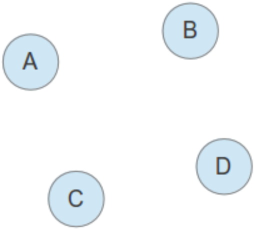
Dynamique d'un graphe

Igor Colin

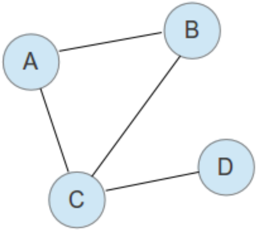
18 février 2014

- Données de départs : graphe utilisateurs

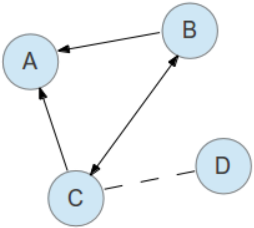
- Données de départs : graphe utilisateurs
- Nœuds : utilisateurs



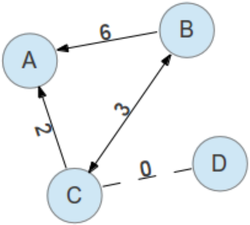
- Données de départs : graphe utilisateurs
- Nœuds : utilisateurs
- Arrêtes :
    - relations d'amitié



- Données de départs : graphe utilisateurs
- Nœuds : utilisateurs
- Arrêtes :
  - relations d'amitié
  - influences



- Données de départs : graphe utilisateurs
- Nœuds : utilisateurs
- Arrêtes :
  - relations d'amitié
  - influences
  - fréquences de partage



Plusieurs objectifs :

- Obtenir les paramètres du graphe (influence, sémantique)
- Comprendre la structure du graphe (communautés, dynamique)
- Agir sur le graphe

- Définition d'une mesure de qualité d'une partition d'un graphe
- Maximisation de cette quantité pour trouver la partition optimale

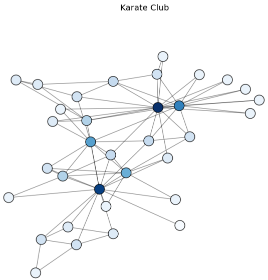


FIGURE: Exemple du club de karaté de Zachary.

Approche spectrale

Définition

Soit $G = (V, E)$ un graphe non-orienté et soit (S, T) une partition de V. La modularité associée à la partition (S, T) correspond taux d'arrêtes de E contenues dans S ou T comparé au taux d'arrêtes qui auraient été contenues dans S ou T si l'on avait distribué les arrêtes du graphe aléatoirement.

- Définition de la modularité dépendante de l'aléa autorisé sur les arrêtes
- Approche usuelle : degrés des nœuds conservés

Approche spectrale

Définition

Soit $G = (V, E)$ un graphe non-orienté et soit (S, T) une partition de V. On note $\mathbf{A}$ la matrice d'adjacence associée à G. Sous les conditions précédentes, la modularité $Q(S, T)$ peut-être exprimée sous la forme suivante :

$$Q(S, T) = \frac{1}{2|E|} \sum_{(i,j) \in V \times V} \left(A_{ij} - \frac{d_i d_j}{2|E|} \right) \mathbf{1}_{\{s_i = s_j\}},$$

où, pour tout $k \in V$,

$$s_k = \begin{cases} +1 & si\ k \in S \\ -1 & si\ k \in T \end{cases}$$

Approche spectrale

Théorème

Soit $G = (V, E)$ un graphe non-orienté. Soit $\mathbf{u}^{(2)}$ le vecteur propre associé à la deuxième plus grande valeur propre $\lambda^{(2)}$ de la matrice de Laplace normalisée. Si $\lambda^{(2)} > 0$, alors la partition (S, T) définie par :

$$\left\{ \begin{array}{rcl} S & = & \{i \in V,\, u_i^{(2)} > 0\} \\ T & = & \{i \in V,\, u_i^{(2)} \leq 0\} \end{array} \right. ,$$

est solution non triviale du problème de maximisation de la modularité.

Approche spectrale

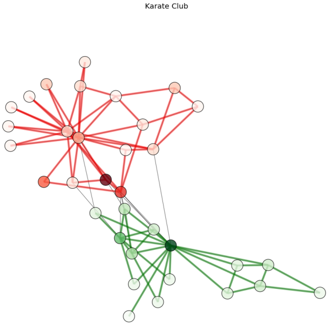
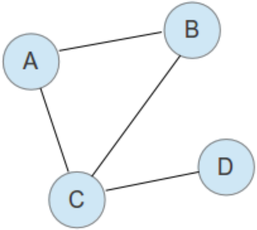


FIGURE: Détection de communauté appliquée au club de karaté de Zachary.

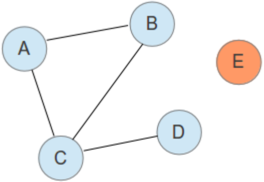
Approche spectrale

- Méthode simple à implémenter
- Critère interprétable
- Condition d'arrêt
- Limites : coût en $\mathcal{O}(|V|^3)$, matrice de Laplace de taille $|V| \times |V|$
- Nécessité d'utiliser une méthode adaptée pour de grands réseaux (méthodes stochastiques)

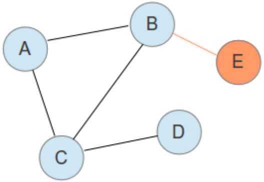
- Graphe d'utilisateurs dynamique : $(G_t = (V_t, E_t))_{t \in \mathcal{T}}$



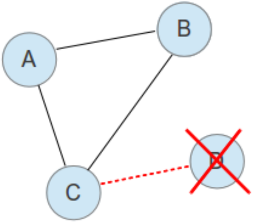
- Graphe d'utilisateurs dynamique : $(G_t = (V_t, E_t))_{t \in \mathcal{T}}$
    - Nouveaux utilisateurs



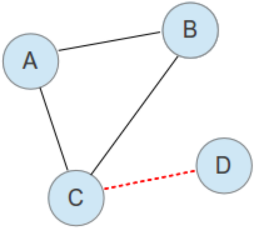
- Graphe d'utilisateurs dynamique : $(G_t = (V_t, E_t))_{t \in \mathcal{T}}$
  - Nouveaux utilisateurs
  - Nouvelles connexions entre utilisateurs



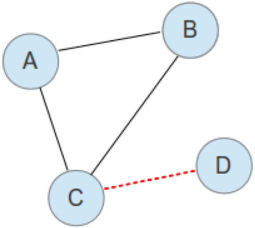
- Graphe d'utilisateurs dynamique : $(G_t = (V_t, E_t))_{t \in \mathcal{T}}$
    - Nouveaux utilisateurs
    - Nouvelles connexions entre utilisateurs
    - Départ d'utilisateurs



- Graphe d'utilisateurs dynamique : $(G_t = (V_t, E_t))_{t \in \mathcal{T}}$
  - Nouveaux utilisateurs
  - Nouvelles connexions entre utilisateurs
  - Départ d'utilisateurs
  - Disparition de connexions



- Graphe d'utilisateurs dynamique : $(G_t = (V_t, E_t))_{t \in \mathcal{T}}$
  - Nouveaux utilisateurs
  - Nouvelles connexions entre utilisateurs
  - Départ d'utilisateurs
  - Disparition de connexions
  - Évolution des connexions existantes (influence dynamique)



Peu de modèles existants sur graphes dynamiques :

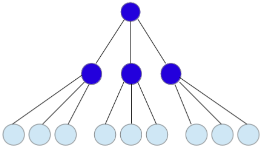
- Épidémiologie (théorie de la survie)
- Génétique (détection de *pattern*)
- Réseaux sociaux, web (citations, Internet)

Community Guided Attachment

- Observation sur des graphes réels (réseaux sociaux, informatiques)
  - Le degré moyen augmente quand le réseau grandit : $|E_t| \propto |V_t|^{\alpha}$, avec $1 < \alpha < 2$
- Possible explication : connexions entre nœuds dépendantes des communautés

Community Guided Attachment

- Structure des communautés en arbre (communautés dans les communautés)
- Distance *h* entre nœuds : distance dans l'arbre



Community Guided Attachment

- Nouveau nœud n_1 se rattache à un nœud n_2 avec une probabilité $c^{d(n_1,n_2)}$
- Sous ce modèle, si arbre homogène de degré b, alors $|E| \propto |V|^{2-\log_b(c)}$

- CGA donne une interprétation réaliste à un phénomène observé
- Mais ne se penche sur un seul aspect : pas d'effet *small world*, pas de créations d'arrêtes parmi nœuds existants, etc.
- Pas de modélisation temporelle
- Nécessité de mixer les modèles

Approche générale

- Objectif : prédire l'évolution du graphe
- Définir un modèle d'évolution dynamique
- Utiliser l'historique des évolutions pour retrouver les paramètres du modèle
- Mettre à jour ces paramètres en ligne ou à intervalles déterminés