Transformation de graphes dynamiques en signaux non stationnaires

Ronan HAMON[1,2], Pierre BORGNAT[1], Patrick FLANDRIN[1], Céline ROBARDET[2]

[1]Laboratoire de Physique de l'ENS de Lyon (UMR CNRS 5672)
46, allée d'Italie 69364 Lyon cedex 07, France

[2]LIRIS, INSA de Lyon (UMR 5205) Bâtiment Blaise Pascal
20, Avenue Albert Einstein 69621 Villeurbanne cedex, France

Avec le soutien des programmes ARC 5 et ARC 6 de la région Rhône-Alpes

ronan.hamon@ens-lyon.fr, pierre.borgnat@ens-lyon.fr
patrick.flandrin@ens-lyon.fr, celine.robardet@insa-lyon.fr

Résumé – De nombreux réseaux, qu'ils soient physiques, biologiques ou sociaux, peuvent se représenter sous la forme d'un graphe qui, si la composante temporelle est prise en compte, devient dynamique. L'étude de ces graphes dynamiques est compliquée du fait qu'il existe peu d'outils pour ce type d'objet. L'objectif ici est de proposer une nouvelle méthode pour visualiser de façon synthétique au cours du temps leur évolution. L'originalité des travaux consiste à aborder l'étude des graphes dynamiques à l'aide d'outils de la théorie du signal, en effectuant des analyses fréquentielles sur des séries représentant les graphes. La méthode proposée permet de transformer un graphe quelconque en une collection de signaux en utilisant le positionnement multidimensionnel classique puis de relier des motifs fréquentiels sur ces séries à des propriétés du graphe. L'application à des graphes dynamiques permet un suivi de ces motifs dans le temps : l'évolution de la structure globale du graphe est ainsi visible. Une méthode de reconstruction du graphe à partir des signaux est également proposée et permet de réduire le graphe en ne gardant que les liens les plus significatifs.

Abstract – Many data associated to networks, whether physical, biological or social, can be described by graphs that can become dynamic if a time evolution is added. These graphs are difficult to study because there exist only a few tools to describe these objects. The objective here is to propose a new method to visualize synthetically their evolution over time. The originality of this work consists of study dynamic graphs using a signal theory approach, by computing spectral analysis on series representing the graphs. The method consists of transforming a graph to a collection of signals using multidimensional scaling then linking frequency patterns of these series with graph properties. The extension to dynamic graph enables us to follow the evolution of these patterns and then track the modification of the structure of the graph over time. A method to reconstruct the graph from the collection of signals is also proposed to reduce the graph by selecting the most significant links.

1 Transformation d'un graphe statique en signaux

Notations Soit un graphe simple non-orienté et non-pondéré noté $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ où $\mathcal{V}$ est l'ensemble de ses sommets (ou nœuds) de taille $\#\mathcal{V} = n$ et $\mathcal{E}$ l'ensemble de ses liens de taille $\#\mathcal{E} = m$. On note $\mathbf{A} = (a_{ij})_{i,j=1,..,n}$ sa matrice d'adjacence.

1.1 Principe

Shimada et al. [**?**] ont proposé une méthode pour transformer un graphe à n nœuds en séries à n points indexées par les sommets en utilisant le positionnement multidimensionnel classique (*classical multidimensional scaling* ou *CMDS*) [**?**]. L'intérêt d'une telle transformation réside dans le fait que si les n signaux permettent de reconstruire exactement le graphe d'origine, il est également possible de réduire la représentation en ne conservant qu'un nombre réduit de signaux.

Transformation d'un graphe en signaux Cette transformation consiste à appliquer le *CMDS* sur la matrice de distance entre nœuds d'un graphe notée $\mathbf{\Delta} = (\delta_{ij})_{i,j\in 1,..n}$ et définie pour $i, j \in \mathcal{V}$ par

$$\delta_{ij} = \begin{cases} 0 & \text{si } i = j \\ 1 & \text{si } a_{ij} = 1 \text{ et } i \neq j \\ w > 1 & \text{si } a_{ij} = 0 \text{ et } i \neq j \end{cases} \quad (1)$$

avec $w > 1$. En suivant [**?**] on prendra $w = 1.1$ par la suite.

Le *CMDS* est une technique permettant d'obtenir les coordonnées de points dans un espace euclidien dont on ne connaît que les distances entre chaque paire de points. La matrice $\mathbf{X}$ des coordonnées peut être calculée analytiquement : on effectue un double centrage de la matrice $\mathbf{\Delta}$ dont les termes sont élevés au carré : $\mathbf{B} = -\frac{1}{2}\mathbf{J}\mathbf{\Delta}^{(2)}\mathbf{J}$ avec $\mathbf{J} = \mathbf{I}_n - \frac{1}{n}\mathbf{1}_n\mathbf{1}_n^T$ où $\mathbf{I}_n$ est la matrice identité de taille n et $\mathbf{1}_n$ est un vecteur colonne de taille n rempli de 1. La solution du *CMDS* est donnée par $\mathbf{X} = \mathbf{Q}_+\mathbf{\Lambda}_+^{\frac{1}{2}}$ avec $\mathbf{\Lambda}_+$ une matrice diagonale dont les termes sont les valeurs propres positives triées par ordre décroissant de la matrice $\mathbf{B}$ et $\mathbf{Q}_+$ la matrice des vecteurs propres correspondants. Les signaux caractérisant le graphe sont les composantes de la matrice $\mathbf{X}$ avec $\mathbf{X}^{(j)}$ j-ème composante de $\mathbf{X}$. Elles sont

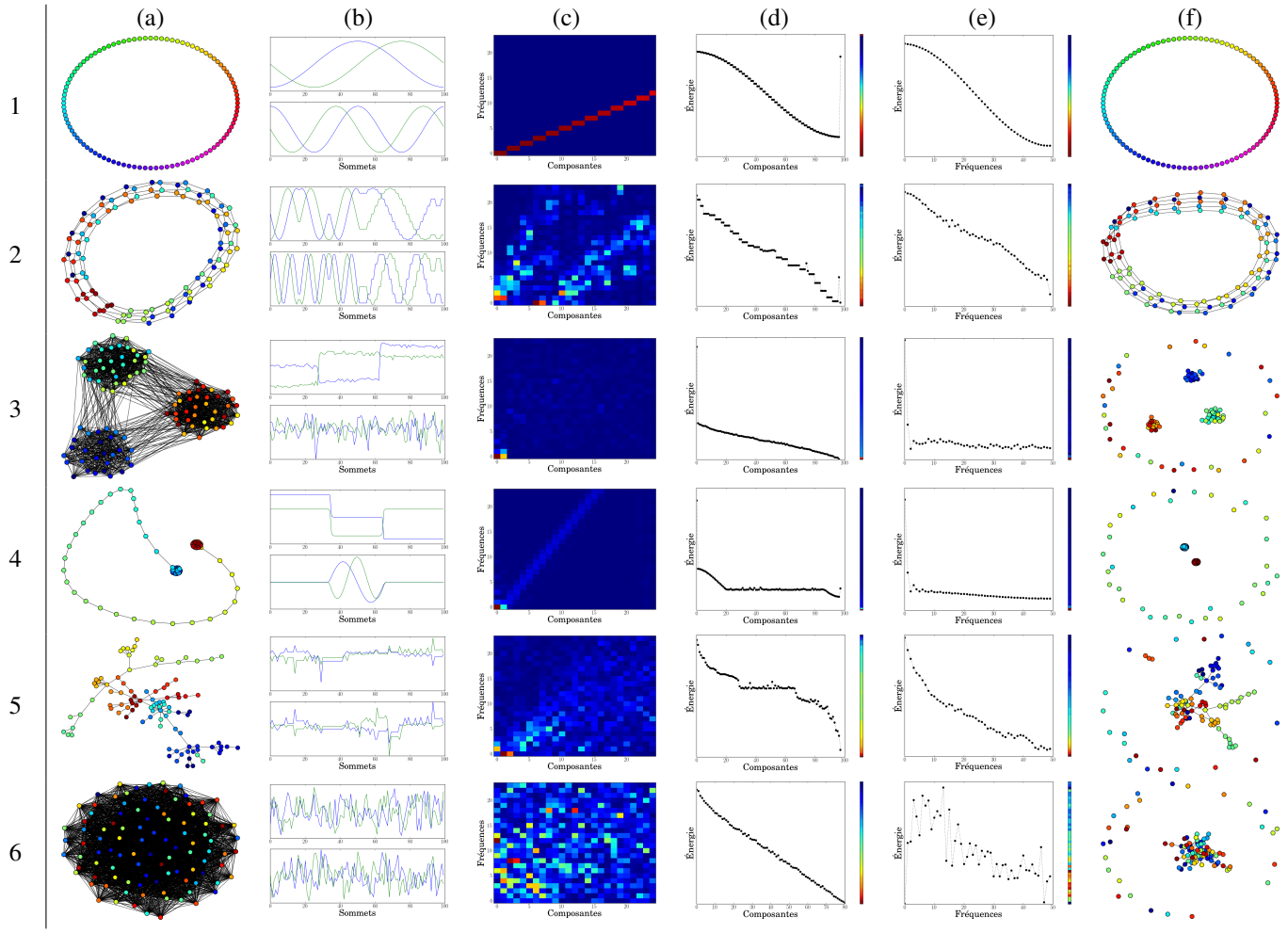


Figure 1: Exemples de graphes et signaux associés. Colonnes : (a) Représentation du graphe (la couleur représente le numéro d'étiquette du noeud). (b) Les 4 premières composantes associées au graphe. (c) Énergie en fonction de la composante et de la fréquence pour les 25 premières composantes et 25 premières fréquences en code couleur. (d) Énergie en fonction de la composante moyennée sur les fréquences avec représentation en code couleur. (e) Énergie en fonction de la fréquence moyennée sur les composantes avec représentation en code couleur. (f) Représentation du graphe après reconstruction (r arbitraire). La couleur est codée du bleu foncé (valeur faible) au rouge foncé (valeur élevée). Lignes : 1. Modèle de Watts-Strogatz : 100 nœuds, degré 2 avec probabilité nulle de changer le lien entre 2 nœuds. 2. Grille 2D 20×5 : 100 nœuds. 3. Graphe de 100 nœuds répartis en 3 communautés. 4. Modèle de Barbell : 2 cliques de 35 noeuds reliées par un chemin de 30 nœuds. 5. Modèle de Barabasi-Albert : 100 nœuds. 6. Modèle de Erdős-Rény : 100 nœuds, $p = 0.6$

indexées par les nœuds.

L'étiquetage des nœuds a une importance puisque deux nœuds très proches dans la numérotation mais distants dans le graphe vont avoir des valeurs sur chaque composante éloignées. Pour éviter des variations brusques du signal, une méthode d'étiquetage est discutée dans la sous-section suivante afin de bien numéroter les sommets avant la transformation.

1.2 Contributions à la méthode de transformation

Étiquetage d'un graphe L'étiquetage d'un graphe (*graph labeling*) consiste à attribuer à chaque sommet d'un graphe un entier naturel compris entre 0 et $n - 1$ en vertu d'un critère à minimiser [**?**]. Le problème qui nous intéresse ici est le *cyclic bandwidth problem*. Il cherche à minimiser la somme des distances dans la numérotation pour chaque paire de sommets voisins, cette distance étant définie pour deux sommets voisins numérotés i et j par $d(i, j) = \min(|i - j|, n - |i - j|)$. Nous avons proposé un nouvel algorithme plus rapide qui n'est pas l'objet de cette communication [**?**]. Il consiste à effectuer une recherche en profondeur du graphe où le choix de l'ordre de parcours des sommets est déterminé par l'indice de Jaccard entre un sommet et ses voisins.

Reconstruction du graphe à partir de signaux La reconstruction d'un graphe à partir d'une collection de signaux est triviale lorsque toutes les composantes sont prises en compte :

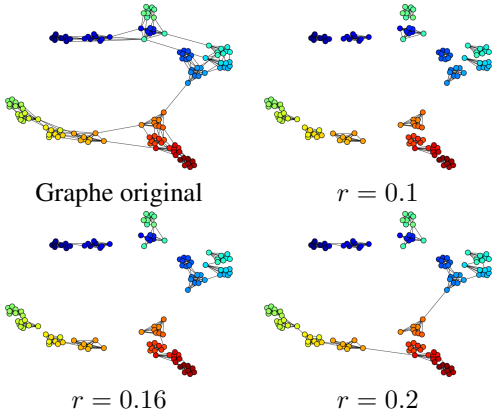


Figure 2: Exemple de reconstructions partielles du graphe sur un graphe avec communautés imbriquées à différentes échelles. Pour $r = 0.1$, seuls les nœuds appartenant aux communautés à petite échelle sont sélectionnés. Quand r augmente, des liens entre ces communautés se forment pour mettre en évidence les communautés à plus grande échelle.

selon le principe du *CMDS*, la matrice de distance $\mathbf{D}$ entre les points est identique à la matrice $\mathbf{\Delta}$ définie sur le graphe. Un simple seuillage suffit alors à retrouver la matrice d'adjacence du graphe et donc le graphe. Si moins de composantes sont retenues, les distances entre points ne sont plus égales à 1 ou à w mais ont une distribution dont la largeur dépend du nombre de composantes retenues. Ces distributions peuvent être utilisées pour sélectionner les liens les plus significatifs : pour une paire de sommets reliés entre eux, plus la distance entre ces sommets dans l'espace euclidien réduit va être faible, plus le lien est considéré comme significatif. Il est alors possible de définir un seuil pour déterminer la significativité d'un lien. Connaissant la matrice d'adjacence du graphe complet, il est possible d'identifier les distributions correspondants aux distances entre les paires de nœuds liées et non-liées, et ainsi de connaître l'intervalle sur lequel les deux distributions se chevauchent. Cet intervalle correspond à une zone où la distance entre deux nœuds n'est pas suffisante pour déterminer leur connectivité avec certitude dans le graphe complet. Le seuil est ainsi choisi de façon à ce que les liens sélectionnées soient «fiables» c'est-à-dire ne correspondent pas à des distances appartement à cet intervalle de chevauchement. Un exemple de reconstruction sur un graphe avec communautés est discuté à la section **??**. On note r la proportion de composantes retenues sur le nombre total de composantes.

1.3 Analyse spectrale des signaux

L'étude des signaux obtenus consiste à effectuer une analyse spectrale sur les composantes de la matrice $\mathbf{X}$. Soit une collection de K signaux indexés par n nœuds, le spectre de la composante k est donné par

$$S(k, f) = |\mathcal{F}\mathbf{X}^{(k)}(f)|^2$$

évalué, pour les fréquences positives, sur $\frac{n}{2} + 1$ échantillons, $\mathcal{F}$ étant la transformée de Fourier. Il est alors possible de calculer les caractéristiques suivantes :

- Énergie des composantes :
  $\forall k \in \{1, \cdots, K\}, E_k = \|\mathbf{X}^{(k)}\|^2 = \sum_{f=1}^{\frac{n}{2}+1} S(k, f)$

- Énergie des fréquences:
  $\forall f \in \{1, \cdots, \frac{n}{2} + 1\}, \bar{S}(f) = \sum_{k=1}^{K} S(k, f)$

Ces caractéristiques peuvent être reliées à des propriétés du graphe comme illustré dans la section suivante.

1.4 Exemples

Transformation et analyse spectrale La Figure **??** propose 5 exemples de graphes de 100 noeuds. Pour chaque exemple, la colonne (a) propose un aperçu du graphe. La colonne (b) affiche les 4 premières composantes de la collection de signaux obtenue après transformation. On peut comparer les caractéristiques de ces signaux avec les graphes dont ils sont issus : la présence de régularité dans le graphe se traduit par des sinusoïdes (graphes 1 et 2) alors que la présence de communautés (graphe 3) donne des signaux avec des paliers correspondant aux communautés du graphe. Il est intéressant de noter que la présence de ces deux propriétés (graphe 4) permet de retrouver des signaux avec des paliers et des signaux sinusoïdaux. Le graphe 5 illustre d'autres types de signaux qu'il est possible d'obtenir et qui présentent également des caractéristiques particulières. Enfin, la transformation sur un graphe aléatoire (graphe 6) montre l'absence de structure particulière sur les signaux si ce n'est celle ajoutée du fait du tri décroissant des composantes par énergie. Ces caractéristiques se retrouvent dans l'analyse spectrale : la colonne (c) représente l'énergie de chaque fréquence pour chaque composante, tandis que les colonnes (d) et (e) sont respectivement les marginales des composantes et des fréquences. Ces deux grandeurs permettent de représenter synthétiquement les caractéristiques des signaux et donc les propriétés du graphe. Enfin la colonne (f) donne un exemple de reconstruction des graphes pour une valeur r fixée pour chaque graphe de façon à obtenir une version réduite, c'est-à-dire avec moins de liens, du graphe.

Reconstruction de graphe à partir de signaux Un exemple détaillé de reconstruction d'un graphe à partir de signaux est donné en Figure **??** pour un graphe hiérarchique avec communautés : à petite échelle, 16 communautés sont visibles qui fusionnent progressivement pour former 8, puis 4 et enfin 2 grandes communautés lorsque l'échelle augmente. La sélection des liens les plus significatifs permet de mettre en évidence, pour différentes proportions de composantes retenues, ces différentes échelles de communautés.

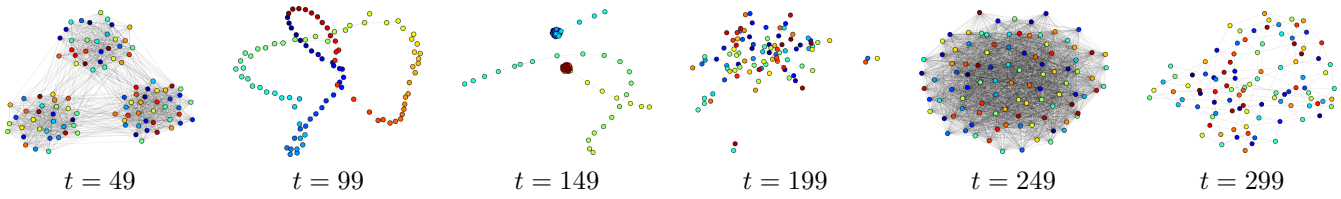


Figure 3: Instantanées du graphe dynamique à différents temps avant un changement du graphe modèle.

	$e \in \mathcal{E}_m$	$e \notin \mathcal{E}_m$
$e \in \mathcal{E}_{t-1}$	$1 - 10^{-4}$	0.90
$e \notin \mathcal{E}_{t-1}$	0.1	10^{-4}

Table 1: Probabilités qu'un lien e soit dans $\mathcal{E}_t$ ensemble des liens du graphe dynamique au temps t en fonction de la présence dans $\mathcal{E}_m$ ensemble des liens du graphe modèle et de $\mathcal{E}_{t-1}$ ensemble des liens du graphe dynamique au temps $t-1$

2 Extension aux graphes dynamiques

2.1 Principe

L'analyse fréquentielle des signaux issus d'une transformation d'un graphe permet de mettre en évidence des motifs caractéristiques des propriétés du graphe transformé. Cette approche peut être étendue aux graphes dynamiques : le suivi dans le temps de ces signatures fréquentielles permet de mettre en évidence l'évolution de la structure globale du graphe. On considère le cas où le nombre de sommets dans le graphe est fixé : seuls les liens évoluent au cours du temps. On note $\mathbf{X}_t^{(k)}$ la k-ème série issue d'un graphe au temps t, son spectre est donné par $S(k, f, t) = |\mathcal{F}\mathbf{X}_t^{(k)}(f)|^2$.

A chaque pas de temps, il est ainsi possible de calculer les caractéristiques détaillées dans la section précédente et de suivre au cours du temps l'évolution de la structure du graphe. On note $\mathcal{G}^t = (\mathcal{V}, \mathcal{E}_t)$ un graphe dynamique où $\mathcal{E}_t$ ensemble des liens à l'instant t.

2.2 Exemple

Un graphe de 100 noeuds est généré à partir d'un graphe modèle : à chaque pas de temps, deux nœuds sont liés avec une probabilité dépendant de la présence ou non du lien dans le graphe modèle et dans le graphe dynamique au temps précédent. La table **??** donne les probabilités utilisées alors que la Figure **??** représente des instantanées du graphe à l'instant précédant un changement du graphe modèle tous les 50 pas de temps.

La relation entre l'évolution des énergies par composante et par fréquence au cours du temps présentées en Figure **??** et les motifs discutés en Figure **??** permet de mettre en évidence les changements de structure du graphe à chaque instant et son évolution au cours du temps. Les ruptures brutales de structure dues au changement de graphe modèle sont visibles, ainsi que les transitions vers le graphe modèle.

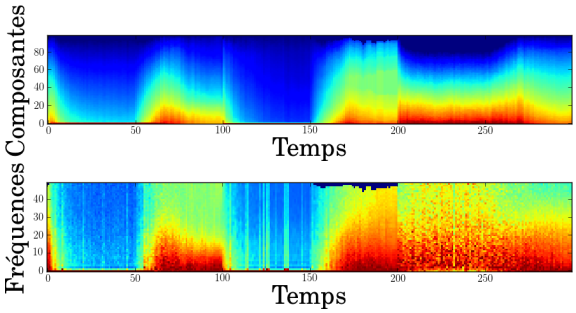


Figure 4: (Haut) : Énergie de chaque composante moyennée en fonction du temps. (Bas) : Énergie de chaque fréquence moyennée sur les composantes en fonction du temps. La couleur est codée du bleu foncé (valeur faible) au rouge foncé (valeur élevée).

3 Conclusion

La transformation de graphe en signaux permet la réalisation de deux objectifs : elle donne tout d'abord la possibilité de suivre la structure globale du graphe au cours du temps en rattachant des motifs fréquentiels des signaux à des propriétés du graphe. Chaque changement de la topologie du graphe est repérable et permet d'identifier des changements brusques dans la structure ou au contraire de lente transition vers une structure qu'il est possible d'identifier. Un deuxième aspect concerne la réduction du graphe en ne considérant qu'une proportion réduite de composantes afin de ne sélectionner que les liens les plus significatifs dans la structure du graphe.