

Suivi de l'évolution de Clusters de Liens dans des Réseaux Sociaux Dynamiques

Erick Stattner*, Martine Collard*

* Laboratoire LAMIA
Université des Antilles
FRANCE

{erick.stattner, martine.collard}@univ-antilles.fr

Résumé. De nombreuses méthodes ont été proposées pour extraire des clusters des réseaux sociaux. Si un travail important est aujourd'hui mené sur la conception de méthodes innovantes capables de rechercher des clusters de nature différente, la plupart des approches font l'hypothèse de réseaux statiques. L'une des récentes méthodes concerne notamment la recherche de liens conceptuels. Il s'agit d'une nouvelle approche de clustering de liens, qui exploite à la fois la structure du réseau et les attributs des noeuds dans le but d'identifier des liens fréquents entre des groupes de noeuds au sein desquels les noeuds partagent des attributs communs. Dans ce travail, nous nous intéressons au suivi des liens conceptuels dans des réseaux dynamiques, c'est-à-dire des réseaux qui connaissent des changements structurels importants. Nous cherchons en particulier à comprendre comment les liens conceptuels se forment et évoluent au cours du développement du réseau. Pour ce faire, nous proposons un ensemble de mesures qui visent à capturer des comportements caractérisant l'évolution de ces clusters. Notre approche est ainsi utilisée pour comprendre l'évolution des liens conceptuels extraits sur deux réseaux réels : un réseau de co-auteurs d'articles scientifiques et un réseau de communications mobiles. Les résultats obtenus permettent de mettre en lumière des tendances significatives dans l'évolution des clusters sur ces deux réseaux.

1 Introduction

Le domaine de l'analyse et l'extraction de connaissances à partir de réseaux sociaux est devenu un des axes de recherche très actif du 21^e siècle (Barbier et Liu, 2011). Si les méthodes fondatrices ont d'abord tiré partie des mesures issues de la théorie des graphes pour analyser ces structures, les approches récentes, aujourd'hui regroupées dans le domaine la "*fouille de réseaux sociaux*" (ou plus simplement "*fouille de liens*") appliquent les principes de la fouille de données traditionnelle aux réseaux (Philip et al., 2010).

L'extraction de clusters est en particulier une des tâches de la modélisation descriptive les plus courantes du domaine de la fouille de réseaux sociaux (Stattner et Collard, 2015). En effet dans de nombreux systèmes, qu'ils soient naturels ou sociaux, les entités impliquées ont

souvent tendance à s'organiser en groupes. Identifier ces groupes s'avère alors être un défi important pour comprendre les structures émergeant des interactions entre les entités, étudier les mécanismes qui y prennent place et déterminer le rôle des agents au sein de ces systèmes.

C'est ainsi que de nombreuses méthodes de clustering de réseaux ont été proposées ces dernières années (Fortunato, 2010). Les approches qui ont vu le jour ont d'ailleurs permis une évolution de la notion même de cluster. En effet, alors que les premières contributions exploitaient uniquement la structure des réseaux pour en extraire des groupes de noeuds densément connectés, aussi appelés *communautés* (Newman et Girvan, 2004), les approches récentes se sont intéressées à l'extraction de clusters de plus en plus complexes définis à la fois par leur structure et par les attributs des noeuds qui les composent. L'une des récentes méthodes de clustering de réseaux concerne notamment la recherche de liens conceptuels. Il s'agit d'une nouvelle approche de clustering de liens, qui exploite à la fois la structure du réseau et les attributs des noeuds, dans le but d'identifier des liens fréquents entre des groupes au sein desquels les noeuds partagent des attributs communs.

Pourtant, si un travail important est actuellement mené sur la conception de méthodes innovantes, capables de rechercher des clusters de nature différente, la plupart de ces approches font l'hypothèse de réseaux statiques. Or, nous savons que la dimension temporelle fait partie inhérente de ces structures (Toivonen et al., 2009). Les réseaux sont en effet des structures vivantes au sein desquelles des noeuds et des liens peuvent apparaître ou disparaître. Certains réseaux, tels que les réseaux de communications (Calabrese et al., 2011) ou les réseaux de proximité géographique (Stehle et al., 2011), connaissent des changements structurels assez importants au cours de leur évolution.

Dans ce travail, nous nous intéressons au suivi des liens conceptuels dans des réseaux dynamiques. Nous cherchons en particulier à comprendre comment les liens conceptuels se forment et évoluent au cours du développement du réseau. Pour ce faire, nous proposons d'abord un ensemble de mesures qui visent à capturer des comportements caractérisant l'évolution de ces clusters, c'est-à-dire certaines transitions qui prennent place au sein de ces clusters. Notre approche est ensuite utilisée pour étudier l'évolution des liens conceptuels extraits sur deux réseaux réels : un réseau de co-auteurs d'articles scientifiques et un réseau de communications mobiles.

Cet article est organisé comme suit. La Section 2 passe en revue les principales méthodes de modélisation descriptive des réseaux et présente les travaux menés sur la dynamique des clusters. La Section 3 décrit formellement la notion de liens conceptuels. La Section 4 détaille l'approche que nous proposons pour suivre l'évolution des liens conceptuels. La Section 5 est consacrée aux résultats obtenus sur deux jeux de données. La Section 6 conclut le papier et présente nos travaux futurs.

2 Travaux antérieurs

Le clustering de réseau fait référence à une famille de méthodes qui vise à extraire des groupes à partir de structures de réseau. Si les approches pionnières ont dans un premier temps exploité uniquement la structure des réseaux, les approches récentes ont tenté d'inclure toutes les informations disponibles sur la structure et les propriétés des noeuds lors de la phase d'extraction. Nous pouvons ainsi distinguer 3 principales approches, qui s'intéressent chacune à un type bien spécifique de cluster.

Le clustering de réseau traditionnel, aussi appelé “recherche de communautés”, a pour objectif de partitionner le réseau en plusieurs composantes (appelées groupes, clusters ou communautés) de sorte que les noeuds au sein de chaque communauté possèdent une forte densité de connexions internes (Fortunato, 2010). L’objectif de ces méthodes est d’identifier des groupes qui maximisent un critère de distance, le plus connu étant la modularité (Newman et Girvan, 2004). Deux types de méthodes sont généralement distinguées : (i) les approches ascendantes, qui fusionnent, à chaque itération, des noeuds ou des groupes de noeuds (Shen et al., 2009). (ii) les approches descendantes, qui partent du réseau complet et le divisent jusqu’à obtenir un partitionnement suffisant (Rattigan et al., 2007).

Le clustering hybride fait référence à de nouvelles approches de clustering de noeuds qui tentent de prendre en compte la structure du réseau et les attributs des noeuds lors du processus de partitionnement (Yang et al., 2013). Ces approches redéfinissent la notion de cluster en s’intéressant à des “groupes de noeuds densément connectés au sein desquels les noeuds partagent des attributs communs” (Zhou et al., 2009). Ces méthodes s’appuient sur de nouvelles mesures de pertinence des clusters qui visent à trouver un équilibre entre la similarité de la structure et la similarité des noeuds. Les clusters mis ainsi en évidence ont une sémantique beaucoup plus forte et adaptée au contexte.

L’extraction de liens conceptuels est une très récente approche de clustering qui cherche, elle, à identifier des clusters de liens (Stattner et Collard, 2015). Elle exploite à la fois les informations disponibles sur la structure du réseau et les attributs des noeuds dans l’objectif d’extraire les groupes de noeuds les plus connectés du réseau, au sein desquels les noeuds partagent des caractéristiques communes. L’extraction des liens conceptuels implique deux étapes (Stattner et Collard, 2012) : (i) une phase de clustering des noeuds, qui regroupe les noeuds sur la base du partage d’attributs communs et (ii) une phase d’identification des liens conceptuels, qui évalue la fréquence des liens entre les clusters identifiés en phase (i).

Bien que toutes ces méthodes soient aujourd’hui largement utilisées pour identifier des clusters des réseaux sociaux, elles opèrent sur des structures supposées statiques. Pourtant, la plupart des réseaux du monde réel connaissent des évolutions structurelles au cours du temps. C’est la raison pour laquelle des travaux sont aujourd’hui menés sur la façon dont un cluster évolue avec le réseau. Une approche naturelle consiste à considérer le réseau comme possédant plusieurs états, selon le moment d’observation, puis à appliquer l’algorithme de clustering sur chacun des états. De cette façon, l’évolution des clusters est étudiée en s’intéressant à 5 grands comportements (Oliveira et Gama, 2010; Greene et al., 2010; Aynaoud et al., 2013) : apparition du cluster, sa disparition, sa survie, la fusion de plusieurs clusters, la division d’un cluster en différents clusters.

Dans ce travail, nous cherchons à comprendre comment évoluent les liens conceptuels au cours du développement du réseau. A notre connaissance, l’étude présentée dans ce papier est la première contribution qui s’intéresse à l’évolution de ce type de clusters sur des réseaux dynamiques.

3 Clustering de liens à travers les liens conceptuels

Soit $G = (V, E)$ un réseau, dans lequel V est l’ensemble des noeuds et E l’ensemble des liens avec $E \subseteq V \times V$. L’ensemble V est défini comme une relation $R(A_1, \dots, A_p)$ où

Clusters de Liens dans des Réseaux Sociaux Dynamiques

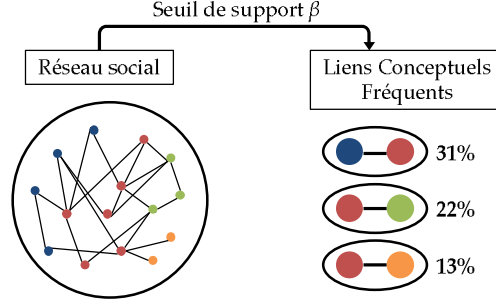


FIG. 1 – Extraction des liens conceptuels à partir d'un réseau social

chaque A_i est un attribut. Ainsi, chaque noeud $v \in V$ est défini par le tuple (a_1, \dots, a_p) où $\forall k \in [1..p], v[A_k] = a_k$ correspond à la valeur de l'attribut A_k du noeud v .

Un item est une expression logique $A = x$ où A est un attribut et x une valeur. L'item vide est noté \emptyset . Un itemset est une conjonction d'items, par exemple $(A_1 = x \text{ et } A_2 = y \text{ et } A_3 = z)$. Quand un itemset est une conjonction k items non-vides, on parle de k -itemset. Soit m un itemset, nous notons V_m , l'ensemble des noeuds de V qui satisfont m .

Posons m_1 et m_2 deux itemsets et V_{m_1}, V_{m_2} , respectivement les ensembles de noeuds dans V qui satisfont m_1 et m_2 . Nous notons $E_{(m_1, m_2)}$ le cluster de liens connectant des noeuds de V_{m_1} à des noeuds de V_{m_2} , i.e.

$$E_{(m_1, m_2)} = \{e \in E; e = (v_1, v_2) \quad v_1 \in V_{m_1} \text{ et } v_2 \in V_{m_2}\} \quad (1)$$

Le cluster $E_{(m_1, m_2)}$ est appelé "lien conceptuel" dans la mesure où il ne s'agit pas d'un lien du réseau, mais d'un cluster de liens entre deux groupes de noeuds qui peuvent être vus comme des "concepts", au sens de l'analyse de concepts formels (Ganter et al., 2005), c'est-à-dire des objets qui partagent des attributs communs.

Le support du lien conceptuel $E_{(m_1, m_2)}$, est le pourcentage de liens appartenant à $E_{(m_1, m_2)}$, i.e. $supp(E_{(m_1, m_2)}) = \frac{|E_{(m_1, m_2)}|}{|E|}$.

Le lien conceptuel entre les itemsets m_1 et m_2 est fréquent si le support de $E_{(m_1, m_2)}$ est plus grand qu'un seuil de support minimum β , i.e. $supp(E_{(m_1, m_2)}) > \beta$. Par simplicité nous notons les liens conceptuels fréquents par les deux itemsets associés (m_1, m_2) .

Soit I l'ensemble des itemsets construits à partir de V , nous définissons LC comme l'ensemble des clusters de liens extraits de G .

$$LC = \bigcup_{m_1 \in I, m_2 \in I} \{E_{(m_1, m_2)}; supp(E_{(m_1, m_2)}) > \beta\} \quad (2)$$

La Figure 1 montre un exemple de liens conceptuels extraits à partir d'un réseau social. Nous pouvons, par exemple, y observer que 31% des liens du réseau connectent des noeuds qui vérifient la propriété "bleu" à des noeuds qui vérifient la propriété "rouge".

4 Caractériser la dynamique des clusters

Notre objectif dans ce papier est d'étudier l'évolution des clusters de liens durant le développement du réseau. Comme l'illustre la Figure 2, il s'agit d'identifier entre deux états du réseau les transitions qui s'opèrent sur les liens conceptuels, c'est-à-dire si les clusters se maintiennent, se divisent, fusionnent, apparaissent ou disparaissent entre les instants t et $(t + 1)$.

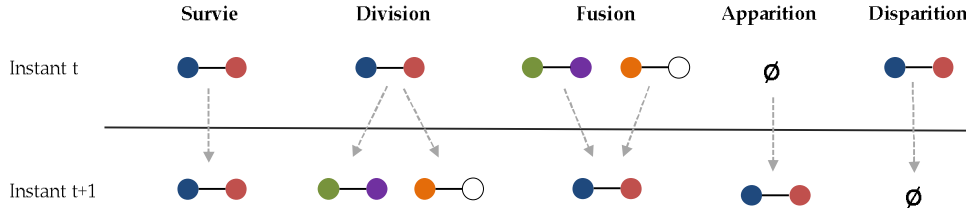


FIG. 2 – Comportement d'évolution possibles des clusters

Ainsi, nous adaptons les règles utilisées pour suivre les communautés dans les réseaux sociaux aux liens conceptuels (Aynaoud et al., 2013). Plus précisément, soient $L = E(m_1, m_2)$ et $L' = E(m'_1, m'_2)$ deux clusters de liens extraits à des états successifs du réseau. Nous définissons la *similarité* entre deux clusters comme étant la quantité de liens partagés entre ces deux ensembles.

$$\text{sim}(L, L') = \min\left(\frac{|L \cap L'|}{|L|}, \frac{|L \cap L'|}{|L'}\right) \quad (3)$$

Cette mesure, comprise entre 0 et 1, fournit une information sur la taille des intersections entre L et L' . Une valeur de 0 indique que les ensembles sont disjoints, alors qu'une valeur proche de 1 indique que les clusters sont très proches ; la valeur de 1 étant atteinte pour des ensembles strictement identiques.

A partir de cette notion de similarité, nous définissons un certain nombre de comportements qui caractérisent les changements qui peuvent prendre place sur un lien conceptuel entre l'état G_t du réseau et son état G_{t+1} . Plus précisément, soit L un cluster de liens extrait de G_t , nous introduisons $\text{match}(L)$ comme l'ensemble des clusters de liens L' dans G_{t+1} dont l'intersection avec L dépasse un certain seuil, c'est-à-dire dont la similitude est supérieure au seuil fixé. Dans nos expériences nous utilisons un seuil de 0.75. S'il n'existe pas dans G_{t+1} un tel cluster, nous avons $\text{match}(L) = \emptyset$. De cette façon, nous utilisons un ensemble de règles pour identifier chacun des comportements d'évolution présentés sur la Figure 2.

- **Fusion** : L dans G_t fusionne avec d'autres liens conceptuels pour former L' dans G_{t+1} si $L' \in \text{match}(L)$ et $\exists Z \neq L$ dans G_t tel que $L' \in \text{match}(Z)$.
- **Division** : L dans G_t se divise en plusieurs liens conceptuels L'_1, L'_2, \dots, L'_k dans G_{t+1} si $\forall i, L'_i \in \text{match}(L)$.
- **Survie** : L dans G_t devient L' dans G_{t+1} si $L' \in \text{match}(L)$ et $\forall Z \neq L$ dans G_t , $L' \notin \text{match}(Z)$.
- **Disparition** : L dans G_t disparaît si aucun des cas précédents ne survient.
- **Apparition** : L' dans G_{t+1} apparaît si $\forall L$ dans G_t , $L' \notin \text{match}(L)$.

Précisons que les événements de fusion et de division ne sont pas disjoints. Une partie d'un cluster peut intervenir dans une fusion alors qu'une autre sera impliquée dans une division.

5 Résultats expérimentaux

Nous avons appliqué notre approche à deux réseaux dynamiques de nature différente pour comprendre comment évoluent les liens conceptuels au cours du développement de ces réseaux et selon différents seuils de support β d'extraction des clusters. Nous détaillons les principales propriétés de ces jeux de données dans la section 5.1 et présentons les résultats obtenus dans la section 5.2.

5.1 Environnement de tests

Deux jeux de données ont été utilisés pour nos expériences. (i) Un réseau de co-auteurs d'articles scientifiques, qui représente l'ensemble des auteurs ayant co-écrit des articles à la conférence EGC¹ de 2004 à 2015. (ii) Un réseau de communications mobiles qui représente les appels téléphoniques effectués par les abonnés d'un opérateur de téléphonie mobile local sur une journée de 5h à 15h.

Ces réseaux sont intéressants pour notre étude puisqu'ils connaissent, tous deux, des évolutions assez significatives de leur structure. Le réseau de co-auteur passe par exemple de 139 à 1349 noeuds de 2004 à 2015 et les liens de 167 à 2366 sur cette même période. Le réseau mobile passe, lui, de 6786 noeuds à 246253 noeuds et de 3799 liens à 255947 sur les dix heures d'étude. Plus généralement, la Figure 3 montre comment évoluent les principales propriétés structurelles de ces réseaux au cours de leur développement : nombre de noeuds et de liens, densité, degré moyen, et distribution des degrés. Précisons que ces réseaux ne font que croître durant leur développement, puisque seuls des ajouts de noeuds ou de liens sont effectués. De plus, nous observons sur la distribution des degrés que le réseau de communications maintient une structure scale-free sur toute la durée de l'étude.

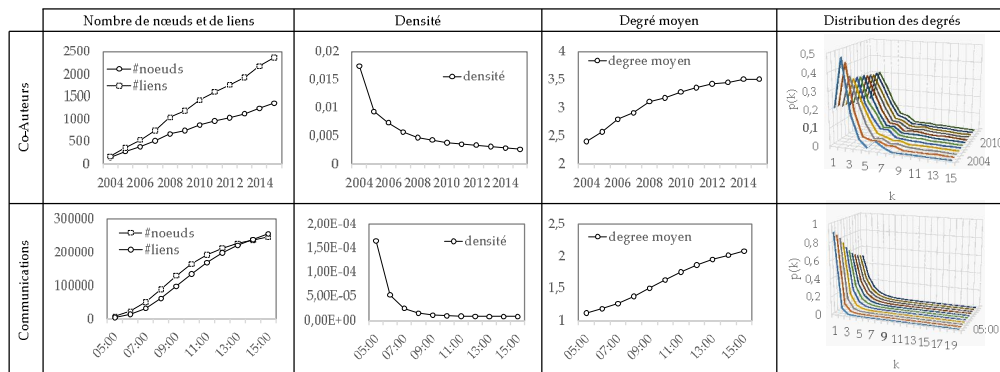


FIG. 3 – Propriétés des jeux de données utilisées

Dans le réseau de co-auteur, chaque noeud est identifié par 7 attributs : (1) label, (2) nombre d'articles publiés, (3) nombre moyen de co-auteurs par papier, (4) taille moyenne des résumés, (5) nombre moyen d'articles par an, (6) position moyenne dans la liste d'auteurs, (7) nombre de fois qu'il apparaît premier auteur.

1. Téléchargeable ici : <http://www.erickstattner.com/datasets/CollaborationNetwork.zip>

Au sein du réseau de communications, chaque noeud est identifié par 10 attributs : (1) numéro, (2) localisation (Martinique, Guadeloupe ou Guyane), (3) tranche horaire sur laquelle il est le plus actif, (4) type de forfait, (5) nombre moyen d'appels passés, (6) durée moyenne des appels passés, (7) nombre moyen d'appels reçus, (8) durée moyenne des appels reçus, (9) nombre de sms envoyés et (10) nombre de sms recus.

Sur les deux jeux de données, les attributs ont été pré-traités de façon à les discrétiser sur cinq classes. Nous avons extrait les liens conceptuels des deux réseaux en utilisant l'algorithme MFCL-Min (Stattner et Collard, 2012) avec 3 seuils de support différents : $\beta = 0.1$, $\beta = 0.2$ et $\beta = 0.3$. Nous ne détaillons pas ici la sémantique des clusters obtenus puisque notre objectif est d'étudier les changements qui prennent place au sein de ces clusters. La Figure 4 montre cependant le nombre de clusters identifiés à chaque instant pour (a) le réseau de co-auteurs et (b) le réseau de communications.



FIG. 4 – Nombre de clusters extraits sur : (a) réseau co-auteurs et (b) réseau communications

Nous observons que les tendances sont différentes pour les deux réseaux. Si le nombre de liens conceptuels extraits semble croître avec le temps sur le réseau de co-auteurs, et ce quel que soit le seuil utilisé, le nombre de clusters identifiés sur le réseau de communications semble se stabiliser après quelques heures d'accumulation des données. Ce résultat est particulièrement intéressant, puisque bien que le nombre de clusters se stabilise sur ce réseau, nous observerons dans la section suivante qu'ils ne sont pourtant pas les mêmes.

5.2 Résultats

Dans un premier temps, nous avons étudié quels étaient les changements qui survenaient sur les clusters. Plus précisément, pour chacun des réseaux, nous dénombrons entre l'état G_t et G_{t+1} du réseau, le pourcentage d'apparition de clusters, de disparition, de survie, de fusion et de division. La Figure 5 détaille les résultats obtenus.

Il est intéressant d'observer que pour les deux réseaux, durant les premiers instants de l'étude, 100% des clusters de liens extraits sont des apparitions. Cela peut s'expliquer par le fait qu'au début de l'étude, les clusters identifiés sont très volatiles du fait du peu de données accumulées qui ne reflètent finalement pas des tendances réelles. Ainsi, les clusters ne sont pas confirmés lors du passage à l'état suivant. Cette observation peut également être faite sur le taux de disparition qui est de 100% sur les premiers instants.

Clusters de Liens dans des Réseaux Sociaux Dynamiques

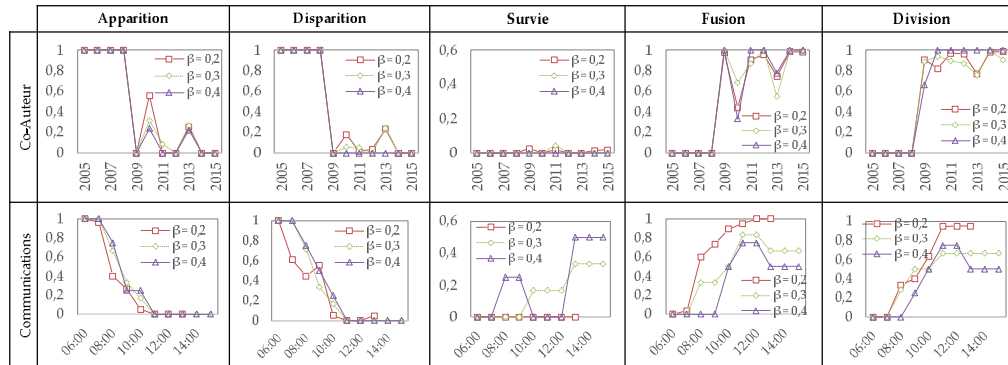


FIG. 5 – *Changements prenant place sur les clusters de liens*

Cependant, après un certain temps d'étude, nous observons une diminution du taux d'apparition et de disparition, ce qui laisse penser que les tendances mises en évidence dans les clusters se confirment d'un état à l'autre. Par exemple, à partir de l'instant $t_0 + 6$, le réseau de communications ne connaît plus aucune apparition quel que soit le seuil β utilisé. Ce taux s'abaisse également fortement pour le réseau de co-auteurs à partir de l'instant $t_0 + 4$.

Ces résultats sur l'apparition et la disparition des clusters laissent à penser que sur ces deux réseaux, après un certain temps d'accumulation des données, les tendances sous-jacentes aux clusters tendent à se confirmer d'un état du réseau à l'autre, ce qui conduit à une diminution des taux d'apparition et de disparition.

Pour comprendre ce que deviennent ces clusters qui semblent ne plus disparaître après un certain temps, intéressons nous dans un premier temps au taux de survie (cf. Figure 5). Les tendances sont cette fois différentes selon le jeu de données. Quel que soit le seuil de support utilisé, le réseau de co-auteurs voit une fraction négligeable de clusters survivre. Ce résultat suggère que ces clusters, qui pourtant ne disparaissent pas, sont présents sous d'autres formes à travers des comportements de fusion ou de division. En ce qui concerne le réseau de communications, nous observons en revanche que le taux de survie varie selon le seuil de support et semble croître avec le temps. Par exemple, à l'instant $t_0 + 8$, nous pouvons observer que 50% des clusters survivent avec un seuil $\beta = 0.4$, contre environ 30% avec un seuil $\beta = 0.3$.

Les résultats obtenus sur la survie des clusters sont intéressants, puisqu'ils montrent que, sur certains jeux de données, des clusters peuvent ne pas disparaître, mais ne pas survivre pour autant entre deux états successifs du réseau. Ce qui laisse penser qu'ils ne se maintiennent qu'à travers des comportements de fusion et de division.

Ainsi, si nous nous concentrons sur les comportements de fusion et de division (cf. Figure 5), les résultats obtenus confirment nos observations précédentes. En effet, sur le réseau de co-auteurs la grande majorité des changements sur les clusters s'avèrent être des fusions et des divisions de clusters (rappelons que ces deux événements ne sont pas disjoints). Sur le réseau de communications, le taux de fusion et de division reste également important, mais varie selon le seuil support β . Par exemple, à l'instant $t_0 + 8$, nous pouvons observer que 50%

des clusters sont impliqués dans une fusion avec un seuil $\beta = 0.4$, contre environ 66% avec un seuil $\beta = 0.3$.

Si les taux de fusion et de division peuvent paraître importants, il convient toutefois de les nuancer au regard du nombre de clusters extraits. Par exemple, dans le réseau de communications avec $\beta = 0.4$, nous observons que 50% des communautés se divisent, mais il s'agit de 50% des 4 clusters extraits (cf. Figure 4).

Les taux de fusion et division observés permettent de mettre en évidence des résultats assez surprenants. En effet, après un certain temps d'accumulation des données, si aucun nouveau cluster n'apparaît, les clusters existants semblent se maintenir pour l'essentiel d'entre eux à travers des comportements de fusion et de division. Ce résultat laisse penser qu'un certain temps d'accumulation est nécessaire pour identifier des tendances fortes, avant de n'observer que des brassages entre les clusters.

Les résultats obtenus ont permis de mettre en évidence que la grande majorité des comportements identifiés sur ces deux réseaux sont des comportements de fusion et de division. Ainsi dans un second temps, nous avons cherché à mieux caractériser ces événements en nous intéressant au pourcentage de clusters impliqués durant les phases de fusion et de division.

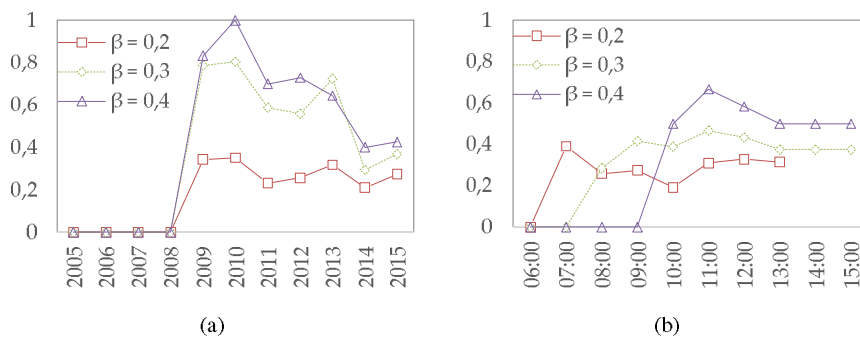


FIG. 6 – Moyenne du pourcentage de clusters impliqués dans une fusion : (a) réseau de co-auteurs, (b) réseau de communications

La Figure 6 montre, pour les deux réseaux, le pourcentage de clusters impliqués lors d'une phase de fusion. Si ce pourcentage varie selon le seuil de support β , nous observons qu'il reste relativement stable et élevé au cours du temps. Par exemple, pour le réseau de co-auteurs, avec un seuil de 0.2, nous observons qu'en moyenne une fusion implique 30% des clusters. Pour le réseau de communication, 50% des clusters sont en moyenne impliqués dans une fusion avec un seuil de $\beta = 0.4$, contre environ 40% pour avec un seuil de $\beta = 0.3$.

De la même façon, la Figure 7 montre le pourcentage de clusters impliqués lors d'une phase de division pour les deux jeux de données. Comme précédemment, nous pouvons observer que le taux de clusters impliqués dans une division reste relativement élevé. Par exemple, pour le réseau de communication, 50% des clusters sont en moyenne impliqués dans une fusion avec un seuil de $\beta = 0.4$, contre environ 40% avec un seuil de $\beta = 0.3$.

A travers ces résultats, nous pouvons observer que si les taux de fusion et de division sont importants (observés sur la Figure 5), ils impliquent également un fort pourcentage des clusters

Clusters de Liens dans des Réseaux Sociaux Dynamiques

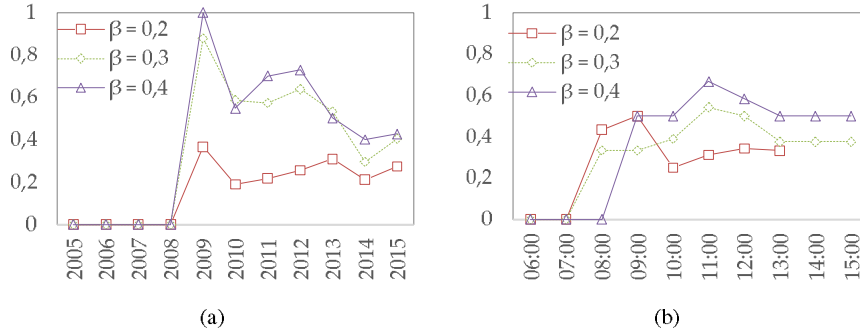


FIG. 7 – Moyenne du pourcentage de clusters impliqués dans une division : (a) réseau de co-auteurs, (b) réseau de communications

présents (cf. Figures 6 et 7). Ils complètent ainsi nos observations précédentes et confirment le fait que, sur ces deux réseaux, bien que le taux de survie des clusters ne soit pas important, les clusters continuent d'exister essentiellement à travers des comportements de fusion et de division impliquant une quantité importante des clusters présents. Il convient cependant de nuancer ces résultats au regard du faible nombre de clusters qui est obtenu selon le seuil de support β utilisé.

6 Conclusion et perspectives

Dans ce travail, nous nous sommes intéressés à la façon dont évoluent des clusters extraits sur des réseaux sociaux dynamiques. L'étude menée dans ce papier s'intéresse à un type bien spécifique de clusters : les liens conceptuels. Il s'agit d'une nouvelle approche de clustering de liens qui exploite à la fois la structure du réseau et les attributs des noeuds, dans le but d'identifier des liens fréquents entre des groupes de noeuds partageant des attributs communs.

Ainsi, nous avons dans un premier temps identifié un ensemble de comportements d'évolution susceptibles de prendre place sur les liens conceptuels : la survie du cluster, la fusion avec d'autres clusters, la division en plusieurs clusters, l'apparition et la disparition. Nous avons ensuite introduit la notion de similarité entre les liens conceptuels afin de décrire formellement ces comportements. Notre approche a été implémentée et testée sur deux réseaux sociaux connaissant des changements importants de leur structure au cours du temps : un réseau de co-auteurs d'articles scientifiques et un réseau de communications téléphoniques. Les résultats obtenus ont permis d'observer qu'un temps d'accumulation des données était nécessaire avant que le taux d'apparition de nouveaux clusters devienne négligeable. Nous avons également pu observer que très peu de comportement de survie de clusters sont identifiés et que la majeure partie des clusters se retrouvent d'un état à l'autre du réseau à travers des comportements de fusion et de division.

Si les résultats obtenus sont encourageants et permettent de mettre en lumière des tendances significatives dans l'évolution des liens conceptuels sur ces réseaux, il est important de pouvoir confirmer ces observations sur d'autres jeux de données. Ainsi, en perspectives nous souhaitons

appliquer notre approche à d'autres jeux de données dans l'espoir de confirmer les tendances ou d'en identifier de nouvelles. Il sera également important de comprendre comment évoluent les liens conceptuels selon la nature des réseaux étudiés. A long terme, la compréhension de l'évolution des liens conceptuels devrait permettre de mieux appréhender les mouvements de groupes au sein des réseaux sociaux, et pourrait permettre de proposer de nouveaux modèles de génération et d'évolution des réseaux sociaux.

Références

- Aynaud, T., E. Fleury, J.-L. Guillaume, et Q. Wang (2013). Communities in evolving networks : definitions, detection, and analysis techniques. In *Dynamics On and Of Complex Networks, Volume 2*, pp. 159–200. Springer.
- Barbier, G. et H. Liu (2011). Data mining in social media. In *Social network data analytics*, pp. 327–352. Springer.
- Calabrese, F., Z. Smoreda, V. D. Blondel, et C. Ratti (2011). Interplay between telecommunications and face-to-face interactions : A study using mobile phone data. *PloS one* 6(7), e20814.
- Fortunato, S. (2010). Community detection in graphs. *Physics Reports* 486, 75–174.
- Ganter, B., G. Stumme, et R. Wille (2005). Formal concept analysis, foundations and applications. *Lecture Notes in Computer Science* 3626.
- Greene, D., D. Doyle, et P. Cunningham (2010). Tracking the evolution of communities in dynamic social networks. In *Advances in social networks analysis and mining (ASONAM), 2010 international conference on*, pp. 176–183. IEEE.
- Newman, M. E. et M. Girvan (2004). Finding and evaluating community structure in networks. *Physical review E* 69(2), 026113.
- Oliveira, M. et J. Gama (2010). Bipartite graphs for monitoring clusters transitions. In *International Symposium on Intelligent Data Analysis*, pp. 114–124. Springer.
- Philip, S. Y., J. Han, et C. Faloutsos (2010). *Link mining : Models, algorithms, and applications*. Springer.
- Rattigan, M. J., M. Maier, et D. Jensen (2007). Graph clustering with network structure indices. In *Proceedings of the 24th international conference on Machine learning*, pp. 783–790. ACM.
- Shen, H., X. Cheng, K. Cai, et M.-B. Hu (2009). Detect overlapping and hierarchical community structure in networks. *Physica A : Statistical Mechanics and its Applications* 388(8), 1706–1712.
- Stattner, E. et M. Collard (2012). Social-based conceptual links : Conceptual analysis applied to social networks. *International Conference on Advances in Social Networks Analysis and Mining*.
- Stattner, E. et M. Collard (2015). Descriptive modeling of social networks. *Procedia Computer Science* 52, 226–233.
- Stehle, J., N. Voirin, A. Barrat, C. Cattuto, L. Isella, J. Pinton, M. Quaggiotto, W. Van Den Broeck, C. Regis, B. Lina, et al. (2011). High-resolution measurements of face-to-face

Clusters de Liens dans des Réseaux Sociaux Dynamiques

- contact patterns in a primary school. *PloS one* 6(8).
- Toivonen, R., L. Kovanen, M. Kivela, J. Onnela, J. Saramaki, et K. Kaski (2009). A comparative study of social network models : network evolution models and nodal attribute models. *Social Networks* 31.
- Yang, J., J. McAuley, et J. Leskovec (2013). Community detection in networks with node attributes. In *Data Mining (ICDM), 2013 IEEE 13th International Conference on*, pp. 1151–1156. IEEE.
- Zhou, Y., H. Cheng, et J. Yu (2009). Graph clustering based on structural/attribute similarities. *VLDB Endowment* 2(1), 718–729.

Summary

Numerous methods have been proposed in order to extract clusters from social networks. While many studies are conducted on the design of innovative methods that are able to search for various kinds of clusters, most approaches make the assumption of static networks. One of the recent methods concerns the search for conceptual links. This is a new approach of link clustering that exploits both network structure and node attributes in order to identify frequent links between groups of nodes in which nodes share common attributes. In this work, we focus on the tracking of conceptual links in dynamic networks, namely networks experiencing important structural changes during their evolution. In particular, we seek to understand how conceptual links appear and evolve during the network development. For this purpose, we propose a set of measures that aim to capture some behaviors characterising the evolution of these clusters. Our approach is thus used to understand the evolution of the conceptual links extracted on two real world networks: a scientific co-author network and a mobile communication network. The results obtained highlight significant trends in the evolution of the conceptual links in these two networks.