

Génération de graphes aléatoires par échanges multiples d'arêtes

Title: Random graph generation using multiple switches of edges

Lionel Tabourier^{1,2}, Jean-Philippe Cointet³ et Camille Roth⁴

Résumé : La génération de graphes aléatoires vérifiant un ensemble de propriétés fixé est un problème majeur pour l'étude des réseaux d'interaction. Pourtant, il n'existe pas de solution générale qui soit satisfaisante dans les cas pratiques, où l'ensemble de propriétés à satisfaire est complexe. Nous proposons une méthode de génération permettant théoriquement d'obtenir un échantillon parfaitement aléatoire de n'importe quel ensemble de graphes, à condition que la distribution des degrés soit fixée et que l'on dispose d'un élément de cet ensemble. Cette méthode dite de k -échanges, généralise les procédures de Monte-Carlo par chaîne de Markov de la littérature, selon lesquelles on échange itérativement les extrémités d'arêtes du graphe. Nous décrivons sa réalisation, les difficultés techniques à résoudre et comment il est possible de les surmonter. Nous appliquons cette méthode sur des réseaux de collaborations scientifiques, et montrons que l'on peut identifier un petit nombre de propriétés suffisantes pour expliquer des caractéristiques typiques du réseau.

Abstract: Generating random graphs which verify a set of predefined properties is a major issue for the analysis of interaction networks. However, there is no general method available for practical cases, where the set of desired properties is complex. We propose a generation method which theoretically allows to obtain a uniform sample of any set of graphs, as long as we have an element of this set and the degree distribution is one of the required properties. This method, called k -edge switching, is a generalization of Monte-Carlo Markov Chain methods of the literature which rely on iterating exchanges of edges ends. We describe its conception and implementation, as well as the technical difficulties encountered and how to overcome them. This method is applied on scientific collaboration networks, and we show that we can point out a small set of properties which can explain typical characteristics of the network.

Mots-clés : graphes aléatoires, génération de graphes, algorithmes MCMC

Keywords: random graphs, graph generation, MCMC algorithms

Classification AMS 2000 : 05C80, 05C82, 65C05, 68W20

1. Introduction

La génération de graphes aléatoires respectant des contraintes quelconques est une question centrale pour notre compréhension des réseaux d'interaction réels. Elle permet en effet de confirmer le rôle d'un ensemble de contraintes pour expliquer l'observation de caractéristiques du graphe. Les contraintes dont il est ici question sont le plus souvent de nature structurelle, telles que la distribution des degrés, qui est souvent associée à l'activité des nœuds du réseau (Tangmunarunkit *et al.*, 2002; Newman, 2004), la connexité du graphe qui traduit l'accessibilité d'un nœud depuis un autre (Viger et Latapy, 2005) ou encore les corrélations entre les degrés des sommets (Mahadevan *et al.*, 2006). Il peut également s'agir de contraintes exogènes, telles qu'imposer l'affiliation de sommets à des groupes spécifiques, puis n'autoriser que certains types d'interactions entre les nœuds, nous pouvons ici citer l'exemple des réseaux de parenté où il n'y a en général pas de relation maritale entre membres d'une même fratrie.

¹ CNRS, UMR 7606, LIP6, 4 Place Jussieu, 75252 Paris

² Sorbonne Universités, UPMC Université Paris 06, UMR 7606, LIP6

E-mail : lionel.tabourier@lip6.fr

³ INRA SenS (IFRIS/UPEM), 5 boulevard Descartes, 77454 Champs-sur-Marne

⁴ CNRS, Centre Marc Bloch Berlin, Friedrichstrasse 191, D-10117 Berlin

Les graphes ainsi produits servent de modèles pour expliquer les caractéristiques observées des réseaux d'interactions, par exemple leur robustesse à la disparition de sommets ou d'arêtes (Callaway *et al.*, 2000).

Cependant, il n'existe pas à l'heure actuelle de méthode pour générer un échantillon de graphes de manière parfaitement aléatoire, excepté pour certains cas très précis, comme par exemple les graphes à distribution des degrés fixée. On peut distinguer deux familles de méthodes pour générer des échantillons aléatoires de graphes \mathcal{G}_C vérifiant un ensemble de contraintes C :

- Soit en construisant les graphes de \mathcal{G}_C en ajoutant des nœuds et des arêtes tout en s'assurant que les contraintes soient respectées. C'est selon ce principe qu'est construit le *modèle configurationnel* décrit dans Bender et Canfield (1978) : on attribue aux nœuds un certain nombre de demi-arêtes, qu'on connecte entre elles, de manière à ce que chaque graphe respecte une distribution des degrés prédéfinie.
- Soit on utilise un graphe appartenant à l'ensemble \mathcal{G}_C , dont on échange itérativement les extrémités d'arêtes, de manière à rester dans l'ensemble \mathcal{G}_C , jusqu'à converger vers un élément quelconque de \mathcal{G}_C après un nombre "suffisant" d'itérations. Cette approche à base d'échanges d'arêtes a été proposée par exemple par Rao *et al.* (1996) ou Roberts (2000).

La première méthode exige de trouver une tactique d'assemblage pour chaque nouvel ensemble de contraintes, cette tâche s'avère complexe, en particulier si l'on cherche à réaliser un échantillon uniformément aléatoire, c'est-à-dire dans lequel chaque graphe de l'ensemble a la même probabilité d'être dans l'échantillon. Nous souhaitons au contraire disposer d'une méthode polyvalente et facilement adaptable en fonction de l'ensemble de contraintes considéré. C'est pourquoi nous allons explorer la seconde famille de méthodes génératives dans la suite de cet article.

Nous allons d'une part (section 2) présenter les limites et inconvénients des méthodes d'échanges basées sur la permutation des extrémités de deux arêtes. Ces méthodes classiques de la littérature ont été discutées dans de nombreux travaux antérieurs, traitant d'aspects théoriques (Eggleton, 1973; Colbourn, 1977; Taylor, 1980, 1982; Rao *et al.*, 1996), de leur mise en œuvre pratique (Milo *et al.*, 2003; Gkantsidis *et al.*, 2003; Stauffer et Barbosa, 2005; Artzy-Randrup et Stone, 2005; Viger et Latapy, 2005; Feder *et al.*, 2006; Coolen *et al.*, 2009) ou d'applications à différents contextes (Kannan *et al.*, 1997; Roberts, 2000; Cooper *et al.*, 2006; Mahadevan *et al.*, 2006; Bansal *et al.*, 2008). Nous introduisons ensuite une méthode à base d'échanges d'ordre supérieur, pour générer aléatoirement des échantillons de graphes vérifiant des contraintes arbitraires (section 3). Enfin, nous présenterons une application pratique à l'utilisation de notre méthode générative en section 4, où nous détaillons le cas de réseaux de collaboration scientifique, où nous cherchons à isoler des caractéristiques suffisantes pour justifier les propriétés structurelles typiques de ces réseaux.

2. Le processus markovien d'échanges d'arêtes

Les processus d'échanges sont souvent utilisés pour générer des échantillons de graphes (Rao *et al.*, 1996; Kannan *et al.*, 1997; Roberts, 2000; Gkantsidis *et al.*, 2003; Artzy-Randrup et Stone, 2005; Stauffer et Barbosa, 2005; Cooper *et al.*, 2006). Par ailleurs ils ont donné lieu à de nombreux approfondissements, par exemple pour obtenir des graphes ayant des propriétés structurelles spécifiques (Milo *et al.*, 2003; Mahadevan *et al.*, 2006; Bansal *et al.*, 2008) ou pour produire ces échantillons plus efficacement (Viger et Latapy, 2005; Feder *et al.*, 2006). Dans la suite de ce travail, nous allons en présenter un plus en détails : *switching & holding*, dont les propriétés vont permettre de définir nos objectifs.

2.1. *Switching & holding*

Miklós et Podani (2004) ont montré qu'il était possible d'utiliser un algorithme d'échanges à base de paires d'arêtes pour générer un échantillon uniformément aléatoire de graphes orientés à distribution des degrés fixée, Artzy-Randrup *et al.* ont dénommé par la suite cette méthode *switching & holding* (Artzy-Randrup et Stone, 2005). Précisément, cette procédure consiste à choisir aléatoirement deux arêtes du graphe, puis à examiner si l'échange des extrémités de ces arêtes conduit à un graphe respectant les contraintes ; si oui, on effectue l'échange, sinon, on conserve l'état dans lequel on se trouve. Notons que *switching & holding* se distingue des procédures simples d'échanges car on y fixe le nombre de tentatives d'échange, plutôt que le nombre d'échanges effectués. Cet algorithme appartient à la famille des méthodes *Markov chain Monte-Carlo* (MCMC), plus précisément il s'agit d'un *échantillonnage de Gibbs*.

2.2. *Marche aléatoire et graphe de Markov*

Nous proposons dans cette partie une description intuitive de la procédure précédente, qui sera utile dans le but d'en construire une forme généralisée par la suite. Celle-ci peut être comprise comme une marche aléatoire dans un *graphe de Markov*. Nous définissons ici ce graphe de la manière suivante : il s'agit d'un multigraphe dirigé, dont chaque sommet représente un élément de l'ensemble à décrire (noté \mathcal{G}_C). Un arc du graphe de Markov représente une tentative d'échange entre deux arêtes du graphe considéré, tel que :

- si l'échange d'arête génère un graphe de \mathcal{G}_C , l'arc relie l'élément considéré du graphe de Markov à celui généré,
- si l'échange d'arête génère un graphe extérieur à \mathcal{G}_C , l'arc est une boucle de l'élément vers lui-même.

Dans ce sens, l'itération de tentatives d'échanges est une chaîne de Markov (Sinclair, 1993) convergeant vers une distribution d'équilibre dont les caractéristiques peuvent être extraites de la matrice de transition du processus markovien. De plus, si le graphe de Markov a un degré constant, c'est-à-dire le même degré entrant et sortant pour chaque élément de l'ensemble, le processus est uniforme. Par ailleurs, si le graphe de Markov est connexe, tous les éléments de l'ensemble sont accessibles. Alors si le graphe de Markov est à la fois connexe et de degré constant, le processus mène à un élément parfaitement aléatoire de \mathcal{G}_C . On donne une illustration de cette situation sur la figure 1.

2.3. *Limites fondamentales de switching & holding*

Pour utiliser des méthodes d'échanges dans le but de générer un échantillon uniforme, il faut s'assurer que les deux conditions précédemment évoquées sont effectivement remplies :

- tous les éléments accessibles ont la même probabilité d'être atteints (*uniformité*),
- tous les éléments de \mathcal{G}_C sont accessibles (*ergodicité*).

L'uniformité est garantie dans le cas de l'approche *switching & holding*. Une démonstration est proposée dans le cas d'une contrainte de distribution des degrés fixée dans Miklós et Podani (2004), mais elle peut être étendue à n'importe quelle contrainte. On peut comprendre simplement pourquoi l'uniformité est vérifiée : le degré d'un nœud du graphe de Markov est égal au nombre de combinaisons possibles d'arêtes dans le graphe sur lequel on effectue la tentative d'échange. Pour tout ensemble de graphes dans lequel le nombre d'arêtes est fixé, ce nombre de combinaisons est strictement identique pour tous les éléments accessibles. Donc dans l'état stationnaire de la chaîne de Markov, le temps de visite est strictement identique pour tous les éléments de l'ensemble de graphes.

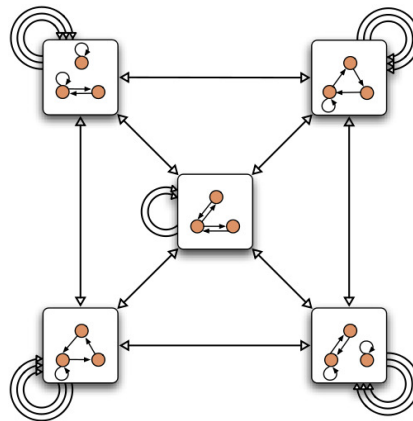


FIGURE 1. Un graphe de Markov simple sur l'ensemble des graphes de trois nœuds à distribution des degrés couplée entrant-sortant fixée : $\{(1,1), (1,1), (2,2)\}$, sans arcs multiples, avec boucles.

L'ergodicité consiste à avoir un graphe de Markov connexe, ou dans la terminologie usuelle, à ce que la chaîne de Markov soit irréductible. À notre connaissance, les théorèmes garantissant l'ergodicité concernent seulement des contraintes simples, essentiellement la conservation de la séquence des degrés, à laquelle on ajoute des contraintes supplémentaires : pour des arbres (Colbourn, 1977), des graphes connexes (Taylor, 1980) et bi-connexes (Taylor, 1982). Dans le cas de contraintes plus élaborées (e.g. Mahadevan *et al.* (2006); Bansal *et al.* (2008)), l'utilisation de *switching & holding* ne serait pas légitime, puisqu'il n'existe pas de garantie sur l'ergodicité. Par exemple Rao *et al.* (1996) montrent que dans le cas de graphes dirigés et sans boucles à distribution des degrés fixée, il est impossible, dans certains cas, d'atteindre tous les éléments de l'ensemble par *switching & holding*. Ainsi, il n'existe pas de combinaisons d'échanges à deux arêtes permettant de passer de l'une à l'autre des matrices d'adjacence suivantes ¹ :

$$\begin{pmatrix} 0^* & 1 & 0 \\ 0 & 0^* & 1 \\ 1 & 0 & 0^* \end{pmatrix} \leftrightarrow \begin{pmatrix} 0^* & 0 & 1 \\ 1 & 0^* & 0 \\ 0 & 1 & 0^* \end{pmatrix}$$

2.4. Convergence de la marche aléatoire

En plus de ces limitations, la question de la vitesse de convergence des algorithmes markoviens est un problème théorique difficile (Rao *et al.*, 1996; Guruswami, 2000), que l'on résout souvent à l'aide d'heuristiques (Gkantsidis *et al.*, 2003; Viger et Latapy, 2005). En effet, la convergence est souvent estimée empiriquement à l'aide de mesures que nous qualifierons d'*observables-cibles*. Il s'agit de mesures scalaires sur le graphe, en général structurelles mais pas nécessairement, que l'on va réaliser régulièrement dans le but de caractériser la dynamique du processus markovien. Les mesures de nombre de motifs (sous-graphes connexes de petite taille) ou de corrélations de degré sont rapides à réaliser et donc un choix privilégié d'observables-cibles. Précisément, le principe consiste à effectuer la mesure de ces observables au cours du processus, et on considère avoir effectué un nombre "suffisant" d'itérations lorsque toutes ces observables ont atteint un plateau, correspondant à leur mesure dans l'état stationnaire de la chaîne de Markov.

¹ Les 0^* de la diagonale ne peuvent être modifiés car les graphes à générer ne contiennent pas de boucle.

3. Procédure d'échange généralisée

Dans cette section, nous proposons une procédure d'échange généralisée, également du type MCMC, qui nous permettra de réaliser des échantillons d'ensembles uniformément aléatoires dans une plus grande variété de cas que *switching & holding*. Pour une question de clarté de l'exposé, nous ne présentons que le cas des graphes dirigés, il est néanmoins aisé de généraliser la méthode aux graphes non-dirigés.

3.1. Procédure de k -échange

La non-connexité du graphe de Markov provient de l'impossibilité de modifier un graphe en un autre au moyen d'un échange simple d'arêtes. Pour surmonter ce problème, nous proposons une méthode expérimentale qui se base sur des échanges d'ordre supérieur : étant donné $G \in \mathcal{G}_C$, on considère k arcs $(a_i, b_i)_{i \in \{1, \dots, k\}}$, correspondant aux nœuds $(a_1, \dots, a_k, b_1, \dots, b_k)$, éventuellement non-distincts. Le processus de k -échange consiste à choisir aléatoirement une permutation σ parmi les $k!$ permutations possibles des extrêmités (b_1, \dots, b_k) des arcs. On obtient alors le graphe où les arcs $(a_i, b_i)_{i \in \{1, \dots, k\}}$ sont remplacés par $(a_i, \sigma(b_i))_{i \in \{1, \dots, k\}}$ (pour un exemple d'algorithme, voir algo. 1). Notons que cette procédure suppose un ensemble de contraintes minimales : \mathbf{C}^\emptyset ; en effet, le type de graphe est fixé (graphe simple, dirigé, multigraphe etc) et les k -échanges conservent nécessairement les degrés des nœuds, ce qui implique que la distribution des degrés reste inchangée.

Par ailleurs, les voisins de G dans le graphe de Markov au sens d'une procédure d'échange classique sont aussi voisins de G pour le graphe de Markov correspondant au k -échange, puisque la permutation qui échange seulement deux extrêmités est incluse dans l'ensemble des permutations à k arcs. Aussi, quand $k = 2$, on retrouve la procédure *switching & holding*. Plus généralement, lorsque la valeur de k augmente, la nouvelle procédure crée de nouveaux liens dans le graphe de Markov. Par exemple, pour le cas de la figure 1, on peut vérifier que le graphe de Markov devient complet pour $k = 3$. Il est possible de cette manière de connecter des composantes du graphe de Markov qui ne l'étaient pas pour le processus *switching & holding*.

Notons également que la procédure de k -échange est uniforme dans l'ensemble de graphes accessibles quelle que soit la valeur de k , pour une raison analogue à celle qui permet de démontrer l'uniformité de *switching & holding* : le nombre de combinaisons possibles de k arêtes est strictement identique pour tous les éléments accessibles de l'ensemble de graphes dans lequel est réalisé la chaîne de Markov.

3.2. Relation entre k et l'ergodicité

Théorème. *Il existe une borne supérieure à k telle que le graphe de Markov associé soit nécessairement connexe, et la marche aléatoire associée irréductible (ergodique). Plus précisément, étant donné deux graphes G_1 et G_2 de \mathcal{G}_C , il existe toujours une permutation de taille au plus M (nombre d'arcs du graphe), permettant de transformer G_1 en G_2 .*

Démonstration. En effet, on peut écrire les M arcs de G_1 ainsi : $\{(a_1, b_1); (a_2, b_2); \dots; (a_M, b_M)\}$. De plus, nous pouvons écrire les M arcs de G_2 , sous la forme : $\{(a_1, b'_1); (a_2, b'_2); \dots; (a_M, b'_M)\}$. Et comme la distribution des degrés est identique dans les deux cas, ces deux listes d'arcs contiennent exactement les mêmes nœuds, qui apparaissent le même nombre de fois. Ainsi, en définissant σ comme la permutation : $(b_1, b_2, \dots, b_M) \xrightarrow{\sigma} (b'_1, b'_2, \dots, b'_M)$, on a défini un M -échange permettant de transformer G_1 en G_2 . □

Algorithme 1: Algorithme de k -échange, dans le cas d'un graphe dirigé vérifiant les contraintes : distribution des degrés fixée, pas de boucle, pas d'arcs multiples et un ensemble de contraintes \mathbf{C}^+ , à fixer selon le contexte.

Entrées: graphe $G_0 = (V_0, E_0)$; nombre de tentatives : n ; taille des échanges : k ;
Sortie: graphe G produit par n tentatives d'échanges ;
 $G = (V, E) \leftarrow G_0$;
 $j = 1$; // initialisation
tant que $j \leq n$
 tirer aléatoirement k arcs : $\{(a_i, b_i)\}_{i \in I} \in E$;
 tirer aléatoirement σ , une permutation de l'ensemble des extrémités $\{b_i\}_{i \in I}$;
 construire l'ensemble des arcs échangés $\{(a_i, \sigma(b_i))\}_{i \in I}$;
 $E' \leftarrow E \cup \{(a_i, \sigma(b_i))\} \setminus \{(a_i, b_i)\}$;
 $G' = (V, E')$;
 $\forall i \in I, \mathcal{W}_i = \{b : \exists (a_i, b) \in E\} \setminus \{b_i\}$;
 si $\forall i, a_i \neq \sigma(b_i)$ // tester pas de boucles
 et $\forall i, \sigma(b_i) \notin \mathcal{W}_i$ // tester pas d'arêtes multiples
 et $G' \in \mathcal{G}_{\mathbf{C}^+}$ // tester contraintes \mathbf{C}^+
 alors $G \leftarrow G'$;
 $j \leftarrow j + 1$;
fin

Par conséquent, le nombre de composantes connexes du graphe de Markov est une fonction décroissante de k , convergeant vers 1 (i.e., le processus est ergodique), et cette valeur est atteinte au plus pour $k = M$.

3.3. Complexité algorithmique et contraintes pratiques

Réaliser un k -échange dans $G \in \mathcal{G}_{\mathbf{C}}$ consiste à :

1. tirer aléatoirement k arêtes dans G ;
2. permuter leurs extrémités pour produire un graphe G' ;
3. vérifier que G' respecte l'ensemble des contraintes, c'est-à-dire vérifier que $G' \in \mathcal{G}_{\mathbf{C}}$.

Avec une structure de données appropriée (e.g., tableau de liste d'adjacence), il est possible d'effectuer les deux premières étapes en $O(k)$. La troisième étape dépend de l'ensemble de contraintes \mathbf{C} , l'optimisation de cette étape est souvent cruciale pour l'utilisation pratique de cette procédure. Réaliser n tentatives de k -échanges est donc au mieux en $O(nk)$. De plus, il faut aussi mesurer à intervalles réguliers les observables-cibles, pour vérifier la convergence de la chaîne de Markov ; ces observables et la fréquence de mesure doivent donc être choisies de manière adéquate pour ne pas dominer la complexité du processus.

Par ailleurs, la vitesse de convergence de la procédure dépend de manière non-triviale de k . En effet, un k -échange modifie la structure du graphe plus en profondeur à plus grand k , mais la probabilité d'échec du k -échange augmente avec k (sauf exception). En pratique dans la plupart des cas, nous constatons qu'à nombre de tentatives fixé, le taux de succès des k -échanges décroît (voir par exemple la dernière ligne de le tableau 1). Augmenter la valeur de k assure donc que le mélange sera d'une qualité supérieure ou égale, au prix d'une efficacité computationnelle diminuée.

Enfin, nous n'avons pas pu identifier expérimentalement de relations simples entre le taux de succès des k -échanges et la vitesse de convergence de l'algorithme. En effet, exprimer la complexité en fonction de cette grandeur semble être un choix pertinent, il est cependant difficile de proposer une

famille de graphes sur laquelle tester l'influence du taux de succès sans modifier des caractéristiques structurelles qui jouent également un rôle dans le calcul de complexité (taille, densité, etc).

3.4. Échantillonner des ensembles de graphes à l'aide de k -échanges

Le nombre de composantes connexes du graphe de Markov décroît de manière monotone avec la valeur de k , mais on ne sait pas, en général, quel k est nécessaire pour que le processus soit ergodique. Nous proposons alors une approche empirique pour réaliser pratiquement les échantillons. Nous prenons une valeur de k en commençant par le cas usuel $k = 2$, puis nous réalisons le processus markovien, qui est arrêté lorsque l'état stationnaire est atteint. Comme on l'a dit précédemment, l'état stationnaire est décrété par l'observation des plateaux des observables-cibles. Nous augmentons ensuite la valeur de k , et itérons la procédure jusqu'à ce que les valeurs des observables-cibles tendent vers des états stationnaires identiques pour $k = k^*$ et les valeurs suivantes (en pratique, nous vérifions le plus souvent la stabilité pour trois valeurs consécutives : k^* , $k^* + 1$ et $k^* + 2$). Ici, « états stationnaires identiques » signifie que nous réalisons pour chaque valeur de k un certain nombre de simulations (typiquement de l'ordre de la centaine) et que les intervalles de confiance à 95% obtenus sont cohérents pour les différentes valeurs de k . Nous qualifions k^* de *valeur seuil* et nous verrons par la suite qu'une petite valeur seuil est souvent suffisante en pratique. La validité de cette procédure repose sur l'hypothèse qu'augmenter k d'une unité permet une augmentation de la couverture de l'ensemble telle que l'ergodicité est très rapidement atteinte. Nos observations expérimentales soutiennent largement cette hypothèse.

En pratique. Pour plus de détails sur la réalisation pratique d'échantillons de graphes à l'aide des k -échanges, les codes correspondant à plusieurs applications et leur guide d'utilisation peuvent être téléchargés à l'adresse : <http://lioneltabourier.fr/program.html>

3.5. Contrainte de triangles orientés-colorés

Afin d'illustrer le fonctionnement de la méthode, nous proposons dans cette section un exemple sur un graphe et des contraintes artificielles, mais dont le rôle est de mettre en évidence le fonctionnement et l'intérêt des k -échanges.

Pour ce qui suit, rappelons que $\mathbf{C} = \mathbf{C}^0 \cup \mathbf{C}^+$ avec \mathbf{C}^0 l'ensemble des contraintes de base : type de graphe, nombre de nœuds, d'arêtes et distribution des degrés ; et \mathbf{C}^+ désigne les contraintes complémentaires à respecter.

Sur cet exemple les contraintes \mathbf{C} associées sont les suivantes :

1. \mathbf{C}^0 : le graphe est dirigé et constitué de $3N$ nœuds, la distribution des couples de degré (entrant, sortant) est fixée, chaque nœud ayant un arc entrant et un sortant ;
2. \mathbf{C}^+ :
 - les nœuds sont répartis en trois groupes de taille N , chacun correspondant à une couleur : rouge (R), vert (V) ou bleu (B) ;
 - le graphe est constitué de N cycles à trois nœuds, c'est-à-dire N triangles dans lesquels chaque nœud pointe vers un autre.

Nous cherchons à tirer aléatoirement un élément quelconque de $G_{\mathbf{C}}$ en partant du graphe G_0 tel que les triangles soient tous tricolores et "orientés R-B-V", i.e., le nœud rouge pointe vers le bleu qui pointe vers le vert.

Pour $k = 2$, le seul k -échange possible est l'identité, donc dans le graphe de Markov, il n'est pas possible de sortir de l'état G_0 . Pour $k = 3$, les k -échanges valides permettent d'échanger les arcs au

sein d'un même triangle, comme indiqué sur la figure 2; la chaîne de Markov associée mène donc exclusivement à des triangles "R-V-B" ou "R-B-V". Pour $k = 4$, les échanges d'arcs sont possibles entre les triangles, de manière à ce que toutes les combinaisons chromatiques soient possibles ("R-R-R", "R-V-V", etc.). Notons d'ailleurs que le graphe de Markov est connexe pour $k = 4$, et donc le processus est ergodique pour $k = 4 \ll M$ (la borne supérieure théorique de la valeur du seuil).

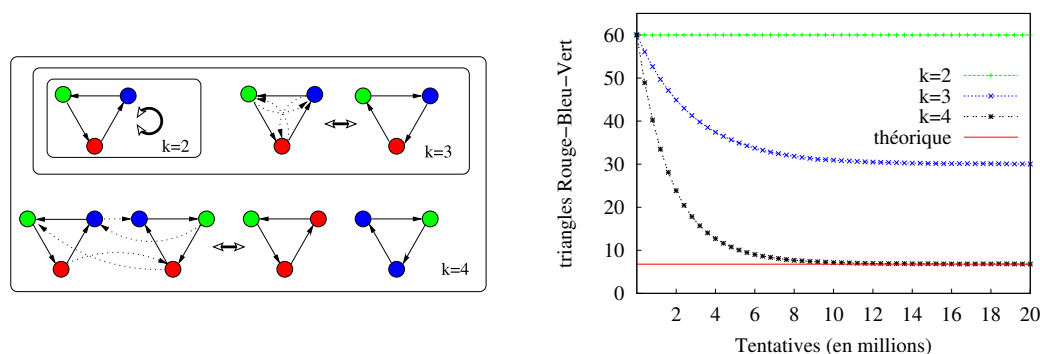


FIGURE 2. Gauche : Représentation schématique des types possibles de k -échanges pour $k \in \{2, 3, 4\}$, partant de la configuration "R-B-V". Droite : Nombre de triangles "R-B-V" en fonction du nombre de k -échanges, pour $k \in \{2, 3, 4\}$ (moyennes calculées sur 10 000 simulations pour chaque k , sur un graphe de 180 nœuds).

En pratique, nous réalisons des simulations à partir d'un graphe de $3N = 180$ nœuds constitué initialement de 60 triangles "R-B-V". Nous utilisons ici pour observable-cible la proportion de triangles de chaque combinaison de couleurs possibles et comparons les résultats des marches à k -échanges pour $k \in \{2, 3, 4, 5, 6\}$. Les résultats sont résumés dans la figure 2 et le tableau 1, où les moyennes sont obtenues sur 10.000 réalisations pour chaque valeur de k .

TABEAU 1. Fraction dans l'état stationnaire des triangles de chaque couleur-orientation en fonction de k , moyennée sur 10.000 simulations de 10^8 tentatives d'échanges, et nombre moyen de k -échanges réussis. La dernière colonne donne la valeur théorique moyenne attendue pour l'ensemble des graphes de \mathcal{G}_C .

	$k = 2$	$k = 3$	$k = 4$	$k = 5$	$k = 6$	Théorique (\mathcal{G}_C)
R-R-R	0.	0.	0.036	0.036	0.036	0.036
V-V-V	0.	0.	0.036	0.036	0.036	0.036
B-B-B	0.	0.	0.036	0.036	0.036	0.036
R-V-V	0.	0.	0.111	0.111	0.111	0.111
R-B-B	0.	0.	0.111	0.111	0.111	0.111
V-V-B	0.	0.	0.111	0.111	0.111	0.111
V-B-B	0.	0.	0.111	0.111	0.111	0.111
R-R-B	0.	0.	0.111	0.111	0.111	0.111
R-R-V	0.	0.	0.111	0.111	0.111	0.111
R-B-V	0.	0.500	0.113	0.113	0.113	0.113
R-V-B	1.000	0.500	0.113	0.113	0.113	0.113
Réussites	0	997 ± 74	2643 ± 108	2067 ± 132	936 ± 55	-

Nous pouvons observer que les états stationnaires mesurés pour $k \geq 4$ sont cohérents avec les valeurs théoriquement attendues pour un échantillon uniformément aléatoire, que l'on peut calculer analytiquement pour cet ensemble de contraintes \mathcal{C} . En revanche, les valeurs obtenues pour $k = 2$ (*switching & holding*) et $k = 3$ sont clairement différentes des valeurs théoriques, indiquant que le processus de Markov correspondant ne permet pas d'accéder à tous les éléments de \mathcal{G}_C . La procédure d'échanges multiples d'arêtes s'avère donc fiable sur cet exemple où les méthodes usuelles ne seraient pas utilisables.

4. Structure des réseaux de collaborations

Nous développons dans cette section un cas pratique réaliste dans lequel l'application de la méthode des k -échanges s'avère indispensable. Il s'agit de rechercher un ensemble de contraintes structurelles qui seraient suffisantes pour rendre compte de la structure de certains réseaux sociaux collaboratifs, plus précisément de réseaux scientifiques de co-signature d'articles. Nous allons tester deux hypothèses différentes correspondant à des ensembles distincts de contraintes.

4.1. Contrainte de triangles

La présence de structures triangulaires en grand nombre est connue pour être une caractéristique typique des réseaux sociaux, celle-ci peut-être interprétée comme une trace du phénomène de transitivité, identifié de longue date (Wasserman et Faust, 1994). Une première hypothèse consiste alors à tester si les contraintes de distribution des degrés et du nombre de triangles du graphe sont suffisantes pour expliquer la structure du réseau. L'ensemble de graphes à générer vérifie alors les contraintes suivantes :

- C^0 : le graphe est non-dirigé et simple (sans boucle, ni arête multiple), sa distribution des degrés est fixée ;
- C^+ : le nombre de triangles (non-dirigés) est fixé.

Le graphe G_0 représente un réseau de collaborations scientifiques extrait de l'*Anthropological Index Online database*². La base de données que nous considérons regroupe les articles relatifs à l'archéologie scandinave publiés pendant la période 2000–2009 : les nœuds sont les auteurs, les liens représentent l'existence de publications communes sur cet ensemble d'articles. G_0 comporte 273 individus et 280 liens, le jeu de données sera dénommé *Scan* par la suite.

4.1.1. Génération des échantillons

Nous utilisons comme observables-cibles les motifs connexes à quatre sommets et la distribution des distances entre les nœuds de la plus grande composante connexe. Les motifs à quatre sommets présentent deux intérêts ici : d'un point de vue technique, on dispose d'algorithmes efficaces pour les calculer au cours du processus (e.g., Wernicke (2006)) ; d'autre part, ils sont significatifs sociologiquement pour la description de l'environnement local d'un nœud. En effet, si le rôle prépondérant des motifs à trois sommets (ou triades) est bien connu dans les réseaux sociaux (Holland et Leinhardt, 1971), les motifs à quatre sommets permettent de décrire la structure locale de manière plus exhaustive, notamment en prenant en compte des séquences élaborées d'événements, comme celles donnant lieu aux chaînes $\hat{\diamond}$. Comme on va le voir, leur nombre ne peut pas être expliqué par les seules triades.

Les résultats de ces générations sont rassemblés dans la figure 3 et le tableau 2. On observe que l'état stationnaire pour $k = 2$ n'est pas compatible avec les états stationnaires pour $k \in \{3, 4, 5, 6\}$, qui ne sont pas discernables entre eux. Selon notre méthode, nous considérons donc qu'il est nécessaire d'utiliser une valeur de $k \geq 3$ pour générer un échantillon de graphes uniformément aléatoire de l'ensemble de graphes \mathcal{G}_C recherché. Au contraire, en faisant usage de la méthode *switching & holding* classique, on obtiendrait un échantillon fortement biaisé.

Nous observons également l'incompatibilité des états stationnaires pour $k = 2$ et $k \geq 3$ en traçant la distribution des distances³ des graphes produits (c.f. figure 4). En effet, alors que la distribution correspondant à l'état stationnaire du processus *switching & holding* est très proche de celle du graphe

² accessible à <http://aio.therai.org.uk>

³ C'est-à-dire la distribution des longueurs des plus courts chemins pour l'ensemble des paires de sommets.

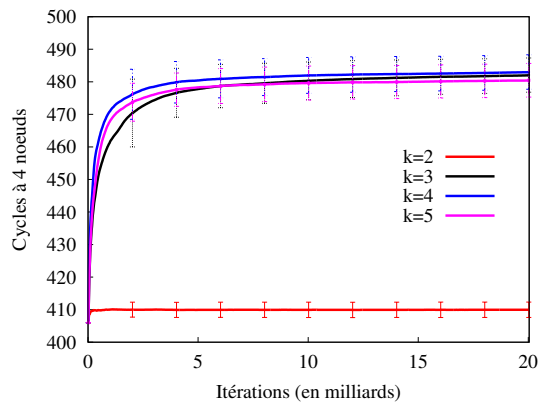


FIGURE 3. Moyenne cumulée du nombre de cycles à 4 nœuds (sur 20 réalisations pour chaque valeur de k), les barres d'erreur correspondent à l'intervalle de confiance à 95 % sur la valeur moyenne.

TABLEAU 2. Nombre moyen de motifs de taille 4 après 20 simulations et 10 milliards de tentatives de k -échanges à partir de G_0 (base de données Scan), erreurs : intervalles de confiance à 95 %.

Observables-cibles	initial G_0	$k = 2$	$k = 3$	$k = 4$	$k = 5$	$k = 6$
\diamond	2794	2799 ± 2	2907 ± 24	2933 ± 15	2942 ± 29	2894 ± 19
\diamond	406	410 ± 2	483 ± 3	483 ± 3	481 ± 3	482 ± 3
\diamond	730	734 ± 2	843 ± 5	841 ± 5	841 ± 3	840 ± 4
\diamond	108	108 ± 0	120 ± 1	120 ± 1	120 ± 1	119 ± 2

initial, ce n'est pas le cas de la distribution obtenue pour $k \geq 3$, qui correspondrait à un échantillonnage uniformément aléatoire. À nouveau, cette observation met en lumière qu'un échantillon généré par les méthodes d'échanges simples peut être fortement biaisé statistiquement et ainsi mener à des interprétations erronées.

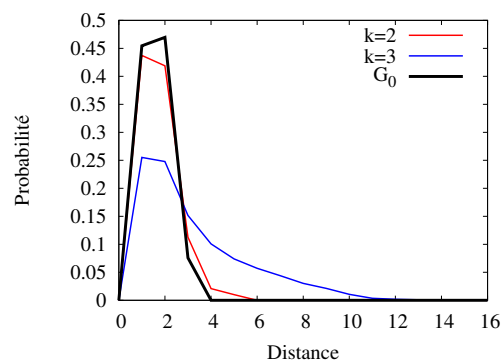


FIGURE 4. Distribution des distances du graphe réel G_0 et de l'état stationnaire pour les processus d'échanges à $k = 2$ (switching & holding) et $k = 3$ (toute valeur de $k > 3$ produit une allure identique à $k = 3$).

4.1.2. Test de l'hypothèse

Nous testons l'hypothèse sur plusieurs jeux de données de collaborations scientifiques, également extraits de la section archéologie de l'*Anthropological Index Online*, mais pour différentes régions géographiques, permettant de manipuler des réseaux de tailles variées, mais de connectivités moyennes comparables :

- *Brit* : correspondant aux publications concernant les îles britanniques durant la période 2000-2009 (807 auteurs, 1020 arêtes),
- *Euro* : concernant la totalité de l'Europe pour la même période (5289 auteurs, 5309 arêtes),
- *Scan* : c.f. ci-avant (273 auteurs, 280 arêtes).





En réalisant la génération des échantillons par la méthode des k -échanges, nous observons que $k^* = 3$ pour *Scan* (c.f. ci-avant), 2 pour *Brit* et 3 pour *Euro* (non-détaillés ici). Une fois que nous disposons d'échantillons satisfaisant les exigences de la méthode (et donc supposément uniformément aléatoires), nous revenons à l'objectif de cette application : vérifier si les contraintes de connectivité et du nombre de triangles sont suffisantes pour justifier certains éléments de la structure des graphes de collaboration.

La qualité de la reconstruction est évaluée à l'aide de métriques structurelles caractérisant les graphes à petite et grande échelles :

- a. Nous dénombrons les motifs connexes à 4 sommets qui décrivent l'environnement local des nœuds du graphe. Notons que par construction, le modèle génère des graphes comportant strictement le même nombre de motifs connexes à 3 sommets (triangles et chemins de longueur 2) que le graphe initial.
- b. Par ailleurs, nous mesurons la distribution des distances dans la plus grande composante connexe du graphe, ce qui informe sur la structure globale de l'ensemble.

Pour chacun des cas précédents, les mesures sont regroupées dans le tableau 3 et la figure 5.

TABLEAU 3. Nombre de motifs de taille 4 pour le graphe réel et la moyenne sur 50 réalisations du modèle.

Observables	<i>Brit</i> réel	<i>Brit</i> modèle	<i>Euro</i> réel	<i>Euro</i> modèle	<i>Scan</i> réel	<i>Scan</i> modèle
	18509	24867	72329	87998	2794	2924
	1460	3368	7084	13401	406	488
	2631	5961	13200	23697	730	850
	391	865	2045	3432	108	121

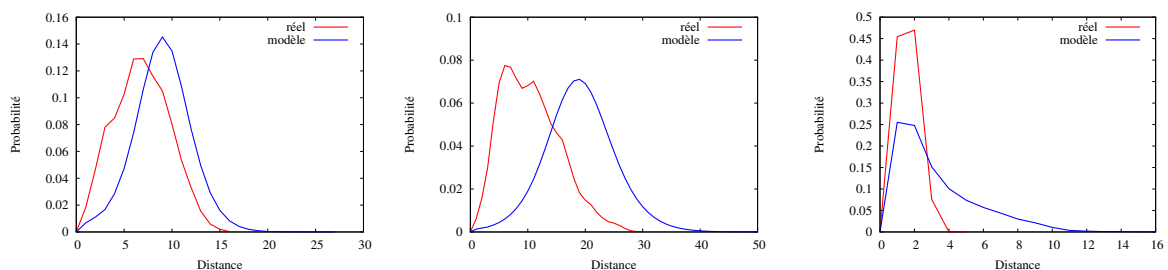


FIGURE 5. Distribution normalisée des distances dans la plus grande composante connexe du graphe, en comparant le graphe original à la moyenne pour 50 réalisations du modèle à nombre de triangles fixés. Gauche : *Brit*, milieu : *Euro*, droite : *Scan*.

Nous constatons que le nombre de motifs de taille 4 peut varier du simple à plus du double entre le modèle et le graphe réel. Les distributions des distances dans la plus grande composante connexe sont également très différentes pour les trois jeux de données : on observe ainsi que la valeur moyenne varie substantiellement entre le modèle et le graphe réel.

Même s'il n'y a pas de critère indiscutable pour juger de la qualité du modèle, ces seules métriques montrent que la structure des graphes produits n'est pas satisfaisante pour reproduire les propriétés classiques des réseaux de collaboration réels. Par conséquent la distribution des degrés et le nombre de triangles ne sont pas des éléments suffisants pour expliquer la structure du réseau dans le contexte des collaborations scientifiques.

4.2. Contraintes sur la structure bipartie des réseaux collaboratifs

Les triangles et la distribution des degrés ne peuvent donc pas rendre compte à eux seuls de certaines spécificités structurelles des réseaux de collaboration. Nous allons donc chercher à définir un nouveau jeu de contraintes qui seraient plus pertinent.

4.2.1. Structure bipartie sous-jacente

Les réseaux sociaux ont pour briques élémentaires des *événements* rassemblant un nombre quelconque de participants, parfois seulement deux mais éventuellement plus, comme dans le cas de conversations par exemple. Autrement dit, la représentation naturelle de ce type de réseaux serait un hypergraphe (un ensemble d'ensembles d'agents), ou de manière isomorphe, un graphe biparti (un ensemble d'agents liés à un ensemble d'événements auxquels ils ont participé).

Dans la plupart des cas, les études se concentrent sur la structure de la projection du réseau biparti, c'est-à-dire le réseau obtenu en connectant entre eux les agents participant au même événement, qui est souvent considéré comme *le* réseau social de référence. Cette approche permet de saisir aisément qui interagit avec qui, mais masque des effets plus complexes tels que la transitivité dont nous faisons état précédemment ; le graphe biparti sous-jacent pourrait donc être un choix plus pertinent pour expliquer la structure observée. Nous allons alors explorer les contraintes de connectivité des différents modèles possibles des graphes de collaborations scientifiques, et pour ce faire nous devons employer la méthode des k -échanges.

Plus précisément nous allons étudier des ensembles de graphes définis exclusivement par des contraintes de connectivité : sur le graphe projeté des interactions entre participants (contrainte monopartie, qui conserve la distribution du nombre cumulé de collaborateurs), sur le graphe biparti rendant compte des événements sous-jacents (contrainte bipartie, qui conserve la distribution des participations à des événements), et sur la combinaison des deux (contrainte monopartie-bipartie).

4.2.2. Contrainte monopartie C_M

Nous examinons d'abord la contrainte sur la connectivité du graphe projeté, le modèle correspondant étant un graphe monopartie dont la distribution des degrés est identique à celle du graphe réel ; on la schématise sur la figure 6. Avec les notations définies précédemment, l'ensemble C_M de contraintes correspondant est :

- C^0 : le graphe est non-dirigé et simple (sans boucle, ni arête multiple), sa distribution des degrés est fixée ;
- $C^+ = \emptyset$, il n'y a pas de contraintes additionnelles.

Les graphes vérifiant ces contraintes s'identifient aux graphes produits par le très classique *modèle configurationnel* (Bender et Canfield, 1978; Molloy et Reed, 1995; Newman *et al.*, 2002), évoqué précédemment. Il a été montré que cette contrainte suffisait à rendre compte de certaines propriétés des réseaux d'interaction, telles que la taille des composantes connexes (Newman *et al.*, 2002; Tangmunarunkit *et al.*, 2002), ainsi que, dans une certaine mesure, la distribution des distances (Mahadevan *et al.*, 2006) et des tailles de voisinage (Tangmunarunkit *et al.*, 2002). En revanche, elle échoue en général à rendre compte de la cohésion locale des réseaux, par exemple, la densité en triangles et en cliques (Newman *et al.*, 2002; Guillaume et Latapy, 2004). Notons également que cette famille de graphes est employée comme base de comparaison dans les méthodes de détection communautaire faisant usage de la notion de *modularité* (Girvan et Newman, 2002).

Il existe un grand nombre de méthodes pour générer de tels graphes, dont le modèle configurationnel ou la stratégie *switching & holding*, qui permettent de créer des échantillons uniformément aléatoires.

4.2.3. Contrainte bipartie C_B

Nous examinons ensuite la contrainte de connectivité sur le graphe biparti (ou contrainte hypergraphique) dans lequel on considère deux types de sommets : les acteurs du réseau associés aux événements auxquels ils participent. On utilise alors une variante du modèle précédent adaptée à ce contexte : la distribution des degrés des deux familles de nœuds est fixée, en d'autres mots, la taille de chaque événement ainsi que le nombre d'événements auxquels chaque acteur participe sont fixés, comme représenté sur la figure 6. La mise en correspondance des agents avec les événements est aléatoire. La contrainte C_B peut être décomposée avec les notations précédentes :

- C^0 : le graphe est non-dirigé biparti et sans arête multiple, la distribution des degrés des sommets de chaque famille de la partition est fixée ;
- $C^+ = \emptyset$, il n'y a pas de contraintes additionnelles.

Diverses études mettent en évidence la pertinence d'une telle structure, dans le contexte de réseaux sociaux de collaboration (Newman, 2004), ou encore dans celui de réseaux d'échanges de fichiers pair-à-pair (Guillaume *et al.*, 2005). Ce type de modèle implique par construction l'existence de cliques dans le graphe projeté et permet ainsi de rendre compte en partie de la cohésion locale des réseaux sociaux. Toutefois ce modèle ne reproduit la cohésion locale qu'assez grossièrement, et s'avère souvent insuffisant pour expliquer la structure du graphe à longue portée (par exemple, la distribution des distances).

Le problème technique de leur génération est proche de celui des graphes monopartis à distribution des degrés fixée, il est possible d'adapter le modèle configurationnel à ce contexte ; il est également possible d'utiliser une stratégie de type *switching & holding*, c'est ce que nous faisons ici pour produire les échantillons.

4.2.4. Contrainte monopartie-bipartie C_{MB}

Les interactions dans les réseaux sociaux sont déterminées à la fois par le nombre et la taille des groupes auxquels on participe, mais également par la façon dont les interactions entre paires d'individus sont possibles. Autrement dit, les contraintes de degré dans le graphe biparti et dans sa projection peuvent jouer un rôle dans la structure du graphe observé. Nous allons chercher à générer des échantillons uniformément aléatoires de graphes dont le degré biparti des acteurs et des événements est fixé, mais également le degré de chaque acteur dans la projection monopartie associée. La contrainte C_{MB} correspondante se décompose en :

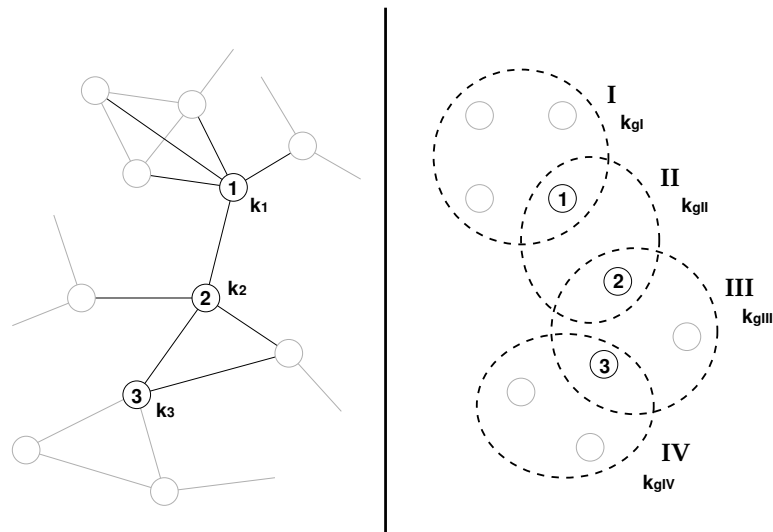


FIGURE 6. Gauche : représentation de la contrainte monopartie C_M , portant sur le degré des sommets acteurs (la distribution des k_i est conservée). Droite : représentation de la contrainte bipartie C_B , portant sur la taille des événements et à la participation des acteurs aux événements (la distribution des k_{gI} est conservée).

- $C^0 = C_B$: le graphe est non-dirigé biparti et sans arête multiple, la distribution des degrés des sommets de chaque famille de la partition est fixée ;
- $C^+ = C_M^a$: le degré des nœuds de la projection monopartie sur la famille des acteurs est fixé.

Notons qu'on impose aux nœuds de la famille des acteurs un couple de degrés biparti et monoparti : les deux distributions sont conjointes et non indépendantes.

Par ailleurs, il n'existe pas de méthode générative dans la littérature pour créer des graphes vérifiant cette contrainte, les échantillons seront donc générés ici avec la méthode des k -échanges.

4.3. Applications aux données de collaboration

Nous appliquons les modèles à connectivité contrainte C_M , C_B et C_{MB} sur les jeux de données de collaboration décrits dans la sous-section 4.1.2 : *Brit*, *Euro* et *Scan*.

4.3.1. Mesures sur les échantillons

Suivant une démarche identique à celle décrite en 4.1.2, le nombre de motifs connexes à quatre nœuds dans la projection du graphe des acteurs est utilisé pour quantifier la structure à courte portée du graphe. Les résultats sont reportés dans le tableau 4, et montrent que, conformément à ce que l'on sait des modèles C_M et C_B , C_M ne reproduit parmi ces motifs que le nombre de chemins de longueur 3, les autres motifs sont largement sous-estimés ; à l'inverse C_B reconstruit de manière fiable les motifs à quatre nœuds, excepté les chemins de longueur 3, qui sont surestimés. Nous observons ici que C_{MB} produit le bon ordre de grandeur pour les motifs à quatre nœuds de tout type, les écarts entre le nombre de motifs réels et le modèle n'excédant pas 20%.

À nouveau comme en 4.1.2, nous utilisons la distribution normalisée des distances de la plus grande composante connexe des graphes pour décrire leur structure globale. L'allure de ces distributions pour les différents modèles et le graphe d'origine est reportée sur la figure 7. On observe que C_M et C_B ne

TABLEAU 4. Nombre de motifs connexes de taille 4 pour le graphe réel et la moyenne sur 50 réalisations des différents modèles.

Obs.	Brit				Euro				Scan			
	réel	C_M	C_B	C_{MB}	réel	C_M	C_B	C_{MB}	réel	C_M	C_B	C_{MB}
$\langle \rangle$	18509	18582	26483	18861	72329	68836	109729	71011	2794	2947	3210	2792
\diamond	1460	40	1229	1201	7084	20	6114	5807	406	14	330	342
\triangleleft	2631	2	2412	2264	13200	0	12199	11442	730	1	657	647
\diamondleftarrow	391	0	396	355	2045	0	2030	1878	108	0	109	102

sont pas satisfaisants pour reproduire ces métriques, comme on pouvait l'attendre. En revanche, C_{MB} produit des distributions plus proches de celles du graphe correspondant au réseau réel. L'amélioration est marginale pour le jeu de données *Scan*, ceci peut-être considéré comme un effet de petite taille : la plus grande composante du graphe d'origine compte seulement 12 nœuds, et une faible modification de sa structure implique de fortes variations de la distribution des distances. En revanche pour les plus grands jeux de données *Brit* et *Euro*, on observe un bien meilleur accord entre le modèle C_{MB} et le graphe original en comparaison des modèles C_M et C_B .

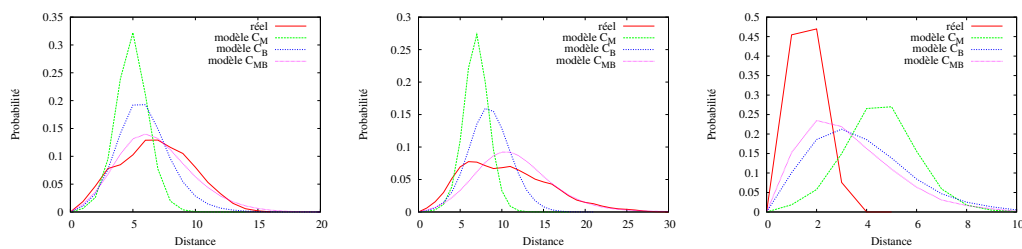


FIGURE 7. Distributions des distances dans la plus grande composante connexe. Gauche : Brit ; centre : Euro ; droite : Scan.

4.3.2. Discussion

Nous observons que la qualité de la reconstruction produite par le modèle C_{MB} est bien supérieure à celles de C_M et C_B , et ce résultat est confirmé pour les différents jeux de données employés. Pour les données de collaborations archéologiques, nous avons donc identifié un petit ensemble de caractéristiques suffisantes pour justifier des caractéristiques structurelles majeures du réseau réel. Nous pouvons alors formuler l'hypothèse que cette conclusion s'applique à un champ plus large que ces collections de données archéologiques. Autrement dit, le nombre d'articles publiés et le nombre de collaborateurs des agents d'une part, le nombre de coauteurs des articles d'autre part, seraient des ingrédients suffisants pour justifier un nombre accru de propriétés structurelles des graphes de collaboration scientifique.

Cependant, il est difficile de déterminer le domaine de validité de cette hypothèse. En effet, la structure du graphe dépend des habitudes de collaborations propres à chaque domaine, ainsi que d'autres paramètres tels que le mode de collecte des données (période, exhaustivité, etc). Ainsi, le modèle C_{MB} appliqué à des données de collaborations scientifiques issues de la physique (non-détaillé ici) produit des reconstructions plus proches des réseaux réels que C_M et C_B mais de qualité moindre que celles observées ici (e.g., les motifs à 4 nœuds sont reconstruits à plus de 20% près).

L'approfondissement de ces résultats nécessite de nouvelles campagnes de collectes et de mesures, mais nous sommes limités par les performances des algorithmes de type MCMC. En effet, pour atteindre l'état stationnaire sur les données collaboratives étudiées, on peut avoir à réaliser jusqu'à 10^{12} itérations, ce qui mobilise un microprocesseur standard pendant plusieurs jours. La création de modèles de graphes pertinents est donc étroitement liée aux contraintes techniques de génération de graphes aléatoires. C'est pourquoi l'accélération de la convergence du processus markovien permettrait d'élargir notablement la gamme d'application possible de cette méthode. Une piste dans cette direction consiste à étudier le *spectral gap* de la matrice de transition du processus markovien, c'est-à-dire la différence entre ses deux plus grandes valeurs propres, qui permet de borner le nombre d'itérations nécessaire à la convergence du processus (e.g., [Levin et al. \(2009\)](#)).

Conclusion

Nous avons présenté dans cet article une méthode expérimentale pour générer des échantillons uniformément aléatoires de graphes vérifiant un ensemble de propriétés fixé. Cette méthode peut être théoriquement employée quelles que soient les propriétés de l'ensemble, dans la limite où celles-ci contiennent la distribution des degrés et que l'on dispose d'un graphe appartenant à l'ensemble. Elle généralise les méthodes classiques d'échange en élargissant la procédure de permutation des extrémités de 2 liens du graphe à un ensemble de k liens, le paramètre k étant ajusté expérimentalement pour satisfaire la contrainte d'ergodicité de l'échantillon à produire.

La gamme d'utilisation d'une telle méthode est potentiellement très large. Elle permet notamment de déterminer un ensemble d'éléments fondamentaux qui rende compte de propriétés topologiques classiques d'un réseau donné. Nous avons développé une application de ce type pour le cas des réseaux de collaboration scientifique, et mis en évidence que certaines contraintes de connectivité des agents et des événements pouvaient justifier l'essentiel de la structure de plusieurs réseaux collaboratifs. Ce résultat nous guide pour proposer des mécanismes d'évolution du réseau de collaboration scientifique : nous pouvons y ajouter des publications artificielles qui respectent les contraintes de connectivité, et ainsi prédire comment la structure du réseau devrait évoluer.

Parmi les autres utilisations possibles, cette méthode permet de générer des jeux de données artificiels et parfaitement aléatoires ; une application de ce type à des réseaux d'échanges économiques a ainsi été proposée ([Henriet et al., 2012](#)). Enfin, la génération de graphes aléatoires permet également d'explorer les propriétés statistiques d'ensemble de graphes complexes vérifiant des propriétés fixées, par exemple évaluer leur taille, souvent méconnue et pourtant essentielle à la description des grands ensembles de graphes complexes.

Références

- Artzy-Randrup, Y. et Stone, L. (2005). Generating uniformly distributed random networks. *PRE*, 72(5):056708.
- Bansal, S., Khandelwal, S. et Meyers, L. (2008). Evolving Clustered Random Networks. *Arxiv preprint cs.DM/0808.0509*.
- Bender, E. et Canfield, E. (1978). The asymptotic number of labeled graphs with given degree sequences. *J. Combin. Theory Ser. A*, 24(3):296–307.
- Callaway, D. S., Newman, M. E., Strogatz, S. H. et Watts, D. J. (2000). Network robustness and fragility : Percolation on random graphs. *Physical review letters*, 85(25):5468.
- Colbourn, C. (1977). *Graph generation*. University of Waterloo.
- Coolen, A., De Martino, A. et Annibale, A. (2009). Constrained Markovian dynamics of random graphs. *Journal of Statistical Physics*, 136(6):1035–1067.
- Cooper, C., Dyer, M. et Greenhill, C. (2006). Sampling regular graphs and a peer-to-peer network. *Combinatorics, Probability and Computing*, 16(04):557–593.
- Eggleton, R. (1973). Graphic sequences and graphic polynomials : a report. *Infinite and Finite Sets*, 1:385–392.

- Feder, T., Guetz, A., Mihail, M. et Saberi, A. (2006). A local switch Markov chain on given degree graphs with application in connectivity of peer-to-peer networks. *In Proc. of FOCS*, volume 6, pages 69–76.
- Girvan, M. et Newman, M. (2002). Community structure in social and biological networks. *Proceedings of the National Academy of Sciences*, 99(12):7821.
- Gkantsidis, C., Mihail, M. et Zegura, E. (2003). The markov chain simulation method for generating connected power law random graphs. *In Proc. 5th Workshop on Algorithm Engineering and Experiments (ALENEX)*.
- Guillaume, J. et Latapy, M. (2004). Bipartite structure of all complex networks. *Information Processing Letters*, 90(5):215–221.
- Guillaume, J., Latapy, M. et Le-Blond, S. (2005). Statistical analysis of a P2P query graph based on degrees and their time-evolution. *Distributed Computing-IWDC 2004*, pages 439–465.
- Guruswami, V. (2000). Rapidly mixing markov chains : A comparison of techniques. MIT Laboratory for Computer Science. Available on cs.washington.edu/homes/venkat/pubs/papers.html.
- Henriet, F., Hallegatte, S. et Tabourier, L. (2012). Firm-network characteristics and economic robustness to natural disasters. *Journal of Economic Dynamics and Control*, 36(1):150–167.
- Holland, P. et Leinhardt, S. (1971). Transitivity in structural models of small groups. *Comparative Group Studies*.
- Kannan, R., Tetali, P. et Vempala, S. (1997). Simple Markov-chain algorithms for generating bipartite graphs and tournaments. *In Proceedings of the eighth annual ACM-SIAM symposium on Discrete algorithms*, pages 193–200. Society for Industrial and Applied Mathematics Philadelphia, PA, USA.
- Levin, D., Peres, Y. et Wilmer, E. (2009). *Markov chains and mixing times*. AMS Bookstore.
- Mahadevan, P., Krioukov, D., Fall, K. et Vahdat, A. (2006). Systematic topology analysis and generation using degree correlations. *In Proc. SIGCOMM'06*. ACM.
- Miklós, I. et Podani, J. (2004). Randomization of presence-absence matrices : comments and new algorithms. *Ecology Archives*, 85(1):86–92. Appendix A available on <http://esapubs.org/archive/ecol/E085/001/appendix-A.htm>.
- Milo, R., Kashtan, N., Itzkovitz, S., Newman, M. et Alon, U. (2003). On the uniform generation of random graphs with prescribed degree sequences. *Arxiv preprint cond-mat/0312.028*.
- Molloy, M. et Reed, B. (1995). A critical point for random graphs with a given degree sequence. *Random Structures and Algorithms*, 16(6):161–179.
- Newman, M. (2004). Coauthorship networks and patterns of scientific collaboration. *Proceedings of the National Academy of Sciences of the United States of America*, 101(Suppl 1):5200.
- Newman, M., Strogatz, S. et Watts, D. (2002). Random graphs models of social networks. *PNAS*, 99:2566–2572.
- Rao, A., Jana, R. et Bandyopadhyay, S. (1996). A Markov chain Monte Carlo method for generating random (0, 1)-matrices with given marginals. *Sankhyā : The Indian Journal of Statistics, Series A*, pages 225–242.
- Roberts, J. (2000). Simple methods for simulating sociomatrices with given marginal totals. *Social Networks*, 22(3):273–283.
- Sinclair, A. (1993). *Algorithms for random generation and counting : a Markov chain approach*. Springer.
- Stauffer, A. et Barbosa, V. (2005). A study of the edge-switching Markov-chain method for the generation of random graphs. *Arxiv preprint cs.DM/0512.105*.
- Tangmunarunkit, H., Govindan, R., Jamin, S., Shenker, S. et Willinger, W. (2002). Network topology generators : Degree-based vs. structural. *In Proc. SIGCOMM'02*, pages 147–159.
- Taylor, R. (1980). Constrained switchings in graphs. *Combinatorial Mathematics*, 8:314–336.
- Taylor, R. (1982). Switchings constrained to 2-connectivity in simple graphs. *SIAM Journal on Algebraic and Discrete Methods*, 3:114.
- Viger, F. et Latapy, M. (2005). Efficient and simple generation of random simple connected graphs with prescribed degree sequence. *Lecture Notes in Computer Science*, 3595:440.
- Wasserman, S. et Faust, K. (1994). *Social network analysis : Methods and applications*. Cambridge university press.
- Wernicke, S. (2006). Efficient detection of network motifs. *Computational Biology and Bioinformatics, IEEE/ACM Transactions on*, 3(4):347–359.