

# Des graphes de documents aux réseaux sociaux

Michel Plantié\*, Michel Crampes\*

\*EMA-LGI2P, Parc Scientifique Georges Besse, 30035 Nîmes Cedex, France  
michel.plantie,michel.crampes@mines-ales.fr, <http://www.lgi2p.ema.fr>

## 1 Introduction

Dans ce papier, nous cherchons à constituer des réseaux sociaux à partir de documents partagés par des personnes dans des plateformes collaboratives. Puis nous cherchons à constituer des listes de diffusions personnalisées.

## 2 Construction du réseau social

Organiser socialement des documents et élaborer des réseaux sociaux à partir du contenu des documents est depuis un certain temps un domaine actif de recherche. Beaucoup de travaux comme par exemple (Velardi et al., 2008; Mika, 2007) cherchent à utiliser les relations entre personnes dans un collectif afin de découvrir des connaissances communes et les exploiter pour différents objectifs.

Le concept de tribu : Nous appelons Tribu un sous-ensemble de personnes  $x_i$  concernées par un partage de document personnalisé. Tous les individus qui appartiennent à la même tribu possèdent un réseau social commun qui est mis en évidence par l'apparition de ces individus sur différents documents sous des formes différentes. Nous résonnerons ici sur un couple d'individus  $x, y$ , plus précisément sur le lien entre  $x$  et  $y$ .

### 2.1 Modèles de Forces de couples et tribus

Ci-dessous nous définissons différentes lois qui constitueront des tribus différentes et nous comparerons ensuite ces différents découpages. Nous définissons d'abord trois forces brutes, et nous définirons ensuite une mesure de pondération de ces forces.

La force simple d'un couple : Ici nous définissons la "fréquence" d'un couple (deux individus dans une certaine forme de relation) comme le nombre d'occurrences de ce couple parmi les concepts divisé par le nombre de concepts :

La proximité ou force pondérée d'un couple : La proximité d'une tribu (respectivement d'un couple) est la somme des inverses des nombres d'individus présents dans les concepts qui contiennent la tribu, divisée par le nombre total de concepts. Donc, plus le couple est "dilué" dans un grand nombre d'individus pour un concept, moins la contribution de ce concept à la proximité est forte. Quand une tribu (un couple) n'apparaît dans aucun concept, la force est

nulle. Deux personnes apparaissant dans un seul concept, donne la proximité maximum de  $2/2/1 = 1$ .

Cohésion d'un couple : Plus deux personnes sont vues séparément, plus elles sont indépendantes l'une de l'autre. Inversement quand deux personnes sont observées toujours ensemble, cela renforce l'hypothèse de liens particuliers entre eux. Nous l'appelons "cohésion".

Nous exprimons ensuite une distance sémantique entre les documents du corpus de la plateforme collaborative. Ce graphe sert à élaborer le graphe d'auteurs dont les noeuds du graphe représentent les acteurs. Les arcs représentent les liaisons entre acteur, et leur longueur est proportionnelle à cette distance. Nous calculons toutes les distances entre tous les documents des deux ensembles, puis nous choisissons la distance minimum parmi toutes ces distances.

Expressions de la sémantique par pondération des forces précédentes : Les liens exprimeront donc à la fois les valeurs de co-écriture et co-intérêt en même temps que l'aspect sémantique. Ainsi nous combinons les forces exprimées avec la distance sémantique.

## 2.2 Extraction des tribus en procédant par filtrage sur les composantes connexes

Dans cette méthode, le principe est de classer les arêtes du graphe de concepts par distances croissantes. Puis à chaque échelle de valeur du seuil de distance en débutant par la valeur de distance la plus faible, nous conservons les arêtes dont la distance est inférieure ou égale au seuil considéré, ainsi que les sommets qui leur sont reliés. Nous obtenons ainsi un graphe constitué des arêtes et sommets dont le seuil de distance est supérieur au seuil fixé. Les diverses solutions de diffusion personnalisées donnent des résultats variés : 75 exemplaires, 18 exemplaires, 69 exemplaires . Les stratégies évaluées permettent de diffuser un nombre de documents plus restreint en augmentant la pertinence des documents pour chaque participant. La diffusion de documents est souvent délicate si l'on souhaite envoyer uniquement des documents intéressants, et trouve ici une solution élégante et efficace.

## Références

- Mika, P. (2007). Ontologies are us : A unified model of social networks and semantics. *Web Semantics : Science, Services and Agents on the World Wide Web* 5(1), 5 – 15. Selected Papers from the International Semantic Web Conference, International Semantic Web Conference (ISWC2005).
- Velardi, P., R. Navigli, A. Cucchiarelli, et F. d` Antonio (2008). A new content-based model for social network analysis. *008 IEEE International Conference on Semantic Computing, Santa Clara, CA*, 18–25.