



**HAL**  
open science

# Énumération de motifs dans un graphe d'évolution de communautés

Raphaël Charbey, Cécile Bothorel, Laurent Brisson

► **To cite this version:**

Raphaël Charbey, Cécile Bothorel, Laurent Brisson. Énumération de motifs dans un graphe d'évolution de communautés. MARAMI 2020, Oct 2020, Virtual Conference, France. hal-03028901

**HAL Id: hal-03028901**

**<https://hal.science/hal-03028901>**

Submitted on 27 Nov 2020

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License

# Énumération de motifs dans un graphe d'évolution de communautés

## Pattern mining in community evolution graph

Raphaël Charbey, Cécile Bothorel, and Laurent Brisson

IMT Atlantique  
Lab-STICC UMR CNRS 6285  
F-29238 Brest, France  
firstname.name@imt-atlantique.fr

**Résumé** Comprendre la dynamique des interactions entre individus dans les réseaux sociaux est un enjeu important, notamment via la dynamique communautaire. Le résultat d'algorithmes de clustering temporel aboutit à la production de complexes graphes d'évolution de communautés que nous identifions à des graphes dirigés acycliques aux arêtes colorées et avec attribut sur les nœuds. Dans cet article, nous proposons une approche basée sur l'énumération de motifs dans les graphes d'évolution de communautés, permettant de décrire le cycle de vie des communautés via la mise en évidence de différentes structures à l'échelle mésoscopique, c'est-à-dire à un niveau intermédiaire entre celui des sommets et celui du graphe lui-même. Nous proposons également une méthode originale de coloration des arêtes qui permet de réduire le nombre de motifs possibles et rendre ainsi plus aisée l'analyse de ces structures. Enfin, une expérimentation basée sur un jeu de données d'échanges de commentaires sur YouTube permet d'illustrer la mise en oeuvre de notre approche.

**Abstract** To understand the dynamic of individual interaction through dynamic communities is an important matter in Social Networks. Temporal clustering algorithms produce complex community evolution graphs that we treat as directed acyclic graphs with labelled edges and attributed nodes. In this article, we propose a pattern-based approach in such graphs, that provides a description of the life cycle of communities through the exhibition of several mesoscopic structures. We also propose an original way of labelling the edges that allows to reduce the list of possible patterns and so does the analysis more simple to conduct. Finally, an experiment based on a Youtube comment dataset illustrates the processes we execute to this aim.

**Keywords:** Graphlet, Motifs, Communautés, Graphe temporel, Sankey

## 1 Introduction

Les réseaux, ou graphes, qui permettent d'étudier les interactions entre un ensemble d'individus ou d'objets, font désormais figure d'outil d'analyse classique

dans de nombreux domaines de recherche. Détecter des communautés dans ces réseaux permet d'en avoir une vue synthétique. Il s'agit alors d'étudier l'organisation des individus au sein des différents groupes, ou les relations entre ces groupes, ou encore la dynamique de ces groupes dans les réseaux temporels.

Rossetti et Cazabet présentent de nombreuses approches pour détecter des communautés évolutives dans leur revue de littérature [30]. Le résultat de ces méthodes est un ensemble de communautés évolutives (appelées parfois communautés dynamiques). Elles sont décrites comme une séquence d'observations d'une même communauté au fil du temps. Les diagrammes alluviaux ou diagrammes Sankey, qui sont des outils de visualisation de flux, permettent de représenter ces différents états des communautés et, via les arêtes, les flux d'individus entre observations d'un pas de temps à l'autre.

Au-delà de la visualisation, le formalisme lui-même — un diagramme Sankey est un graphe orienté pondéré sans cycle, avec attributs sur les nœuds — permet d'appliquer l'ensemble des outils d'analyse de graphe et de réseaux sociaux. Plusieurs travaux se sont d'ores et déjà intéressés à l'analyse de communautés évolutives [7,23], et parfois déjà en utilisant des représentations proches des diagrammes Sankey [18,32], sans toutefois exploiter le formalisme de graphes ou les outils dédiés à l'analyse des réseaux sur ces objets. Nous montrons ici, que la détection de motifs est particulièrement appropriée pour suivre avec précision l'ensemble des étapes de la vie d'une communauté, pour détecter des similarités entre les formes d'évolution ou encore pour cerner certains moments ayant impacté l'ensemble des communautés présentes.

L'objectif de nos travaux est d'exploiter le formalisme des diagrammes Sankey à travers l'utilisation de techniques d'énumération de motifs. Nous cherchons à détecter les transformations classiques de naissance, mort, fusion, scission, croissance et décroissance [30]. L'originalité de notre approche réside dans le fait qu'elle est ne repose sur aucun a priori concernant les données et les structures qui s'y trouveraient. Il s'agit d'énumérer de manière exhaustive les différentes variantes que prennent ces six transformations du cycle de vie des communautés présentes, et éventuellement d'en détecter d'autres, moins intuitives en s'appuyant sur la vision fine, à l'échelle mésoscopique (c'est-à-dire à un niveau intermédiaire entre celui des sommets et celui du graphe lui-même), des structures qui composent un graphe.

L'objectif de ce papier est d'apporter une nouvelle méthode adaptée à l'analyse de l'évolution de communautés basée sur l'énumération de motifs dont les arêtes sont orientées *et* colorées :

- Nous proposons un formalisme inspiré des diagrammes Sankey, le graphe d'évolution de communautés, qui permet de décrire la vie des communautés, grâce à un graphe dirigé acyclique aux arêtes colorées et avec attribut sur les nœuds.
- La coloration des arêtes est une fonction qui associe un label aux arêtes. Cette fonction peut selon les cas refléter un poids (ex. petit, moyen et grand) ou une fonction plus complexe. Nous proposons notamment d'utiliser une double pondération décrivant la proportion de membres quittant

la communauté de départ et la part de ces individus arrivant dans la communauté d'arrivée au pas de temps suivant.

- Grâce à cette coloration, nous appliquons des techniques existantes de canonisation puis de construction de l'arbre des parcours de graphlets G-Trie [28,21], de façon à énumérer l'ensemble des motifs présents dans un graphe d'évolution de communautés.
- Enfin, à travers nos expériences préliminaires, nous montrons le pouvoir descriptif des motifs énumérés pour caractériser la fusion de communautés.

Cet article est structuré de la manière suivante : en section 2 nous présentons un état de l'art sur les études relatives à la dynamique des communautés et à l'énumération de motifs. En section 3, nous présentons des définitions, le processus au coeur de notre contribution et plus particulièrement notre proposition de coloration des arêtes du graphe. En section 4, nous proposons un scénario-type d'utilisation de notre méthode en utilisant un jeu de données dédié à l'analyse de commentaires publiés sur YouTube. Enfin, en section 5 nous concluons et faisons le bilan de nos contributions.

## 2 État de l'art

### 2.1 Étude des dynamiques de communautés

La dynamique des communautés en ligne demeure une question importante de la recherche actuelle, tant au niveau de la détection de celles-ci que de leur analyse [7].

Plusieurs études se sont penchées sur l'analyse de l'évolution des communautés sociales via une représentation en réseaux. Palla et al. analysent l'évolution de communautés dynamiques au sein de graphes de co-publications et d'appels téléphoniques [23]. Ils décrivent notamment 6 formes d'événements qui jalonnent l'existence de ces communautés : la naissance et la mort, la croissance et la perte de membres, la fusion et la scission. Goldberg et al. mesurent plusieurs paramètres au long de la vie de communautés détectées dans des graphes temporels et concluent que certaines de ces mesures, prises aux premiers pas de temps de chaque communautés, sont fortement corrélées avec leur durée de vie [9].

Plusieurs études s'appuient, avec plus ou moins d'insistance, sur des visualisations proches des diagrammes de Sankey, comme Muyata et Moriyasu qui représentent sous la forme d'un réseau orienté (mais non pondéré) l'évolution des communautés temporelles qu'ils détectent [22]. Lin et al. proposent, par ailleurs, un framework d'analyse de détection de l'évolution des communautés leur permettant de suivre l'évolution des membres de plusieurs réseaux au fil du temps [18]. Leur framework aboutit en un *Evolution Net*, un graphe dont les arêtes sont pondérées par la probabilité de passer d'une communauté à l'autre entre deux pas de temps. Les auteurs n'utilisent pas, par la suite, le formalisme des graphes pour analyser ces objets. Rosvall et Bergstrom utilisent également la

visualisation en diagramme de Sankey pour représenter l'évolution des disciplines scientifiques, vues comme des communautés de chercheurs [32].

## 2.2 Énumération de motifs

Énumérer les motifs, ou graphlets [20], de taille  $k$  d'un graphe revient à énumérer les sous-graphes induits de l'ensemble de ses combinaisons de  $k$  sommets connectés. Cette méthode offre un outil riche et versatile, dont l'efficacité comme descripteur de réseaux a été démontrée dans différents contextes [10,33]. Depuis l'article qui l'a popularisée [20] en proposant qu'elle soit un outil pertinent pour l'analyse de tous types de réseaux, la méthode a effectivement été appliquée à de nombreux domaines. Dans un premier temps, en biologie [25,19] où elle s'est avérée très efficace dans le cadre de l'alignement de réseaux. Son utilisation s'est depuis diversifiée puisqu'elle est également employée en analyse des réseaux sociaux [8,6], des réseaux trophiques [2], etc. Une revue de littérature sur l'énumération des sous-graphes a récemment été proposée [27].

Afin d'être appliqués à des domaines aussi variés, les graphlets s'adaptent aux réseaux dans lesquels ils sont énumérés. La majorité des études s'appuient sur les graphlets simples, non-pondérés et non-colorés, qu'ils soient orientés ou non. Initialement, les triades de Holland et Leinhardt [13] sont ainsi les graphes orientés à trois sommets. Par la suite, les travaux en biologie ont massivement employés les graphlets non-orientés [25,24]. La majorité des méthodes d'énumération de motifs se sont ainsi dans un premier temps appliquées à énumérer les graphlets non-orientés [34,12]. Plusieurs modèles ont été proposés dans le cadre de l'analyse de graphes temporels avec des arêtes horodatées [15] ou de réseaux successifs [14], mais aussi graphes nivelés [26], hypergraphes [16], etc.

Concernant les graphlets pondérés, Rocha et Blondel effectuent des marches aléatoires dans des graphes en énumérant des motifs dont le poids des arêtes est le nombre de fois où elle a été empruntée par la marche [29]. Guo et al. énumèrent les sous-graphes uniquement selon des critères topologiques tout en maintenant une moyenne de la distribution des poids des arêtes de chaque graphlet [11]. Comme ici, des études utilisent déjà la discrétisation des poids des arêtes pour détecter les motifs de réseaux [17]. Des graphlets hétérogènes, aux arêtes et sommets étiquetés ont récemment été proposés dans le cas non-orienté [31].

Témoins privilégiés de l'évolution des graphlets, leurs algorithmes d'énumération ont été adaptés au fur et à mesure que ceux-ci se diversifiaient pour interpréter de nouveaux réseaux. La méthode du G-Trie, qui base son parcours des sous-graphes sur une structure adjacente de phylogénie des sous-graphes [28] a, par exemple, été dans un premier temps adaptée aux graphlets orientés [1] puis aux graphlets aux arêtes colorées [21]. Des chercheurs se sont par ailleurs focalisés sur l'étape de canonisation dans un contexte de graphes colorés mais non orientés [5] : cette étape, préalable à l'énumération, consiste à produire les formes canoniques de l'ensemble des motifs potentiels de taille  $k$ .

L'état de l'art nous fournit ainsi une méthode d'énumération de graphlets labellisés et orientés que nous utiliserons pour dénombrer les sous-structures des graphes d'évolution de communautés.

### 3 Méthode

Afin de procéder à l'analyse de la dynamique de communautés évolutives, notre méthode suppose que nous soyons en possession d'un graphes d'évolution de communautés tel que décrit en section 3.1, soit un graphe orienté et pondéré sans cycle, avec attribut sur les nœuds.

Après avoir présenté notre terminologie (3.1), nous proposons un processus générique à appliquer (3.2) et détaillons comment nous procédons à la coloration des arêtes de notre graphe de communautés (3.3) dans le cadre de l'expérience que nous présentons en section 4.

#### 3.1 Définitions

**Definition 1.** *Communauté ponctuelle.* Soit un ensemble d'individus ou d'objets à un instant donné  $t$ , on appelle communautés ponctuelles à l'instant  $t$  les regroupements de ces individus détectés par une méthode de partitionnement.

**Definition 2.** *Communauté évolutive.* On appelle communauté évolutive une séquence chronologique de communautés ponctuelles.

**Definition 3.** *Graphe d'évolution de communautés.* Un graphe d'évolution de communautés est un graphe  $G = (V, E, T)$  où  $V$  est l'ensemble des sommets,  $E$  est l'ensemble des arêtes et  $T$  est un ensemble d'instant. Un sommet est de la forme  $(v_t, \phi)$  où  $t \in T$ .  $v_t$  représente une communauté ponctuelle à l'instant  $t$  et  $\phi$  est le nombre de membres qu'elle contient. Une arête  $(u_t, v_{t+1}, \rho)$  est dirigée de  $u_t$  vers  $v_{t+1}$  et  $\rho$  est le nombre de membres communs à  $u_t$  et  $v_{t+1}$ .

Notons que puisque les arêtes d'un graphe d'évolution de communautés sont orientées de  $t$  vers  $t + 1$  pour tout  $t$ , ceux-ci ne contiennent pas de cycle : on peut ainsi déduire un ordre partiel sur les sommets à partir de la topologie du graphe. Par ailleurs, le nombre de membres d'une communauté, la somme de ses flux entrants et la somme de ses flux sortants sont trois valeurs éventuellement différentes puisque des individus ou objets peuvent disparaître de l'univers à chaque instant.

#### 3.2 Processus

Les étapes successives à réaliser pour obtenir une distribution de graphlets d'un graphe d'évolution de communautés sont les suivantes :

1. **Coloration des arêtes :** L'étape de coloration des arêtes permet de diminuer la taille de l'alphabet de graphlets afin de se positionner à un niveau d'information pertinent permettant d'analyser les relations entre graphlets. Il s'agit ici d'encoder différentes informations, la pondération des arêtes du graphes mais également les étiquettes des noeuds.
2. **Canonisation des graphlets :** Comme suggéré par les auteurs du gLab-Trie [21], nous recherchons la matrice d'adjacence d'ordre lexicographique maximal pour calculer un ordre canonique des sommets.

3. **Énumération des graphlets** : Nous appliquons finalement la méthode gLabTrie [21] au graphe d'évolution des communautés ainsi coloré. Ce type de graphes étant un sous-ensemble des graphes orientés et colorés, il rentre tout à fait dans le champs d'application de l'algorithme.

### 3.3 Coloration des arêtes

Pour traiter des graphlets dans le cadre d'un graphe d'évolution de communautés tout en maintenant une taille d'alphabet de graphlets raisonnable, nous proposons d'effectuer une étape préalable de coloration des arêtes.

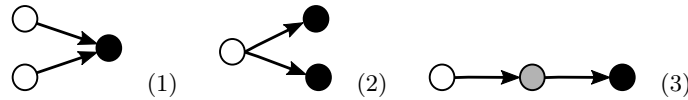
**Definition 4.** *Coloration.* Une fonction de coloration est une fonction

$$\begin{aligned} \text{coloration} : W &\longrightarrow L \\ w_e &\longmapsto l \end{aligned}$$

où  $L$  est un ensemble de labels discrets tel que  $|L| \ll |W|$ .

Outre l'objectif de diminuer la taille de l'alphabet de graphlets, il s'agit de se positionner à un niveau d'information pertinent permettant d'analyser les relations entre graphlets, leur ordre d'apparition, leurs co-occurrences, etc. L'étape de coloration est donc essentielle pour éviter de se trouver dans une situation où chaque sous-graphe du graphe d'évolution de communautés correspondrait à un graphlet unique, rendant impossible ainsi toute analyse mésoscopique.

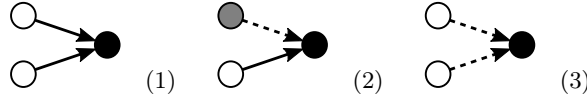
A titre d'illustration, la figure 1 présente les motifs de taille 3 des graphes d'évolution de communautés. Les couleurs des sommets décrivent les classes d'équivalence, ou positions, de chaque graphlet. Seule la structure est représentée. Il existe un graphlet représentant la fusion (11), la scission (12) ou la continuité (13).



**FIGURE 1.** La structure des trois motifs de taille 3 des graphes d'évolution de communautés, la fusion (1), la scission (2) et la continuité (3).

Cependant, le nombre de graphlets différents augmente avec l'ajout des labels aux arêtes. Si on se focalise, par exemple, sur le graphlet de fusion (figure 1(1)) dans le cas où il existe deux labels (flèche pleine ou en pointillés), nous aboutissons à 3 variantes (figure 2). Dans le cas du graphlet (figure 2(2)) les deux sommets représentant les communautés qui fusionnent ne sont plus équivalents : ce graphlet possède donc trois positions différentes.

Dans le cas des graphes d'évolution de communautés, une arête est pondérée par le nombre d'individus formant le flux qu'elle représente. Afin d'analyser ce



**FIGURE 2.** Coloration des positions pour différents graphlets de fusion

réseau, nous souhaitons que le rôle des arêtes vis à vis de leur source et de leur cible soit mis en lumière. Nous proposons alors de les colorer par un label défini par le couple formé par la proportion  $0 \leq \rho_{out} \leq 1$  de membres de la communauté source et la proportion  $0 \leq \rho_{in} \leq 1$  de membres de la communauté cible inclus dans le flux.

$$label(e) = (\rho_{out} = \frac{\rho}{\phi_u}, \rho_{in} = \frac{\rho}{\phi_v})$$

## 4 Illustration

### 4.1 Données

Nous basons nos expériences sur un jeu de données extraites d'une étude longitudinale de la plate-forme de vidéos en ligne YouTube. Dans le cadre du projet ANR PIL<sup>1</sup>, nous nous intéressons au Pluralisme de la Presse en Ligne. Nous analysons en particulier les pratiques informationnelles des internautes et notamment le phénomène des chambres d'écho [3]. Nous avons sélectionné 60 chaînes appartenant à des médias français (presse écrite, télévision, médias alternatifs) et collecté six ans de commentaires.

### 4.2 Résultats

Pour chaque semaine d'observation, un réseau d'interaction entre les commentateurs des différentes vidéos est construit, au sein duquel des communautés ponctuelles sont calculées (algorithme Louvain [4]). Nous obtenons finalement un graphe d'évolution de communautés à partir des séquences hebdomadaires de partitions, via un mécanisme d'appariement de communautés ponctuelles similaires (distance de Jaccard) entre pas de temps successifs (cf. catégorie de méthodes *Instant Optimal Communities Discovery/Iterative Similarity-Based Approaches* dans la taxonomie de Rossetti et Cazabet [30]).

Nous obtenons ainsi un graphe d'évolution de communautés de 77 541 sommets, soit autant de communautés ponctuelles réparties sur 315 pas de temps. La taille de ces communautés ponctuelles est variable et suit la classique distribution en loi de puissance puisqu'elles regroupent une moyenne de 15 membres mais que la moitié d'entre elles n'en comptent que trois. Le réseau contient par ailleurs 92 468 flux de membres entre ces communautés.

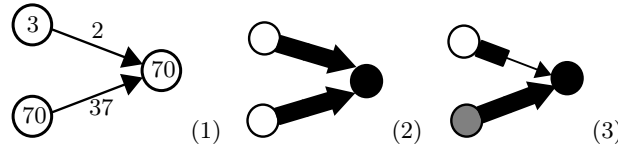
1. [www.anr-pil.org](http://www.anr-pil.org)



Nous cherchons dans cette première expérience à démontrer l'intérêt croisé de la double pondération des arêtes et de l'énumération des graphlets du graphe d'évolution de communauté ainsi construit. Pour cela, nous nous plaçons dans un cadre simple et contraint : chaque poids  $\rho_{out}$  et  $\rho_{in}$  sont discrétisés selon qu'ils soient strictement inférieurs ou bien supérieurs ou égaux à 0,5. De plus, nous ne nous appuyons ici que sur le graphlet de taille 3 représentant le fait que deux communautés ponctuelles fournissent des membres à une troisième, tel que présenté en Figure 2.

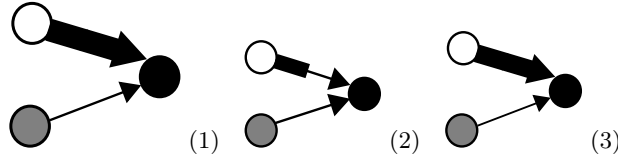
Dans le cas d'une coloration selon un unique seuil d'une double pondération, nous obtenons quatre labels et 10 graphlets différents représentant le regroupement de membres de communautés ponctuelles différentes. Ces 10 graphlets permettent de décrire chaque communauté ponctuelle du réseau selon son nombre d'apparitions. Le cas de la simple pondération est moins riche puisqu'il ne produit que deux labels et trois graphlets de même topologie. Dans les deux cas, l'algorithme trouve 832 136 graphlets dans le réseau.

Une illustration pertinente de la supériorité de la double pondération est le cas présenté en Figure 3. Dans cet exemple, une communauté ponctuelle de 70 membres et une de trois membres fournissent des membres à une communauté ponctuelle de 70 membres également. Dans les deux cas, plus de 50% des membres de chaque communauté ponctuelle transitent vers la communauté ponctuelle du pas de temps suivant. Dans cet exemple, le graphlet à simple pondération ne permet pas de distinguer l'importance de chacune des deux communautés ponctuelles sources tandis que le graphlet à double pondération marque effectivement l'importance plus grande du sommet du bas.



**FIGURE 3.** (1) Une situation du jeu de données YouTube. Les valeurs représentent respectivement le nombre de membres des sommets et la taille des flux. (2) le graphlet à simple pondération énuméré dans cette situation. (3) le graphlet à double pondération énuméré.

En fait, pour chaque sous-graphe énuméré nous pouvons comparer le graphlet détecté dans le cas de la simple pondération avec ceux dans le cas double. Cette approche généralisée au réseau YouTube nous fournit les valeurs présentées en Table 1. Nous pouvons ainsi décrire 10 scénarios possibles impliquant le regroupement de membres de plusieurs communautés ponctuelles. On note que de manière générale, tous les motifs où les flux représentent des faibles proportions de membres par rapport à leurs communautés ponctuelles de départ ou d'arrivée sont d'assez loin les plus courants.



**FIGURE 4.** 1) Un graphlet à pondération simple. 2) et 3) Deux de ses variantes à pondération double.

À l'inverse, parmi les motifs qui possèdent une arête dont les deux valeurs de pondération sont supérieures à 50%, le motif représenté en Figure 4(3) est le plus courant. Les communautés ponctuelles qui apparaissent dans la position blanche et celles qui apparaissent dans la position noire de ce motif ont un lien privilégié puisque elles ont plus de 50% de membres en commun. Or, cette observation n'aurait pas été possible avec le motif à simple pondération de la Figure 4(1). En effet, il indique simplement qu'une part importante des membres du sommet blanc se retrouvent dans le sommet noir, sans précisions sur l'importance qu'ils représentent pour ce dernier : il se peut que l'apport du sommet blanc soit négligeable ou au contraire reflète la majorité du sommet noir. De fait, dans notre jeu de données, c'est le cas d'un apport minoritaire qui se trouve être le plus fréquent (cf. le graphlet en Figure 4(2) énuméré 150K fois).

669 508			152 059			10 569			
668K	1.6K	206	150K	200	1K	58	10K	212	6

**TABLE 1.** Première ligne : les graphlets simplement pondérés, avec leur nombre d'apparition en deuxième ligne. Les quatrième et troisième lignes correspondent aux graphlets doublement pondérés issus de chaque graphlet simplement pondéré et leur nombre d'apparitions.

Ces quelques remarques démontrent, même dans un cas contraint, l'intérêt en terme descriptif de la double pondération des arêtes des graphes d'évolution de communautés. Les graphlets à simple pondération ne permettent ainsi généralement pas de discerner toutes les situations remarquables et potentiellement importantes de la vie des communautés évolutives.

Les graphlets permettent d'étudier les différentes étapes qui jalonnent l'existence des communautés évolutives. Il est par exemple possible de restreindre l'énumération des graphlets au voisinage des communautés ponctuelles qui leur appartiennent afin de les comparer entre elles. À chaque pas de temps, la com-

munauté ponctuelle représentant une communauté évolutive apparaît dans un ensemble de positions et il serait très utile d’observer la séquence de ces ensembles de positions par lesquelles passe une communauté évolutive au cours de son existence.

## 5 Conclusion

Dans ces travaux, nous proposons une approche permettant une description fine de l’évolution de communautés. De nombreux travaux se sont intéressés à la manière d’extraire des communautés sur un graphe temporel et la capacité de pouvoir comparer et interpréter les résultats issus d’algorithmes de clustering temporel reste un enjeu important.

Nous proposons une méthode de caractérisation de l’évolution de communautés qui se situe à un niveau mésoscopique. Nous décrivons les communautés, non pas par l’agrégation des propriétés de leurs membres, non pas par des métriques sur le graphe global (approches cependant complémentaires et utiles) mais par une énumération des graphlets permettant ainsi de décrire un réseau par les différentes formes de cycle de vie des structures intermédiaires que sont les communautés évolutives.

Dans un scénario-type, basé sur un jeu de données issu de commentaires échangés sur YouTube, nous montrons comment cette méthode peut se mettre en oeuvre. Afin de gérer la complexité engendrée par un graphe pondéré en terme de nombre de graphlets possibles, et de préserver la capacité pour un expert-métier d’interpréter les résultats obtenus, nous proposons une méthode de coloration des arêtes originale transformant le poids des arêtes du graphe de communautés (le flux d’individus circulant d’une communauté ponctuelle à une autre) et les étiquettes sur les noeuds (le nombre de membres des communautés) en un label qui encode une double pondération transformant ainsi le flux absolu en proportion de la taille de la communauté de départ (et respectivement de la communauté d’arrivée).

En perspectives futures, nous prévoyons de mener des tests de montée en charge en augmentant la taille  $k$  des graphlets. Concernant les graphes de communautés évolutives, ils sont des cas particuliers de graphes pondérés et orientés, et sont en pratique peu denses. Il nous faut tout d’abord examiner la nature de tels graphes générés à partir de différents types graphes temporels (interaction, amitiés, co-publication, etc.), de différentes méthodes de détection de communautés dynamiques ; ensuite, nous chercherons à optimiser la méthode d’énumération de graphlets pour exploiter les propriétés de nos graphes spécifiques.

Enfin, dans le cadre applicatif de l’étude sociologique de la consommation de presse en ligne, notre approche permettra de décrire l’évolution structurelle des communautés dites « chambres d’écho ». Grâce au pouvoir descriptif des graphlets, nous pourrions ainsi décrire de manière fine le cycle de vie de ces communautés particulières, polarisées sur un seul sujet d’intérêt ne laissant pas la place au débat et aux brassage d’idées.

## Références

1. David Aparicio, Pedro Ribeiro, and Fernando Silva. Extending the Applicability of Graphlets to Directed Networks. *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, 14(6) :1302–1315, 2017.
2. Jordi Bascompte and Carlos J Melián. Simple trophic modules for complex food webs. *Ecology*, 86(11) :2868–2873, 2005.
3. Alessandro Bessi, Fabiana Zollo, Michela Del Vicario, Michelangelo Puliga, Antonio Scala, Guido Caldarelli, Brian Uzzi, and Walter Quattrociocchi. Users polarization on facebook and youtube. *PloS one*, 11(8), 2016.
4. Vincent D Blondel, Jean-Loup Guillaume, Renaud Lambiotte, and Etienne LeFebvre. Fast unfolding of communities in large networks. *Journal of statistical mechanics : theory and experiment*, 2008(10) :P10008, 2008.
5. Nicolas Bloyet, Pierre-François Marteau, and Emmanuel Frenod. Scott : A method for representing graphs as rooted trees for graph canonization. In *International Conference on Complex Networks and Their Applications*, pages 578–590. Springer, 2019.
6. Raphaël Charbey and Christophe Prieur. Stars, holes, or paths across your facebook friends : A graphlet-based characterization of many networks. *Network Science*, pages 1–22, 2019.
7. Hocine Cherifi, Gergely Palla, Boleslaw K Szymanski, and Xiaoyan Lu. On community structure in complex networks : challenges and opportunities. *Applied Network Science*, 4(1) :1–35, 2019.
8. Alexandra Duma and Alexandru Topirceanu. A network motif based approach for classifying online social networks. In *2014 IEEE 9th IEEE International Symposium on Applied Computational Intelligence and Informatics (SACI)*, pages 311–315. IEEE, 2014.
9. Mark Goldberg, Malik Magdon-Ismaïl, Srinivas Nambirajan, and James Thompson. Tracking and predicting evolution of social communities. In *2011 IEEE Third International Conference on Privacy, Security, Risk and Trust and 2011 IEEE Third International Conference on Social Computing*, pages 780–783. IEEE, 2011.
10. Shawn Gu and Tijana Milenkovic. Graphlets versus node2vec and struc2vec in the task of network alignment. pages 1–11, 2018.
11. Hongyu Guo, Khalique Newaz, Scott Emrich, Tijana Milenkovic, and Jun Li. Weighted graphlets and deep neural networks for protein structure classification, 2019.
12. Tomaž Hočevar and Janez Demšar. A combinatorial approach to graphlet counting. *Bioinformatics*, 30(4) :559–565, 2014.
13. Paul W Holland and Samuel Leinhardt. Local structure in social networks. *Sociological methodology*, 7 :1–45, 1976.
14. Yuriy Hulovatyy, Huili Chen, and Tijana Milenković. Exploring the structure and function of temporal networks with dynamic graphlets. *Bioinformatics*, 31(12) :i171–i180, 2015.
15. Lauri Kovanen, Márton Karsai, Kimmo Kaski, János Kertész, and Jari Saramäki. Temporal motifs in time-dependent networks. *Journal of Statistical Mechanics : Theory and Experiment*, 2011(11) :P11005, 2011.

16. Geon Lee, Jihoon Ko, and Kijung Shin. Hypergraph motifs : Concepts, algorithms, and discoveries. *arXiv preprint arXiv :2003.01853*, 2020.
17. Jiandun Li, Dingyu Yang, and Chunlei Ji. Mine weighted network motifs via bayes' theorem. In *2017 4th International Conference on Systems and Informatics (ICSAI)*, pages 448–452. IEEE, 2017.
18. Yu-Ru Lin, Yun Chi, Shenghuo Zhu, Hari Sundaram, and Belle L Tseng. Facetnet : a framework for analyzing communities and their evolutions in dynamic networks. In *Proceedings of the 17th international conference on World Wide Web*, pages 685–694, 2008.
19. Tijana Milenković and Nataša Pržulj. Uncovering biological network function via graphlet degree signatures. *Cancer informatics*, 6 :CIN–S680, 2008.
20. Ron Milo, Shai Shen-Orr, Shalev Itzkovitz, Nadav Kashtan, Dmitri Chklovskii, and Uri Alon. Network motifs : simple building blocks of complex networks. *Science*, 298(5594) :824–827, 2002.
21. Misael Mongioví, Giovanni Micale, Alfredo Ferro, Rosalba Giugno, Alfredo Pulvirenti, and Dennis Shasha. glabtrie : A data structure for motif discovery with constraints. *Graph Data Management*, page 71, 2018.
22. Tsuyoshi Murata and Sakiko Moriyasu. Link prediction of social networks based on weighted proximity measures. In *Proceedings of the IEEE/WIC/ACM international conference on web intelligence*, pages 85–88. IEEE Computer Society, 2007.
23. Gergely Palla, Albert-László Barabási, and Tamás Vicsek. Quantifying social group evolution. *Nature*, 2007.
24. Nataša Pržulj. Biological network comparison using graphlet degree distribution. *Bioinformatics*, 23(2) :e177–e183, 2007.
25. Nataša Pržulj, Derek G Corneil, and Igor Jurisica. Modeling interactome : scale-free or geometric ? *Bioinformatics*, 20(18) :3508–3515, 2004.
26. Yuanfang Ren, Aisharjya Sarkar, Ahmet Ay, Alin Dobra, and Tamer Kahveci. Finding conserved patterns in multilayer networks. In *Proceedings of the 10th ACM International Conference on Bioinformatics, Computational Biology and Health Informatics*, pages 97–102, 2019.
27. Pedro Ribeiro, Pedro Paredes, Miguel EP Silva, David Aparicio, and Fernando Silva. A survey on subgraph counting : Concepts, algorithms and applications to network motifs and graphlets. *arXiv preprint arXiv :1910.13011*, 2019.
28. Pedro Ribeiro and Fernando Silva. G-tries : an efficient data structure for discovering network motifs. In *Proceedings of the 2010 ACM symposium on applied computing*, pages 1559–1566, 2010.
29. Luis EC Rocha and Vincent D Blondel. Flow motifs reveal limitations of the static framework to represent human interactions. *Physical Review E*, 87(4) :042814, 2013.
30. Giulio Rossetti and Rémy Cazabet. Community discovery in dynamic networks : A survey. *ACM Comput. Surv.*, 51(2), February 2018.
31. Ryan A Rossi, Nesreen K Ahmed, Aldo Carranza, David Arbour, Anup Rao, Sungchul Kim, and Eunye Koh. Heterogeneous network motifs. *arXiv preprint arXiv :1901.10026*, 2019.
32. Martin Rosvall and Carl T Bergstrom. Mapping change in large networks. *PloS one*, 5(1), 2010.

33. Mattia Tantardini, Francesca Ieva, Lucia Tajoli, and Carlo Piccardi. Comparing methods for comparing networks. *Scientific reports*, 9(1) :1–19, 2019.
34. Sebastian Wernicke and Florian Rasche. Fanmod : a tool for fast network motif detection. *Bioinformatics*, 22(9) :1152–1153, 2006.