Une méthode mixte d'analyse d'un réseau social: classification prétopologique et centralité d'intermédiarité

Vincent Levorato *

*LIFO (Laboratoire d'Informatique Fondamentale d'Orléans)
Batiment IIIA, Rue Léonard de Vinci, B.P. 6759 F-45067 ORLEANS Cedex 2
vincent.levorato@univ-orleans.fr

Résumé. Dans cet article, nous proposons de modéliser les réseaux sociaux par la théorie de la prétopologie comme une généralisation de la théorie des graphes. Après quelques définitions, nous expliquons comment nous pouvons généraliser par la prétopologie des méthodes d'analyse connues (k-moyennes, centralité d'intermédiarité) dans l'optique d'obtenir des résultats nouveaux. Pour appuyer notre modélisation, nous donnerons un exemple d'application sur un réseau social réel du Web.

1 Introduction

Actuellement, la modélisation des réseaux complexes est utilisée dans de nombreux domaines scientifiques, et se base principalement sur la théorie des graphes. Les graphes sont utilisés, par exemple en sciences sociales, afin de modéliser les interactions entre entités. La plupart de ces études considèrent les individus comme des entités uniques, un groupe étant formé par plusieurs individus, les uns interagissant avec les autres. En effet, la plupart des travaux portant sur l'analyse des réseaux sociaux modélisent un groupe comme une combinaison d'individus, non comme une entité propre. Les réseaux sociaux étant des réseaux complexes (Newman et al. (2006)), un phénomène d'émergence peut apparaître, et le comportement d'un groupe de personnes peut être différent de la "somme" des comportements de chaque individu. De notre point de vue, la théorie des graphes paraît insuffisante pour modéliser toutes les interactions complexes qui ont lieu dans un réseau social : nous proposons l'utilisation d'une théorie plus générale, la théorie de la *prétopologie* (Belmandt (1993)).
Cet article est structuré en trois parties :

- dans la première partie, nous donnons les définitions de la prétopologie et la définition d'un réseau (social ou non) qui en découle.
- dans la deuxième partie, nous explicitons notre apport : une nouvelle méthode d'analyse d'un réseau social en se basant sur l'algorithme des k-moyennes et un indice de centralité connu (ici la centralité d'intermédiarité) adapté au cas prétopologique général.
- enfin, pour illustrer notre discours, nous montrerons les résultats que l'on obtient sur un réseau social réel du Web.

2 Modélisation prétopologique d'un réseau social

Avant d'entrer dans la définition des concepts prétopologiques, nous allons étayer notre propos par un exemple simple d'interactions dans un petit réseau social composé de quatre individus : John, Tim, Ben et Ed (voir Fig. 1). Si on considère qu'une arête représente une relation d'amitié, il est aisé de savoir qui est l'ami de qui. On peut facilement trouver les amis de John par exemple qui sont Tim, Ben et Ed. En revanche, peut-on trouver facilement les amis du groupe $\{John, Ed\}$? Cela pose un problème car Tim est ami avec John mais pas avec Ed. Pour résoudre ce problème, nous devons nous référer à une certaine proximité entre les éléments, ce qui revient à définir la notion de *voisinage* du modèle. On peut définir par exemple que les amis d'un groupe de personnes sont ceux qui ont au moins une relation d'amitié avec un des individus du groupe. Mais on peut également définir qu'un ami d'un groupe de personnes doit être en relation avec tous les membres du groupe. On se rend bien compte de la complexité des interactions qui peuvent intervenir dans un tel modèle, d'où l'intérêt d'utiliser une théorie permettant de modéliser ces phénomènes : la théorie de la *prétopologie*.
La prétopologie est un outil mathématique définissant la proximité entre les éléments d'un espace discret. Cette théorie généralise la topologie, permettant d'analyser un système complexe *pas à pas*, grâce à des processus d'adhérence, d'intérieur, que nous définissons ci-après.

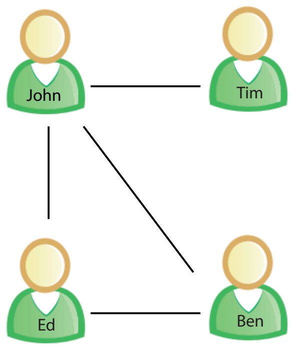

FIG. 1 – *Exemple d'un petit réseau social avec une relation d'amitié*

2.1 Concepts prétopologiques

Soit E un ensemble non vide, et soit $\mathcal{P}(E)$ l'ensemble des parties de E.

Soit une application $a : \mathcal{P}(E) \rightarrow \mathcal{P}(E)$ appelée *adhérence* et définie comme suit :
$\forall A$, $A \subseteq E$ l'adhérence de A, $a(A) \subseteq E$ est telle que :

– $a(\emptyset) = \emptyset$ $\quad (P_1)$
– $A \subseteq a(A)$ $\quad (P_2)$

L'adhérence est associée au processus de *dilatation*. De plus, $a(.)$ peut être appliquée à A selon une séquence : $A \subseteq a(A) \subseteq a^2(A) \subseteq \dots$. Cela signifie que l'on peut suivre le processus pas à pas, ce qui n'est pas possible avec la topologie, qui conserve la propriété d'idempotence ($a(A) = a^2(A)$) (Bourbaki (1971)). Grâce à l'adhérence, on peut directement modéliser la notion de *proximité*.

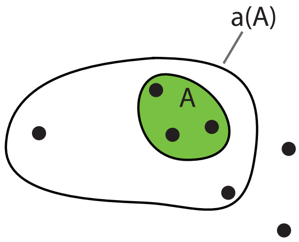

FIG. 2 – *Adhérence de A*

Soit une application $i : \mathcal{P}(E) \rightarrow \mathcal{P}(E)$ appelée *intérieur* et définie comme suit : $\forall A$, $A \subseteq E$ l'intérieur de A, $i(A) \subseteq E$ est telle que :

- $i(A) = [a(A^c)]^c$ $\quad (P_1)$
- $i(A) \subseteq A$ $\quad (P_2)$

avec A^c le complémentaire de A soit $E - A$.

L'intérieur est quant à lui associé au processus *d'érosion*. Notons que la propriété 1 de l'intérieur amenant la dualité n'est pas toujours vraie. Il est possible de définir une application intérieur indépendamment de l'adhérence.

On appelle *espace prétopologique* le triplet (E, i, a) dont les applications i et a sont définies précédemment.

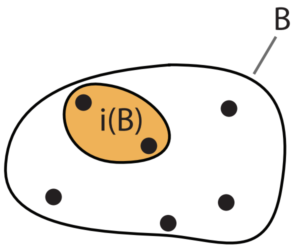

FIG. 3 – *Intérieur de B*

L'intérêt des précédentes définitions pour la modélisation des réseaux sociaux peut être expliqué ainsi : on peut dire que les éléments de $a(A)$ sont proches de A (voisins "directs"), et pour chaque adhérence, on absorbe de nouveaux éléments. On est capable de modéliser des dynamiques complexes comme la diffusion d'une information dans un réseau par exemple.

Nous avons également l'application intérieur : celle-ci permet d'exclure des éléments en périphérie d'un groupe social. Pour reprendre l'exemple précédent, on pourrait retrouver grâce à une série d'intérieurs, l'origine de la diffusion de l'information.

Le processus de dilatation généré par l'adhérence s'arrête à un instant donné et n'évolue plus. Dans ce cas, on a $a^{k+1}(A) = a^k(A)$. On nomme A comme étant un sous ensemble *fermé*. De la même manière, l'évolution de l'intérieur va cesser, ce qui nous donne $i^{k+1}(A) = i^k(A)$. Cette fois, on nomme A comme étant un sous ensemble *ouvert*.
Respectivement, on utilise les notations $F(A)$ pour la fermeture de A et $O(A)$ pour l'ouverture de A.

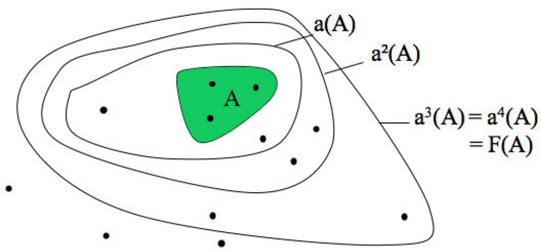

FIG. 4 – *Illustration d'adhérences successives menant au fermé*

On appelera fermé élémentaire et on notera F_x, la fermeture d'un singleton $\{x\}$ de E. On note $\mathcal{F}_e(E, a)$ ou $\mathcal{F}_e$, l'ensemble des fermés élémentaires de E :

$$\mathcal{F}e(E, a) = \{F_x, x \in E\}$$

On appelle fermé minimal de E, tout élément de $F(E, a)$, minimal au sens de l'inclusion. L'ensemble des fermés minimaux est noté : $\mathcal{F}_m(E, a)$ ou $\mathcal{F}_m$.
Un résultat important est que tout fermé minimal est obligatoirement élément de $\mathcal{F}_e$, c'est à dire un fermé élémentaire. Déterminer les fermés minimaux revient donc à explorer les éléments de $\mathcal{F}_e$ et en extraire les éléments minimaux par la relation d'inclusion.

2.1.1 Espace prétopologique de type $\mathcal{V}$

Un espace prétopologique général comme défini ultérieurement ne présente que peu d'intérêt en l'état, car il est difficile d'en faire une analyse. Il faut donc amener une nouvelle propriété pour rendre cet espace prétopologique plus "intéressant", d'où la définition d'un nouvel espace prétopologique : le type $\mathcal{V}$.

Un espace prétopologique de type $\mathcal{V}$ (E, a) est défini comme suit :

$$\forall A, B, A \subseteq E, B \subseteq E \; et \; A \subset B \; avec \; a(A) \subseteq a(B)$$

2.1.2 Espace prétopologique de type $\mathcal{V}_d$

Un espace prétopologique de type $\mathcal{V}_d$ (E, a) est défini comme suit :

$$\forall A, B, A \subseteq E, B \subseteq E \; et \; A \subset B \; avec \; a(A \cup B) = a(A) \cup a(B)$$

Tout espace de type $\mathcal{V}_d$ est de type $\mathcal{V}$.

2.1.3 Espace prétopologique de type $\mathcal{V}_s$

Un espace prétopologique de type $\mathcal{V}_s$ (E, a) est défini comme suit :

$$\forall A, A \subseteq E, avec \; a(A) = \bigcup_{\forall \; x \in A} a(\{x\})$$

Un espace de type $\mathcal{V}_s$ est clairement de type $\mathcal{V}_d$. Les applications a et i ne sont pas forcément idempotentes. On ne doit pas confondre une prétopologie de type $\mathcal{V}_s$ et une topologie. Les types d'espaces les plus utilisés dans nos études sont les types $\mathcal{V}$ et $\mathcal{V}_s$.

2.2 Définition d'un réseau en prétopologie

Maintenant les concepts prétopologiques présentés, nous sommes en mesure de définir un réseau (social) de manière prétopologique. Un réseau social peut être défini comme une famille de relations binaires ou valuées définies sur une population donnée (Degenne (2004)). La dynamique d'un réseau est basée sur des opérations telle que l'arrivée de nouveaux éléments, l'éviction d'éléments existants, la formation de groupe ou la séparation en sous-groupes. Ces phénomènes sont souvent observables dans les réseaux sociaux sous forme de communautés (Backstrom et al. (2006)) mais également dans le cas des réseaux de manière plus générale.

Dans le cadre de la prétopologie, un réseau est une famille de prétopologies sur un ensemble donné (Fig. 5), d'où la définition suivante (Dalud-Vincent (1994)) :

Soit X un ensemble :
soit I une famille dénombrable d'indices ;
soit $\{a_i, i \in I\}$ une famille de prétopologies sur X_i ;
la famille d'espaces prétopologiques $\{(X, a_i), i \in I\}$ constitue un réseau sur X.

On peut représenter ainsi des relations de natures différentes : par exemple, on pourra modéliser un réseau social où les individus sont reliés entre eux par une relation d'amitié (relation binaire) et où leur emplacement géographique est nécessaire (métrique). Le voisinage d'un individu pourra être défini selon les besoins de la problématique : sont voisins ceux qui sont amis et qui habitent dans un rayon de x km. La définition de l'adhérence et/ou de l'intérieur dépend donc de la nature de la problématique : un certain nombre de travaux dans ce domaine ont déjà montré d'intéressants résultats (Bonnevay et al. (1999); Largeron et Bonnevay (1997); Levorato et Bui (2007); Levorato et al. (2009)).

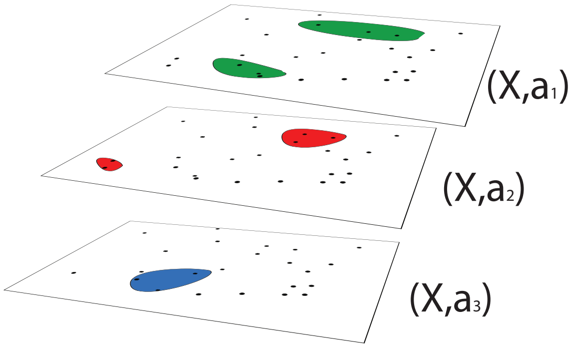

FIG. 5 – *Exemple de trois prétopologies différentes sur X*

3 Analyse d'un réseau social par une méthode mixte : k-moyennes et centralité d'intermédiarité

L'idée est la suivante :
- on partitionne notre réseau social grâce à une méthode des k-moyennes,
- on classe les partitions obtenues selon leur centralité d'intermédiarité.

Dans cette section, nous présentons un algorithme de partionnement des k-moyennes basé sur la théorie de la prétopologie qui a déjà fait l'objet de travaux et qui a été introduit par Le et al. (2008). Nous présenterons ensuite nos contributions : la centralité d'intermédiarité prétopologique, puis la méthode finale permettant une analyse de réseau social.

3.1 MCPR : Méthode de Classification Prétopologique avec Réallocation

La base de cette méthode reprend l'algorithme original de MacQueen (1967). Cet algorithme assigne chaque objet au sous-ensemble dont le centre est le plus proche de l'objet en question :
- Choisir le nombre de sous-ensembles k à obtenir.
- Choisir k groupes de manière aléatoire et en déterminer les centres, ou prendre aléatoirement k objets comme étant les centres initiaux.
- Assigner chaque objet au groupe dont le centre est le plus proche.
- Recalculer les centres de chaque groupe.
- Répéter les deux étapes précédentes tant que la composition des groupes change.

La performance de cette technique est "proportionnelle" à la qualité de la fonction de mesure de distance utilisée. En prétopologie, nous ne sommes pas forcément dans un espace métrique, donc nous ne disposons pas d'une distance à proprement parler. Une pseudo-distance doit être définie : Le et al. (2008) définissent $\delta(A, B)$ comme la pseudo-distance entre deux sous-ensembles A et B d'un ensemble fini E. Celle-ci est calculée comme la distance de Hausdorff.

- $k_0 = min(min\{k|A \subset a^k(B)\}, \infty)$
- $k_1 = min(min\{k|B \subset a^k(A)\}, \infty)$
- $\delta(A, B) = min(k_0, k_1)$

La famille $\mathcal{F}_m$ des fermés minimaux de E représente le nombre k de partitions à obtenir. Concernant la détermination du centre d'un ensemble F, on procède comme suit :
On note x_0 le centre de l'ensemble.
Avec $F = \bigcup_{x \in F} \{x\}$, nous devons décider quel $\{x_i\}$ choisir. Pour cela, nous calculons $Card(a(x_i))$ avec $i \in [1, Card(F)]$. Nous choisissons x_0 tel que $Card(a(x_i))$ soit l'adhérence contenant le plus grand nombre d'éléments.
Au cas où plus d'un x_0 existe, on choisit x_0 de manière à ce que celui-ci minimise la pseudo-distance avec *le plus grand fermé élémentaire qui le contient*.

L'algorithme MCPR se basant sur l'algorithme des *k-moyennes*, on retrouve ainsi son déroulement dans ce qui suit :

1. Choisir k groupes initiaux par les fermés minimaux puis en calculer les centres en formant ainsi k classes.
2. (Ré)affecter chaque objet x à la classe C_i de centre M_i tel que $\delta(x, M_i)$ soit minimale
3. Recalculer le centre M_i de chaque classe.
4. Aller à l'étape 2 jusqu'à ce que les objets ne changent plus de classe.

Nous avons là une méthode prétopologique de partionnement, applicable à des espaces non-métriques ou mixtes. Nous voulons aller au-delà du simple partitionnement d'un réseau social, en classant ces partitions selon leur importance. Ici, nous nous penchons sur le rôle que chaque partition peut jouer dans le réseau en terme de diffusion de l'information.

3.2 Centralité d'intermédiarité prétopologique

La centralité d'intermédiarité a été proposée par Freeman (1977) et défend l'idée qu'un individu peut bien être faiblement connecté aux autres et même relativement éloigné, mais servir d'intermédiaire dans bon nombre des échanges entre les autres membres du groupe. Plus il sert ou peut servir d'intermédiaire pour tous les membres, plus il est en position de contrôler la communication ou d'être indépendant des autres pour communiquer. Un tel individu peut influencer le groupe plus facilement en filtrant ou distordant les informations qui y circulent. Sa position lui permet également d'assurer la coordination du groupe. D'où la définition suivante :

Centralité d'intermédiarité : *Soit n le nombre de sommets d'un graphe, g_{jk} le nombre de chemins géodésiques*[1] *reliant le nœud j au nœud k, et $g_{jk}(i)$ le nombre de ces chemins passant par le nœud i, on définit $C_{AI(i)}$ l'indice de centralité absolu d'intermédiarité du sommet i par :*

$$C_{AI}(i) = \sum_{j}^{n} \sum_{k=1}^{n} \frac{g_{jk}(i)}{g_{jk}}$$

1. plus courts chemins.

avec : $j \neq k \neq i$ *et* $j < k$

La propriété de Freeman est intéressante, et il nous a paru intéressant d'en adapter une version prétopologique plus générale :

Algorithme 1 ***Algorithme d'intermédiarité prétopologique***
Méthode : ***PretopoBetweeness****(Ensemble A)*
Variables :
A *: ensemble de départ tel que* $A \subset E$
g_{jk}, $g_{jk}i$, $g_{jk}tmp$ *: entier*
$Bdeg$ *: réel*

Début
$Bdeg \leftarrow 0$
$g_{jk} \leftarrow 0$
$g_{jk}i \leftarrow 0$
Pour *i de 0 à* $Card(E)$ ***Faire***
    Pour *j de 0 à* $Card(E)$ ***Faire***
        $elt_i \leftarrow$ *singleton de* E
        $elt_j \leftarrow$ *singleton de* E
        $g_{jk}tmp \leftarrow$ *nb_chemins_geo(*elt_i*,*elt_j*)*
        Si $g_{jk}tmp > 0$ ***Alors***
            $g_{jk} \leftarrow g_{jk} + g_{jk}tmp$
            $g_{jk}i \leftarrow g_{jk}i$ + *(nb_chemins_geo(*elt_i*,A)* $\times$ *nb_chemins_geo(A,*elt_j*))*
        FinSi
    FinPour
FinPour
Si $g_{jk} > 0$ ***Alors***
    $Bdeg \leftarrow g_{jk}i/g_{jk}$
FinSi
Renvoyer $Bdeg$
Fin

Exemple
Voici un exemple concret sur un passage de la boucle ci-dessus. Soit E un espace prétopologique de type $\mathcal{V}_s$ avec des relations inter-éléments de nature binaire. Pour une plus grande facilité de lecture, nous représentons l'espace E comme un graphe (Fig. 6). Après avoir classé les éléments selon l'adhérence dans laquelle il se trouvent, en supposant que la classe 1 représente les éléments contenus dans l'adhérence de degré 1 (moins l'élément j), nous avons :

1. {A,B,C}
2. {D,E,F}
3. {G,k}

De manière intuitive, nous excluons d'emblée l'élément G. Ainsi, les plus courts chemins entre j et k sont : j-A-D-k, j-A-E-k, j-B-E-k.
Dans ce cas précis, il y a 3 chemins géodésiques entre j et k. Nous voulons calculer la centralité d'intermédiarité d'un élément, par exemple E (nommé i dans l'algorithme). On remarque qu'il y a 2 chemins géodésiques entre j et k passant par E. Donc, la centralité d'intermédiarité de l'élément E, pour un seul passage de la boucle est de $\frac{2}{3}$. Pour avoir le résultat final, il faut bien sûr finir l'algorithme en prenant toutes les paires (j, k) du réseau avec E comme élément i. L'intérêt de cet algorithme est que si pour l'exemple et la compréhension, on ne travaille qu'avec des singletons, dans la pratique, on peut calculer l'intermédiarité d'un ensemble, permettant ainsi de généraliser l'algorithme original.

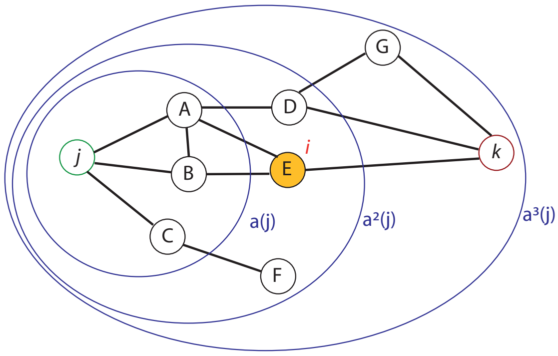

FIG. 6 – *Adhérences successives de j*

3.3 Méthode mixte : MCPR & centralité d'intermédiarité

La méthode d'analyse qui consiste à coupler MCPR et la centralité d'intermédiarité généralisée reprend l'idée énoncée en début de section : on partitionne tout d'abord notre espace en groupes grâce à MCPR puis on classe ces partitions selon leur centralité d'intermédiarité entre elles. C'est à dire qu'on considère uniquement les chemins d'une classe à une autre. Pour une classe étudiée, plus il y a de chemins entre deux autres classes passant par cette classe, plus celle-ci est intermédiaire.

Algorithme 2 ***Algorithme d'analyse mixte***
Méthode : ***PretopoMCPR&Between****(Espace E)*
Variables :
E *: espace prétopologique*
$listeMCPR$ *: liste d'ensembles (partitions)*
$listeResultats$ *: couple ensemble-réél (composition de l'ensemble et score associé)*

Début
$listeMCPR \leftarrow MCPR(E)$
Pour *i de 0 à* $listeMCPR.taille - 1$ ***Faire***
    $listeResultats.index(i) \leftarrow couple(listeMCPR.get(i), PretopoBetweenessClasses(listeMCPR.get(i))$
FinPour
Renvoyer $listeResultats$
Fin

On peut donc associer un poids d'intermédiarité à chaque classe. Nous avons appliqué cette méthode sur un réseau social Web : YouTube. Les données ont été extraites par Cheng et al. (2008) et représentent les liens "Vidéos similaires" qu'il peut y avoir entre les vidéos. Notre méthode a été appliquée sur un réseau non-connexe de 953 vidéos et 3037 liens orientés. Le sens de la centralité d'intermédiarité dans ce cas est que plus une vidéo est intermédiaire, plus son rôle dans le ou les plus courts chemins entre deux vidéos quelconque est important. Si on exclut ce genre d'éléments du réseau, il est probable que la taille du plus court chemin entre couples d'éléments augmente ou même qu'il n'y ait plus du tout de chemin. En terme d'interprétation, cela signifie que certaines vidéos permettent de faire découvrir un maximum d'autres vidéos par leur biais. Avoir cette information peut, par exemple, permettre de promouvoir toute une catégorie de vidéos en mettant en avant seulement quelques vidéos clés. Voici comment est défini notre espace prétopologique avec R des relations binaires réflexives non-symétriques :

$$R_1(x) = \{y \in E, xRy\}, R_2(x) = \{y \in E, yRx\}$$

$$\forall A \in \mathcal{P}(E), a_1(A) = \{x \in E, R_1(x) \cap A \neq \emptyset\} \text{ et}$$
$$\forall A \in \mathcal{P}(E), a_2(A) = \{x \in E, R_2(x) \cap A \neq \emptyset\}$$

On utilise la première adhérence pour MCPR car ce qui nous intéresse dans un premier temps, ce sont les voisins d'une vidéo qui pointent vers celle-ci. Puis on utilise la deuxième adhérence pour la partie intermédiarité car ce sont les chemins dans le réseau qui nous intéressent par la suite. Nous donnons une illustration de ce que l'on veut obtenir sur une partie de réseau YouTube Fig. 7. Les classes sont en vert, il peut y avoir des éléments non classés, et les éléments colorés appartiennent aux classes les plus intermédiaires. La non-connexité est tout à fait compréhensible pour un réseau tel que YouTube, en tout cas en ce qui concerne la récupération des données. Evidemment, cette exemple, servant uniquement à la visualisation du problème, n'est pas très représentatif puisque l'on a trop peu d'éléments pour pouvoir faire une quelconque interprétation. Néanmoins, cela permet de comprendre le principe : après avoir partitionné le réseau, on recherche les classes les plus intermédiaires, et à fortiori les éléments les composant.

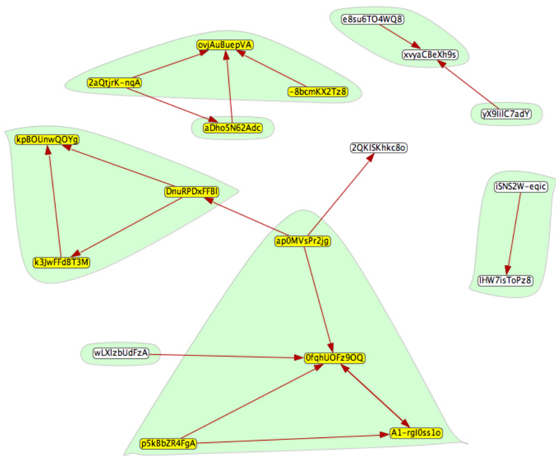

FIG. 7 – *Exemple de résultats obtenus*

Sur le réseau de 953 vidéos et 3037 liens, nous obtenons 171 groupes dans un premier temps, composés au maximum de 20 éléments. Pour ces groupes, seuls 29 d'entre eux possèdent une intermédiarité supérieure à 0. Sur ces 29 groupes, 5 groupes se détachent avec des valeurs supérieures aux 24 restants, représentant 42 éléments du réseau. Les scores se situent entre 0.023 et 0.011, ordre de grandeur que l'on retrouve avec la version standard de la centralité d'intermédiarité. D'ailleurs, en comparant les scores noeud par noeud du réseau avec la centralité d'intermédiarité originale, et le score calculé par notre méthode, si on retrouve

Noeud	Score
$_d8nWIqxo0U$	0.028
$L95Sv5aLtZg$	0.015
$Q7Cpi5t-YQI$	0.015
$drW1zIv4wnA$	0.013
$PyC3Byv_0mM$	0.012
$4LK3KVSKebI$	0.011
$cGY5lhFZFpc$	0.011
$6QVtiaBImlw$	0.011
...	...
Partition	**Score**
[dl4-v0oK7PY, bCZZnJYcPZ4, QBvyHHwQLdw]	0.02353
[mUko-hI7rV8]	0.02353
[Oh40Lm4sRoU, aHer7D0USEc, L95Sv5aLtZg, Q1U6LjALlgo, 4LK3KVSKebI, hECwBuJN2us, eiApTvWkkBI, PyC3Byv_0mM, aWpkJIOwH5c, 2J-Moz5j_do, U1I6H3hTzeI, vLRWAnJQjbU, RYETO3-FTek, Q7Cpi5t-YQI, JWp_gxJORAQ, uRkkmVR6ATU, 9HTzWE8WeCc, S5rU71HA69Y, 9RDCWaSERoY]	0.02353
[_d8nWIqxo0U, 2_W1u-K7dS8, UPumjbRBqZM, XllwXjlKIvA, kBYoYeIPBUc, cGY5lhFZFpc, vPXKq2HWHJM, hI4ixSkCgas, P2JYz22fwAI, du1SK1SYgEo, drW1zIv4wnA, b5RVxE4jLD0, PX7Yujz6Kj8, NmzplkH4q88]	0.02353
[SsHjkq-_RCw, 68gwSEpog6g, 81UN9izhrjM, 6QVtiaBImlw, V8pVrKvofpY]	0.02353
...	...

TAB. 1 – *Extrait de résultats de centralités d'intermédiarité : méthode standard et méthode prétopologique mixte*

une certaine cohérence, on a cependant des noeuds qui, seuls, ont une centralité d'intermédiarité proches de zéro, et qui, en groupe, ont un score placé dans les premiers (Tab. 1). Nous observons une propriété émergente que l'on voit apparaître seulement si on regroupe certains éléments qui, ensemble, auront un comportement différent que s'ils agissaient chacun séparément. Notre méthode nous permet dans ce cas de détecter des phénomènes que l'on n'aurait pas pu déceler avec une modélisation et des méthodes d'analyse classiques.

4 Conclusion

Le travail présenté dans ce document participe à la généralisation de la modélisation des réseaux sociaux du Web, de part la modélisation utilisée (théorie de la prétopologie, qui allie qualitatif et quantitatif), et des méthodes algorithmiques proposées. Outre le fait d'avoir donné un algorithme alliant les k-moyennes et la centralité d'intermédiarité généralisés, notre méthode amène une vision différente de celles que l'on connaît habituellement dans ce domaine, en prenant en compte l'émergence de propriétés qui apparaît quand plusieurs éléments forment un groupe, celui-ci ayant un comportement différent des seuls éléments le composant. Il nous est donc susceptible d'obtenir des résultats "plus fins". Bien entendu, le problème de l'interprétation des résultats ne disparaît pas pour autant, nécessitant l'avis d'experts du domaine (sociologues), mais en nous plaçant à un niveau de modélisation plus général, nous sommes désormais capables d'analyser de manière plus précise et plus fine la dynamique et la structure des réseaux complexes, nous permettant de mieux comprendre les phénomènes émergents qui s'y déroulent.

Références

Backstrom, L., D. Huttenlocher, J. Kleinberg, et X. Lan (2006). Group formation in large social networks : Membership, growth, and evolution. *Proc. 12th ACM SIGKDD Intl. Conf. on Knowledge Discovery and Data Mining*.

Belmandt, Z. (1993). *Manuel de prétopologie et ses applications : Sciences humaines et sociales, réseaux, jeux, reconnaissance des formes, processus et modèles, classification, imagerie, mathématiques*. Hermes Sciences Publications.

Bonnevay, S., M. Lamure, C. Largeron, et N. Nicoloyannis (1999). A pretopological approach for structuring data in non-metric spaces. *Electronic Notes in Discrete Mathematics 2*.

Bourbaki, N. (1971). *Topologie générale*. Hermann.

Cheng, X., C. Dale, et J. Liu (2008). Dataset for "statistics and social network of youtube videos". School of Computing Science Simon Fraser University British Columbia, Canada. http ://netsg.cs.sfu.ca/youtubedata/.

Dalud-Vincent, M. (1994). *Modèle prétopologique pour une méthodologie d'analyse des réseaux : concepts et algorithmes*. Ph. D. thesis, Université Claude Bernard - Lyon 1.

Degenne, A. (2004). Entre outillage et théorie, les réseaux sociaux. *Réseaux Sociaux de l'Internet*.

Freeman, L. (1977). A set of measures of centrality based on betweenness. *Sociometry 40*, 35–41.

Largeron, C. et S. Bonnevay (1997). Une méthode de structuration par recherche de fermés minimaux : application à la modélisation de flux de migrations inter-villes. In *5ème rencontres de la Société Française de Classification*, Lyon, France.

Le, T. V., N. Kabachi, et M. Lamure (2008). Pretopology and a homogeneous method for data clustering. In *RIVF'08 conference*, Hochiminh city, Vietnam.

Levorato, V. et M. Bui (2007). Modeling the complex dynamics of distributed communities of the web with pretopology. In *I2CS*, Munich, Germany.

Levorato, V., T. V. Le, M. Lamure, et M. Bui (2009). Classification prétopologique basée sur la complexité de kolmogorov. *Studia informatica universalis 7.1*, 199–222.

MacQueen, J. (1967). Some methods for classification and analysis of multivariate observations. In *5th Berkeley Symposium on Mathematical Statistics and Probability*, Berkeley, pp. 281–297.

Newman, M., A.-L. Barabási, et D. J. Watts (2006). *The Structure and Dynamics of Networks*. Princeton University Press.

Summary

In this paper, we propose to model social networks by applying the pretopology theory as a generalization of the graph theory. After giving some definitions and examples, we explain how measures used in social network analysis (k-means, betweenness centrality) can be generalized with pretopology theory in order to obtain new interesting results. To argue in this sense, our work will be supported by an example of application obtained on a real Web social network.