Community Detection Algorithms I

Leonid E. Zhukov

National Research University Higher School of Economics

12.03.2012

Community Detection Algorithms

Edge Betweenness, Newman-Girvan, 2004

Edge betweenness -number of shortest paths $\sigma_{st}(e)$ going through edge e

$$C_B(e) = \sum_{s \neq t} \frac{\sigma_{st}(e)}{\sigma_{st}}$$

Algorithm: Edge Betweenness

Data: adjacency matrix **A**

repeat

  compute edge betweenness $C_B(e)$;

  remove edge with largest betweenneess;

until *no edges left* ;

If bi-partition, then stop when graph splits in two components

Community Detection Algorithms

Spectral Modularity Maximization (Newman, 2006)

Modularity:

$$Q = \frac{1}{2m} \sum_{ij} \left(A_{ij} - \frac{k_i k_j}{2m} \right) \delta(c_i, c_j)$$

c_i- class, $\delta(c_i, c_j)$- kronecker delta
For two classes C_1, C_2 indicator variable $s = \pm 1$

$$\delta(c_i, c_j) = \frac{1}{2}(s_i s_j + 1) = \begin{cases} 1 : & i, j \in C_1 \text{ or } i, j \in C_2 \\ 0 : & i \in C_1, j \in C_2 \text{ or } i \in C_2, j \in C_1 \end{cases}$$

Community Detection Algorithms

$$Q = \frac{1}{4m} \sum_{ij} \left(A_{ij} - \frac{k_i k_j}{2m} \right) (s_i s_j + 1) = \frac{1}{4m} \sum_{i,j} B_{ij} s_i s_j$$

where

$$B_{ij} = A_{ij} - \frac{k_i k_j}{2m}$$

Qudratic form:

$$Q(\mathbf{s}) = \frac{1}{4m} \mathbf{s}^T \mathbf{B} \mathbf{s}$$

Let $\mathbf{B}\mathbf{u}_i = \lambda_i \mathbf{u}_i$, then

$$Q(\mathbf{s}) = \sum_i (\mathbf{u}_i^T \mathbf{s})^2 \lambda_i$$

for $\max(Q(\mathbf{s}))$ select $\mathbf{s} \parallel \mathbf{u}_1$, where $\lambda_1 = \lambda_{max}$

Community Detection Algorithms

Algorithm: Spectral Moduldarity Maximization

Data: adjacency matrix $\mathbf{A}$

Result: class indicator vector $\mathbf{s}$

compute $\mathbf{k} = \mathit{degree}(\mathbf{A})$;

compute $\mathbf{B} = \mathbf{A} - \frac{1}{2m}\mathbf{k}\mathbf{k}^T$;

solve for maximal eigenevector $\mathbf{B}\mathbf{u} = \lambda\mathbf{u}$;

set $\mathbf{s} = \mathit{sign}(\mathbf{u}_1)$