



HAL
open science

Un modèle de génération de graphes “ petit monde ” imitant les réseaux sociaux

Jason Vallet, Bruno Pinaud, Guy Melançon

► To cite this version:

Jason Vallet, Bruno Pinaud, Guy Melançon. Un modèle de génération de graphes “ petit monde ” imitant les réseaux sociaux. 7ème conférence sur les Modèles et l'Analyse des Réseaux: Approches Mathématiques et Informatiques (MARAMI), Oct 2016, Cergy-Pontoise, France. hal-01416524

HAL Id: hal-01416524

<https://hal.science/hal-01416524v1>

Submitted on 14 Dec 2016

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.

Un modèle de génération de graphes « petit monde » imitant les réseaux sociaux

Jason Vallet, Bruno Pinaud, et Guy Melançon

Univ. Bordeaux, LaBRI, UMR CNRS 5800
F-33405 Talence, France
prenom.nom@labri.fr

Résumé

Lors de la construction ou l'évaluation de méthodes d'analyse de réseaux sociaux, il est important de pouvoir tester sa solution sur des données viables. Cependant, et malgré leur grande popularité, les structures internes de ces réseaux, indiquant les relations inter-individus, ne sont que rarement ouvertes ou diffusées, forçant les chercheurs à se tourner vers des réseaux construits par des modèles de génération aléatoire. Nous présentons un nouveau modèle de génération permettant la création de graphes. Notre solution est paramétrable et peut être utilisée pour construire un large éventail de graphes, allant des réseaux sociaux petit monde –avec certaines caractéristiques des réseaux invariants d'échelle– à des graphes complètement aléatoires.

1 Introduction

Bien que de nombreuses plate-formes de réseaux sociaux existent, les données les concernant, et plus particulièrement la structure exhaustive de ces réseaux, ne sont que rarement mises à disposition du public ou des chercheurs. Afin de valider leurs modèles, algorithmes ou approches analytiques, les spécialistes de l'analyse des réseaux sociaux n'ont souvent d'autre choix que de se contenter de réseaux générés aléatoirement. Cette approche n'est toutefois pas sans avantages puisque les propriétés des réseaux générés, telles que le nombre d'utilisateurs ou la densité de connexions, peuvent alors être paramétrées afin de générer des réseaux aux caractéristiques souhaitées. Cette création sur-mesure permet ainsi d'évaluer dans différentes conditions les solutions que l'on souhaite tester.

L'utilisation de générateurs de données n'est pas réservée aux seuls réseaux sociaux, en effet, de nombreux modèles permettant de créer des graphes aux caractéristiques spécifiques existent. Le plus connu d'entre eux est sans nul doute le modèle *ER*, proposé par Erdős et Rényi [ER59] à la fin des années 50. Bien que nous définissions un réseau social comme un graphe $G = (V, E)$, où V désigne l'ensemble des membres du réseau comme ses nœuds et $E \subseteq V \times V$ l'ensemble de liens reliant certains nœuds ensemble, n'importe quel générateur ne peut être utilisé pour obtenir un graphe s'apparentant à un réseau social.

Nous sommes intéressés par deux types de graphe particuliers souvent associés aux réseaux sociaux : les réseaux « petit monde » et les réseaux invariants d'échelle. La première catégorie a été introduite par Watts et Strogatz [WS98] et baptisé d'après le

phénomène du même nom décrit par Milgram [Mil67]. Un graphe peut être qualifié de « petit monde » lorsque son coefficient d'agglomération moyen est haut tandis que la longueur moyenne de ses chemins s'avère faible.¹ Les réseaux invariants d'échelle, présenté par Barabási et Albert [BA99], sont, quant à eux, identifiables par la distribution des degrés de leur nœuds : cette dernière devant suivre une loi de puissance pour que le réseau soit désigné ainsi. De nombreuses variations de ces deux modèles initiaux existent et plusieurs proposent notamment de générer des réseaux présentant à la fois les caractéristiques des réseaux « petit monde » et invariant d'échelle (par exemple : [HK02, KE02, SZM13, WDDS06]). Chacun de ces modèles peut ainsi générer des réseaux aux caractéristiques données. Néanmoins, les démarches suivies lors des processus de génération sont complètement orientées vers une approche mathématique et manquent du caractère organique inhérent aux réseaux de relations entre êtres humains.

Nous présentons dans ce papier un nouveau modèle de génération de graphe. Celui-ci à la particularité d'être conçu de façon à construire un « simili » réseau social car suivant les règles établies par les plate-formes de réseaux sociaux actuelles (relations pair-à-pair et visibilité des voisins de voisins). Les transformations et constructions que nous orchestrons suivent également les comportements sociaux observés et décrits dans la littérature (voir fin de section 2). Le modèle propose un nombre de nœuds et d'arêtes paramétrable et l'utilisateur peut choisir le pourcentage de connexions aléatoires à rajouter dans le réseau. Ce point permet à notre modèle de générer un large spectre de graphes allant du graphe complètement aléatoire au réseau « petit monde » avec une distribution de degrés s'approchant d'un réseau invariant d'échelle.

Dans la suite de ce papier, nous présentons notre modèle et décrivons la démarche de génération utilisée. Une analyse des réseaux générés est ensuite proposée pour établir leurs caractéristiques en fonction des paramètres de génération choisis et quelques exemples de graphes construits par notre modèle sont présentés.

2 Présentation de notre solution

Afin de créer un modèle se rapprochant au plus près d'un réseau social réel, il nous faut cerner puis recréer au mieux le comportement des membres de ces réseaux. À l'opposé des approches mathématiques des autres modèles, nous utilisons ici une approche imitant les démarches suivies dans un véritable réseau social. Le processus génératif est divisé en trois phases, chacune permettant d'accomplir un aspect différent de l'évolution des liens entre personnes au sein du réseau. Dans un premier temps, nous générons un graphe simple et acyclique, connectant chaque individu rejoignant le réseau à la personne qui l'a invité. Lors de la seconde phase, des arêtes additionnelles sont créées aléatoirement lorsque les individus se connectent avec les personnes qu'ils connaissent à l'extérieur du réseau. La dernière phase permet finalement de créer ou renforcer des communautés : les membres du réseau se familiarisent et se connectent avec les amis de leurs amis.

Initialisation et paramètres Avant de démarrer, les paramètres du modèle doivent être établis. Par défaut, le processus que nous proposons permet de créer un graphe $G = (V, E)$, de $|V|$ nœuds et $|E|$ arêtes, simple, connexe et qui sera orienté ou non selon le choix de l'utilisateur. Nous introduisons également le paramètre $\rho \in [0; 1]$ qui spécifiera le ratio d'arêtes aléatoires qui seront créées par le modèle lors de la génération.

1. Les formules pour calculer ces mesures sont spécifiées dans la section 3.

Le graphe que nous générons dans cette approche et les analyses que nous réalisons par la suite ont été réalisées sur des graphes non orientés, néanmoins, l'implémentation du modèle discutée en section 3 gère l'orientation des arêtes. Nous créons G de façon à obtenir une seule composante connexe, néanmoins, plusieurs composantes non connexes peuvent être construites en réutilisant le modèle successivement et en adaptant les valeurs $|V|$ et $|E|$ passées comme paramètres. Nous pouvons établir que les limites autorisées du nombre d'arêtes $|E|$ sont telles que suit : (1) puisque G est simple et non orienté², sa borne supérieure est définie telle que $|E| \leq \frac{|V| \times (|V| - 1)}{2}$, et (2) comme G est connexe, sa borne inférieure est telle que : $|E| \geq |V| - 1$. Notre modèle peut garantir le nombre de nœuds et d'arêtes qui sera généré tant que les valeurs demandées restent dans les limites permises.

Phase 1 : création des nœuds La première étape permet l'ajout de nouveaux individus au sein du réseau social. Chaque fois qu'une personne rejoint un réseau social, son nombre de connexions est très limité, et souvent ne représente que l'utilisateur qui l'a introduit dans le réseau. Cette nouvelle personne a néanmoins maintenant l'opportunité d'inviter un ou plusieurs amis à la rejoindre. Bien évidemment, chaque ami joignant le réseau est par la suite considéré comme un nouvel utilisateur qui peut également à son tour inviter d'autres personnes.

Nous utilisons l'algorithme 1 pour initier la génération de G . Le graphe obtenu en résultat comporte le nombre attendu de nœuds ($|V|$) mais n'atteint pas encore le nombre d'arêtes escomptées. Cette première étape nous retourne en effet un graphe connexe, simple et acyclique, présentant donc une topologie similaire à un arbre. Celui-ci représente le réseau initial évoluant alors que de nouveaux utilisateurs sont rajoutés et connectés à la personne les ayant invité. Les phases suivantes vont nous permettre d'enrichir le graphe en arêtes.

Algorithme 1 : Création du graphe et ajout des individus dans le réseau social

Entrées : n : nombre de noeuds à créer

Sorties : $G = (V, E')$ avec $|V| = n$ et $|E'| = n - 1$

initialisation de G

création d'un noeud dans G

$i \leftarrow 1$

// G n'aura jamais plus de n noeuds

tant que $i < n$ **faire**

$i \leftarrow i + 1$

$u \leftarrow$ un noeud aléatoire de G

$v \leftarrow$ création d'un noeud dans G

 création d'une arête (u, v) dans G

fin

Phase 2 : ajout d'arêtes aléatoires La seconde phase permet de mettre en contact des membres du réseau social de manière aléatoire. Cette étape sert à reproduire le processus de mise en relation entre individus se connaissant soit avant de rejoindre le réseau ou en dehors de celui-ci. L'ajout de ces liens pré-existants va permettre de réduire le diamètre du graphe généré, ainsi que la longueur moyenne de ses chemins (voir Déf. 1). Le nombre d'arêtes aléatoires ajoutées au réseau de cette manière est décidé par la valeur choisie pour le paramètre ρ . Ce dernier est un ratio, compris entre

2. Respectivement, $|E| \leq |V| \times (|V| - 1)$ si G est orienté.

Algorithme 2 : Ajout des arêtes aléatoires

Entrées : $G = (V, E')$,
 ρ : ratio d'arêtes à rajouter aléatoirement à G ,
 $|E|$: nombre d'arêtes à atteindre à la fin de la génération

Sorties : $G = (V, E'')$ avec $|E''| = |E'| + \rho \cdot (|E| - |E'|)$

$limite \leftarrow \rho \cdot (|E| - |E'|)$
 $noeudsU \leftarrow$ ensemble des noeuds de G
 $i \leftarrow 0$
// limite arêtes sont ajoutées à G

tant que $i < limite$ **ET** nombre d'éléments dans $noeudsU > 0$ **faire**

$u \leftarrow$ un élément aléatoire de $noeudsU$
 $noeudsV \leftarrow$ ensemble des noeuds de $G - u$
 $arêteCréée \leftarrow faux$

tant que nombre d'éléments dans $noeudsV > 0$ **ET** $arêteCréée = faux$ **faire**

faire

$v \leftarrow$ un élément aléatoire de $noeudsV$
 suppression de v de $noeudsV$
 si l'arête (u, v) n'existe pas dans G **alors**

 création d'une arête (u, v) dans G
 $i \leftarrow i + 1$
 $arêteCréée \leftarrow vrai$

fin

fin

si $arêteCréée = faux$ **alors**

 suppression de u de $noeudsU$

fin

fin

0 et 1 inclus, qui permet d'établir, parmi les arêtes restant à ajouter au début de la phase actuelle, combien seront construites aléatoirement.

L'algorithme 2 présente la démarche suivie lors de cette phase. Les connexions sont réalisées entre deux individus pris au hasard dans le graphe, mais il convient néanmoins de s'assurer que le nombre d'arêtes créées est en adéquation avec le résultat demandé afin de garantir les propriétés du graphe généré à la fin. La modulation du paramètre ρ nous permet de générer différents types de réseaux, allant d'un graphe complètement aléatoire lorsque $\rho = 1$, à un graphe où la phase actuelle sera complètement ignorée si $\rho = 0$, laissant la génération de toutes les arêtes à la phase 3. Dans ce dernier cas, le modèle construira un réseau où les membres ne peuvent se reconnaître que par l'intermédiaire du réseau et où leurs cercles de connaissance ne pourront donc grandir qu'en passant par des amis communs déclarés comme appartenant eux aussi au réseau social.

Phase 3 : renforcement des communautés Cette étape correspond au moment où les membres du réseau commencent à nouer des relations avec les autres membres qui partagent les mêmes amis. Deux personnes qui ne se connaissaient pas auparavant peuvent ainsi potentiellement devenir amis, ce qui mènera à la création d'un lien entre eux. La construction d'arêtes suivant ce schéma permet le renforcement des communautés, augmentant en conséquence le coefficient d'agglomération (voir Déf. 2) du

Algorithme 3 : Ajout des arêtes renforçant les communautés

Entrées : $G = (V, E'')$,
 $|E|$: nombre d'arêtes à atteindre à la fin de la génération
Sorties : $G = (V, E)$

$limite \leftarrow |E| - |E''|$
 $noeuds \leftarrow$ ensemble des noeuds de G
 $i \leftarrow 0$
// limite arêtes sont ajoutées à G
tant que $i < limite$ *ET* nombre d'éléments dans $noeuds > 0$ **faire**

$u \leftarrow$ un élément aléatoire de $noeuds$
 $voisins \leftarrow$ ensemble formé par les noeuds voisins de u , ordonnés aléatoirement
 $arêteCréée \leftarrow faux$
 tant que $|voisins| > 0$ *ET* $arêteCréée = faux$ **faire**

$v \leftarrow$ un élément aléatoire de $voisins$
 suppression de v de $voisins$
 pour chaque $w \leftarrow$ un élément de $voisins$ **faire**

si l'arête (v, w) n'existe pas dans G **alors**

 création d'une arête (v, w) dans G
 $i \leftarrow i + 1$
 $arêteCréée \leftarrow vrai$
 sortie de la boucle

fin

fin

si $arêteCréée = faux$ **alors**

 suppression de u de $noeuds$

fin

fin

réseau généré.

L'algorithme 3 décrit les opérations réalisées pour assurer la bonne construction des arêtes. Le nombre d'éléments à générer durant cette étape est directement dépendant du nombre de connexions ajoutées dans la phase 2, donc de ρ , et du nombre total d'arêtes à générer. Tout comme précédemment, l'algorithme est implémenté de manière à assurer la création du nombre attendu de liens afin de garantir que le graphe généré G ait bien les caractéristiques attendues.

Généralisation aux réseaux orientés Les instructions spécifiées dans les algorithmes concernant la création des arêtes peuvent aisément être adaptées de manière à permettre leur orientation. Dans l'algorithme 1, il suffit ainsi de créer aléatoirement, entre les noeuds u et v , soit l'arête (u, v) , soit (v, u) . Dans les deux autres algorithmes, il faut évidemment vérifier l'existence de l'arête orientée (u, v) –respectivement (v, w) pour l'algorithme 3– mais il est possible d'ajouter l'arête inverse (v, u) (resp. (w, v)), une fois son existence non avérée, si la première existe déjà.

Lors de la phase 3, le processus de pensée suivit pour la construction des arêtes orientées repose sur le principe de l'accointance entre personnes, ainsi que sur l'évolution des relations entre individus appartenant à des cercles sociaux proches. Plus par-

ticulièrement, nous nous sommes inspirés de l'idée de balance sociale initialement développée par Heider [Hei46] puis reprise et généralisée par Cartwright *et al.* [CH56]. Dans leurs travaux, ces auteurs décrivent les relations à trois individus (triades) et étudient les différents scénarios pouvant se réaliser lorsqu'une personne réagit à l'opinion de deux autres. Bien que, dans ces études, les liens entre individus peuvent être marqués comme positifs ou négatifs, peu de plate-formes de réseaux sociaux proposent cette distinction. Nous pouvons noter tout de même comme exceptions celles proposant des systèmes de vote (comme pour l'élection des administrateurs de Wikipédia [LHK10a, WPLP14]) ou de transmission d'opinion (comme la notation de produits ou la confiance en un revendeur [CCC+11, LHK10b, RAD03]). En conséquence, nous nous basons sur un travail plus récent de Leskovec *et al.* [LHK10b] afin de sélectionner les différents contextes de triades à signe positif uniquement et les utiliser dans notre modèle.

3 Analyse des graphes générés

Afin d'analyser les réseaux créés par notre modèle, nous devons étudier certaines caractéristiques des graphes construits. Bien que la connaissance du nombre de nœuds et de liens nous indique la densité du réseau, ce paramètre ne permet en aucun cas de définir entièrement le graphe créé. Nous proposons donc de nous appuyer sur deux mesures bien connues de la communautés afin d'étudier les graphes générés en fonction des paramètres soumis à l'algorithme. Les deux mesures utilisées sont la *longueur moyenne des chemins* et le *coefficient d'agglomération* (respectivement, et plus communément, désignés en anglais sous les termes de *Characteristic Path Length* et *Clustering Coefficient*) telles que définies dans [WS98].

DÉFINITION 1 (Longueur moyenne des chemins). *La longueur moyenne des chemins est définie par la moyenne du nombre de liens dans le chemin le plus court entre toute paire de nœuds. Soit un graphe $G = (V, E)$, la longueur moyenne des chemins L se calcule par la formule :*

$$L = \frac{1}{|V| \times (|V| - 1) / 2} \sum_{\forall a, b \in V} d_{a,b}$$

où $d_{a,b}$ représente le nombre minimal de liens à emprunter pour relier a à b .

DÉFINITION 2 (Coefficient d'agglomération). *Dans un graphe $G = (V, E)$, un nœud $v \in V$ possède k_v voisins. En conséquence, un maximum de $\frac{k_v(k_v-1)}{2}$ arêtes peuvent exister entre tous ses voisins ; ce cas ne se produit que lorsque tous les voisins de v , noté $N(v)$, sont entre-connectés deux à deux. Nous désignons C_v comme le pourcentage existant de ces liens entre les k_v voisins de v ,*

$$C_v = \frac{1}{k_v(k_v - 1) / 2} \sum_{\forall a, b \in N(v)} \Delta(a, b)$$

$$\text{avec } \Delta(a, b) = \begin{cases} 1 & \text{si } d_{a,b} \text{ est égal à } 1 \\ 0 & \text{sinon} \end{cases}$$

et nous définissons le coefficient d'agglomération C comme la valeur moyenne de ce pourcentage pour l'ensemble des nœuds du graphe :

$$C = \frac{1}{|V|} \sum_{\forall v \in V} C_v$$

Afin d'étudier l'évolution de ces deux mesures pour notre modèle en fonction des paramètres choisis, nous avons générés plusieurs graphes avec, pour paramètre fixé, leur nombre de nœuds ($|V| = 1000$) et, comme paramètres variables, le nombre d'arêtes ($|E|$ est porté à 2 000, 5 000, 10 000, 20 000, 50 000 et 100 000) ainsi que le ratio ρ d'arêtes créées aléatoirement. Concernant cette dernière variable, nous avons initialement prévu de la tester en avançant par tranche de 10% mais nos premiers tests nous ont permis d'établir que beaucoup de changements se produisaient dans le premier décile, aussi avons-nous réalisé des itérations par pas de 1% pour les dix premières mesures puis par pas de 10% par la suite. Enfin, nous avons réalisé 10 mesures pour chaque tuple de ces paramètres et utilisé les valeurs moyennes obtenues pour étudier les résultats.

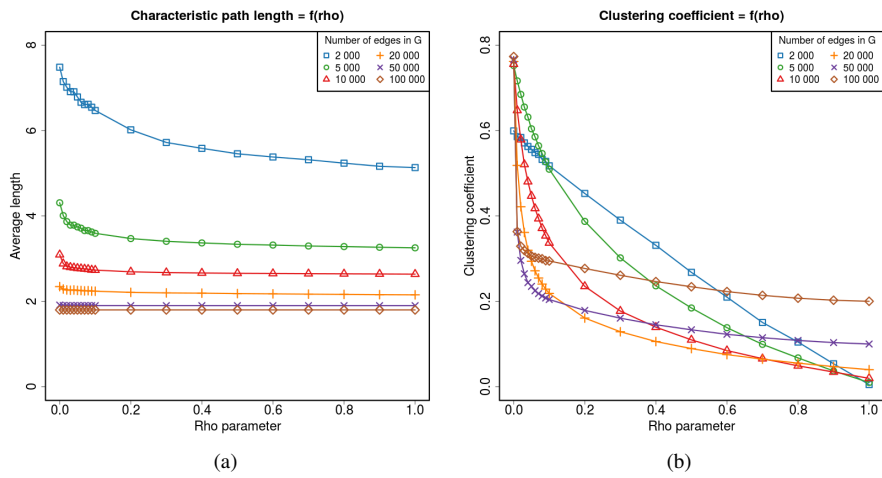


FIGURE 1 – Suivi, en fonction de la valeur de ρ , de l'évolution de (a) la longueur moyenne des chemins et (b) du coefficient d'agglomération pour des graphes de 1 000 nœuds. Le nombre de liens pour chaque graphe testé est spécifié dans la légende.

Les courbes (a) et (b) de la Figure 1 représentent respectivement les suivis de l'évolution de la longueur moyenne des chemins ainsi que du coefficient d'agglomération, tous deux calculés pour chacun des graphes testés en fonction de la valeur de ρ choisie. L'analyse des longueurs moyennes observées, voir Fig. 1a, ne présente pas de réelle surprise, en effet, il est connu que plus la densité d'arêtes est élevée dans un graphe simple, plus la distance du chemin moyen en est réduite. Lorsque ρ est initialisé à 1, le graphe généré sera l'équivalent d'un graphe aléatoire tel que défini par le modèle d'Erdős-Rényi [ER59], et la longueur moyenne atteint alors son minimum, quel que soit son nombre d'arêtes considérées. Nous nous apercevons tout de même que plus la densité d'arêtes ajoutées au graphe est importante, plus ce minimum sera atteint rapidement, et ce pour des valeurs de plus en plus petites de ρ .

L'étude de l'évolution du coefficient d'agglomération en fonction de ρ et de la densité d'arêtes (voir Fig. 1b) nous renseigne sur les conditions limites du modèle. Lorsque ρ est égal à zéro, le coefficient d'agglomération atteint pratiquement $\frac{1}{5}$ de sa valeur maximale pour tous les graphes sauf celui avec 2 000 arêtes où le nombre limité de connexions de ce dernier l'empêche d'atteindre une valeur aussi haute que les autres graphes. Pour les valeurs intermédiaires de ρ , le coefficient d'agglomération décroît

plus rapidement pour les graphes avec une forte densité d'arêtes. Si nous poussons la valeur du paramètre ρ à 1, le coefficient d'agglomération retombe aux alentours de 0, à part pour les graphes les plus denses où leur grand nombre d'arêtes garantit des valeurs minimales d'agglomération plus élevées que pour les autres.

Avec ces résultats, nous nous apercevons que le modèle permet de générer des graphes avec une longueur de chemin moyenne courte et un coefficient d'agglomération élevé si l'on utilise des petites valeurs pour le paramètre ρ . Ces deux caractéristiques combinées nous indiquent que les réseaux ainsi obtenus sont de type « petit monde ».

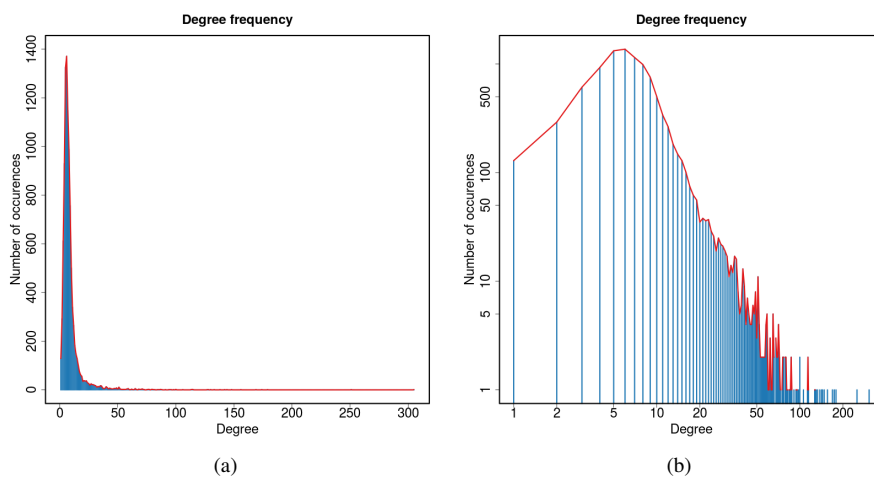


FIGURE 2 – Histogramme de la fréquence d'apparition des degrés pour un graphe G de 10 000 nœuds et 50 000 arêtes. Les deux graphiques représentent la même information mais (b) utilise une échelle logarithmique.

Nous souhaitons également découvrir si le réseau généré pourrait être qualifié de réseau invariant d'échelle. Pour avoir une idée indicative avant tout calcul poussé, nous pouvons observer la distribution des fréquences des degrés dans un des graphes générés. La Figure 2 montre une telle distribution pour un réseau de 10 000 nœuds et 50 000 arêtes, généré pour $\rho = 0$. Le détail donnée par la sous-figure 2b nous montre que la fréquence des degrés ne suit pas exactement une loi de puissance, qui devrait former une pente décroissante assez régulière lorsque représentée sur une échelle log-log. La distribution observée s'en approche néanmoins énormément si l'on ne tient pas compte des nœuds de degré inférieur à 6. Plusieurs graphes ont été générés avec différents paramètres afin d'observer l'évolution de la distribution des fréquences mais sans résultat concluant. Néanmoins, nous avons remarqué qu'augmenter la valeur de ρ fait transiter le réseau vers un graphe aléatoire, changeant en conséquence sa distribution de degré en une loi de Poisson typique de ce type de graphe. Au bout du compte, nous ne pouvons donc pas qualifier notre modèle comme générant des réseaux invariants d'échelle mais seulement des graphes s'en rapprochant lorsque ρ est très petit.

Pour finir, nous présentons quelques représentations de graphes générés à l'aide de notre modèle dans la Figure 3. Les réseaux proposés ici sont volontairement laissés petits afin de les rendre facile à visualiser et de permettre l'observation de l'impact qu'a le paramètre ρ sur les structures obtenues. Cependant, l'implémentation du modèle

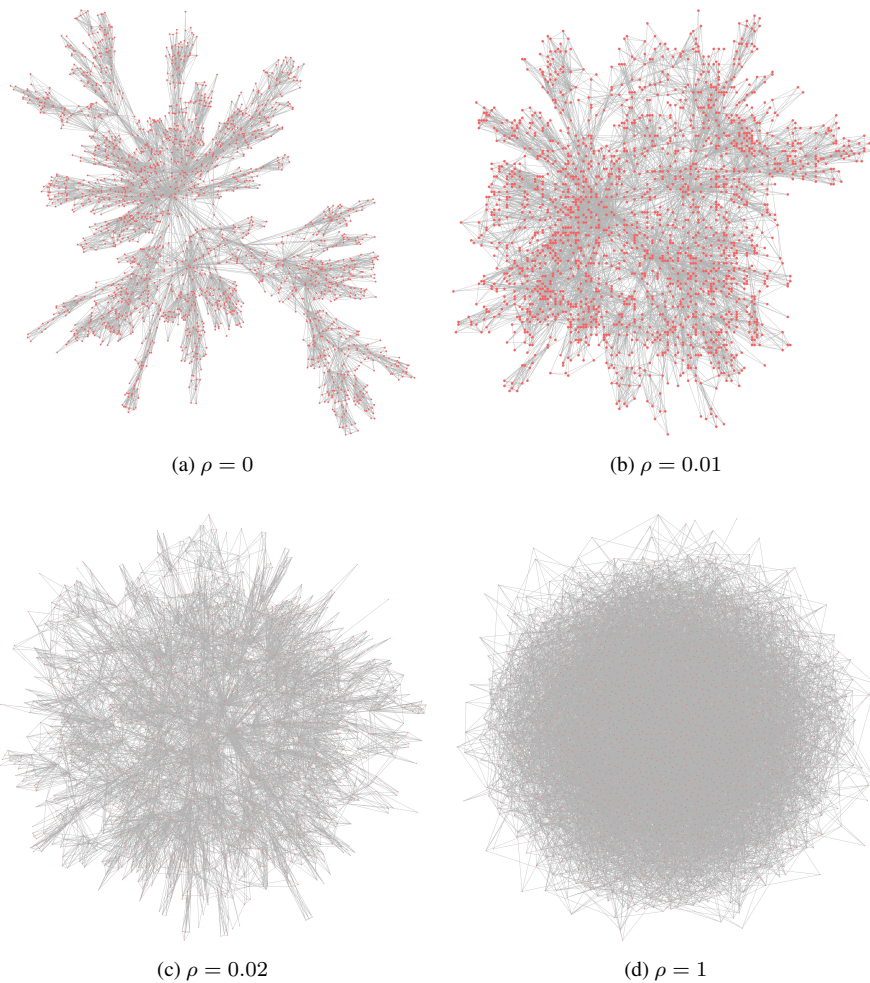


FIGURE 3 – Exemples de réseaux générés par notre modèle. Chacun des graphes possède 2 000 nœuds et 10 000 arêtes. Le ratio ρ utilisé lors de chacune des générations est donné en légende pour chaque graphe.

réalisée en C++ à travers le logiciel de visualisation d'information Tulip³ [AAB⁺14] permet la génération de grands graphes très facilement. Par exemple, la création d'un réseau comptant un million de nœuds et dix millions d'arêtes (avec $\rho = 0.05$) a été réalisée en environ 15 secondes sur une station de travail relativement récente.

4 Conclusion

Nous avons présenté dans cet article un nouveau modèle de génération de graphes. Bien que celui-ci ait été créé en s'inspirant des démarches existantes sur les plateformes de réseaux sociaux, notre solution reste modulaire de telle façon que les graphes

3. Tulip est disponible sur le site <http://tulip.labri.fr>

produits peuvent proposer différentes caractéristiques selon les paramètres utilisés lors de la génération. Nous avons montré que le modèle permettait notamment de construire des réseaux « petit monde » et que, lorsque la valeur de ρ était adaptée, la distribution de ses degrés pouvait se rapprocher de celles rencontrées dans les réseaux invariants d'échelle.

Ce modèle a été produit en se basant entièrement sur une méthode empirique, reproduisant de manière automatique les comportements qui avaient été observés dans des cas réels. Le résultat s'est avéré néanmoins des plus satisfaisant, marquant ce type d'approche comme tout à fait fonctionnel. Pour nos travaux futurs, nous prévoyons de nous intéresser à la modélisation d'autres types de réseaux sociaux et notamment ceux supportant les relations positives, négatives, ou marquées par un spectre de ressentis des membres entre-eux.

Références

- [AAB⁺14] David Auber, Daniel Archambault, Romain Bourqui, Maylis Delest, Jonathan Dubois, Bruno Pinaud, Antoine Lambert, Patrick Mary, Morgan Mathiaut, and Guy Melançon. Tulip III. In Reda Alhajj and Jon Rokne, editors, *Encyclopedia of Social Network Analysis and Mining*, pages 2216–2240. Springer New York, 2014.
- [BA99] Albert-László Barabási and Réka Albert. Emergence of scaling in random networks. *Science*, 286(5439) :509–512, 1999.
- [CCC⁺11] Wei Chen, Alex Collins, Rachel Cummings, Te Ke, Zhenming Liu, David Rincón, Xiaorui Sun, Yajun Wang, Wei Wei, and Yifei Yuan. Influence maximization in social networks when negative opinions may emerge and propagate. In *Proc. of the 11th SIAM Int. Conf. on Data Mining, SDM*, pages 379–390, 2011.
- [CH56] Dorwin Cartwright and Frank Harary. Structural balance : a generalization of Heider's theory. *Psychological Review*, 63 :277–293, 1956.
- [ER59] Paul Erdős and Alfréd Rényi. On Random Graphs I. *Publicationes Mathematicae (Debrecen)*, 6 :290–297, 1959.
- [Hei46] Fritz Heider. Attitudes and cognitive organization. *The Journal of Psychology*, 21(1) :107–112, 1946.
- [HK02] Petter Holme and Beom Jun Kim. Growing scale-free networks with tunable clustering. *Phys. Rev. E*, 65 :026107, Jan 2002.
- [KE02] Konstantin Klemm and Víctor M. Eguíluz. Growing scale-free networks with small-world behavior. *Phys. Rev. E*, 65 :057102, May 2002.
- [LHK10a] Jure Leskovec, Daniel Huttenlocher, and Jon Kleinberg. Predicting positive and negative links in online social networks. In *Proc. of the 19th Int. Conf. on World Wide Web, WWW '10*, pages 641–650. ACM, 2010.
- [LHK10b] Jure Leskovec, Daniel Huttenlocher, and Jon Kleinberg. Signed networks in social media. In *Proc. of the SIGCHI Conf. on Human Factors in Computing Systems, CHI '10*, pages 1361–1370. ACM, 2010.
- [Mil67] Stanley Milgram. The small world problem. *Psychology Today*, 2 :60–67, 1967.
- [RAD03] Matthew Richardson, Rakesh Agrawal, and Pedro Domingos. *Trust Management for the Semantic Web*, pages 351–368. Springer Berlin, 2003.

- [SZM13] Arnaud Sallaberry, Faraz Zaidi, and Guy Melançon. Model for generating artificial social networks having community structures with small-world and scale-free properties. *Social Network Analysis and Mining*, 3(3) :597–609, 2013.
- [WDDS06] Lei Wang, F. Du, H. P. Dai, and Y. X. Sun. Random pseudofractal scale-free networks with small-world effect. *The European Physical Journal B - Condensed Matter and Complex Systems*, 53(3) :361–366, 2006.
- [WPLP14] Robert West, Hristo S. Paskov, Jure Leskovec, and Christopher Potts. Exploiting social network structure for person-to-person sentiment analysis. *CoRR*, abs/1409.2450, 2014.
- [WS98] Duncan J. Watts and Steven H. Strogatz. Collective dynamics of ‘small-world’ networks. *Nature*, 393 :440–442, 1998.

Abstract

Whenever one wishes to try out or fine tune a social network analysis solution, being able to apply the said solution to an appropriate data set is of the utmost importance. However, and despite their established popularity, such network structures are rarely open or available in their entirety, thus developing the need for random generative models able to create graphs with properties close to those of real-world networks. We introduce a new generative model allowing the creation of such graphs. Our solution can be fine-tuned and used to build a wide array of graphs, ranging from small-world networks –with some of the free-scale networks characteristics– to completely random graphs.