Networks

Modelling Complex Systems

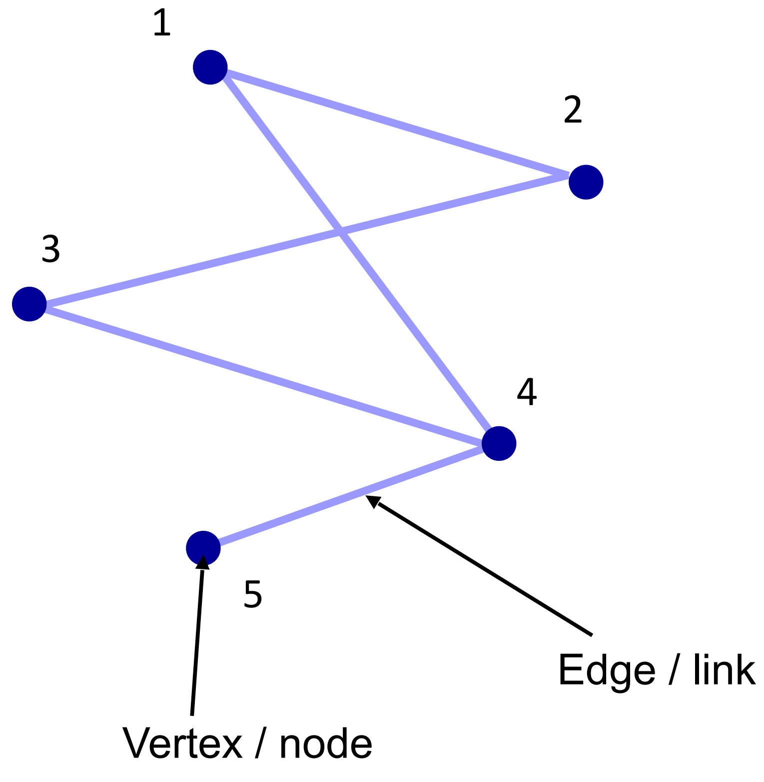
Some of this lecture is adapted from:
Albert and Barabasi, Reviews of Modern Physics 74 (2002)
M. Barthelemy, Physics Reports 499 (2011)
Newman, Networks (2011)
-previous slides of David Sumpter.

Networks

- Things with connections

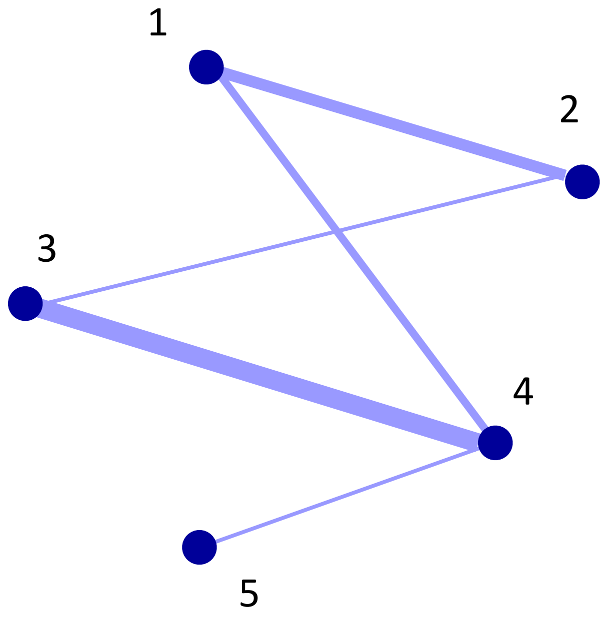
Networks

- Things with connections
- Or, “real life” graphs



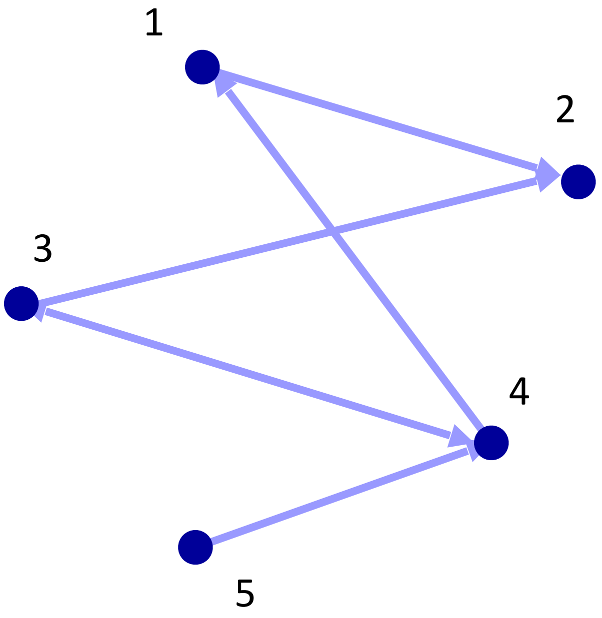
Networks

Can be weighted or unweighted



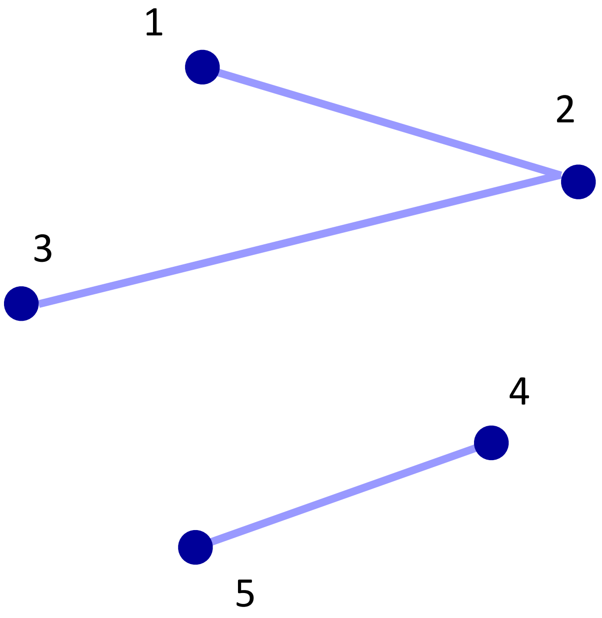
Networks

Can be directed or undirected



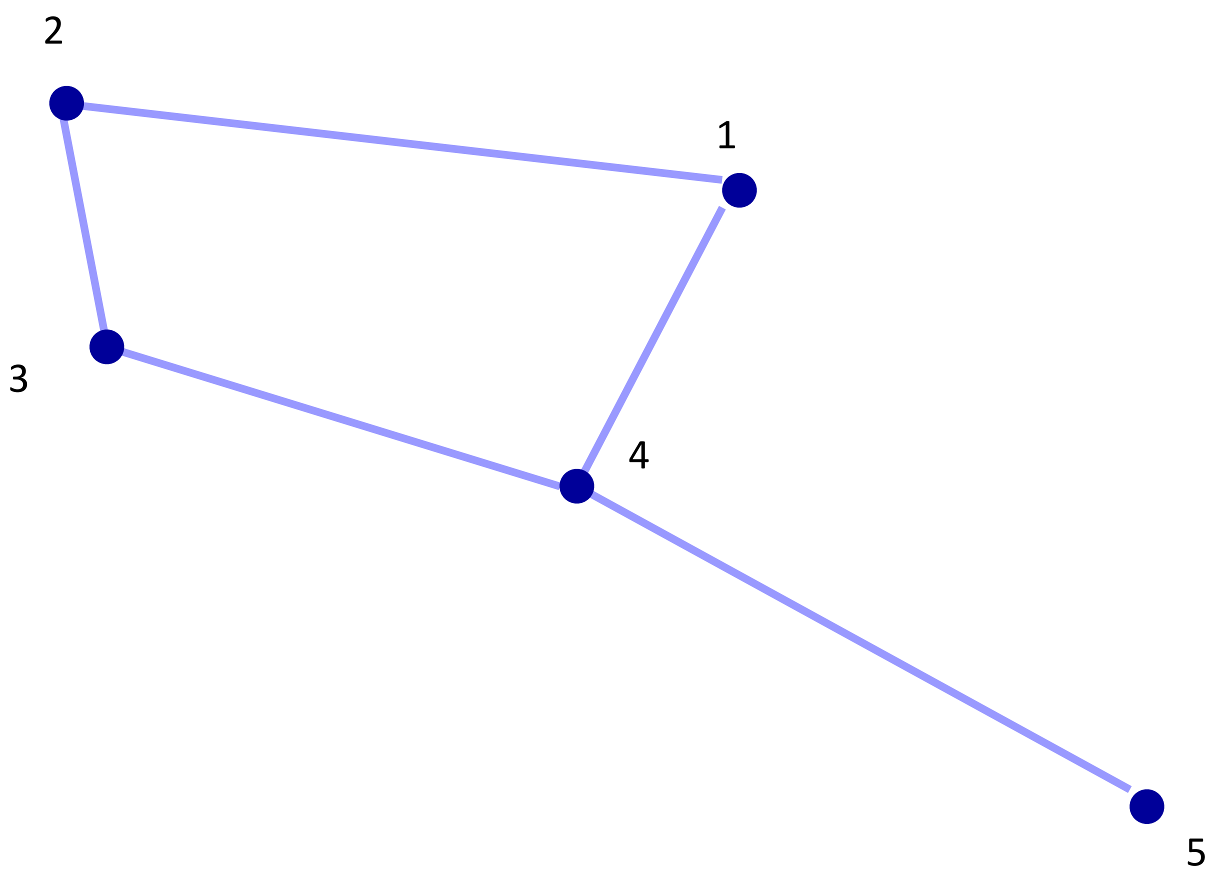
Networks

Can be connected or disjoint



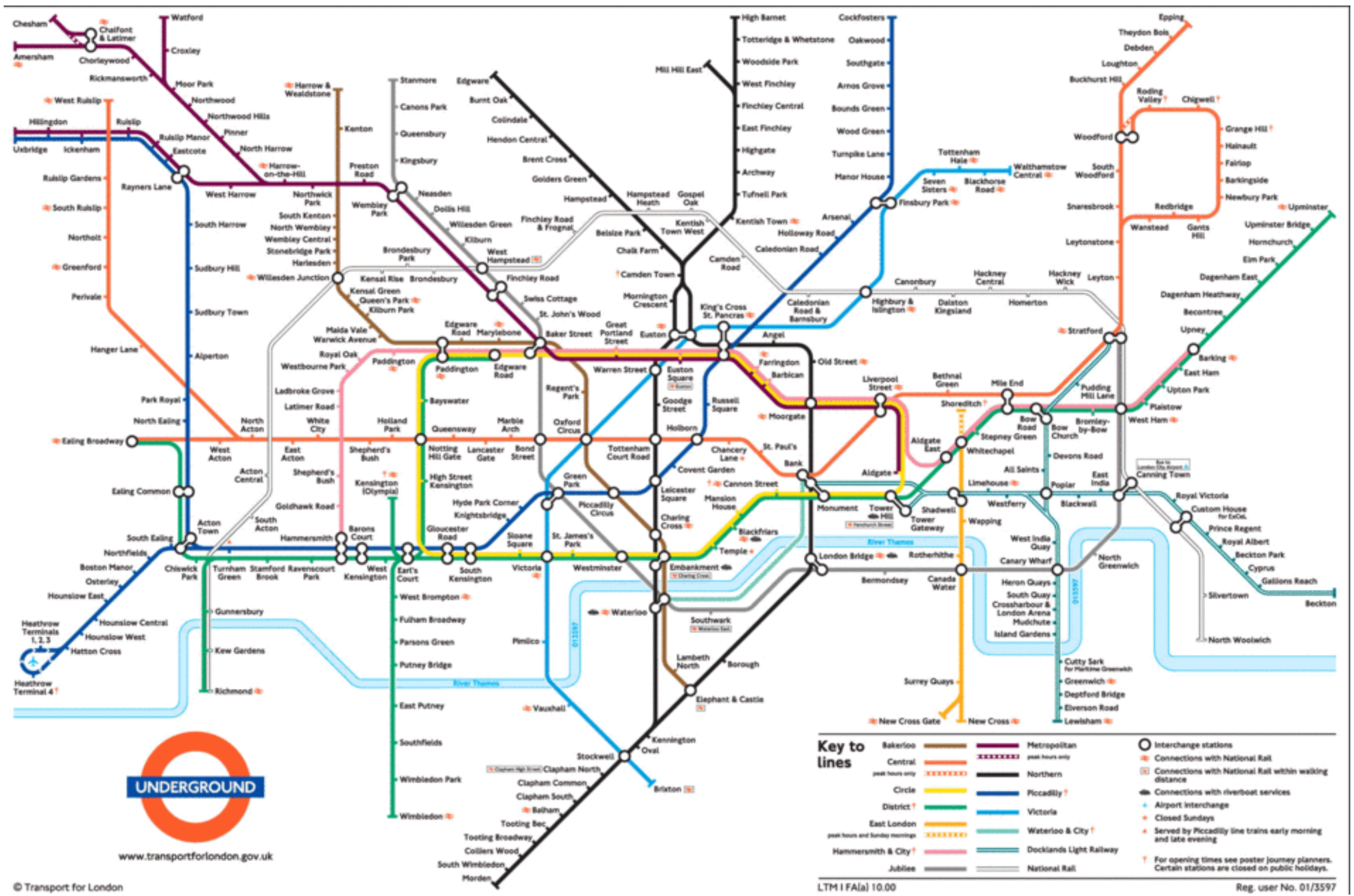
Networks

Can be planar or non-planar



Real-world Networks

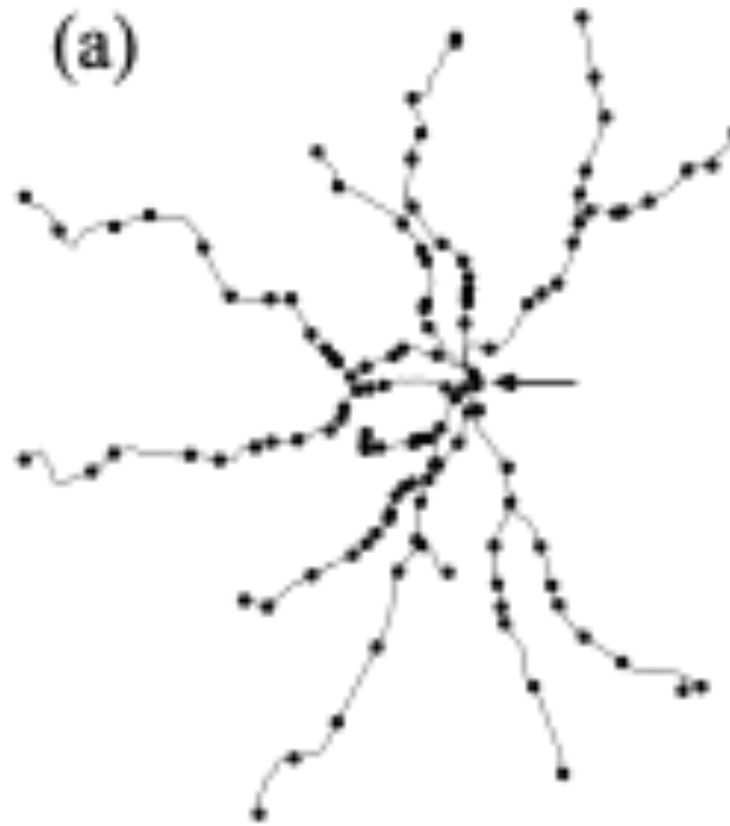
Planned networks

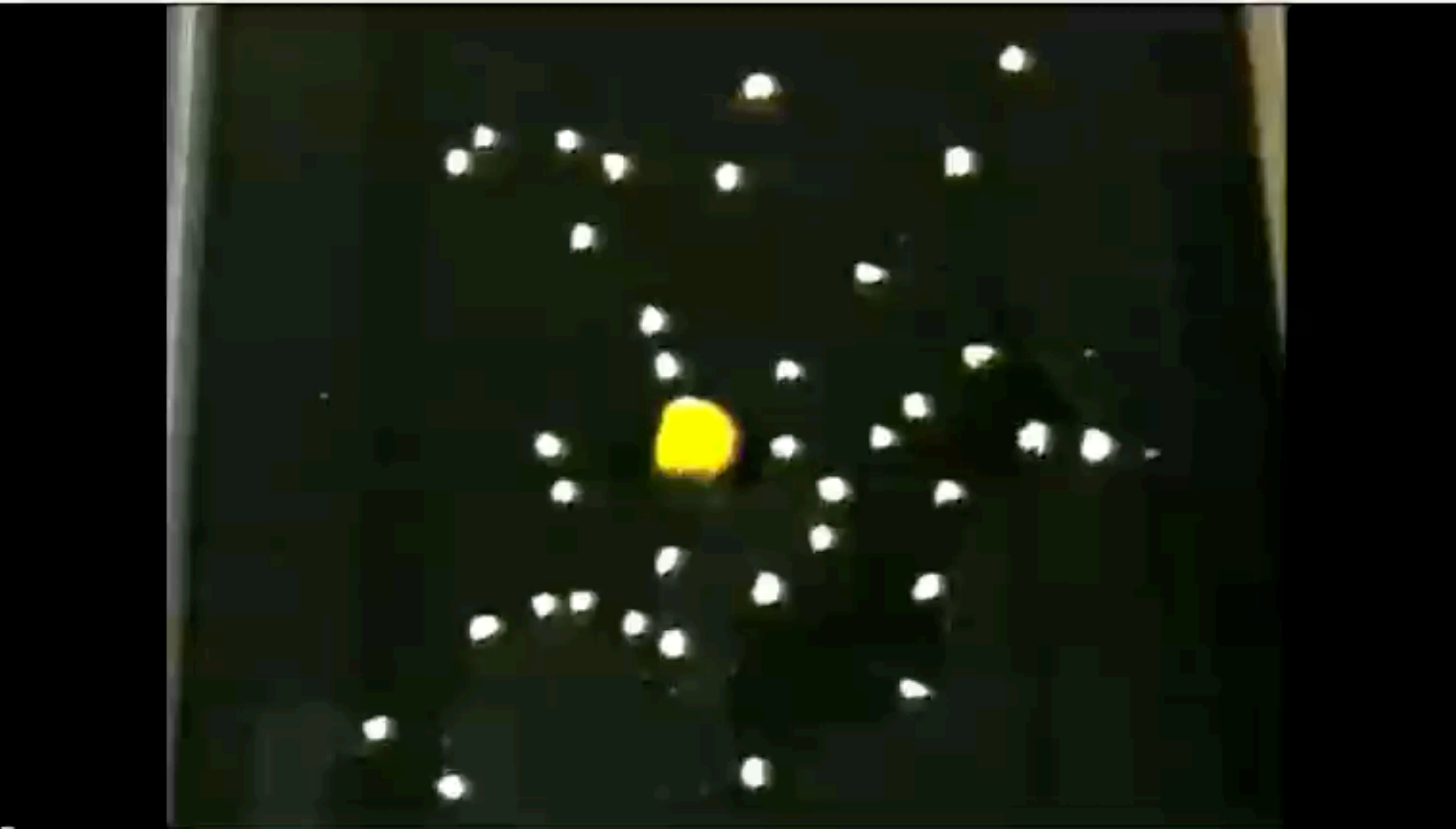


Real-world Networks

Commuter rail network in Boston area.

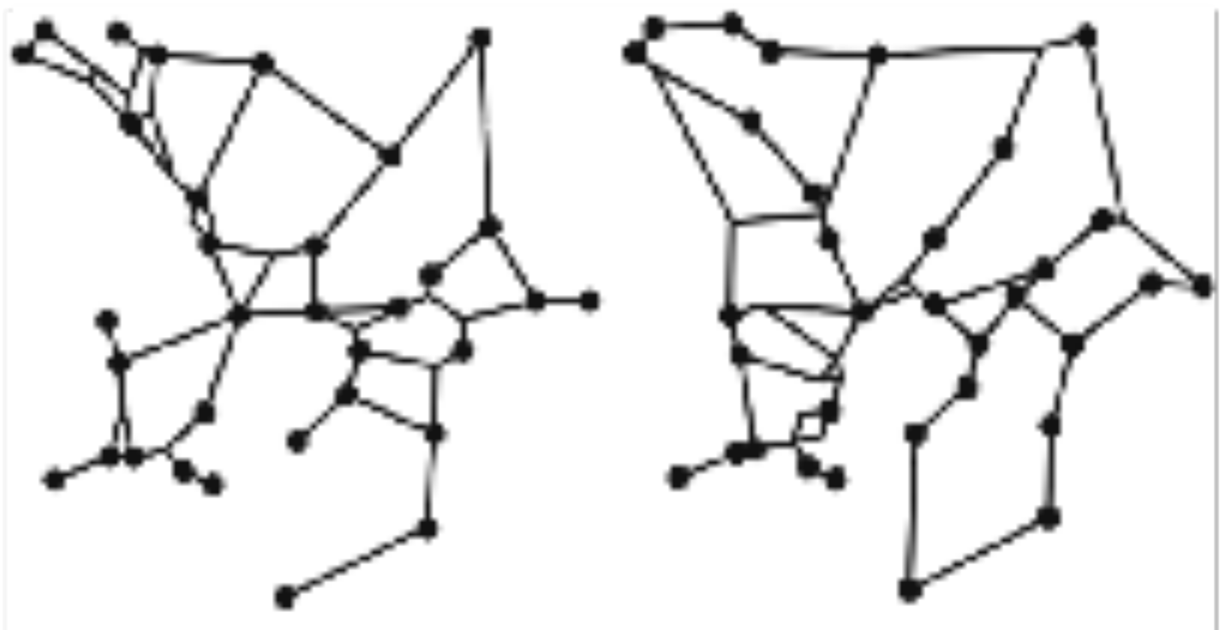
Physical and planar.





Toshi Nakagaki and co-workers

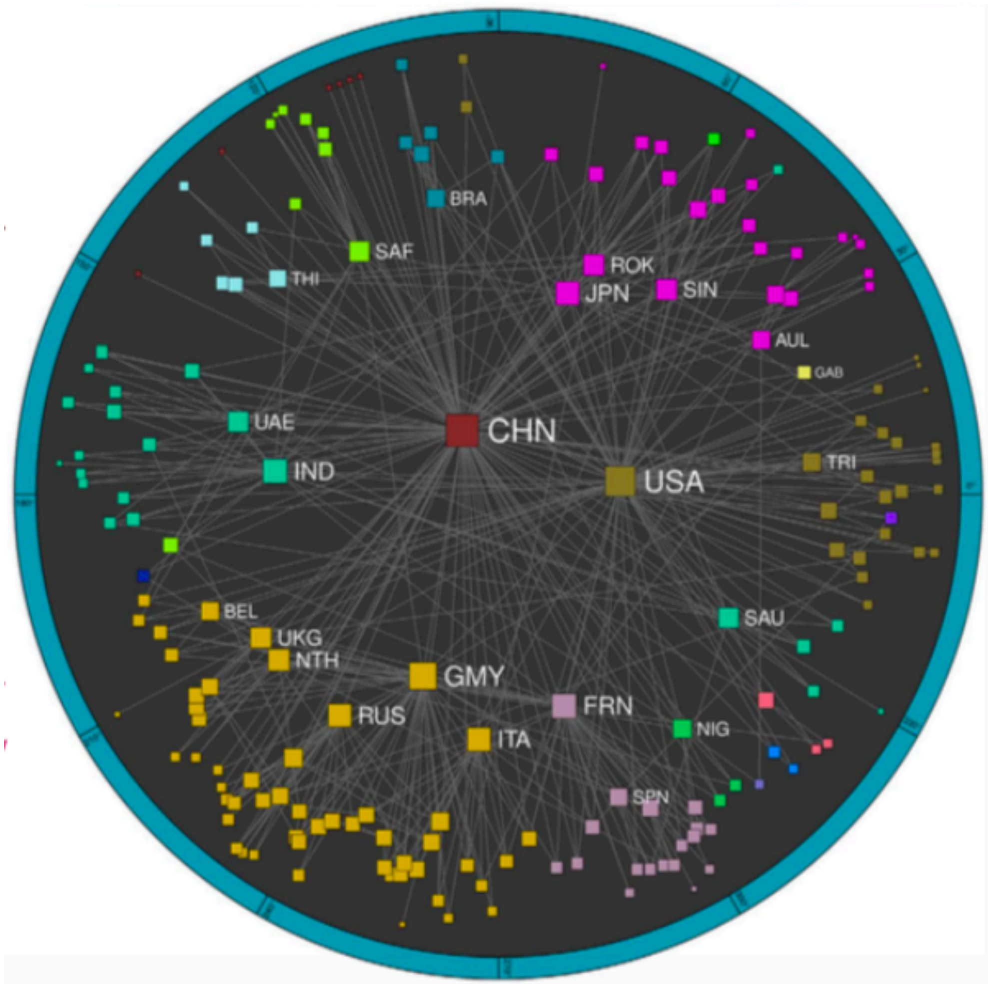
Real-world Networks



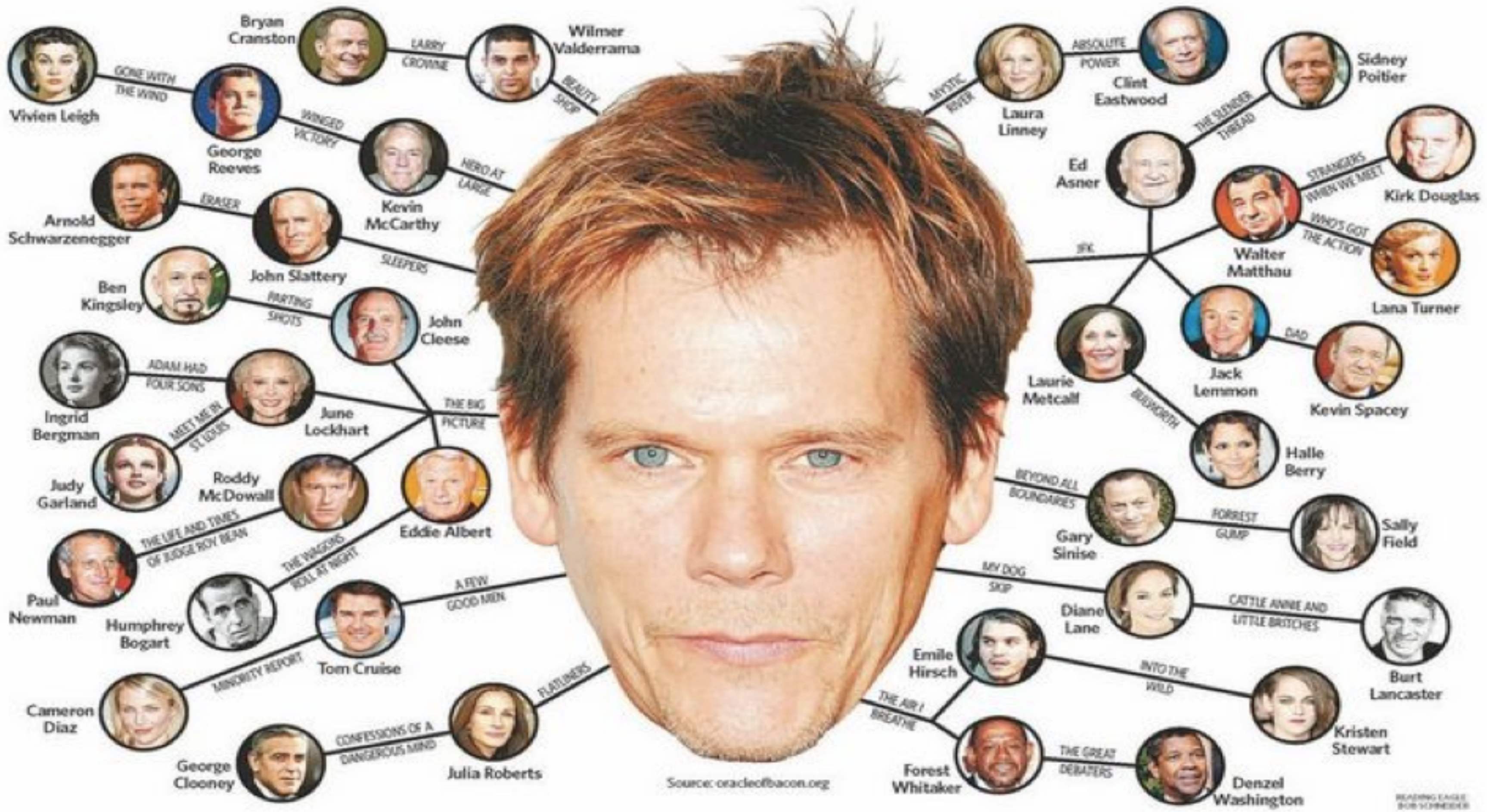
Slime mould

Tokyo Engineers

Real-world Networks

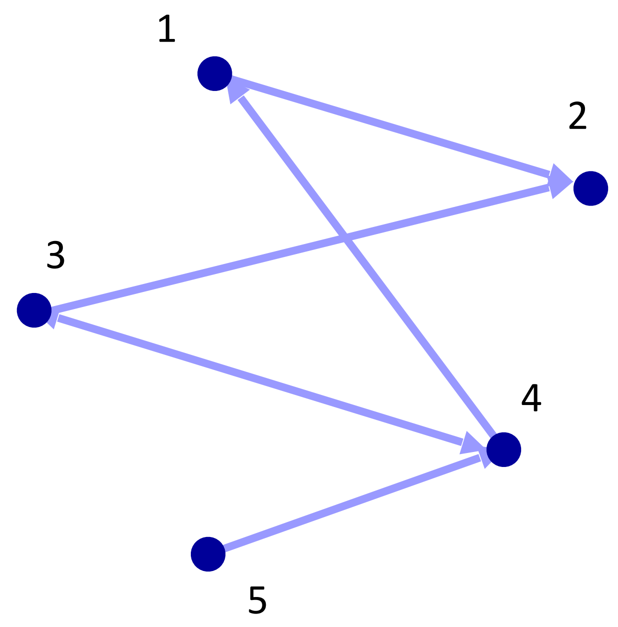


Real-world Networks



Representing Networks

source	destination	weight
1	2	1
4	1	1
3	2	1
3	4	1
4	3	1
5	4	1



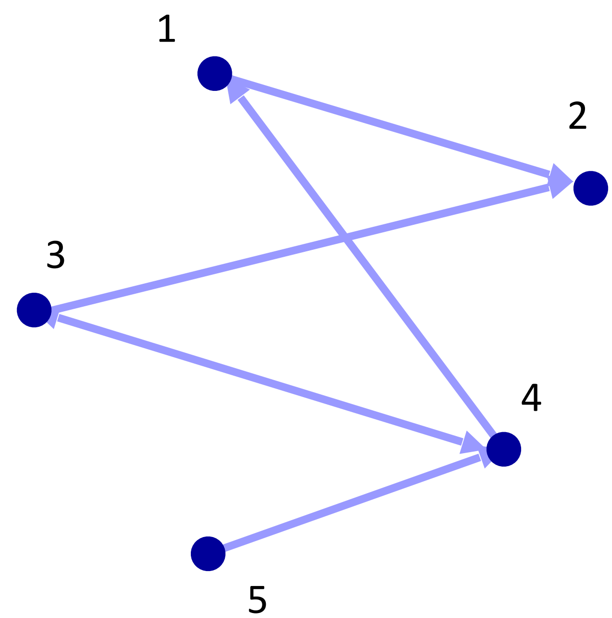
n=5 nodes

Representing Networks

- Adjacency matrix A_{ij}

Destination

		j=1				j=5
	i=1	0	1	0	0	0
		0	0	0	0	0
Source		0	1	0	1	0
		1	0	1	0	0
	i=5	0	0	0	1	0



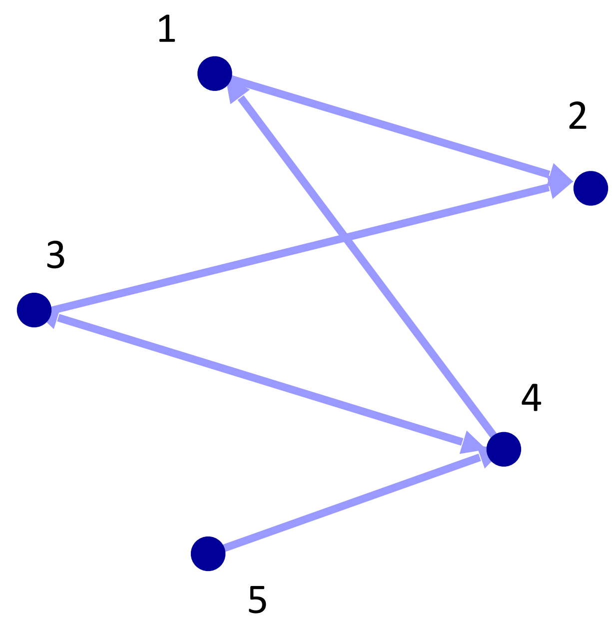
Representing Networks

- Adjacency matrix A_{ij}

out-degree

Destination

		j=1				j=5	out-degree
	i=1	0	1	0	0	0	1
		0	0	0	0	0	0
Source		0	1	0	1	0	2
		1	0	1	0	0	2
	i=5	0	0	0	1	0	1
In-degree →		1	2	1	2	0	



Another handy property: $(A^n)_{ij}$ tells us whether you can go from i to j in n steps

Other networks

- Hypergraph
- Multi-layer Network
- Temporal Network

Five (of many) network measures

- Average degree
- Degree distribution
- Mean path length
- Clustering coefficient
- Maximum modularity/
  Community partitions

....

Degree and average degree

The in in and out degrees are

$$k_i^{in} = \sum_{j=1} A_{ij} \qquad\qquad k_i^{out} = \sum_{i=1} A_{ij}$$

The average degree is

$$\mathrm{c} = \frac{1}{n}\sum_{i,j} A_{ij}$$

same for in and out degree

Degree distribution

How many people follow you on Twitter.

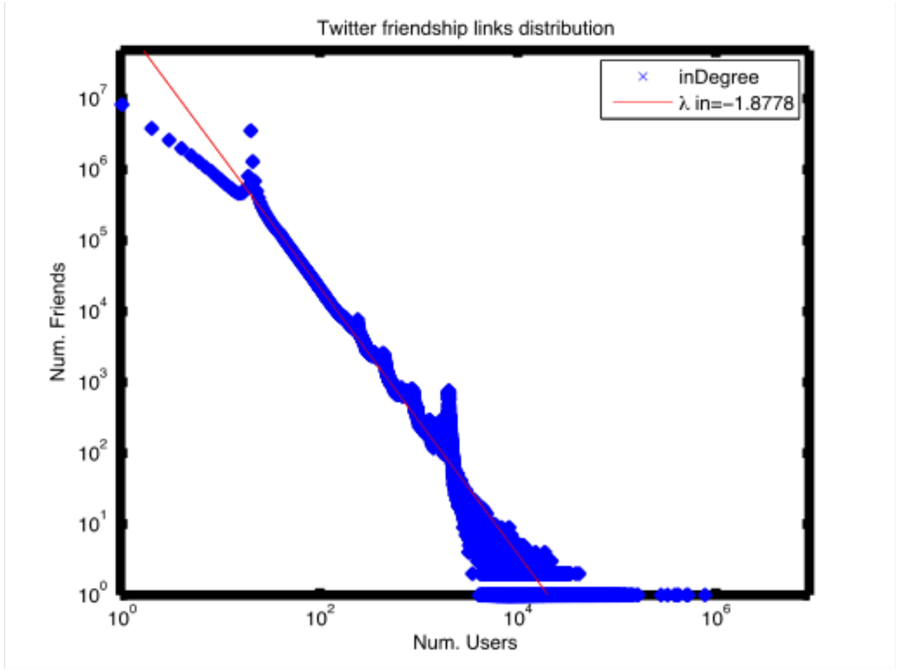


Figure 2. Incoming degree distribution of Twitter's network. As the figure shows, there are a few users with an enormous degree (number of followers). On the contrary, the majority of them have less than 100 followers.

Degree distribution $p(k)$ tells us how the connectedness varies between nodes

Degree distribution

How many people you follow on Twitter.

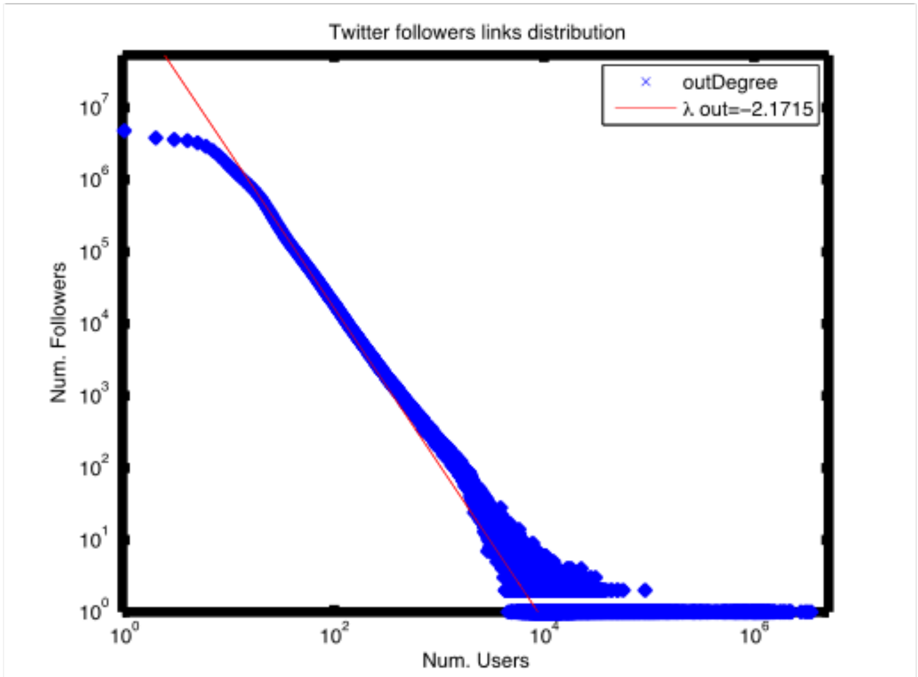
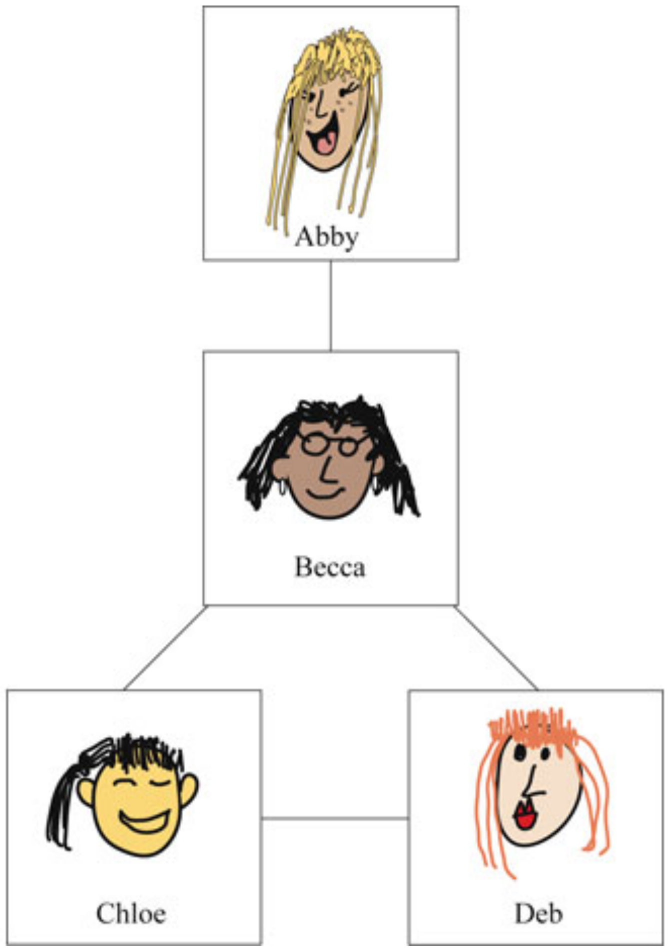


Figure 1. Outgoing degree distribution of Twitter's network. As the figure shows, there are a few users with an enormous degree (number of friends). On the contrary, the majority of them have just at most 1000 friends.

Degree distribution $p(k)$ tells us how the connectedness varies between nodes

Friendship Paradox



https://opinionator.blogs.nytimes.com/2012/09/17/friends-you-can-count-on/

Proceedings of the Seventh International AAAI Conference on Weblogs and Social Media

Friendship Paradox Redux: Your Friends Are More Interesting Than You

Nathan O. Hodas
USC Information Sciences Institute
4676 Admiralty Way
Marina del Rey, CA 90292
nhodas@isi.edu

Farshad Kooti
USC Information Sciences Institute
4676 Admiralty Way
Marina del Rey, CA 90292
kooti@usc.edu

Kristina Lerman
USC Information Sciences Institute
4676 Admiralty Way
Marina del Rey, CA 90292
lerman@isi.edu

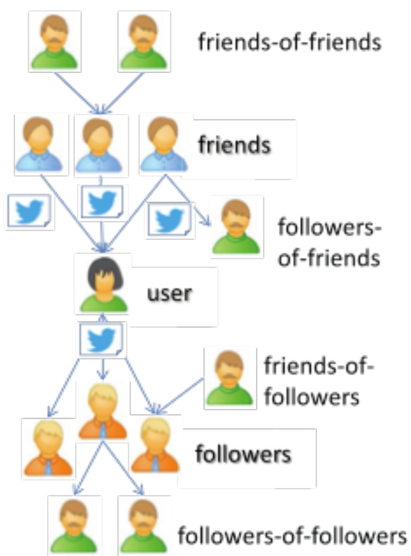
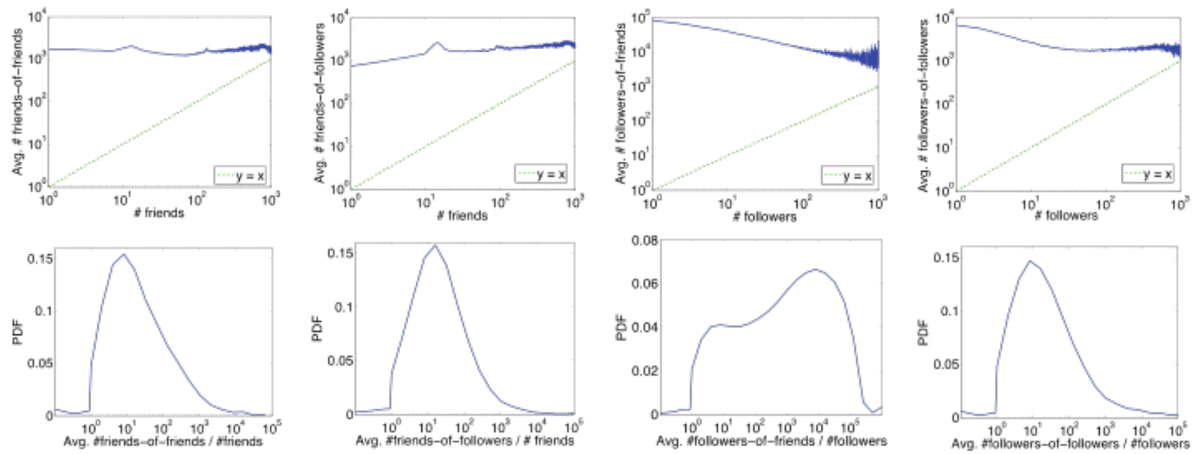


Figure 1: *An example of a directed network of a social media site with information flow links. Users receive information from their friends and broadcast information to their followers.*



(i) (ii) (iii) (iv)

Mean path length

- Find shortest path between all pairs i,j
- The mean path length *l* is the mean of each
- Measures degrees of separation

(**Diameter** = longest path length)

Distance between two random individuals

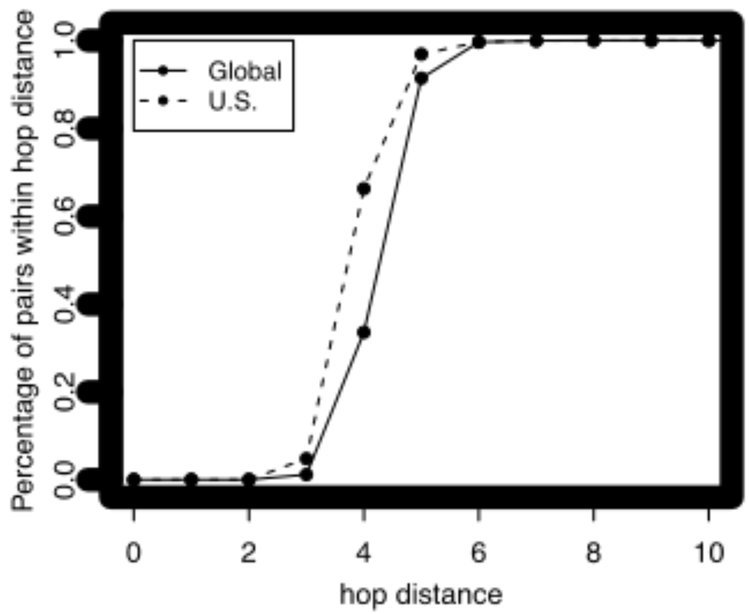
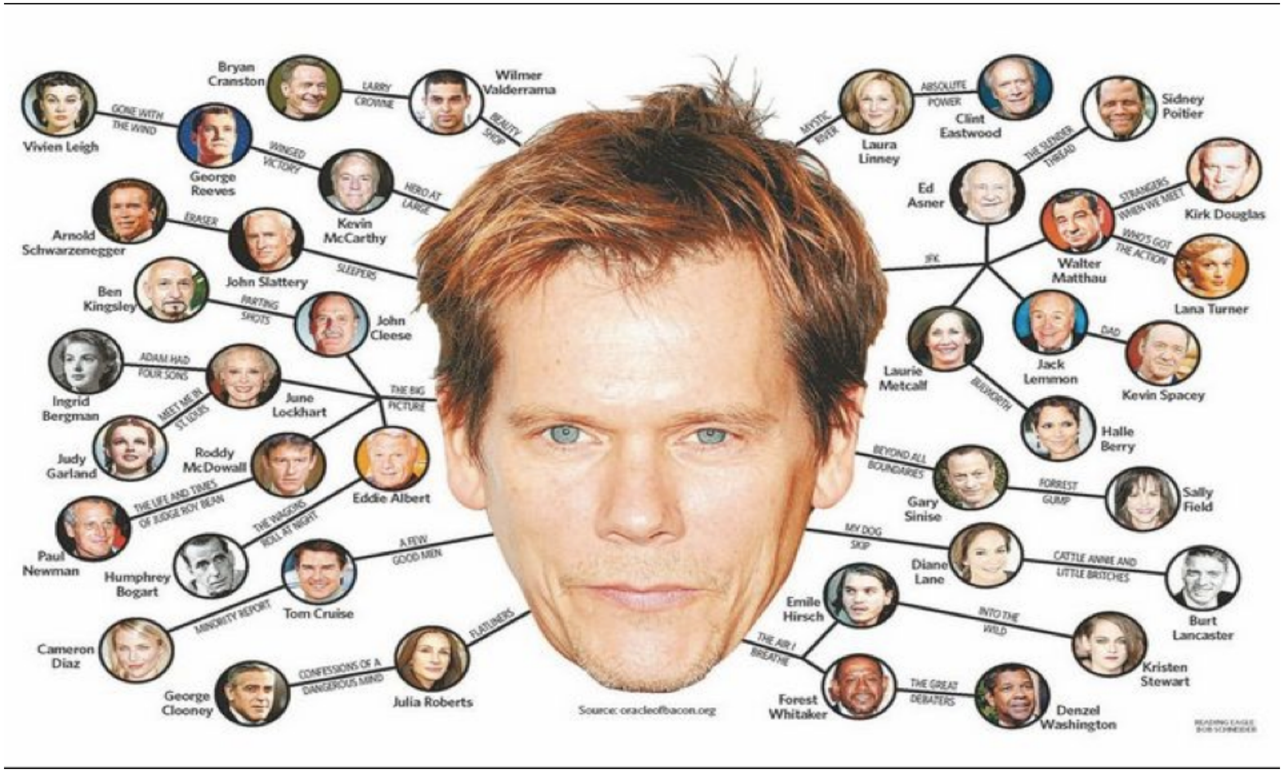
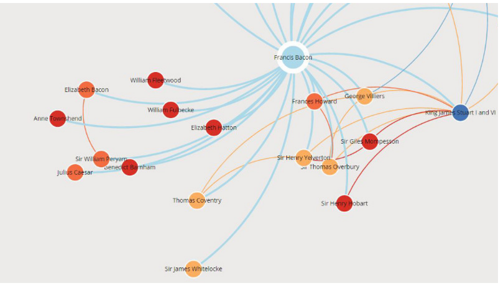


Figure 2. Diameter. The neighborhood function $N(h)$ showing the percentage of user pairs that are within h hops of each other. The average distance between users on Facebook in May 2011 was 4.7, while the average distance within the U.S. at the same time was 4.3.

Mean path length

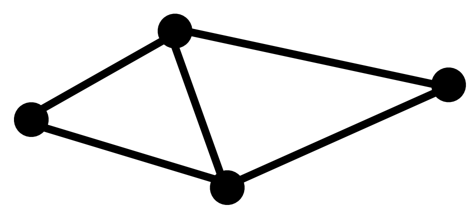




Network Measures

Clustering coefficient

- $$C = \frac{3 \times \text{number of closed triangles}}{\text{number of connected triplets}}$$



C = 6 / 8

= probability that nodes **a** and **b** are connected if both have a common neighbour **c**

- High in social networks. You are friends with your friends friends.

Lattice networks

- All internal nodes have the same degree
- High C (~ constant)
- High mean path length (increases as $n^{1/d}$)

Networks - community partition

Communities of interest

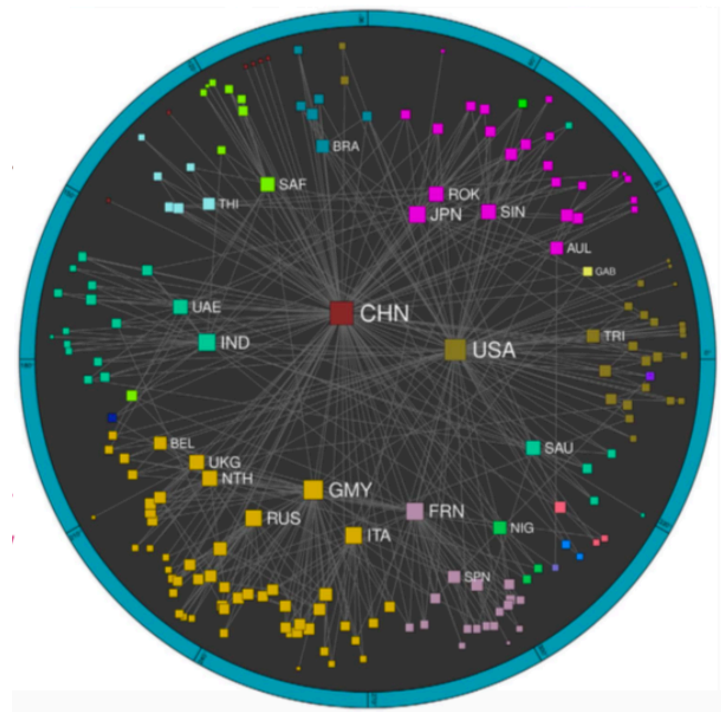
Network: nodes are countries, weight of each link is volume of trade between countries.

Garcia-Pérez 2016

USA, Canada, Bahamas, Haiti, Dominican Republic, Jamaica, Grenada, Mexico, Honduras, Venezuela, Peru

China, North Korea, Gambia, Sierra Leone, Togo, South Sudan

Japan, South Korea, Taiwan, Singapore, Sri Lanka, Philippines, New Zealand, Fiji, Papua New Guinea

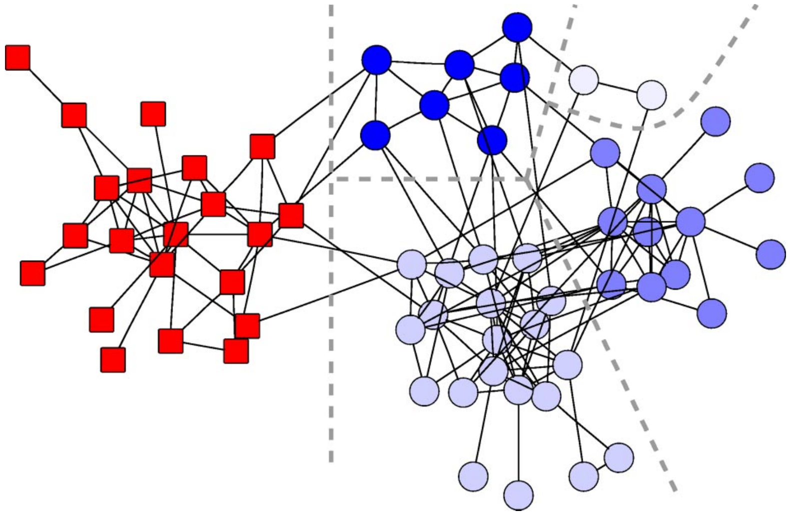


Garcia-Pérez, 2016

Networks - community partition

Communities of interest

Network: dolphins of doubtful sound, NZ, links between dolphins ‘often’ seen together.



Lusseau PhD Thesis,
Newman & Girvan, Finding and evaluating community structure in networks, *Phys Rev E*, 2004

Networks - community partition

As stepping stone: - analyse use of language in climate change debate

Network: links between blogs on climate change

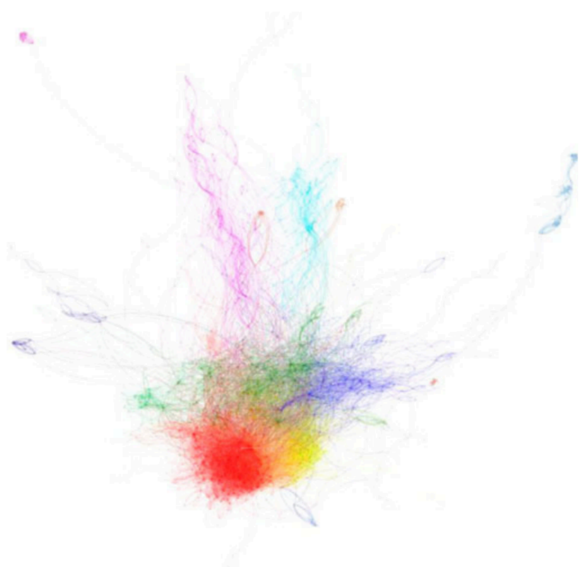


Figure 1. The network of climate change blogs, colored by community.

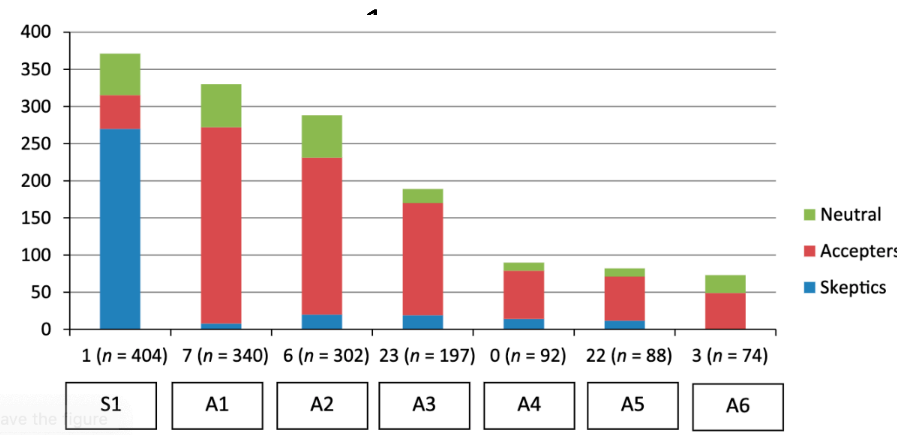


Figure 3. The distribution of skeptical, accepting, and neutral blogs in the seven largest among the central groups of blogs concerned with climate change.

Elgesem, Steskal & Diakopoulos 2015

Networks - community partition

As stepping stone: - analyse use of language in climate change debate

Network: links between blogs on climate change

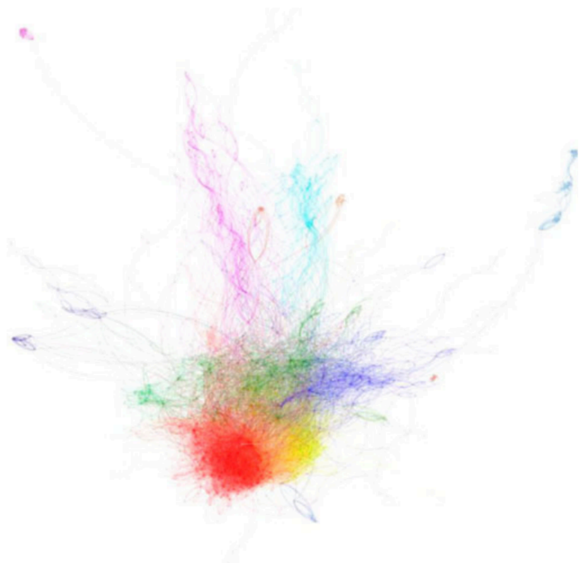


Figure 1. The network of climate change blogs, colored by community.

Table 5. The top 15 collocates around "climate" in communities 1 (skeptic), 23 (accepter), and 7 (accepter) computed with the point-wise mutual information metric.

Top collocates of "CLIMATE" in the skeptical community S1	Top collocates of "CLIMATE" in the accepter community A3	Top collocates of "CLIMATE" in the accepter community A1
1 CLIMATE	1 DENIERS	1 POPPIN
2 SKEPTICS	2 SKEPTICS	2 DENIERS
3 ALARMISM	3 CLIMAT	3 SKEPTICS
4 DENIERS	4 DECADAL	4 OBAMA
5 IPCC	5 CONTRARIANS	5 WWW
6 DECADAL	6 OBAMA	6 EU'S
7 ALARMISTS	7 NOAA'S	7 CLIMATE
8 CLIMAT	8 AGW	8 YVO
9 CHANGE	9 WWW	9 NOAA'S
10 INTERGOVERNMENTAL	10 DENIER	10 WILDFIRES
11 OBAMA	11 CLIMATE	11 CHANGE'S
12 ANTHROPOGENIC	12 VAPOR	12 IPCC
13 AGW	13 ANTHROPOGENIC	13 ALARMISM
14 IPCC'S	14 ALARMISM	14 PACHAURI
15 WARMING	15 CONTRARIAN	15 DENIER

Reference corpus: The British National Corpus, approximately 100 million words.

Elgesem, Steskal & Diakopoulos 2015

Mathematics of community partitions

Define a score! "Modularity"

$$q^*(G) = \max_{\mathcal{A}} q_{\mathcal{A}}(G) = \sum_{A \in \mathcal{A}} \frac{e(A)}{m} - \frac{\mathrm{vol}(A)^2}{4m^2}$$

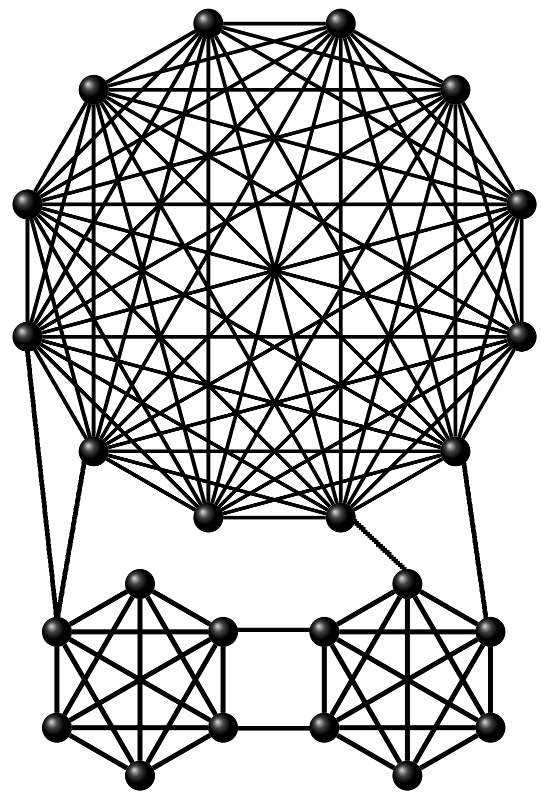
Edge contribution/Coverage

$$q^E_{\mathcal{A}}(G) = \sum_{A \in \mathcal{A}} \frac{e(A)}{m}$$

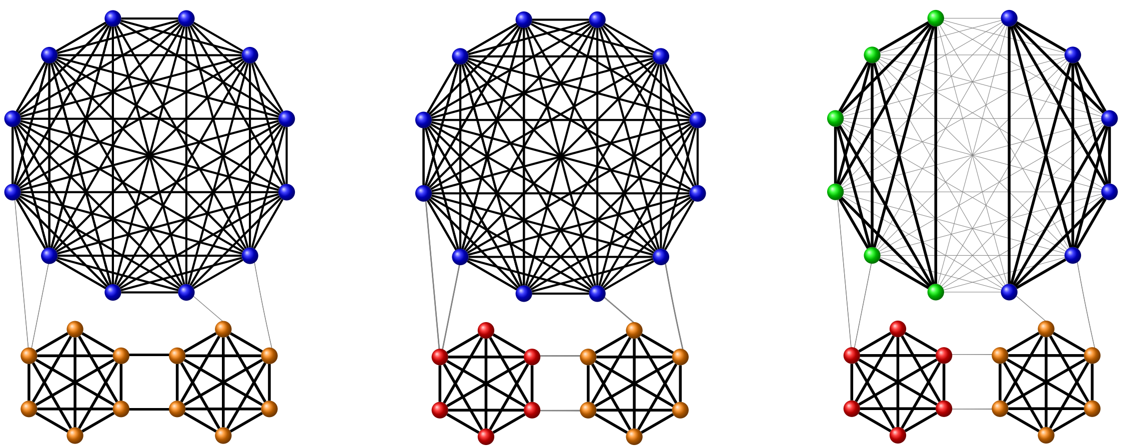
Degree tax

$$q^D_{\mathcal{A}}(G) = \sum_{A \in \mathcal{A}} \frac{\mathrm{vol}(A)^2}{4m^2}$$

Networks - community partition



Networks - community partition



$q^E_{\mathcal{A}_1} = 0.96, \quad q^D_{\mathcal{A}_1} = 0.56$

$q_{\mathcal{A}_1} = 0.40$

$q^E_{\mathcal{A}_2} = 0.94, \quad q^D_{\mathcal{A}_2} = 0.50$

$q_{\mathcal{A}_2} = 0.44$

$q^E_{\mathcal{A}_3} = 0.59, \quad q^D_{\mathcal{A}_3} = 0.29$

$q_{\mathcal{A}_3} = 0.30$

Modelling Networks with (random) graphs

- Lattice graphs
- Erdos-Renyi random graph
- Chung-Lu random graph
- Configuration model
- Preferential attachment model
- KPKVB model - random hyperbolic graph

- KPKVB model - random hyperbolic graph
  - Krioukov-Papadopoulos-Kitsak-Vahdat-Boguñá
  - Power law degree distribution
  - Clustering coefficient

Müller and Fountoulakis, Law of large numbers for the largest component in a hyperbolic model of complex networks, 2018

KPKVB model - random hyperbolic graph

- Hyperbolic plane curvature -alpha^2

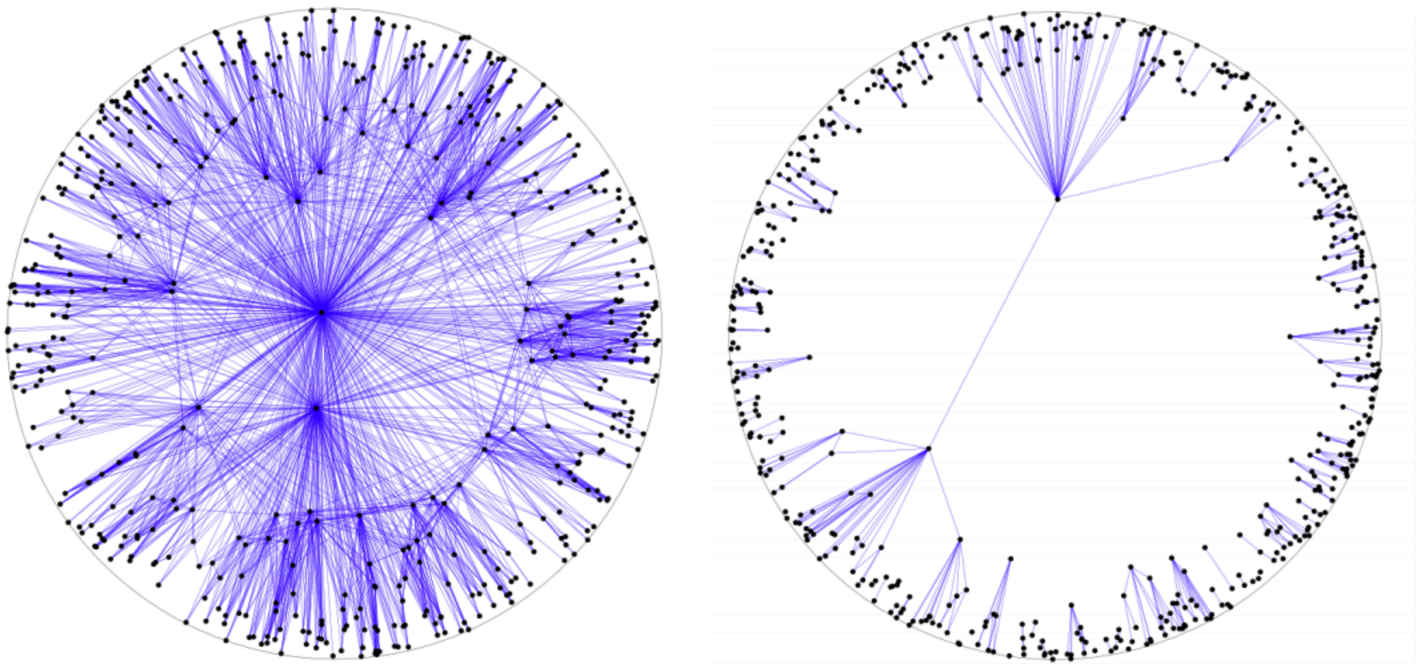


Figure 1: The random graph $G(N; \alpha, \nu)$ with $N = 500$ vertices, $\nu = 2$ and $\alpha = 0.7$ and $3/2$.

Müller and Fountoulakis, Law of large numbers for the largest component in a hyperbolic model of complex networks, 2018

Random graph

Every pair of nodes i,j is connected with probability q. *Total of* n nodes

- Binomial degree distribution, $c = q(n-1)$
- Low C = c/n
- Low mean path length l ~ $\log(n)$

Real networks

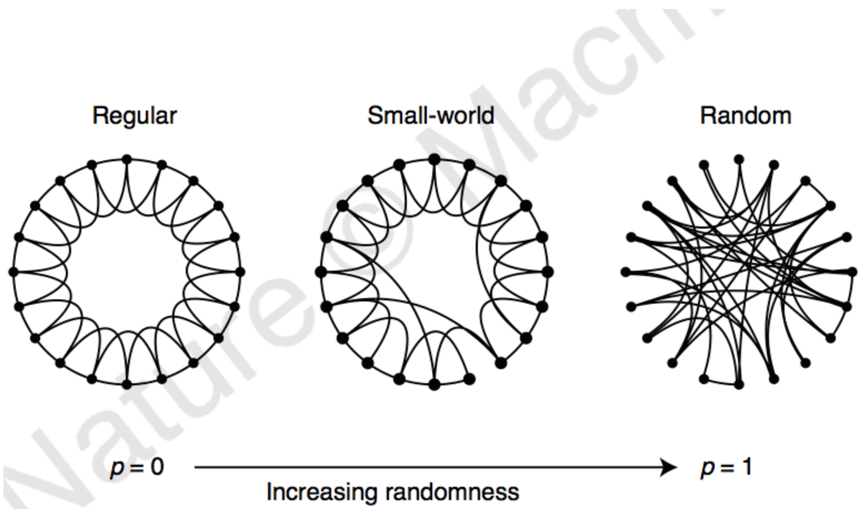
TABLE I. The general characteristics of several real networks. For each network we have indicated the number of nodes, the average degree $\langle k \rangle$, the average path length ℓ, and the clustering coefficient C. For a comparison we have included the average path length ℓ_{rand} and clustering coefficient C_{rand} of a random graph of the same size and average degree. The numbers in the last column are keyed to the symbols in Figs. 8 and 9.

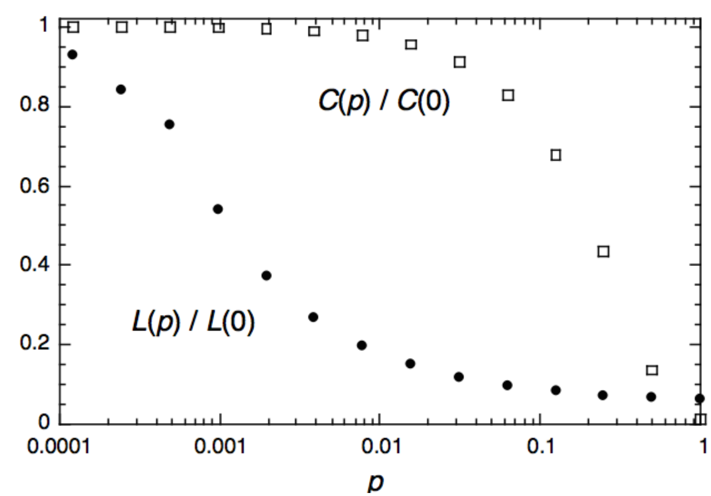
Network	Size	$\langle k \rangle$	ℓ	ℓ_{rand}	C	C_{rand}	Reference	Nr.
WWW, site level, undir.	153 127	35.21	3.1	3.35	0.1078	0.00023	Adamic, 1999	1
Internet, domain level	3015–6209	3.52–4.11	3.7–3.76	6.36–6.18	0.18–0.3	0.001	Yook *et al.*, 2001a, Pastor-Satorras *et al.*, 2001	2
Movie actors	225 226	61	3.65	2.99	0.79	0.00027	Watts and Strogatz, 1998	3
LANL co-authorship	52 909	9.7	5.9	4.79	0.43	1.8×10^{-4}	Newman, 2001a, 2001b, 2001c	4
MEDLINE co-authorship	1 520 251	18.1	4.6	4.91	0.066	1.1×10^{-5}	Newman, 2001a, 2001b, 2001c	5
SPIRES co-authorship	56 627	173	4.0	2.12	0.726	0.003	Newman, 2001a, 2001b, 2001c	6
NCSTRL co-authorship	11 994	3.59	9.7	7.34	0.496	3×10^{-4}	Newman, 2001a, 2001b, 2001c	7
Math. co-authorship	70 975	3.9	9.5	8.2	0.59	5.4×10^{-5}	Barabási *et al.*, 2001	8
Neurosci. co-authorship	209 293	11.5	6	5.01	0.76	5.5×10^{-5}	Barabási *et al.*, 2001	9
E. coli, substrate graph	282	7.35	2.9	3.04	0.32	0.026	Wagner and Fell, 2000	10
E. coli, reaction graph	315	28.3	2.62	1.98	0.59	0.09	Wagner and Fell, 2000	11
Ythan estuary food web	134	8.7	2.43	2.26	0.22	0.06	Montoya and Solé, 2000	12
Silwood Park food web	154	4.75	3.40	3.23	0.15	0.03	Montoya and Solé, 2000	13
Words, co-occurrence	460.902	70.13	2.67	3.03	0.437	0.0001	Ferrer i Cancho and Solé, 2001	14
Words, synonyms	22 311	13.48	4.5	3.84	0.7	0.0006	Yook *et al.*, 2001b	15
Power grid	4941	2.67	18.7	12.4	0.08	0.005	Watts and Strogatz, 1998	16
C. Elegans	282	14	2.65	2.25	0.28	0.05	Watts and Strogatz, 1998	17

Small world network

- Watts & Strogatz model interpolates between a structured and random network

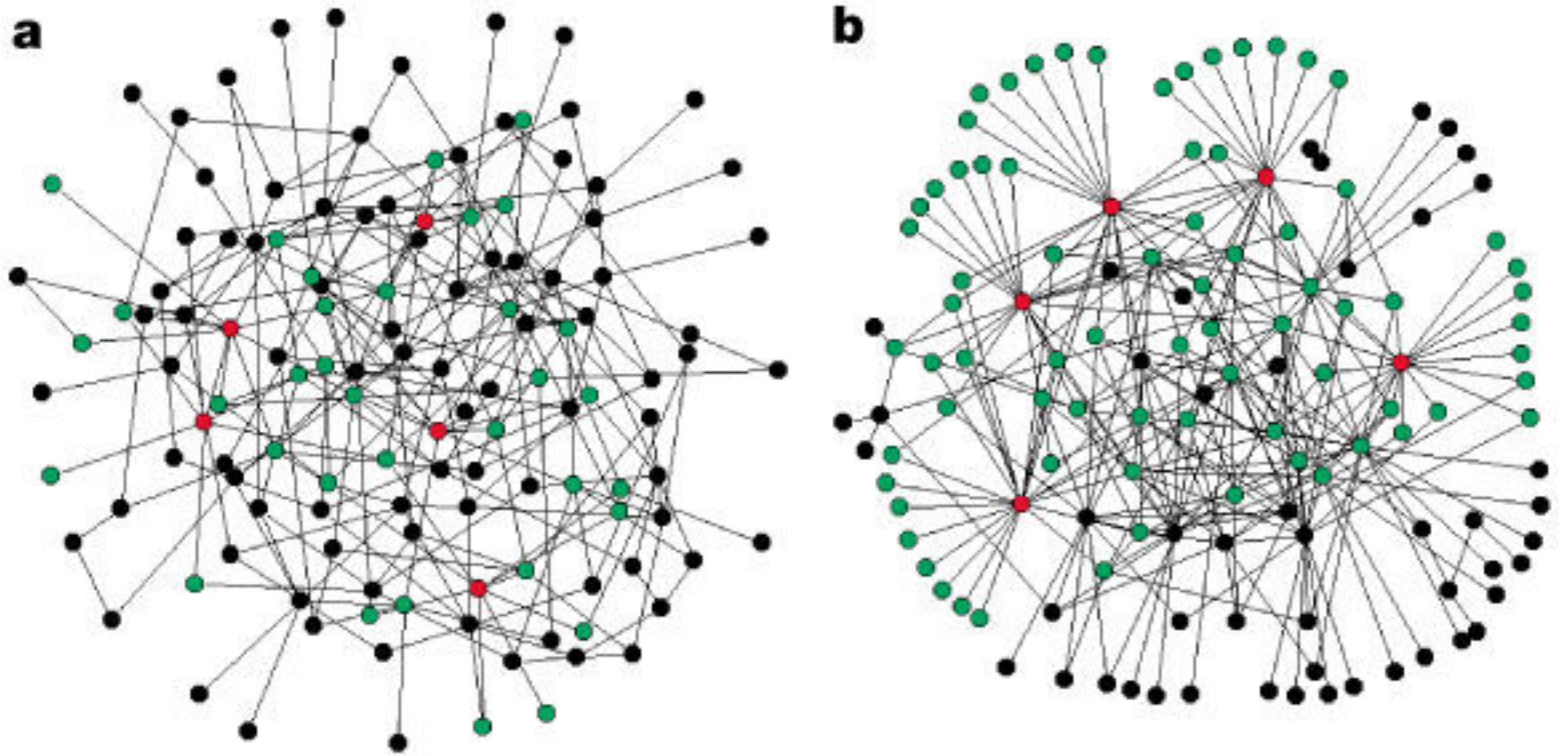
- Low diameter + high clustering = small world





Watts and Strogatz, *Nature* 393 (1998)

Power law network



Exponential

Scale-free

Power law network

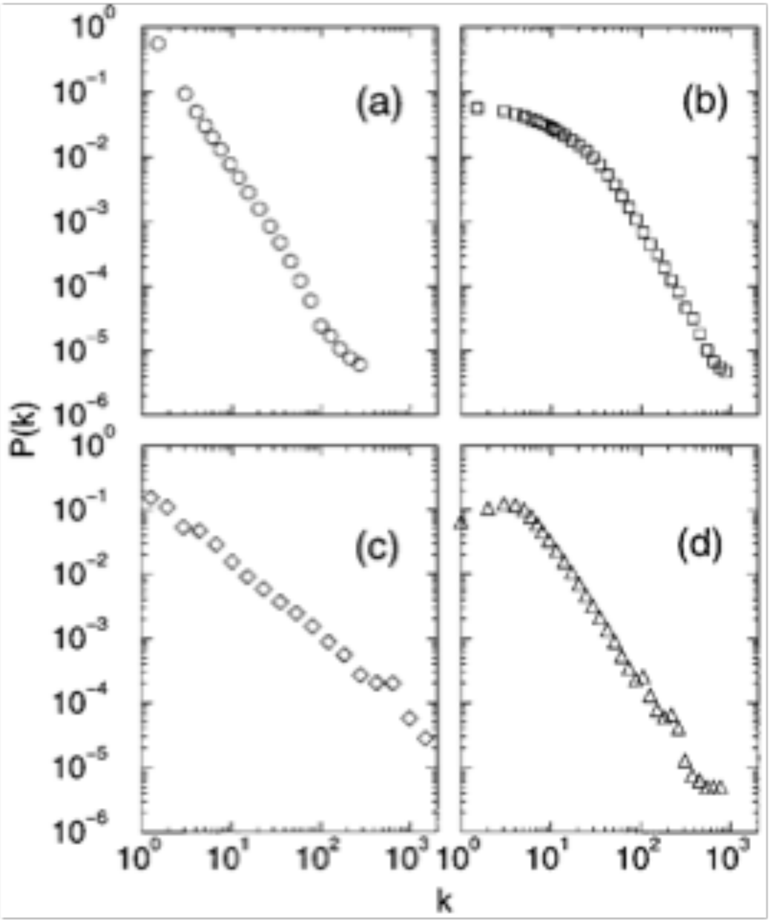
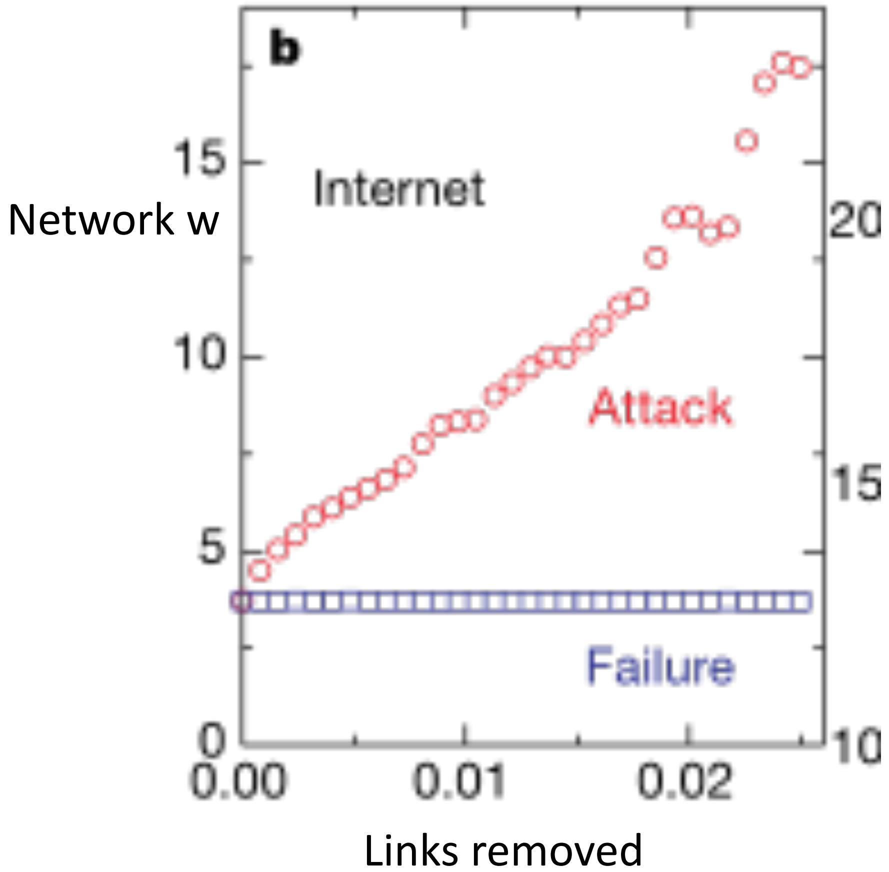


FIG. 3. The degree distribution of several real networks: (a) Internet at the router level. Data courtesy of Ramesh Govindan; (b) movie actor collaboration network. After Barabási and Albert 1999. Note that if TV series are included as well, which aggregate a large number of actors, an exponential cut-off emerges for large *k* (Amaral *et al.*, 2000); (c) co-authorship network of high-energy physicists. After Newman (2001a, 2001b); (d) co-authorship network of neuroscientists. After Barabási *et al.* (2001).

Power law network



- Robust to random failures.
- Susceptible to attack