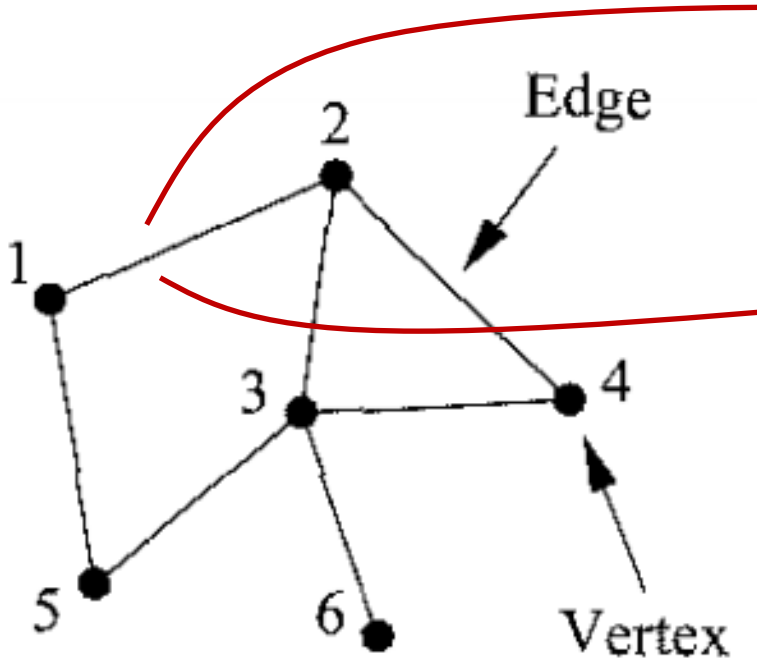
Computational Social Science
Week 9

Samuel Martin-Gutierrrez

Informatics faculty, TU Wien
Complexity Science Hub Vienna

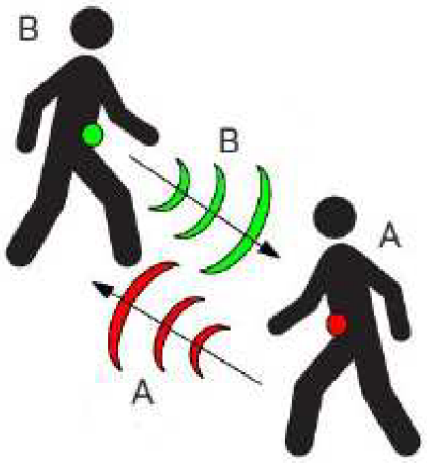
e-mail: samuel.martin-gutierrez@tuwien.ac.at
web: www.networkinequality.com

Adjacency matrix

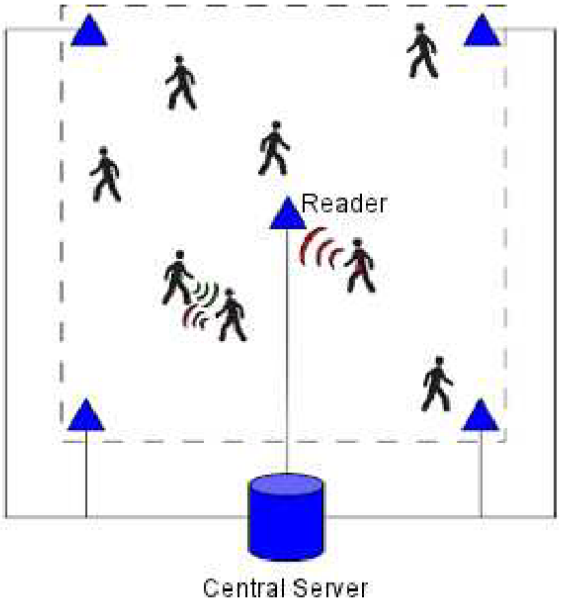


$$\mathbf{A} = \begin{pmatrix} 0 & 1 & 0 & 0 & 1 & 0 \\ 1 & 0 & 1 & 1 & 0 & 0 \\ 0 & 1 & 0 & 1 & 1 & 1 \\ 0 & 1 & 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \end{pmatrix}$$

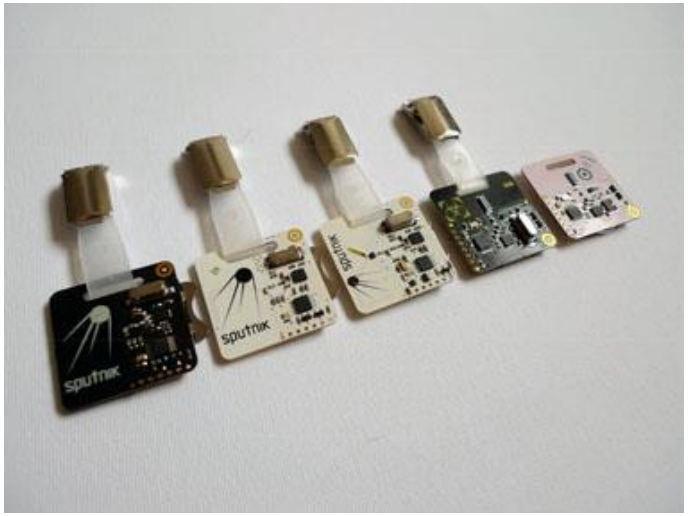
Dataset



(a) When two user meet, they exchange their Id.



(b) Data on tags are periodically broadcasted to readers.



Dataset

tij_Thiers13.dat - Notepad

File Edit Format View Help

```
43220 454 640
43220 1 939
43220 185 258
43220 55 170
43220 9 453
43220 9 45
43220 14 190
43220 400 637
43220 255 275
43220 176 533
43220 116 533
43220 151 866
43220 280 484
43220 243 687
```

t Node1 Node 2

metadata_Thiers13.dat - Notepad

File Edit Format View Help

```
650     2BIO1
498     2BIO1
627     2BIO1
857     2BIO1
487     2BIO1
28      2BIO1
927     2BIO1
72      2BIO1
400     2BIO1
945     2BIO1
```

Node Group

Degree centrality

- Popularity?
- Friendliness?
- Sociability?

$$x_i = \sum_j A_{ij}$$

Eigenvector centrality

A node is important if its neighbors are.

$$x_i' = \sum_j A_{ij} x_j$$

$$\mathbf{x}' = \mathbf{A}\mathbf{x}$$

Eigenvectors of A

$$\mathbf{x}(t) = \mathbf{A}^t \mathbf{x}(0) \qquad \mathbf{x}(0) = \sum_i c_i \mathbf{v}_i$$

$$\mathbf{x}(t) = \mathbf{A}^t \sum_i c_i \mathbf{v}_i = \sum_i c_i \kappa_i^t \mathbf{v}_i = \kappa_1^t \sum_i c_i \left[\frac{\kappa_i}{\kappa_1}\right]^t \mathbf{v}_i$$

Eigenvalues of A

$k_1 > k_2 > \cdots > k_n$

$$\mathbf{x}(t) \longrightarrow c_1 \kappa_1^t \mathbf{v}_1$$

$$\mathbf{A}\mathbf{x} = \kappa_1 \mathbf{x}$$

PageRank centrality

PageRank of site A

$$PR(A) = \frac{1-d}{N} + d\left(\frac{PR(B)}{L(B)} + \frac{PR(C)}{L(C)} + \frac{PR(D)}{L(D)} + \cdots\right)$$

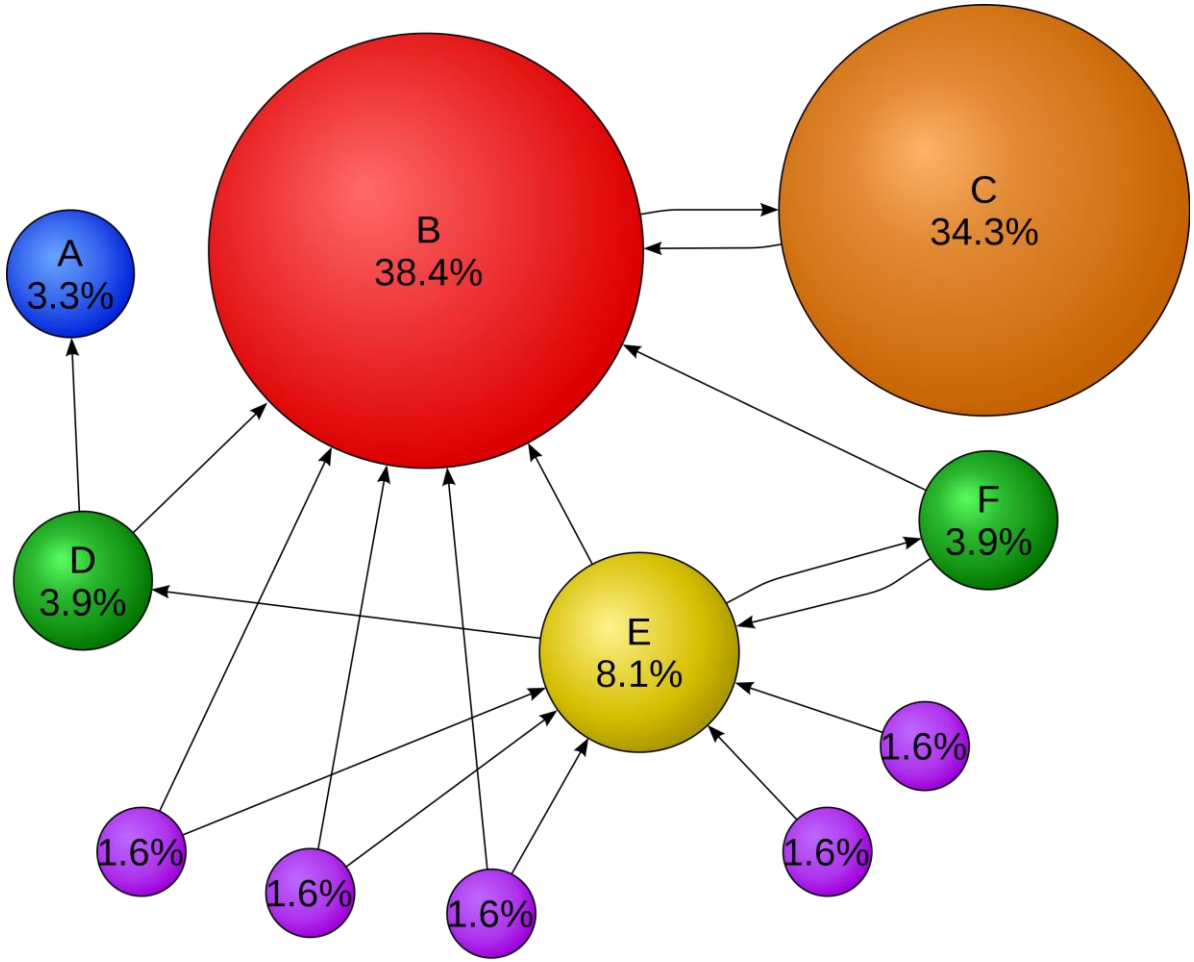
Number of out-Links in page B

$$PR(p_i) = \frac{1-d}{N} + d\sum_{p_j \in M(p_i)} \frac{PR(p_j)}{L(p_j)}$$

Set of pages that link to p_i

$$x_i = \alpha \sum_j A_{ij} \frac{x_j}{k_j^{out}} + \beta$$

PageRank centrality



Closeness centrality

How close is a node to everyone else.

$$C_i = \frac{1}{\ell_i} = \frac{n}{\sum_j d_{ij}}$$

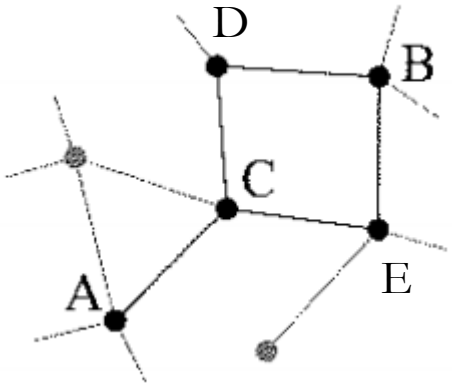
Betweenness centrality

How many shortest paths pass through a node: central nodes are bridges.

$$x_i = \sum_{st} \frac{n_{st}^i}{g_{st}}$$

Total number of geodesic paths between *s* and *t* that go through *i*

Total number of geodesic paths between *s* and *t*



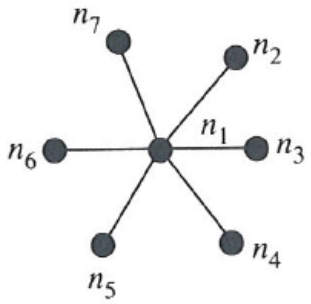
There are 2 geodesic paths between A and B.

- C is in both.
- D and E are in one each.

Centrality and Prestige in Undirected Social Graphs [Wasserman Faust 1994]

degree = closeness = betweenness centrality:

n1>n2,n3,n4,n5,n6,n7

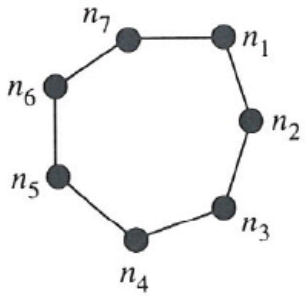


(a) Star graph

$$\begin{matrix} 0 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 \end{matrix}$$

degree= Betweeness centrality = Closeness centrality:

n1=n2=n3=n4=n5=n6=n7

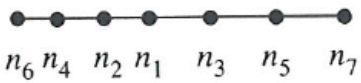


(b) Circle graph

$$\begin{matrix} 0 & 1 & 0 & 0 & 0 & 0 & 1 \\ 1 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 1 \\ 1 & 0 & 0 & 0 & 0 & 1 & 0 \end{matrix}$$

Betweeness centrality:

n1>n2,n3>n4,n5>n6,n7



(c) Line graph

$$\begin{matrix} 0 & 1 & 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 \end{matrix}$$

Fig. 5.1. Three illustrative networks for the study of centrality and prestige

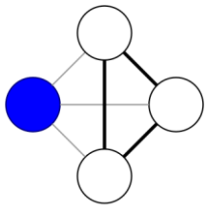
Clustering coefficient

Investigating friends of friends.

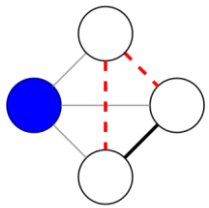
Number of closed triangles

$$C_i = \frac{\text{(number of pairs of neighbors of } i \text{ that are connected)}}{\text{(number of pairs of neighbors of } i\text{)}}$$

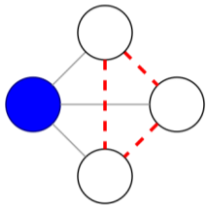
$$\frac{1}{2} k_i (k_i - 1)$$



c = 1

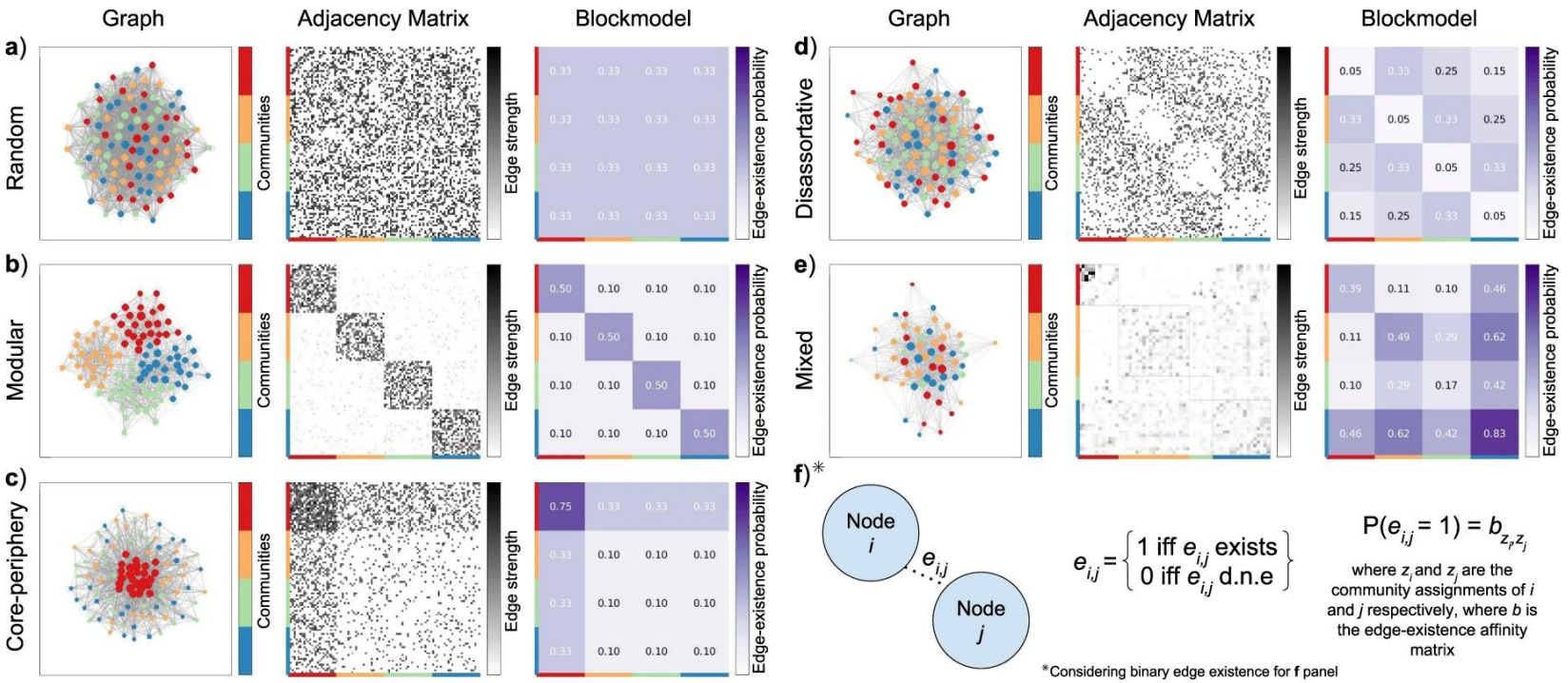


c = 1/3



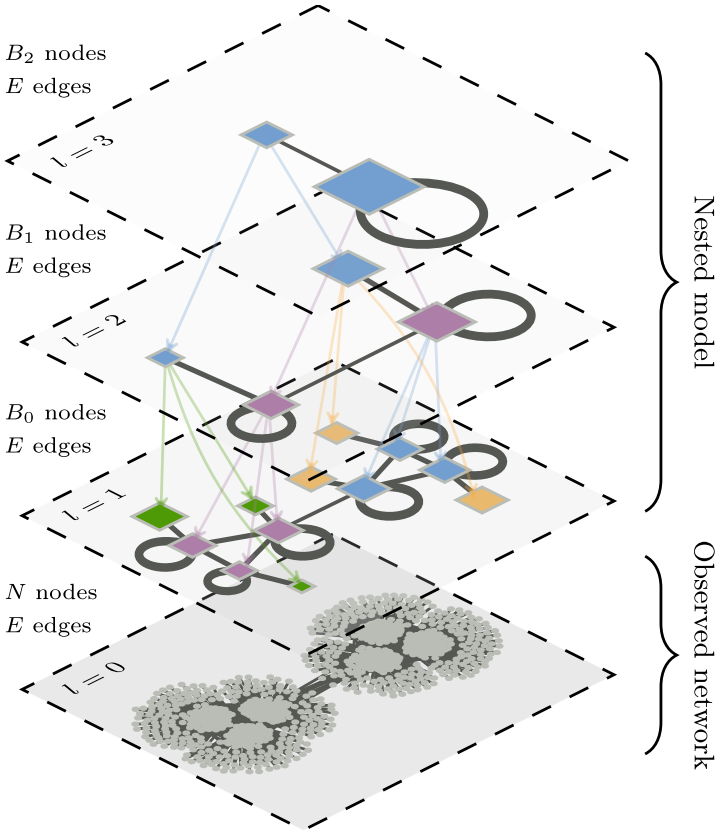
c = 0

Community detection with Stochastic Block Models (SBM)

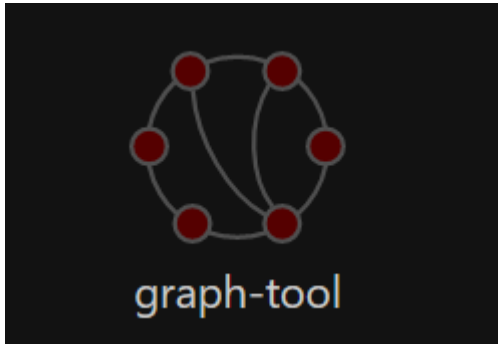


Scientific Reports volume 8, Article number: 12997 (2018)

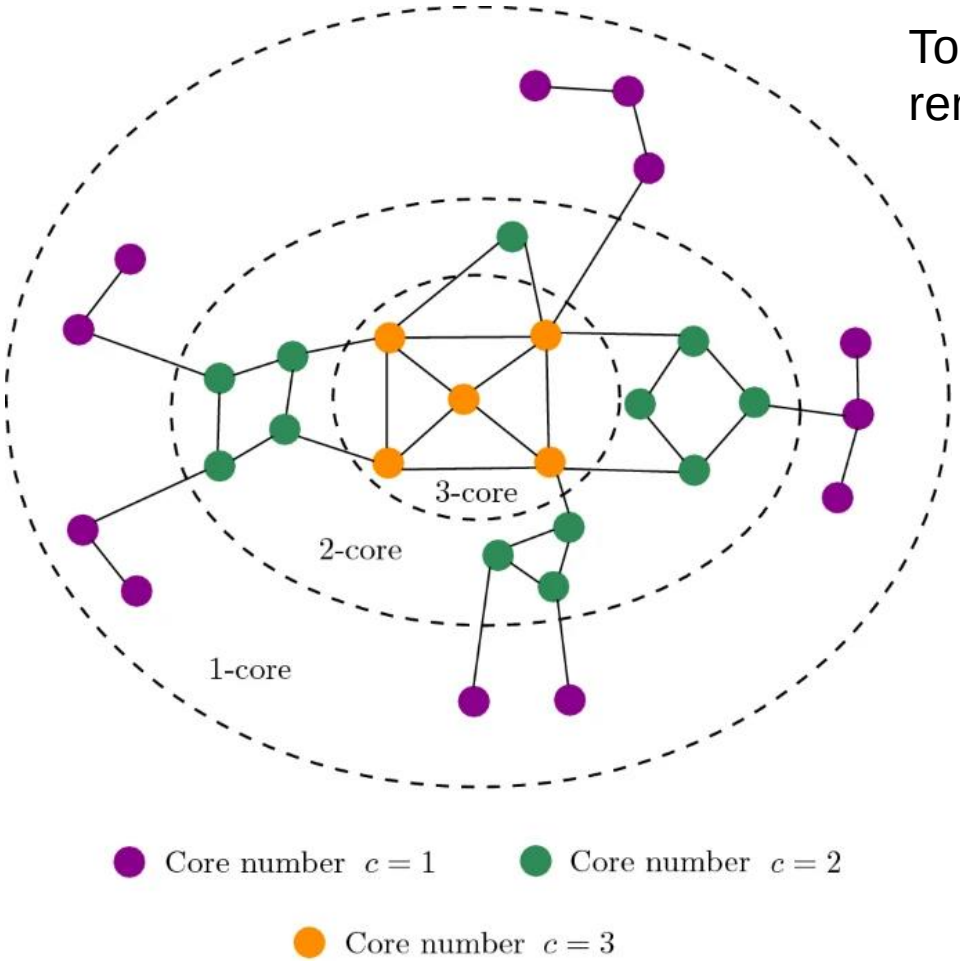
Community detection with Stochastic Block Models (SBM)



Minimum Description Length =
Quality of fit – **Complexity** of the model



K-core decomposition



To find the k-core iteratively remove all nodes of degree $\leq$ k

The VLDB Journal volume 29, pages 61–92 (2020)

Miniproject 1

- Distribute in pairs.
- Find an interesting dataset.
- Analyze it using the techniques we have seen (plus any other technique that you find relevant).
- Write a report (max. 2000 words) presenting the dataset, explaining the methodology you have used and why you chose to use it, and discuss the implications of the results in terms of the context of the data.
- Submit the dataset, the code, and the report.

Data sources

http://www.sociopatterns.org/

https://networks.skewed.de/

https://snap.stanford.edu/data/

Beware of large datasets! They are cool but may be difficult to deal with using NetworkX.