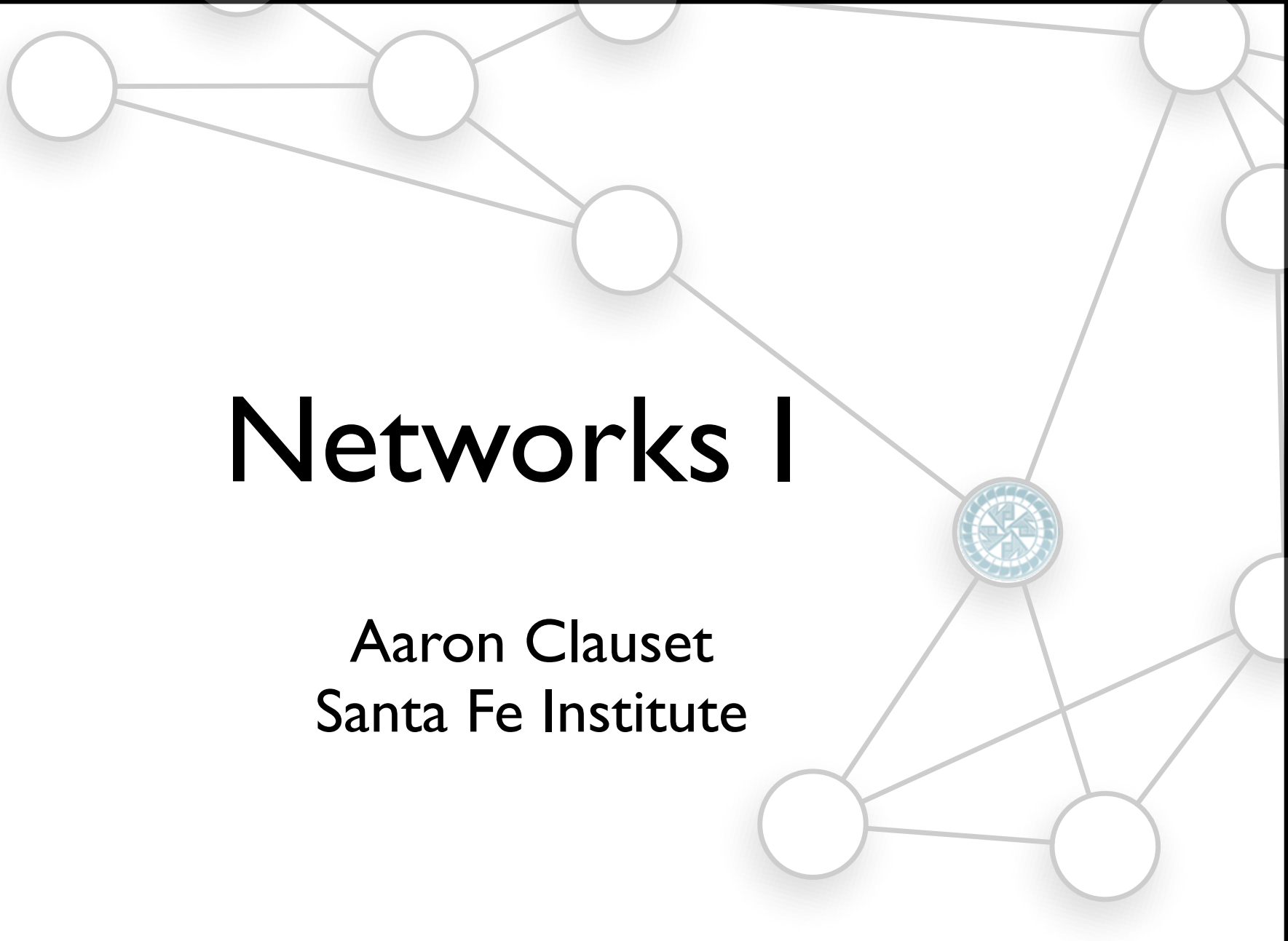
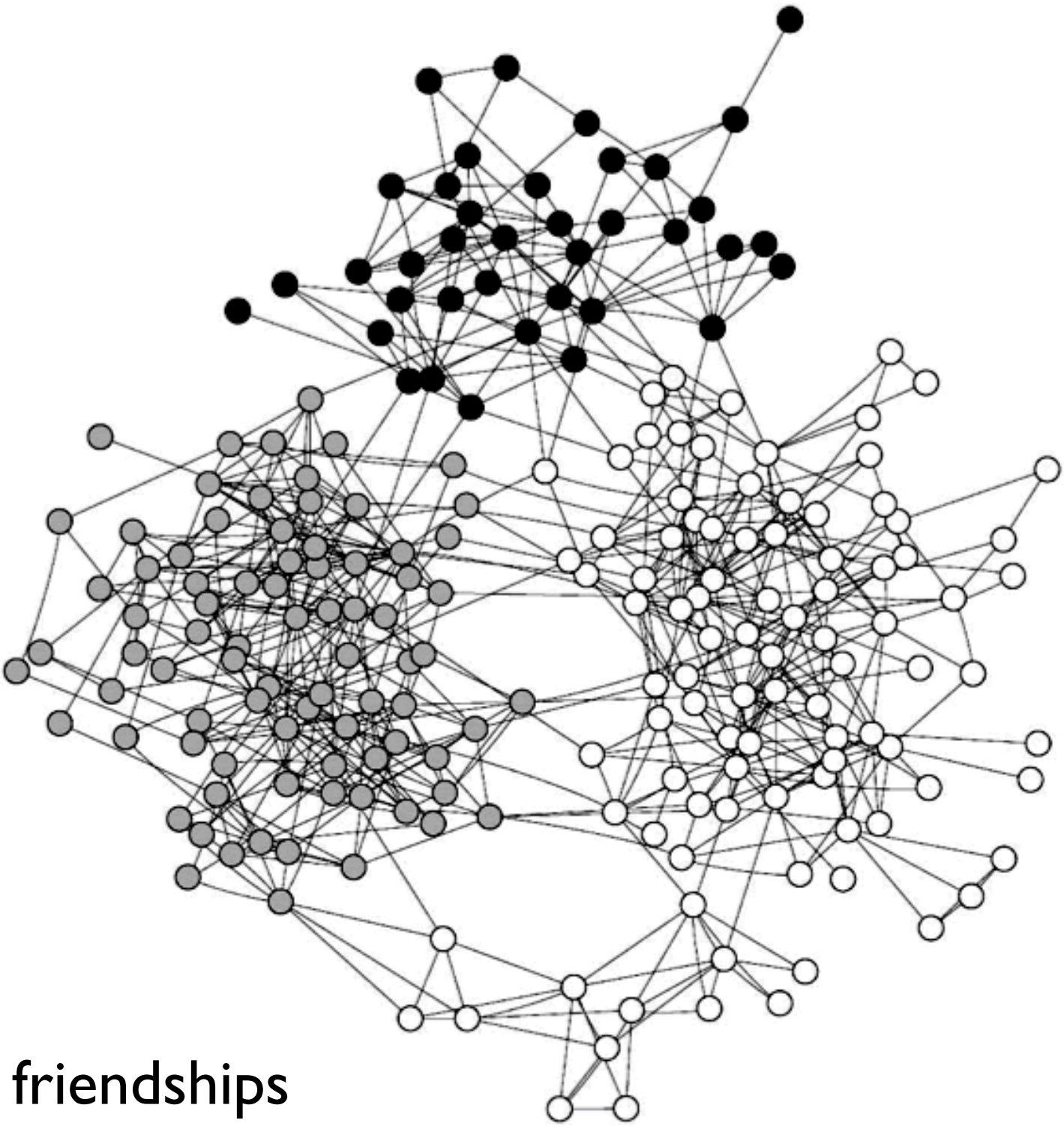
Networks I

Aaron Clauset
Santa Fe Institute

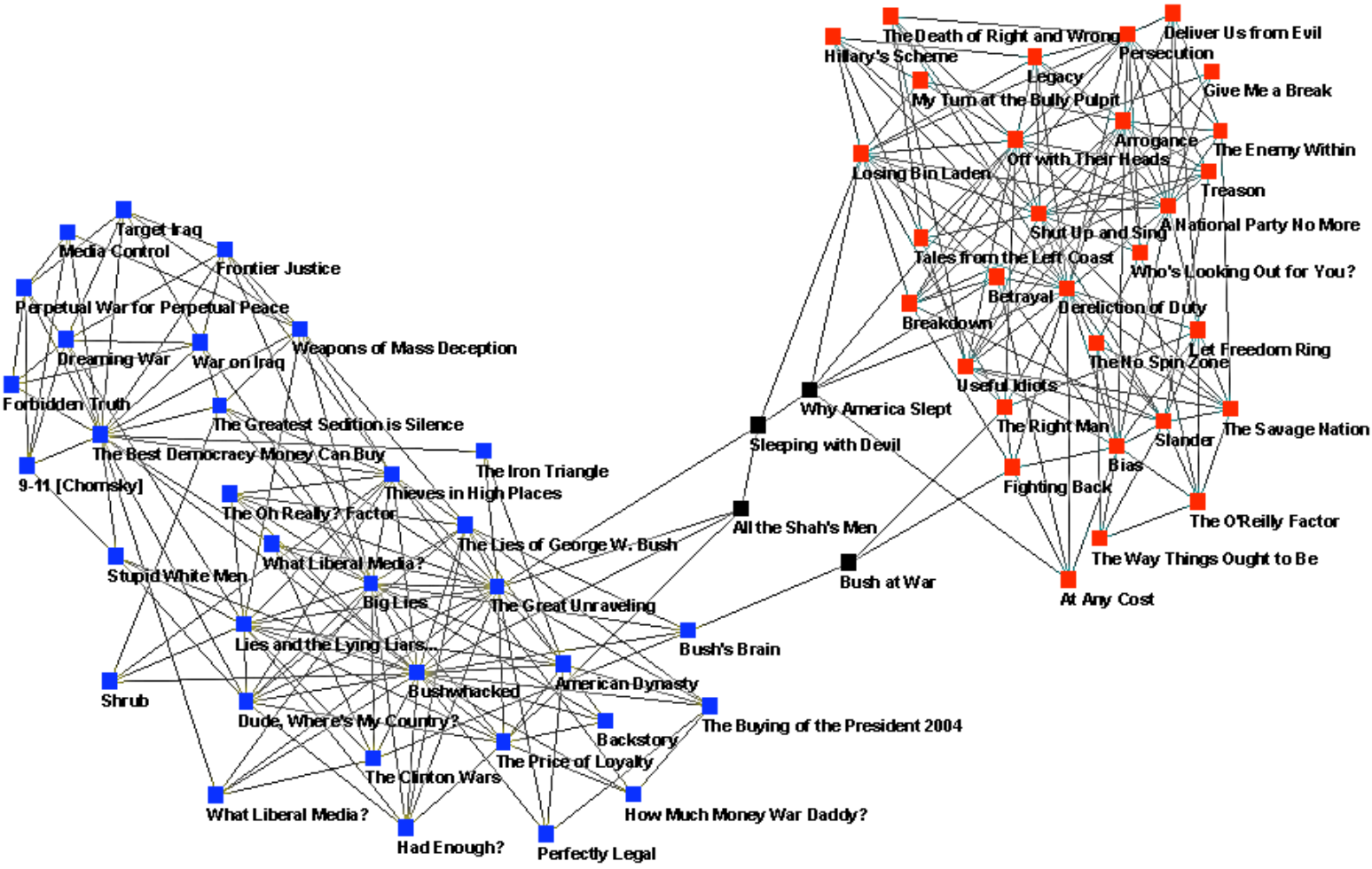


social



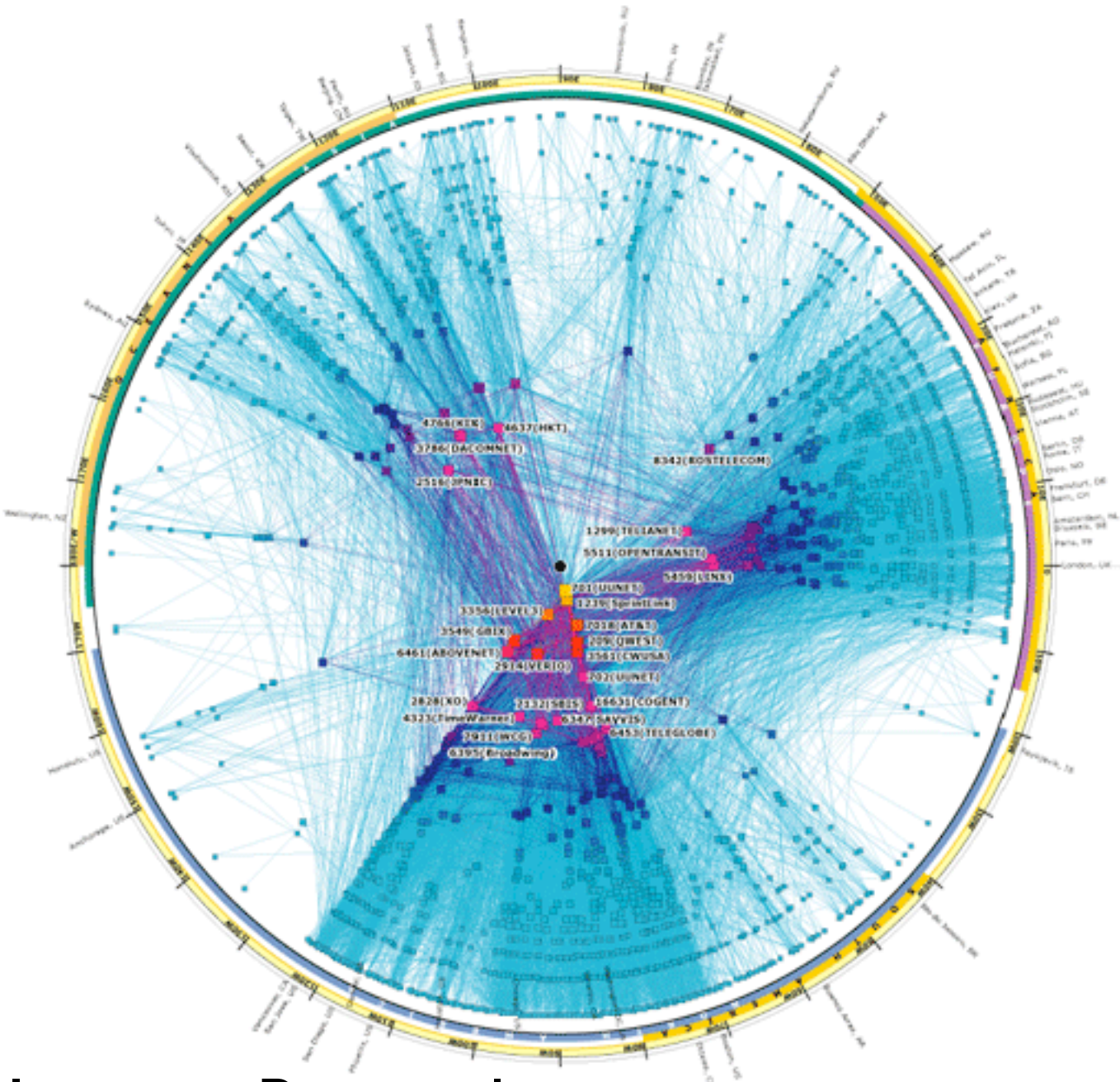
- friendships
- coauthorship
- sexual contacts
- etc.

informational



- citations
- recommendations
- hyperlinks
- similarity
- etc.

technological



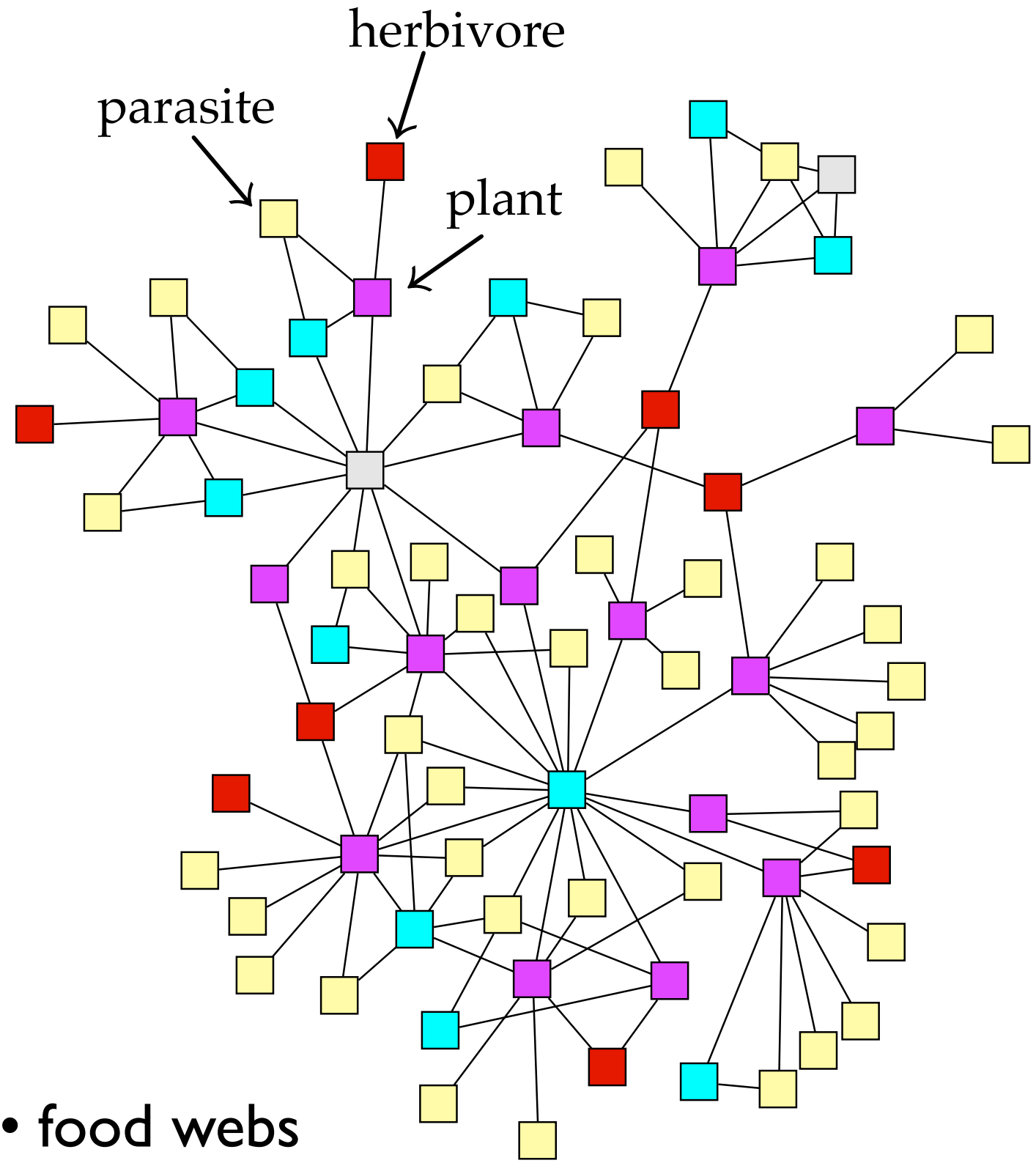
- Internet Protocol
- Bluetooth
- BGP
- etc.

caida

spatial

- transportation
- distribution
- etc.

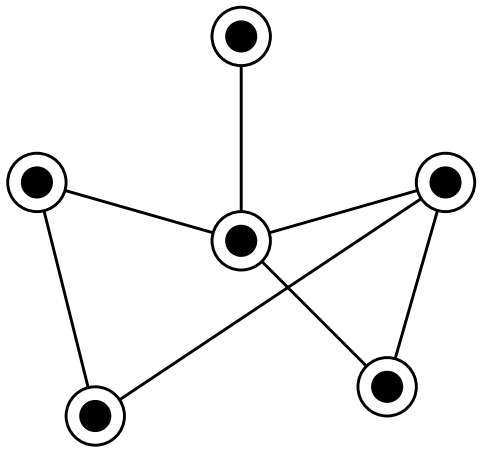
biological



- food webs
- biochemical reactions
- gene regulations
- neurons
- etc.

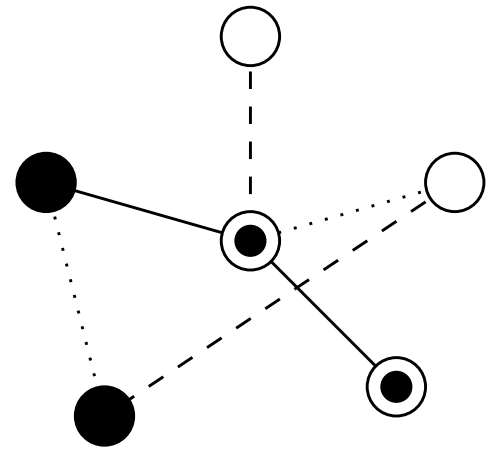
undirected
unweighted
one edge type
one node type

(a)



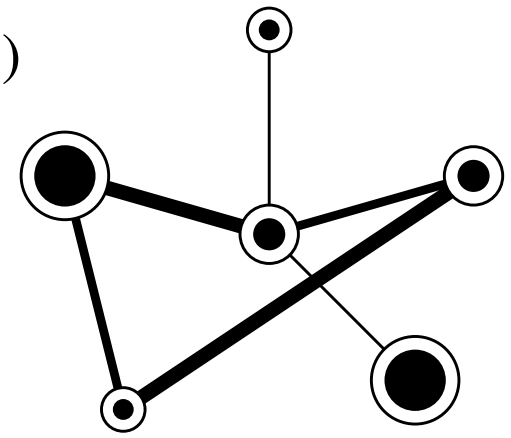
undirected
unweighted
several edge types
several node types

(b)



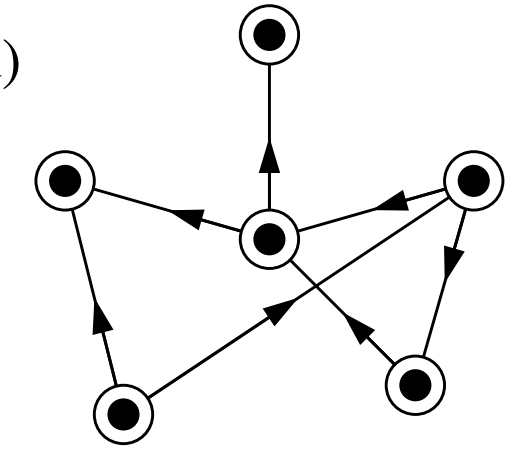
undirected
one edge type
one node type
weighted edges
weighted nodes

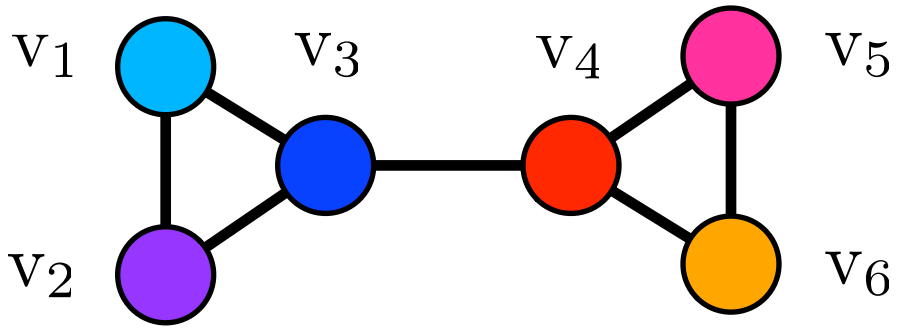
(c)



directed edges
unweighted
one node type
one edge type

(d)





adjacency matrix

	v_1	v_2	v_3	v_4	v_5	v_6
v_1	0	1	1	0	0	0
v_2	1	0	1	0	0	0
v_3	1	1	0	1	0	0
v_4	0	0	1	0	1	1
v_5	0	0	0	1	0	1
v_6	0	0	0	1	1	0

adjacency list

1	2
1	3
2	3
3	4
4	5
4	6
5	6

network science: goals

1. basic principles
   - modularity, hierarchy, core-periphery
2. structure + dynamics = function
   - evolution, robustness, adaptability, computation
3. application
   - prediction, manipulation, design

network science: methods

1. summary statistics

   degrees, centrality, etc.

2. physics-style models

   preferential attachment, etc.

3. model-based inference

   hierarchical random graphs, etc.

4. others

summary statistics

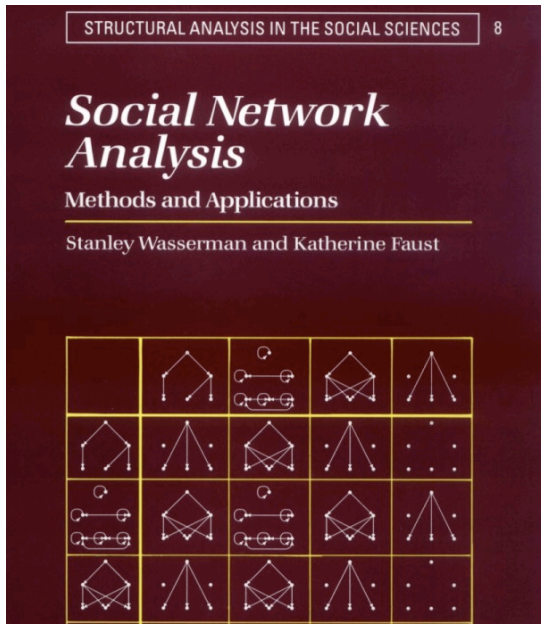
degree distribution

clustering coefficient

eigenvector centrality

betweenness

closeness



Wasserman & Faust

pair-wise distance distribution

rich-club coefficient

diameter

degree-correlations

motif frequency

etc.

summary statistics : nodes

degree distribution

$$d_i = \sum_j A_{i,j} \qquad \Pr(x = d_i)$$

clustering coefficient

$$C_i = \frac{\#\,\text{triangles connected to } i}{\#\,\text{triples connected to } i} \qquad C = \frac{1}{n}\sum_i C_i$$

eigenvector centrality (PageRank)

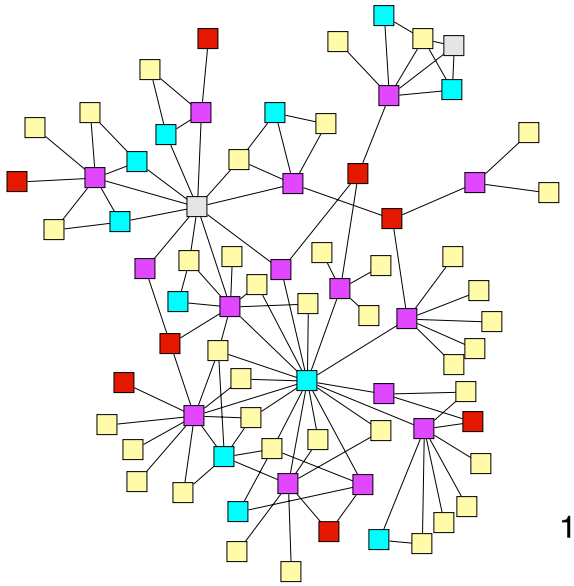
$$A\mathbf{x} = \lambda\mathbf{x} \qquad \text{or} \qquad x_i = \frac{1}{\lambda}\sum_{j=1}^{n} A_{i,j}\, x_j$$

betweenness

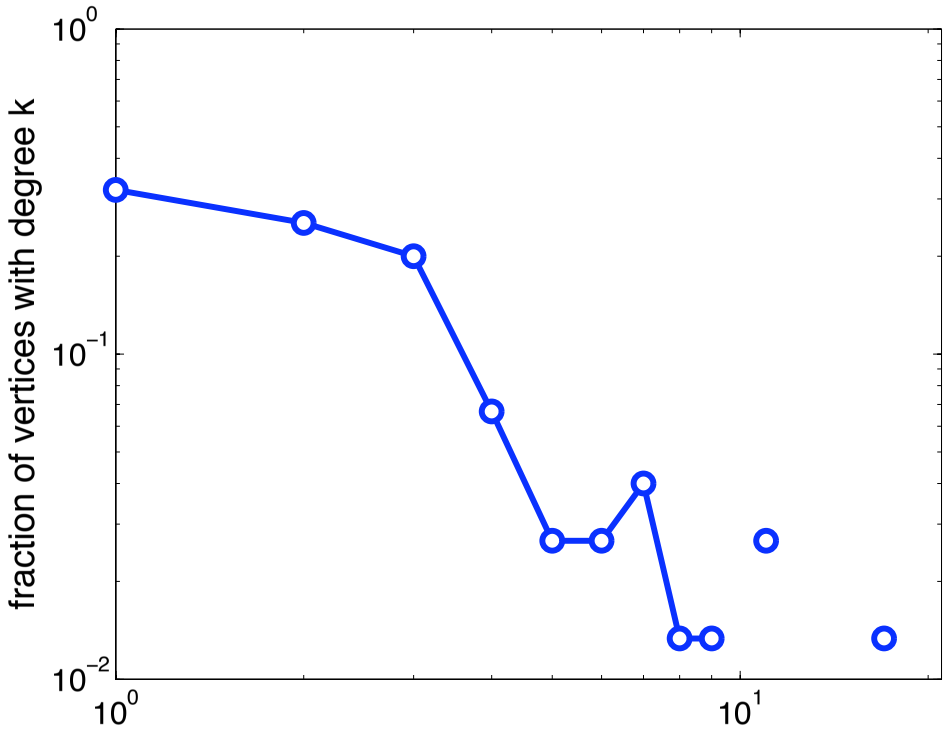
$$B_k = \sum_{i,j} \frac{\#\text{shortest paths}(i \to \cdots \to k \to \cdots \to j)}{\#\text{shortest paths}(i \to \cdots \to j)}$$

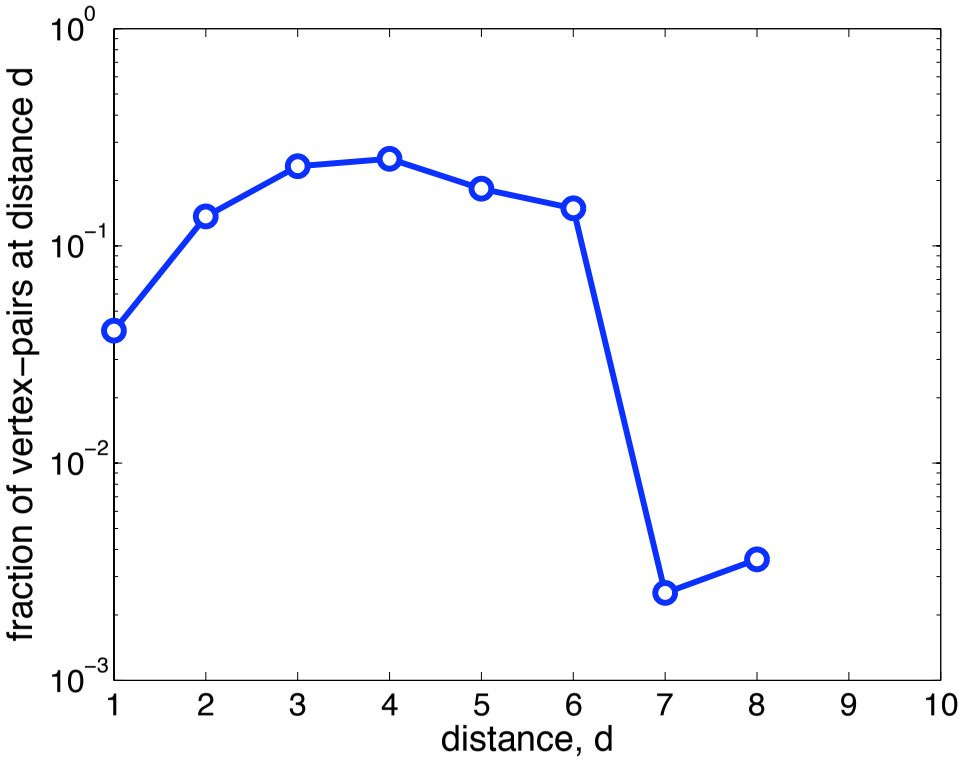
closeness

$$R_i = \frac{1}{n-1}\sum_{j \neq i} \text{shortest-distance}(i \to j)$$



$$n = 75$$
$$m = 113$$
$$\langle k \rangle = 3.013$$
$$\langle C \rangle = 0.174$$





	network	type	n	m	z	ℓ	α	$C^{(1)}$	$C^{(2)}$	r
social	film actors	undirected	449 913	25 516 482	113.43	3.48	2.3	0.20	0.78	0.208
	company directors	undirected	7 673	55 392	14.44	4.60	–	0.59	0.88	0.276
	math coauthorship	undirected	253 339	496 489	3.92	7.57	–	0.15	0.34	0.120
	physics coauthorship	undirected	52 909	245 300	9.27	6.19	–	0.45	0.56	0.363
	biology coauthorship	undirected	1 520 251	11 803 064	15.53	4.92	–	0.088	0.60	0.127
	telephone call graph	undirected	47 000 000	80 000 000	3.16		2.1			
	email messages	directed	59 912	86 300	1.44	4.95	1.5/2.0		0.16	
	email address books	directed	16 881	57 029	3.38	5.22	–	0.17	0.13	0.092
	student relationships	undirected	573	477	1.66	16.01	–	0.005	0.001	−0.029
	sexual contacts	undirected	2 810				3.2			
information	WWW `nd.edu`	directed	269 504	1 497 135	5.55	11.27	2.1/2.4	0.11	0.29	−0.067
	WWW Altavista	directed	203 549 046	2 130 000 000	10.46	16.18	2.1/2.7			
	citation network	directed	783 339	6 716 198	8.57		3.0/–			
	Roget's Thesaurus	directed	1 022	5 103	4.99	4.87	–	0.13	0.15	0.157
	word co-occurrence	undirected	460 902	17 000 000	70.13		2.7		0.44	
technological	Internet	undirected	10 697	31 992	5.98	3.31	2.5	0.035	0.39	−0.189
	power grid	undirected	4 941	6 594	2.67	18.99	–	0.10	0.080	−0.003
	train routes	undirected	587	19 603	66.79	2.16	–		0.69	−0.033
	software packages	directed	1 439	1 723	1.20	2.42	1.6/1.4	0.070	0.082	−0.016
	software classes	directed	1 377	2 213	1.61	1.51	–	0.033	0.012	−0.119
	electronic circuits	undirected	24 097	53 248	4.34	11.05	3.0	0.010	0.030	−0.154
	peer-to-peer network	undirected	880	1 296	1.47	4.28	2.1	0.012	0.011	−0.366
biological	metabolic network	undirected	765	3 686	9.64	2.56	2.2	0.090	0.67	−0.240
	protein interactions	undirected	2 115	2 240	2.12	6.80	2.4	0.072	0.071	−0.156
	marine food web	directed	135	598	4.43	2.05	–	0.16	0.23	−0.263
	freshwater food web	directed	92	997	10.84	1.90	–	0.20	0.087	−0.326
	neural network	directed	307	2 359	7.68	3.97	–	0.18	0.28	−0.226

E II Basic statistics for a number of published networks. The properties measured are: type of graph, directed or undirected; total numb
er of edges m; mean degree z; mean vertex–vertex distance ℓ; exponent α of degree distribution if the distribution follows a power law (or "–
ents are given for directed graphs); clustering coefficient $C^{(1)}$ from Eq. (3); clustering coefficient $C^{(2)}$ from Eq. (6); and degree correlation c
ast column gives the citation(s) for the network in the bibliography. Blank entries indicate unavailable data.

summary statistics : caveats

sensitive to noise, perturbations

strong correlations

not “sufficient” statistics

indirect relation to function

assume null-model (often)

large-scale patterns

core-periphery

- high-low centrality
- *social*: importance, influence, power?
- *spatial*: efficiency?

modularity

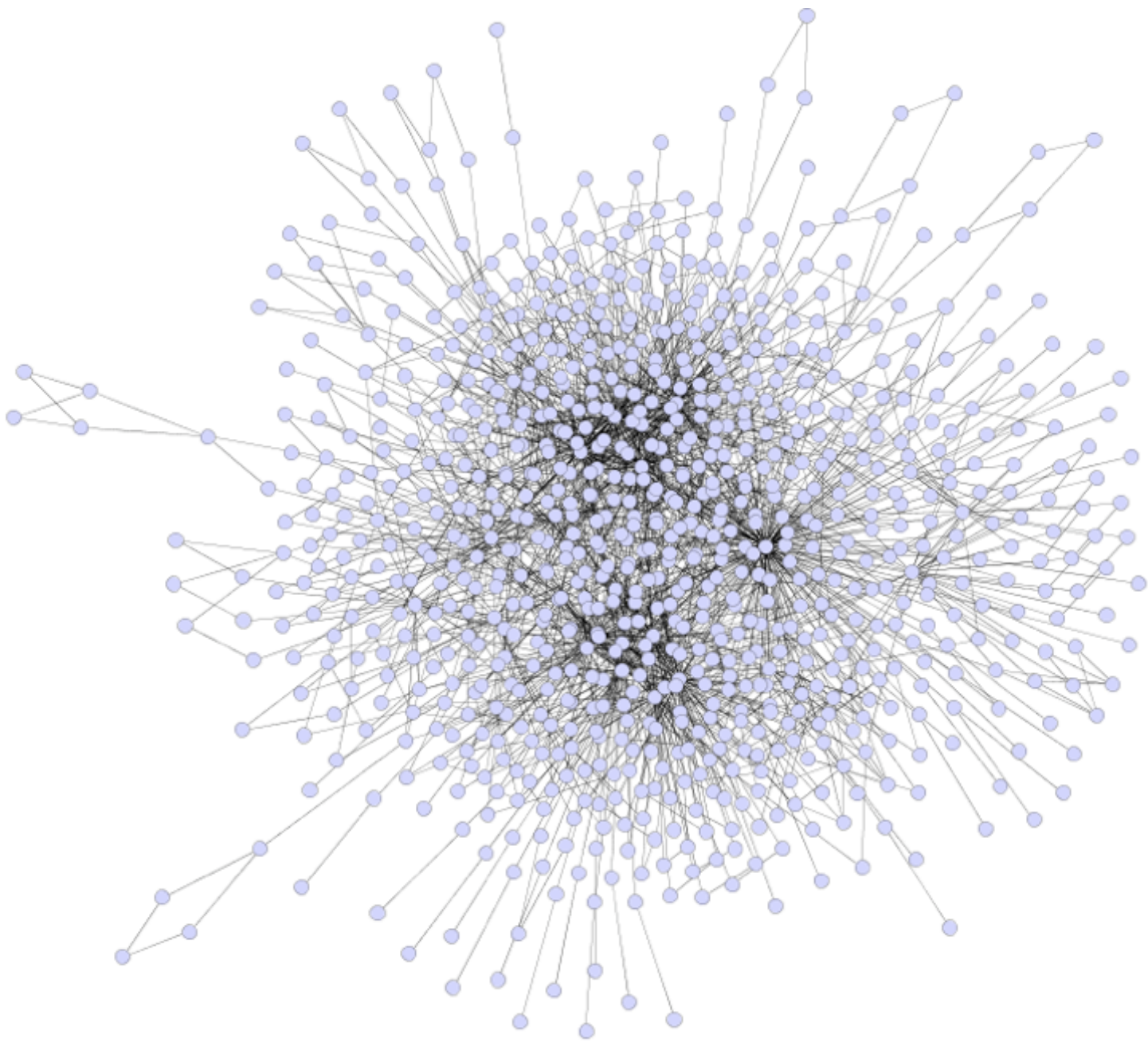
- "community structure"
- cluster node statistics
- functionality? robustness?
- modularity $Q = \sum_i (e_{ii} - a_i^2)$

hierarchies

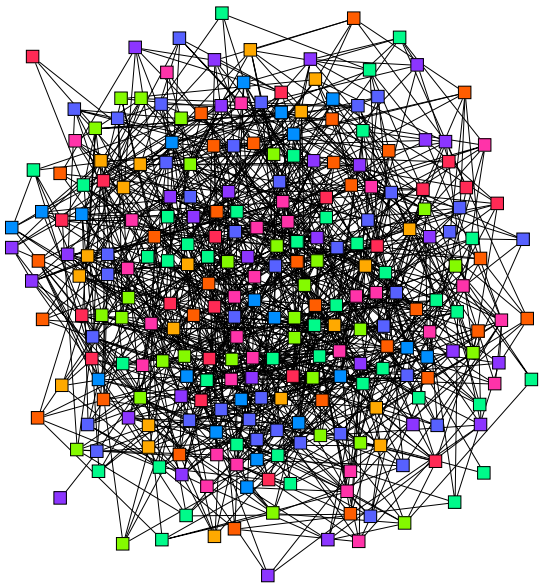
- nodes
- modules

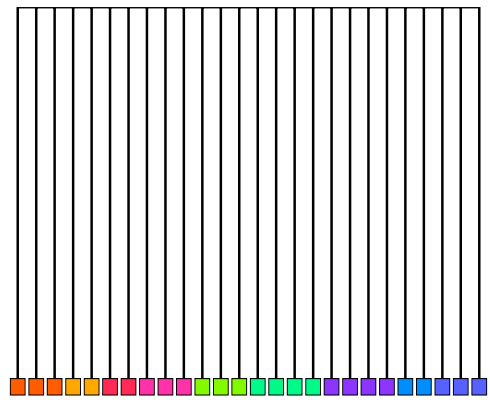
navigability

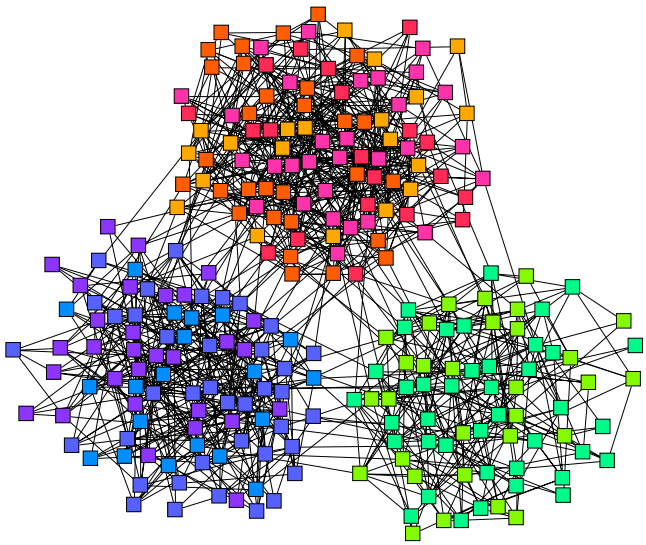
core-periphery

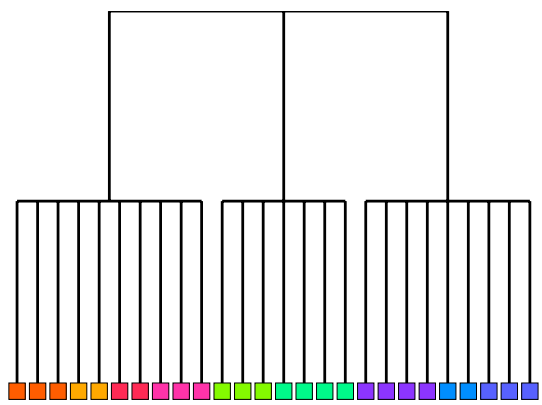


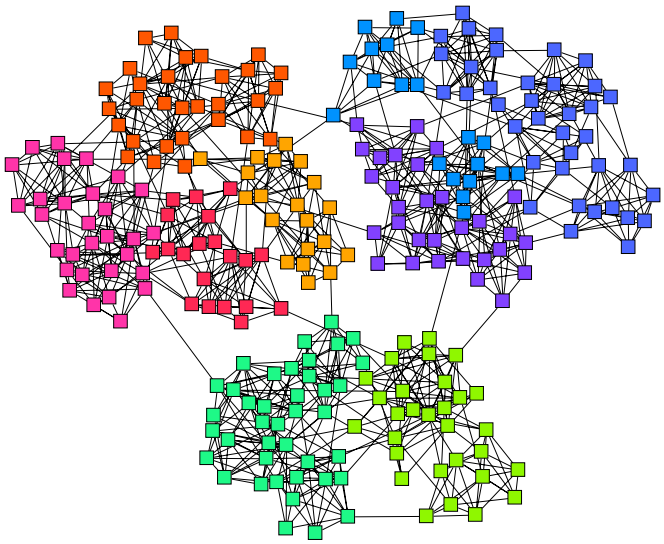
modularity

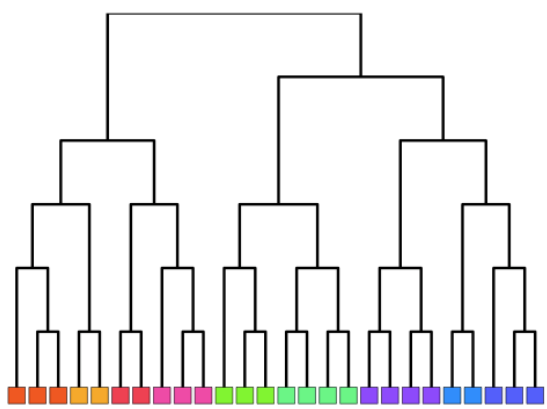




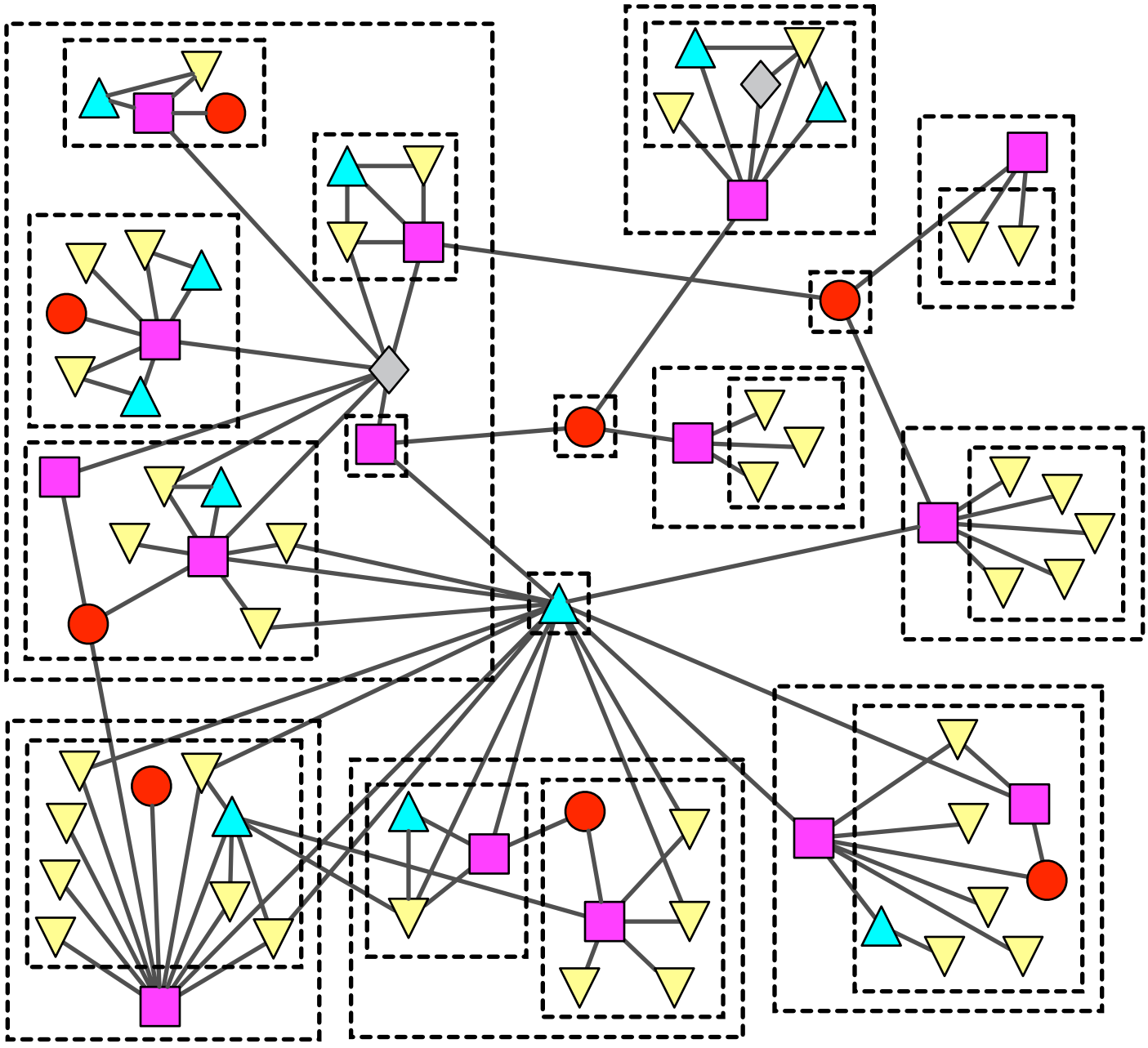








hierarchical modules



selected references

- The structure and function of complex networks. M. E. J. Newman, *SIAM Review* **45**, 167–256 (2003).
- *The Structure and Dynamics of Networks*. M. E. J. Newman, A.-L. Barabási, and D. J. Watts, Princeton University Press (2006).
- Hierarchical structure and the prediction of missing links in networks. A. Clauset, C. Moore, and M. E. J. Newman, *Nature* **453**, 98–101 (2008).
- Modularity and community structure in networks. M. E. J. Newman, *Proc. Natl. Acad. Sci. USA* **103**, 8577–8582 (2006).
- Why social networks are different from other types of networks. M. E. J. Newman and J. Park, *Phys. Rev. E* **68**, 036122 (2003)
- Random graphs with arbitrary degree distributions and their applications. M. E. J. Newman, S. H. Strogatz, and D. J. Watts, *Phys. Rev. E* **64**, 026118 (2001).
- Comparing community structure identification. L. Danon, A. Diaz-Guilera, J. Duch and A. Arenas. *J. Stat. Mech*. P09008 (2005).
- Characterization of Complex Networks: A Survey of measurements. L. daF. Costa, F. A. Rodrigues, G. Travieso and P. R. VillasBoas. arxiv:cond-mat/050585 (2005).
- Evolution in Networks. S.N. Dorogovtsev and J. F. F. Mendes. *Adv. Phys*. **51**, 1079 (2002).
- Revisting "scale-free" networks. E. F. Keller. *BioEssays* **27**, 1060-1068 (2005).
- Currency metabolites and network representations of metabolism. P. Holme and M. Huss. arxiv:0806.2763 (2008).
- Functional cartography of complex metabolic networks. R. Guimera and L. A. N. Amaral. *Nature* **433**, 895 (2005).
- Graphs over Time: Densification Laws, Shrinking Diameters and Possible Explanations. J. Leskovec, J. Kleinberg and C. Faloutsos. *Proc. 11th ACM SIGKDD Intl. Conf. on Knowledge Discovery and Data Mining* 2005.
- The Structure of the Web. J. Kleinberg and S. Lawrence. *Science* 294, 1849 (2001).
- Navigation in a Small World. J. Kleinberg. *Nature* 406 (2000), 845.
- Towards a Theory of Scale-Free Graphs: Definitions, Properties and Implications. L. Li, D. Alderson, J. Doyle, and W. Willinger. *Internet Mathematics 2(4), 2006*.
- A First-Principles Approach to Understanding the Internet's Router-Level Topology. L. Li, D. Alderson, W. Willinger, and J. Doyle. *ACM SIGCOMM* 2004.
- Inferring network mechanisms: The Drosophila melanogaster protein interaction network. M. Middendorf, E. Ziv and C. H. Wiggins. *Proc. Natl. Acad. Sci. USA* 102, 3192 (2005).
- Robustness Can Evolve Gradually in Complex Regulatory Gene Networks with Varying Topology. S. Ciliberti, O. C. Martin and A. Wagner. *PLoS Comp. Bio*. 3, e15 (2007).
- Simple rules yield complex food webs. R. J. Williams and N. D. Martinez. *Nature* 404, 180 (2000).
- A network analysis of committees in the U.S. House of Representatives. M. A. Porter, P. J. Mucha, M. E. J. Newman and C. M. Warmbrand. *Proc. Natl. Acad. Sci. USA* 102, 7057 (2005).
- On the Robustness of Centrality Measures under Conditions of Imperfect Data. S. P. Borgatti, K. M. Carley and D. Krackhardt. *Social Networks* **28**, 124 (2006).