Network Communities

Leonid E. Zhukov

School of Applied Mathematics and Information Science
National Research University Higher School of Economics

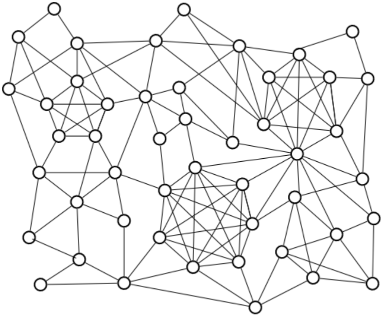
07.03.2013

НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ УНИВЕРСИТЕТ

Cliques, plexes and cores

Definition

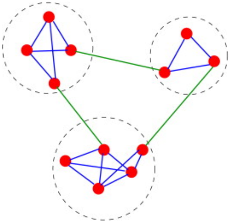
A *clique* is a maximal subset of the vertices such that every member of the subset is connected by an edge to every other



Cliques, plexes and cores

Definition

A *k-plex of size n* is a maximal subset of n vertices such that each vertex is connected to no fewer than $n - k$ of the others.



clique is a k-plex with $k = 1$

Cliques, plexes and cores

Definition

A *k-core* is a maximal subset vertices such that each vertex is connected to at least k others in subset

k-core is an $(n-k)$-plex of n vertices

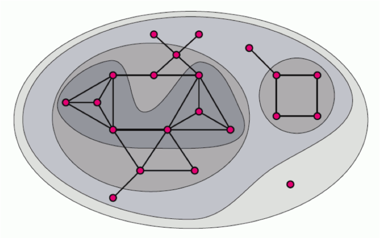


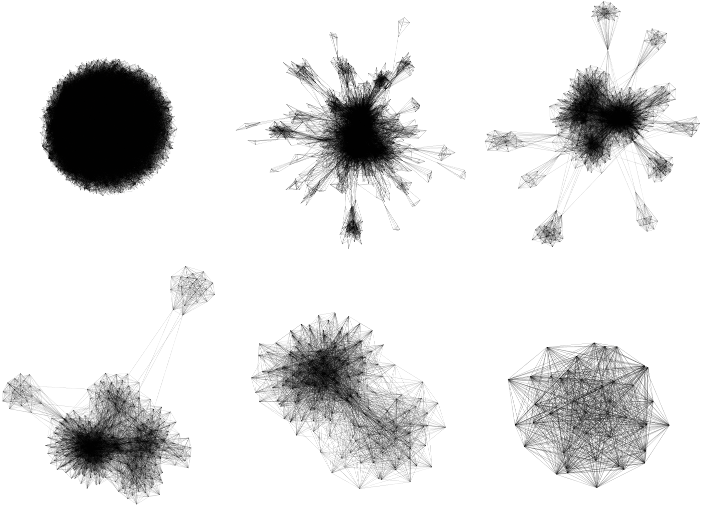
Figure: 0-core, 1-core, 2-core, 3-core

Every vertex in k-core has a degree $k_i \geq k$

K-core

- Tree: 1-core
- Cycle: 2-core
- *n*-Clique: $n - 1$- core

k-core

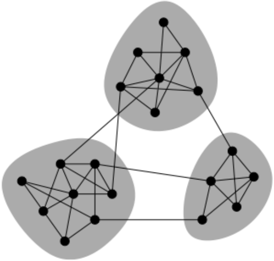


k-core: k = 2,5,10,15,20,31

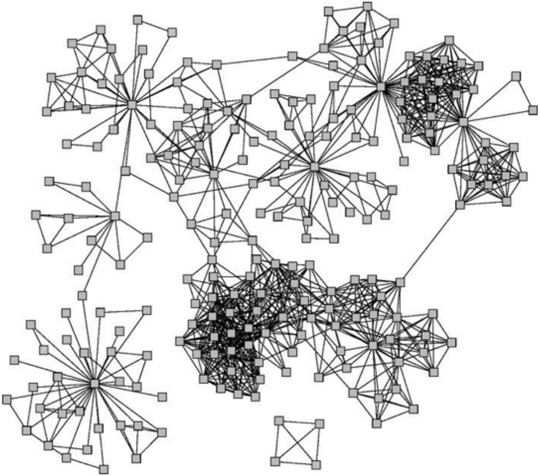
Network Communities

Definition

Network community is a group of verticies such that vertices inside the group are connected with many more edges than between groups.



Network Communities



- graph density

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

- community (cluster) density

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

- external edges density

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

- community (cluster): $\delta_{int} > \rho$, $\delta_{ext} < \rho$
- cluster detection

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

Graph partitioning

Graph: $G(E, V)$

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

- 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

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

Graph partitioning

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

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

References

- An O(m) Algorithms for Cores Decomposition of Networks. V. Batagelj, M. Zaversnik, 2002
- Community detection in graphs, S. Fortunato, 2010