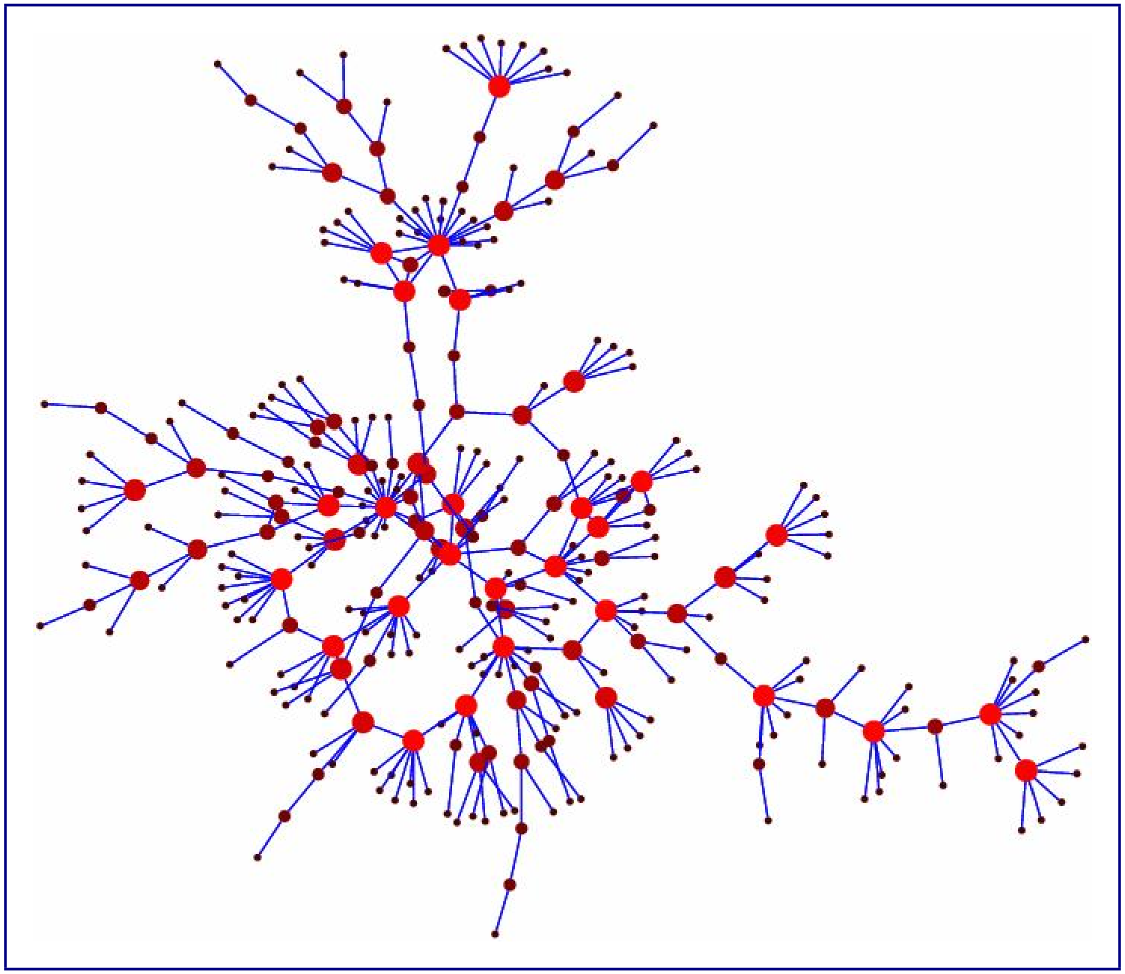
MAE 298, Lecture 12

May 11, 2006



“More network measures and modeling”

Basic network metrics

- Number of vertices and edges
- Average degree
- Degree distribution
- Clustering coefficient
- Spectral gap

More basic measures

From Social Network Analysis:

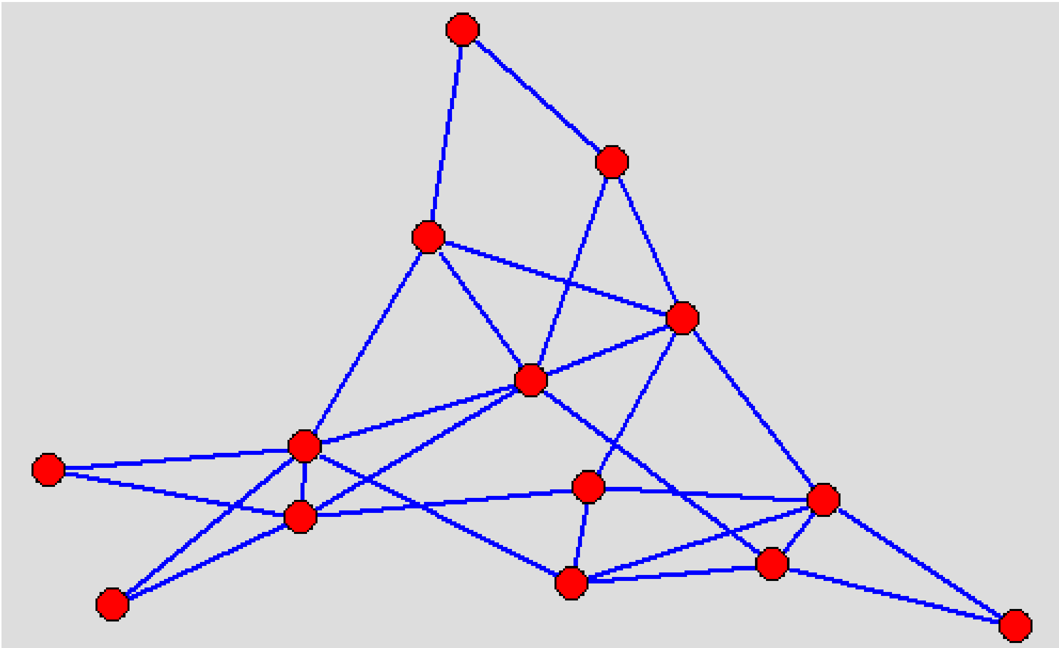
- Betweenness (betweenness centrality)
- Structural Equivalence

From network theory:

- Mixing patterns

Betweenness

[Freeman, L. C. "A set of measures of centrality based on betweenness." *Sociometry* **40** 1977]



A measure of how many shortest paths between all other vertices pass through a given vertex.

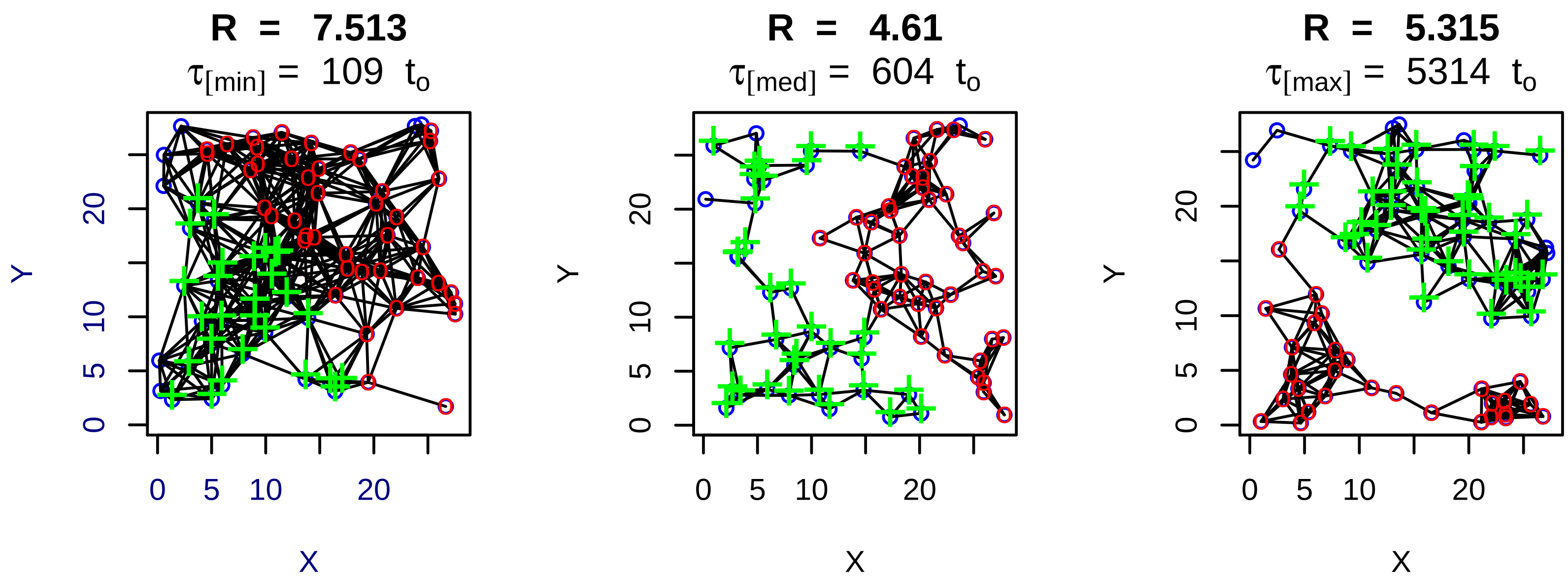
Betweenness (formal definition)

For a given vertex i:

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

- Where σ_{st} is the number of shortest geodesic paths between s and t.
- And $\sigma_{st}(i)$ are the number of those passing through vertex i.

Betweenness and eigenvalues (bottlenecks)

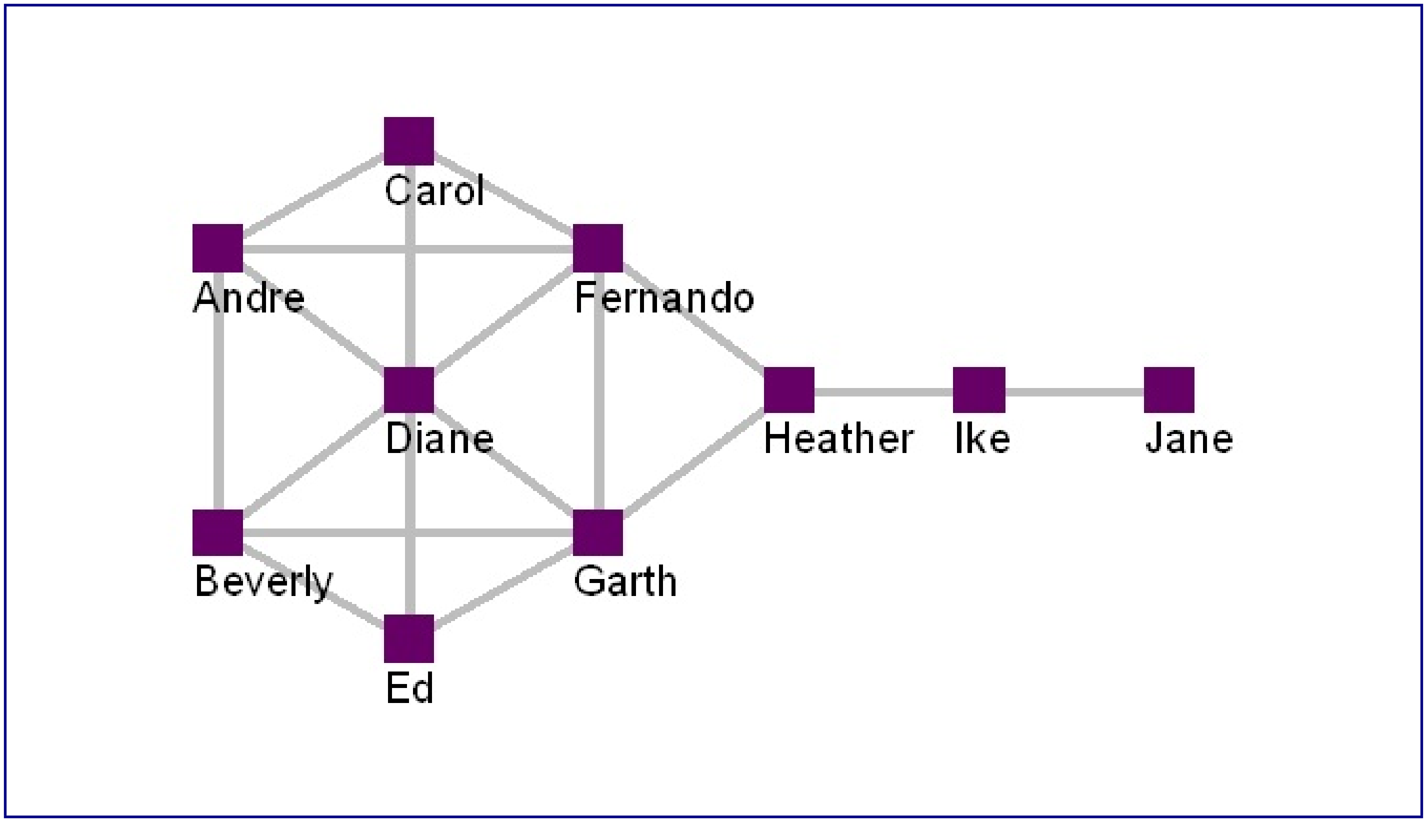


- Bottlenecks have large betweenness values.
- In social networks betweenness is a measure of a nodes "centrality" and importance (could be a proxy for influence).
- In a road network, high betweenness could indicate where alternate routes are needed.
- Also a measure of the resilience of a network (remove high betweenness nodes and destroy connectivity). More on this at end!!

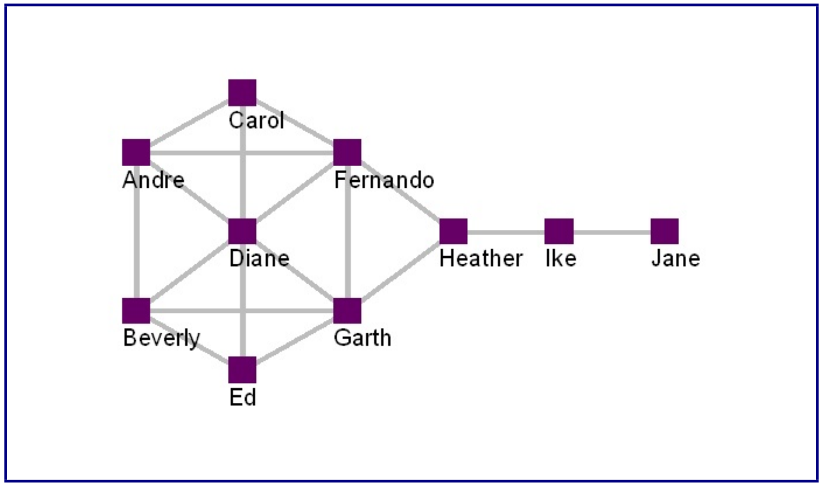
A classic example from Social Network Analysis (SNA)

[http://www.fsu.edu/∼spap/water/network/intro.htm]

The "Kite Network"



The Kite Network



- Degree – Diane looks important (a “hub”).
- Betweenness – Heather looks important (a “connector”/“broker”).
- Closeness – Fernando and Garth can access anyone via a short path.
- Boundary spanners – as Fernando, Garth, and Heather are well-positioned to be “innovators”.
- Peripheral Players – Ike and Jane may be an important resources for fresh information.

Other measures for SNA

- Structural Equivalence- determine which nodes play similar roles in the network
- Cluster Analysis- find cliques and other densely connected clusters
- Structural Holes- find areas of no connection between nodes that could be used for advantage or opportunity
- External/Internal Ratio- find which groups in the network are open or closed to others
- Small Worlds- find node clustering, and short path lengths, that are common in networks exhibiting highly efficient small-world behavior

Structural Equivalence

Narrow definition: Two vertices in a network are structurally equivalent if they have all the same neighbors.

Broader definition: identifying groups of nodes that are similar in their patterns of ties to all other nodes.

How to determine?

Measures of similarity and structural equivalence

- Valued relations
  - Pearson correlations covariances and cross-products
  - Euclidean, Manhattan, and squared distances

- Binary relations
  - Matches: Exact, Jaccard, Hamming

http://www.faculty.ucr.edu/~hanneman/nettext/
C13_%20Structural_Equivalence.html#measure

Some SNA resources

- International Network for Social Network Analysis (http://www.insna.org/)

- InFlow 3.1 - Social Network Mapping Software (http://www.orgnet.com/inflow3.html)
  – Network centrality, cluster analysis, structural equivalence, prestige/influence....

- UCI net (http://www.analytictech.com/ucinet.htm)

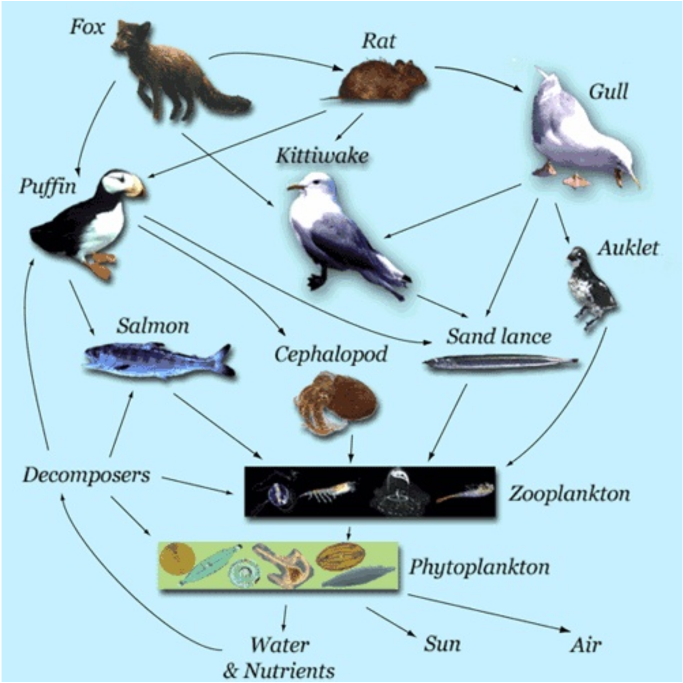
- S. Wasserman and K. Faust, "Social Network Analysis: Methods and Applications", Cambridge University Press, Cambridge, UK, 1994

Mixing

- In almost all networks, nodes of different types (e.g., gender, race, function).
- Does probability of connection between two vertices depend on their types?
- In other words: Mixing by scalar characteristics.

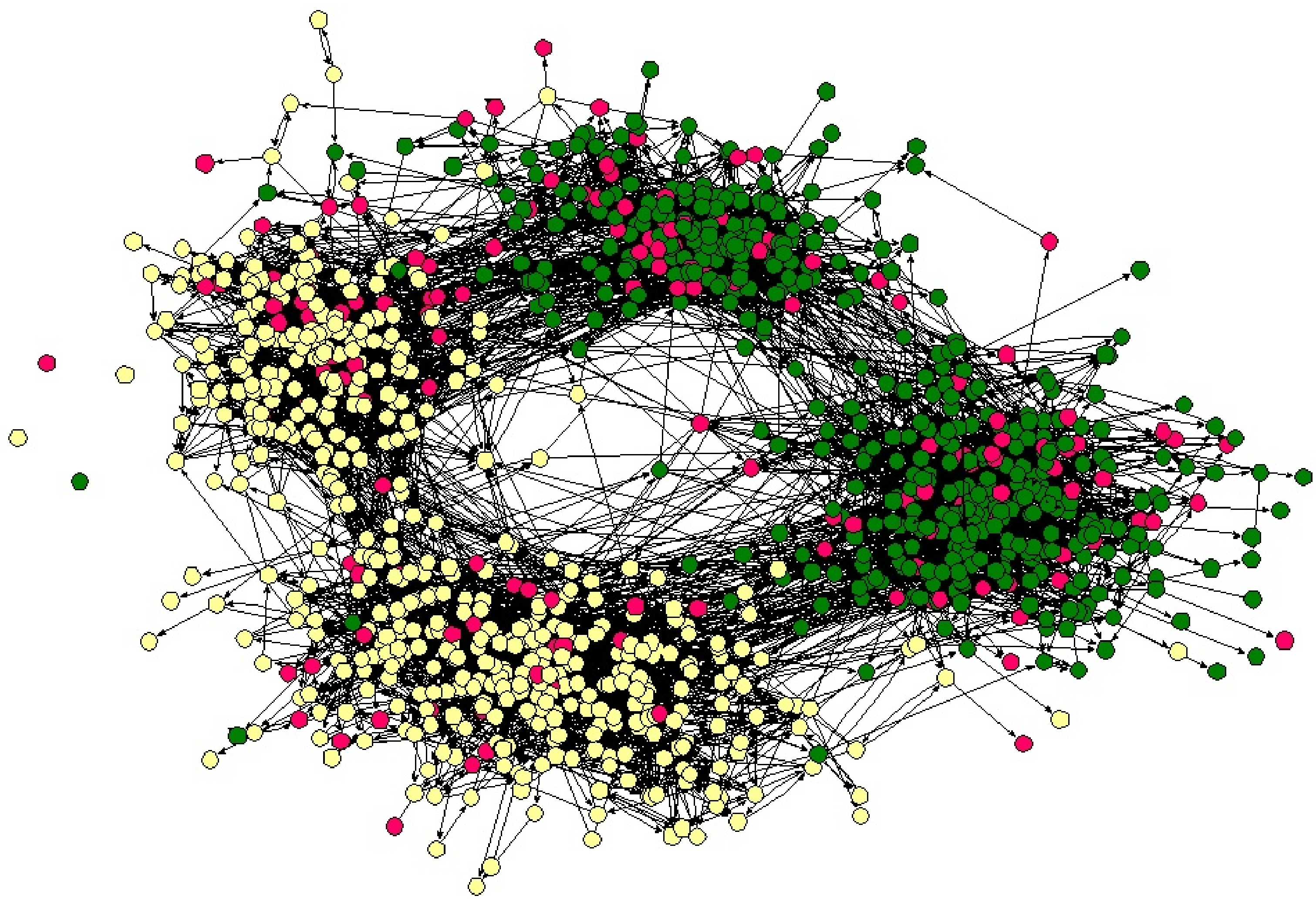
Example: food web

- Types of nodes: plants, herbivores and carnivores.
- Many links between plants and herbivores.
- Many links between herbivores and carnivores.
- Almost no plant-plant or herbivore-herbivore edges.



Assortative mixing

Instead consider a case with many liketype-liketype edges.
Classic example is mixing by race in a social network:



Measure of Assortativity

- Define j different classes/types
- Let E_{ij} be the number of edges connecting types i and j.
- Let $||E||$ be the total number of edges between all classes.
- Define the mixing *matrix* with matrix elements:
  $m_{ij} = E_{ij}/E$

Getting a scalar quantity from m_{ij}

The assortativity coefficient:

$$\boxed{r = (Tr\mathbf{m} - ||\mathbf{m^2}||)/(1 - ||\mathbf{m^2}||)}$$

- $r = 0$ for a randomly mixed network.
- $r = 1$ for a perfectly assortative one.

[MEJ Newman, “Assortative mixing in networks”, Phys Rev Lett. 89(20): 2002]

Degree correlation

What is the scalar quantity of interest is the degree?

Found that social networks are assortative, while technological and biological ones are dissortative.

Why? Just an observation for now.

Table 3.1 *Basic statistics for a number of published networks. The properties measured are as follows: total number of vertices n; total number of edges m; mean degree z; mean vertex–vertex distance ℓ; type of graph, directed or undirected; exponent α of degree distribution if the distribution follows a power law (or "–" if not; in/out-degree exponents are given for directed graphs); clustering coefficient $C^{(1)}$ from* (3.3)*; clustering coefficient $C^{(2)}$ from* (3.6)*; degree correlation coefficient r, section* 3.6*. The last column gives the citation for the network in the bibliography. Blank entries indicate unavailable data.*

	Network	Type	n	m	z	ℓ	α	$C^{(1)}$	$C^{(2)}$	r	Ref(s).
Social	film actors	undirected	449 913	25 516 482	113.43	3.48	2.3	0.20	0.78	0.208	[20, 415]
	company directors	undirected	7 673	55 392	14.44	4.60	–	0.59	0.88	0.276	[105, 322]
	math coauthorship	undirected	253 339	496 489	3.92	7.57	–	0.15	0.34	0.120	[107, 181]
	physics coauthorship	undirected	52 909	245 300	9.27	6.19	–	0.45	0.56	0.363	[310, 312]
	biology coauthorship	undirected	1 520 251	11 803 064	15.53	4.92	–	0.088	0.60	0.127	[310, 312]
	telephone call graph	undirected	47 000 000	80 000 000	3.16		2.1				[8, 9]
	email messages	directed	59 912	86 300	1.44	4.95	1.5/2.0		0.16		[136]
	email address books	directed	16 881	57 029	3.38	5.22	–	0.17	0.13	0.092	[320]
	student relationships	undirected	573	477	1.66	16.01	–	0.005	0.001	−0.029	[45]
	sexual contacts	undirected	2 810				3.2				[264, 265]
Information	WWW `nd.edu`	directed	269 504	1 497 135	5.55	11.27	2.1/2.4	0.11	0.29	−0.067	[14, 34]
	WWW Altavista	directed	203 549 046	2 130 000 000	10.46	16.18	2.1/2.7				[74]
	citation network	directed	783 339	6 716 198	8.57		3.0/–				[350]
	Roget's Thesaurus	directed	1 022	5 103	4.99	4.87	–	0.13	0.15	0.157	[243]
	word co-occurrence	undirected	460 902	17 000 000	70.13		2.7		0.44		[119, 157]
Technological	Internet	undirected	10 697	31 992	5.98	3.31	2.5	0.035	0.39	−0.189	[86, 148]
	power grid	undirected	4 941	6 594	2.67	18.99	–	0.10	0.080	−0.003	[415]
	train routes	undirected	587	19 603	66.79	2.16	–		0.69	−0.033	[365]
	software packages	directed	1 439	1 723	1.20	2.42	1.6/1.4	0.070	0.082	−0.016	[317]
	software classes	directed	1 377	2 213	1.61	1.51	–	0.033	0.012	−0.119	[394]
	electronic circuits	undirected	24 097	53 248	4.34	11.05	3.0	0.010	0.030	−0.154	[155]
	peer-to-peer network	undirected	880	1 296	1.47	4.28	2.1	0.012	0.011	−0.366	[6, 353]
Biological	metabolic network	undirected	765	3 686	9.64	2.56	2.2	0.090	0.67	−0.240	[213]
	protein interactions	undirected	2 115	2 240	2.12	6.80	2.4	0.072	0.071	−0.156	[211]
	marine food web	directed	135	598	4.43	2.05	–	0.16	0.23	−0.263	[203]
	freshwater food web	directed	92	997	10.84	1.90	–	0.40	0.48	−0.326	[271]
	neural network	directed	307	2 359	7.68	3.97	–	0.18	0.28	−0.226	[415, 420]

Detecting Community structure given a network

- “Finding and evaluating community structure in networks”, MEJ Newman, M Girvan - Physical Review E, 2004.
- “Detecting community structure in networks”, MEJ Newman - The European Physical Journal B-Condensed Matter, 2004.
- “Community structure in social and biological networks” M Girvan, MEJ Newman - Proceedings of the National Academy of Sciences, 2002.

Alternately, given a node, can you identify which community it belongs to?

- G. W. Flake, S. R. Lawrence, C. L. Giles, and F. M. Coetzee, “Self-organization and identification of Web communities”, IEEE Computer, 35 (2002).

More basic measures, summary

From Social Network Analysis:

- Betweenness (betweenness centrality)
- Structural Equivalence

From network theory:

- Mixing patterns

“Layered complex networks”

[M. Kurant and P. Thiran, “Layered Complex Networks”, Phys Rev Lett. 89, 2006.]

- Offer a simple formalism to think about two coexisting network topologies.
- The physical topology.
- And the virtual (application) topology.

Example 1:
WWW and IP layer views of the Internet

- Each WWW link virtually connects two IP addresses.
- Those two IP nodes are typically far apart in the underlying IP topology, so the virtual connection is realized as a multihop path along IP routers.
- (Of course the IP network is then mapped onto the physical layer of optical cables and routers.)

Example 2:
Transportation networks

Up until now separate studies of:

1. Physical topology (of roads)

2. Real-life traffic patterns

Want a comprehensive view analyzing them both together.

The formalism

Consider two different networks:

- $G^{\phi} = (V^{\phi}, E^{\phi})$; the physical graph.
- $G^{\lambda} = (V^{\lambda}, E^{\lambda})$; the logical/application-layer graph.

Assume both sets of nodes identical, $V^{\phi} = V^{\lambda}$.

The load on a node

- Load on node i, $l(i)$, is the sum of the weights of all logical edges whose paths traverse i.

- E.g., in a transportation network $l(i)$ is the total amount of traffic that flows through node i.

Application

Study three transportation systems:

1. Mass transit system of Warsaw Poland.

2. Rail network of Switzerland.

3. Rail network of major trains in the EU.

Load

They can estimate the real load from the timetables (some assumptions; decompose into units (one train, one bus, etc), independent of number of people).

Two load estimators:

1. The node degree of the physical network.

2. Betweenness of the physical network.

(Note, these estimators are the ones currently in use in almost all cases: 1) Resilience of networks to edge removal, 2) Modeling cascading failures, etc.....)

Findings

[M. Kurant and P. Thiran, "Layered Complex Networks", Phys Rev Lett. 89, 2006.]

- All three estimators 1) real load, 2) degree, 3) betweenness differ from one-another.
- Using the two-layer view can see the logical graphs may have radically different properties than the physical graphs.
- May lead to reexamination of network robustness (previous studies on Internet, power grid, etc, based on physical layer).