

# Network Structures

Leonid E. Zhukov

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

10.03.2014

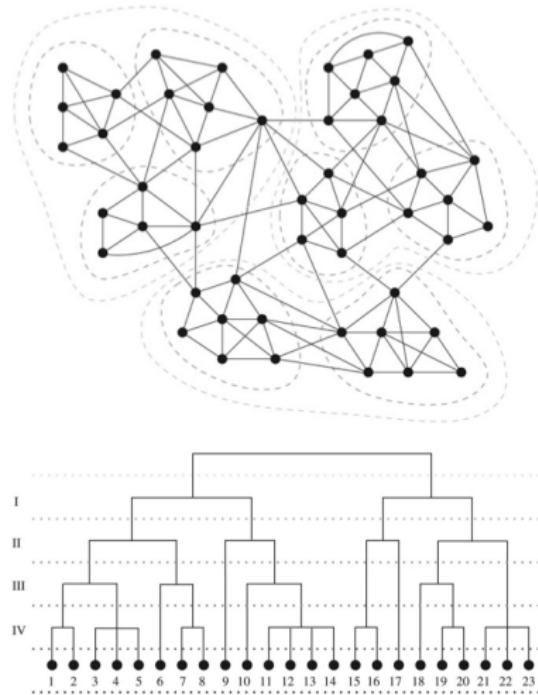


НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ  
УНИВЕРСИТЕТ

# Community detection algorithms

- Global methods (divisive):
  - Sparse cuts: min cut, normalized cut, min conductance
  - Modularity optimization: spectral, direct
  - Heuristics: edge betweenness
  - Randomized cuts
- Local methods: random walks
- Flat clustering and hierarchical clustering
- Hierarchical clustering: top down, bottom up

# Hierachical communities



Schaeffer, 2007

# Community detection algorithms

Author	Ref.	Label	Order
Eckmann & Moses	(Eckmann and Moses, 2002)	EM	$O(m(k^2))$
Zhou & Lipowsky	(Zhou and Lipowsky, 2004)	ZL	$O(n^3)$
Latapy & Pons	(Latapy and Pons, 2005)	LP	$O(n^3)$
Clauset et al.	(Clauset <i>et al.</i> , 2004)	NF	$O(n \log^2 n)$
Newman & Girvan	(Newman and Girvan, 2004)	NG	$O(nm^2)$
Girvan & Newman	(Girvan and Newman, 2002)	GN	$O(n^2 m)$
Guimerà et al.	(Guimerà and Amaral, 2005; Guimerà <i>et al.</i> , 2004)	SA	parameter dependent
Duch & Arenas	(Duch and Arenas, 2005)	DA	$O(n^2 \log n)$
Fortunato et al.	(Fortunato <i>et al.</i> , 2004)	FLM	$O(m^3 n)$
Radicchi et al.	(Radicchi <i>et al.</i> , 2004)	RCCLP	$O(m^4 / n^2)$
Donetti & Muñoz	(Donetti and Muñoz, 2004, 2005)	DM/DMN	$O(n^3)$
Bagrow & Boltt	(Bagrow and Boltt, 2005)	BB	$O(n^3)$
Capocci et al.	(Capocci <i>et al.</i> , 2005)	CSCC	$O(n^2)$
Wu & Huberman	(Wu and Huberman, 2004)	WH	$O(n + m)$
Palla et al.	(Palla <i>et al.</i> , 2005)	PK	$O(\exp(n))$
Reichardt & Bornholdt	(Reichardt and Bornholdt, 2004)	RB	parameter dependent

Author	Ref.	Label	Order
Girvan & Newman	(Girvan and Newman, 2002; Newman and Girvan, 2004)	GN	$O(nm^2)$
Clauset et al.	(Clauset <i>et al.</i> , 2004)	Clauset et al.	$O(n \log^2 n)$
Blondel et al.	(Blondel <i>et al.</i> , 2008)	Blondel et al.	$O(m)$
Guimerà et al.	(Guimerà and Amaral, 2005; Guimerà <i>et al.</i> , 2004)	Sim. Ann.	parameter dependent
Radicchi et al.	(Radicchi <i>et al.</i> , 2004)	Radicchi et al.	$O(m^4 / n^2)$
Palla et al.	(Palla <i>et al.</i> , 2005)	Cfinder	$O(\exp(n))$
Van Dongen	(Dongen, 2000a)	MCL	$O(nk^2)$ , $k < n$ parameter
Rosvall & Bergstrom	(Rosvall and Bergstrom, 2007)	Infomod	parameter dependent
Rosvall & Bergstrom	(Rosvall and Bergstrom, 2008)	Infomap	$O(m)$
Donetti & Muñoz	(Donetti and Muñoz, 2004, 2005)	DM	$O(n^3)$
Newman & Leicht	(Newman and Leicht, 2007)	EM	parameter dependent
Ronhovde & Nussinov	(Ronhovde and Nussinov, 2009)	RN	$O(m^\beta \log n)$ , $\beta \sim 1.3$

# Graph clustering

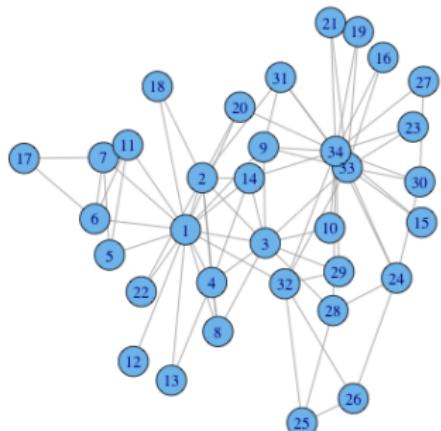
- Similarity based clustering
- Vertex structural equivalence (blockmodeling)
- Vertex approximate equivalence
  - Cosine similarity (vectors in  $n$ -dim space)

$$\sigma(v_i, v_j) = \cos(v_i, v_j) = \frac{v_i \cdot v_j}{\|v_i\| \|v_j\|} = \frac{\sum_k A_{ik} A_{kj}}{\sqrt{\sum A_{ik} A_{ki}} \sqrt{\sum A_{jk} A_{kj}}} = \frac{n_{ij}}{\sqrt{k_i k_j}}$$

- Jaccard similarity

$$J(v_i, v_j) = \frac{|\mathcal{N}(v_i) \cap \mathcal{N}(v_j)|}{|\mathcal{N}(v_i) \cup \mathcal{N}(v_j)|}$$

# Graph clustering



Graph



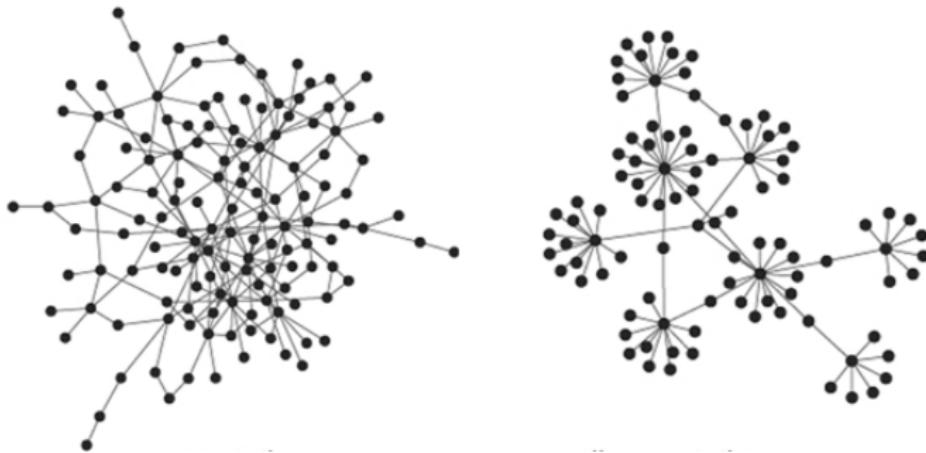
Similarity matrix

# Graph clustering

Clustering algorithms based on vertex similarity

- Agglomerative clustering
- k-means
- Spectral clustering etc

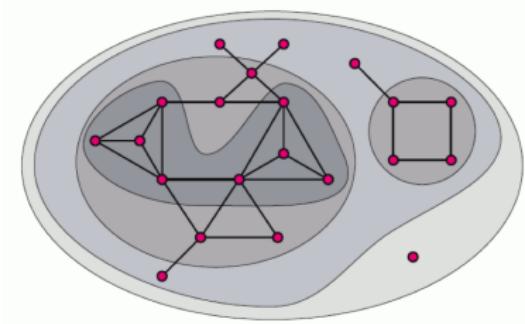
# Graph cores



# Graph cores

## Definition

A  $k$ -core is the largest subgraph such that each vertex is connected to at least  $k$  others in subset



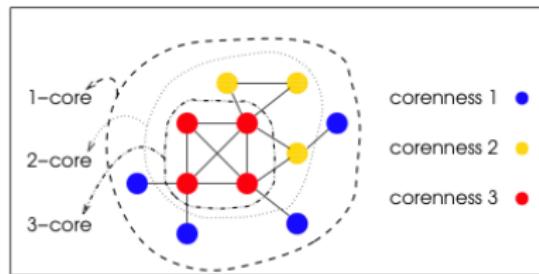
Every vertex in  $k$ -core has a degree  $k_i \geq k$

$(k+1)$ -core is always subgraph of  $k$ -core

Batagelj, 2003

# Finding k-cores

- The core number of a vertex is the highest order of a core that contains this vertex



# Finding k-cores

---

**Algorithm:** Core decomposition

**Input:** graph  $G(V, E)$

**Output:**  $\text{core}[v]$  - core number for each vertex

compute  $\deg[v]$ ;

sort  $V$  :  $\deg[v_{i+1}] \geq \deg[v_i]$ ;

**for** each  $v \in V$  **do**

$\text{core}[v] = \text{degree}[v]$ ;

**for** each  $u \in \text{NN}(v)$  **do**

**if**  $\deg[u] > \deg[v]$  **then**

$\deg[u] = \deg[u] - 1$ ;

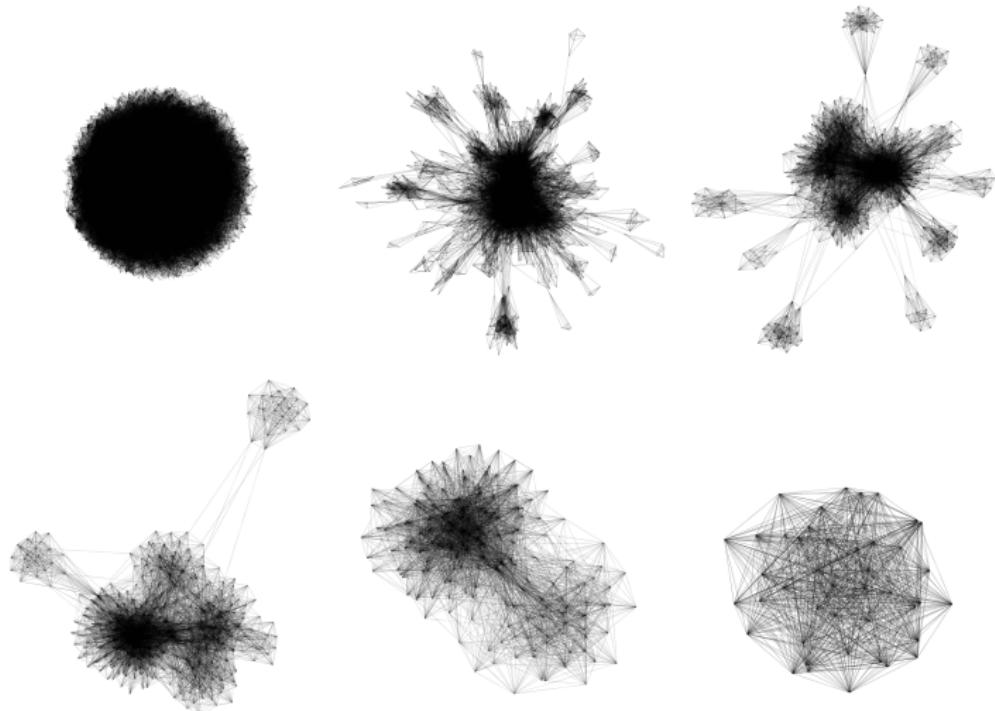
sort  $V$

**end**

**end**

**end**

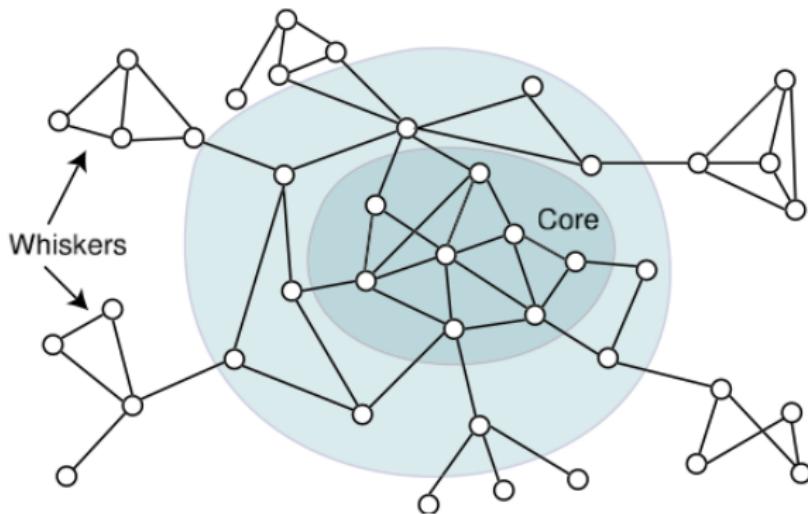
# k-cores in real networks



Zhukov, 2005

k-cores:  $k = 2, 5, 10, 15, 20, 31$

# Communities in real networks

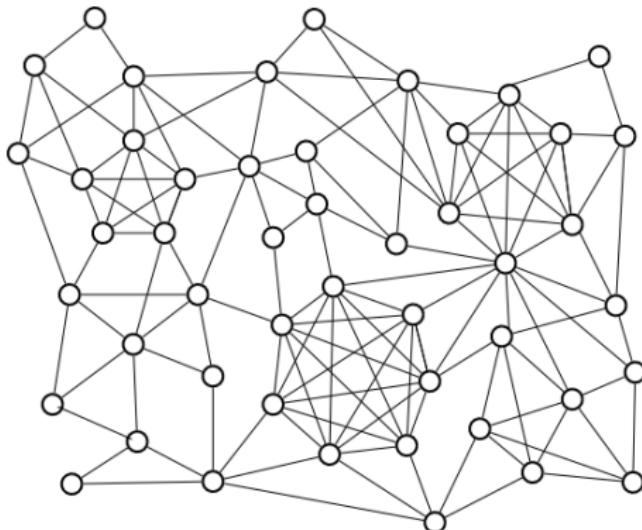


Leskovec et al, 2008

# Graph cliques

## Definition

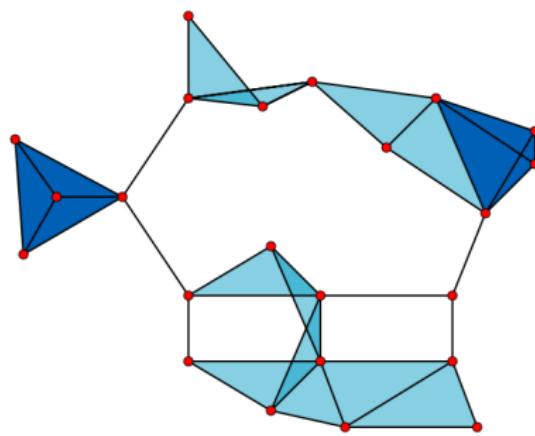
A *clique* is a complete subgraph, i.e. a set of elements where each pair of elements is connected.



Finding click of fixed given size  $k$  -  $O(n^k k^2)$

# Graph cliques

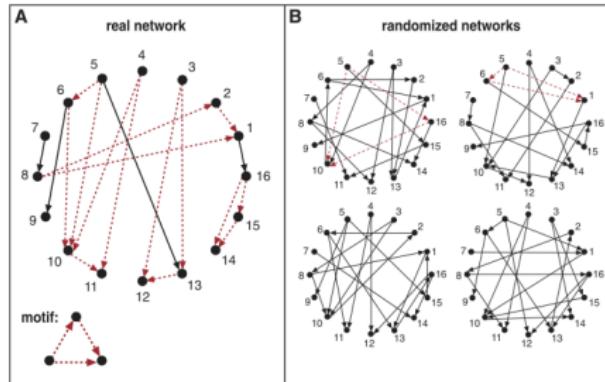
- A *maximal clique* is a clique that cannot be extended by including one more adjacent vertex (not included in larger one)
- A *maximum clique* is a clique of the largest possible size in a given graph -  $O(3^{n/3})$



# Network motifs

## Definition

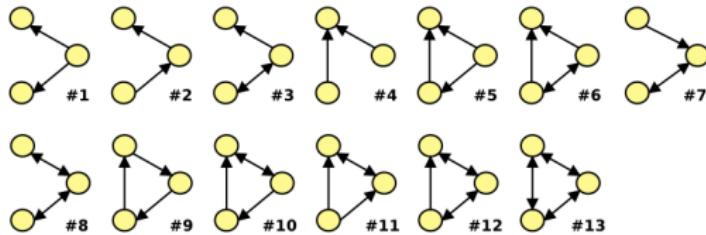
Network motifs are subgraphs that appear more frequently in a real network than could be statistically expected (compare to random network)



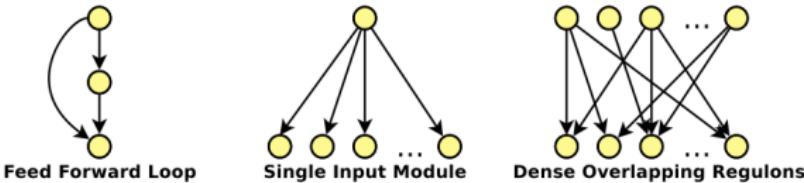
Milo et. al, 2002

# Network motifs

Triads - three node connected subgraphs:



More complicated motifs:



Ribeiro, 2011, Shen-Orr, 2002

# Network motifs

Motif discovery algorithms:

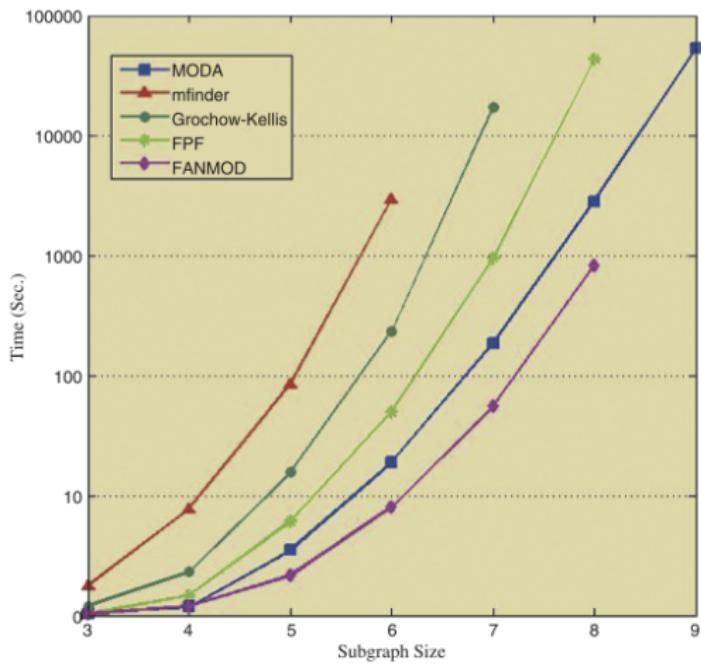
- calculate the number of occurrences of a sub graph
- evaluate the significance

For  $G'$  subgraph (motif candidate) of  $G$ ,

$$Z_{score}(G') = \frac{F_G(G') - \mu_R(G')}{\sigma_R(G')}$$

$R$  - random graph,  $\mu$  - mean frequency,  $\sigma$ -standard deviation

# Network motifs

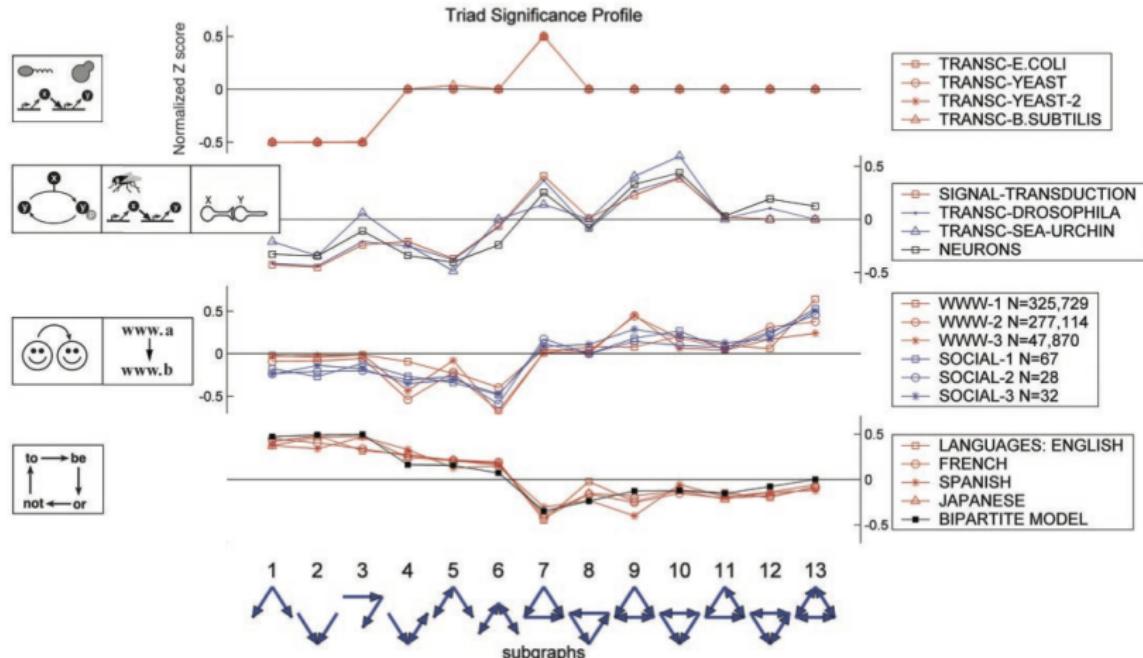


# Network motifs

Network	Nodes	Edges	$N_{real}$	$N_{rand}$	Z-Score	$N_{real}$	$N_{rand}$	Z-Score	$N_{real}$	$N_{rand}$	Z-Score
Gene Regulation (transcription)											
E. coli	424	519	40	7±3	10	203	47±12	13			
S. cerevisiae	685	1052	70	11±4	14	1812	300±40	41			
Food Webs											
Little Rock	92	984	3219	3120±50	2.1	7295	2220±210	25			
Ythan	83	391	1182	1020±20	7.2	1357	230±50	23			
Electronic Circuits (digital fract. multipliers)											
s208	122	189	10	1±1	9	4	1±1	3.8	5	1±1	5
s420	252	399	20	1±1	18	10	1±1	10	11	1±1	11

Ribeiro, 2011, Milo, 2002

# Network motifs



## References

- S. E. Schaeffer. Graph clustering. Computer Science Review, 1(1):27–64, 2007.
- S. Fortunato. Community detection in graphs, Physics Reports, Vol. 486, Iss. 3–5, pp 75-174, 2010
- V. Batagelj and M. Zaversnik. An  $O(m)$  algorithm for cores decomposition of networks. arXiv, October 2003.
- Milo R, Shen-Orr SS, Itzkovitz S, Kashtan N, Chklovskii D, Alon U. Network motifs: simple building blocks of complex networks. Science 298 (5594): 824–827, 2002