NATIONAL RESEARCH UNIVERSITY

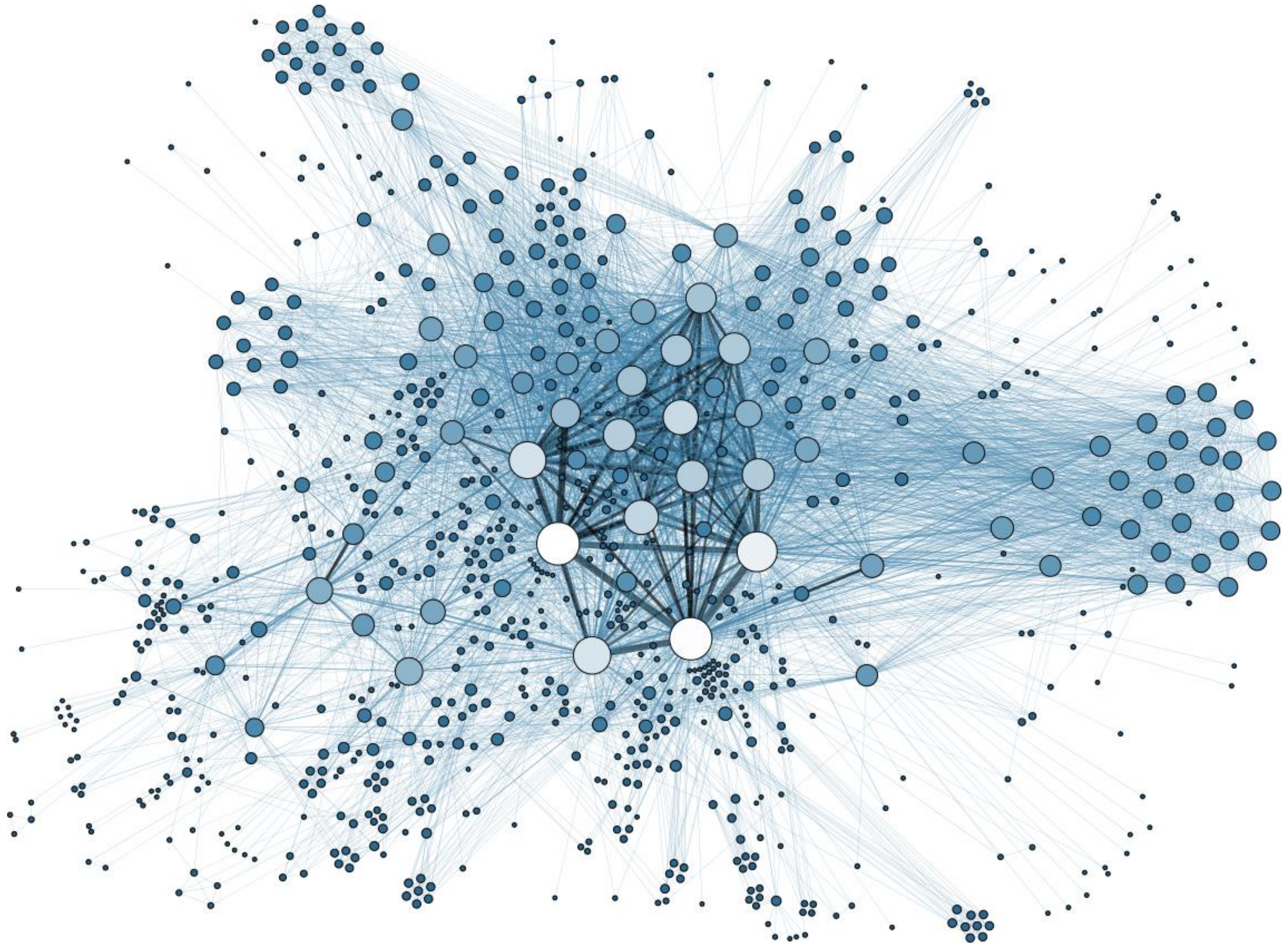
NODE CENTRALITIES

Ilia Karpov (karpovilia@gmail.com)

Moscow, 2022

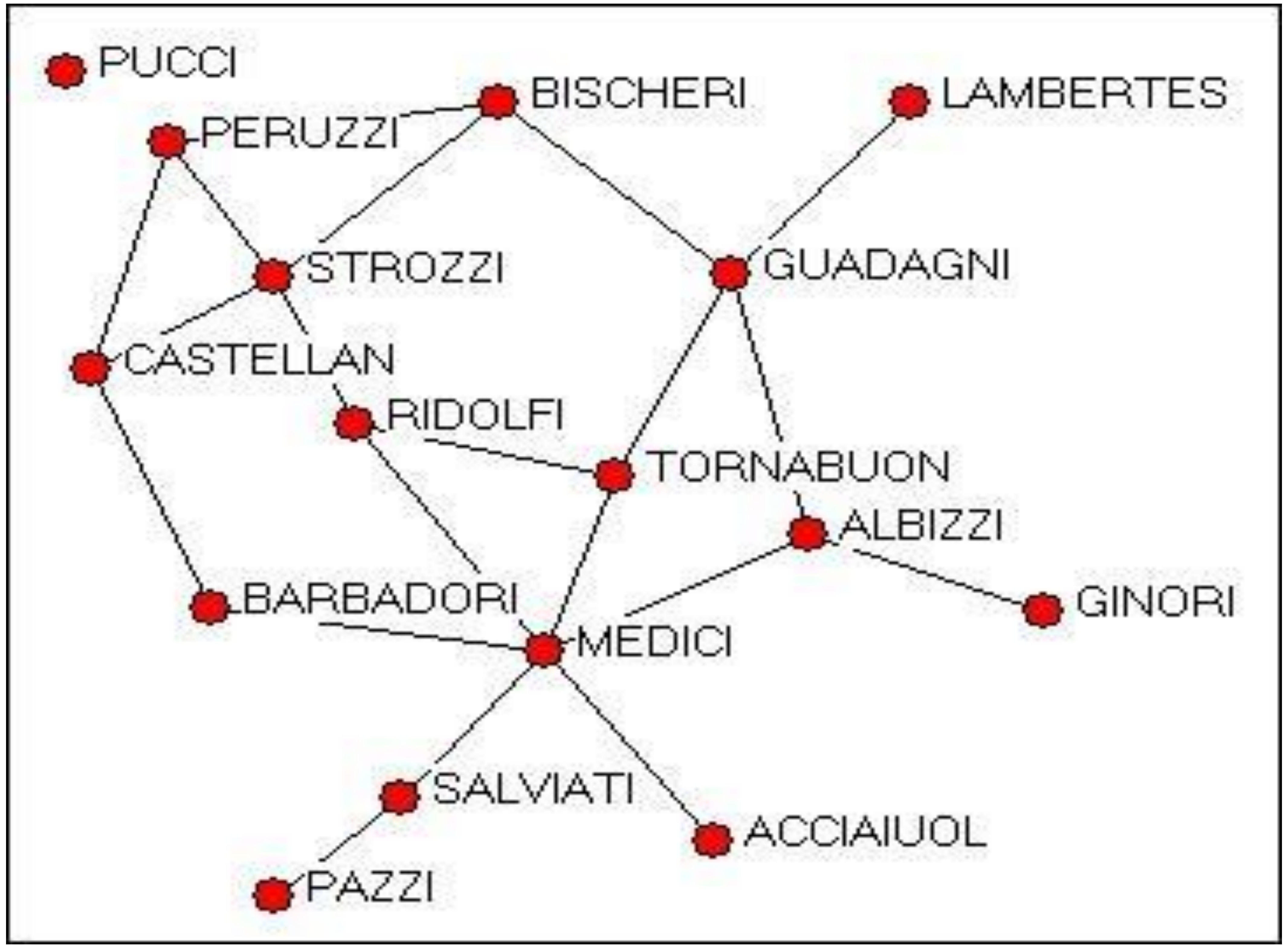
INTRODUCTION
CENTRALITIES EXAMPLES

Which vertices are important?



INTRODUCTION

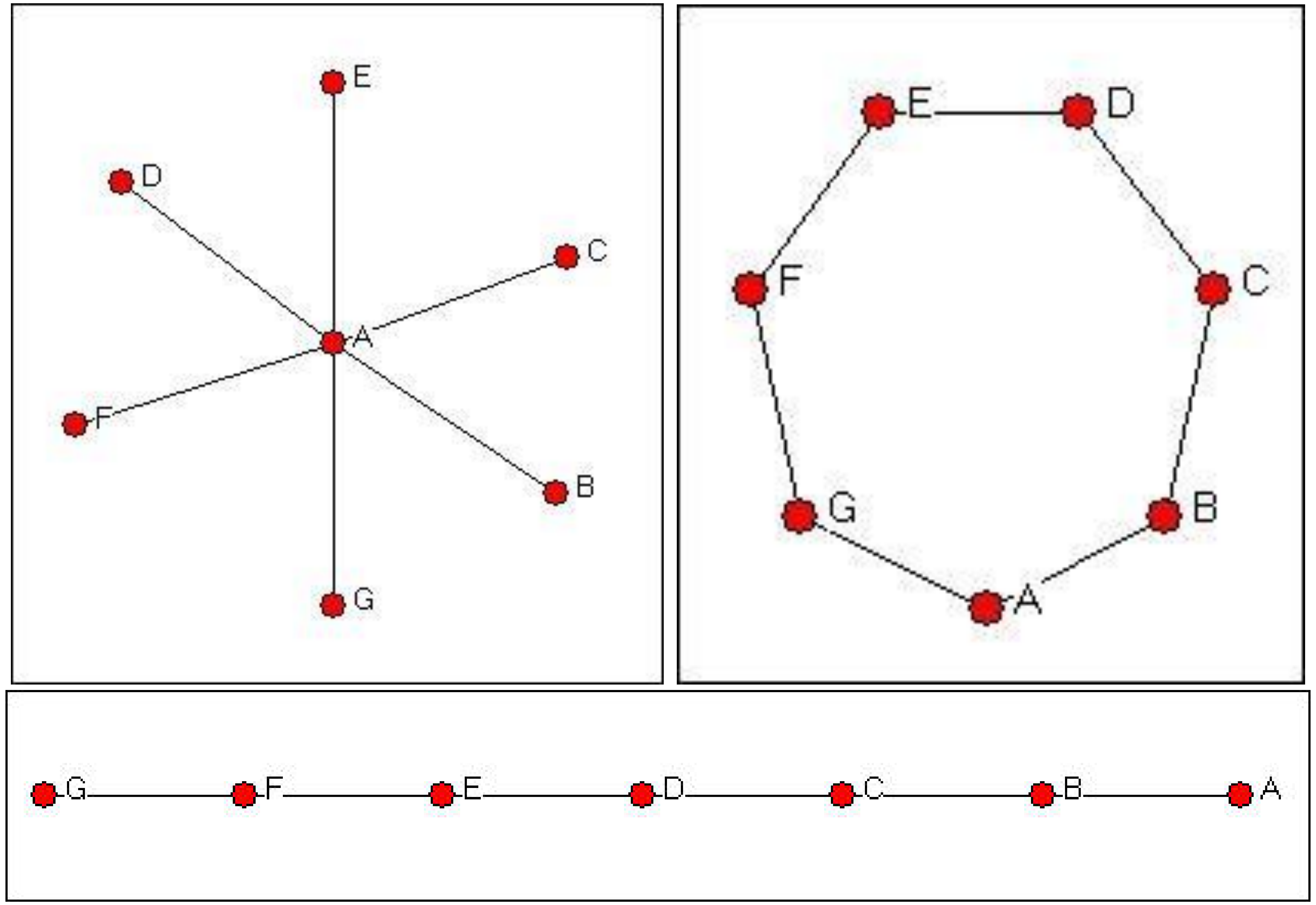
CENTRALITIES EXAMPLES



Marriage alliances among leading Florentine families 15th century.

Determine the most ”important” or ”prominent” actors in the network based on actor location.

INTRODUCTION



Star graph Circle graph Line Graph

DEGREE CENTRALITY

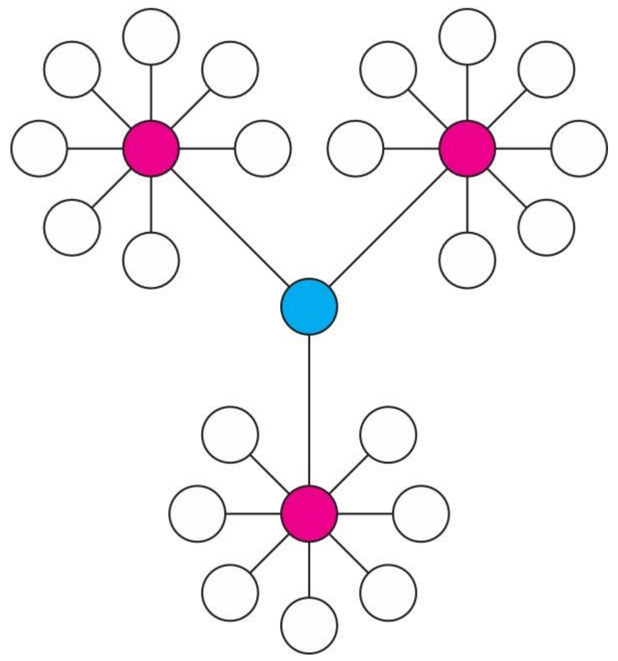
Degree centrality: number of nearest neighbors

$$C_D(i) = k(i) = \sum_j A_{ij} = \sum_j A_{ji}$$

Normalized degree centrality

$$C_D^*(i) = \frac{1}{n-1} C_D(i) = \frac{k(i)}{n-1}$$

High centrality degree -direct contact with many other actors



CLOSENESS CENTRALITY

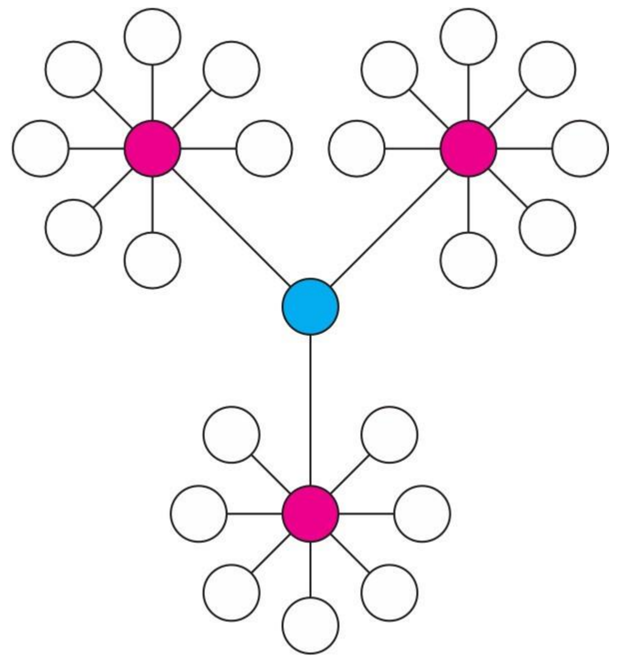
Closeness centrality: how close an actor to all the other actors in network

$$C_C(i) = \frac{1}{\sum_j d(i,j)}$$

Normalized closeness centrality

$$C_C^*(i) = (n-1)C_C(i) = \frac{n-1}{\sum_j d(i,j)}$$

High closeness centrality - short communication path to others, minimal number of steps to reach others



BETWEENNESS CENTRALITY

Betweenness centrality: number of shortest paths going through the actor $\sigma_{st}(i)$

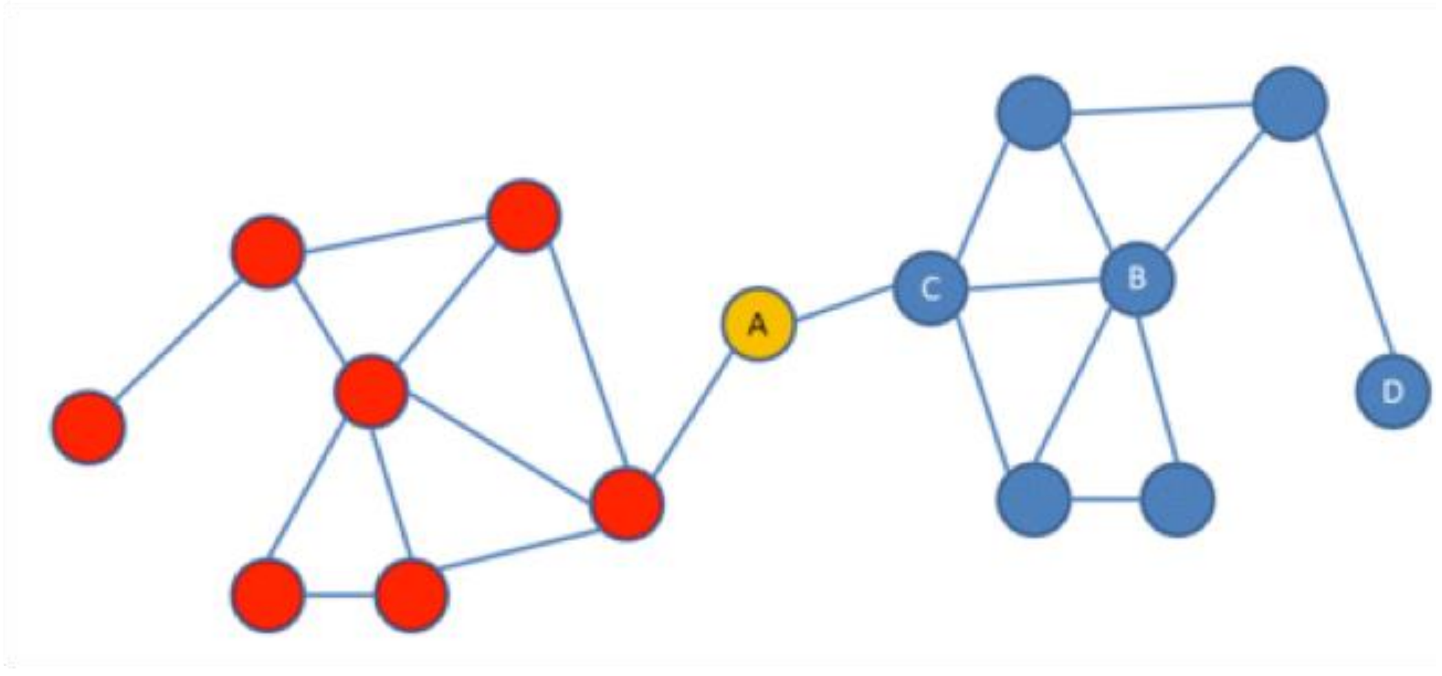
$$C_B(i) = \sum_{s \neq t \neq i} \frac{\sigma_{st}(i)}{\sigma_{st}}$$

Normalized betweenness centrality

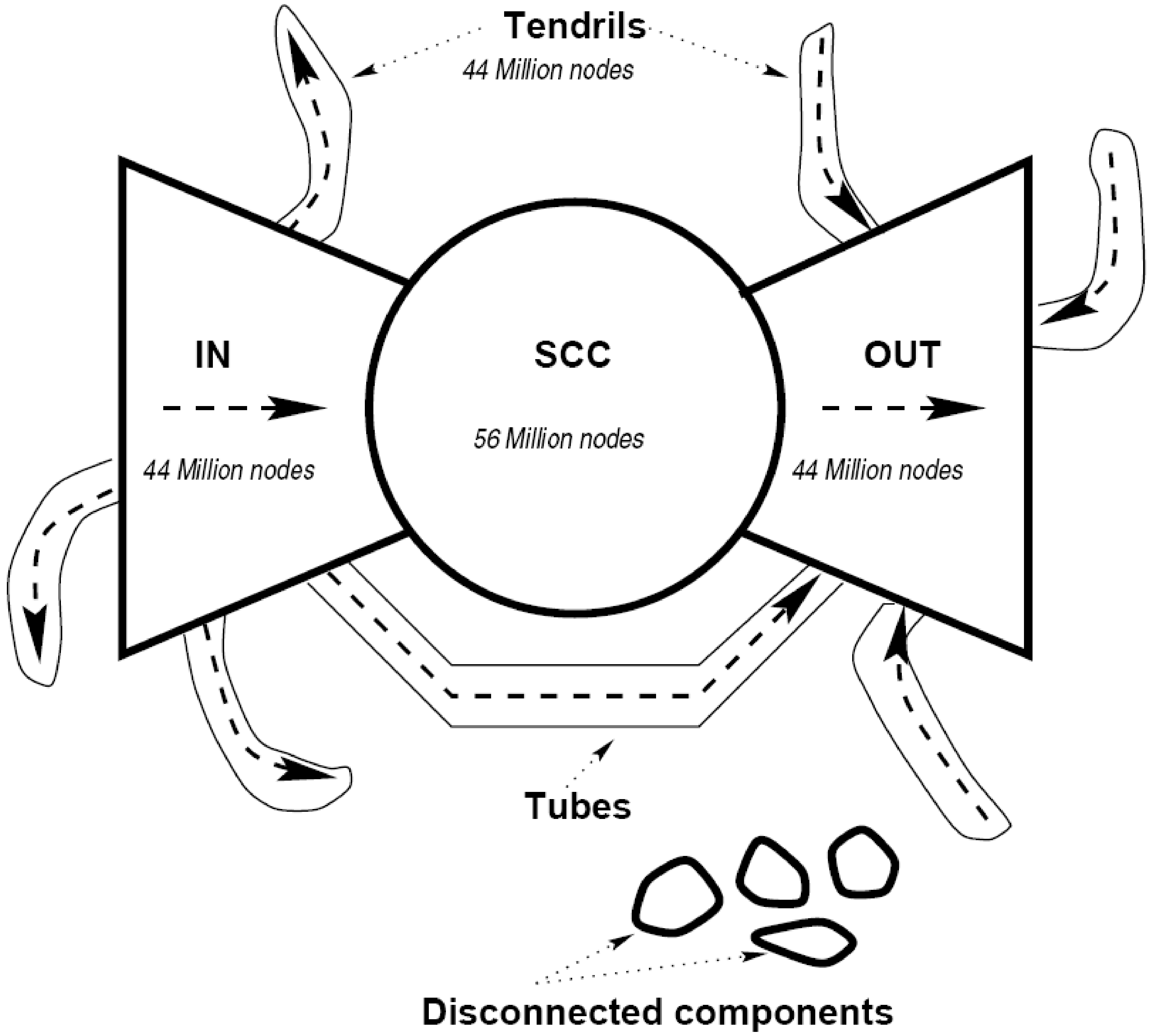
$$C_B^*(i) = \frac{2}{(n-1)(n-2)} C_B(i) = \frac{2}{(n-1)(n-2)} \sum_{s \neq t \neq i} \frac{\sigma_{st}(i)}{\sigma_{st}}$$

Hight betweenness centrality - vertex lies on many shortest paths

Probability that a communication from *s* to *t* will go through *i*



BOW-TIE STRUCTURE



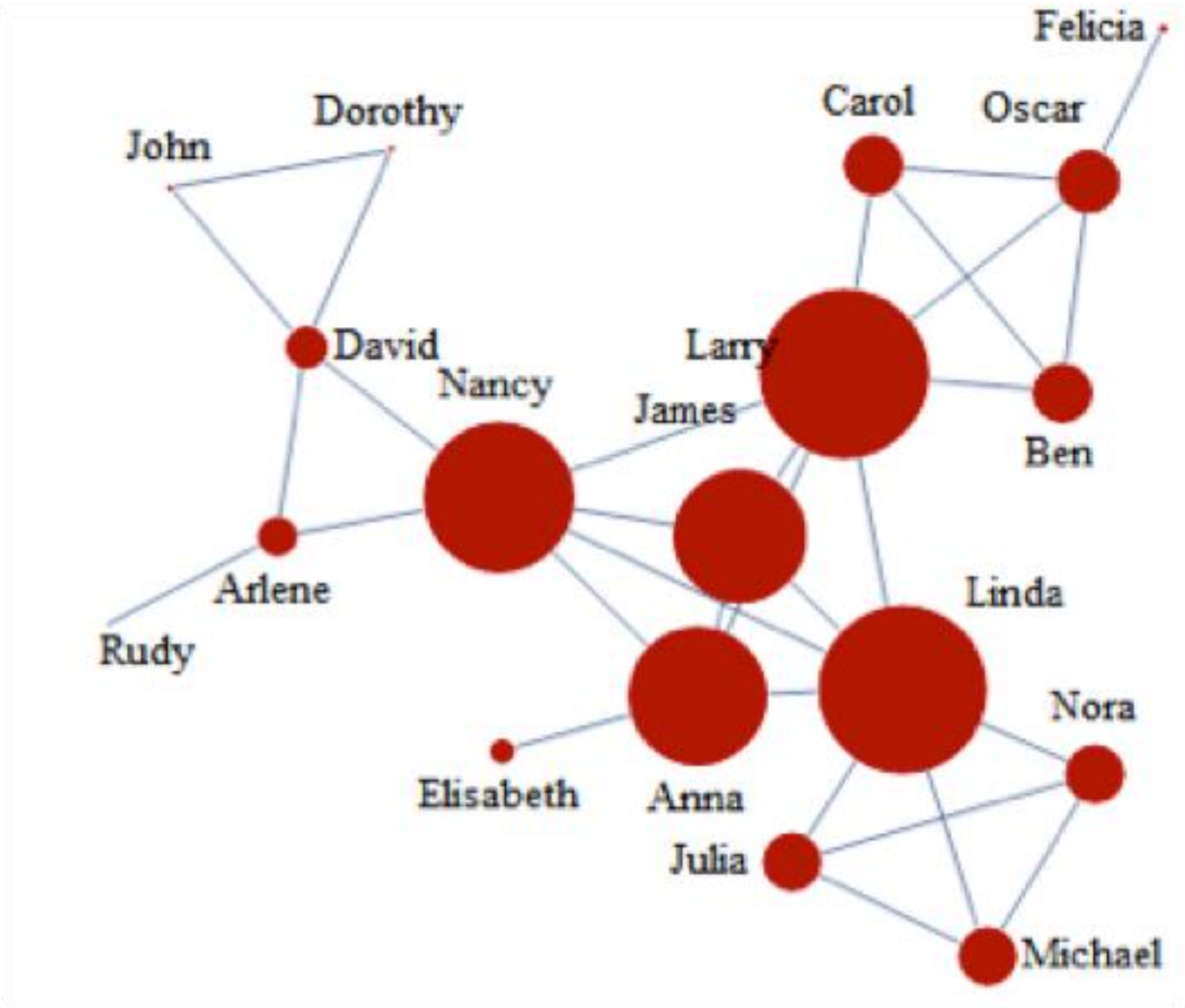
EIGENVECTOR CENTRALITY

Importance of a node depends on the importance of its neighbors (recursive definition)

$$v_i \leftarrow \sum_j A_{ij} v_j$$

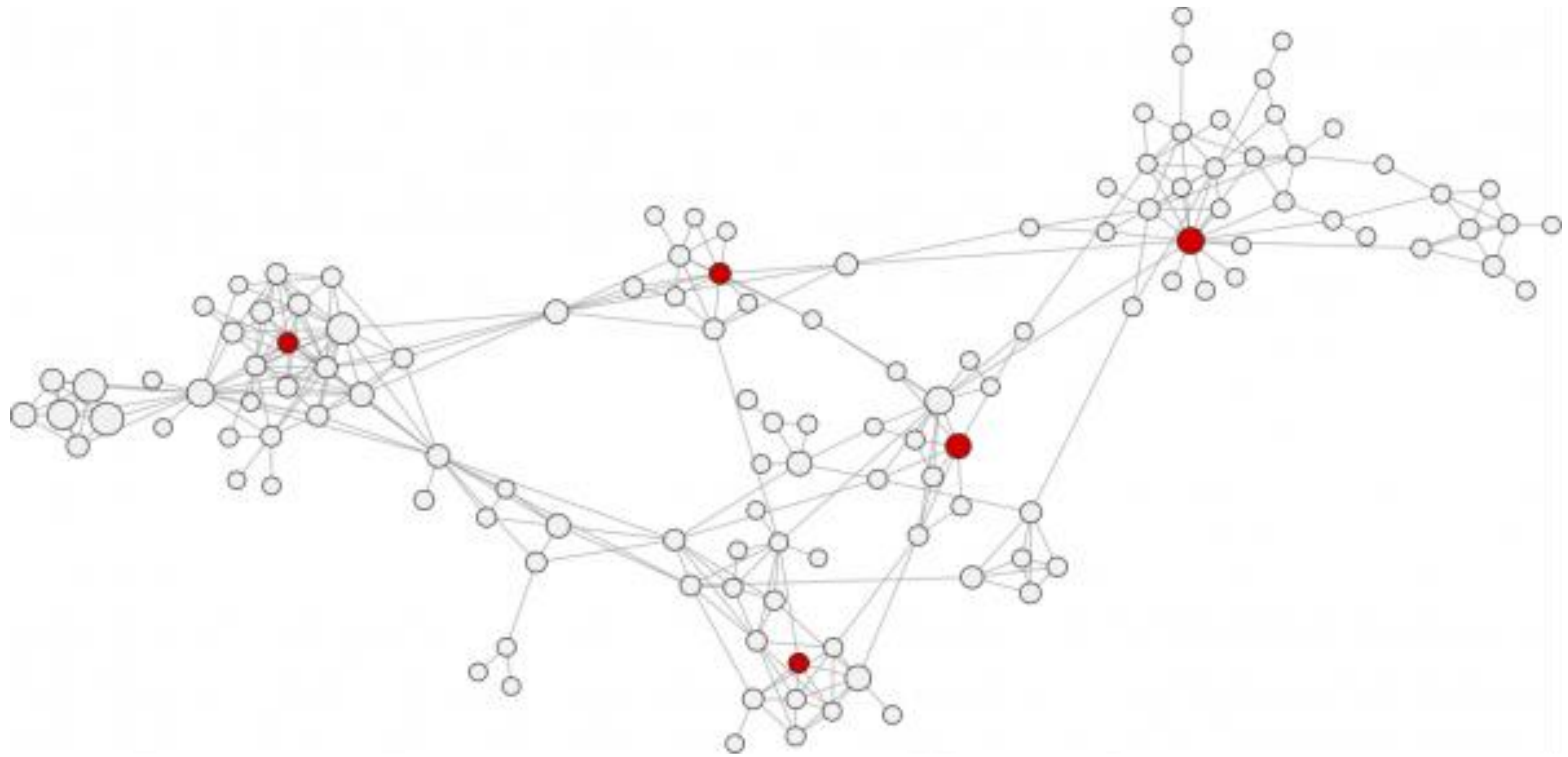
$$v_i = \frac{1}{\lambda} \sum_j A_{ij} v_j$$

$$\mathbf{A}\mathbf{v} = \lambda \mathbf{v}$$



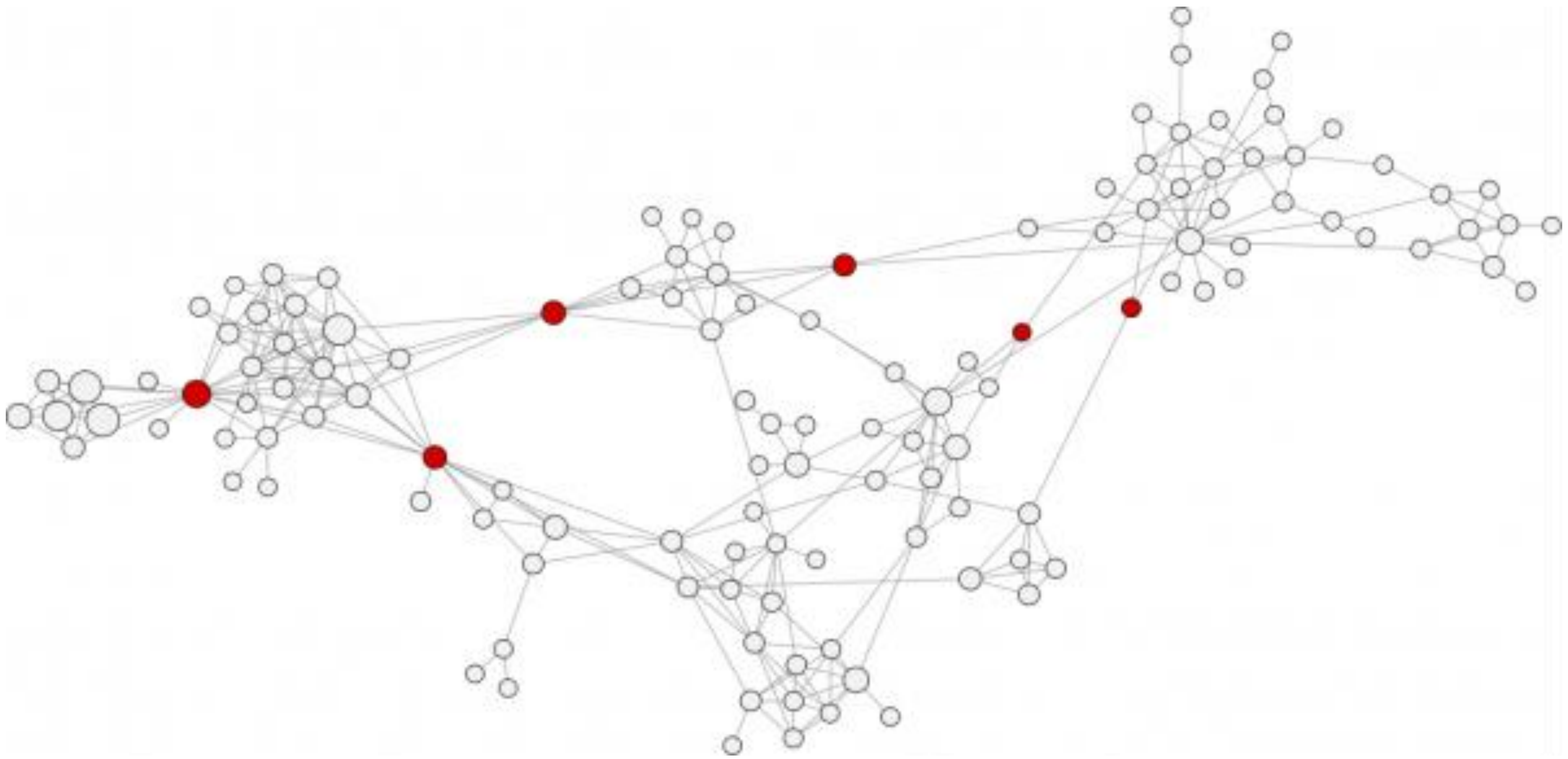
Select an eigenvector associated with largest eigenvalue $\lambda = \lambda_1$, $\mathbf{v} = \mathbf{v}_1$

CLOSENESS CENTRALITY



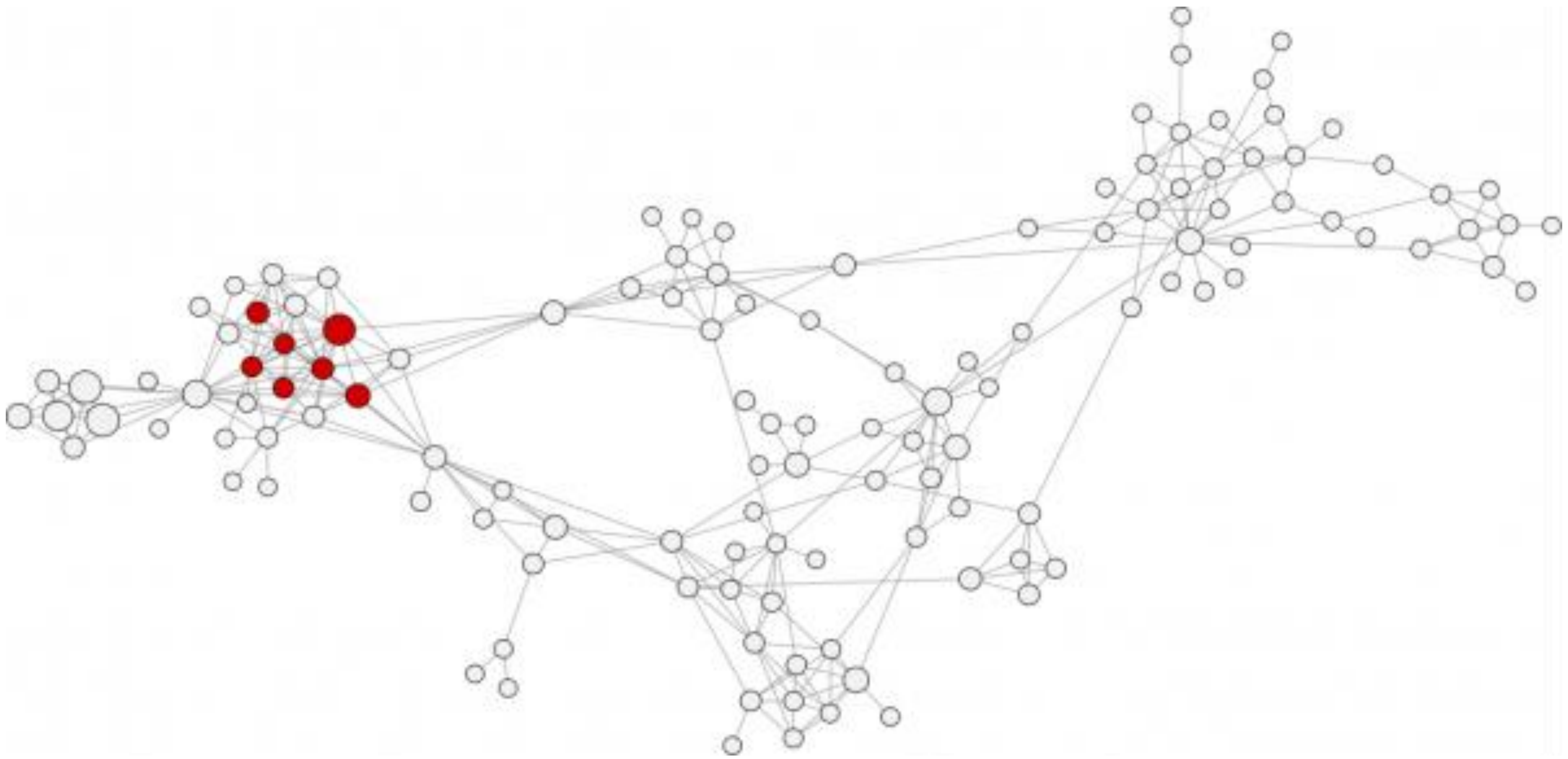
from www.activenetworks.net

BETWEENNESS CENTRALITY



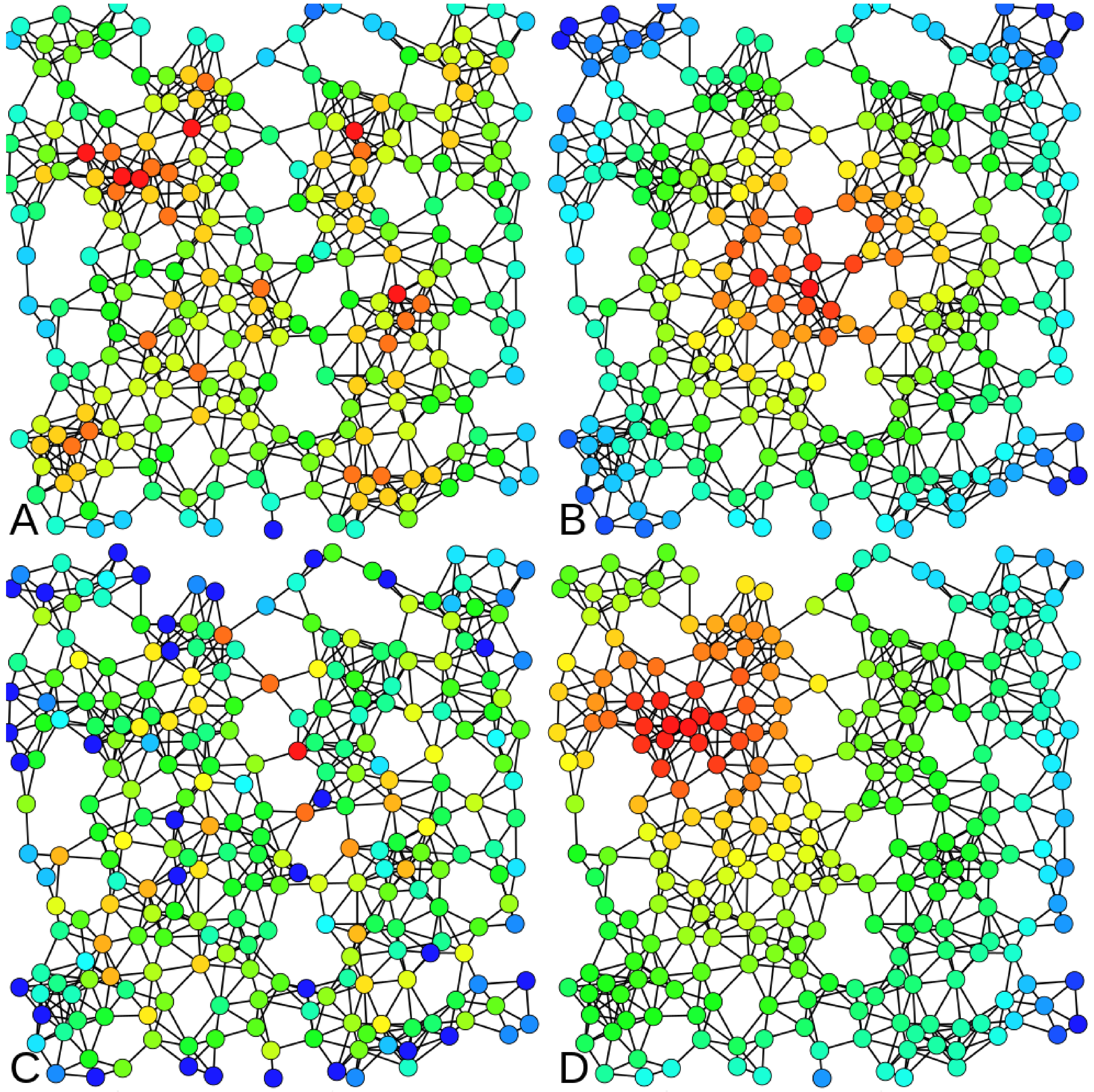
from www.activenetworks.net

EIGENVECTOR CENTRALITY



from www.activenetworks.net

CENTRALITY EXAMPLES



from Claudio Rocchini

- A) Degree centrality
- B) Closeness centrality
- C) Betweenness centrality
- D) Eigenvector centrality

CENTRALIZATION

Centralization (network measure) - how central the most central node in the network in relation to all other nodes.

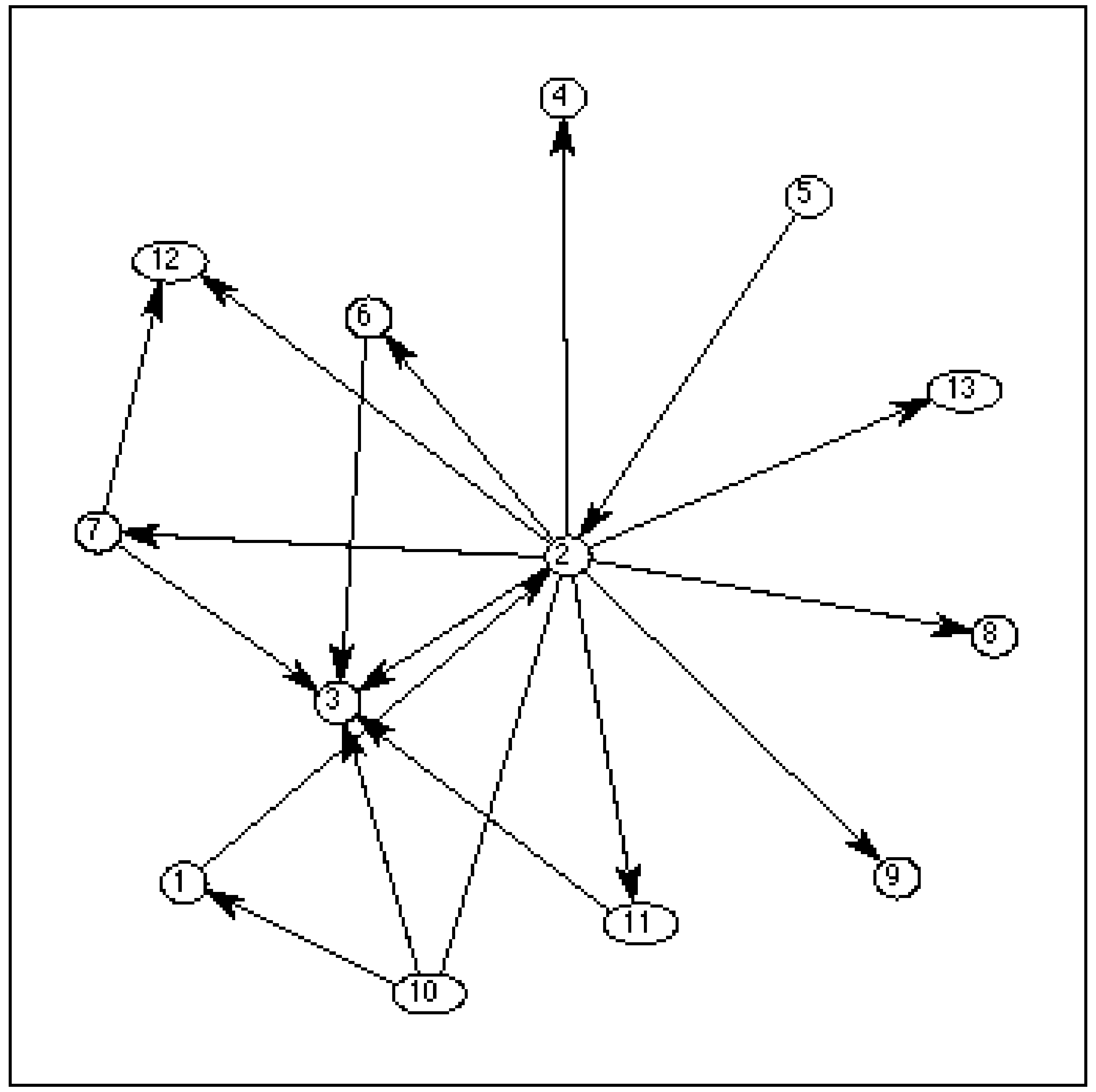
$$C_x = \frac{\sum_i^N [C_x(p_*) - C_x(p_i)]}{\max \sum_i^N [C_x(p_*) - C_x(p_i)]}$$

C_x - one of the centrality measures
p_* - node with the largest centrality value
max - is taken over all graphs with the same number of nodes (for degree, closeness and betweenness the most centralized structure is the star graph)

Linton Freeman, 1979

DIRECTIONAL RELATIONS

Directed graph: distinguish between choices made (outgoing edges) and choices received (incoming edges)



sending - receiving
export - import
cite papers - being cited

CENTRALITY MEASURES

All based on outgoing edges

- Degree centrality (normalized):

$$C_D^*(i) = \frac{k^{out}(i)}{n-1}$$

- Closeness centrality (normalized):

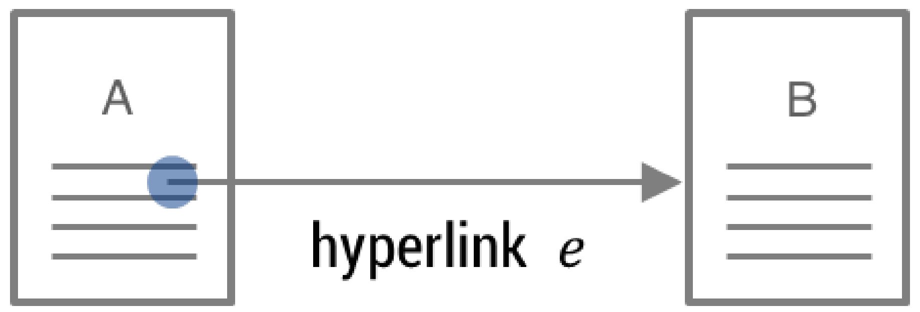
$$C_C^*(i) = \frac{n-1}{\sum_j d(i,j)}$$

- Betweenness centrality (normalized):

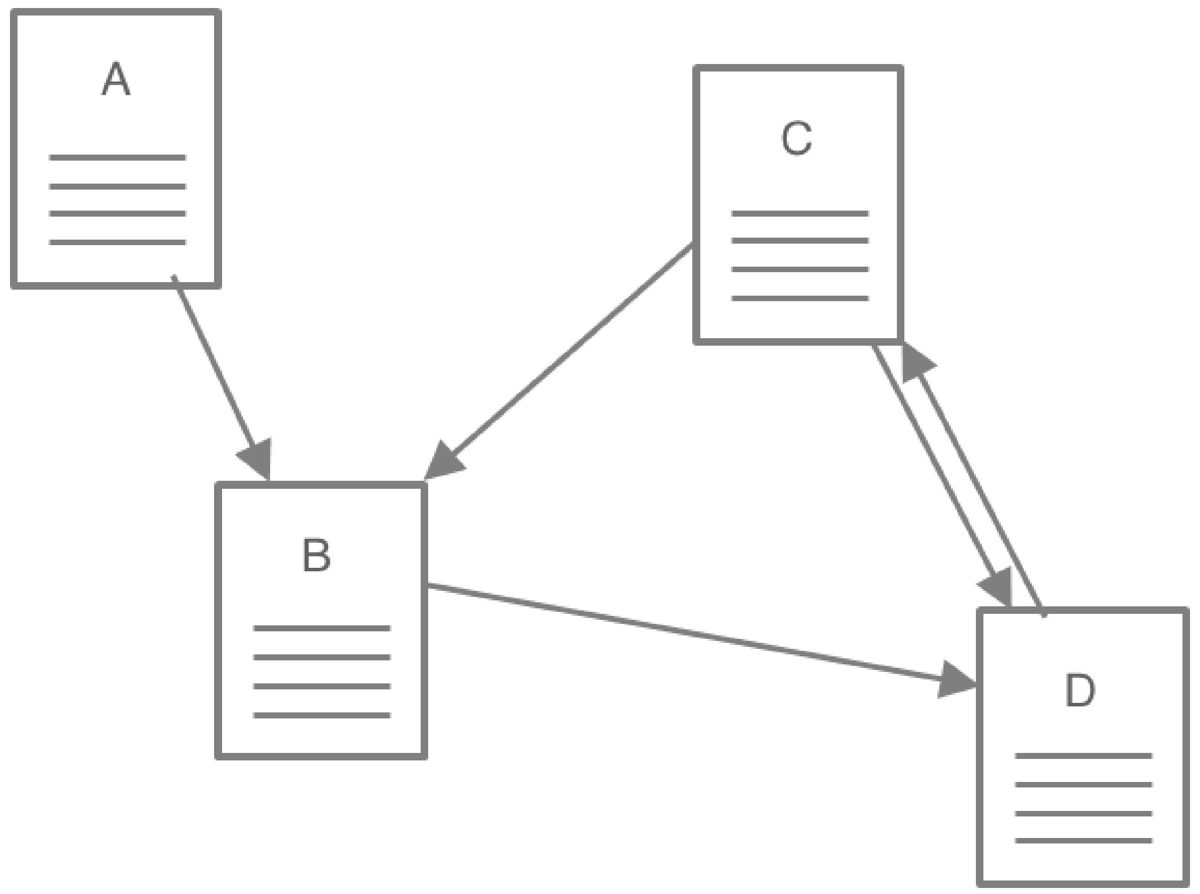
$$C_B^*(i) = \frac{1}{(n-1)(n-2)} \sum_{s \neq t \neq i} \frac{\sigma_{st}(i)}{\sigma_{st}}$$

WEB AS A GRAPH

- Hyperlinks - implicit endorsements

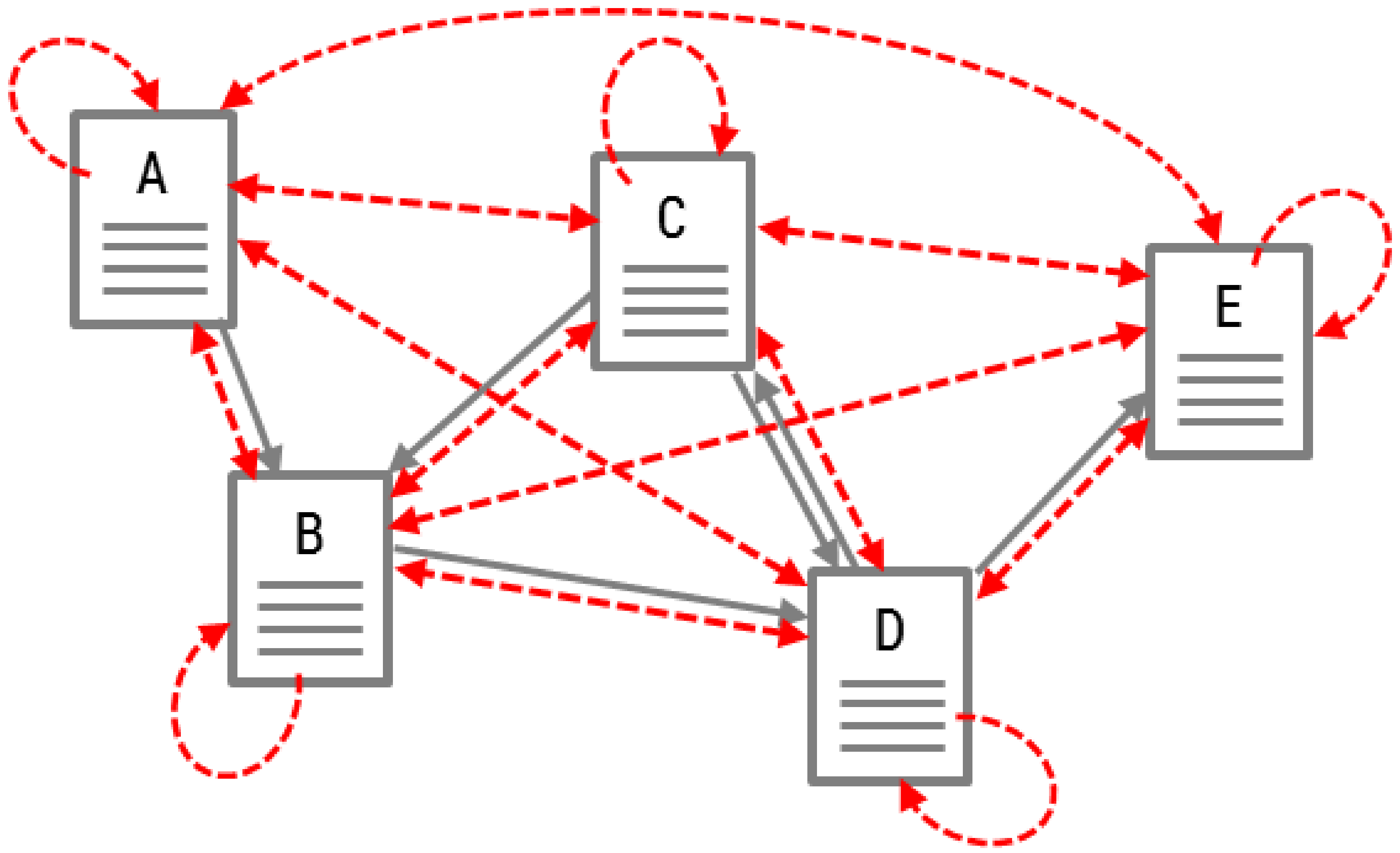


- Web graph - graph of endorsements (sometimes reciprocal)



PAGERANK

”PageRank can be thought of as a model of user behavior. We assume there is a ”random surfer” who is given a web page at random and keeps clicking on links, never hitting ”back” but eventually gets bored and starts on another random page. The **probability** that the random surfer visits a page is its **PageRank**.”



Sergey Brin and Larry Page, 1998

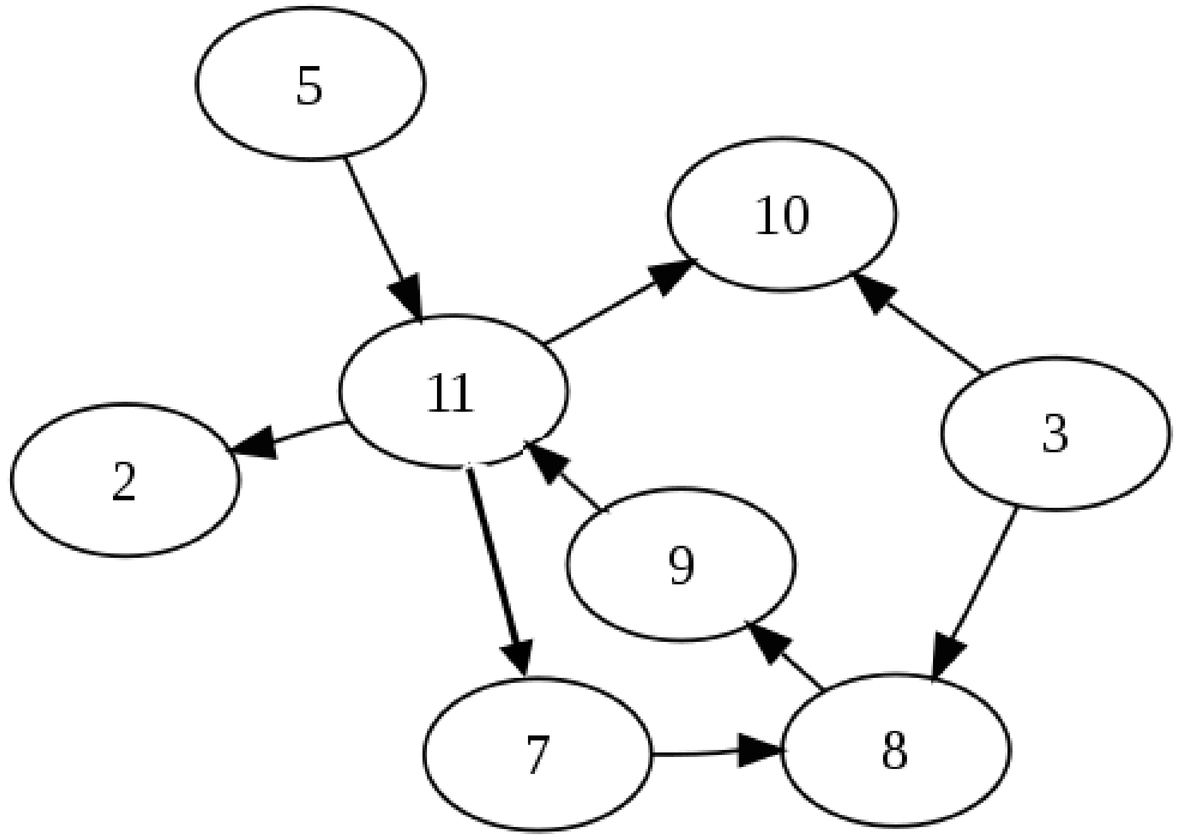
RANDOM WALK

- Random walk on graph

$$p_i^{t+1} = \sum_{j \in N(i)} \frac{p_j^t}{d_j^{out}} = \sum_j \frac{A_{ji}}{d_j^{out}} p_j$$

$$\mathbf{P} = \mathbf{D}^{-1}\mathbf{A}, \quad \mathbf{D}_{ii} = diag\{d_i^{out}\}$$

$$\mathbf{p}^{t+1} = \mathbf{P}^T \mathbf{p}^t$$



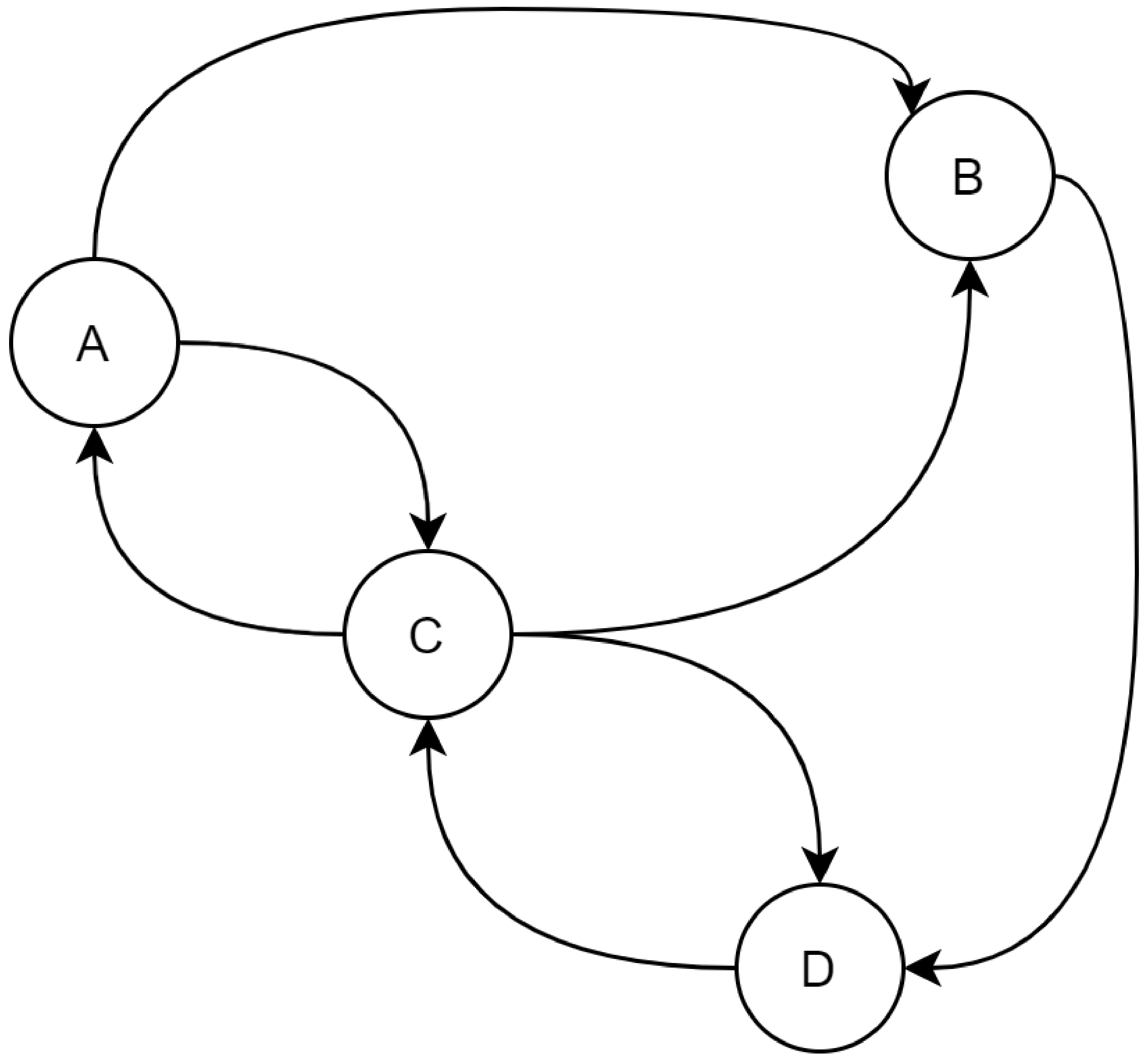
- with teleportation

$$\mathbf{p}^{t+1} = \alpha \mathbf{P}^T \mathbf{p}^t + (1-\alpha)\frac{\mathbf{e}}{n}$$

Perron-Frobenius Theorem guarantees existence and uniqueness of the solution $\lim_{t\to\infty} \mathbf{p} = \pi$

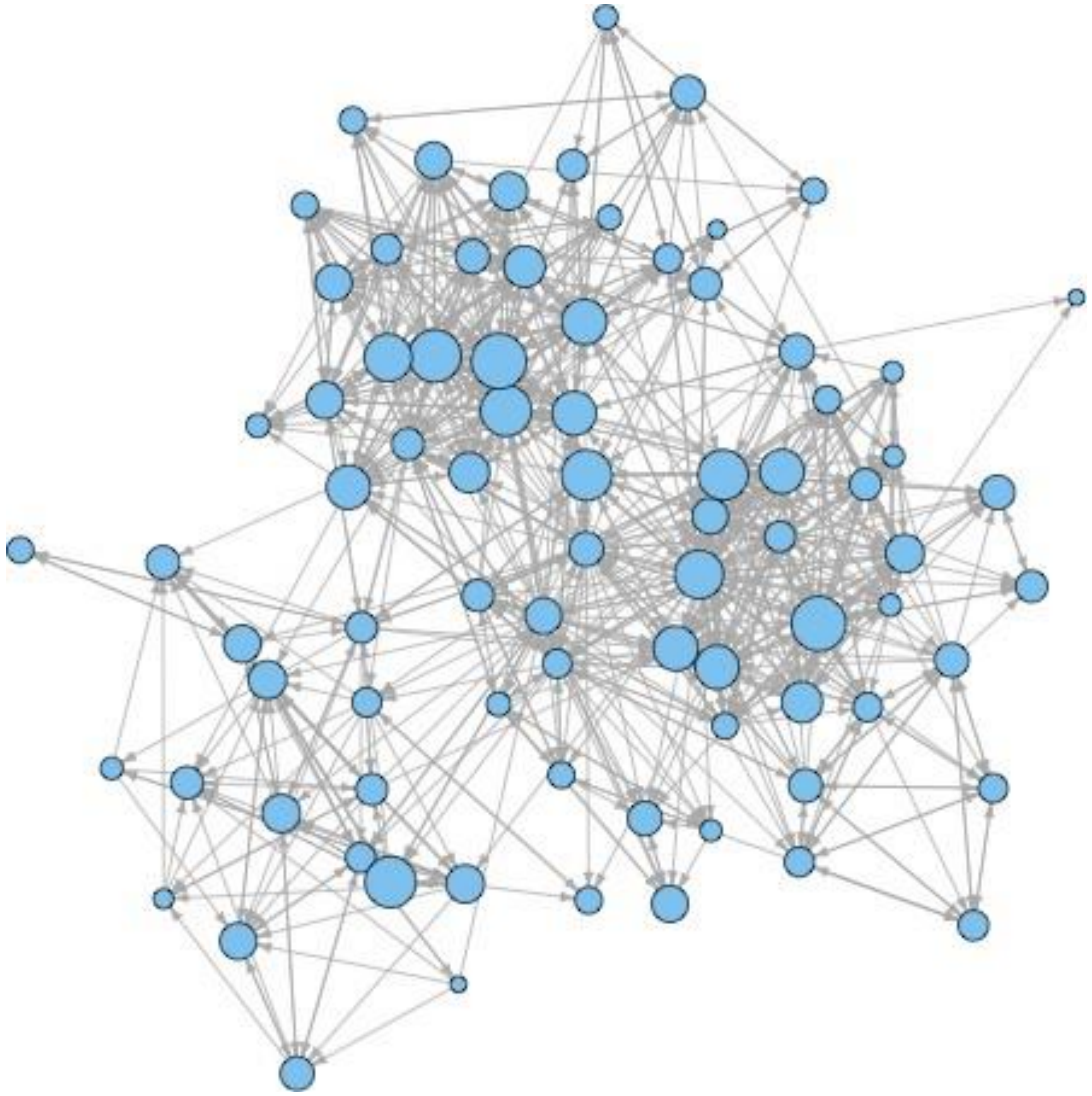
$$\mathbf{p} = \alpha \mathbf{P}^T \mathbf{p} + (1-\alpha)\frac{\mathbf{e}}{n}$$

PAGERANK



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

	it. 0	it 1	it 2	PR
A	¼	1/12	1.5/12	1
B	¼	2.5/12	2/12	2
C	¼	4.5/12	4.5/12	4
D	¼	4/12	4/12	3



1. GeneRank
2. ProteinRank
3. FoodRank
4. SportsRank
5. HostRank
6. TrustRank
7. BadRank
8. ObjectRank
9. ItemRank
10. ArticleRank
11. BookRank
12. FutureRank
13. TimedPageRank
14. SocialPageRank
15. DiffusionRank
16. ImpressionRank
17. TweetRank
18. TwitterRank
19. ReversePageRank
20. PageTrust
21. PopRank
22. CiteRank
23. FactRank
24. InvestorRank
25. ImageRank
26. VisualRank
27. QueryRank
28. BookmarkRank
29. StoryRank
30. PerturbationRank
31. ChemicalRank
32. RoadRank
33. PaperRank
34. Etc…

HUBS AND AUTHORITIES (HITS)

Citation networks. Reviews vs original research (authoritative) papers

- authorities, contain useful information, a_i
- hubs, contains links to authorities, h_i

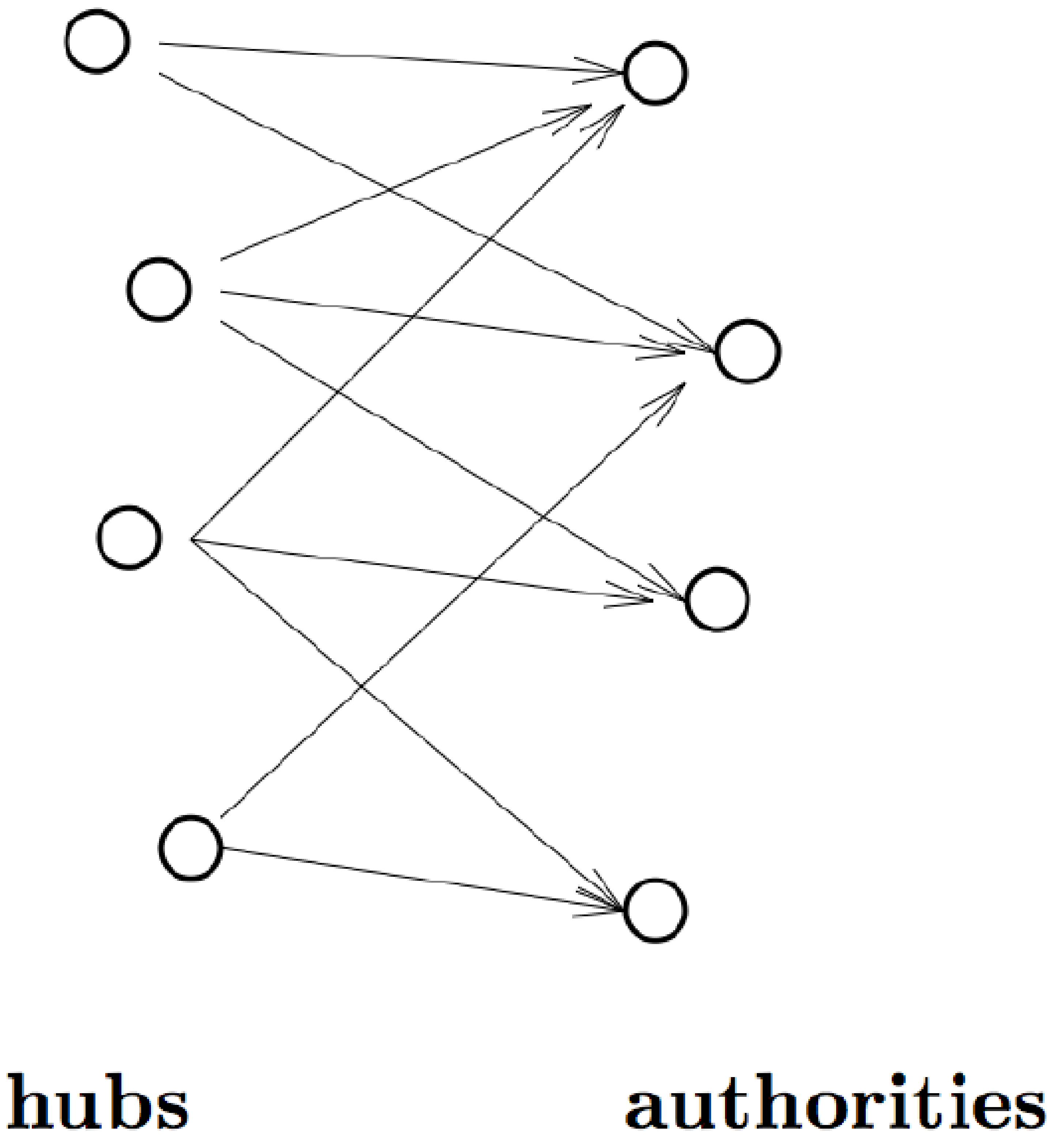
Mutual recursion

- good authorities referred by good hubs

$$a_i \leftarrow \sum_j A_{ji} h_j$$

- good hubs point to good authorities

$$h_i \leftarrow \sum_j A_{ij} a_j$$



HUBS AND AUTHORITIES (HITS)

System of linear equations

$$\begin{aligned} \mathbf{a} &= \alpha \mathbf{A}^T \mathbf{h} \\ \mathbf{h} &= \beta \mathbf{A} \mathbf{a} \end{aligned}$$

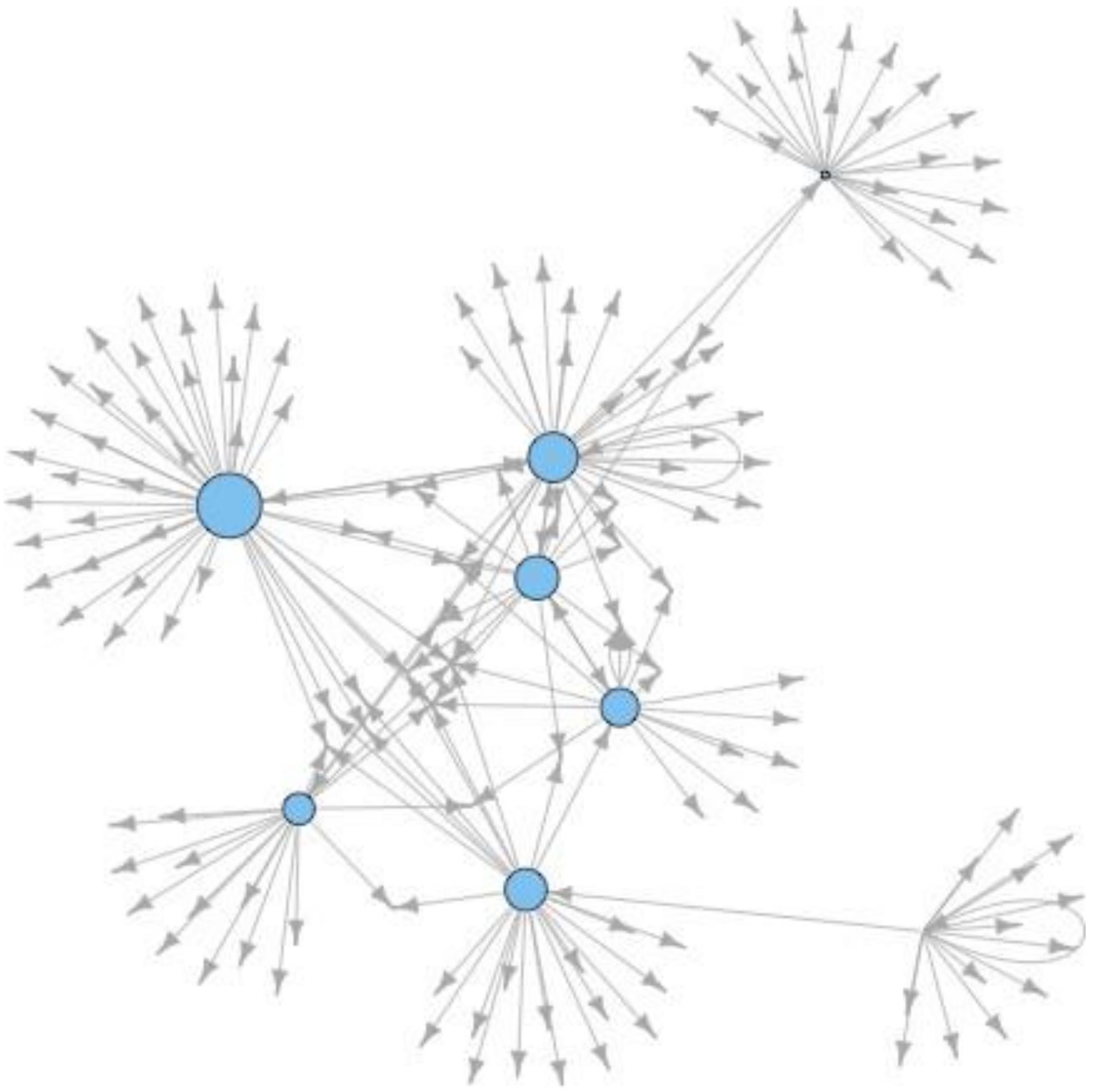
Symmetric eigenvalue problem

$$\begin{aligned} (\mathbf{A}^T \mathbf{A})\mathbf{a} &= \lambda \mathbf{a} \\ (\mathbf{A}\mathbf{A}^T)\mathbf{h} &= \lambda \mathbf{h} \end{aligned}$$

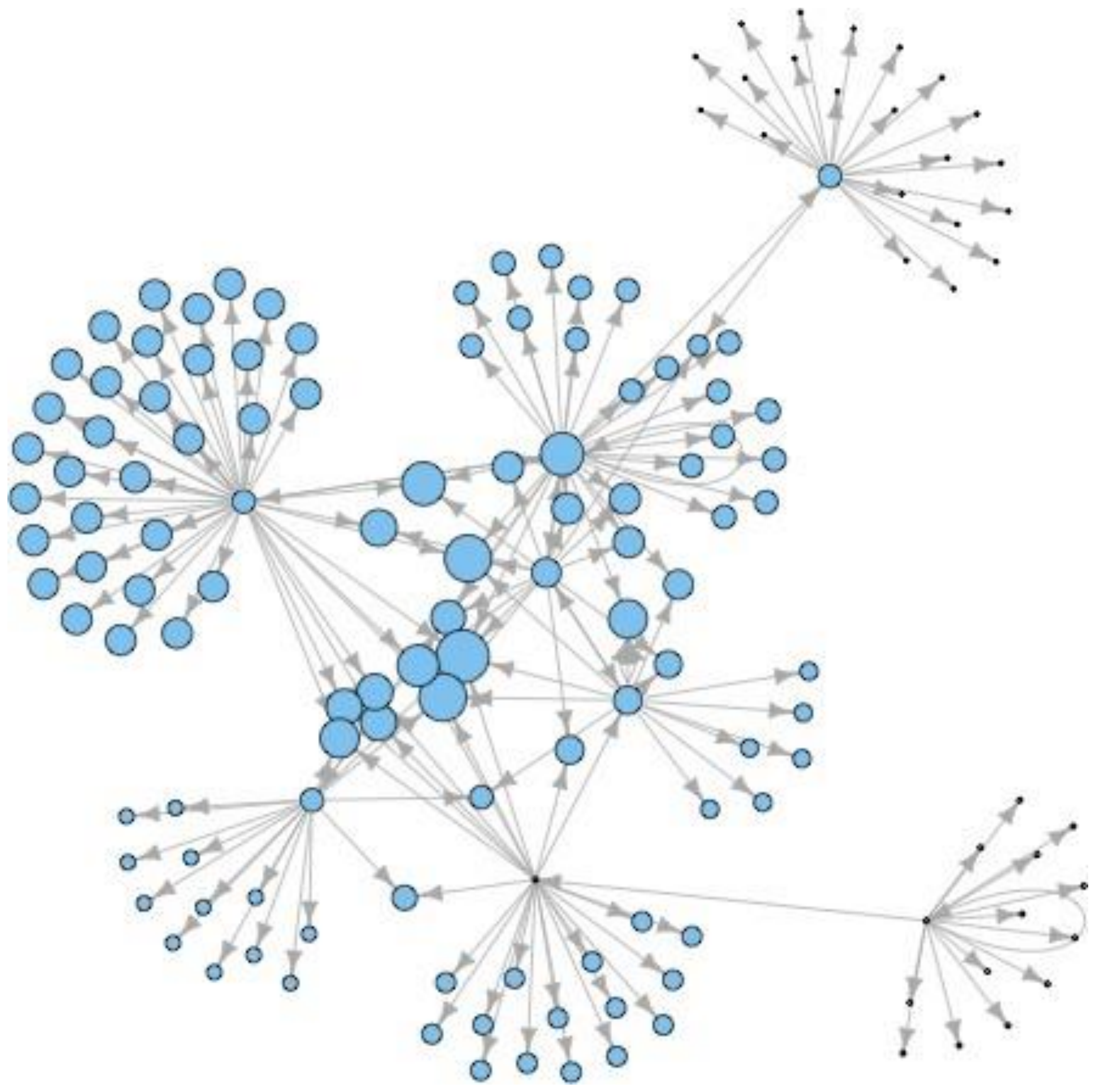
where eigenvalue $\lambda = (\alpha\beta)^{-1}$

HUBS AND AUTHORITIES (HITS)

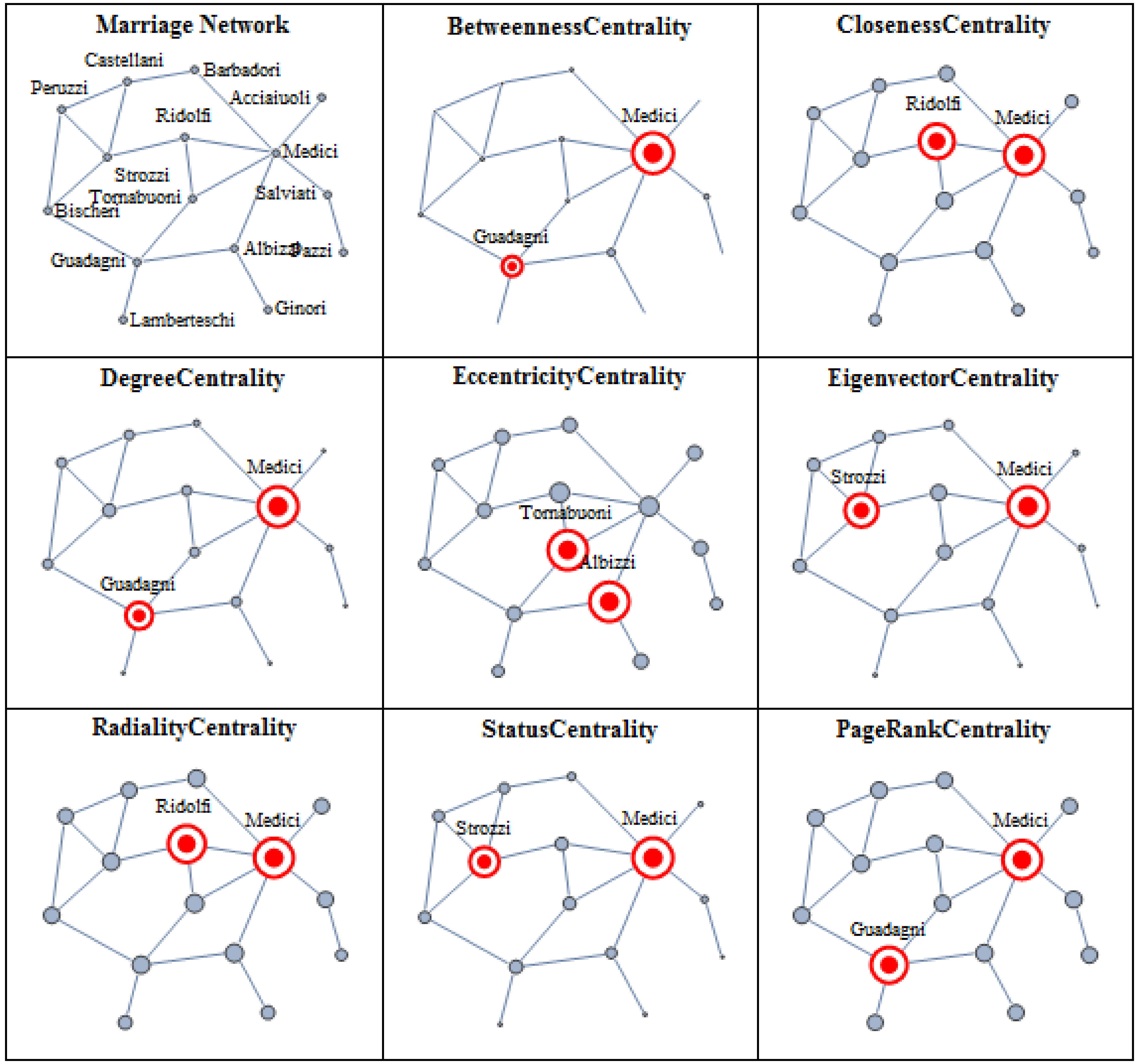
Hubs



Authorities



CENTRALITIES



CORE-PERIPHERY STRUCTURE

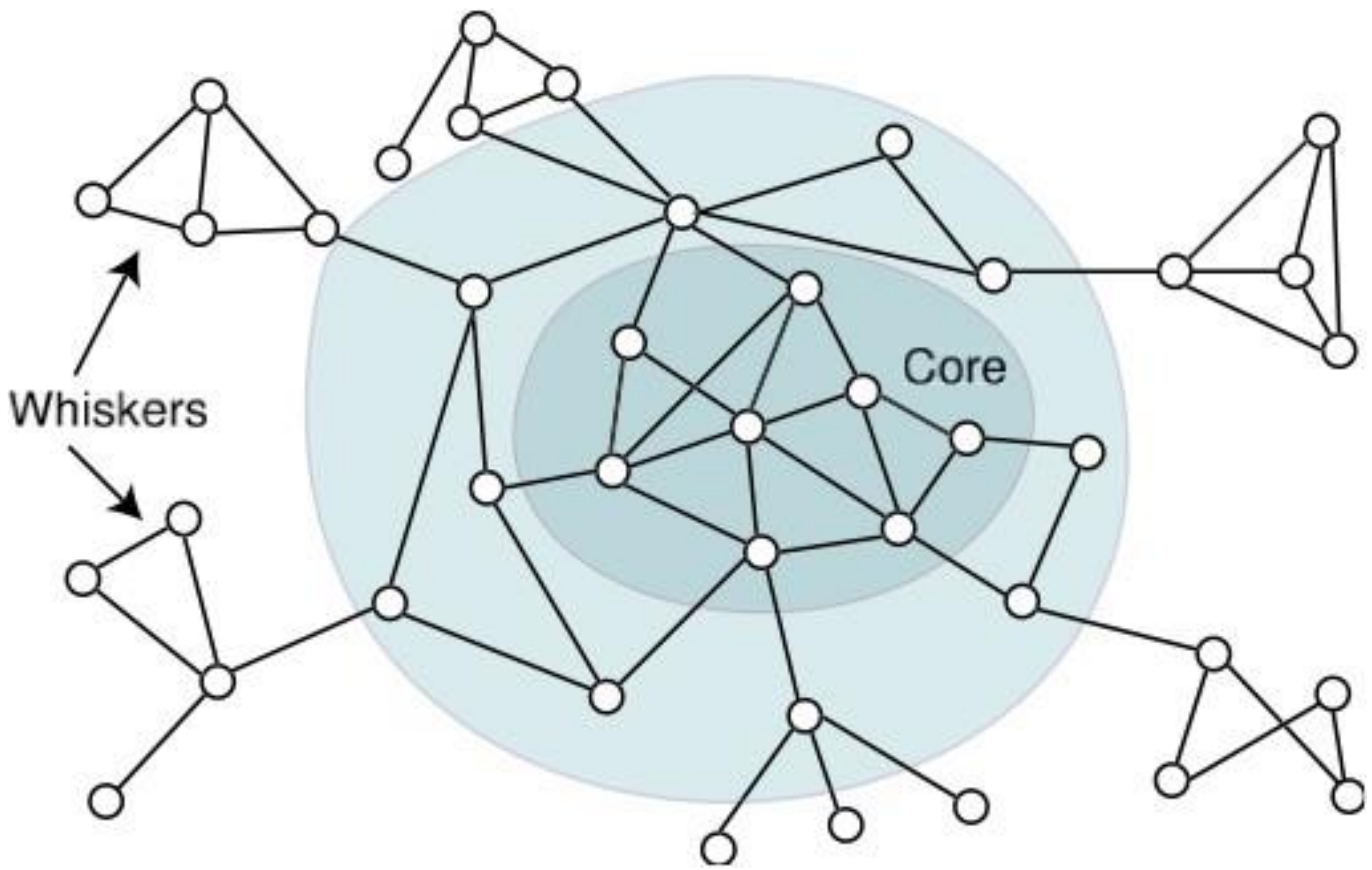
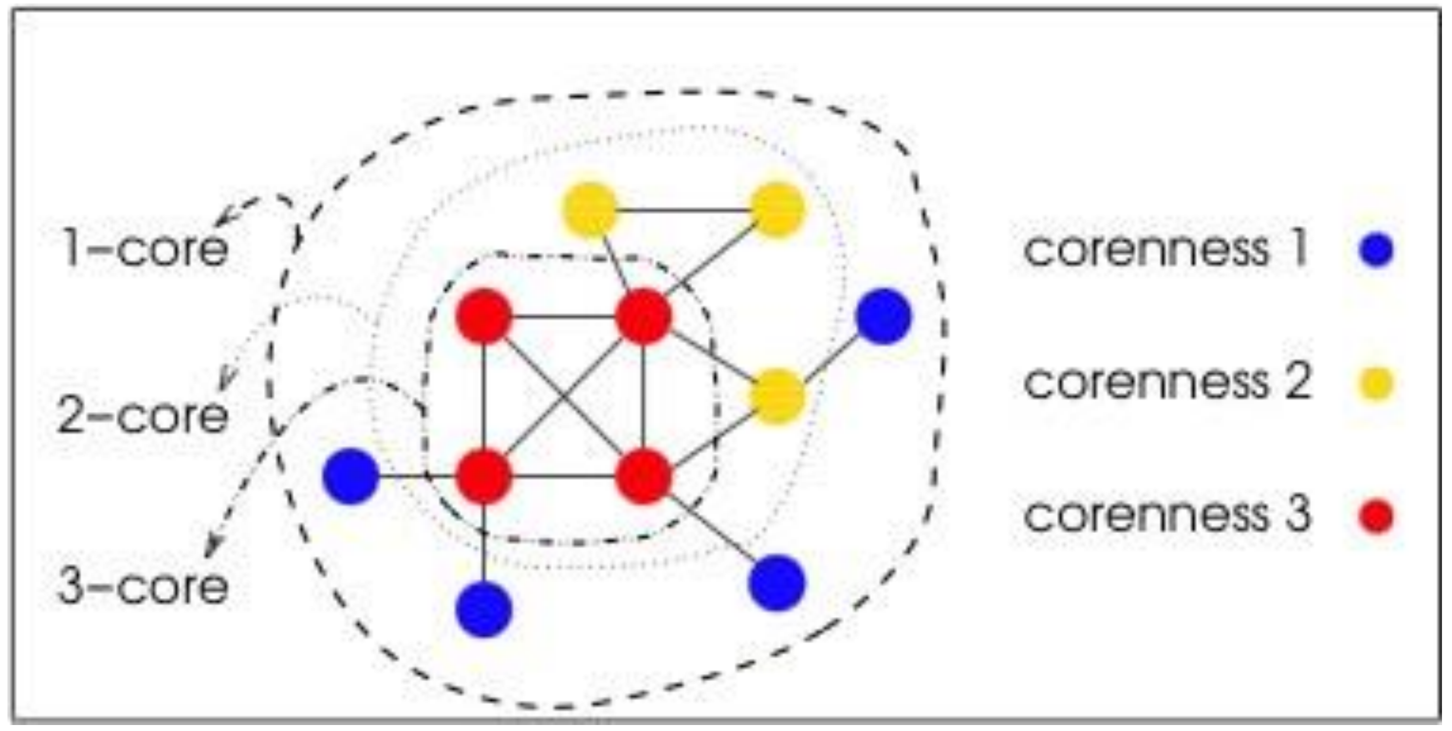
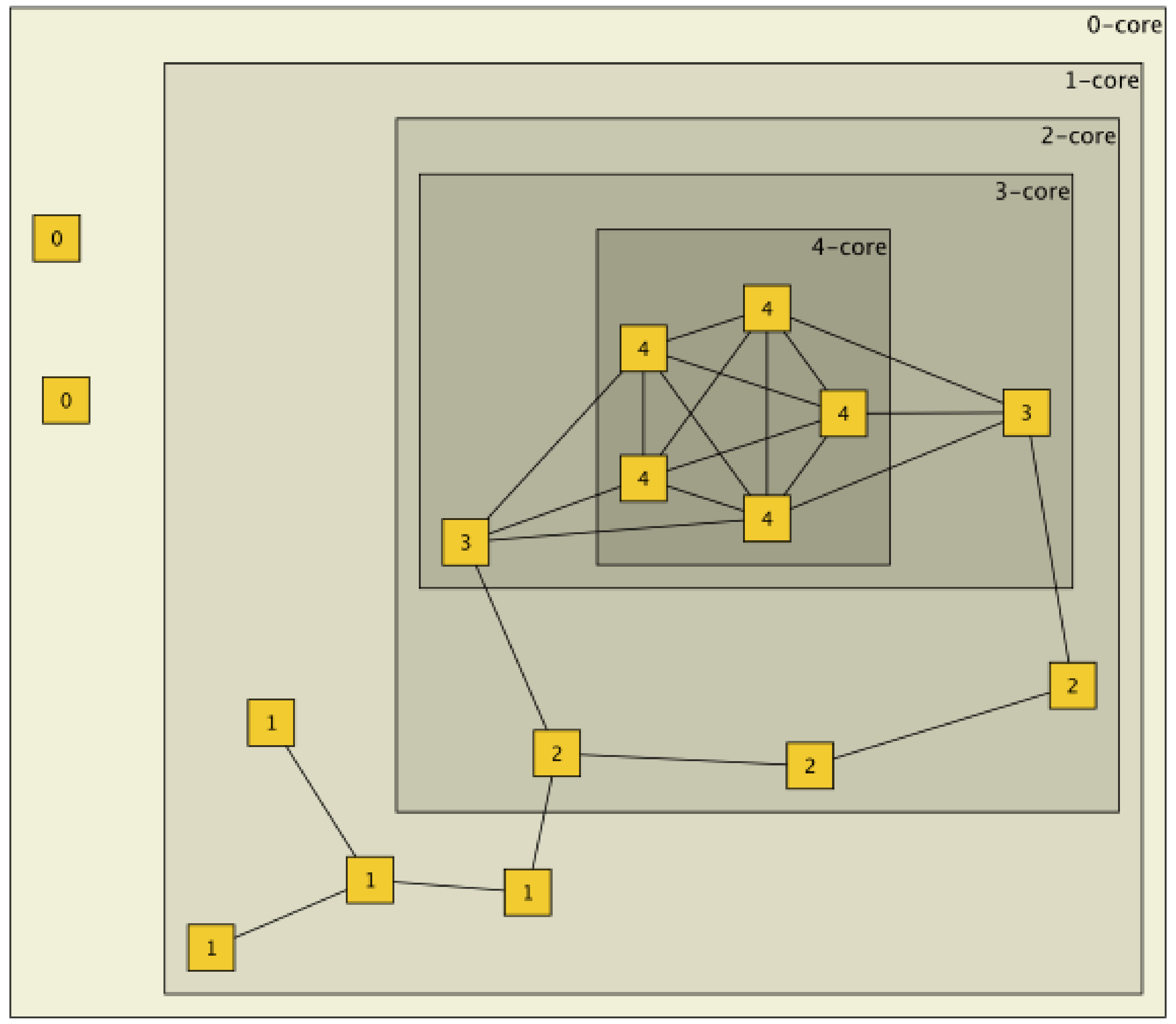


image from J. Leskovec, K. Lang, 2010

GRAPH CORES

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



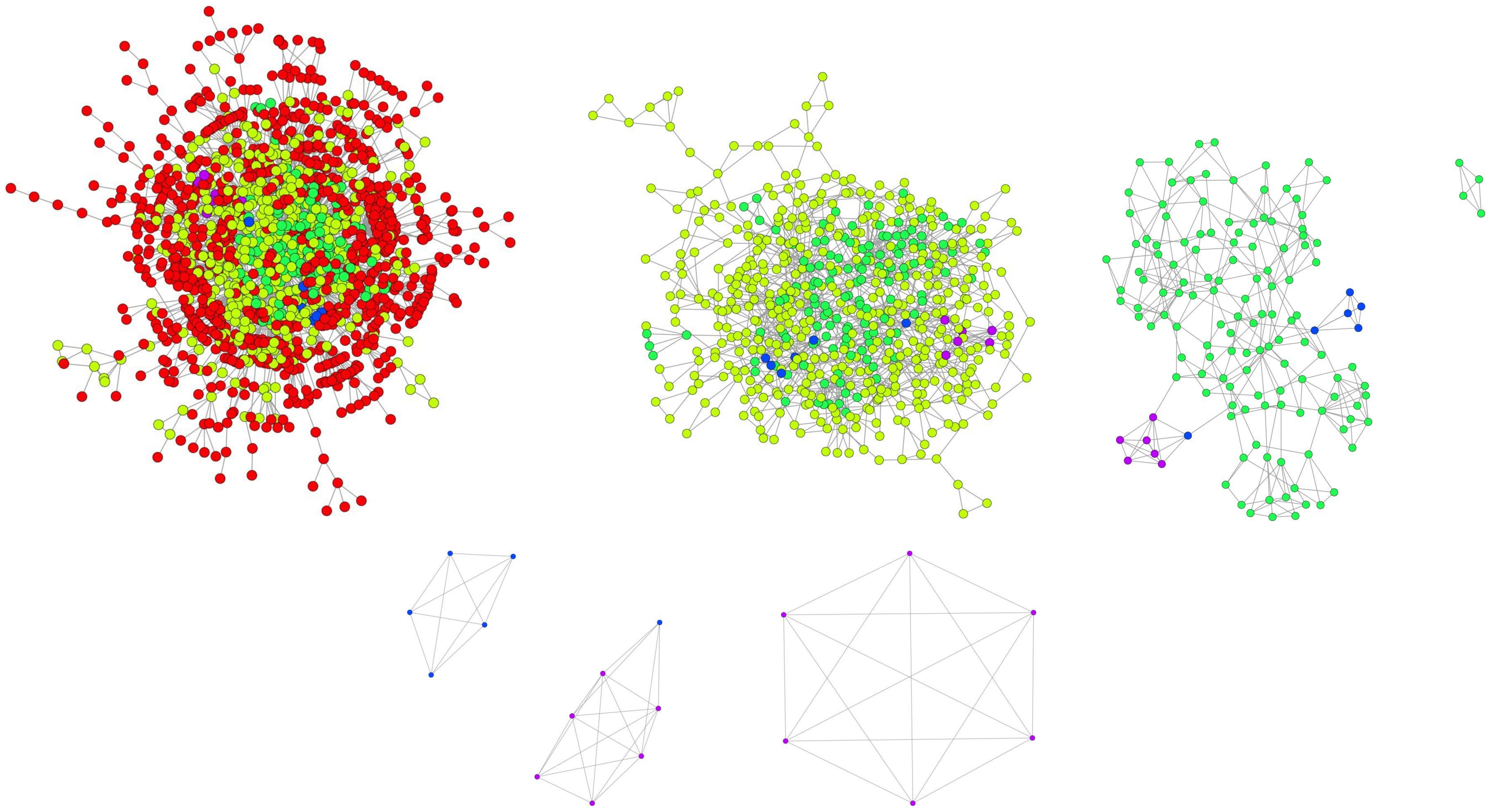


Every vertex in k-core has a degree $k_i \geq k$

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

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

K-CORES



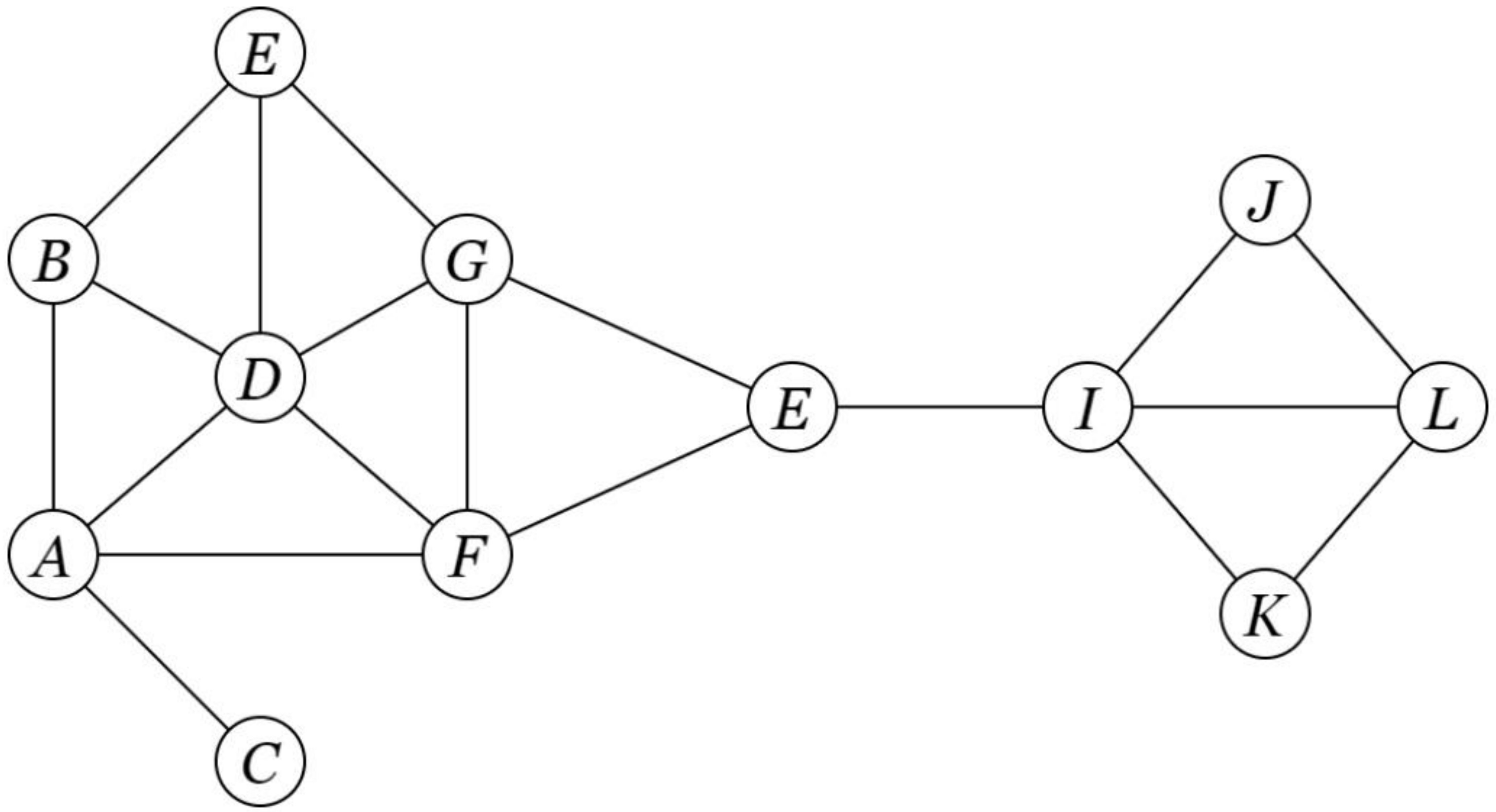
k-cores: 1:1458, 2:594, 3:142, 4:12, 5:6

k-shells: 1:864-red, 2:452-pale green, 3:130-green, 5:6-blue, 6:6-purple

R:graph.coreness(gcc)

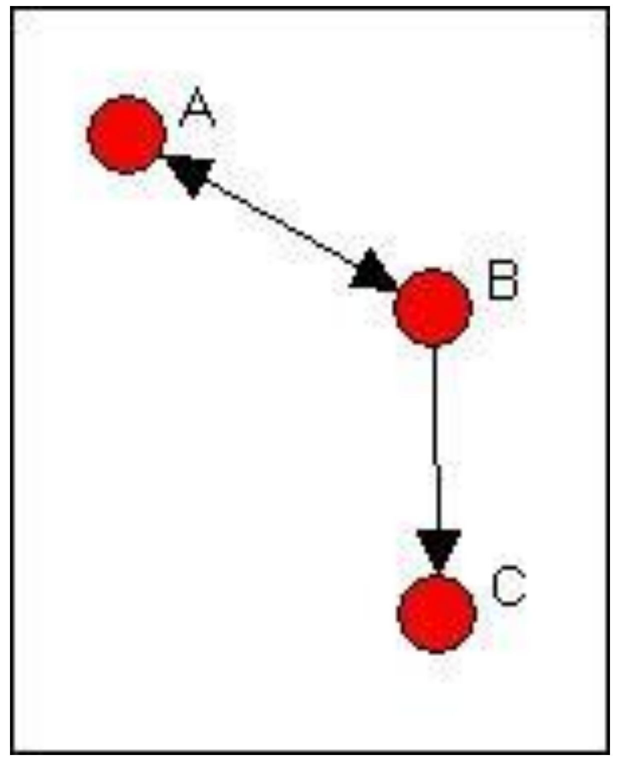
K-CORES

Find 3-core of the given network



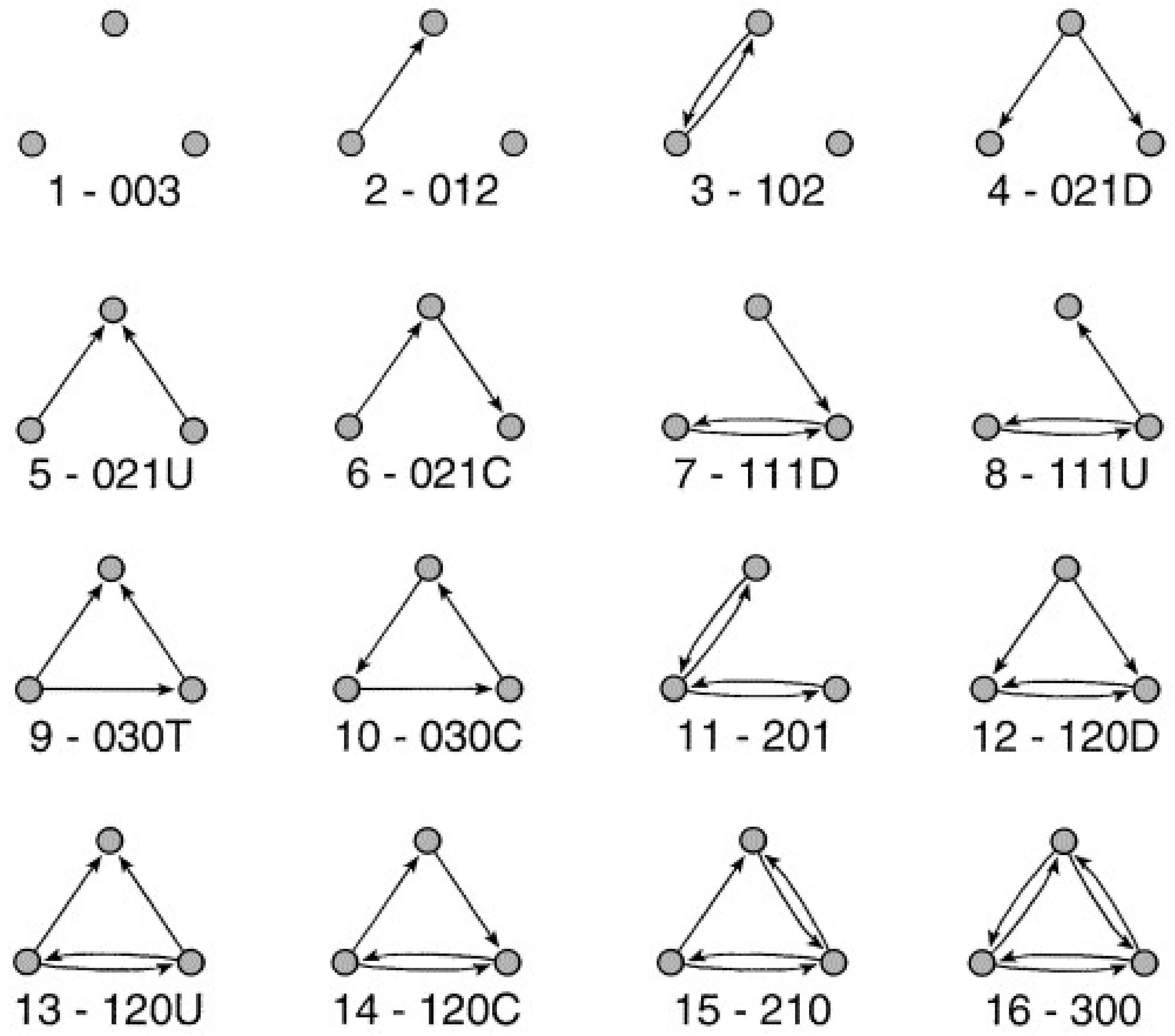
DYADS

Dyad is a pair of vertices and possible relational ties between them:
– mutual
– asymmetric
– null (non-existent)



TRIADS

Triad is a subgraph of three vertices and possible ties between them:



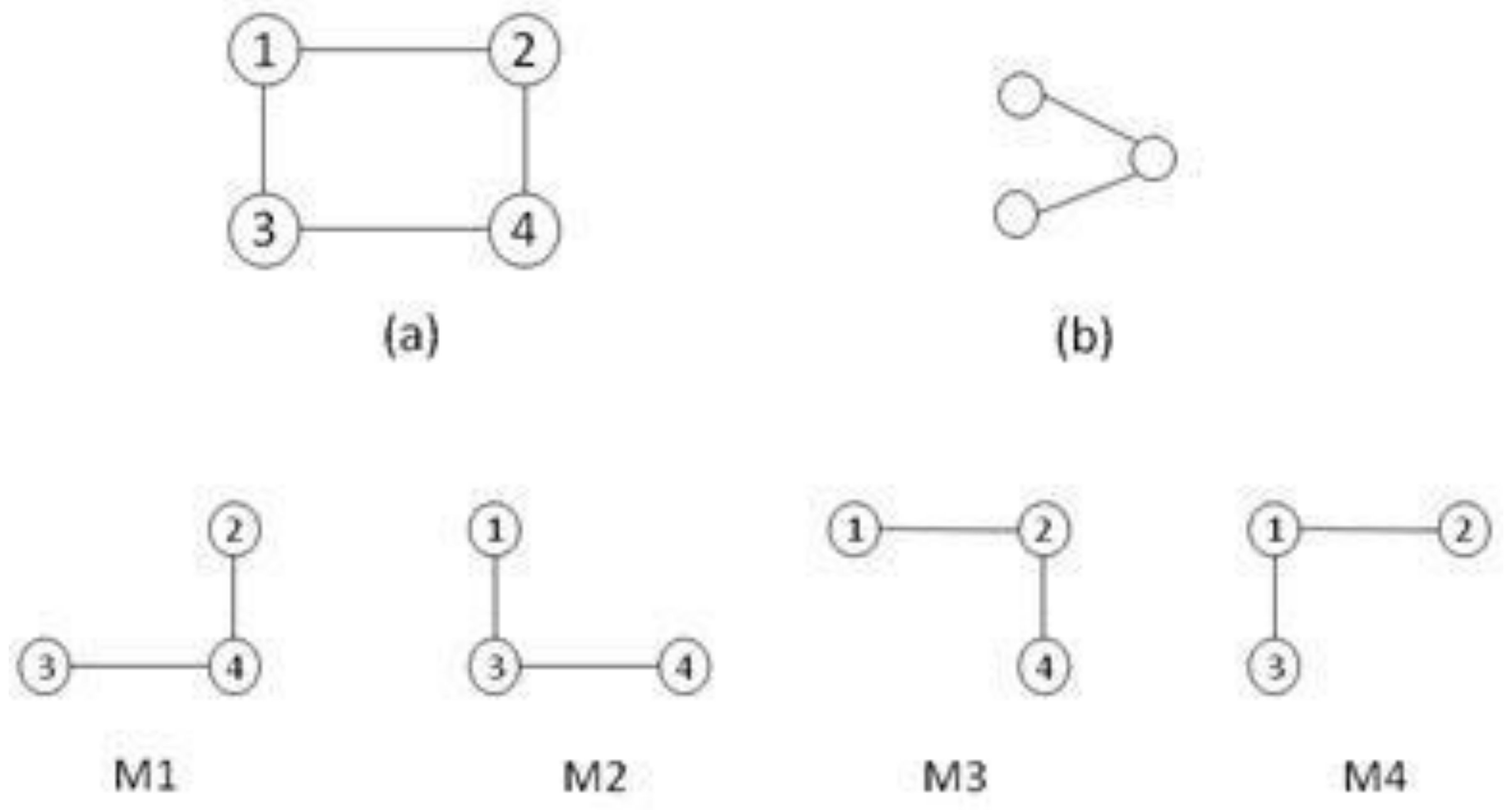
Triad census :16 isomorphism classes

D - down, U - up, T - transitive, C - cyclic.

| mutual diads | assymetric dyads | null dyads |

NETWORK MOTIFS

Network motifs are recurrent statistically significant subgraphs or patterns in graphs connected subgraphs that (compare to random network)



Motifs are not induced subgraphs, i.e. they do not contain all the graph edges between selected vertices.

NETWORK MOTIFS

Motifs appear in a network more frequently than in a comparable random network

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

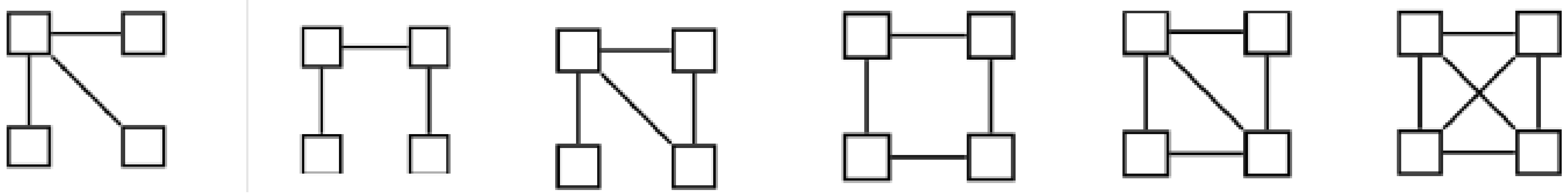
For G^t subgraph (motif candidate) of G ,

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

R - random graph, μ - mean frequency, σ-standard deviatiom

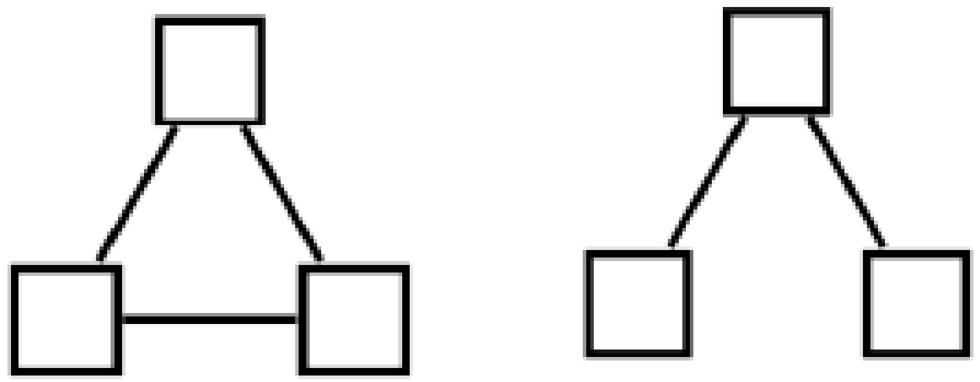
Undirected graphs: motifs of size 3 and 4

Connected Motifs of size 4



Star Chain 3-loop-out Box Semi-Clique Clique

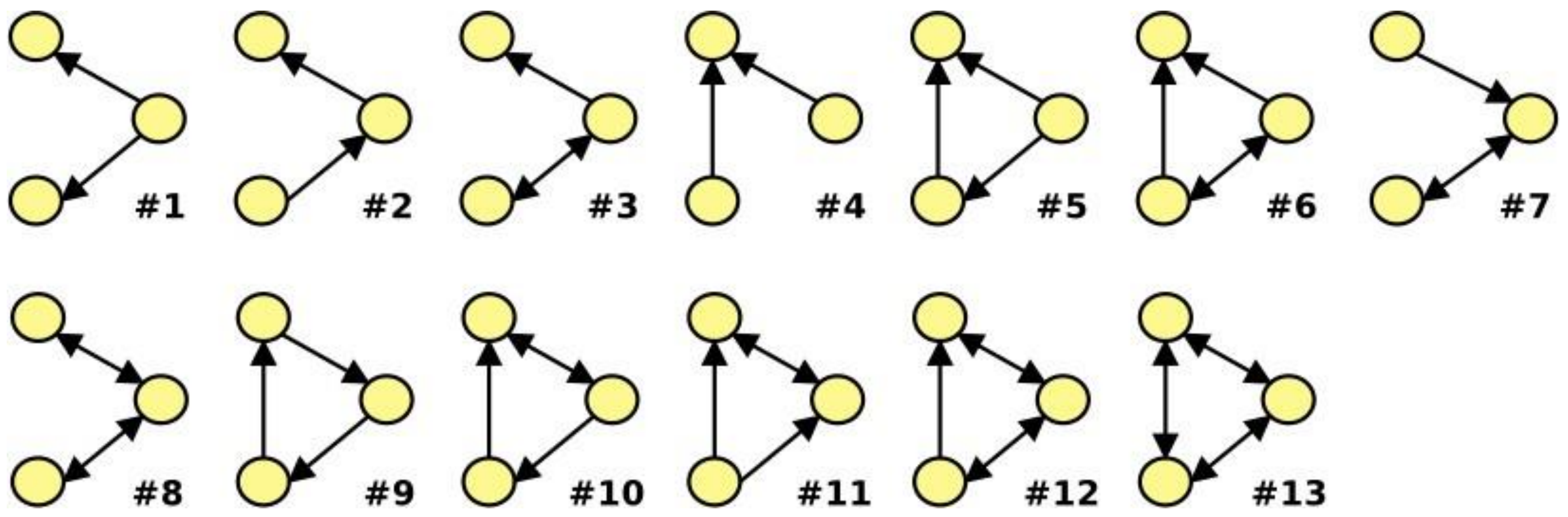
Connected Motifs of size 3



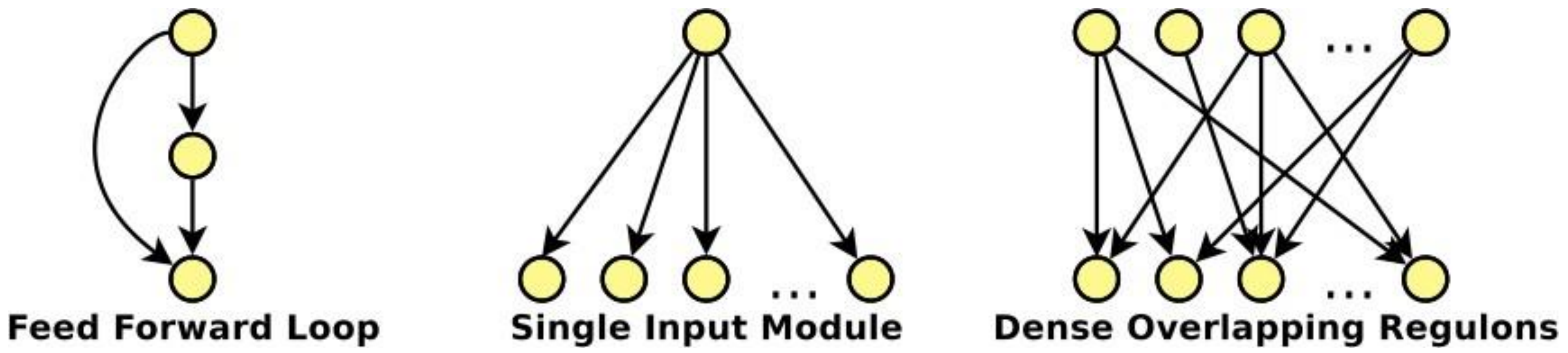
Clique Chain

NETWORK MOTIFS

Connected triads - motifs of size 3

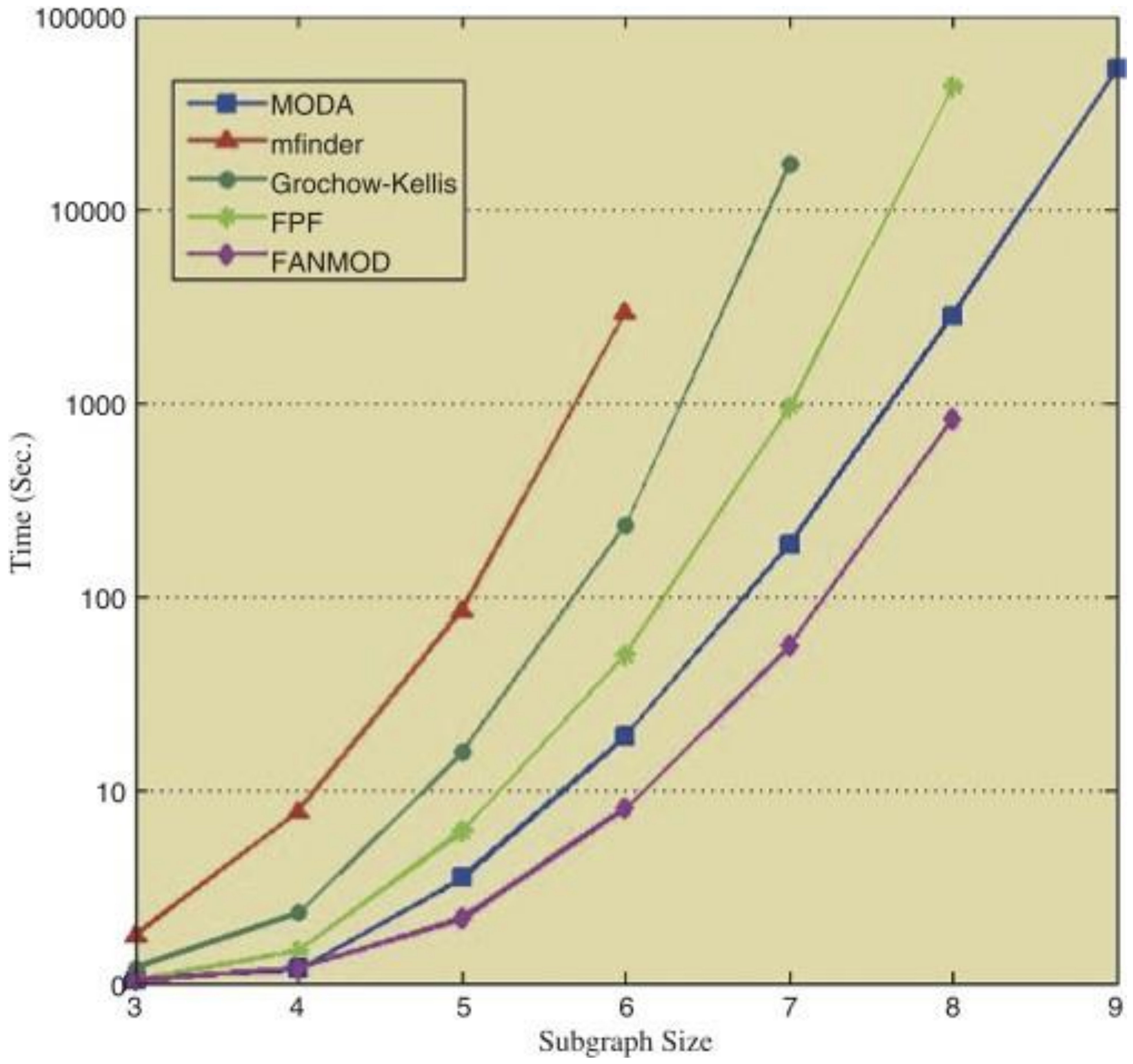


More complicated motifs:



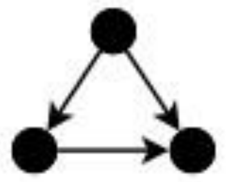
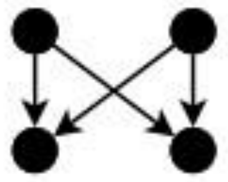
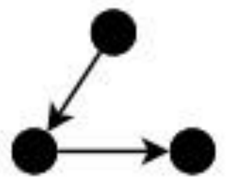
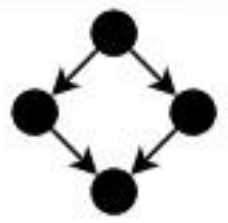
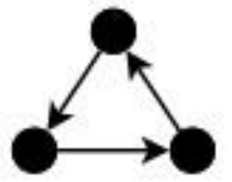
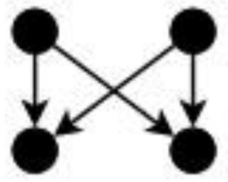
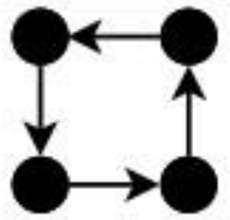
Ribeiro, 2011, Shen-Orr, 2002

NETWORK MOTIFS



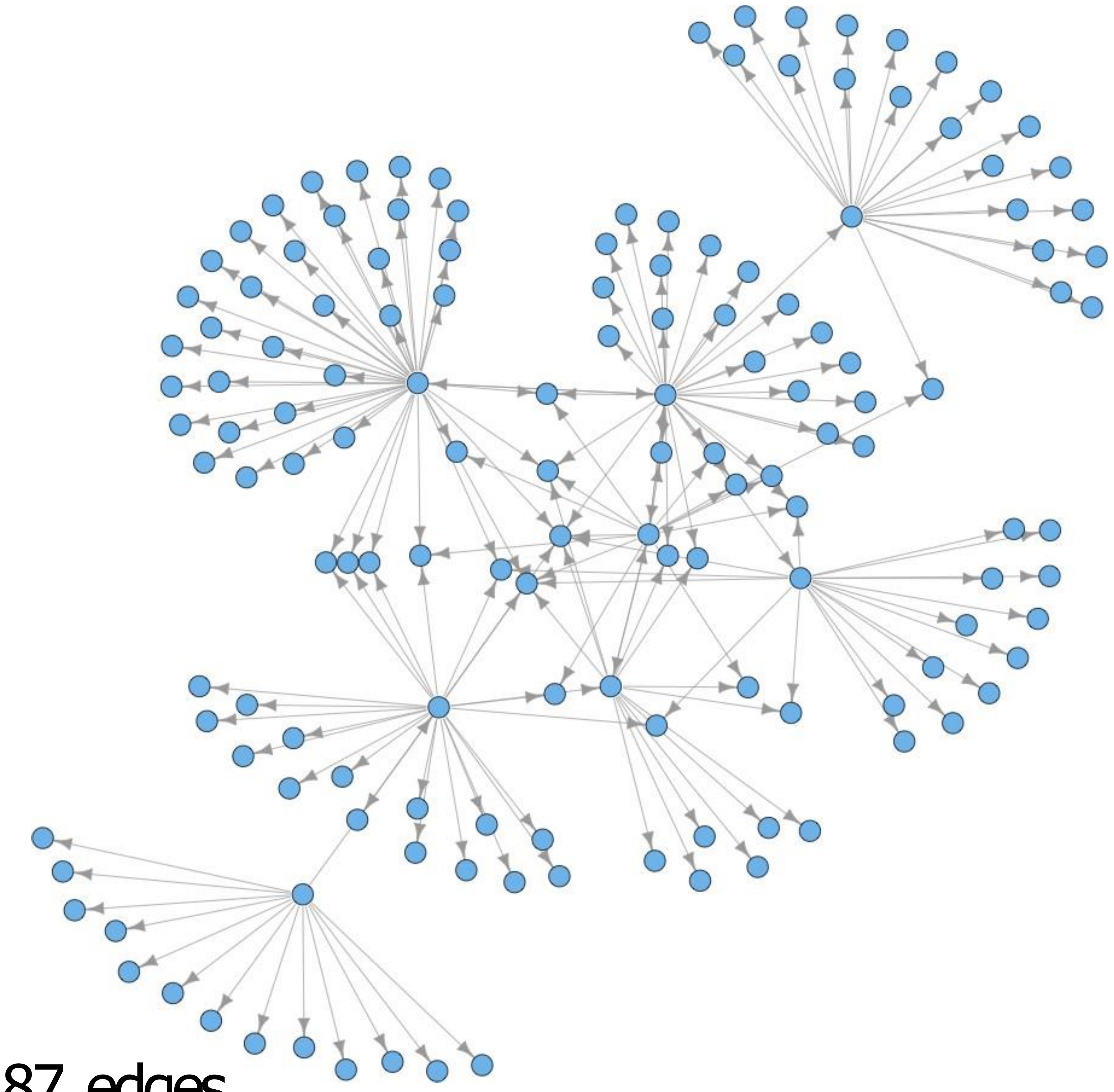
S. Omidi, 2009

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)					Feed-forward loop			Bi-Fan			
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					Three chain			Bi-Parallel			
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)					3-node loop			Bi-Fan			4-node loop
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

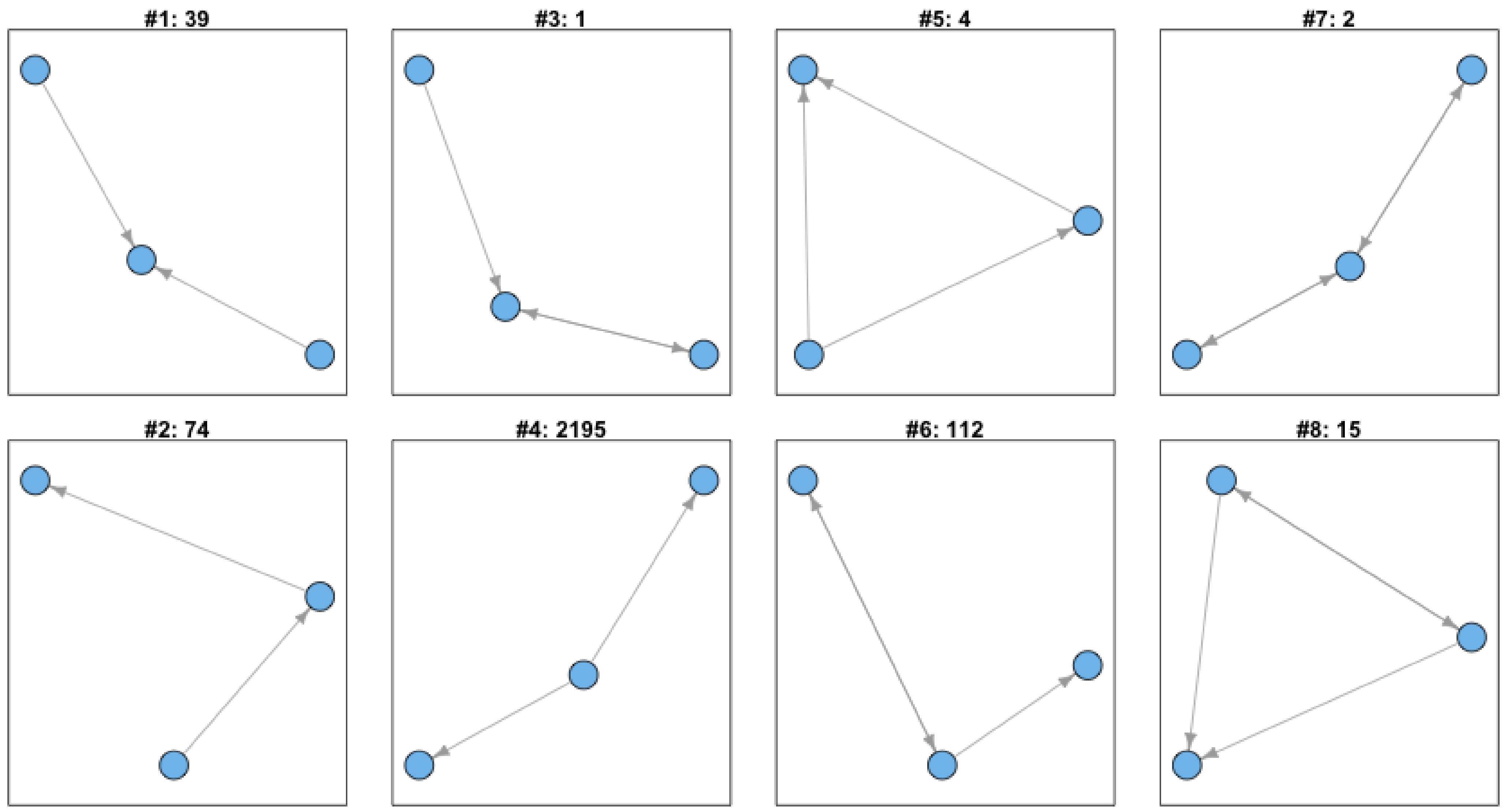
Ribeiro, 2011, Milo, 2002

NETWORK MOTIFS

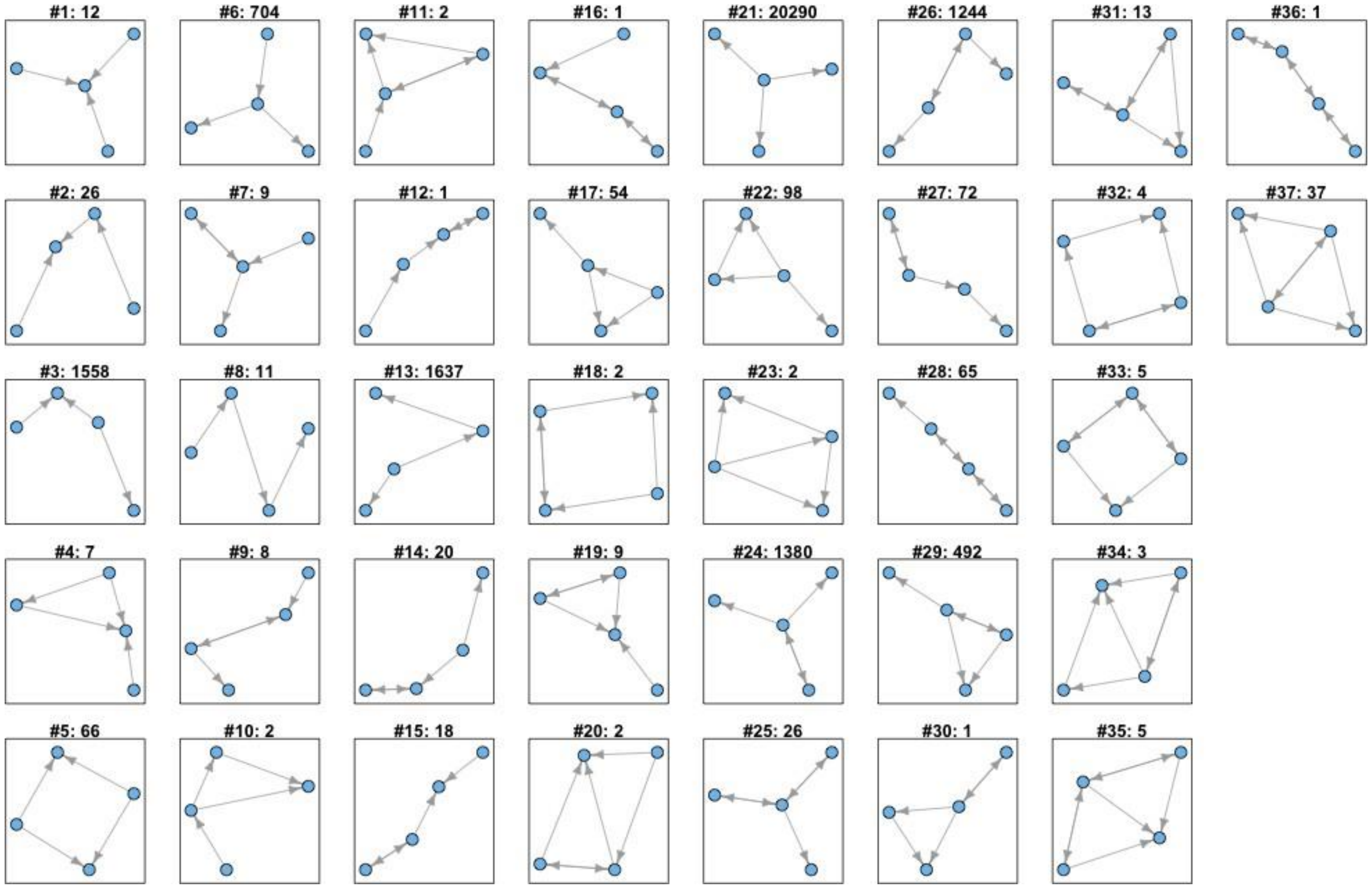


146 nodes, 187 edges

NETWORK MOTIFS



NETWORK MOTIFS



REFERENCES

- Centrality in Social Networks. Conceptual Clarification, Linton C. Freeman, Social Networks, 1, 215-239, 1979
- Power and Centrality: A Family of Measures, Phillip Bonacich, The American Journal of Sociology, Vol. 92, No. 5, 1170-1182, 1987
- A new status index derived from sociometric analysis, L. Katz, Psychometrika, 19, 39-43, 1953.
- Eigenvector-like measures of centrality for asymmetric relations, Phillip Bonacich, Paulette Lloyd, Social Networks 23, 191-201, 2001
- The PageRank Citation Ranknig: Bringing Order to the Web. S. Brin, L. Page, R. Motwany, T. Winograd, Stanford Digital Library Technologies Project, 1998
- Authoritative Sources in a Hyperlinked Environment. Jon M. Kleinberg, Proc. 9th ACM-SIAM Symposium on Discrete Algorithms
- A Survey of Eigenvector Methods of Web Information Retrieval. Amy N. Langville and Carl D. Meyer, 2004
- PageRank beyond the Web. David F. Gleich, arXiv:1407.5107, 2014