

Introduction to Social Network Analysis and Models

范耀中

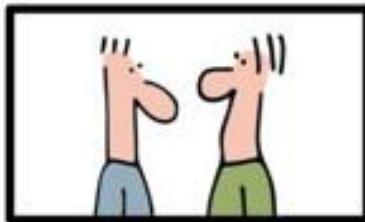
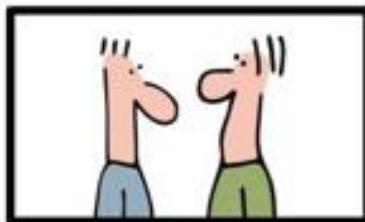
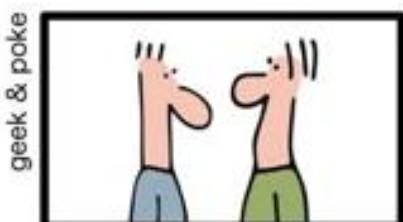
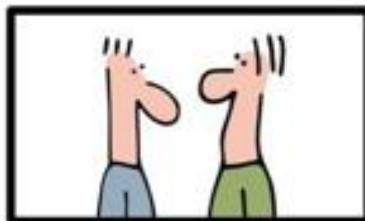
Yao-Chung Fan

yfan@nchu.edu.tw,

Dept. of Computer Science,

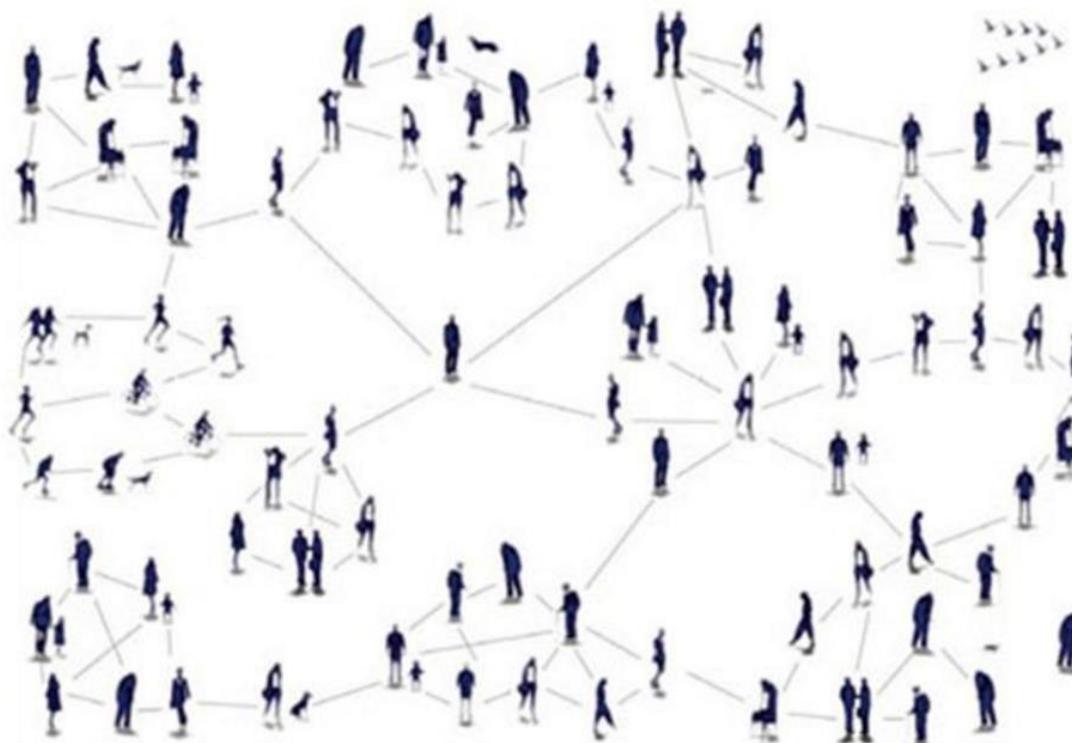
National Chung Hsing University

TODAY: HOW TO SURVIVE EMBARRASSING SITUATIONS



BE CONSISTENT

- **Six Degrees of Separation** is the theory that everyone and everything is six or fewer steps away from any other person in the world, so that the chain of “a friend of a friend” statements can be made to connect any two people in a maximum of six steps.



6.6

每個在這個星球上的人都被另外六個人所隔開。

六度分隔。在我們與星球上的其他人之間。

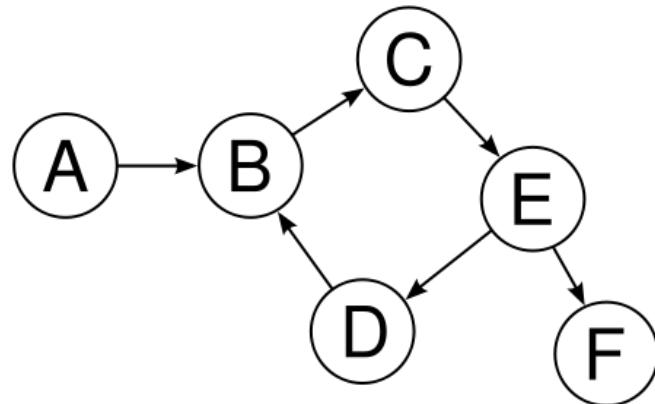
美國總統與威尼斯的貢多拉船手，不只是因為誰有名而已，管你是雨林裡的原住民、火地島的居民或愛斯基摩人都一樣！我與星球上的每位居民都隔著六個人被綁在網上...每個人都像一扇新門，開啟另一個世界

Lecture Outline

- **Part I:**
 - Social Network Analysis
- **Part II:**
 - Social Network Models

Graph/Network

A graph data structure consists of a finite set of ordered pairs, called edges or arcs, of certain entities called nodes or vertices. An edge (x, y) is said to point or go from x to y .



There are thousands of graph applications if not hundreds in computer science

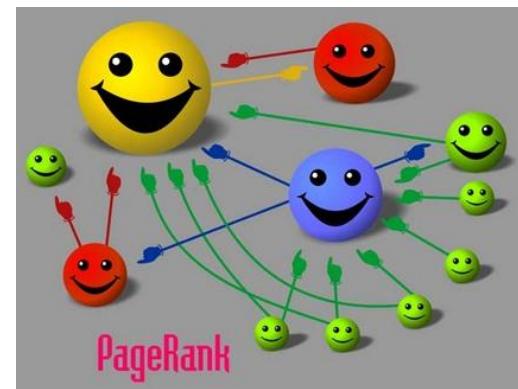
Traditional Graph Problems and Applications

- Traveling Salesman Problem
- Topological Sorting
- Shortest Path Problem
- Hamilton Cycle Problem
- ...



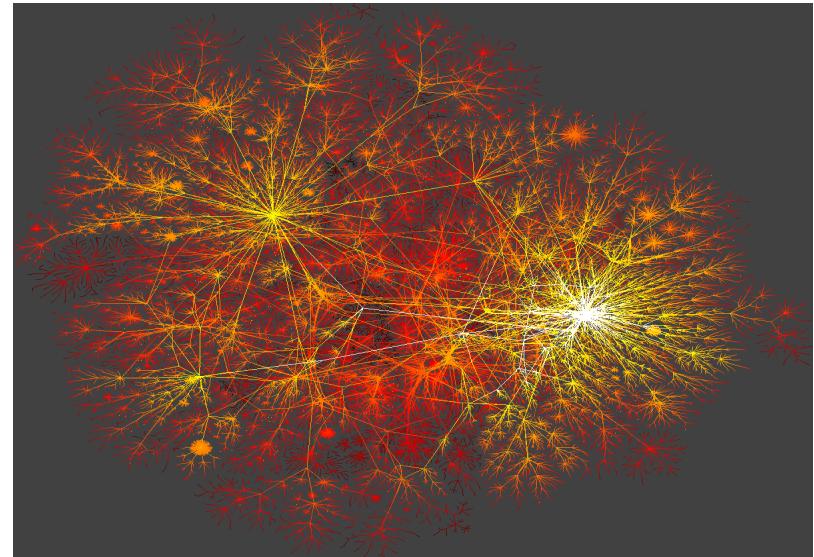
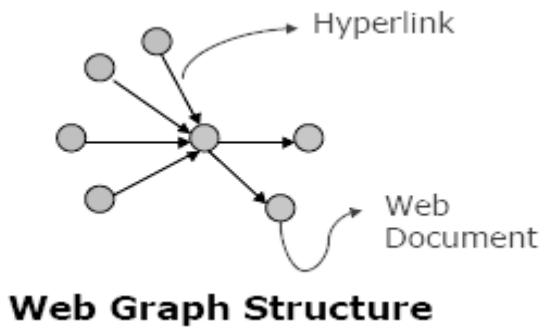
Today's Graph Problems and Applications

- Web Graph for PageRank
- Social Graph for Marketing



What are the differences between traditional and nowadays graph?

Huge, Complex



There are 2 billions vertexes, 15 billions edges, and 7.3 millions pages are added every day in July, 2000.

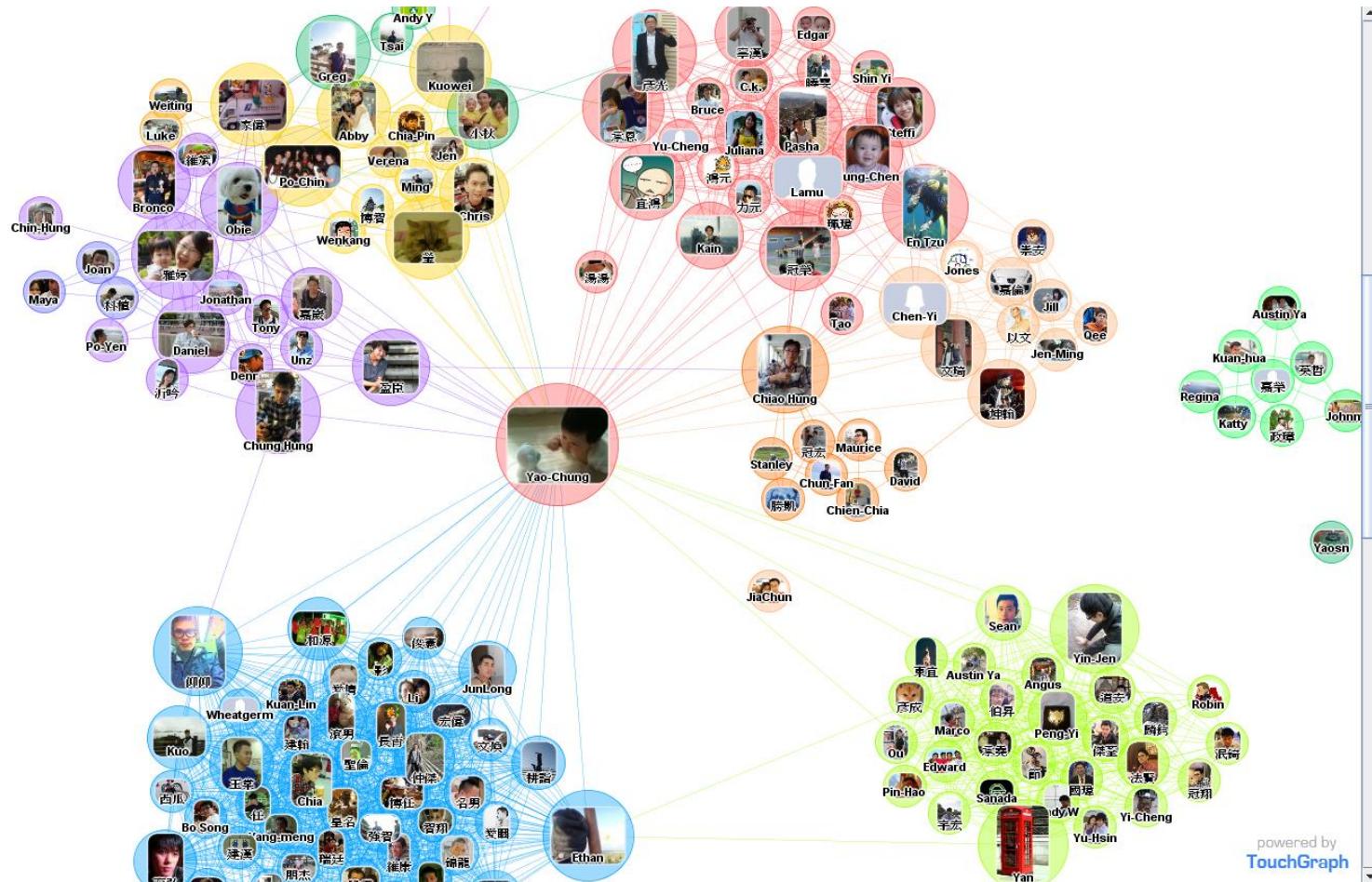
Require about 360G storage space

Social Networks Analysis

- Social network analysis views **social relationships** in terms of **network/graph theory** consisting of *nodes* and *ties*
- Nodes: the individual actors within the networks
- Ties: the relationships between the actors.
 - friendship,
 - kinship,
 - common interest,
 - information exchange,
 - dislike,
 - sexual relationships

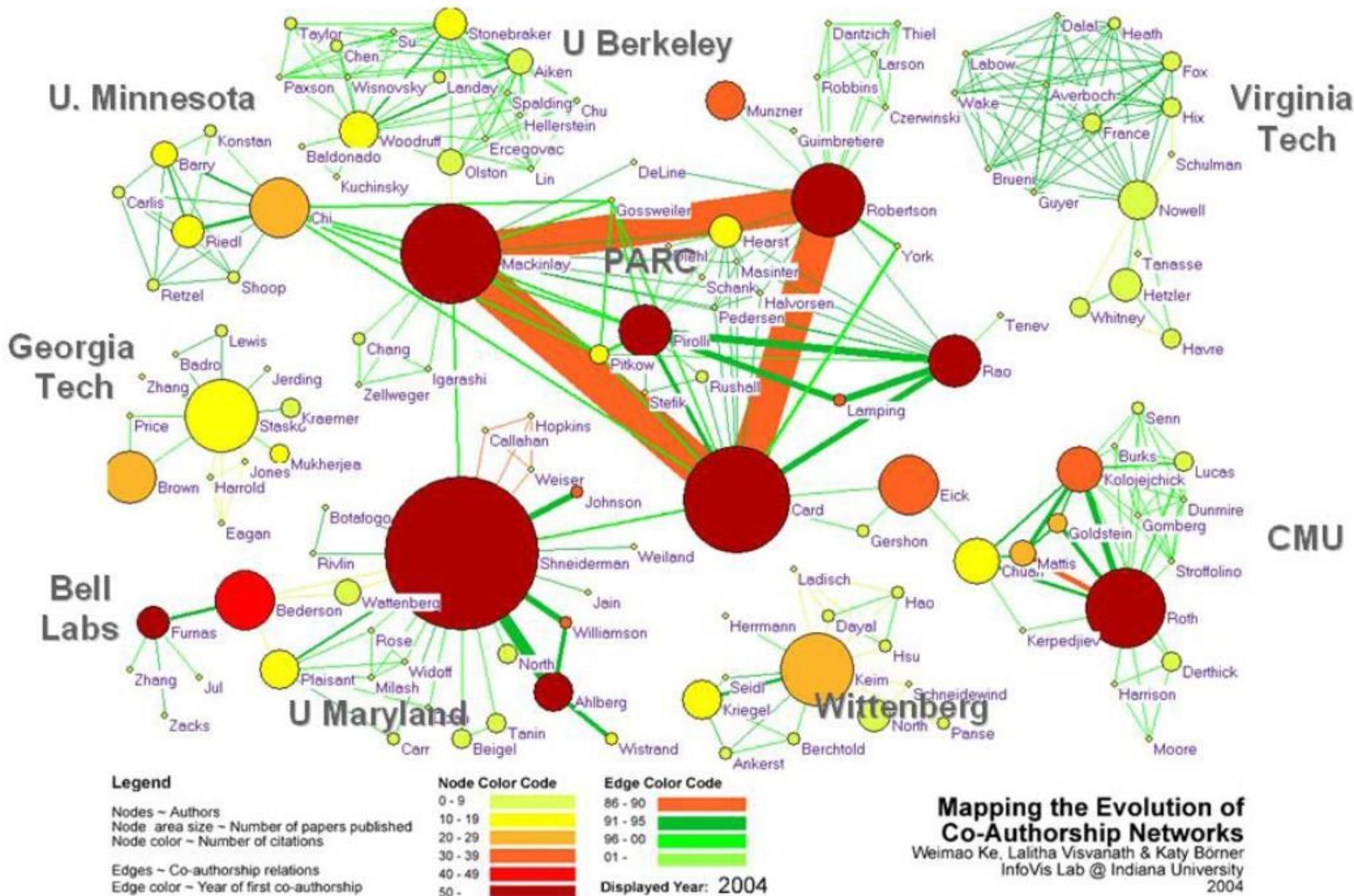
Friendship Network

- My personal friendship network by *touchgraph*:
<http://www.touchgraph.com/facebook#instructions>



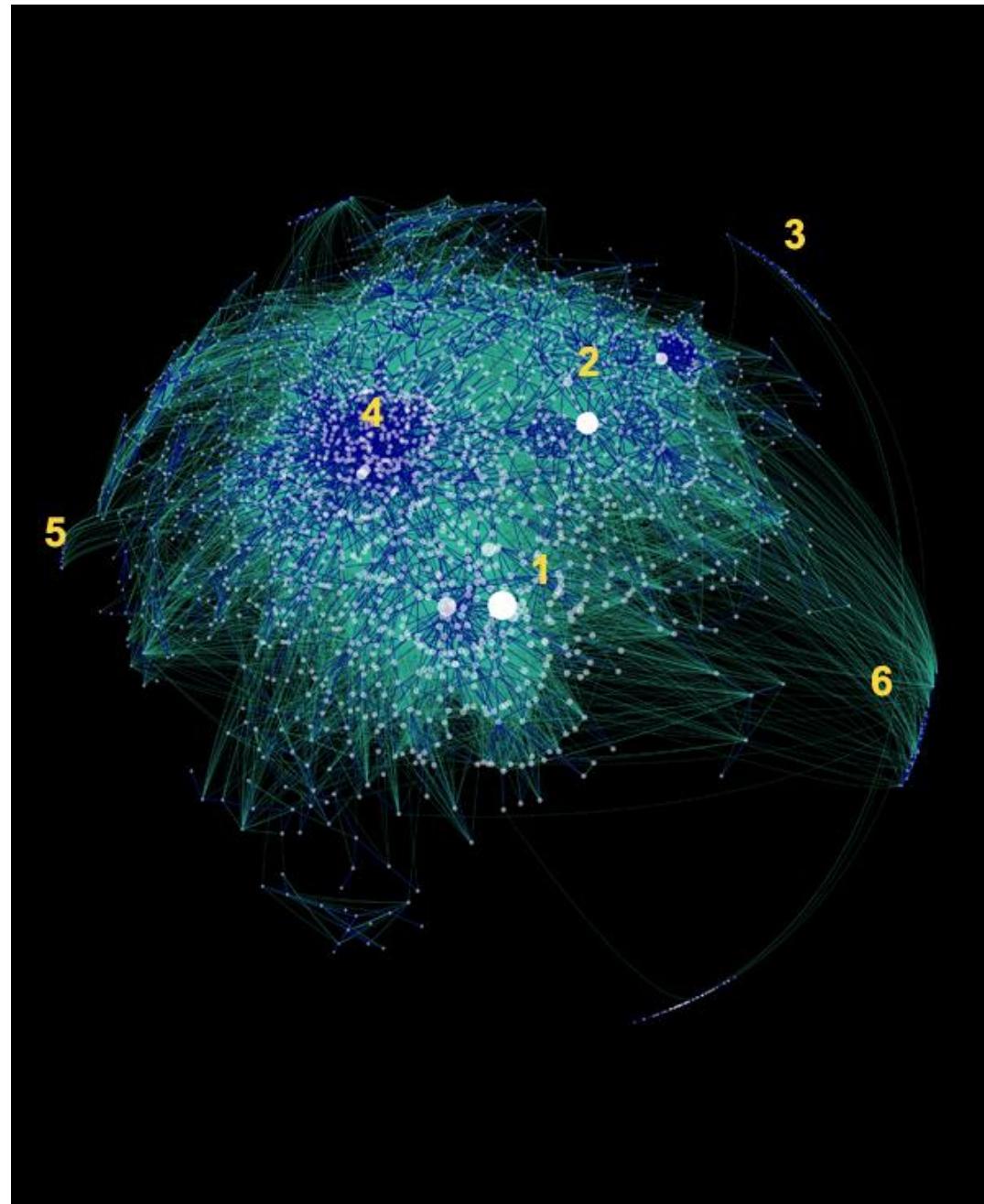
Mapping the Evolution of Co-Authorship Networks

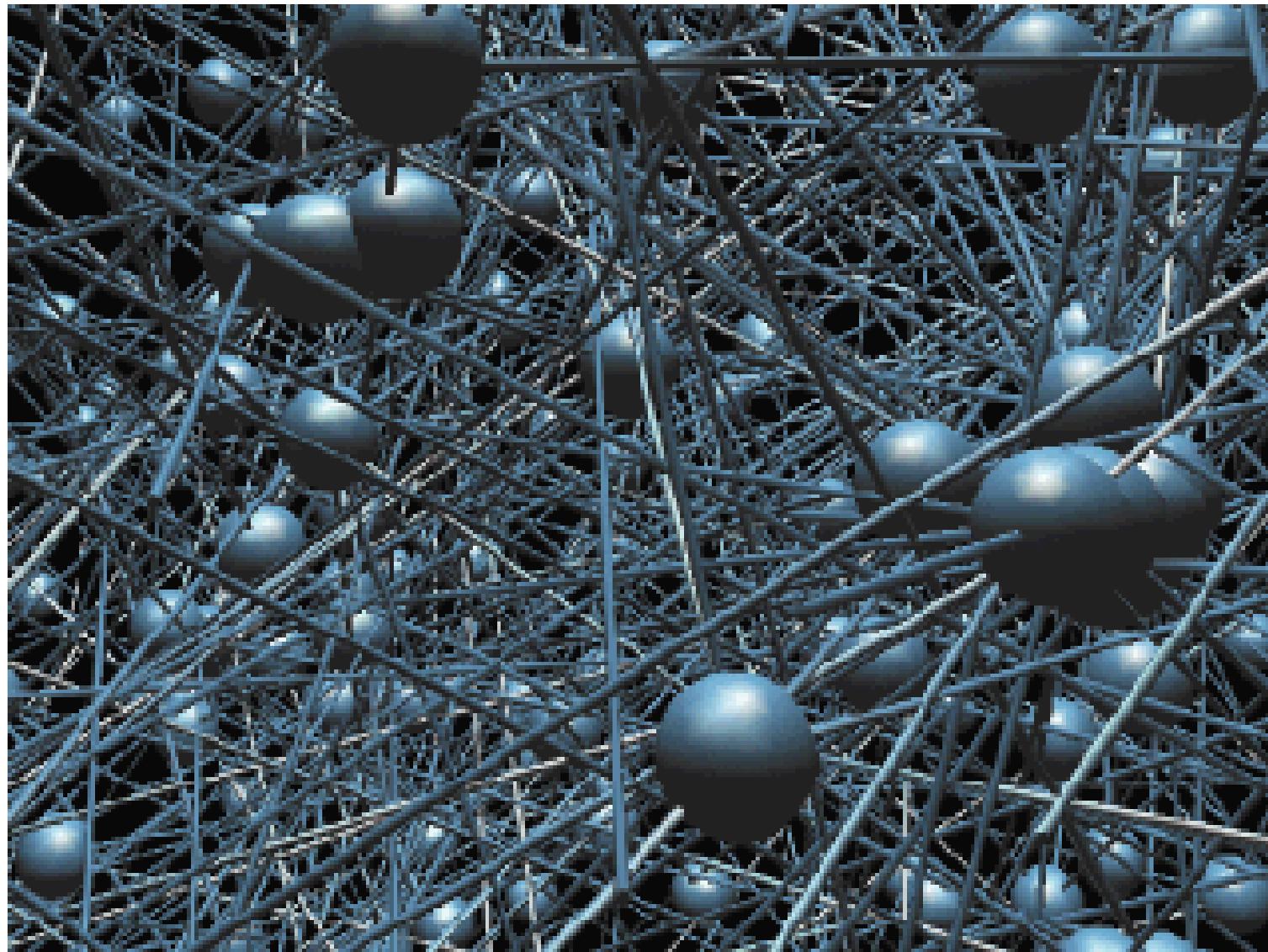
Ke, Viswanath & Börner, (2004) Won 1st price at the IEEE InfoVis Contest.



Blogsphere

- White dots represent individual blogs,
- Sized according to number of links
- Green links represent one-way links
- Blue links indicate reciprocal links

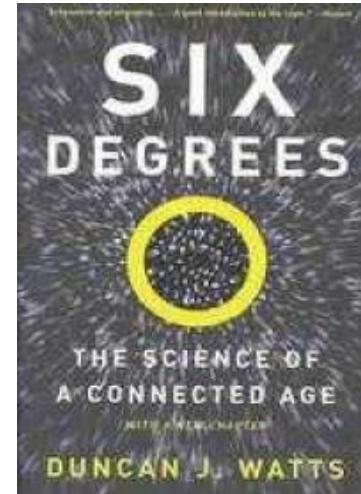
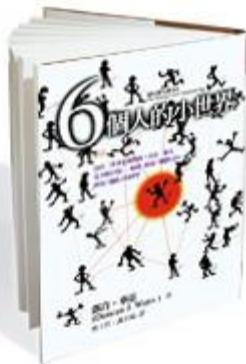




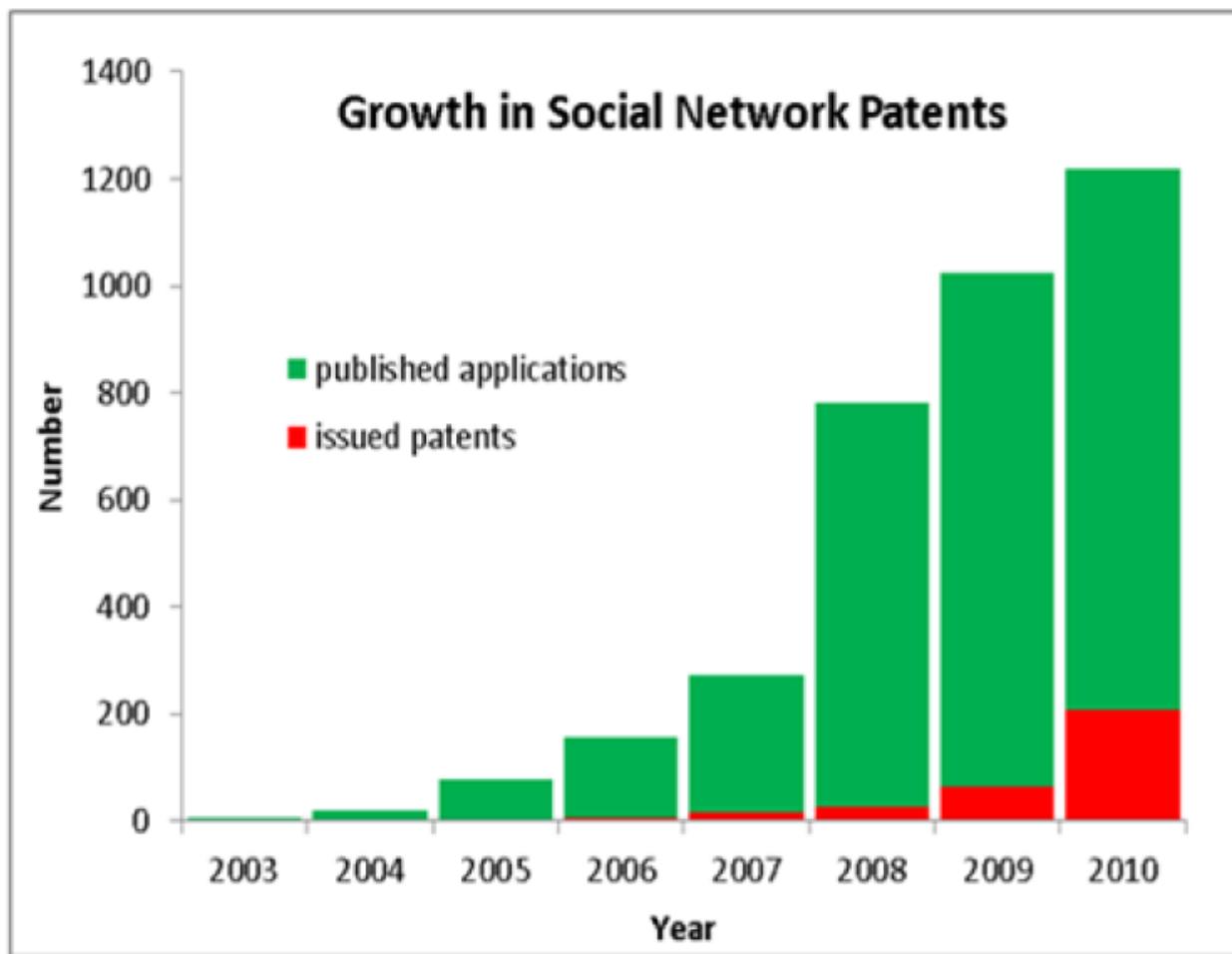
Order ← complex → Chaos

The groundbreaking paper

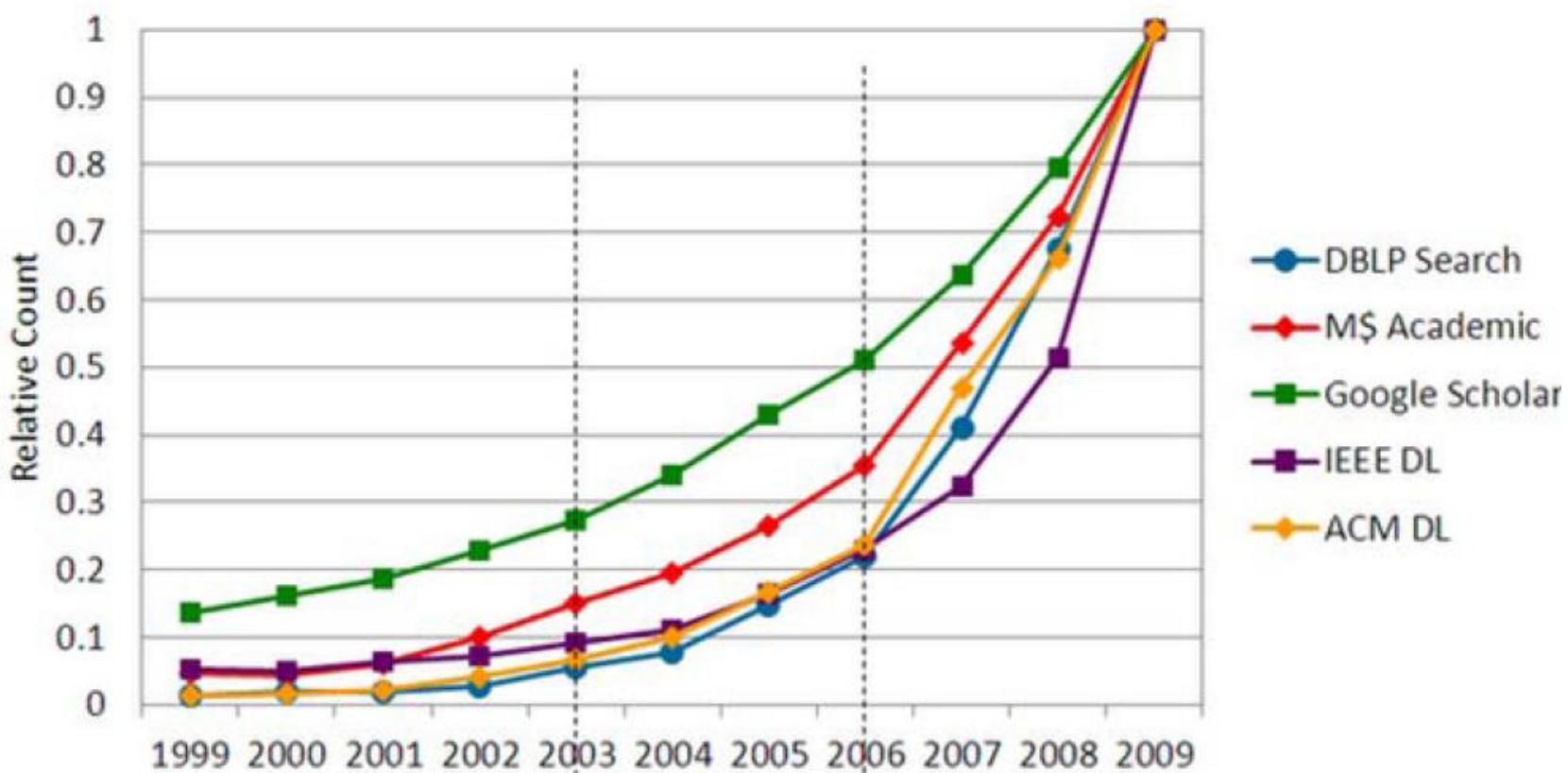
- The groundbreaking paper by Watt and Strogatz,
 - Watt and Strogatz, Collective dynamics of small world networks, Nature 393, 440-442, 1998



Growth in Social Network Patents



Growth in Social Network Papers (Computer Science)



Categories of Networks

- M. E. J. Newman summarizes four categories of networks
 - Information Network
 - Social Network
 - Technological Network
 - Biological Network

[The Structure and Function of Complex Networks - SIAM - Society for ...](https://www.siam.org/doi/abs/10.1137/S003614450342480)

[epubs.siam.org/doi/abs/10.1137/S003614450342480](https://www.siam.org/doi/abs/10.1137/S003614450342480) ▾

由 MEJ Newman 著作 - 2003 - 被引用 16909 次 - 相關文章

Inspired by empirical studies of networked systems such as the Internet, social **networks**, and biological **networks**, researchers have in recent years developed a variety of techniques and models to help us understand or predict the behavior of these systems. Here we review developments in this field, including such ...

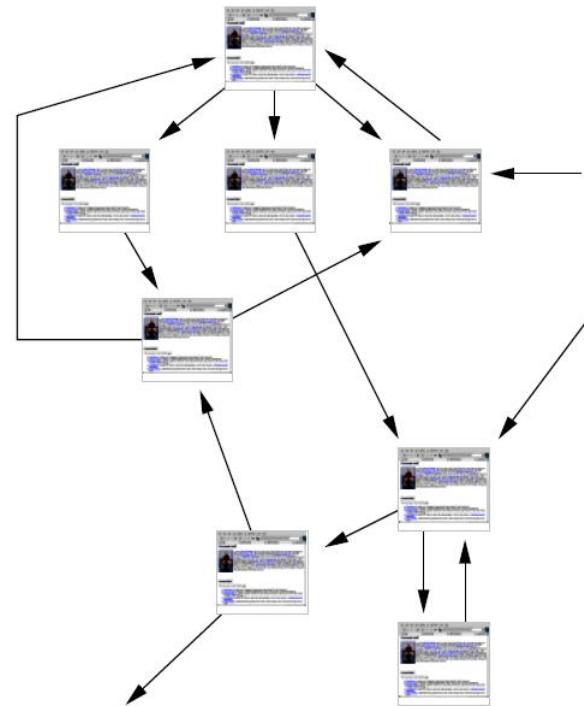
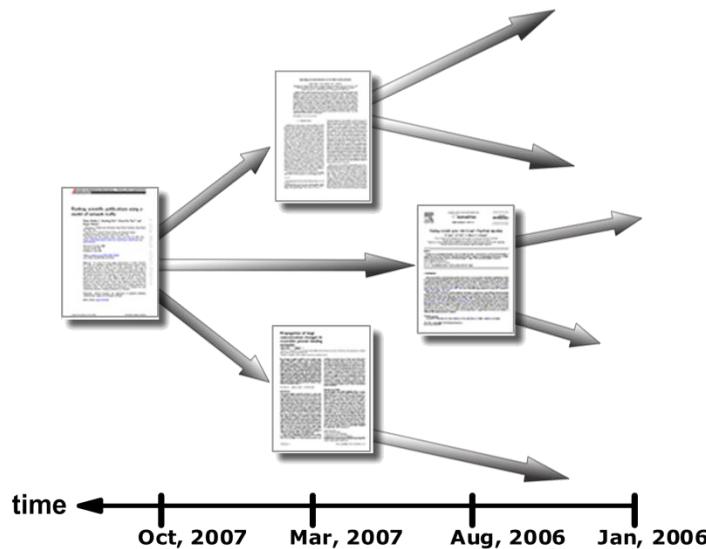
M. E. J. Newman, the structure and function of complex network,
SIAM review 2003, citation number: 13705

	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, 416
	company directors	undirected	7 673	55 392	14.44	4.60	—	0.59	0.88	0.276	105, 323
	math coauthorship	undirected	253 339	496 489	3.92	7.57	—	0.15	0.34	0.120	107, 182
	physics coauthorship	undirected	52 909	245 300	9.27	6.19	—	0.45	0.56	0.363	311, 313
	biology coauthorship	undirected	1 520 251	11 803 064	15.53	4.92	—	0.088	0.60	0.127	311, 313
	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	321
	student relationships	undirected	573	477	1.66	16.01	—	0.005	0.001	—0.029	45
	sexual contacts	undirected	2 810	—	—	—	3.2	—	—	—	265, 266
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/—	—	—	—	351
	Roget's Thesaurus	directed	1 022	5 103	4.99	4.87	—	0.13	0.15	0.157	244
	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	416
	train routes	undirected	587	19 603	66.79	2.16	—	—	0.69	—0.033	366
	software packages	directed	1 439	1 723	1.20	2.42	1.6/1.4	0.070	0.082	—0.016	318
	software classes	directed	1 377	2 213	1.61	1.51	—	0.033	0.012	—0.119	395
	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, 354
biological	metabolic network	undirected	765	3 686	9.64	2.56	2.2	0.090	0.67	—0.240	214
	protein interactions	undirected	2 115	2 240	2.12	6.80	2.4	0.072	0.071	—0.156	212
	marine food web	directed	135	598	4.43	2.05	—	0.16	0.23	—0.263	204
	freshwater food web	directed	92	997	10.84	1.90	—	0.20	0.087	—0.326	272
	neural network	directed	307	2 359	7.68	3.97	—	0.18	0.28	—0.226	416, 421

TABLE II Basic statistics for a number of published networks. The properties measured are: type of graph, directed or undirected; total number of vertices n ; total number of edges m ; mean degree z ; mean vertex–vertex distance ℓ ; 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 Eq. (3); clustering coefficient $C^{(2)}$ from Eq. (6); and degree correlation coefficient r , Sec. III.F. The last column gives the citation(s) for the network in the bibliography. Blank entries indicate unavailable data.

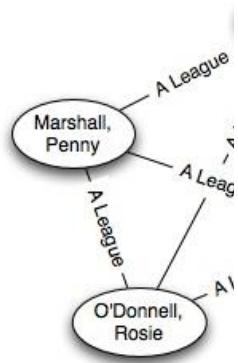
Information Network

- Wikipedia Wordnet, Citation Network, World Wide Web



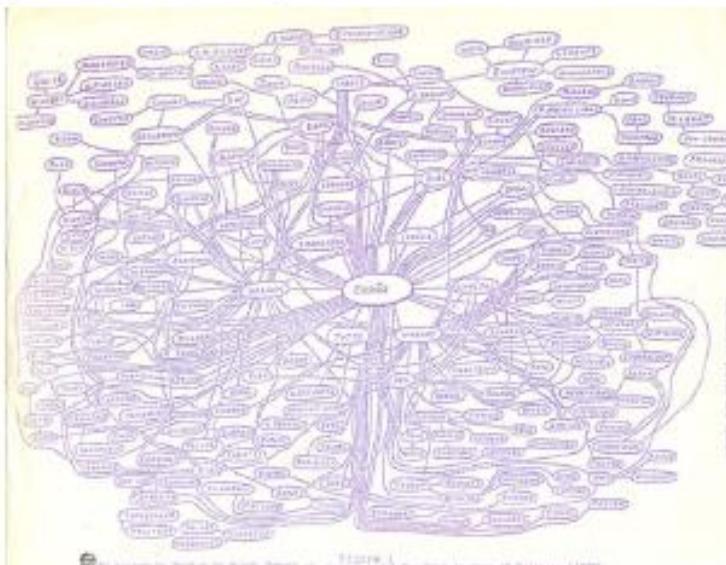
Social Network

- Co-acting Network, Co-authorship Network

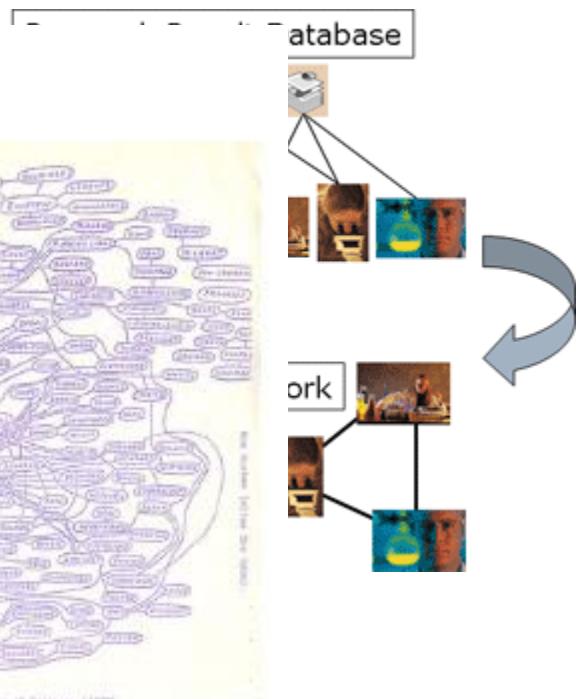


Co-authors of Paul Erdős have number 1,
co-authors of co-authors number 2, etc.

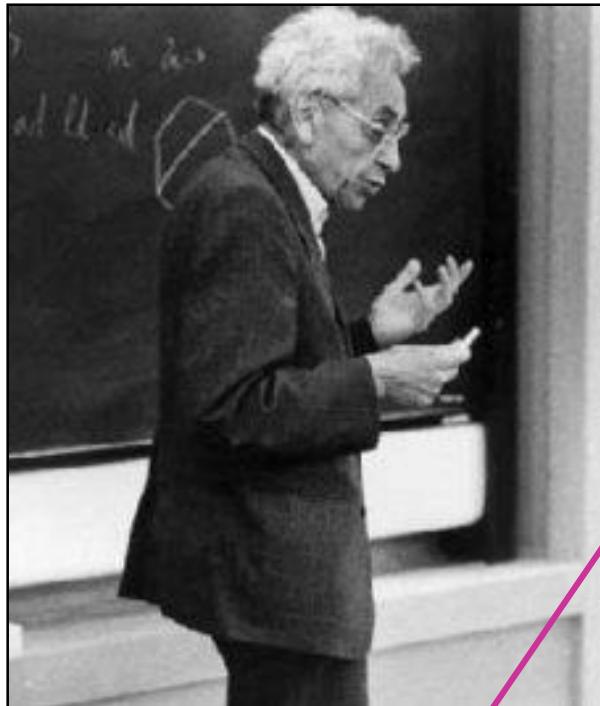
"The Erdős Number Project"



*Mathematicians form a highly clustered
($C = 0.14$) small world ($L = 7.64$)*



Erdős Number (Bacon game for Brainiacs ☺)



Paul Erdős (1913-1996)

Unlike Bacon, Erdos has
better centrality in his network

Number of links required to connect scholars to Erdős, via co-authorship of papers

Erdős wrote 1500+ papers with 507 co-authors.

Jerry Grossman's (Oakland Univ.) website allows mathematicians to compute their Erdos numbers:

<http://www.oakland.edu/enp/>

Connecting path lengths, among mathematicians only:

- average is 4.65
- maximum is 13

MR Collaboration Distance Results - Microsoft Internet Explorer

File Edit View Favorites Tools Help

Back Forward Stop Home Search Favorites Print Find Copy Paste Send

Address http://www.ams.org/msnmain/cgd/index.html?preferred_language=en&group_source=351252&AuthorTargetName=turing%2C+alan&group_target=& Go Links

Google erdos number Search 912 blocked ABC Check Look for Map Options

MathSciNet Mathematical Reviews on the Web

Support Mail | Help | FAQ

MR Collaboration Distance = 7

Subbarao Kambhampati	coauthored with	Dana S. Nau	MR1391058 (97d:68214)
Dana S. Nau	coauthored with	Vipin Kumar	MR0760321 (86e:68094)
Vipin Kumar	coauthored with	Ahmed H. Sameh	MR1639069 (99e:65165)
Ahmed H. Sameh	coauthored with	Gene Howard Golub	MR1836514 (2002c:65045)
Gene Howard Golub	coauthored with	James Hardy Wilkinson	MR0212984 (35 #3849)
James Hardy Wilkinson	coauthored with	M. H. A. Newman	MR0028103 (10,405f)
M. H. A. Newman	coauthored with	Alan Mathison Turing	MR0006332 (3,290b)

[Change First Author](#) [Change Second Author](#) [New Search](#)

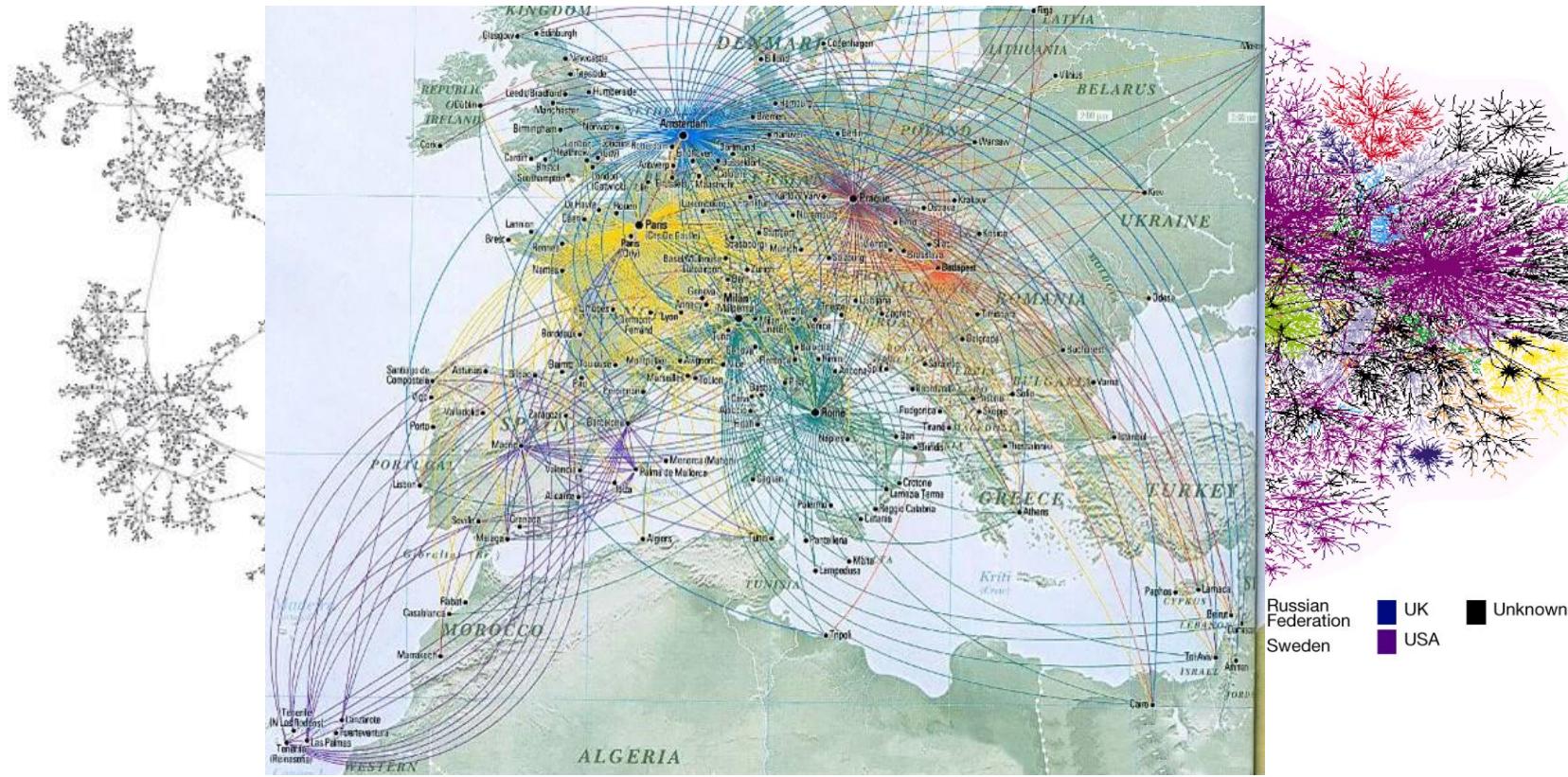
<http://www.ams.org/mathscinet-getitem?mr=28103> Internet

[Change First Author](#) [Change Second Author](#) [New Search](#)

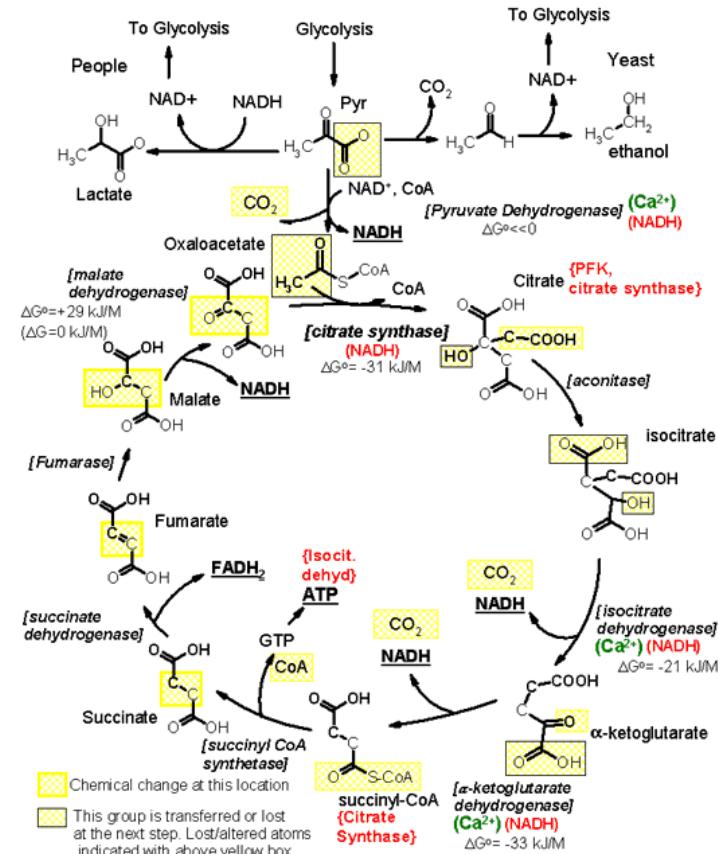
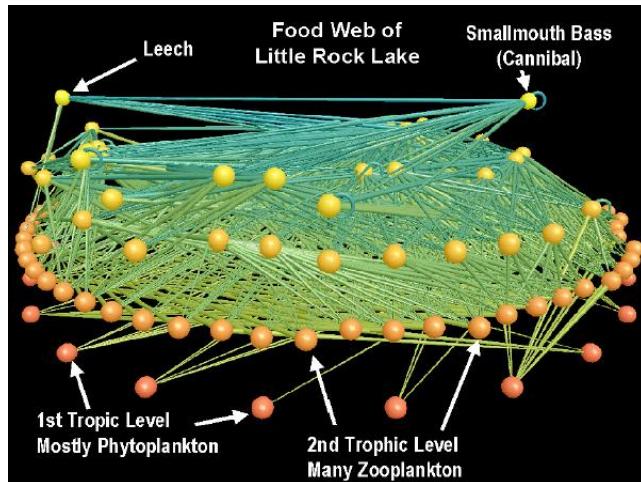
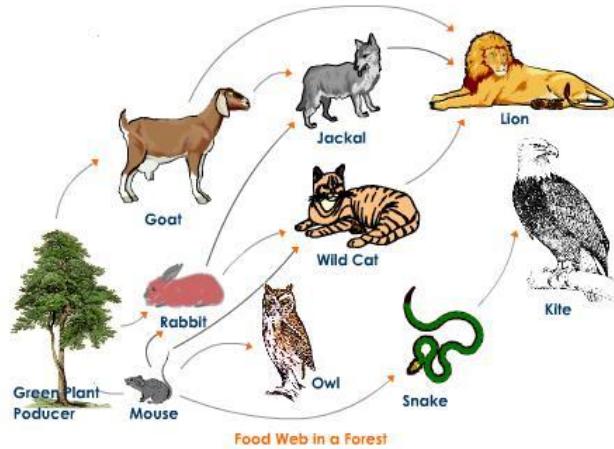
Internet

Technological Network

- Power Grid, Internet Router, Air Flight



Biological Network



Key: (Feedback/product inhibition) (Activation) (Inhibition)

	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, 416
	company directors	undirected	7 673	55 392	14.44	4.60	—	0.59	0.88	0.276	105, 323
	math coauthorship	undirected	253 339	496 489	3.92	7.57	—	0.15	0.34	0.120	107, 182
	physics coauthorship	undirected	52 909	245 300	9.27	6.19	—	0.45	0.56	0.363	311, 313
	biology coauthorship	undirected	1 520 251	11 803 064	15.53	4.92	—	0.088	0.60	0.127	311, 313
	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	321
	student relationships	undirected	573	477	1.66	16.01	—	0.005	0.001	—0.029	45
	sexual contacts	undirected	2 810	—	—	—	3.2	—	—	—	265, 266
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/—	—	—	—	351
	Roget's Thesaurus	directed	1 022	5 103	4.99	4.87	—	0.13	0.15	0.157	244
	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	416
	train routes	undirected	587	19 603	66.79	2.16	—	—	0.69	—0.033	366
	software packages	directed	1 439	1 723	1.20	2.42	1.6/1.4	0.070	0.082	—0.016	318
	software classes	directed	1 377	2 213	1.61	1.51	—	0.033	0.012	—0.119	395
	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, 354
biological	metabolic network	undirected	765	3 686	9.64	2.56	2.2	0.090	0.67	—0.240	214
	protein interactions	undirected	2 115	2 240	2.12	6.80	2.4	0.072	0.071	—0.156	212
	marine food web	directed	135	598	4.43	2.05	—	0.16	0.23	—0.263	204
	freshwater food web	directed	92	997	10.84	1.90	—	0.20	0.087	—0.326	272
	neural network	directed	307	2 359	7.68	3.97	—	0.18	0.28	—0.226	416, 421

TABLE II Basic statistics for a number of published networks. The properties measured are: type of graph, directed or undirected; total number of vertices n ; total number of edges m ; mean degree z ; mean vertex–vertex distance ℓ ; 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 Eq. (3); clustering coefficient $C^{(2)}$ from Eq. (6); and degree correlation coefficient r , Sec. III.F. The last column gives the citation(s) for the network in the bibliography. Blank entries indicate unavailable data.

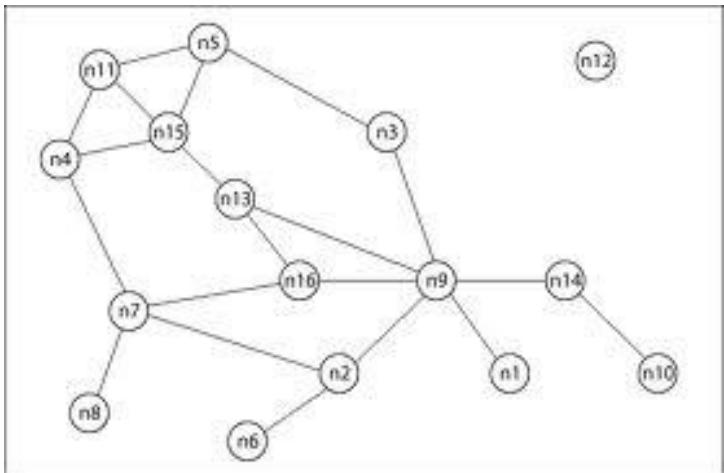
Network Properties

- Average Distance between Pairs
- Transitivity
- Degree Distribution
- Network Resilience
- Clustering Structure
 - Mixing Patterns
 - Degree Correlations

Basic Network Measures

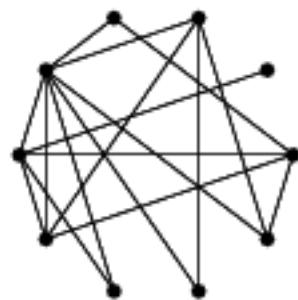
- d_{ij} = shortest path between node i and node j
- Diameter = Max d_{ij}
- Averaged shortest path =

$$\frac{1}{\frac{1}{2}n(n + 1)} \sum_{i \geq j} d_{ij}$$



Diameter

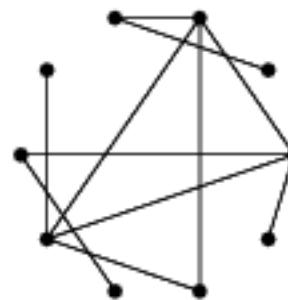
Diameter = Max d_{ij}



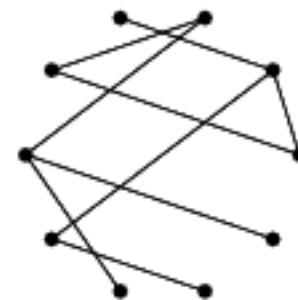
3



4

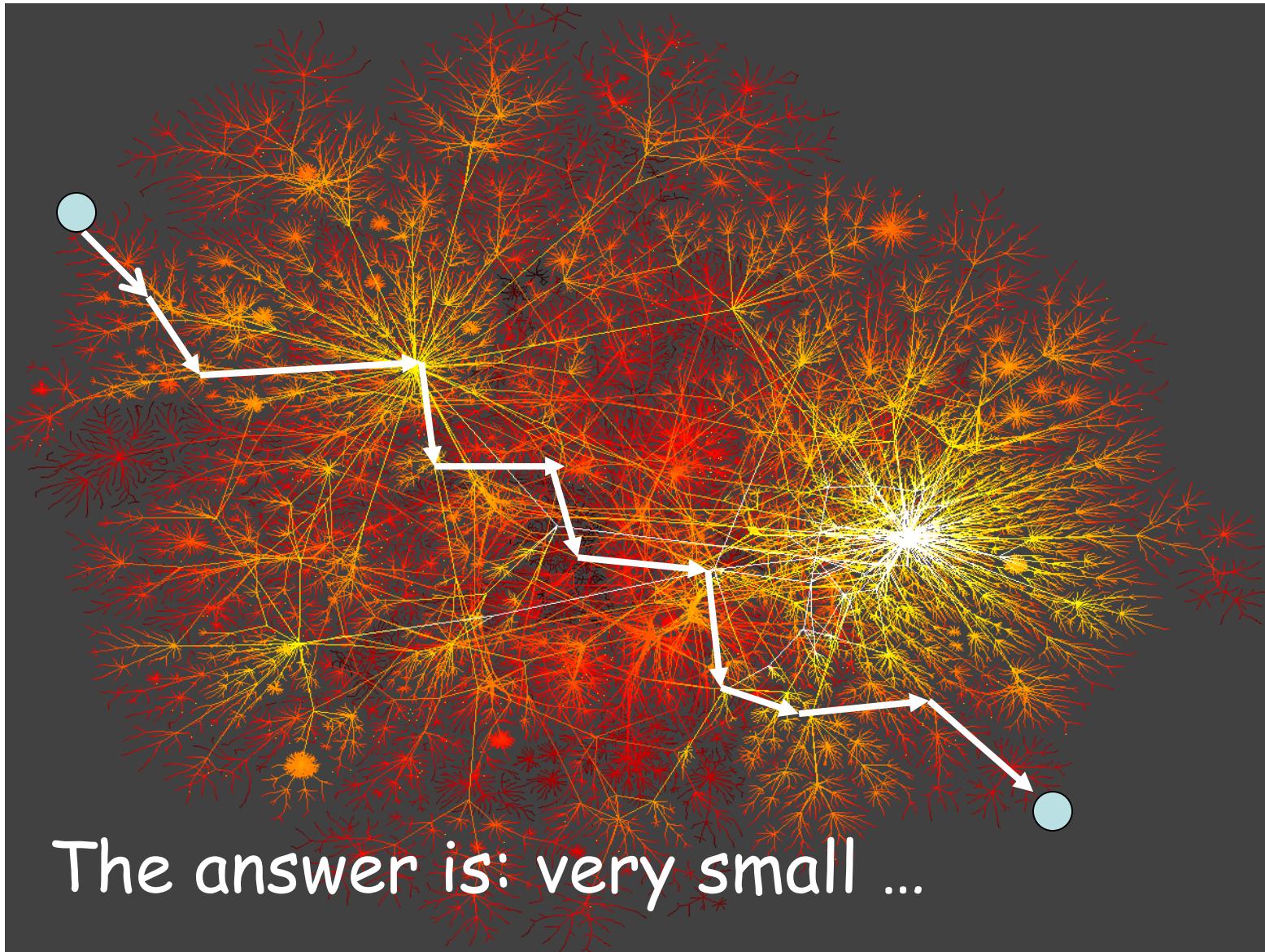


5



7

How far apart are things?



	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, 416
	company directors	undirected	7 673	55 392	14.44	4.60	—	0.59	0.88	0.276	105, 323
	math coauthorship	undirected	253 339	496 489	3.92	7.57	—	0.15	0.34	0.120	107, 182
	physics coauthorship	undirected	52 909	245 300	9.27	6.19	—	0.45	0.56	0.363	311, 313
	biology coauthorship	undirected	1 520 251	11 803 064	15.53	4.92	—	0.088	0.60	0.127	311, 313
	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	321
	student relationships	undirected	573	477	1.66	16.01	—	0.005	0.001	—0.029	45
	sexual contacts	undirected	2 810	—	—	—	3.2	—	—	—	265, 266
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/—	—	—	—	351
	Roget's Thesaurus	directed	1 022	5 103	4.99	4.87	—	0.13	0.15	0.157	244
	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	416
	train routes	undirected	587	19 603	66.79	2.16	—	—	0.69	—0.033	366
	software packages	directed	1 439	1 723	1.20	2.42	1.6/1.4	0.070	0.082	—0.016	318
	software classes	directed	1 377	2 213	1.61	1.51	—	0.033	0.012	—0.119	395
	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, 354
biological	metabolic network	undirected	765	3 686	9.64	2.56	2.2	0.090	0.67	—0.240	214
	protein interactions	undirected	2 115	2 240	2.12	6.80	2.4	0.072	0.071	—0.156	212
	marine food web	directed	135	598	4.43	2.05	—	0.16	0.23	—0.263	204
	freshwater food web	directed	92	997	10.84	1.90	—	0.20	0.087	—0.326	272
	neural network	directed	307	2 359	7.68	3.97	—	0.18	0.28	—0.226	416, 421

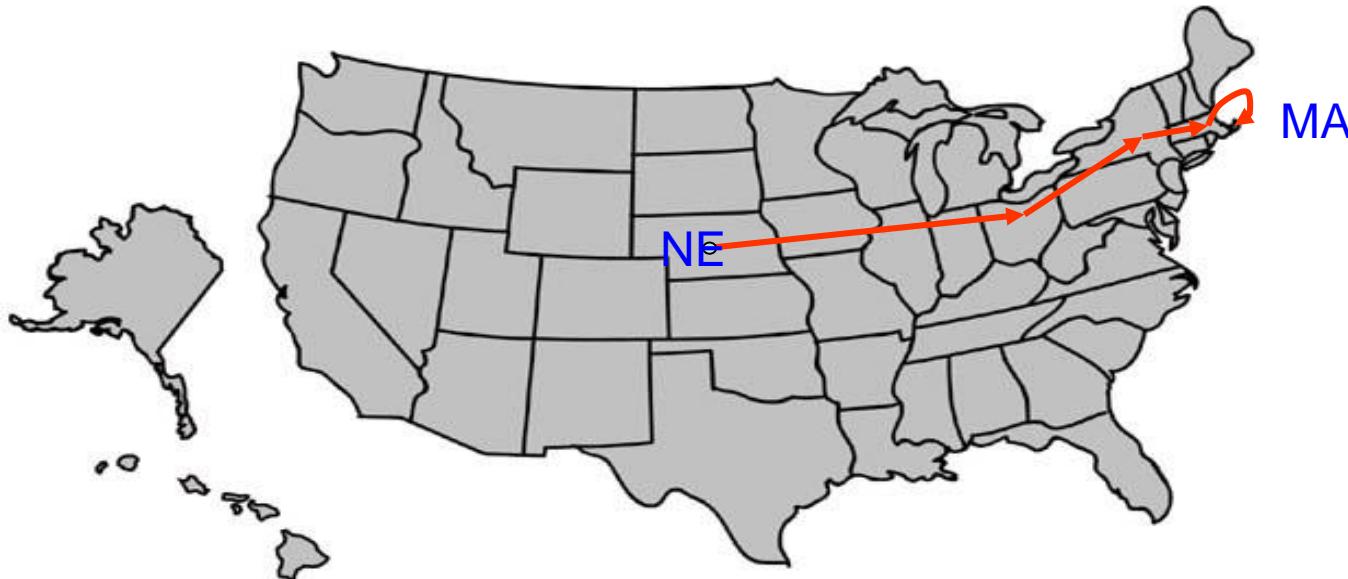
TABLE II Basic statistics for a number of published networks. The properties measured are: type of graph, directed or undirected; total number of vertices n ; total number of edges m ; mean degree z ; mean vertex–vertex distance ℓ ; exponent α of degree distribution if the distribution follows a power law (or “—” if not; in/out-degree exponents are given for directed graphs); clustering coefficient $C^{(1)}$ from Eq. (3); clustering coefficient $C^{(2)}$ from Eq. (6); and degree correlation coefficient r , Sec. III.F. The last column gives the citation(s) for the network in the bibliography. Blank entries indicate unavailable data.

Small World Phenomenon

The Milgram's experiment (1967)



- Random People in Omaha were asked to send letters to a stockbroker in Boston
- Letters can only be passed to acquaintances
- Only 64 out of 296 letters successfully delivered
- The average step is six



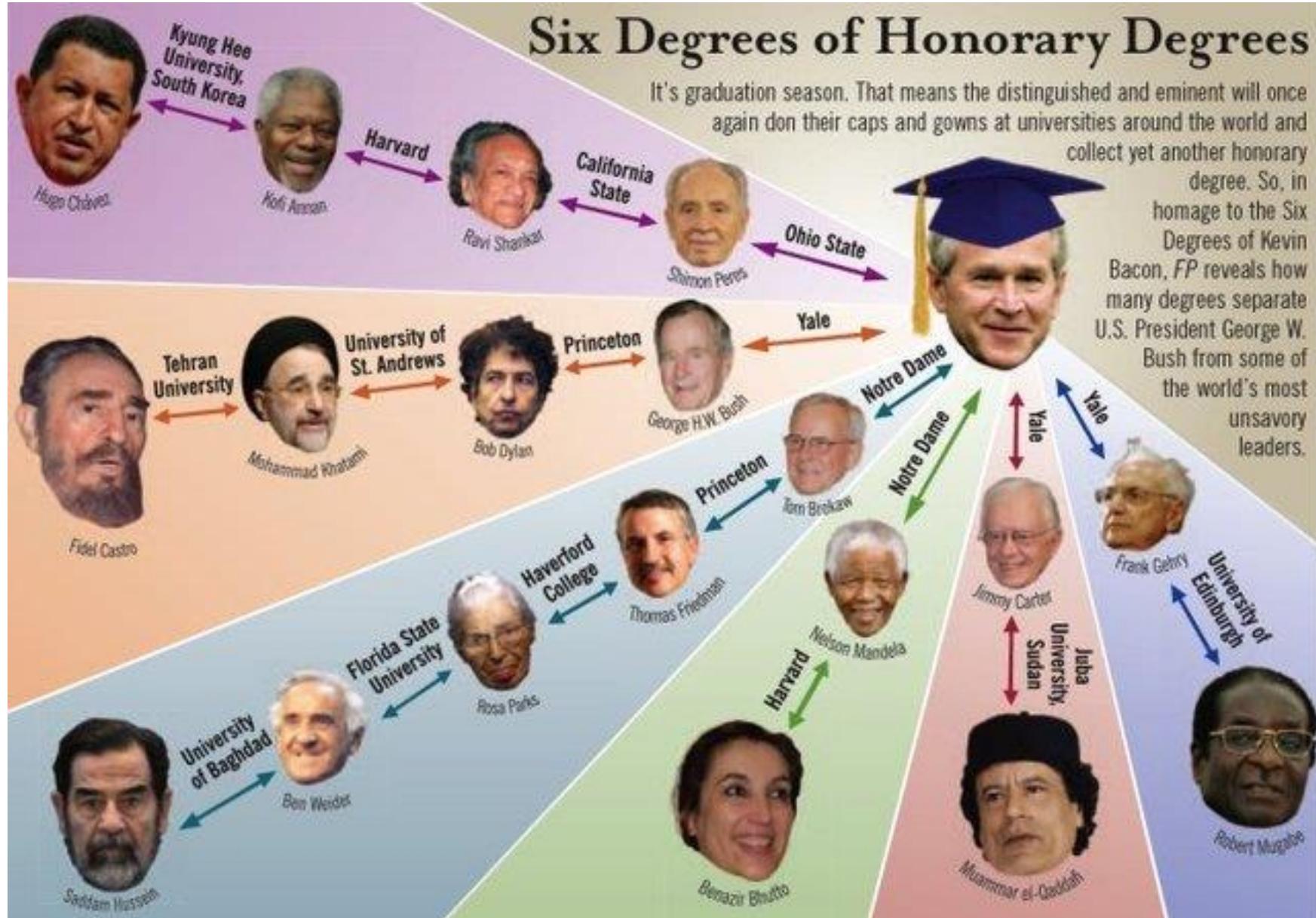
Small World Phenomenon

- 「我從某處得知，在地球上，人與人之間只被六個人隔絕。六度的分隔，正是這個星球的人際距離。」 — 舞台劇《六度分離》中歐莎之語
- *I read somewhere that everybody on this planet is separated by only six other people. Six degrees of separation between us and everyone else on this planet. The President of the United States, a gondolier in Venice, just fill in the names. Ouisa, Six Degree of Separation (Play)*
- *I find that extremely comforting, that we're so close, but I also find it like Chinese water torture that we're so close because you have to find the right six people to make the connection.*
- *It's not just big names -- it's anyone. A native in a rain forest, a Tierra del Fuegan, an Eskimo. I am bound -- you are bound -- to everyone on this planet by a trail of six people. It's a profound thought -- how Paul found us, how to find the man whose son he claims to be, or perhaps is, although I doubt it. How everyone is a new door, opening into other worlds.* -- John Guare



Six Degrees of Honorary Degrees

It's graduation season. That means the distinguished and eminent will once again don their caps and gowns at universities around the world and collect yet another honorary degree. So, in homage to the Six Degrees of Kevin Bacon, *FP* reveals how many degrees separate U.S. President George W. Bush from some of the world's most unsavory leaders.



The Kevin Bacon Game

An Example

Kevin Bacon

Mystic River (2003)

Tim Robbins

Code 46 (2003)

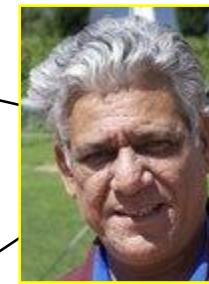
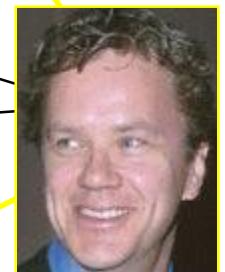
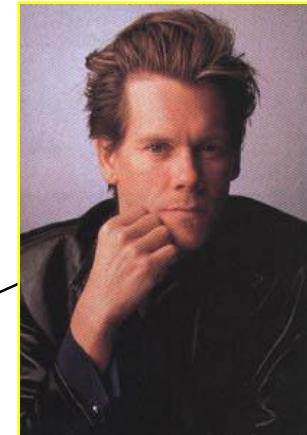
Om Puri

Yuva (2004)

Rani Mukherjee

Black (2005)

Amitabh Bachchan



Small World Experiments by Facebook

- 你與 的距離？

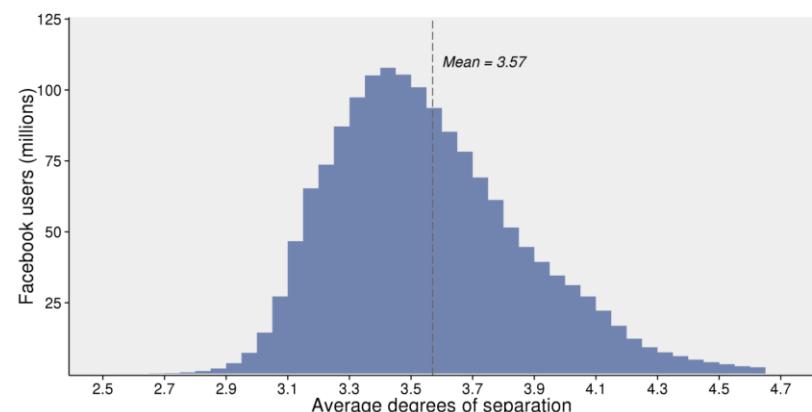


Mark Zuckerberg
3.17 degrees of separation



Sheryl Sandberg
2.92 degrees of separation

The majority of the people on Facebook have averages between 2.9 and 4.2 degrees of separation. Figure 1 (below) shows the distribution of averages for each person.



https://research.fb.com/three-and-a-half-degrees-of-separation/?hc_location=ufi

Network Properties

- Average Distance between Pairs (Small)
- **Transitivity**
- Degree Distribution
- Network Resilience
- Mixing Patterns
- Degree Correlations
- Clustering Structure

Transitivity

- Clustering Coefficient
- If vertex A is connected to vertex B and vertex B to vertex C, then likely vertex A will also be connected to vertex C.



Transitivity

$$C = \frac{1}{n} \sum_i C_i. \quad C_i = \frac{\text{number of triangles connected to vertex } i}{\text{number of triples centered on vertex } i}.$$

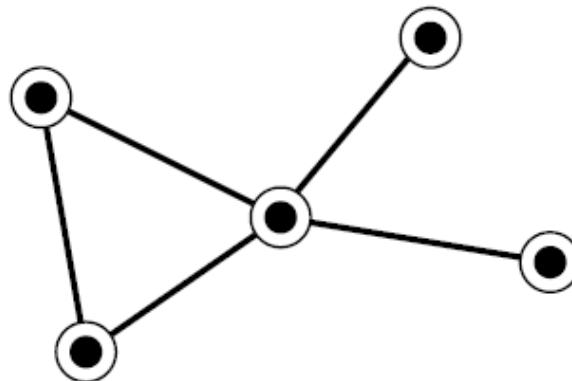


FIG. 5 Illustration of the definition of the clustering coefficient C , Eq. (3). This network has one triangle and eight connected triples, and therefore has a clustering coefficient of $3 \times 1/8 = \frac{3}{8}$. The individual vertices have local clustering coefficients, Eq. (5), of 1, 1, $\frac{1}{6}$, 0 and 0, for a mean value, Eq. (6), of $C = \frac{13}{30}$.

	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, 416
	company directors	undirected	7 673	55 392	14.44	4.60	—	0.59	0.88	0.276	105, 323
	math coauthorship	undirected	253 339	496 489	3.92	7.57	—	0.15	0.34	0.120	107, 182
	physics coauthorship	undirected	52 909	245 300	9.27	6.19	—	0.45	0.56	0.363	311, 313
	biology coauthorship	undirected	1 520 251	11 803 064	15.53	4.92	—	0.088	0.60	0.127	311, 313
	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	321
	student relationships	undirected	573	477	1.66	16.01	—	0.005	0.001	—0.029	45
	sexual contacts	undirected	2 810	—	—	—	3.2	—	—	—	265, 266
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/—	—	—	—	351
	Roget's Thesaurus	directed	1 022	5 103	4.99	4.87	—	0.13	0.15	0.157	244
	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	416
	train routes	undirected	587	19 603	66.79	2.16	—	—	0.69	—0.033	366
	software packages	directed	1 439	1 723	1.20	2.42	1.6/1.4	0.070	0.082	—0.016	318
	software classes	directed	1 377	2 213	1.61	1.51	—	0.033	0.012	—0.119	395
	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, 354
biological	metabolic network	undirected	765	3 686	9.64	2.56	2.2	0.090	0.67	—0.240	214
	protein interactions	undirected	2 115	2 240	2.12	6.80	2.4	0.072	0.071	—0.156	212
	marine food web	directed	135	598	4.43	2.05	—	0.16	0.23	—0.263	204
	freshwater food web	directed	92	997	10.84	1.90	—	0.20	0.087	—0.326	272
	neural network	directed	307	2 359	7.68	3.97	—	0.18	0.28	—0.226	416, 421

TABLE II Basic statistics for a number of published networks. The properties measured are: type of graph, directed or undirected; total number of vertices n ; total number of edges m ; mean degree z ; mean vertex–vertex distance ℓ ; 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 Eq. (3); clustering coefficient $C^{(2)}$ from Eq. (6); and degree correlation coefficient r , Sec. III.F. The last column gives the citation(s) for the network in the bibliography. Blank entries indicate unavailable data.

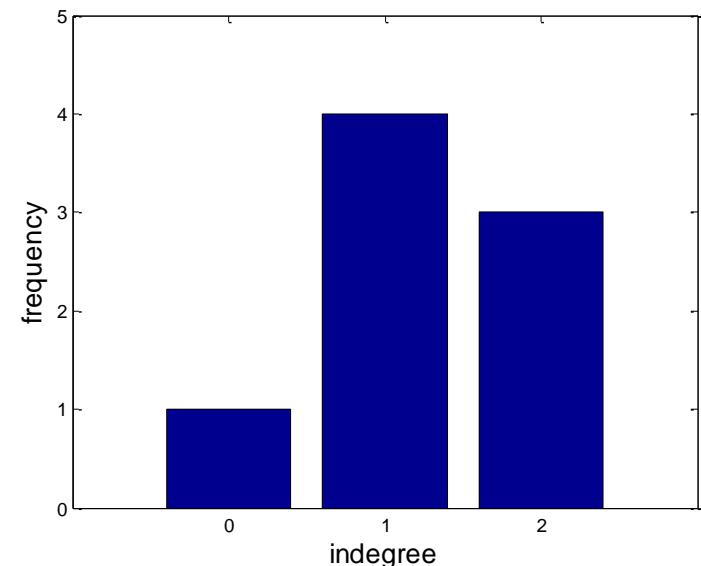
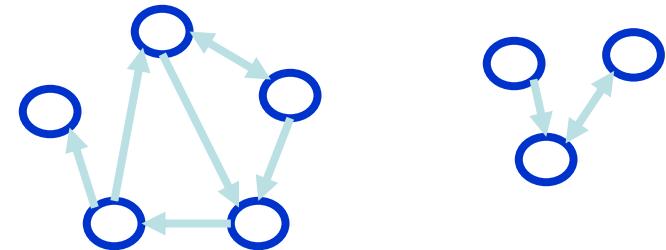
$$C = \frac{3 \times \text{number of triangles in the network}}{\text{number of connected triples of vertices}},$$

Network Properties

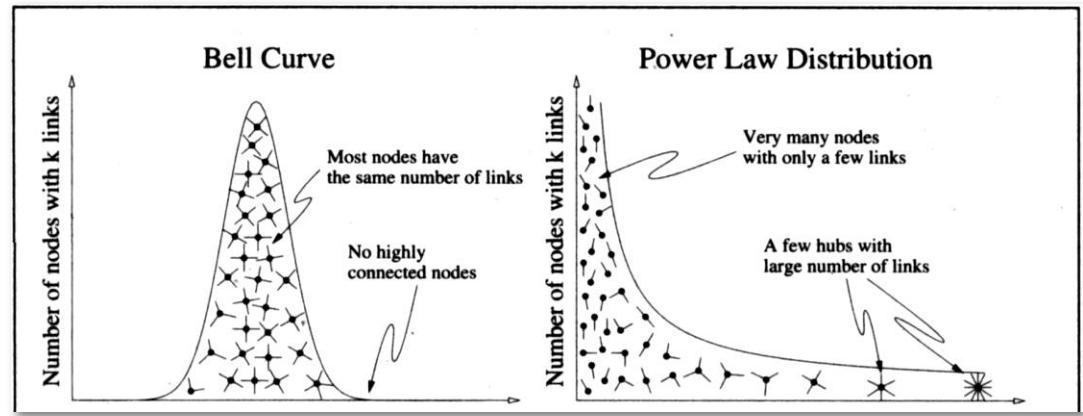
- Average Distance between pairs (Small)
- Transitivity (High)
- **Degree Distribution**
- Network Resilience
- Mixing Patterns
- Degree Correlations
- Clustering Structure

Degree Distribution

- Degree sequence: An ordered list of the (in,out) degree of each node
 - In-degree sequence:
 - [2, 2, 2, 1, 1, 1, 1, 0]
 - Out-degree sequence:
 - [2, 2, 2, 2, 1, 1, 1, 0]
 - (undirected) degree sequence:
 - [3, 3, 3, 2, 2, 1, 1, 1]
- Degree distribution: A frequency count of the occurrence of each degree
 - In-degree distribution:
 - [(2,3) (1,4) (0,1)]
 - Out-degree distribution:
 - [(2,4) (1,3) (0,1)]
 - (undirected) distribution:
 - [(3,3) (2,2) (1,3)]

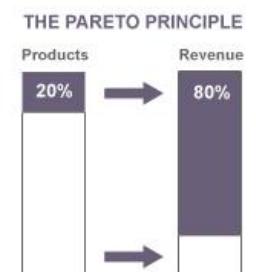


Degree Distribution



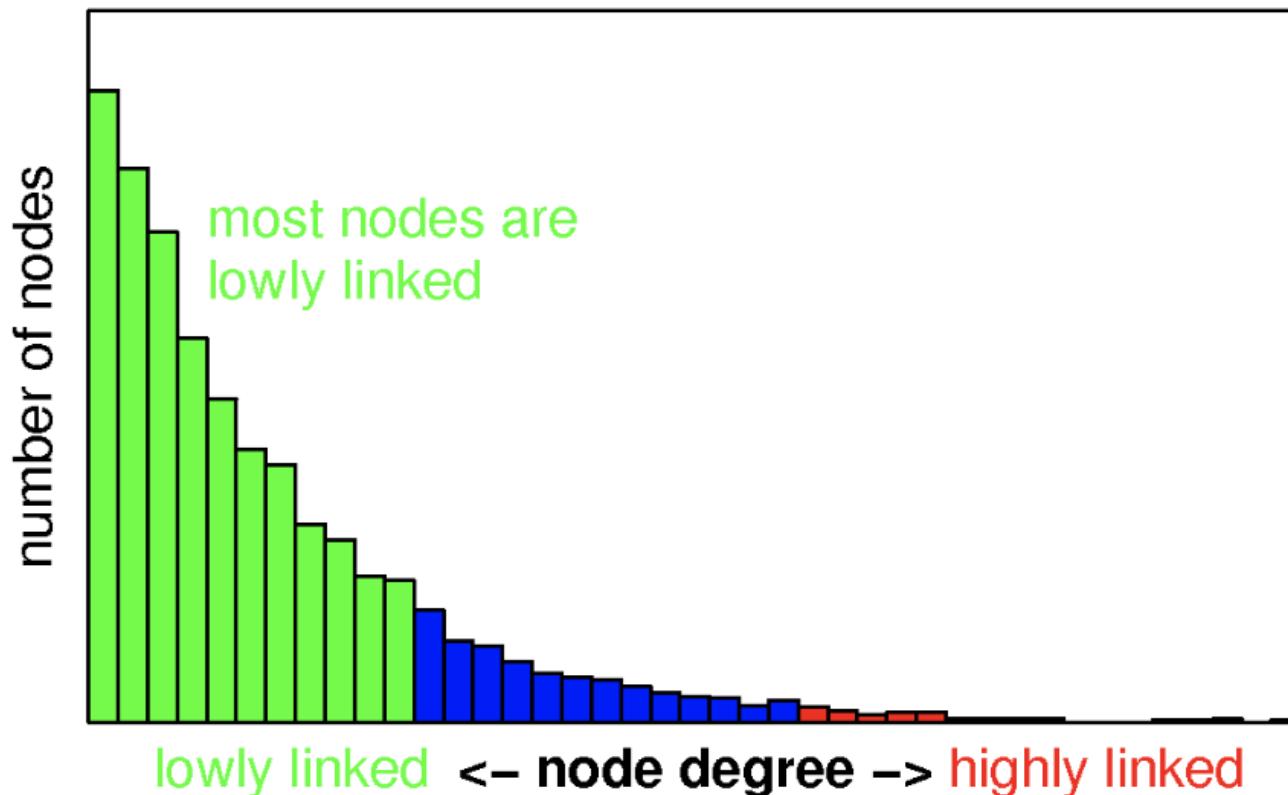
- Pareto Principle
 - 80/20 Rule

$$P_k \sim \sum_{k'=k}^{\infty} k'^{-\alpha} \sim k^{-(\alpha-1)}.$$



Degree Distribution

real networks (power-law, scale-free)



	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, 416
	company directors	undirected	7 673	55 392	14.44	4.60	—	0.59	0.88	0.276	105, 323
	math coauthorship	undirected	253 339	496 489	3.92	7.57	—	0.15	0.34	0.120	107, 182
	physics coauthorship	undirected	52 909	245 300	9.27	6.19	—	0.45	0.56	0.363	311, 313
	biology coauthorship	undirected	1 520 251	11 803 064	15.53	4.92	—	0.088	0.60	0.127	311, 313
	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	321
	student relationships	undirected	573	477	1.66	16.01	—	0.005	0.001	—0.029	45
	sexual contacts	undirected	2 810	—	—	—	3.2	—	—	—	265, 266
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/—	—	—	—	351
	Roget's Thesaurus	directed	1 022	5 103	4.99	4.87	—	0.13	0.15	0.157	244
	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	416
	train routes	undirected	587	19 603	66.79	2.16	—	—	0.69	—0.033	366
	software packages	directed	1 439	1 723	1.20	2.42	1.6/1.4	0.070	0.082	—0.016	318
	software classes	directed	1 377	2 213	1.61	1.51	—	0.033	0.012	—0.119	395
	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, 354
biological	metabolic network	undirected	765	3 686	9.64	2.56	2.2	0.090	0.67	—0.240	214
	protein interactions	undirected	2 115	2 240	2.12	6.80	2.4	0.072	0.071	—0.156	212
	marine food web	directed	135	598	4.43	2.05	—	0.16	0.23	—0.263	204
	freshwater food web	directed	92	997	10.84	1.90	—	0.20	0.087	—0.326	272
	neural network	directed	307	2 359	7.68	3.97	—	0.18	0.28	—0.226	416, 421

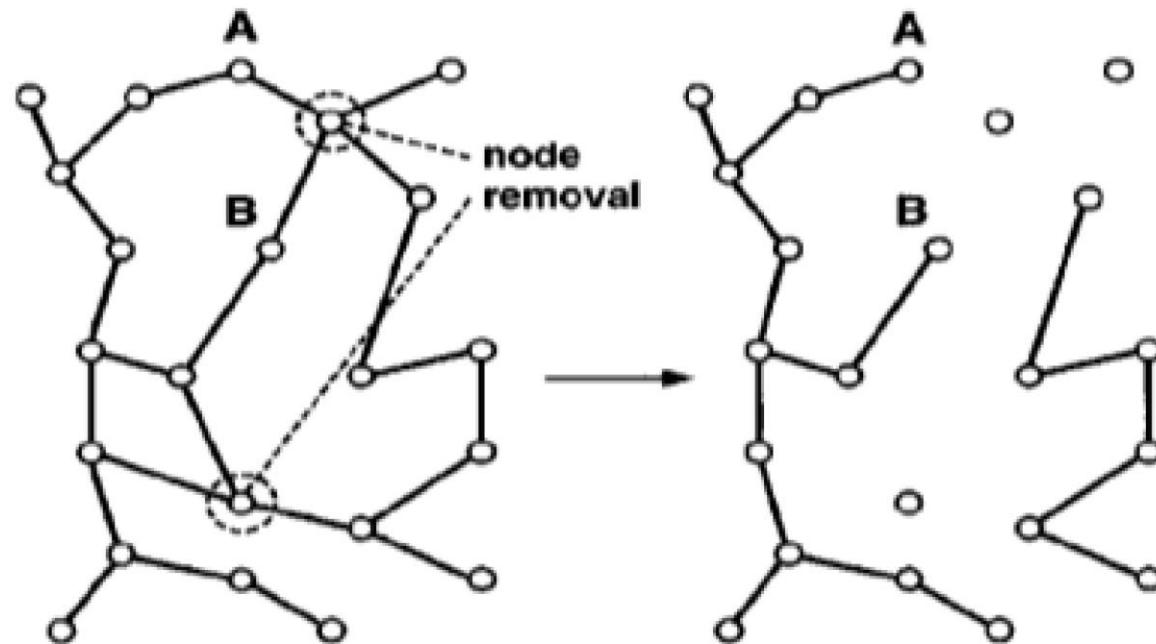
TABLE II Basic statistics for a number of published networks. The properties measured are: type of graph, directed or undirected; total number of vertices n ; total number of edges m ; mean degree z ; mean vertex–vertex distance ℓ ; exponent α of degree distribution if the distribution follows a power law (or “—” if not; in/out-degree exponents are given for directed graphs); clustering coefficient $C^{(1)}$ from Eq. (3); clustering coefficient $C^{(2)}$ from Eq. (6); and degree correlation coefficient r , Sec. III.F. The last column gives the citation(s) for the network in the bibliography. Blank entries indicate unavailable data.

Network Properties

- Average Distance between pairs (small)
- Transitivity (high)
- Degree Distribution (power law)
- Network Resilience
- Clustering Structure
 - Mixing Patterns
 - Degree Correlations

Network Resilience

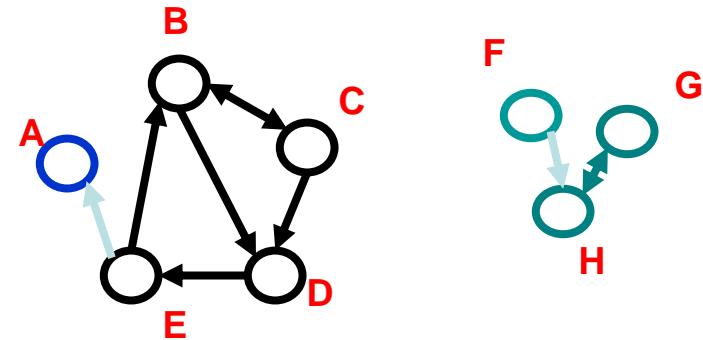
- The connectivity by removing nodes
 - Two removal ways
 - Attack by removing high centrality nodes
 - Failure by randomly removing nodes



- **Strongly connected components**
 - Each node within the component can be reached from every other node in the component by following directed links

- **Strongly connected components**

- B C D E
- A
- G H
- F



- **Weakly connected components:** every node can be reached from every other node by following links in either direction

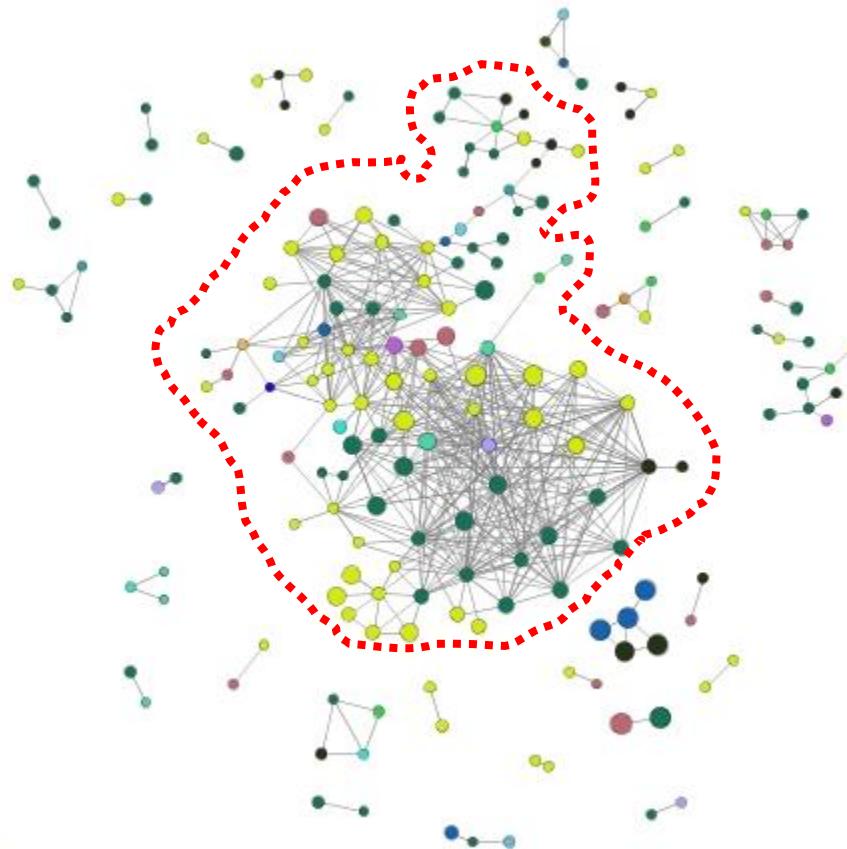
- **Weakly connected components**

- A B C D E
- G H F

- In undirected networks one talks simply about 'connected components'

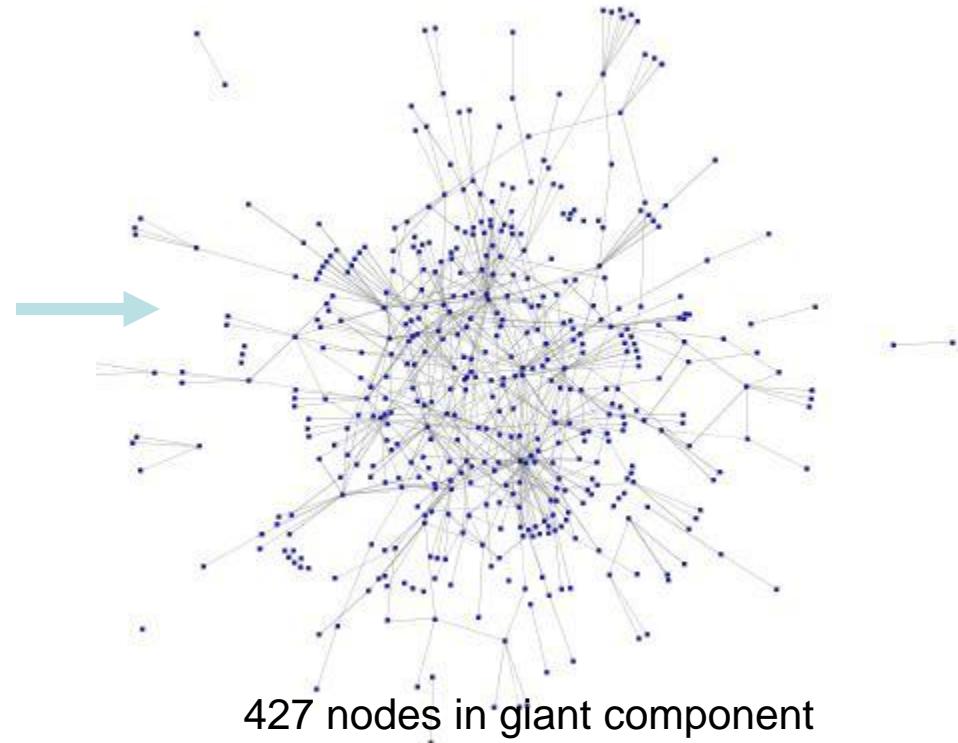
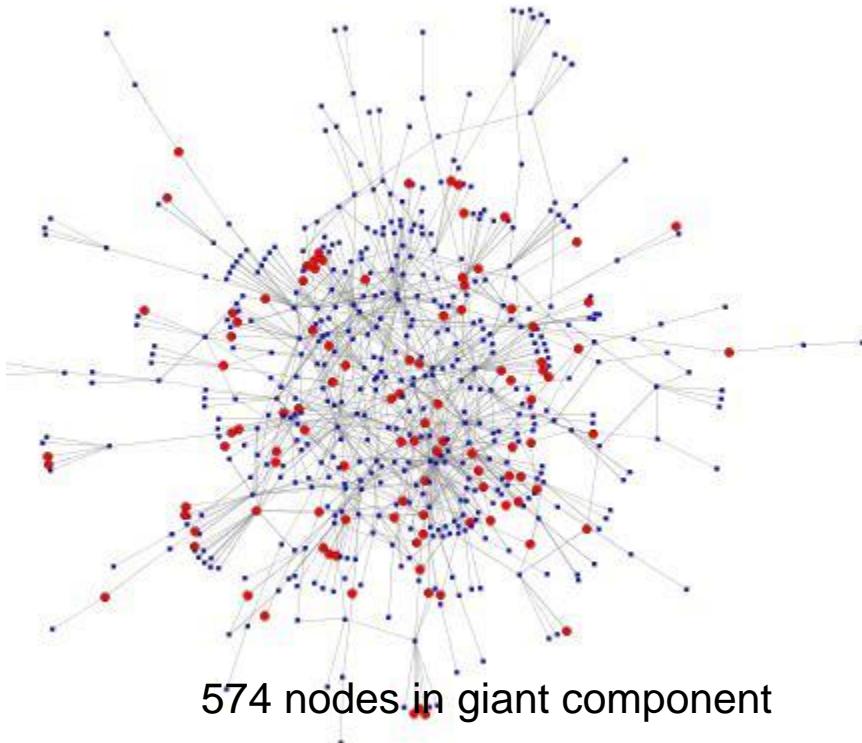
Size of Giant Component

- if the largest component encompasses a significant fraction of the graph, it is called the **giant component**



Network Resilience under failure

- gnutella network
 - 20% of nodes randomly removed



Centrality

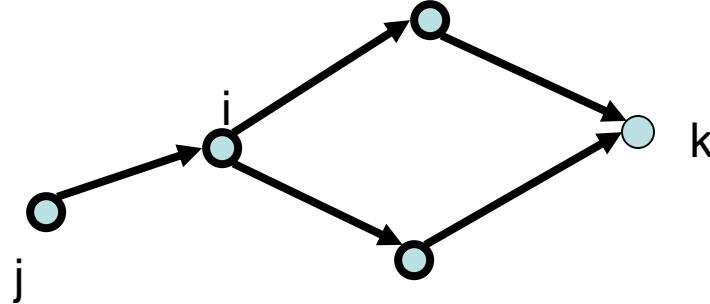
- The fraction of all directed paths between any two vertices that pass through a node

betweenness of vertex i

shortest paths between j and k that pass through i

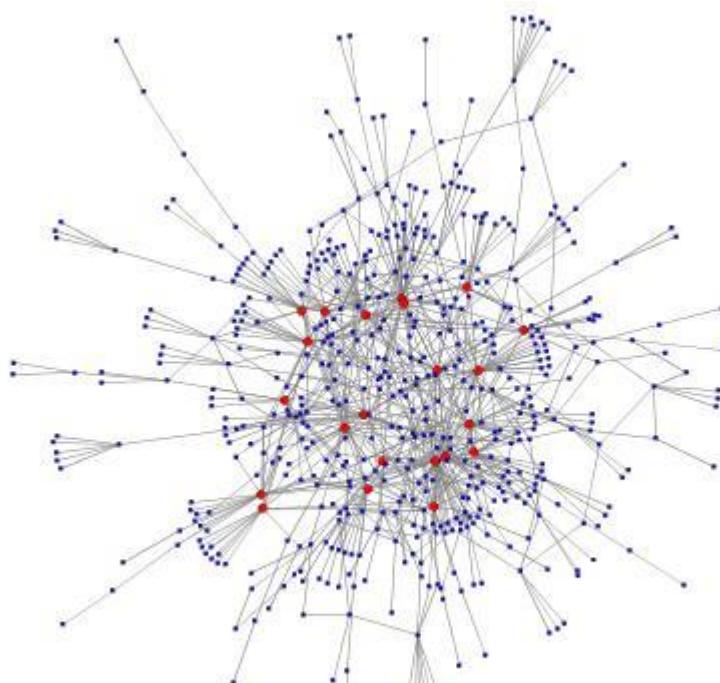
$$C_B(n_i) = \sum_{j,k} g_{jk}(i) / g_{jk}$$

all shortest paths between j and k

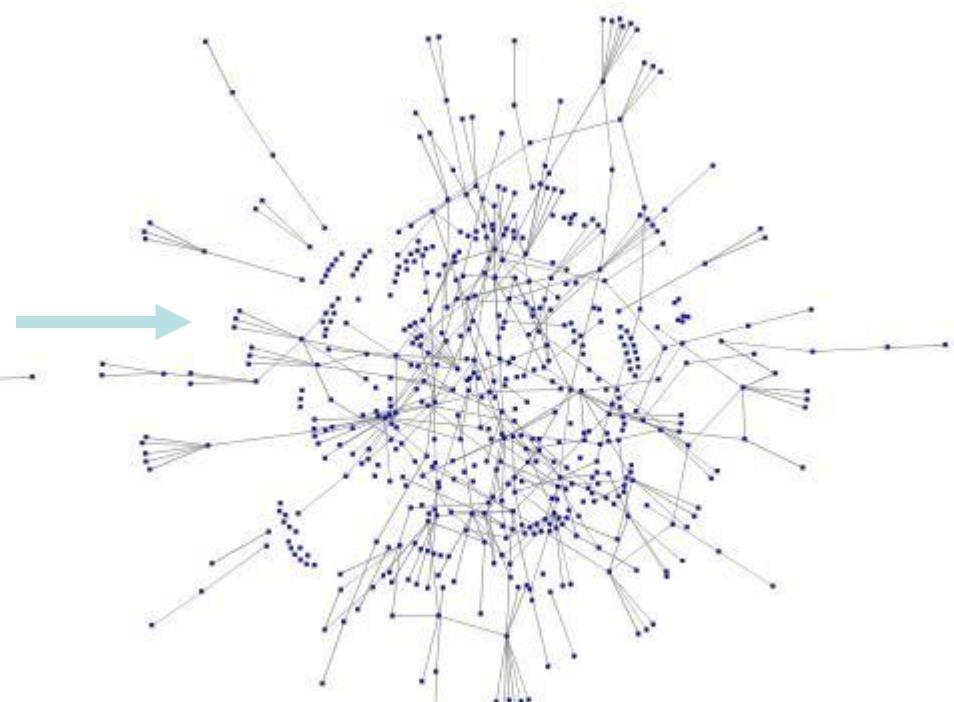


Network Resilience under attack

- gnutella network,
 - 22 most connected nodes removed (2.8% of the nodes)

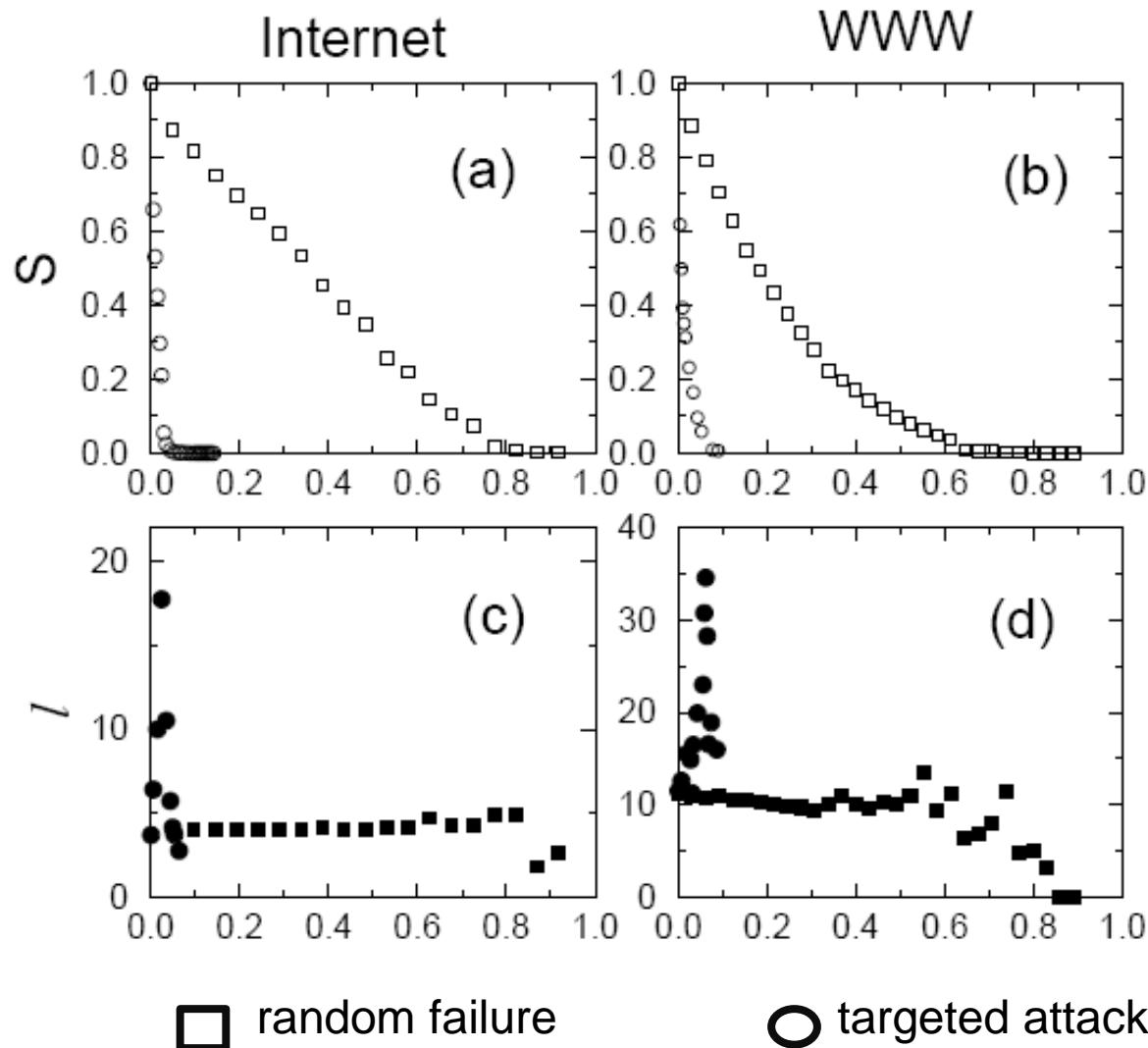


574 nodes in giant component



301 nodes in giant component

Network Resilience

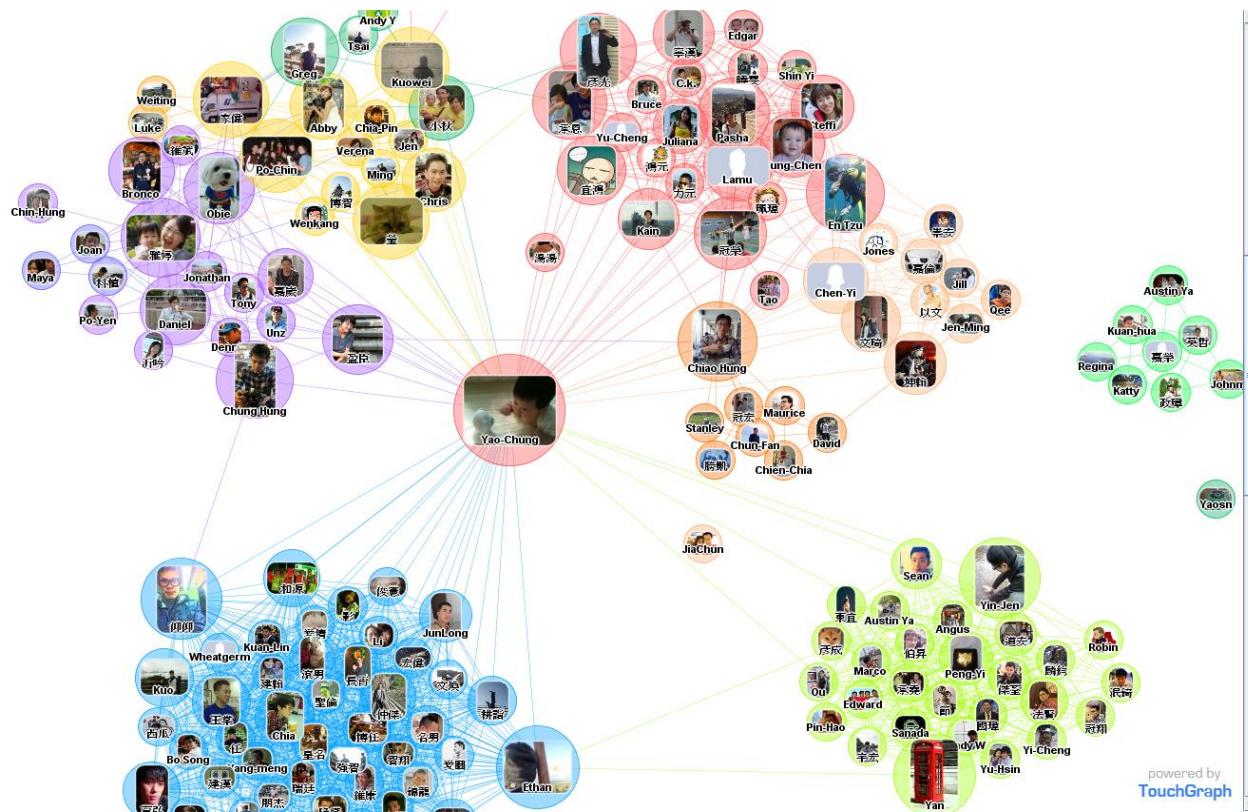


Network Properties

- Average Distance between pairs (small)
- Transitivity (high)
- Degree Distribution (power law)
- Network Resilience (Weak Under Attack)
- Clustering Structure
 - Mixing Patterns
 - Degree Correlations

Clustering Structure

- Groups of vertices that have a high density of edges within them, with a lower density of edges between groups

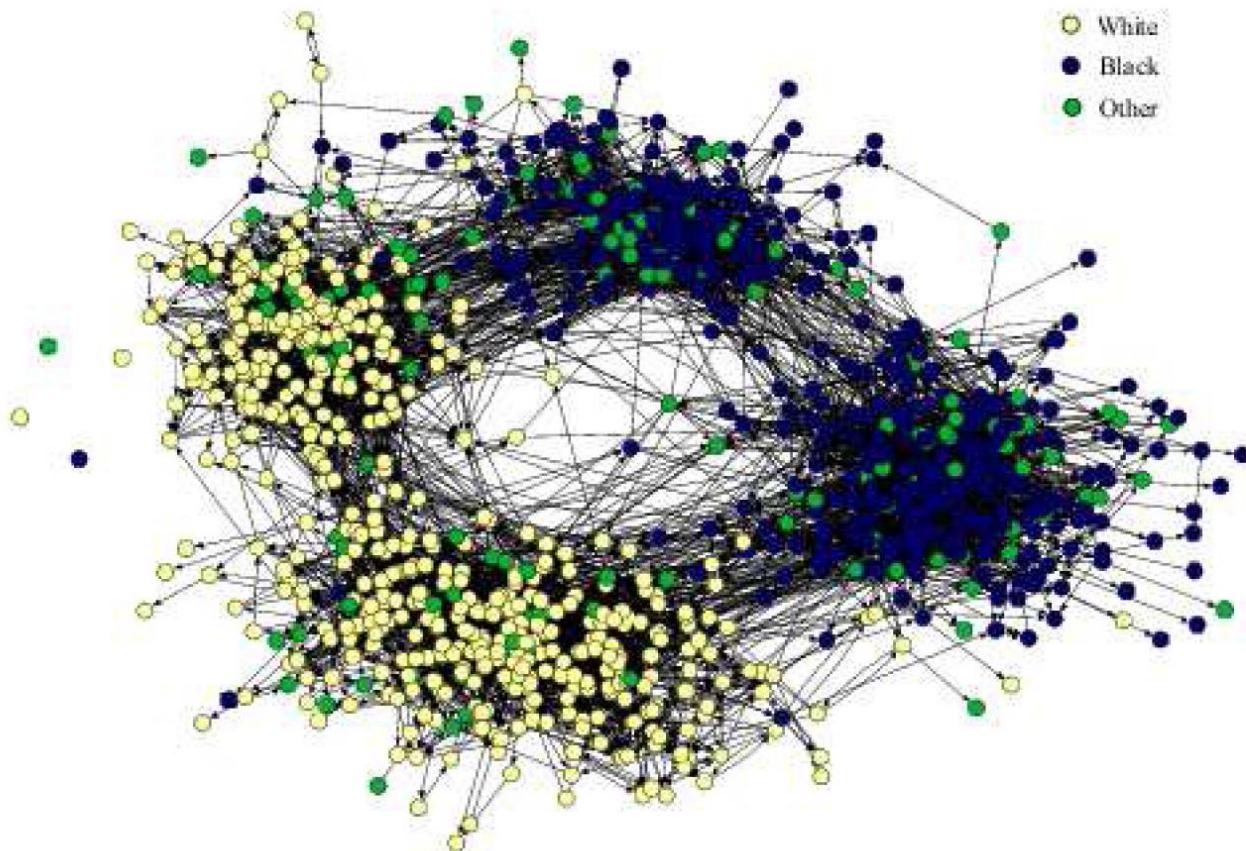


Mixing Patterns (1/2)

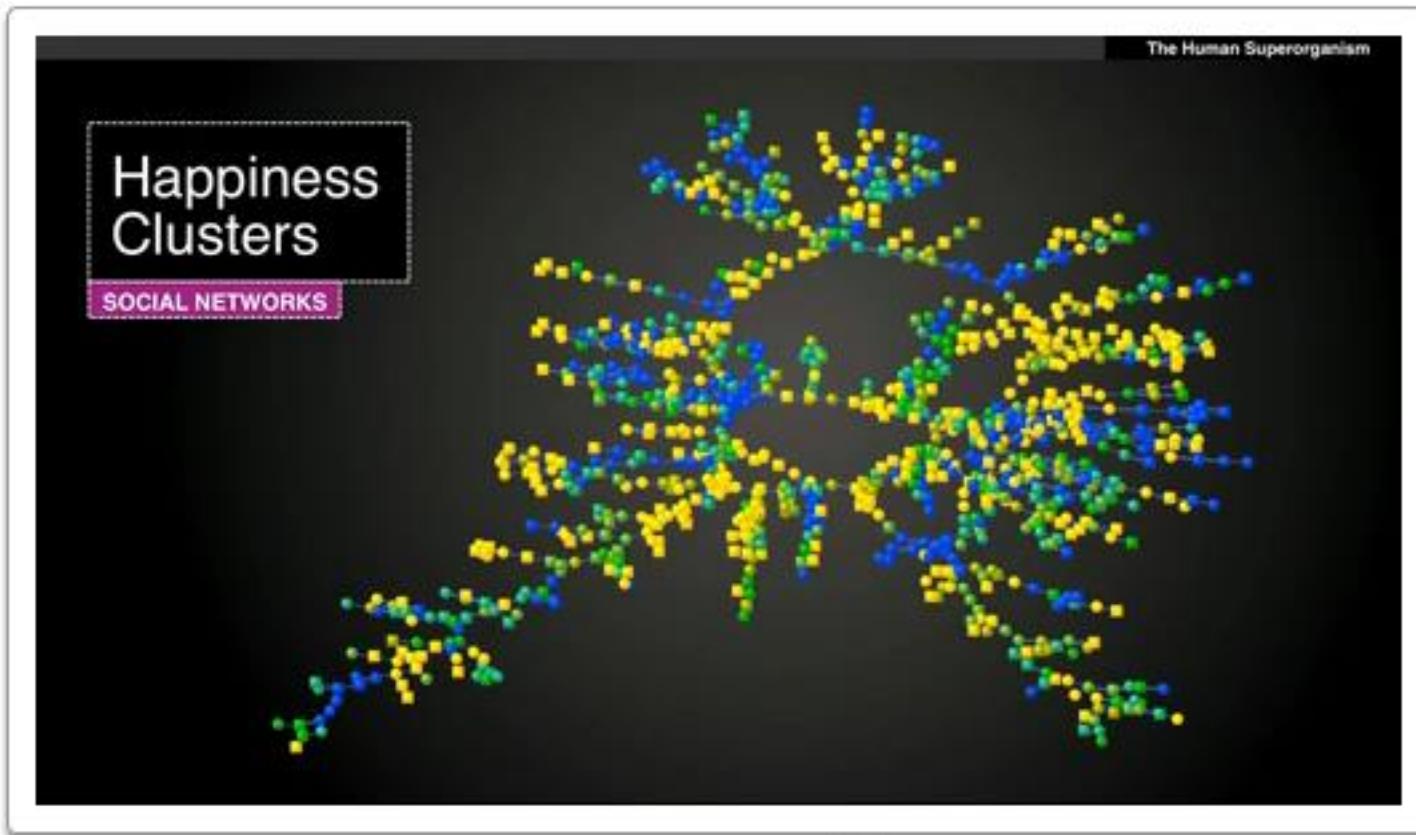
- It indicates which types of nodes are connected to which others
- The probability of connection between nodes depends on the types of them
 - assortative mixing

		women			
		black	hispanic	white	other
men	black	506	32	69	26
	hispanic	23	308	114	38
	white	26	46	599	68
	other	10	14	47	32

Mixing Patterns (2/2)



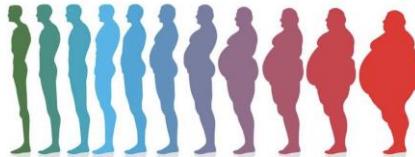
Mixing Patterns



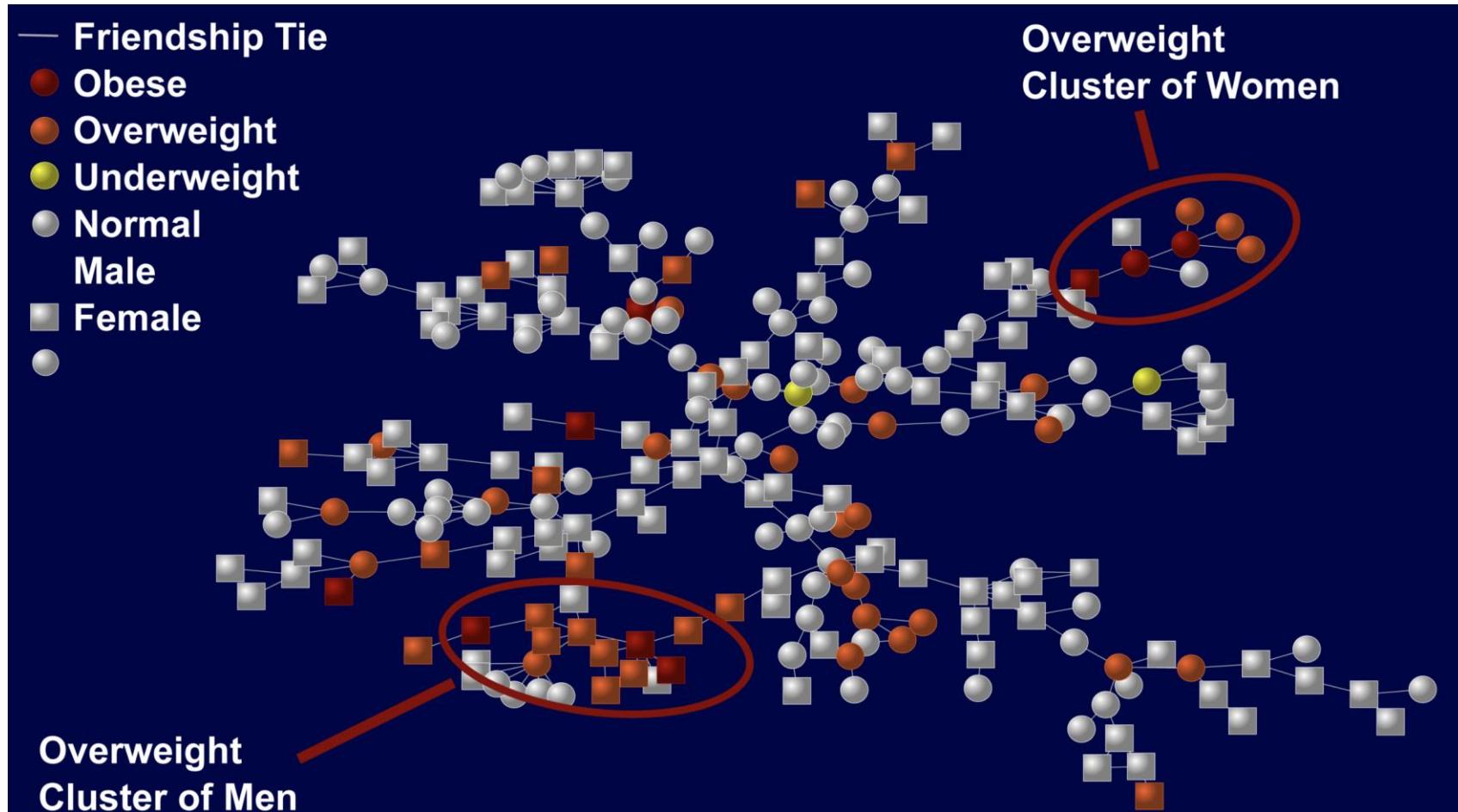
- Happy people
- In between people
- Unhappy people

Christakis & Fowler, 2011

**OBESITY IS NOW A
GLOBAL EPIDEMIC!**



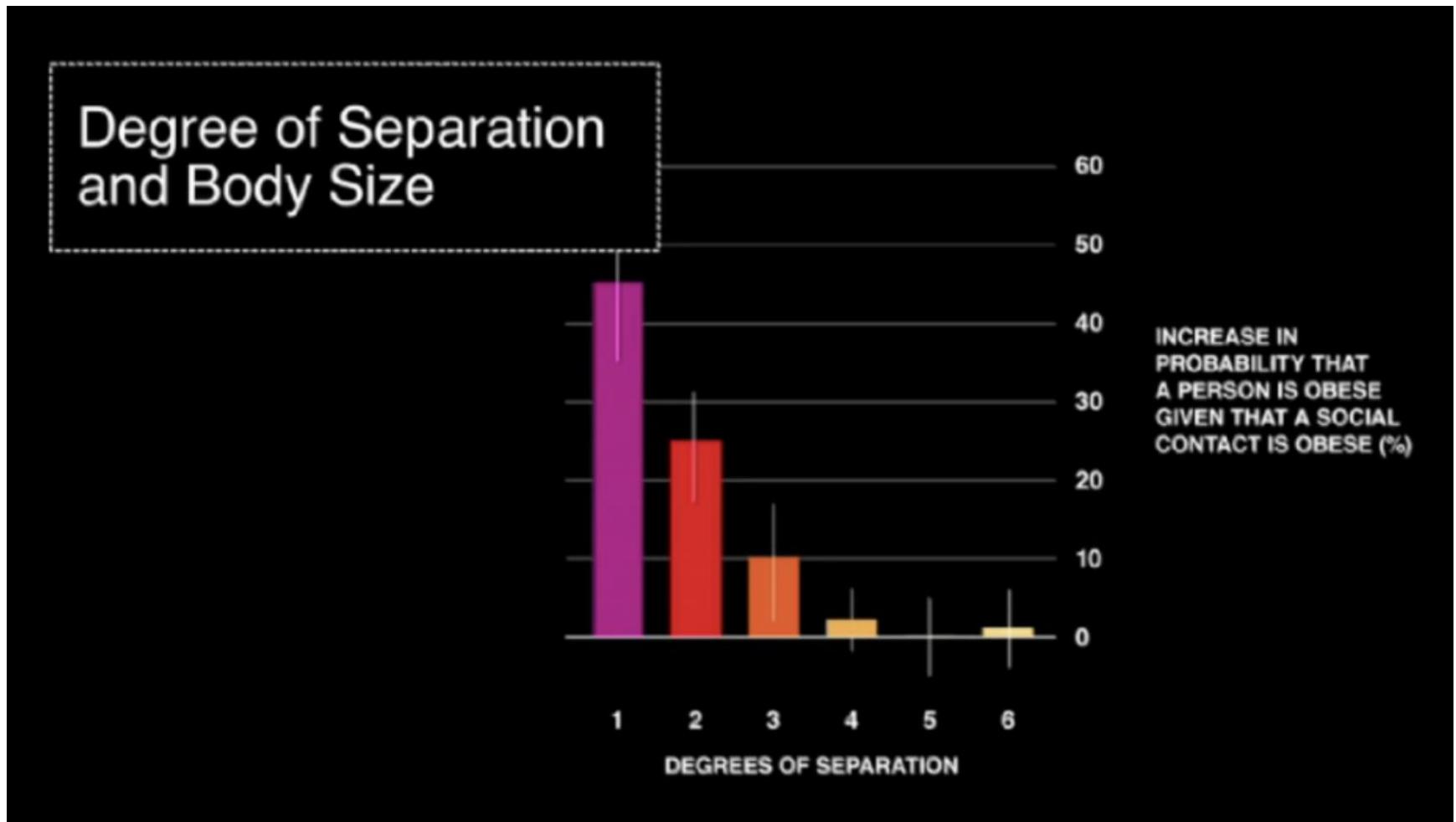
Mixing Patterns



<http://humannaturelab.net/resources/images/>

<https://www.youtube.com/watch?v=pJfq-o5nZQ4>

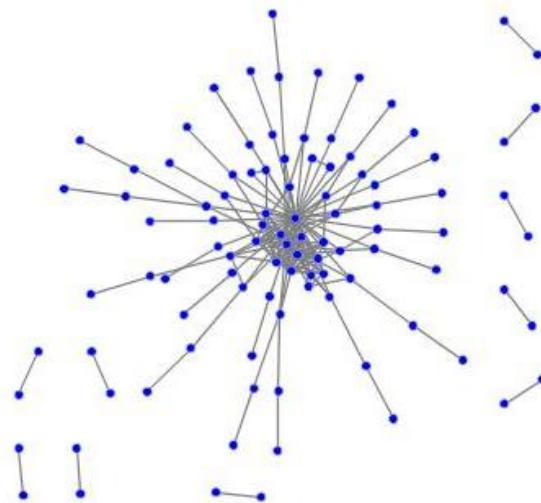
肥胖會傳染？



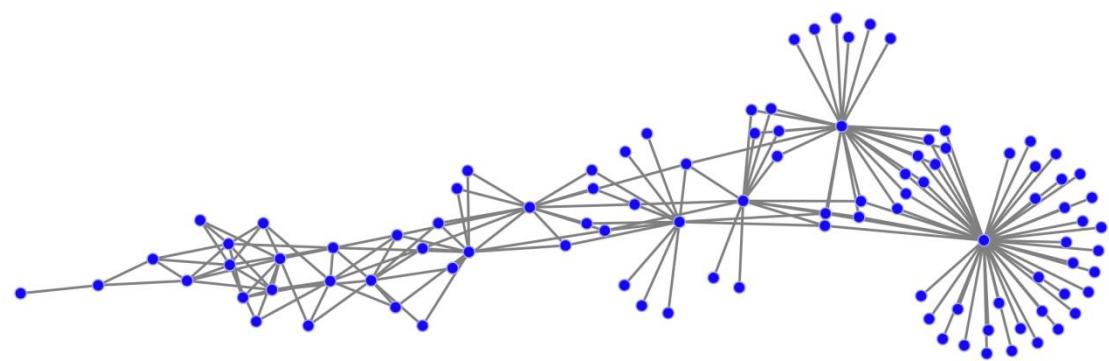
<https://www.youtube.com/watch?v=2U-tOghblfE>

Degree Correlation

Mixing Patterns by Degrees



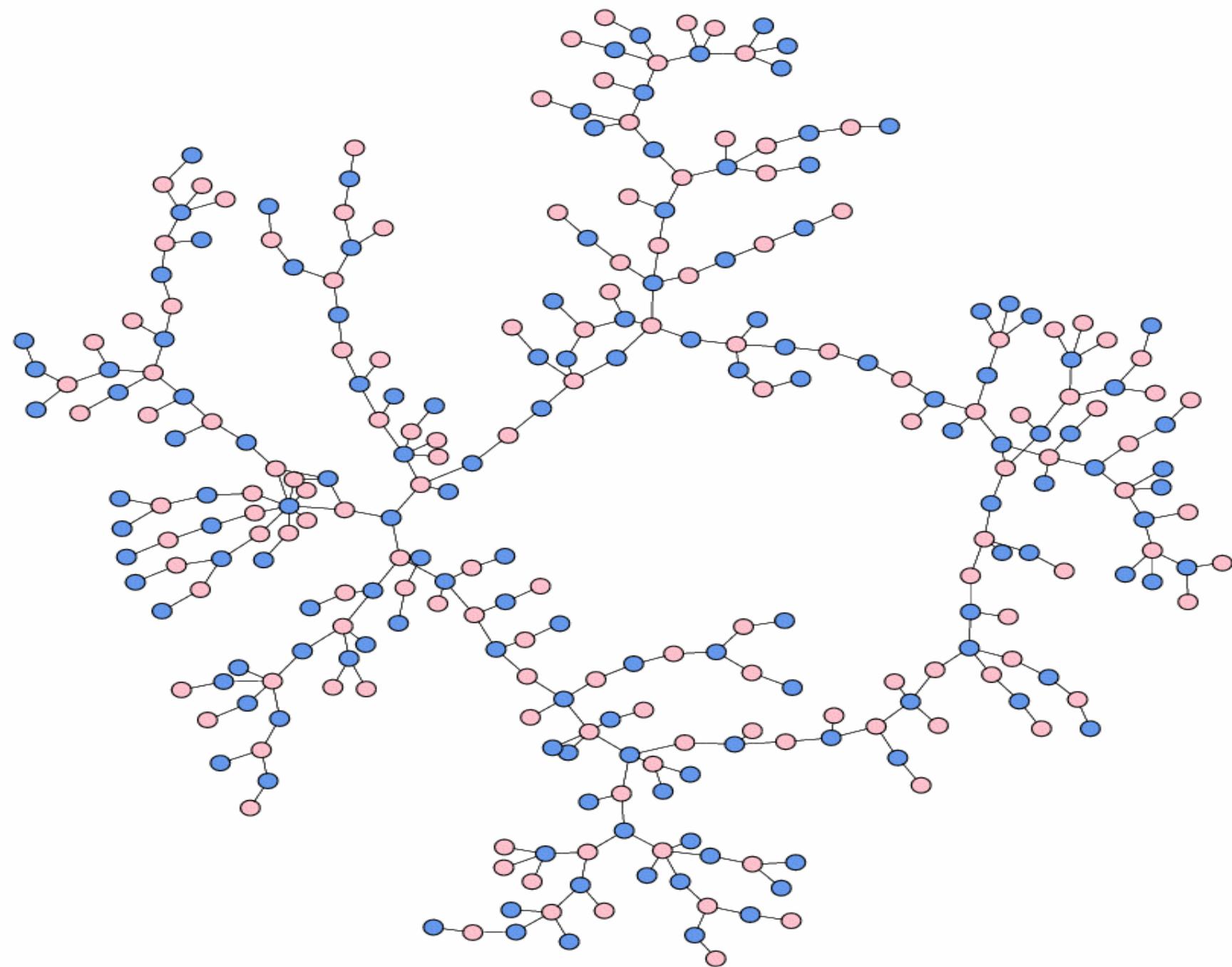
assortative



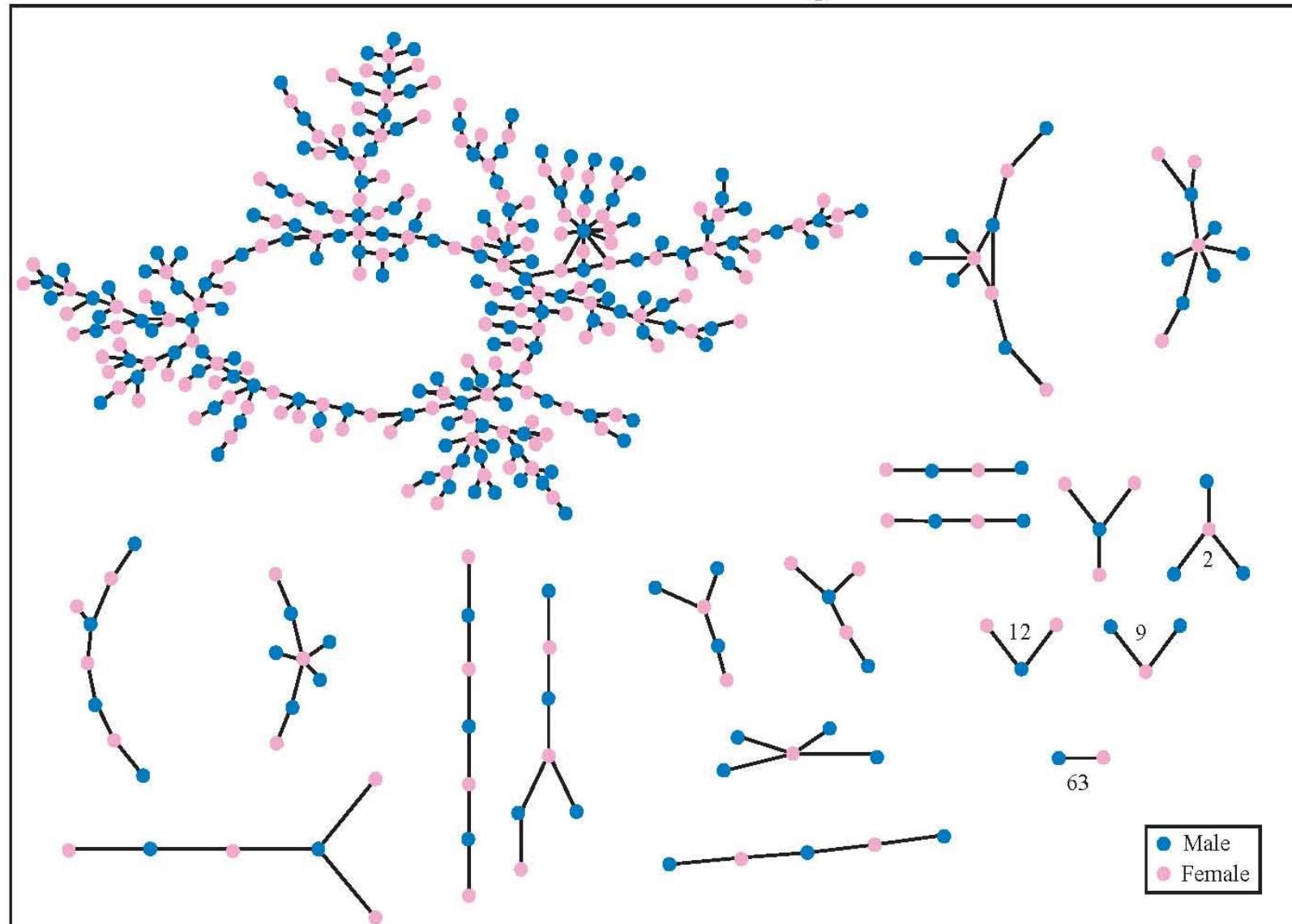
disassortative

	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, 416
	company directors	undirected	7 673	55 392	14.44	4.60	—	0.59	0.88	0.276	105, 323
	math coauthorship	undirected	253 339	496 489	3.92	7.57	—	0.15	0.34	0.120	107, 182
	physics coauthorship	undirected	52 909	245 300	9.27	6.19	—	0.45	0.56	0.363	311, 313
	biology coauthorship	undirected	1 520 251	11 803 064	15.53	4.92	—	0.088	0.60	0.127	311, 313
	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	321
	student relationships	undirected	573	477	1.66	16.01	—	0.005	0.001	—0.029	45
	sexual contacts	undirected	2 810	—	—	—	3.2	—	—	—	265, 266
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/—	—	—	—	351
	Roget's Thesaurus	directed	1 022	5 103	4.99	4.87	—	0.13	0.15	0.157	244
	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	416
	train routes	undirected	587	19 603	66.79	2.16	—	—	0.69	—0.033	366
	software packages	directed	1 439	1 723	1.20	2.42	1.6/1.4	0.070	0.082	—0.016	318
	software classes	directed	1 377	2 213	1.61	1.51	—	0.033	0.012	—0.119	395
	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, 354
biological	metabolic network	undirected	765	3 686	9.64	2.56	2.2	0.090	0.67	—0.240	214
	protein interactions	undirected	2 115	2 240	2.12	6.80	2.4	0.072	0.071	—0.156	212
	marine food web	directed	135	598	4.43	2.05	—	0.16	0.23	—0.263	204
	freshwater food web	directed	92	997	10.84	1.90	—	0.20	0.087	—0.326	272
	neural network	directed	307	2 359	7.68	3.97	—	0.18	0.28	—0.226	416, 421

TABLE II Basic statistics for a number of published networks. The properties measured are: type of graph, directed or undirected; total number of vertices n ; total number of edges m ; mean degree z ; mean vertex–vertex distance ℓ ; 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 Eq. (3); clustering coefficient $C^{(2)}$ from Eq. (6); and degree correlation coefficient r , Sec. III.F. The last column gives the citation(s) for the network in the bibliography. Blank entries indicate unavailable data.



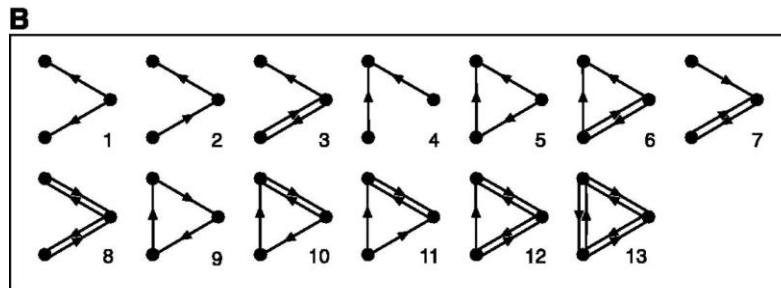
The Structure of Romantic and Sexual Relations at "Jefferson High School"



Each circle represents a student and lines connecting students represent romantic relations occurring within the 6 months preceding the interview. Numbers under the figure count the number of times that pattern was observed (i.e. we found 63 pairs unconnected to anyone else).

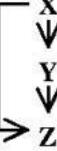
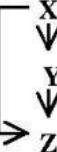
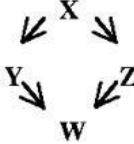
Network Motif

- The basic building pattern of a network

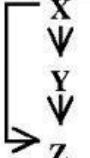
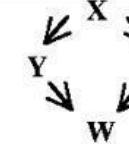
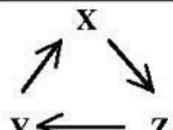
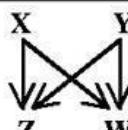
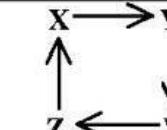
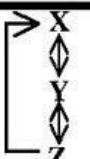
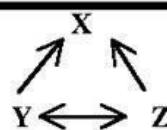


- Using brute force, searched target network for every possible subgraph, counting results
- Similarly, searched random network
- **Motifs** are patterns that occur greater or equal number of times in random networks more than 1% of the time.

Results

Network	Nodes	Edges	N_{real}	$N_{\text{rand}} \pm \text{SD}$	Z score	N_{real}	$N_{\text{rand}} \pm \text{SD}$	Z score	N_{real}	$N_{\text{rand}} \pm \text{SD}$	Z score
Gene regulation (transcription)											
				Feed-forward loop			Bi-fan				
<i>E. coli</i>	424	519	40	7 ± 3	10	203	47 ± 12	13			
<i>S. cerevisiae</i> *	685	1,052	70	11 ± 4	14	1812	300 ± 40	41			
Neurons											
				Feed-forward loop			Bi-fan			Bi-parallel	
<i>C. elegans</i> †	252	509	125	90 ± 10	3.7	127	55 ± 13	5.3	227	35 ± 10	20
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			
St. Martin	42	205	469	450 ± 10	NS	382	130 ± 20	12			
Chesapeake	31	67	80	82 ± 4	NS	26	5 ± 2	8			
Coachella	29	243	279	235 ± 12	3.6	181	80 ± 20	5			
Skipwith	25	189	184	150 ± 7	5.5	397	80 ± 25	13			
B. Brook	25	104	181	130 ± 7	7.4	267	30 ± 7	32			

Results (cont.)

Network	Nodes	Edges	N_{real}	$N_{\text{rand}} \pm \text{SD}$	Z score	N_{real}	$N_{\text{rand}} \pm \text{SD}$	Z score	N_{real}	$N_{\text{rand}} \pm \text{SD}$	Z score
Electronic circuits (forward logic chips)				Feed-forward loop			Bi-fan			Bi-parallel	
s15850	10,383	14,240	424	2 ± 2	285	1040	1 ± 1	1200	480	2 ± 1	335
s38584	20,717	34,204	413	10 ± 3	120	1739	6 ± 2	800	711	9 ± 2	320
s38417	23,843	33,661	612	3 ± 2	400	2404	1 ± 1	2550	531	2 ± 2	340
s9234	5,844	8,197	211	2 ± 1	140	754	1 ± 1	1050	209	1 ± 1	200
s13207	8,651	11,831	403	2 ± 1	225	4445	1 ± 1	4950	264	2 ± 1	200
Electronic circuits (digital fractional multipliers)				Three-node feedback loop			Bi-fan			Four-node feedback 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
s838‡	512	819	40	1 ± 1	38	22	1 ± 1	20	23	1 ± 1	25
World Wide Web				Feedback with two mutual dyads			Fully connected triad			Uplinked mutual dyad	
nd.edu§	325,729	1.46e6	1.1e5	$2e3 \pm 1e2$	800	6.8e6	$5e4 \pm 4e2$	15,000	1.2e6	$1e4 \pm 2e2$	5000