

Détection de communautés et génération de graphes

Igor Colin

14 février 2014

1 Détection de communautés

2 Génération de graphe

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

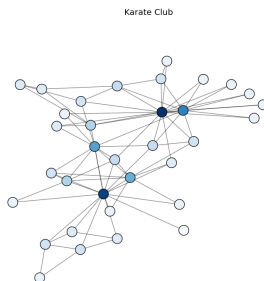


FIGURE: Exemple du club de karaté de Zachary.

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 au 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 dépendantes de l'aléa autorisé sur les arrêtes
- Approche usuelle : degrés des nœuds conservés

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 $k \in V$,

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

Définition

Soit $G = (V, E)$ un graphe non-orienté et soit \mathbf{A} la matrice d'adjacence associée. On note $\mathbf{D} = \text{diag}((d_i)_{i \in V})$. La matrice de Laplace normalisée \mathbf{L} associée au graphe G est définie par :

$$\mathbf{L} = \mathbf{D}^{-1/2} \mathbf{A} \mathbf{D}^{-1/2}.$$

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 :

$$\begin{cases} S &= \{i \in V, u_i^{(2)} > 0\} \\ T &= \{i \in V, u_i^{(2)} \leq 0\} \end{cases},$$

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

Karate Club

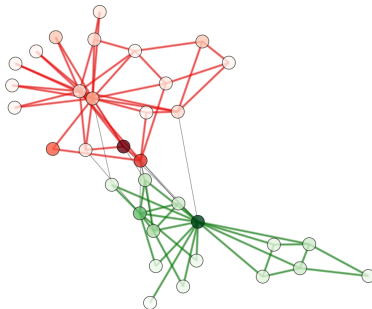


FIGURE: Détection de communauté appliquée au club de karaté de Zachary.

- 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 randomisées

1 Détection de communautés

2 Génération de graphe

- Actuellement, peu d'utilisateurs sur l'application
- Nécessité de créer des graphes plus grands pour tester les méthodes employées
- Comment générer des graphes capable de « prédire » le futur de l'application ?

- MCMC
 - Switching : degré, degré joint
 - Matching
 - Garantie d'indépendance
- Social, épidémiologique, etc.
 - Attachement préférentiel
 - Kronecker
 - Forest fire