Network Communities

Leonid E. Zhukov

National Research University Higher School of Economics

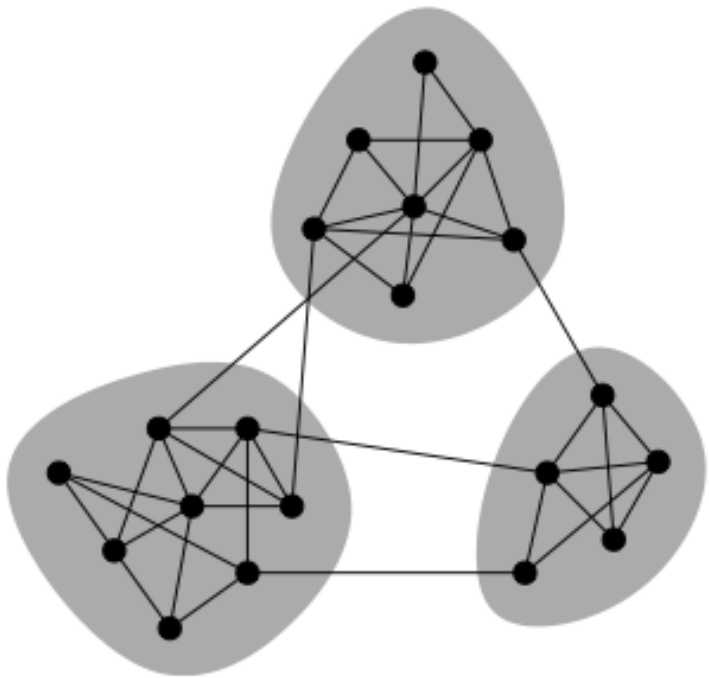
05.03.2012

Network Communities

Community detection / Graph partitioning/ Dividing into clusters

Definition

Division of network vertices into groups according to the pattern of the edges such that vertices inside the group are connected with many more edges than between groups.



Network Communities

- средняя плостнось в графе

$$\rho = \frac{m}{n(n-1)/2}$$

- плотность связей в кластере

$$\delta_{int}(C) = \frac{m_c}{n_c(n_c-1)/2}$$

- с внешними узлами

$$\delta_{ext}(C) = \frac{m_{ext}}{n_c(n-n_c)}$$

- кластер: $\delta_{int} > \rho$, $\delta_{ext} < \rho$

$$\max(\delta_{int} - \delta_{ext})$$

Network Communities

Graph partitioning: $G(E, V)$

$||E|| = m$, $||V|| = n$, $V = V_1 + V_2$

- min cut

$$Q = cut = \sum_{i \in V_1, j \in V_2} e_{ij}$$

- quotent cut

$$Q = \frac{cut(V_1, V_2)}{||V_1||} + \frac{cut(V_1, V_2)}{||V_2||}$$

- normalized cut

$$Q = \frac{cut(V_1, V_2)}{\sum_{i \in V_1, j \in V} e_{ij}} + \frac{cut(V_1, V_2)}{\sum_{i \in V_2, j \in V} e_{ij}}$$

Network Communities

Graph Partitioning:

bi-partition: $n = n_1 + n_2$, combinations $= (n_1!n_2!)/n!$,

Algorithms:

- Kernighan-Lin
- Spectral
- Flow
- Multilevel

Community detection:

- Betweenness (Newman-Girvin)
- Modularity (spectral)

Modularity:

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

Network Communities

Nodes structural similarity matrix:

- cosine similarity

$$M_{ij} = \cos(A_i, A_j)$$

- correlation matrxix (Pearson correlation)

$$M_{ij} = r_{ij} = \frac{cov(A_i, A_j)}{\sqrt{var(A_i)}\sqrt{var(A_j)}}$$

Clustering:

- Agglomerative
- K-means