

Un générateur de réseaux dynamiques attribués avec structure communautaire

Oualid Benyahia*, Christine Largeron*
Baptiste Jeudy*, Osmar R. Zaiane**

*Univ Lyon, UJM-Saint-Etienne, CNRS, Institut d'Optique Graduate School,
Laboratoire Hubert Curien UMR 5516, F-42023, SAINT-ETIENNE, France.
oualid.benyahia,christine.largeron,baptiste.jeudy@univ-st-etienne.fr

**Department of Computer Science, University of Alberta, Edmonton, Canada.
zaiane@cs.ualberta.ca

Résumé. Nous proposons une nouvelle approche pour générer des graphes dynamiques avec attributs munis d'une structure communautaire reflétant les propriétés connues des graphes de terrain comme l'attachement préférentiel ou l'homophilie. Le générateur développé permet de construire une suite de graphes formant ainsi un réseau dynamique. Il offre la possibilité de visualiser l'évolution de ces graphes à travers une interface dédiée. Cette interface présente aussi plusieurs mesures évaluées sur chacun des graphes du réseau pour vérifier dans quelle mesure les propriétés du réseau sont préservées au cours de son évolution.

1 Introduction

La prolifération des réseaux d'information complexes dans divers domaines d'application a engendré une panoplie de méthodes pour l'analyse et la détection de motifs dans ces réseaux. Cependant, l'évaluation de ces méthodes et la validation des résultats qu'elles produisent n'est pas une tâche aisée étant donné le manque ou l'indisponibilité de grands réseaux réels munis d'une vérité de terrain accessible librement aux chercheurs. Une alternative consiste alors à exploiter des données synthétiques produites par des générateurs. Il existe une large bibliographie en ce qui concerne la génération de graphes statiques ; à l'instar du modèle classique d'Erdős-Rényi (ER) qui construit des graphes aléatoires ou encore du modèle de Barabási-Albert (BA) qui génère des réseaux aléatoires invariants d'échelle. Mais rares sont les générateurs qui permettent de construire des graphes évolutifs, exhibant ou non une structure communautaire, et qui associent des attributs aux nœuds.

L'intérêt des tâches de détection de communautés, de prédiction de liens ou en général de détection de motifs dans les réseaux dynamiques où les nœuds sont décrits par des attributs nous a amené à développer un générateur de graphes dynamiques avec attributs munis d'une structure communautaire. Ce dernier est une version étendue d'un générateur dédié à la génération de graphes statiques (c.f.(Largeron et al., 2015; Benyahia et al., 2016)).

Ce générateur permet de construire une séquence de graphes attribués avec une structure communautaire bien définie qui évolue au court du temps tout en préservant les propriétés bien connues des réseaux réels. Ainsi il peut être utilisé pour évaluer des méthodes de fouille de

graphes ou de réseaux puisque l'utilisateur peut obtenir un premier réseau de référence puis, en changeant les paramètres, en générer d'autres dans lesquels la structure communautaire ou les propriétés seront moins bien vérifiées, et ainsi il peut évaluer la robustesse de sa méthode sur des données bruitées.

2 Modèle

Un réseau dynamique avec attributs et structure communautaire généré par DANCer est représenté par : (1) une séquence de T graphes avec attributs $\mathcal{G}_i = (\mathcal{V}_i, \mathcal{E}_i)$, $i \in \{1, \dots, T\}$, où \mathcal{V}_i est l'ensemble des nœuds, \mathcal{E}_i l'ensemble des arêtes et où pour chaque nœud $v \in \mathcal{V}_i$ et chaque attribut $A \in \mathcal{A}$, v_A représente la valeur de l'attribut A assignée au nœud v et (2) une séquence de T partitions \mathcal{P}_i de \mathcal{V}_i , $i \in \{1, \dots, T\}$ qui attribuent une communauté pour chaque nœud dans le graphe correspondant \mathcal{G}_i , $i \in \{1, \dots, T\}$. Chaque partition permet de définir une structure communautaire pour chacun des graphes de la séquence (i.e., le réseau à un instant donné) de manière à avoir dans une même partie (i.e. communauté) des nœuds parfaitement reliés entre eux et relativement homogènes vis à vis des attributs, tandis que les nœuds appartenant à des parties différentes seront moins connectés et plus dissimilaires vis à vis des attributs.

La génération du réseau passe par deux étapes. Dans la première, un graphe initial $\mathcal{G}_1 = (\mathcal{V}_1, \mathcal{E}_1)$ est construit en respectant les propriétés bien connues des réseaux telles que l'attachement préférentiel ou encore l'homophilie. Dans la deuxième étape, le graphe initial précédemment construit subit des modifications selon deux types d'opérations. Les opérations du premier type, appelées "micro opérations", consistent à enlever des nœuds (respectivement des arêtes) existants ou à ajouter des nouveaux nœuds (respectivement de nouvelles arêtes) ou à modifier les attributs. Dans le deuxième type d'opérations (i.e., "macro opérations"), les modifications sont appliquées au niveau des communautés. Ces opérations consistent à : (1) migrer les membres d'une communauté vers une nouvelle communauté ou vers une communauté déjà existante, (2) scinder une communauté en deux sous communautés distinctes ou (3) fusionner deux communautés en une seule.

3 Aperçu de l'interface du générateur

Ce générateur est également présenté dans la vidéo jointe à cette soumission. L'interface utilisateur dispose de trois panneaux visibles dans la figure 1. Dans le panneau de gauche, l'utilisateur peut modifier les différents paramètres du générateur présentés dans le tableau 1¹. Il faut souligner qu'une graine est utilisée lors du processus de génération, i.e., pour la génération des nombres aléatoires. L'utilisation d'une même graine permet notamment de reproduire exactement le même graphe initial \mathcal{G}_1 .

Le panneau contient deux onglets. le premier permet d'afficher le réseau généré. On peut accéder directement à chaque graphe via la barre de défilement, ce qui permet de voir séparément chaque graphe de la séquence (figure 1). Dans chaque graphe, la couleur des noeuds indique leur communauté. Le deuxième onglet ("Community Dynamics"), permet d'afficher

1. Pour une description détaillée des paramètres le lecteur est invité à consulter Largeron et al. (2015)

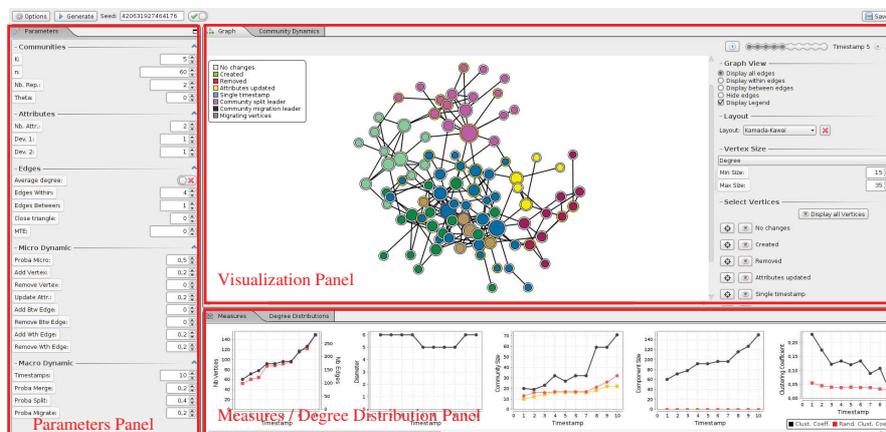


FIG. 1 – L'interface principale du générateur.

Parameter	Domain	Description
First timestamp		
K	\mathbb{N}^+	Number of communities
N	\mathbb{N}^+	Number of vertices
p	\mathbb{N}^+	Number of numerical attributes
$\mathcal{A} = \{\sigma_1, \dots, \sigma_p\}$		Standard deviations of the attributes generated using centered normal distributions
E_{wth}^{max}	\mathbb{N}	Maximum number of edges connecting a new vertex to vertices in its community
E_{btw}^{max}	$\{0, \dots, E_{wth}^{max}\}$	Maximum number of edges connecting a new vertex to vertices in a different community
$NbRep$	\mathbb{N}^+	Maximum number of representatives of each community
MTE	\mathbb{N}	Minimum number of total edges
$P_{randomCommunity}$	$[0, 1]$	A threshold to decide if a new vertex joins a randomly selected community or not
Micro operations		
$Proba\ Micro$	$[0, 1]$	A threshold to select if the micro dynamic updates are performed or not
$Add\ Vertex$	$[0, 1]$	Ratio defining the number of vertices inserted
$Remove\ Vertex$	$[0, 1]$	Ratio defining the number of vertices removed
$Update\ Attr.$	$[0, 1]$	Ratio defining the number of attributes updated
$Add\ Btw.\ Edges$	$[0, 1]$	Ratio defining the number of between edges inserted
$Remove\ Btw.\ Edges$	$[0, 1]$	Ratio defining the number of between edges removed
$Add\ Wth.\ Edges$	$[0, 1]$	Ratio defining the number of within edges inserted
$Remove\ Wth.\ Edges$	$[0, 1]$	Ratio defining the number of within edges removed
Macro operations		
$P_{removeEdgeSplit}$	$[0, 1]$	Proba. to remove an edge between two vertices in the previously same community when splitting a community
$Timestamps$	\mathbb{N}^+	Number of graphs generated
$Proba\ Merge$	$[0, 1]$	Probability to perform the merge operation
$Proba\ Split$	$[0, 1]$	Probability to perform the split operation
$Proba\ Migrate$	$[0, 1]$	Probability to perform the migrate vertices operation

TAB. 1 – Description of the dynamic network generator parameters

l'évolution des communautés induite par les opérations macro dynamiques (scission, fusion et migration).

Afin de vérifier que les graphes de la séquence sont construits en préservant les propriétés des réseaux réels, l'utilisateur peut consulter différentes mesures comme : la modularité,

le coefficient de clustering, le diamètre, l'homophilie attendue et observée ou encore le taux d'inertie calculés pour chaque graphe du réseau dynamique². Les mesures sont calculées après la génération et affichées dans le panneau en bas de l'interface.

Le panneau du bas inclut aussi un onglet pour afficher la distribution des degrés des nœuds pour chaque graphe de la séquence.

Le générateur offre la possibilité de sauvegarder le réseau dynamique en une collection de fichiers. Un premier fichier décrit chaque graphe (nœuds et arêtes ainsi que pour chaque nœud ses attributs et la communauté à laquelle il a été affecté dans la séquence). Un fichier "parameters" énumère tous les paramètres utilisés pour la génération du dit réseau, y compris la graine utilisée pour le générateur aléatoire. Les mesures et l'évolution des communautés peuvent aussi être sauvegardées dans des fichiers séparés.

4 Conclusion

Le générateur ainsi qu'un manuel d'utilisation détaillé³ sont disponibles sous les termes de la licence GNU GPL. Il faut souligner que le générateur peut être adapté pour produire des réseaux multiplexes où tous les nœuds sont présents à chaque niveau et où il existe une connexion entre les représentations d'un nœud d'un niveau à un autre. L'adaptation consiste à considérer que les pas de temps correspondent aux niveaux et à ajouter les liens inter-niveaux.

Références

Benyahia, O., C. Largeron, B. Jeudy, et O. R. Zaïane (2016). Dancer : Dynamic attributed network with community structure generator. In *Machine Learning and Knowledge Discovery in Databases - European Conference, ECML PKDD 2016, Riva del Garda, Italy, September 19-23, 2016, Proceedings, Part III*, pp. 41–44.

Largeron, C., P.-N. Mougel, R. Rabbany, et O. R. Zaïane (2015). Generating attributed networks with communities. *PLoS one* 10(4), e0122777.

Summary

We propose a new generator for dynamic attributed networks with community structure which follow known properties of real-world networks such as preferential attachment, small world or homophily. After the generation, the different graphs forming the dynamic network as well as its evolution can be displayed in the interface.

2. Pour plus de détails sur les propriétés et les mesures utilisées voir Largeron et al. (2015)

3. http://perso.univ-st-etienne.fr/largeron/DANCer_Generator/