

Identification de communautés locales dans les réseaux sociaux

Blaise NGONMANG

L2TI- UP13 et LIRIMA/UMMISCO - UY1

23 Mars 2012



Plan

- 1 Introduction
- 2 Détection de communautés globales
- 3 Identification de communautés locales
- 4 Méthode de l'ajout simultané et de la proximité
- 5 Méthode du balayage itératif
- 6 Conclusion
- 7 Bibliographie

Plan

- 1 Introduction
- 2 Détection de communautés globales
- 3 Identification de communautés locales
- 4 Méthode de l'ajout simultané et de la proximité
- 5 Méthode du balayage itératif
- 6 Conclusion
- 7 Bibliographie

Introduction

- De nombreux systèmes complexes peuvent être modélisés par des réseaux

Introduction

- De nombreux systèmes complexes peuvent être modélisés par des réseaux
- Un réseau est un ensemble de nœuds (individus) reliés entre eux par des liens (relations)

Introduction

- De nombreux systèmes complexes peuvent être modélisés par des réseaux
- Un réseau est un ensemble de nœuds (individus) reliés entre eux par des liens (relations)
- Un réseau est donc un graphe mais:

Introduction

- De nombreux systèmes complexes peuvent être modélisés par des réseaux
- Un réseau est un ensemble de nœuds (individus) reliés entre eux par des liens (relations)
- Un réseau est donc un graphe mais:
 - ▶ nœuds porteurs d'attributs

Introduction

- De nombreux systèmes complexes peuvent être modélisés par des réseaux
- Un réseau est un ensemble de nœuds (individus) reliés entre eux par des liens (relations)
- Un réseau est donc un graphe mais:
 - ▶ nœuds porteurs d'attributs
 - ▶ liens valués pouvant porter des données (messages)

Introduction

- De nombreux systèmes complexes peuvent être modélisés par des réseaux
- Un réseau est un ensemble de nœuds (individus) reliés entre eux par des liens (relations)
- Un réseau est donc un graphe mais:
 - ▶ nœuds porteurs d'attributs
 - ▶ liens valués pouvant porter des données (messages)
 - ▶ similarité entre deux nœuds = f (attributs, liens)

Introduction

- De nombreux systèmes complexes peuvent être modélisés par des réseaux
- Un réseau est un ensemble de nœuds (individus) reliés entre eux par des liens (relations)
- Un réseau est donc un graphe mais:
 - ▶ nœuds porteurs d'attributs
 - ▶ liens valués pouvant porter des données (messages)
 - ▶ similarité entre deux nœuds = f (attributs, liens)
 - ▶ caractéristiques du réseau (effet petit monde, loi de puissance, clustering...)

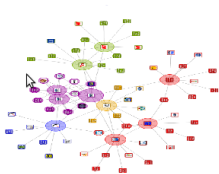
Introduction

- De nombreux systèmes complexes peuvent être modélisés par des réseaux
- Un réseau est un ensemble de nœuds (individus) reliés entre eux par des liens (relations)
- Un réseau est donc un graphe mais:
 - ▶ nœuds porteurs d'attributs
 - ▶ liens valués pouvant porter des données (messages)
 - ▶ similarité entre deux nœuds = f (attributs, liens)
 - ▶ caractéristiques du réseau (effet petit monde, loi de puissance, clustering...)
- Plusieurs types d'analyses peuvent être effectuées sur les réseaux parmi lesquelles la décomposition en communautés

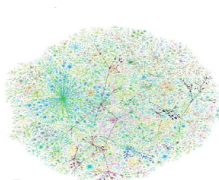
Introduction

- De nombreux systèmes complexes peuvent être modélisés par des réseaux
- Un réseau est un ensemble de nœuds (individus) reliés entre eux par des liens (relations)
- Un réseau est donc un graphe mais:
 - ▶ nœuds porteurs d'attributs
 - ▶ liens valués pouvant porter des données (messages)
 - ▶ similarité entre deux nœuds = f (attributs, liens)
 - ▶ caractéristiques du réseau (effet petit monde, loi de puissance, clustering...)
- Plusieurs types d'analyses peuvent être effectuées sur les réseaux parmi lesquelles la décomposition en communautés
- Nous nous intéressons ici à l'identification de la communauté locale d'un nœud.

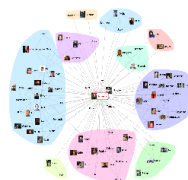
Exemples de réseaux



Pages Web



Routeurs Internet



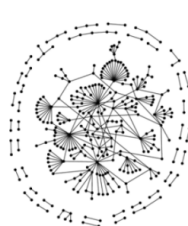
Facebook



Communications

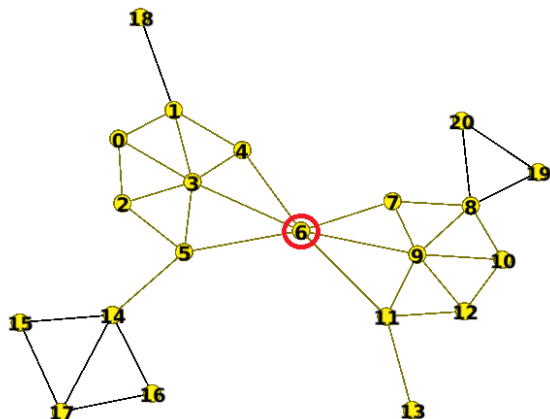


Citations

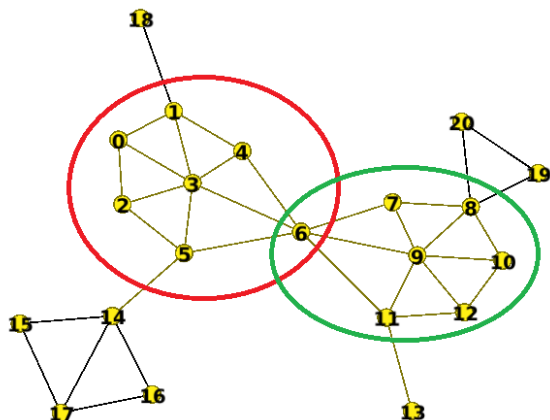


Biologie

Contexte



Contexte



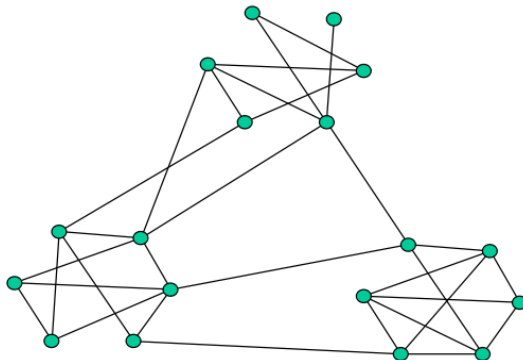
Plan

- 1 Introduction
- 2 Détection de communautés globales
 - Définitions
 - Méthodes séparatives
 - Méthodes agglomeratives
 - Communautés recouvrantes
- 3 Identification de communautés locales
- 4 Méthode de l'ajout simultané et de la proximité
- 5 Méthode du balayage itératif
- 6 Conclusion

Definitions

Définition liée à la topologie du réseau :

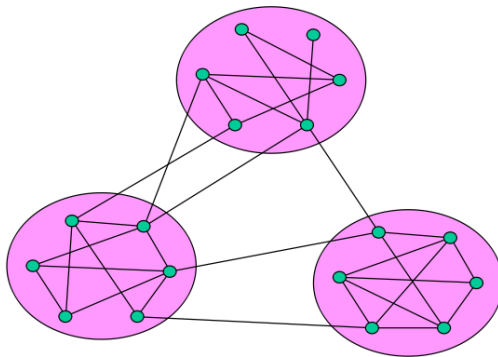
- Groupes de sommets densément connectés.
- Peu de liens entre les groupes.



Definitions

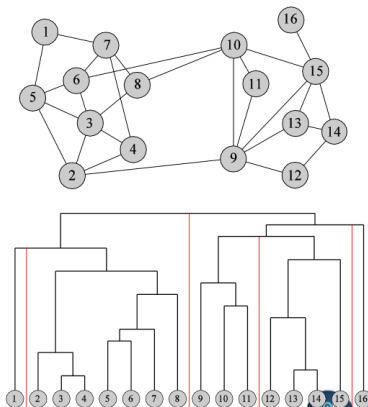
Définition liée à la topologie du réseau :

- Groupes de sommets densément connectés.
- Peu de liens entre les groupes.



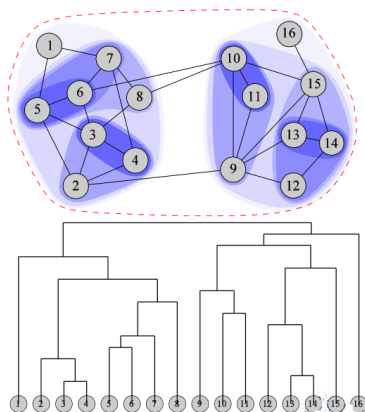
Méthodes séparatives

- Au départ on a une seule communauté
- A chaque étape on essaye de séparer les communautés
- Exemple algorithme de Girvan et Newman



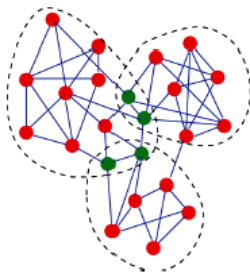
Méthodes agglomératives

- Au départ on a autant de communautés que de noeuds
- A chaque étape on choisit 2 communautés à regrouper
- Exemple algorithme de louvain



Communautes recouvrantes

- Dans la réalité un individu appartient à plusieurs communautés
- Un contact Facebook a par exemple des camarades de classe et les membres de sa famille
- Exemple de méthode : percolation de cliques



Discussion

- Santo Fortunato a proposé en 2010 une revue assez complète des algorithmes de détection de communautés dans les graphes

Discussion

- Santo Fortunato a proposé en 2010 une revue assez complète des algorithmes de détection de communautés dans les graphes
- Supposition: structure complète du graphe connue

Discussion

- Santo Fortunato a proposé en 2010 une revue assez complète des algorithmes de détection de communautés dans les graphes
- Supposition: structure complète du graphe connue
- Impossible pour les très grands graphes (Web, Facebook,...)

Discussion

- Santo Fortunato a proposé en 2010 une revue assez complète des algorithmes de détection de communautés dans les graphes
- Supposition: structure complète du graphe connue
- Impossible pour les très grands graphes (Web, Facebook,...)

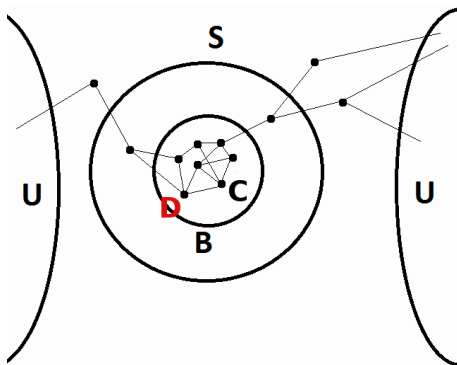
=> Recherche de communautés locales autour d'un nœud: utilisation exclusive d'informations locales

Plan

- 1 Introduction
- 2 Détection de communautés globales
- 3 Identification de communautés locales
 - Définition du problème
 - Fonction de qualité de Clauset
 - Fonction de qualité de Luo
 - Méthode de Chen
- 4 Méthode de l'ajout simultané et de la proximité
- 5 Méthode du balayage itératif
- 6 Conclusion

Définitions

- On part d'un nœud de départ
- A chaque instant on a une vue locale du réseau.
- Les nœuds sont évalués un à un.



Définition

Algorithme: identification de communautés locales

Données: Un graphe R et un nœud de départ n_0 .

Résultat: Un sous-graphe D qui représente la communauté locale de n_0 .

Initialiser D avec n_0

Initialiser B avec n_0

Initialiser C avec l'ensemble vide

Initialiser S avec les voisins de n_0

$Q = 0$

Répéter

pour chaque $n_i \in S$ **faire**

calculer la qualité de n_i par rapport à D

fin pour

Déterminer le n_i^* de qualité Q^* qui maximise la qualité en départageant aléatoirement les ex-aequo.

Si $Q^* > Q$ **alors**

Ajouter n_i^* à D et le retirer de S .

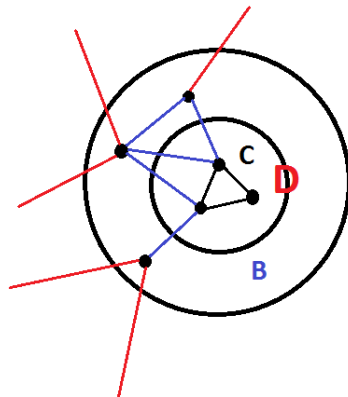
Mettre à jour B , S , C .

Fin si

Jusqu'à ($Q^* \leq Q$)

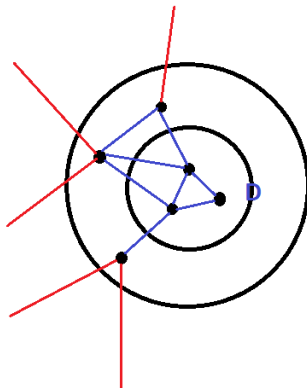
Fonction de qualité de Clauset

- Idée: Les nœuds situés à la Bordure de la communauté doivent avoir plus de liens avec **les autres membres** de leur communauté qu'avec **l'extérieur**.
- Fonction de qualité: $R = \frac{B_{in}}{B_{in} + B_{out}}$

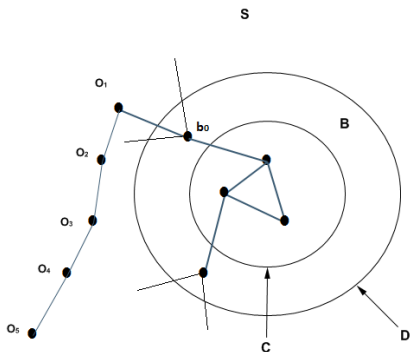


Fonction de qualité de Luo

- Idée: Le rapport des **liens internes** sur les **liens externes** doit être grand (supérieur à 1).
- Fonction de qualité: $M = \frac{D_{in}}{D_{out}}$
- On a des communautés plus cohésives

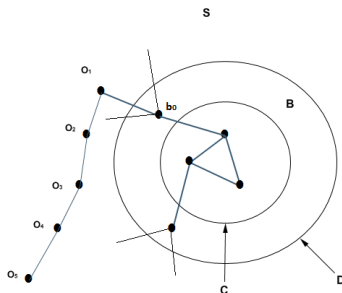


- Limite: avec R et M, et l'algorithme présenté, des chaines de nœuds reliés entre eux sont ajoutés systématiquement.



Méthode de Chen

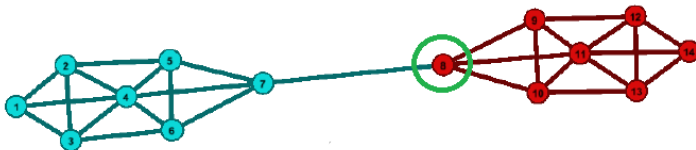
- Idée: insérer un nœud uniquement si $L'_{in} > L_{in}$ et $L'_{ex} < L_{ex}$
- Fonction de qualité : $L = \frac{L_{in}}{L_{ex}}$
- $L_{in} = \frac{D_{in}}{|D|}$ et $L_{ex} = \frac{B_{out}}{|B|}$
- Ajout d'une phase de confirmation pour maintenir ou non les nœuds



- L'ajout de O_1 fait augmenter L_{in} mais sans changer L_{ex} .

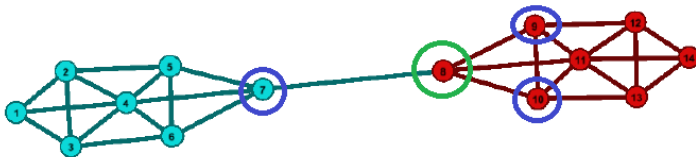
Problème de la Méthode de Chen

- On Recherche la communauté du nœud 8.



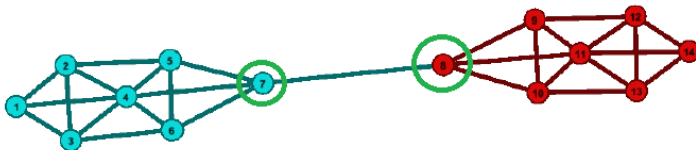
Problème de la Méthode de Chen

- Les nœuds 7, 9 et 10 maximisent la fonction de qualité



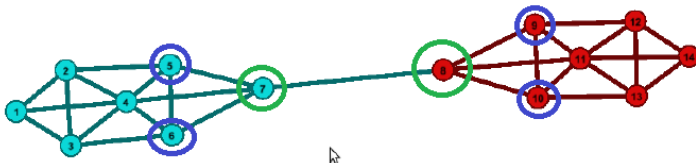
Problème de la Méthode de Chen

- Le choix aléatoire est porté sur le nœud 7.



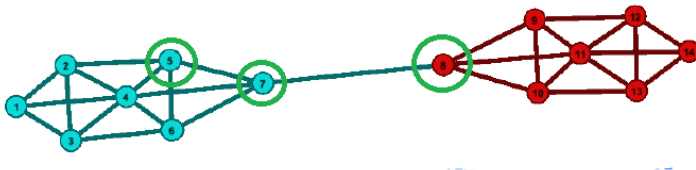
Problème de la Méthode de Chen

- Les nœuds 5, 6, 9 et 10 maximisent la fonction de qualité



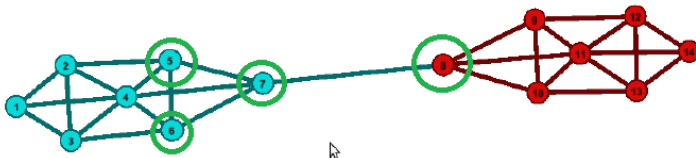
Problème de la Méthode de Chen

- Le choix aléatoire est porté sur le nœud 5.



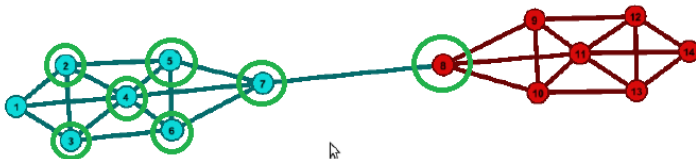
Problème de la Méthode de Chen

- Les nœuds de gauche sont tour à tour ajoutés



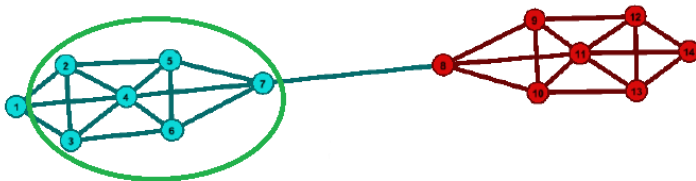
Problème de la Méthode de Chen

- Les nœuds de gauche sont tour à tour ajoutés



Problème de la Méthode de Chen

- Le 8 nœud est expulsé.

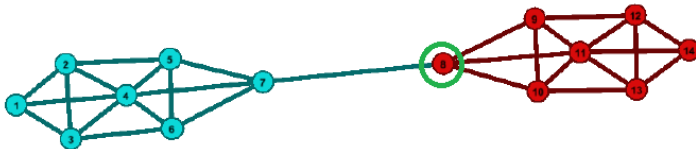


Plan

- 1 Introduction
- 2 Détection de communautés globales
- 3 Identification de communautés locales
- 4 Méthode de l'ajout simultané et de la proximité**
- 5 Méthode du balayage itératif
- 6 Conclusion
- 7 Bibliographie

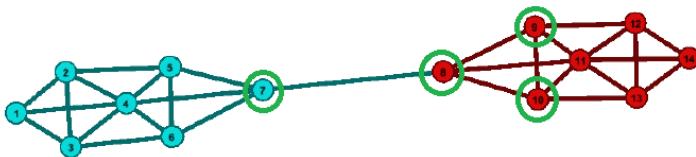
Ajout simultané

- Idée: modification de la méthode de Chen. Considérer l'ensemble D pour L_{ex} et lorsque plusieurs nœuds maximisent la fonction de qualité on les ajoute **simultanément**
- Rechercher les communautés recouvrantes sur le sous-graphe résultat



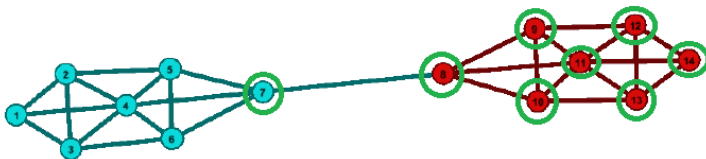
Ajout simultané

- Idée: modification de la méthode de Chen. Considérer l'ensemble D pour L_{ex} et lorsque plusieurs nœuds maximisent la fonction de qualité on les ajoute **simultanément**
- Rechercher les communautés recouvrantes sur le sous-graphe résultat



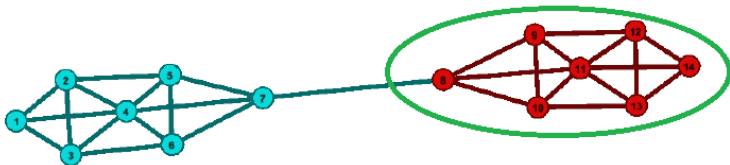
Ajout simultané

- Idée: modification de la méthode de Chen. Considérer l'ensemble D pour L_{ex} et lorsque plusieurs nœuds maximisent la fonction de qualité on les ajoute **simultanément**
- Rechercher les communautés recouvrantes sur le sous-graphe résultat



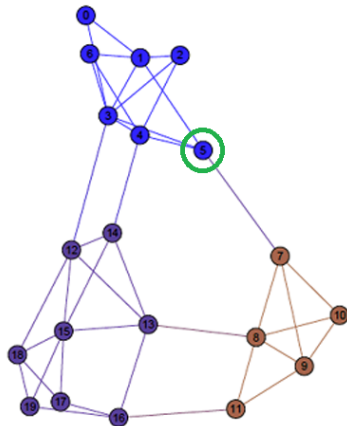
Ajout simultané

- Idée: modification de la méthode de Chen. Considérer l'ensemble D pour L_{ex} et lorsque plusieurs nœuds maximisent la fonction de qualité on les ajoute **simultanément**
- Rechercher les communautés recouvrantes sur le sous-graphe résultat



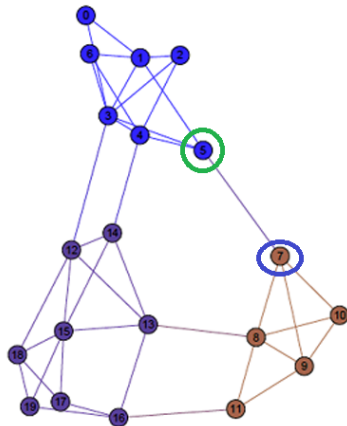
Ajout simultané

- Il y a des cas où l'algorithme ne trouvera pas de communauté



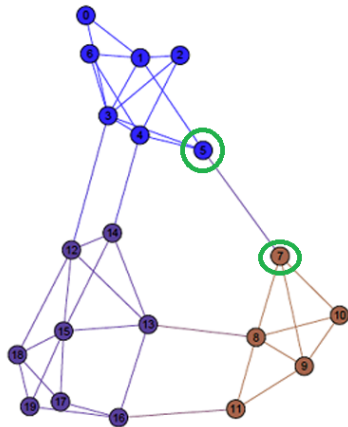
Ajout simultané

- Il y a des cas où l'algorithme ne trouvera pas de communauté



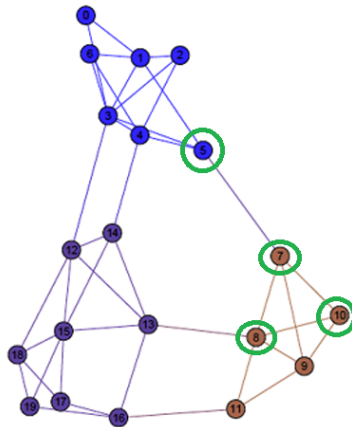
Ajout simultané

- Il y a des cas où l'algorithme ne trouvera pas de communauté



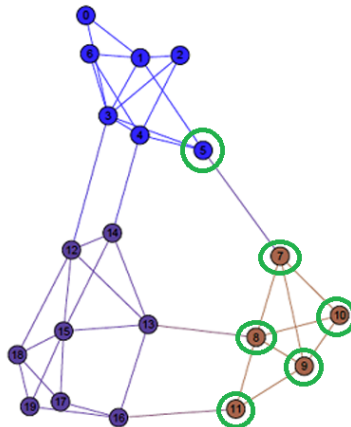
Ajout simultané

- Il y a des cas où l'algorithme ne trouvera pas de communauté



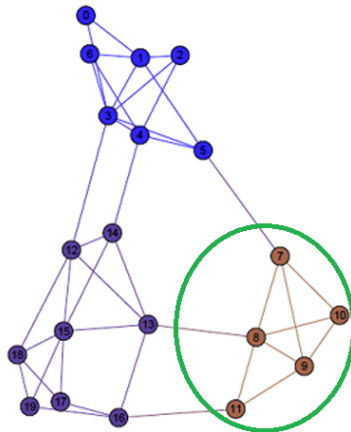
Ajout simultané

- Il y a des cas où l'algorithme ne trouvera pas de communauté



Ajout simultané

- Il y a des cas où l'algorithme ne trouvera pas de communauté

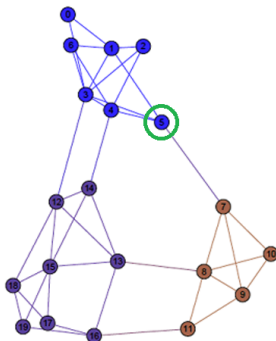


Proximité

- Idée: un nœud est plus susceptible d'appartenir à la communauté du nœud de départ si leur degré de séparation est faible.

- Fonction de qualité:

$$T_{in} = \sum_{i \in D} \frac{|\Gamma(i) \cap D|}{(1+d_i)} \text{ et } T_{ex} = \sum_{i \in D} |\Gamma(i) \cap S|(1+d_i)$$

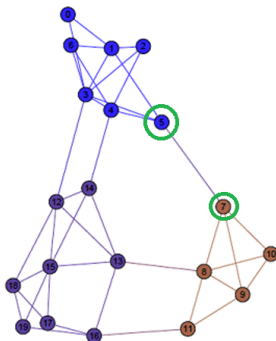


Proximité

- Idée: un nœud est plus susceptible d'appartenir à la communauté du nœud de départ si leur degré de séparation est faible.

- Fonction de qualité:

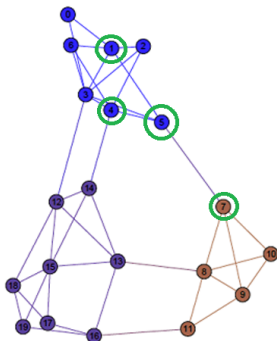
$$T_{in} = \sum_{i \in D} \frac{|\Gamma(i) \cap D|}{(1+d_i)} \text{ et } T_{ex} = \sum_{i \in D} |\Gamma(i) \cap S|(1+d_i)$$



Proximité

- Fonction de qualité:

$$T_{in} = \sum_{i \in D} \frac{|\Gamma(i) \cap D|}{(1+d_i)} \text{ et } T_{ex} = \sum_{i \in D} |\Gamma(i) \cap S|(1+d_i)$$

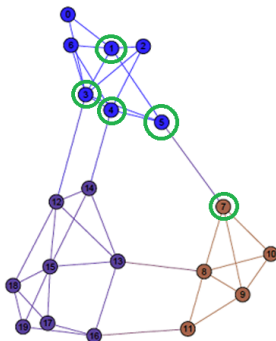


Proximité

- Idée: un nœud est plus susceptible d'appartenir à la communauté du nœud de départ si leur degré de séparation est faible.

- Fonction de qualité:

$$T_{in} = \sum_{i \in D} \frac{|\Gamma(i) \cap D|}{(1+d_i)} \text{ et } T_{ex} = \sum_{i \in D} |\Gamma(i) \cap S|(1+d_i)$$

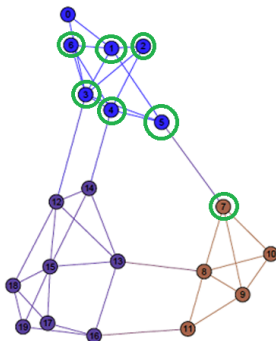


Proximité

- Idée: un nœud est plus susceptible d'appartenir à la communauté du nœud de départ si leur degré de séparation est faible.

- Fonction de qualité:

$$T_{in} = \sum_{i \in D} \frac{|\Gamma(i) \cap D|}{(1+d_i)} \text{ et } T_{ex} = \sum_{i \in D} |\Gamma(i) \cap S|(1+d_i)$$

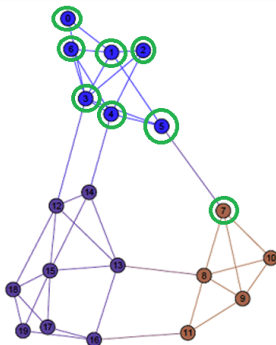


Proximité

- Idée: un nœud est plus susceptible d'appartenir à la communauté du nœud de départ si leur degré de séparation est faible.

- Fonction de qualité:

$$T_{in} = \sum_{i \in D} \frac{|\Gamma(i) \cap D|}{(1+d_i)} \text{ et } T_{ex} = \sum_{i \in D} |\Gamma(i) \cap S|(1+d_i)$$

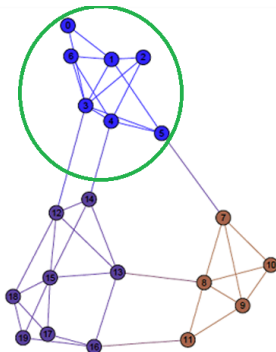


Proximité

- Idée: un nœud est plus susceptible d'appartenir à la communauté du nœud de départ si leur degré de séparation est faible.

- Fonction de qualité:

$$T_{in} = \sum_{i \in D} \frac{|\Gamma(i) \cap D|}{(1+d_i)} \text{ et } T_{ex} = \sum_{i \in D} |\Gamma(i) \cap S|(1+d_i)$$



Comparaison avec Chen

- Jeux de données NCAA Football 2000.
- 115 équipes organisées en conférences
- Structure de communautés connue : les conférences.

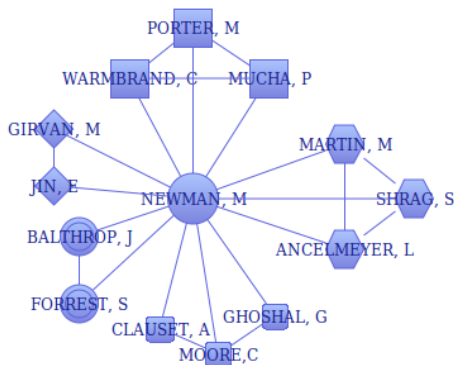
2000 NCAA Football	Résultats des algorithmes	
	Algorithme de Chen	Notre algorithme
Pas de Communautés	41(35,65%)	0(0%)
Précision	0,98	0,93
Rappel	0,82	0,90
F-Mesure	0,87	0,91

Identification de sous-communautés recouvrantes

- Très souvent la communauté locale d'un individu contient plusieurs sous-structures
- Ces sous-structures présentent parfois des recouvrements
- Les amis Facebook par exemple peuvent être regroupés en famille, collègues, ...
- Un collègue peut aussi être de la famille

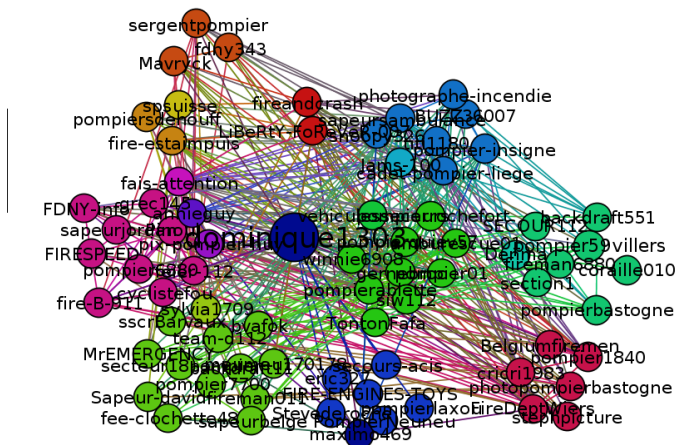
Sous-communautés recouvrantes (Exemple 1)

- Exemple sur jeu de données réel.
- Dans le graphe de co-auteurs dans le domaine des grands réseaux par exemple la communauté locale de M.E.J. Newman comporte plusieurs (5) sous-communautés.



Sous-communautés recouvrantes (Exemple 2)

- Exemple sur jeu de données réel.
- Jeux de données Skyrock (5 millions de noeuds). Communauté de Pompiers en majorité.



Plan

- 1 Introduction
- 2 Détection de communautés globales
- 3 Identification de communautés locales
- 4 Méthode de l'ajout simultané et de la proximité
- 5 Méthode du balayage itératif**
- 6 Conclusion
- 7 Bibliographie

Contexte

- L'algorithme précédent nécessite d'utiliser un algorithme de recherche de communautés recouvrantes

Contexte

- L'algorithme précédent nécessite d'utiliser un algorithme de recherche de communautés recouvrantes
- La complexité en temps des algorithmes de recherche de communautés recouvrantes actuels est élevée

Contexte

- L'algorithme précédent nécessite d'utiliser un algorithme de recherche de communautés recouvrantes
- La complexité en temps des algorithmes de recherche de communautés recouvrantes actuels est élevée
- Nous proposons une méthode qui permet d'identifier directement les communautés locales recouvrantes

Algorithme

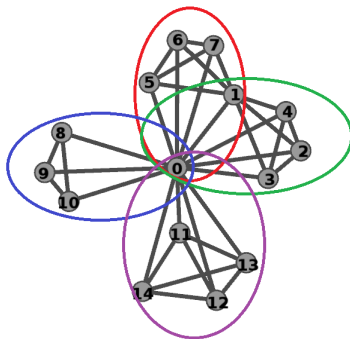
- Idée: lorsqu'un nœud se fait expulser lors de la recherche d'une communauté on vérifie s'il est membre de cette sous-communauté

Algorithme

- Idée: lorsqu'un nœud se fait expulser lors de la recherche d'une communauté on vérifie s'il est membre de cette sous-communauté
- On supprime les liens internes de la communauté trouvée

Algorithme

- Idée: lorsqu'un nœud se fait expulser lors de la recherche d'une communauté on vérifie s'il est membre de cette sous-communauté
- On supprime les liens internes de la communauté trouvée
- On fait un balayage itératif de toutes les communautés voisines et on garde celles auxquelles le nœud de départ est "assez connecté"



Evaluation

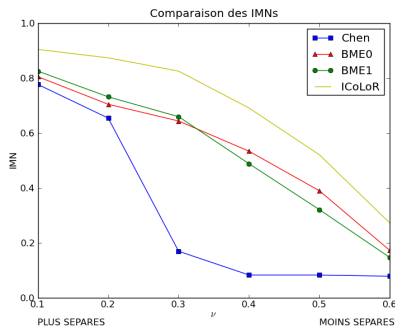
- IMN: Information Mutuelle Normalisée

Evaluation

- IMN: Information Mutuelle Normalisée
- L'IMN permet d'évaluer les communautés obtenues

Evaluation

- IMN: Information Mutuelle Normalisée
- L'IMN permet d'évaluer les communautés obtenues
- Le Benchmark de Fortunato permet de générer des réseaux sociaux réalistes



Evaluation

- Jeu de données de co-occurrences d'achats sur Amazon
- Un lien existe entre deux produit s'ils sont fréquemment achetés ensemble

Table: Communauté locale élargie du livre "Merlin Trilogy"

Num.	Membres	description
1	The Wicked Day (Arthurian Saga, Book 4) The Last Enchantment (Arthurian Saga, Bk. 3.) The Crystal Cave (Arthurian Saga, Book 1)	Saga du roi Arthur
2	The Sword in the Stone The Book of Merlyn The Once and Future King	Merlin et Arthur

Applications

- Recommandation d'amis dans les réseaux sociaux : On recherche la communauté locale d'un individu et on lui propose comme nouveaux amis tous les membres de sa communauté locale avec lesquels il n'a pas encore de liens.

Applications

- Recommandation d'amis dans les réseaux sociaux : On recherche la communauté locale d'un individu et on lui propose comme nouveaux amis tous les membres de sa communauté locale avec lesquels il n'a pas encore de liens.
- Recommandation d'articles (news, produits, ...) dans les réseaux sociaux: On propose un membre d'une communauté des articles en fonction des choix des autres membres de la communauté.

Applications

- Recommandation d'amis dans les réseaux sociaux : On recherche la communauté locale d'un individu et on lui propose comme nouveaux amis tous les membres de sa communauté locale avec lesquels il n'a pas encore de liens.
- Recommandation d'articles (news, produits, ...) dans les réseaux sociaux: On propose un membre d'une communauté des articles en fonction des choix des autres membres de la communauté.
- Propagation d'épidémies (ou diffusion d'information) : La communauté locale d'un individu infecté donne une idée sur les personnes qu'il pourrait contaminer dans un réseau de contact (cercle d'influence).

Plan

- 1 Introduction
- 2 Détection de communautés globales
- 3 Identification de communautés locales
- 4 Méthode de l'ajout simultané et de la proximité
- 5 Méthode du balayage itératif
- 6 Conclusion**
- 7 Bibliographie

Conclusion et perspectives

Perspectives

- Caractériser les communautés découvertes
- Étudier la dynamique des communautés
- Utiliser le résultat pour faire de la recommandation

Publications

Le travail de ce mémoire a donné lieu aux publications suivantes:

- Blaise NGONMANG, Maurice TCHUENTE, Emmanuel VIENNET.
Identification de communautés locales dans les réseaux sociaux. AGS May 2011.
- Blaise NGONMANG, Maurice TCHUENTE, Emmanuel VIENNET.
Recherche de communautés locales. FGG Juillet 2011.
- Blaise NGONMANG, Maurice TCHUENTE, Emmanuel VIENNET.
Local community identification in social networks. PPL 2012.

Financement

Ce travail a été partiellement financé par les projets:

- ANR Ex DEUSS
- DG-CIS CEDRES
- POFS du Fonds Francophone des Inforoutes.

Plan

- 1 Introduction
- 2 Détection de communautés globales
- 3 Identification de communautés locales
- 4 Méthode de l'ajout simultané et de la proximité
- 5 Méthode du balayage itératif
- 6 Conclusion
- 7 Bibliographie**

Bibliographie

- ① Blaise NGONMANG, Maurice TCHUENTE, Emmanuel VIENNET. Identification de communautés locales dans les réseaux sociaux. AGS May 2011.
- ② A.Clauset. Finding local communities in network. Physical Review, 2005: 026132.
- ③ Bagrow, J. P. Evaluating local community methods in networks. J.STAT.MECH, 2008.
- ④ Bollt, J. P. Bagrow and E. M. Local method for detecting communities. Physical Review, 2005: 233-239.
- ⑤ Jiyang Chen, Osmar R. Zaine and Randy Goebel. Local Community Identification in Social Networks. ASONAM, 2009.
- ⑥ Santo Fortunato. Community detection in graphs. Physics Reports 486, 75-174 (2010)
- ⑦ Blondel Vincent D., Guillaume Jean-Loup, Lambiotte Renaud, Lefebvre Etienne. Fast unfolding of communities in large networks. Journal of Statistical Mechanics: Theory and Experiment, Issue 10, pp. 10008 (2008).
- ⑧ The structure and function of complex networks, M. E. J. Newman, SIAM Review 45, 167-256 (2003).

Merci

- MERCI POUR VOTRE AIMABLE ATTENTION
- Questions?

