

Partitionnement des liens d'un graphe : Critères et Mesures

Noé Gaumont et François Queyroi

LIP6 - CNRS & UPMC , Université Pierre et Marie Curie – Sorbonne
Universités, Paris, France.

Algotel 2014

Sommaire

- ① Détection de communautés
- ② Définition d'une fonction de qualité pour les liens
- ③ Prise en compte du voisinage
- ④ Méthodes existantes
- ⑤ Conclusion et perspectives

Détection de communautés

Communautés de nœuds

Communautés de liens

Méthodes existantes

Communautés de nœuds

Entrée :

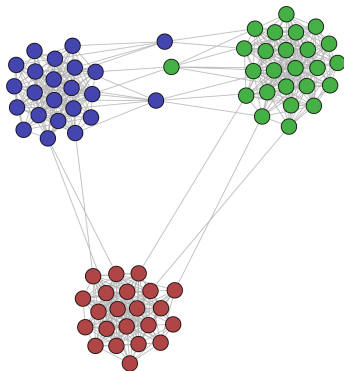
Un graphe, $G = (V, E)$.

Sortie :

Une partition \mathcal{P} de V .

S. Fortunato.

Community detection in graphs.



Exemple dans les courriels.

Communauté : groupe de personnes.

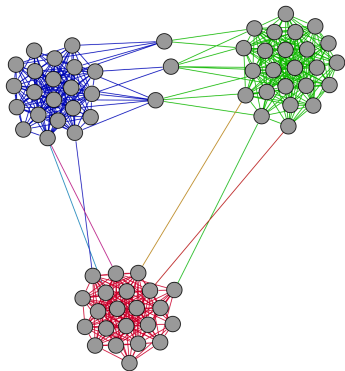
Communautés de liens

Entrée :

Un graphe, $G = (V, E)$.

Sortie :

Une partition \mathcal{P} de E .



T.S. Evans et R. Lambiotte. **Line graphs, link partitions, and overlapping communities.**

Y.-Y. Ahn, J. P. Bagrow, et S. Lehmann.

Link communities reveal multiscale complexity in networks.

Exemple dans les courriels.
Communauté : conversation.

Méthodes existantes

Communautés de liens :

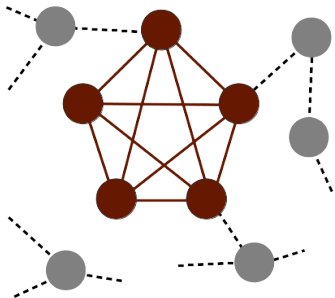
- Méthode à base de densité [Ahn et al.].
- Méthode utilisant un *line graph* [Evans et al.].

Communautés de nœuds :

- Méthodes spectrales.
- Méthode probabilistes.
- Méthodes à base de **fonction de qualités**.

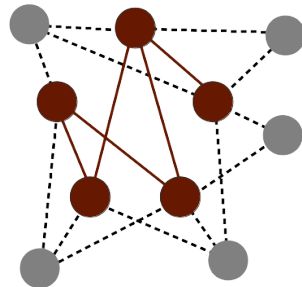
La modularité pour les nœuds

Comparer le nombre d'arêtes induites observées et celui attendu



$$\text{card}(\text{---}) = 10$$

→
Mélange des arêtes



$$\text{card}(\text{---}) = 5$$

M.E.J. Newman, M. Girvan.

Finding and evaluating community structure in networks

Définition d'une fonction de qualité pour les liens

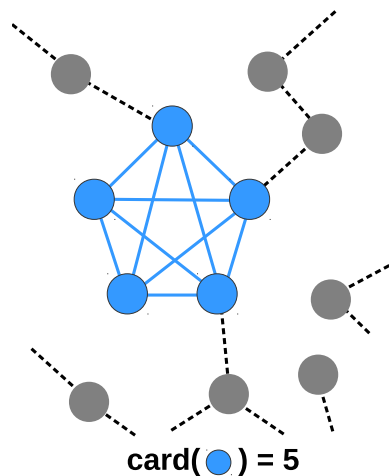
S'inspirer de la modularité

Calcul du nombre de nœuds attendus

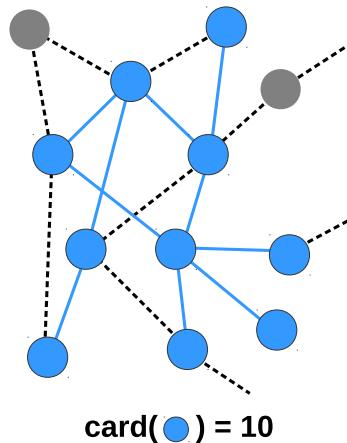
Premiers résultats et limitations

Idée du critère

Comparer le nombre de nœuds induits observés et celui attendu



➔
Mélange des arêtes



S'inspirer de la modularité

G : graphe, L_i : ensemble de *liens*

$V(L_i)$: nœuds induits dans G

- Qualité d'un groupe L_i :

$$Q_{in}(L_i) = \mathbb{E}[V(L_i)] - |V(L_i)|$$

- Qualité d'une partition \mathcal{P} :

$$Q_{in}(\mathcal{P}) = \sum_{L_i \in \mathcal{P}} Q_{in}(L_i)$$

Calcul du nombre de nœuds attendus

Graphe aléatoire avec la même distribution de degrés.

Tirage aléatoire et sans remise de $2|L_i|$ demi-arêtes.

B_u : variable aléatoire correspondant au nombre de fois où le nœud u est tiré, $B_u \sim \mathcal{G}(2|E|, d_G(u), 2|B|)$.

$$\mathbb{P}(B_u = 0) = \frac{\binom{2|E| - d_G(u)}{2m}}{\binom{2|E|}{2m}}$$

$$\mathbb{E}[|V_{in}(B)|] = \sum_{u \in V} 1 - \mathbb{P}(B_u = 0)$$

Algorithme glouton d'optimisation locale adapté aux liens

Algorithme agglomératif ayant deux mouvements :

- Fusion de 2 communautés.
- Changement de groupe d'une arête.

Situation initiale : chaque lien dans une communauté.

Condition d'arrêt : plus d'améliorations possibles.

Premiers résultats

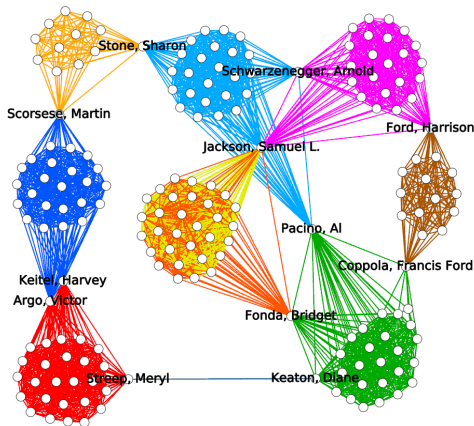
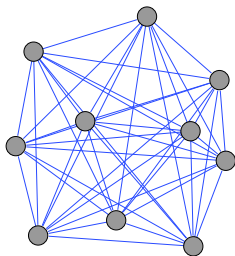


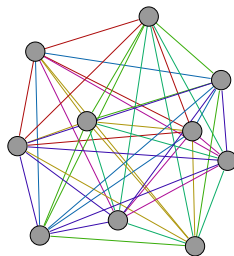
FIGURE: Réseaux d'acteurs et réalisateurs ayant travaillé ensemble.

Limitations du critère

Exemple de la clique :



$$Q_{in}(G) = 0$$



$$Q_{in}(G) = 0.398$$

Prise en compte du voisinage

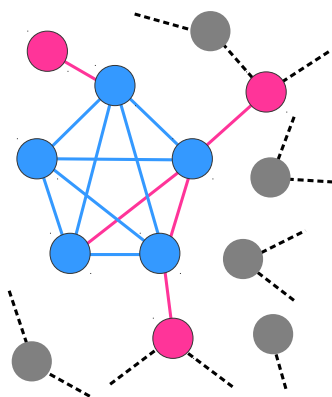
Idée de l'amélioration

Pénalisation du voisinage

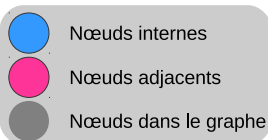
Premiers résultats et limitations

Idée de l'amélioration

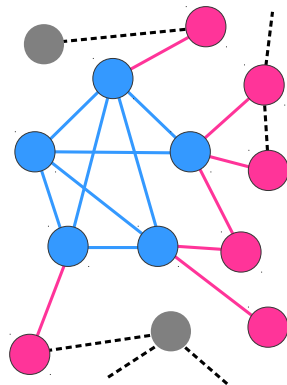
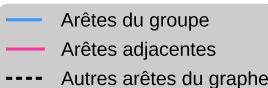
Comparer le nombre de nœuds adjacents observés et celui attendu



$$\text{card}(\text{pink}) = 3$$



➔
Mélange des arêtes
sauf internes



$$\text{card}(\text{pink}) = 6$$

Pénaliser un mauvais voisinage

$V_{out}(L_i)$: nœuds adjacent à L_i dans G .

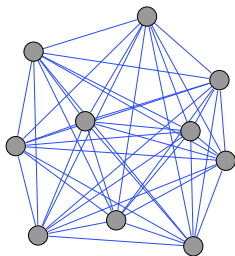
Calcul de la qualité en pénalisant un voisinage comportant plus de nœuds qu'attendus :

$$Q^*(L_i) = Q_{in}(L_i) - \max(0, |V_{out}(L_i)| - \mathbb{E}[|V_{out}(L_i)|]) \quad \text{où } L_i \subseteq E$$

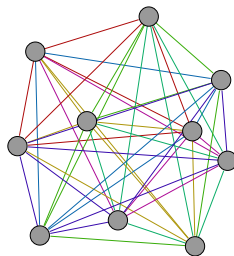
$\mathbb{E}[|V_{out}(L_i)|]$: espérance du nombre de nœuds adjacents à L_i .

Premiers résultats

Exemple de la clique :



$$Q^*(G) = 0$$



$$Q^*(G) = -0.043$$

Premiers résultats et limitations

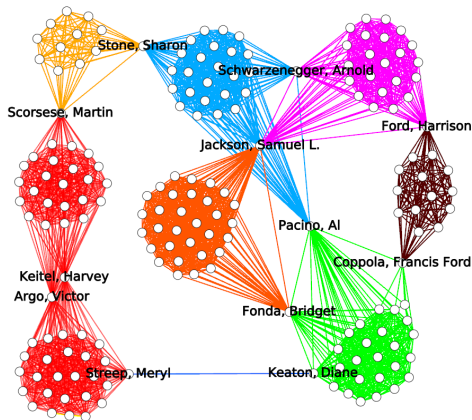


FIGURE: Réseaux d'acteurs et réalisateurs ayant travaillé ensemble.

Méthodes existantes

Auteurs	Y.-Y. Ahn et al.	T.S. Evans et al.
Avantages	Rapidité d'exécution	Utilisation des méthodes existantes
Désavantages	<ul style="list-style-type: none"> • Densité dans les graphes multiples ou valués ? • Capture d'étoiles dans l'exemple précédent 	<ul style="list-style-type: none"> • Utilisation du <i>line graph</i> • Impossible de capturer des petits groupes

Conclusion

- Étude des communautés de liens au lieu de nœuds.
- Transposition d'un problème non trivial.
- Définition d'un critère comparant le nombre de nœuds induits par rapport à la quantité attendue.
- Optimisation locale via un algorithme glouton.
- Problème toujours ouvert.

Perspectives

- Recherche d'autres critères.
- Tests quantitatifs plus complets et diversifiés des mesures.
- Extension aux réseaux temporels.

Merci de votre attention.

Des questions ?