Détection de communautés et génération de graphes

Igor Colin

11 février 2014

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

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 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 $k \in V$,

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

Définition

Soit $G = (V, E)$ un graphe non-orienté et soit $\mathbf{A}$ la matrice d'adjacence associée. On note $\mathbf{D} = \mathbf{diag}\left((d_i)_{i \in V}\right)$. 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 de la matrice de Laplace normalisée. 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.