résultat des détections peut être vue à nouveau comme un graphe biparti, dans lequel les nœuds \top représentent les communautés et les nœuds \perp représentent les sommets d'origine. Les liens du graphe biparti relient les communautés de \top aux sommets qu'elles contiennent dans \perp et permettent d'identifier les liens d'appartenance des nœuds à leur(s) communauté(s).

Il serait alors particulièrement pertinent, dans ce contexte, de comparer la structure bipartite (et en particulier la redondance) observée dans les réseaux réels (tels ceux issus de WIKIPEDIA) à celle obtenue après une phase de détection générant des struc-

tures communautaires recouvrantes. Nous laissons cette perspective intéressante pour un travail ultérieur.

References

- Ahn Y.-Y., Ahnert S. E., Bagrow J. P., Barabási A.-L. (2011). Flavor network and the principles of food pairing. *Scientific reports*, Vol. 1.
- Battiston S., Catanzaro M. (2004). Statistical properties of corporate board and director networks. *The European Physical Journal B-Condensed Matter and Complex Systems*, Vol. 38, No. 2, pp. 345–352.
- Blondel V., Guillaume J., Lambiotte R., Lefebvre E. (2008). Fast unfolding of communities in large networks. *Journal of Statistical Mechanics: Theory and Experiment*, Vol. 2008, No. 10, pp. P10008.
- Cancho R. F. i, Solé R. V. (2001). The small world of human language. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, Vol. 268, No. 1482, pp. 2261–2265.
- Evans T., Lambiotte R. (2009). Line graphs, link partitions, and overlapping communities. *Physical Review E*, Vol. 80, No. 1, pp. 016105.
- Fortunato S. (2010). Community detection in graphs. *Physics Reports*, Vol. 486, No. 3, pp. 75–174.
- Guillaume J.-L., Latapy M. (2006). Bipartite graphs as models of complex networks. *Physica A: Statistical Mechanics and its Applications*, Vol. 371, No. 2, pp. 795–813.
- Lancichinetti A., Fortunato S., Kertész J. (2009). Detecting the overlapping and hierarchical community structure in complex networks. *New Journal of Physics*, Vol. 11, No. 3, pp. 033015.
- Latapy M., Magnien C., Vecchio N. D. (2008). Basic notions for the analysis of large two-mode networks. *Social Networks*, Vol. 30, No. 1, pp. 31–48.
- Le Fessant F., Handurukande S., Kermarrec A.-M., Massoulié L. (2005). Clustering in peer-to-peer file sharing workloads. In *Peer-to-peer systems iii*, pp. 217–226. Springer.
- Mérindol P., Donnet B., Bonaventure O., Pansiot J.-J. (2010, November). On the impact of layer-2 on node degree distribution. In *Proc. ACM/USENIX internet measurement conference (IMC)*.
- Mérindol P., Schriek V. Van den, Donnet B., Bonaventure O., Pansiot J.-J. (2009, November). Quantifying ASes multiconnectivity using multicast information. In *Proc. ACM/USENIX internet measurement conference (IMC)*.
- Newman M. E., Strogatz S. H., Watts D. J. (2001). Random graphs with arbitrary degree distributions. *Physics Reviews E*, Vol. 64.
- Newman M. E., Watts D. J., Strogatz S. H. (2002). Random graph models of social networks. *Proceedings of the National Academy of Sciences of the United States of America*, Vol. 99, No. Suppl 1, pp. 2566–2572.
- Omodei E., Cointet J.-P., Poibeau T. (2013). A symmetric approach to understand the dynamics of scientific collaborations and knowledge production. In *Proc. conf. conférence sur les modèles et l'analyse des réseaux : Approches mathématiques et informatiques*.

- Palla G., Derényi I., Farkas I., Vicsek T. (2005). Uncovering the overlapping community structure of complex networks in nature and society. *Nature*, Vol. 435, No. 7043, pp. 814–818.
- Prieur C., Cardon D., Beuscart J.-S., Pissard N., Pons P. (2008). The strength of weak cooperation: A case study on flickr. *arXiv preprint arXiv:0802.2317*.
- Rosvall M., Bergstrom C. T. (2008). Maps of random walks on complex networks reveal community structure. *Proceedings of the National Academy of Sciences*, Vol. 105, No. 4, pp. 1118–1123.
- Tarissan F., Coupechoux E. (2013). Un modèle pour les graphes bipartis aléatoires avec redondance. In *Acte de la 4ème conférence sur les modèles et l'analyse des réseaux : Approches mathématiques et informatique (marami'13)*, pp. 1–10.
- Tarissan F., Latapy M., Prieur C. (2009). Efficient measurement of complex networks using link queries. In *Proceedings of the ieee international workshop on network science for communication networks (netscom'09)*, pp. 1–6.
- Tarissan F., Nollez-Goldbach R. (2013). The network of the international criminal court decisions as a complex system. In A. Sanayei, I. Zelinka, O. E. Rossler (Eds.), *Iscs 2013: Interdisciplinary symposium on complex systems*, Vol. 8, p. 1-9. Springer.
- Tarissan F., Quoitin B., Mérimodol P., Donnet B., Pansiot J.-J., Latapy M. (2013). Towards a bipartite graph modeling of the internet topology. *Computer Networks*, Vol. 57, No. 11, pp. 2331–2347.
- Tumminello M., Miccichè S., Lillo F., Piilo J., Mantegna R. N. (2011). Statistically validated networks in bipartite complex systems. *PloS one*, Vol. 6, No. 3, pp. e17994.
- Watts D. J., Strogatz S. H. (1998). Collective dynamics of 'small-world' networks. *nature*, Vol. 393, No. 6684, pp. 440–442.