
Un modèle pour les graphes bipartis aléatoires avec redondance

Émilie Coupechoux* — Fabien Tarissan*

* *Laboratoire d'Informatique de Paris 6 (LIP6), Université Pierre et Marie Curie (UPMC) & Centre National de la Recherche Scientifique (CNRS)*

RÉSUMÉ. Actuellement, les modèles de graphes aléatoires ne capturent pas toutes les propriétés connues des réseaux réels. En particulier, ils ne permettent pas à deux cliques d'avoir plusieurs nœuds en commun bien que ce soit une propriété observée dans les réseaux réels. Dans les réseaux sociaux par exemple, deux amis ont souvent plus d'un centre d'intérêt commun. Le modèle présenté ici vise à capturer ce genre de propriété. Plus précisément, nous présentons un modèle de graphe triparti aléatoire dont le projeté biparti possède des distributions de degrés et de redondances proches de ceux d'un graphe biparti donné.

ABSTRACT. Current models of random graphs do not capture all the properties observed in real-world networks. In particular, two cliques in such models generally do not have more than one node in common, while it is intuitive that, in social networks for instance, two friends have more than one interest in common. The model presented here aims at capturing this kind of property. More precisely, we present a model for random tripartite graphs such that the bipartite projection has degree and redundancy distributions close to those of a given bipartite graph.

MOTS-CLÉS : Graphes aléatoires, Modèles générateurs, Graphes bipartis, Redondance.

KEYWORDS: Random graphs, Generative models, Bipartite graphs, Redundancy.

1. Introduction

Les modèles de graphes aléatoires ont pour but de générer des structures ayant des propriétés similaires à celles observées dans les réseaux réels (pour une introduction, voir par exemple (Newman, 2003)). Depuis le premier modèle de graphe aléatoire (Erdős et Rényi, 1960), de nombreux autres modèles ont été introduits : les comparaisons entre modèles et réseaux réels ont permis de dégager des propriétés structurelles communes à ces réseaux, et d'en déduire de nouveaux modèles, prenant en compte de plus en plus de propriétés de ces réseaux.

Le modèle dit *de configuration* (Bollobás, 2001) (l'un des plus utilisés, et l'un des plus étudiés) permet de générer un graphe ayant la même distribution des degrés que celle du réseau considéré. Cependant, ce modèle de graphe ne possède pas (ou très peu) de triangles, contrairement aux réseaux réels (Watts et Strogatz, 1998).

Par ailleurs, plusieurs études ont mis en évidence que la plupart des réseaux réels ont une structure bipartite (voir (Newman *et al.*, 2001) et (Guillaume et Latapy, 2006)). Un graphe biparti est un graphe avec deux types de nœuds (par exemple, des individus et leurs centres d'intérêts), et des liens uniquement entre des nœuds de type différent (un individu a les centres d'intérêt auxquels il est relié). Le modèle de (Newman *et al.*, 2001) permet de générer un graphe biparti ayant une distribution identique à celle du réseau observé, pour les degrés de chacun des deux types de nœuds (c'est-à-dire pour le nombre de centres d'intérêts d'un individu, et le nombre d'individus possédant un centre d'intérêt donné).

Le graphe projeté d'un graphe biparti est obtenu en reliant deux individus ayant au moins un centre d'intérêt commun. Le graphe projeté du modèle de (Newman *et al.*, 2001) possède naturellement des triangles, cependant, sa distribution des degrés est très éloignée de la distribution des degrés dans le réseau observé (Tarissan *et al.*, 2013). Cela s'explique par le fait que, dans (Newman *et al.*, 2001), deux individus n'ont en général pas plus d'un centre d'intérêt commun, contrairement à ce qui est observé dans les réseaux réels (Latapy *et al.*, 2008). Nous proposons ici un modèle de graphe aléatoire qui possède cette dernière propriété des réseaux réels, en se basant sur une représentation tripartite du recouvrement observé au niveau biparti.

Le reste du papier est organisé de la façon suivante. La Section 2 rappelle les définitions qui vont être utilisées. La Section 3 présente le modèle, et la Section 4 constitue une première évaluation du modèle.

2. Définitions

2.1. Graphes bipartis

Un graphe biparti est un triplet $\mathcal{B} = (V_2, V_3, E_{23})$, où V_2 et V_3 sont des ensembles disjoints de nœuds, et $E_{23} \subseteq V_2 \times V_3$ l'ensemble des liens entre V_2 et V_3 .

Dans (Latapy *et al.*, 2008), le coefficient de redondance $r(v)$ d'un nœud $v \in V_2$ est défini comme étant, parmi les paires $(w, w') \in V_3$ de sommets voisins de v , la proportion de celles qui ont un voisin commun différent de v , soit :

$$r(v) = \frac{|\{(w, w') \in N(v)^2 \mid \exists u \neq v, (u, w), (u, w') \in E_{23}\}|}{|N(v)|(|N(v)| - 1)/2}, \quad [1]$$

où $N(v) = \{w \in V_3 \mid (v, w) \in E_{23}\}$ est l'ensemble des voisins de v et $|U|$ le cardinal de l'ensemble U . Par exemple, dans le graphe biparti \mathcal{B}_0 de la Figure 1, les coefficients de redondance des nœuds de V_2 sont : $r(A) = 1/1$, $r(B) = 4/10$ et $r(C) = 3/6$.

Dans (Newman *et al.*, 2001), le modèle de graphe biparti aléatoire est le suivant. Pour des degrés donnés pour les nœuds de V_2 et de V_3 , le graphe est choisi uniformément au hasard parmi tous les graphes bipartis ayant ces degrés. Dans un tel modèle, les coefficients de redondance des nœuds sont proches de zéro, contrairement à ce qui est observé dans les réseaux réels (Latapy *et al.*, 2008). L'utilisation d'un graphe triparti (Tarissan *et al.*, 2013) permet d'encoder le phénomène de redondance observé dans les réseaux réels.

2.2. Graphes tripartis et projection

Un graphe triparti est un uplet $\mathcal{T} = (V_1, V_2, V_3, E_{12}, E_{13}, E_{23})$, où V_1, V_2 et V_3 sont des ensembles disjoints de nœuds, et $E_{ij} \subseteq V_i \times V_j$ l'ensemble des liens entre V_i et V_j , pour tous $1 \leq i < j \leq 3$.

La projection du graphe triparti \mathcal{T} sur (V_2, V_3) est le graphe biparti $\mathcal{B} = (V_2, V_3, E'_{23})$, où $E'_{23} = E_{23} \cup \{(v, w) \mid \exists u \in V_1, (u, v) \in E_{12} \text{ et } (u, w) \in E_{13}\}$, autrement dit une arête est présente entre deux nœuds de \mathcal{B} s'ils sont reliés dans \mathcal{T} ou s'ils ont au moins un voisin commun dans \mathcal{T} . Un exemple de graphe triparti et son projeté sont donnés en Figure 1.

Le modèle défini dans la suite est un modèle de graphe triparti aléatoire permettant de construire, via sa projection, un graphe biparti dont les degrés et les coefficients de redondance sont proches de ceux d'un graphe donné. La génération d'un graphe triparti à partir d'un graphe biparti donné n'est pas discutée ici : on suppose que l'on dispose d'un graphe triparti donné, et on compare sa projection à la projection d'un graphe triparti aléatoire construit comme suit.

3. Modèle

Dans la Section 3.1, on introduit certaines quantités associées aux nœuds d'un graphe triparti. La proposition de la Section 3.2 présente (sous certaines hypothèses) le lien entre les quantités définies dans la section précédente et les quantités (degrés et coefficients de redondance) qui nous intéressent dans le graphe biparti projeté. Ceci permet d'introduire le modèle, finalement défini en Section 3.3.

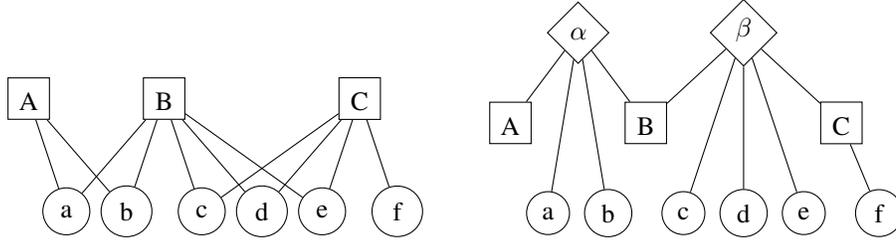


Figure 1. Un graphe biparti \mathcal{B}_0 (à gauche) et un graphe triparti \mathcal{T}_0 (à droite) dont le projeté est \mathcal{B}_0 . Les ensembles de nœuds sont : $V_1 = \{\alpha, \beta\}$, $V_2 = \{A, B, C\}$ et $V_3 = \{a, b, c, d, e, f\}$.

3.1. Notations

Soit $\mathcal{T} = (V_1, V_2, V_3, E_{12}, E_{13}, E_{23})$ un graphe triparti et $\mathcal{B} = (V_2, V_3, E'_{23})$ son projeté. Pour $i \in \{2, 3\}$, on pose $i^c = 3$ (resp. $i^c = 2$) si $i = 2$ (resp. $i = 3$).

Pour $i, j \in \{1, 2, 3\}, i \neq j$, et $v \in V_i$, on note $N_j(v) := N(v) \cap V_j = \{v' \in V_j \mid (v, v') \in E_{ij}\}$ l'ensemble des voisins de v appartenant à V_j , et $d_{ij}(v) := |N_j(v)|$ son cardinal, appelé dans la suite *degré de v vers V_j* .

Soit $i \in \{2, 3\}$. Pour $v \in V_i$ et $k \geq 0$, on note $s_{i1}(v, k) := |\{u \in V_1 \mid (u, v) \in E_{1i}, d_{1i^c}(u) = k\}|$ le nombre de voisins de v appartenant à V_1 et dont le degré vers V_{i^c} est k . Calculons $s_{21}(B, k)$ pour le nœud $B \in V_2$ de \mathcal{T}_0 , pour tout $k \geq 0$. Le nœud B est relié à un nœud (α) dont le degré vers V_3 vaut $k = 2$, donc $s_{21}(B, k = 2) = 1$, et le nœud B est également relié à un nœud (β) dont le degré vers V_3 vaut $k = 3$, donc $s_{21}(B, k = 3) = 1$. Pour $k \notin \{2, 3\}$, on a $s_{21}(B, k) = 0$ car le nœud B n'est relié à aucun nœud de degré k vers V_3 .

3.2. Liens entre le graphe triparti et son projeté

La proposition suivante montre que les degrés et coefficients de redondance dans le graphe biparti projeté peuvent être déduits des quantités définies précédemment sur le graphe triparti, sous certaines conditions (à savoir que le graphe triparti considéré ne possède pas de cycles de faible longueur, car ceux-ci peuvent être à l'origine de multi-arêtes lors de la projection et avoir un impact sur le calcul des coefficients de redondance du graphe projeté : les détails sont donnés dans la preuve).

Proposition 3.1. *Soit \mathcal{T} un graphe triparti ne contenant pas de cycle de longueur inférieure ou égale à 8. On suppose également qu'on a, pour tout $u \in V_1$: $d_{12}(u) \geq 2$. Soit \mathcal{B} le graphe biparti projeté de \mathcal{T} , $i \in \{2, 3\}$ et v un nœud de V_i . Alors on a :*

(a) le degré de v dans \mathcal{B} est donné par : $d^{\mathcal{B}}(v) = d_{i^c}(v) + \sum_{k \geq 0} k s_{i1}(v, k)$;

(b) lorsque $i = 2$, la redondance de $v \in V_2$ dans \mathcal{B} est donnée par : $r(v) = \left(\sum_{k \geq 0} \frac{k(k-1)}{2} s_{21}(v, k) \right) \times \frac{2}{d^{\mathcal{B}}(v)(d^{\mathcal{B}}(v)-1)}$.

Démonstration. Considérons les conditions suivantes :

(i) pour tous $(u, u') \in N_1(v)^2$, on a : $N_{i^c}(u) \cap N_{i^c}(u') = \emptyset$ (quand u parcourt l'ensemble des voisins de v dans V_1 , les voisins de u dans V_{i^c} forment des ensembles disjoints) ;

(ii) $N_{i^c}(v) \cap \left(\bigcup_{u \in N_1(v)} N_{i^c}(u) \right) = \emptyset$ (l'ensemble des voisins de v dans V_{i^c} et l'ensemble des voisins dans V_{i^c} des voisins de v dans V_1 sont disjoints).

La condition (i) (resp. (ii)) est vérifiée : autrement il y aurait un cycle de longueur 4 (resp. 3) dans \mathcal{T} . Ces conditions impliquent qu'aucune multi-arête n'est formée quand \mathcal{T} est projeté sur \mathcal{B} , ce qui prouve (a).

Soit $v \in V_2$ et soient $w, w' \in V_3$, $w \neq w'$, deux voisins de v dans \mathcal{B} . Nous prouvons que les conditions suivantes sont équivalentes (voir la Figure 2 pour une illustration) :

(i') w et w' ont au moins un autre voisin $v' \neq v$ en commun dans \mathcal{B} ;

(ii') il existe $u \in N_1(v)$ dans \mathcal{T} tel que $w, w' \in N_3(u)$.

Supposons que (ii') est vérifiée. Comme $d_{12}(u) \geq 2$, il existe $v' \in N_2(u)$ tel que $v' \neq v$. Les nœuds w, w', v et v' sont tous voisins de $u \in V_1$ dans \mathcal{T} , donc (i') est vérifiée. Supposons maintenant que (i') est vérifiée. Alors w, w', v et v' forment un cycle de longueur 4 dans \mathcal{B} , donc il n'existe pas dans \mathcal{T} ; ce cycle est donc causé par la projection d'au moins un nœud de V_1 , relié dans \mathcal{T} à au moins un des nœuds w, w', v et v' . En fait, il y a seulement quelques configurations dans \mathcal{T} qui conduisent (une fois projetées) à un tel cycle dans \mathcal{B} : soit celle de (ii'), soit une configuration contenant au moins un cycle. Cependant le cycle le plus long possible dans ce cas est celui de la Figure 2, à droite, et qui a longueur 8. Donc la seule configuration possible dans \mathcal{T} est celle de (ii'). Ceci prouve que (i') et (ii') sont équivalents.

Prouvons (b) pour $v \in V_2$. par définition de $r(v)$ dans [1], nous devons compter le nombre de paires $(w, w') \in N_3(v)$ qui vérifient (i'), ou de façon équivalente (ii'). Pour chaque nœud $u \in N_1(v)$ dans \mathcal{T} tel que $d_{13}(u) = k$, il y a $\frac{k(k-1)}{2}$ paires $(w, w') \in N_3^2(u)$, et $s_{21}(v, k)$ tels nœuds u . Ces paires sont disjointes par (ii), ce qui prouve (b).

□

REMARQUE. —

Dans le cas général où \mathcal{T} est un graphe triparti quelconque, l'équation en (a) devient une borne supérieure pour le degré d'un nœud v du graphe biparti projeté :

$$d^{\mathcal{B}}(v) \leq d_{i^c}(v) + \sum_{k \geq 0} k s_{i1}(v, k).$$

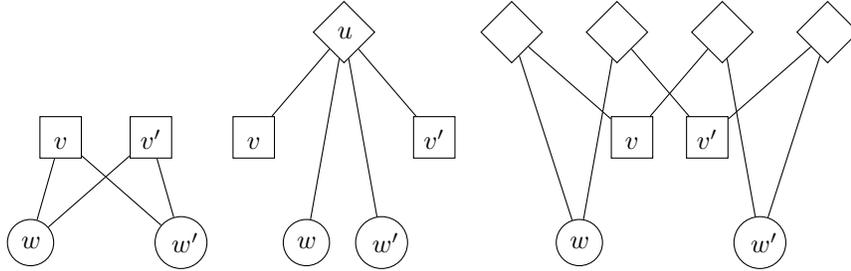


Figure 2. À gauche : une partie \mathcal{P} d'un graphe biparti, avec une paire de nœuds $(w, w') \in V_3^2$ voisins de v et ayant un autre voisin v' en commun. Au milieu : une configuration (la seule sans cycle) dont la projection est \mathcal{P} . À droite : une autre configuration (celle avec le cycle le plus long, de longueur 8) dont la projection est \mathcal{P} .

Le modèle suivant conserve les quantités $d_{ii^c}(v)$ et $s_{i1}(v, k)$ pour tous $i \in \{2, 3\}$, $v \in V_i$ et $k \geq 0$. Lorsque le graphe triparti donné et celui généré ne possèdent pas de cycle de longueur inférieure ou égale à 8, les degrés et redondances dans le graphe biparti projeté sont donc conservés (cf. Proposition 3.1). Cependant, ce ne sera pas le cas en pratique : la Proposition 3.1 permet alors seulement d'espérer que ces quantités sont proches, et une première évaluation de la différence sera donnée en Section 4.

3.3. Modèle de graphe triparti aléatoire

Dans un premier temps, nous définissons le modèle. Dans un second, nous prouvons que les données qui nous intéressent sont bien conservées (cf. Proposition 3.2).

On considère l'ensemble \mathcal{D} de données suivantes (calculées à partir d'un graphe triparti \mathcal{T} donné) : $(d_{12}(u), d_{13}(u))$ pour tout $u \in V_1$, $(d_{23}(v), (s_{21}(v, k))_{k \geq 0})$ pour tout $v \in V_2$, et $(d_{32}(v), (s_{31}(v, k))_{k \geq 0})$ pour tout $v \in V_3$. Dans la suite, on définit un graphe triparti aléatoire $\tilde{\mathcal{T}} = (V_1, V_2, V_3, \tilde{E}_{12}, \tilde{E}_{13}, \tilde{E}_{23})$ qui conserve les données \mathcal{D} . Une illustration des notations suivantes est donnée en Figure 3.

À tout nœud $v \in V_i$, $i \in \{2, 3\}$, on associe $d_{ii^c}(v)$ demi-arêtes de type ii^c et on choisit un appariement uniformément au hasard (parmi tous ceux ne produisant pas de multi-arêtes) entre les demi-arêtes de type 23 de V_2 et les demi-arêtes de type 32 de V_3 . Ceci construit \tilde{E}_{23} .

La construction de \tilde{E}_{1i} , $i \in \{2, 3\}$, nécessite plus de notations. À tout nœud $v \in V_i$, on associe, pour tout $k \geq 0$, $s_{i1}(v, k)$ demi-arêtes de type $(i1, k)$ (qui correspondent aux demi-arêtes initialement associées à des nœuds de V_1 dont le degré vers V_i^c est k). Par exemple, si $\mathcal{T} = \mathcal{T}_0$ (voir Figures 1 et 3), alors le nœud B a une

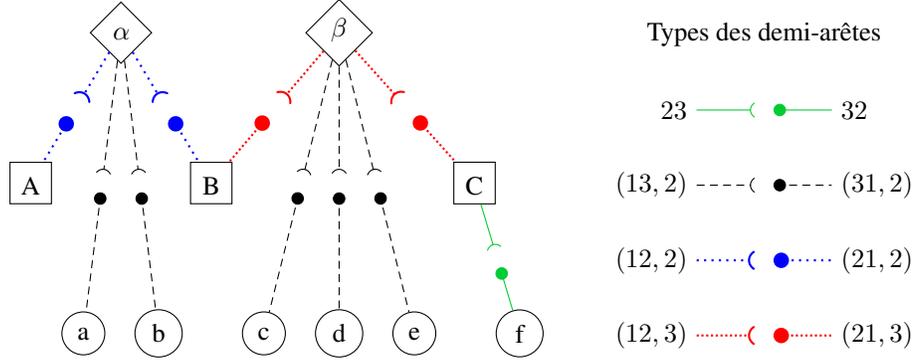


Figure 3. À gauche : illustration des données associées au graphe triparti \mathcal{T}_0 de la Figure 1. À droite : légende indiquant les types des demi-arêtes.

demi-arête de type $(21, k = 2)$ (car il a une demi-arête vers α qui est de degré $k = 2$ vers V_3) et une demi-arête de type $(21, k = 3)$ (car il a une demi-arête vers β qui est de degré $k = 3$ vers V_3). À tout nœud $u \in V_1$, on associe, posant $k_2 = d_{12}(u)$ et $k_3 = d_{13}(u)$, k_2 demi-arêtes de type $(12, k_3)$ et k_3 demi-arêtes de type $(13, k_2)$. Par exemple, si $\mathcal{T} = \mathcal{T}_0$ (voir Figures 1 et 3), alors β a $k_2 = 2$ demi-arêtes de type $(12, k_3 = 3)$ et $k_3 = 3$ demi-arêtes de type $(13, k_2 = 2)$. Pour tout $k \geq 0$, on choisit un appariement uniformément au hasard (parmi tous ceux ne produisant pas de multi-arêtes) entre les demi-arêtes de type $(i1, k)$ de V_1 et les demi-arêtes de type $(i1, k)$ de V_i . Ceci construit \tilde{E}_{1i} .

Proposition 3.2. *Le graphe triparti aléatoire $\tilde{\mathcal{T}}$ conserve les données \mathcal{D} .*

Démonstration. Soit $i \in \{2, 3\}$ et $v \in V_i$. Les $s_{i1}(v, k)$ demi-arêtes initialement associées à des nœuds de V_1 dont le degré vers V_{i^c} est k sont de type $(i1, k)$ par définition. Elles sont donc associées à des demi-arêtes de type $(1i, k)$ qui appartiennent à des nœuds de V_1 dont le degré vers V_{i^c} est k . \square

4. Évaluation et discussion

Dans cette section, on compare les degrés des nœuds de V_2 , les degrés des nœuds de V_3 , et la redondance des nœuds de V_2 entre :

- le projeté biparti \mathcal{B} d'un graphe triparti \mathcal{T} donné (appelé graphe donné),
- et le projeté biparti $\tilde{\mathcal{B}}$ d'un graphe triparti aléatoire $\tilde{\mathcal{T}}$ construit selon la méthode décrite précédemment (appelé graphe aléatoire).

Les graphes tripartis \mathcal{T} et $\tilde{\mathcal{T}}$ ayant les mêmes données \mathcal{D} (cf. Proposition 3.2), la Proposition 3.1 permet d'espérer que ces quantités vont être assez proches. Plus précisément, ces quantités seraient identiques si les graphes tripartis \mathcal{T} et $\tilde{\mathcal{T}}$ ne possédaient

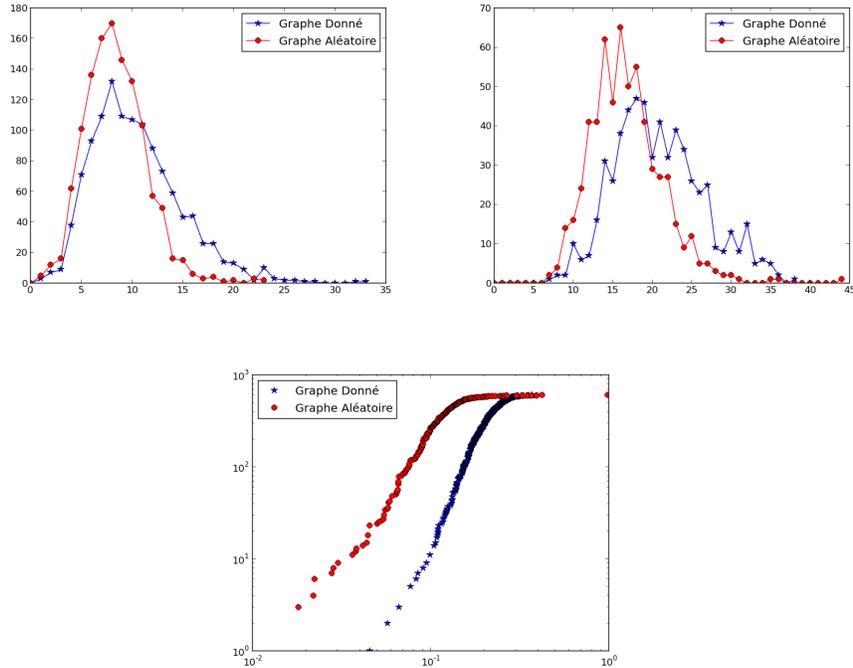


Figure 4. En haut à gauche (resp. en haut à droite) : le nombre de nœuds de V_3 (resp. V_2) en fonction de leur degré. En bas : le nombre de nœuds de V_2 de redondance inférieure ou égale à x en fonction de x .

pas de cycles de longueur inférieure ou égale à 8, et nous évaluons ici l'impact de tels cycles.

Le graphe triparti \mathcal{T} est ici construit de façon aléatoire : l'utilisation de graphes tripartis provenant de données réelles constitue l'étape suivante de ce travail (la construction d'un graphe triparti à partir d'un graphe biparti donné n'est d'ailleurs pas unique, et il sera intéressant de comparer également l'impact du choix de graphe triparti). Toutes les distributions des degrés (de V_1 vers V_2 , de V_1 vers V_3 , etc.) sont ici homogènes. Lorsque la distribution des degrés de V_1 vers V_2 (ou celle de V_1 vers V_3) est changée en une loi de puissance, les résultats obtenus sont identiques. La taille du graphe triparti \mathcal{T} est la suivante : $|V_1| = 200$, $|V_2| = 600$ et $|V_3| = 1200$. Les évaluations sont ici réalisées sur une instance de graphe triparti aléatoire, et les résultats obtenus pour de nombreuses instances sont très similaires.

La Figure 4 (haut) montre que les distributions de degrés dans les deux graphes sont sensiblement les mêmes, avec des degrés légèrement plus faibles dans le modèle

aléatoire. La Figure 4 (bas) montre que la redondance dans le modèle aléatoire a une distribution similaire quoique un peu plus faible. Ces différences s'expliquent de la façon suivante. Le nombre de liens dans le graphe aléatoire $\tilde{\mathcal{B}}$ est plus faible que dans le graphe donné \mathcal{B} (10053 contre 12496 dans \mathcal{B}), ce qui suppose que le graphe triparti aléatoire $\tilde{\mathcal{T}}$ possède davantage de cycles de faible longueur que le graphe triparti donné \mathcal{T} (ces cycles étant à l'origine des multi-arêtes, voir la Section 3.2). Or, plus le graphe triparti possède de cycles de faible longueur, plus les degrés dans le graphe projeté sont susceptibles de baisser (voir la remarque après la Proposition 3.1). Les degrés sont donc logiquement plus faibles dans le graphe aléatoire $\tilde{\mathcal{B}}$ que dans le graphe donné \mathcal{B} , ce qui est en accord avec la Figure 4.

Dans l'ensemble, ce modèle capture les principales propriétés (degrés et redondance) du graphe donné.

Ce travail présente une première étape vers la construction d'un modèle de graphe biparti aléatoire avec des degrés et une redondance proches de ceux d'un graphe biparti donné, l'étape suivante étant la comparaison entre ce modèle et des données réelles.

Remerciements

Ce travail a été financé en partie par la commission européenne à travers le projet EULER (FP7 FIRE, n° 258307).

5. Bibliographie

- Bollobás B., *Random graphs*, vol. 73 of *Cambridge Studies in Advanced Mathematics*, second edn, Cambridge University Press, Cambridge, 2001.
- Erdős P., Rényi A., « On the evolution of random graphs », *Publications of the Mathematical Institute of the Hungarian Academy of Sciences*, vol. 5, p. 17-61, 1960.
- Guillaume J.-L., Latapy M., « Bipartite graphs as models of complex networks », *Physica A*, vol. 371, p. 795-813, 2006.
- Latapy M., Magnien C., Vecchio N. D., « Basic notions for the analysis of large two-mode networks », *Social Networks*, vol. 30, n° 1, p. 31 - 48, 2008.
- Newman M. E. J., *Random graphs as models of networks*, Handbook of Graphs and Networks, wiley-vch, berlin edn, S. Bornholdt, H. G. Schuster, 2003.
- Newman M. E. J., Strogatz S. H., Watts D. J., « Random graphs with arbitrary degree distributions and their applications », *Phys. Rev. E*, vol. 64, p. 026118, Jul, 2001.
- Tarissan F., Quoitin B., Mérimol P., Donnet B., Pansiot J.-J., Latapy M., « Towards a bipartite graph modeling of the internet topology », *Computer Networks*, vol. 57, n° 11, p. 2331 - 2347, 2013.
- Watts D. J., Strogatz S. H., « Collective dynamics of 'small-world' networks. », *Nature*, vol. 393, n° 6684, p. 440-442, June, 1998.