

Génération aléatoire de graphes bipartis

Fabien Tarissan

stages@complexnetworks.fr

<http://www.complexnetworks.fr>

LIP6 – CNRS et UPMC – Paris

Contexte. Dans de très nombreux contextes applicatifs, on rencontre de grands graphes n'ayant aucune structure simple apparente, que nous appellerons graphes de terrain (par opposition aux graphes explicitement construits par un modèle ou une théorie). Citons par exemple la topologie de l'internet (routeurs et câbles entre eux), les graphes du web (pages web et liens hypertextes entre elles), les échanges divers (pair-à-pair, e-mail, etc), mais aussi les réseaux sociaux, biologiques ou linguistiques.

Il est apparu récemment que la plupart de ces grands graphes ont des propriétés statistiques en commun, et que ces propriétés les différencient fortement des graphes aléatoires¹ utilisés jusqu'alors pour les modéliser. Notamment, ils ont une densité très faible, une distance moyenne faible, une distribution de degrés hétérogène, et une densité locale forte. Depuis lors, de nombreux travaux ont été menés visant à capturer ces propriétés dans des modèles, nécessaires tant pour effectuer des simulations que pour étudier formellement ces objets, et bien sûr pour en comprendre la nature.

Les graphes aléatoires classiques capturent la densité faible (qui est en fait un paramètre du modèle) et la distance moyenne faible. Par contre, ils ont une distribution des degrés homogène et une densité locale faible. Nous sommes également en mesure de générer un graphe aléatoire à distribution de degrés donnée. On capture ainsi toutes les propriétés citées ci-dessus sauf la densité locale forte. Malgré de nombreuses tentatives, générer des graphes ayant également une densité locale forte tout en gardant leur caractère aléatoire reste un problème ouvert. Les attentes sont pourtant extrêmement fortes.

Graphes bipartis. 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 ou les réseaux de co-publications qui relient les auteurs publiant ensemble, 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, et \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 et la génération aléatoire des graphes de terrains.

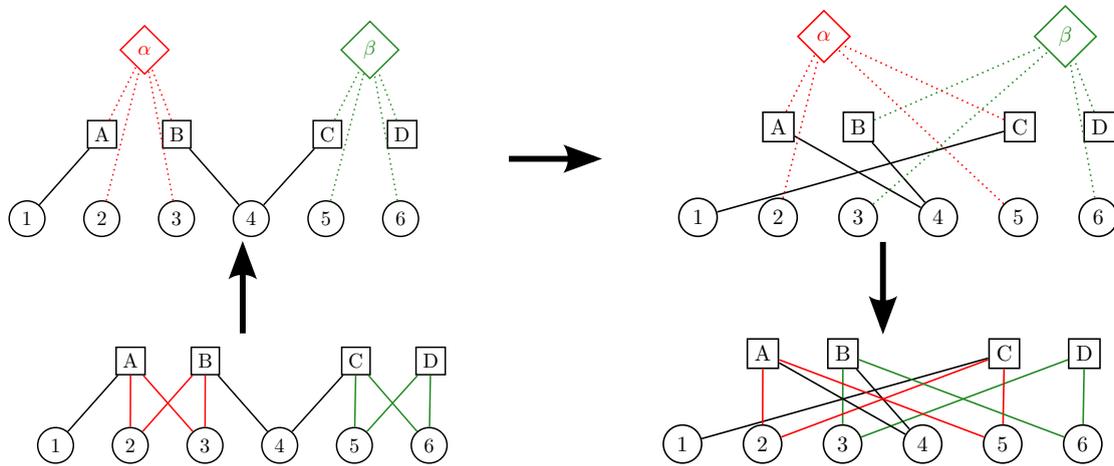
Dans un papier récent [1], cette approche a été exploitée afin de proposer, pour la première fois, un modèle biparti de la topologie de l'Internet dans lequel les nœuds routeurs et les nœuds switches sont distingués. Ce modèle reste simple en ce qu'il ne prend en paramètre que la séquence des degrés 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

1. Nous entendons par *aléatoire* un tirage aléatoire uniforme dans un ensemble donné, ici l'ensemble des graphes ayant une taille donnée.

naturellement. Mais ce travail a aussi montré que le modèle échouait dans la reproduction de certaines propriétés telle que les recouvrements, assimilables à des bicliques maximales². La figure ci-dessous montre un exemple typique de recouvrements (en rouge et en vert) qui sont perdus lors de la phase aléatoire :



Objectifs du stage. Pour tenter d'améliorer la génération des graphes bipartis, et en particulier la conservation des structures de recouvrement présentes dans les réseaux réels, une proposition a été faite [2] qui consiste à détecter dans un premier temps les structures recouvrantes dans le graphe biparti puis les représenter à l'aide d'un troisième niveau, définissant de fait un *graphe triparti* encodant la structure bipartie. Dès lors il devient possible de générer aléatoirement un graphe triparti puis de projeter le graphe généré afin de récupérer un graphe biparti, comme le montre la figure suivante :



Le but de ce stage est d'étendre cette proposition et d'étudier sa pertinence en tant que support pour générer des graphes réalistes. Plusieurs pistes peuvent être envisagées, parmi lesquelles on peut citer :

Algorithme : la phase de détection des structures recouvrantes passe par une énumération des bicliques maximales dans le graphe biparti. Cette phase est coûteuse si l'on fait une énumération exhaustive mais il est possible d'améliorer son efficacité en pratique si on s'appuie sur les propriétés attendues du graphe.

Codage triparti : une fois les bicliques maximales détectées, plusieurs stratégies peuvent être adoptées pour les représenter dans une structure tripartite. Il est donc nécessaire de comprendre quel impact ces stratégies vont avoir sur la phase de génération aléatoire

Génération aléatoire : une fois les bicliques maximales représentées par un graphe triparti, plusieurs générations aléatoires peuvent être envisagées. Il s'agit donc là de comprendre quelles sont les plus efficaces pour générer des graphes tripartis qui, une fois projetés, seront proches du graphe biparti original.

2. ensemble de sommets \perp et \top tous reliés deux à deux.

Références

- [1] Fabien Tarissan, Bruno Quoitin, Pascal Mérindol, Benoit Donnet, Jean-Jacques Pansiot, and Matthieu Latapy. Towards a bipartite graph modeling of the internet topology. *Computer Networks*, 57(11) :2331–2347, 2013.
- [2] Fabien Tarissan and Emilie Coupechoux. 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)*, pages 1–10, 2013.