

Nasserine Benchettara

A3 -LIPN UMR 7030 du C.N.R.S.

Institut Galilée – Univ. Paris-Nord

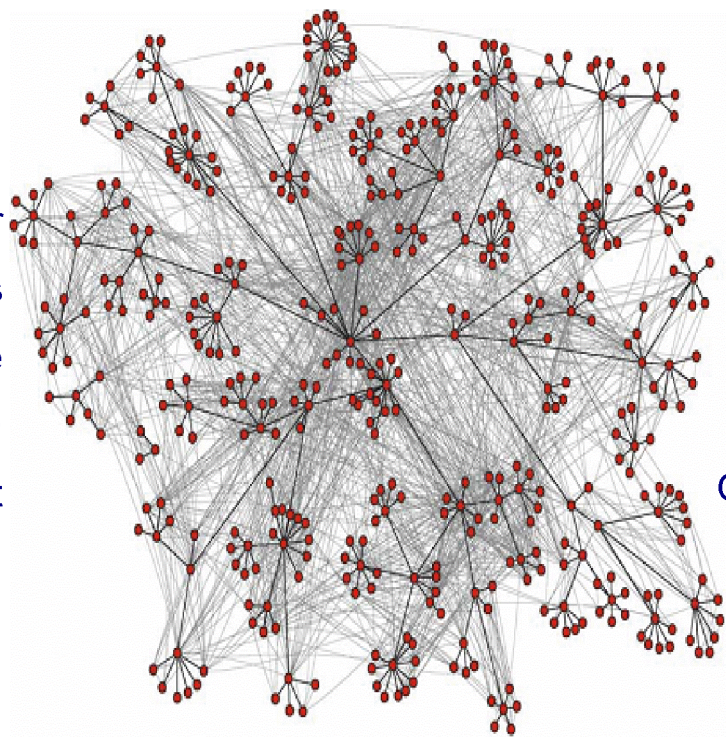
Nasserine.Benchettara@lipn.univ-paris13.fr

Encadreurs : Rushed Kanawati & Céline Rouveirol

Problématique

Les réseaux sociaux

Un réseau social est un ensemble d'entités (ex : individus, pages web,...) reliées entre elles par des relations sociales (ex : collaborations, liens hypertextes ,...). La structure sociale peut être modélisée par un graphe sans échelle $G = \langle V, E \rangle$ où les sommets désignent les entités sociales et les arrêtes les relations sociales.



Prédiction de liens

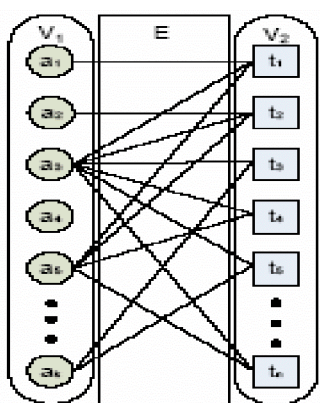
Etant donné un réseau social à l'instant t_i (qu'on note G_i), on cherche à prédire l'apparition de nouveaux liens dans le réseau G_{i+1} à un instant t_{i+1} avec $t_i < t_{i+1}$.

On s'intéresse à la prédiction des liens dans un graphe biparti $G_i = \langle V_1, V_2, E \subseteq V_1 \times V_2 \rangle$

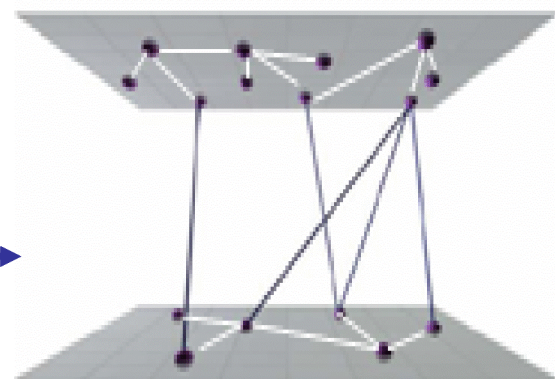
Approche proposée

- Approche dyadique : calculer une score de connexion $\text{score}(x,y)$ pour chaque paire de nœuds (x,y) .

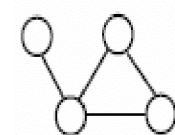
- Projeter le graphe $G_i = \langle V_1, V_2, E \subseteq V_1 \times V_2 \rangle$ sur les ensembles V_1 et V_2 . La projection sur l'ensemble V_1 est donnée par le graphe $G_1' = \langle V_1, E_1 \rangle$ où deux nœuds $u, v \in V_1$ sont liés s'ils sont reliés à au moins n nœuds communs dans V_2 , la projection sur l'ensemble V_2 se fait de façon analogue.



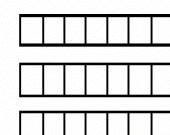
Projection



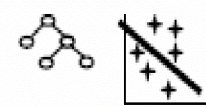
- Approche topologique : calculer les caractéristiques topologiques (ex : les voisins communs, le plus court chemin,...) pour chaque paires de nœuds dans le graphe biparti et les graphes projetés.



Extraction des
attributs
topologiques



Construction des
vecteurs
caractéristiques



Méthodes de
classification
supervisé

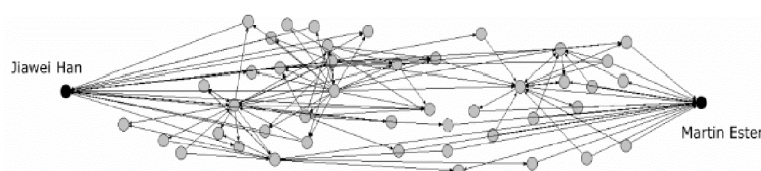
Le graphe initial est découper en deux sous graphes disjoints :

- Le graphe d'apprentissage : utilisé pour la construction des exemples à partir des attributs topologiques.

- Le graphe d'étiquetage : sert à étiqueter les exemples, il attribue un label positif si l'exemple possède un lien dans ce graphe, un label négatif sinon.

Applications

- Recommandation d'achats de musique en ligne (projet ANR CADI : cadi.lipn.fr).
- Recommandation de collaborations académiques dans le cadre du projet LIPS (données DBLP : <http://dblp.unitrier.de>).



Résultats

Nous avons expérimenté notre approche sur les données d'achats de musique en ligne dans une période d'une année, le graphe contient 837 clients et 2239 produits et 18058 transactions. On a utilisé les 9 premiers mois pour l'apprentissage et les 3 derniers mois pour l'étiquetage. Les résultats obtenus donnent une précision de 91% et un rappel de 6% pour la classe positive.

Enseignements :

- Problème de déséquilibre entre les classes.
- Problème des sous concepts mal représentés.

Perspectives

- Utiliser les informations temporelles.
- Utiliser plusieurs réseaux en même temps décrivant différents types d'informations.

Références

- BENCHETTARA N & KANAWATI R & ROUVEIROL C (2009). Calcul de recommandations par prédiction de liens dans un graphe biparti. Atelier AGS, plateforme AFIA Hammamet, Tunis. Mai 2009.
- HASAN M. A., CHAOJI V., SALEM S. & ZAKI M. (2006). Link prediction using supervised learning. In SIAM Workshop on Link Analysis, Counterterrorism and Security avec SIAM Data Mining Conference, Bethesda, MD.
- LIBEN-NOWELL D. & KLEINBERG J. M. (2003). The link prediction problem for social networks. In CIKM, p. 556-559. ACM.

