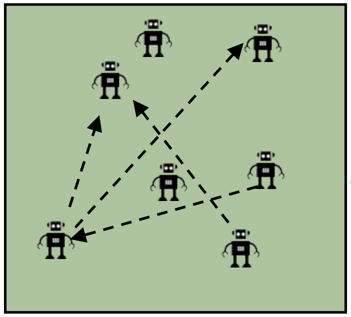
15-382 Collective Intelligence – S18

Lecture 35:
Wisdom of the Crowd
Networks

Instructor:
Gianni A. Di Caro

So far…

- In Game Theory we have considered multi-agent systems with potentially *conflictual utilities*. Solution concept: equilibrium

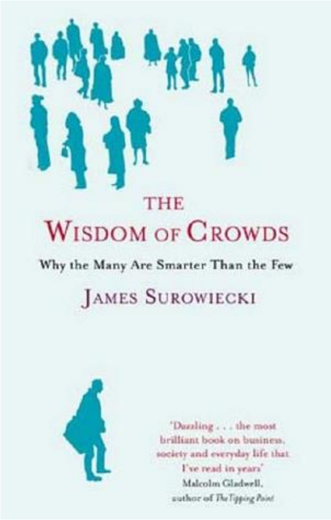


- In PSO and ACO agents do *cooperate online* by continual information sharing (agent-to-agent in PSO, mediated by the environment in ACO)

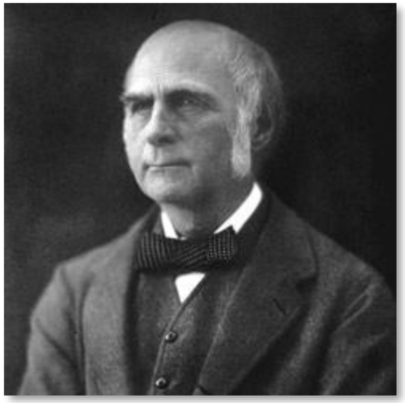
- In Auctioning and Task Allocation, agents can *compete or cooperate*, depending on the context

WISDOM OF THE CROWD

- Let's use a multi-agent system, a *crowd* to solve problems, like making estimates of values, taking decisions, …
- The basic idea is that the **collective opinion of a group of individuals *can* be better than a single expert opinion:**
  - If the individuals in the crowd are experts (or, most of them are), the intuition is quite obvious
  - What if the majority is far from being an expert for the problem domain?
  - A few conditions need to be in place, to make a crowd "wise", otherwise it may fail miserably (e.g., a number of examples from Reddit)

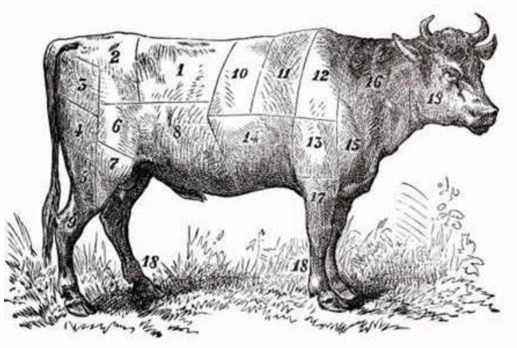


Francis Galton and the Ox weight



- Francis Galton (16 February 1822 – 17 January 1911), cousin of Charles Darwin, was an English Victorian polymath, proto-geneticist, statistician…

- In 1906 Galton visited a livestock fair and stumbled upon an contest. An ox was on display, and the villagers were invited to guess the animal's weight after it was slaughtered and dressed.



- Galton disliked the idea of democracy and wanted to use the competition to show the problems of allowing large groups of people to vote on a topic.

Power of aggregating information

- 787 people guessed the weight of the ox, some were experts, farmers and butchers, others knew little about livestock.
- Some guessed very high, others very low, many guessed fairly sensibly.
- Galton collected the guesses after the competition was over
- The **average** guess from the crowd was **1,197 pounds**
- The **correct weight** was **1,198 pounds**!
- What Dalton discovered was that in actuality crowds of people can make surprisingly good decisions IN THE AGGREATE, even if they have imperfect information
- Many other examples can be found / mentioned …

WHO WANTS TO BE A MILLIONAIRE?

- Compare the lifelines:
  - Phone a friend
  - Ask the Audience
- The correct answer is given:
  - Phone a friend → 65%
  - Ask the Audience → 91%

THE SPACESHUTTLE CHALLENGER

- On January 28, 1986, when the Space Shuttle Challenger broke apart 73 seconds into its flight, leading to the deaths of its seven crew members. The spacecraft disintegrated over the Atlantic Ocean, off the coast of central Florida

- The stock market did not pause to mourn. Within minutes, investors started dumping the stocks of the four major contractors who had participated in the Challenger launch:
  - Rockwell International, which built the shuttle and its main engines;
  - Lockheed, which managed ground support;
  - Martin Marietta, which manufactured the ship's external fuel tank; and
  - Morton Thiokol, which built the solid-fuel booster rocket.

- By the end of the day, Morton Thiokol’s stock was down nearly 12 percent. By contrast, the stocks of the three other firms started to creep back up, and by the end of the day their value had fallen only around 3 percent.

The Spaceshuttle Challenger

- What this means is that the stock market had, almost immediately, labelled Morton Thiokol as the company that was responsible for the *Challenger* disaster.
- Months later it was discovered that it was in fact Morton Thiokol who caused the problem with the production of faulty O-rings.
- How did the stock investors know ?
- A good “explanation” is that, again, this is effect of the the wisdom of crowds.

Google Page Ranking

- How does Google work? (in a ”simple” way)
- How does it classify pages so that typically the page you are looking for is in the first ten links it returns?
- It uses the PageRank algorithm, whose main idea is:
  - The more sites that link to a certain URL with a certain phrase, the higher the rating.
  - This works because each link is a *vote* for the connection between the phrase and the site.
- Again, this can be seen as a form of the wisdom of the crowds

Experts vs. Wisdom of Crowds

- It shows us that groups of people make excellent decisions and can select the correct alternative out of a number of options without any specific expertise ("maybe")
- How could this be?
- One general observation is that *individual experts really aren't as smart as we think*, such that it might be difficult to find the "right" expert when the decision is fairly complex and involves multiple levels of knowledge and abilities
- An interesting experiments in this respect was done by Herbert Simon and W.G. Chase (1973), who explored the nature of expertise in the domain of chess.

EXPERTS ARE NOT KNOW-IT-ALLS

- They showed a chess-board in the middle of a game to an expert chess player and an amateur.
- They asked both to recreate the locations of all of the pieces on another boards, consistency the experts were easily able to reproduce the boards, whereas the amateur rarely could.
- So does this mean experts are smarter?
- No, because when they put the pieces on the board randomly, the expert and amateur both did equally as well.
- This shows how limited might be the scope of expertise.
- We normally assume people who are intelligent at one pursuit are good at all, but in actuality this is not at all the case.
- Chase said the intelligence and expertise is, in fact, "*spectacularly narrow*"

Crowd of Experts / Non-experts

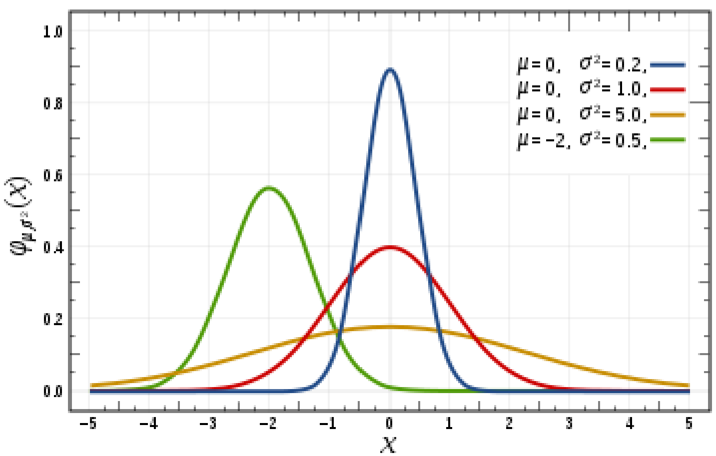
- If a group of multiple experts for the domain is available, it is expected the they collectively provide a better answer than they would do individually

- Value sampling from expert population (the *crowd*), each expert i outputs an estimate s_i, that can be seen as a random variable, and their sample mean c has the same expected value of the population

- If the population is of true experts, the estimates s_i will have (in the limit of large populations) a Gaussian distribution centred at the true value $\bar{s}$, and small variance

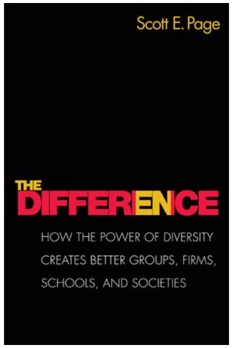
- The less expert the crowd is, the larger the variance

- If the crowd has no expertise at all, there's the risk that estimates will have a wrong bias (e.g., the green distribution)



Diversity Prediction Theorem

- How the crowd issues correct estimates / makes good decisions?
- It's a simple theorem (actually an identity)
- **Diversity Prediction Theorem:**



$$(c-\theta)^2 = \frac{1}{n}\sum_{i=1}^{n}(s_i-\theta)^2 - \frac{1}{n}\sum_{i=1}^{n}(s_i-c)^2$$

c is the crowd estimate, the sample mean of individual estimates s_i

θ is the ground truth

n is the number of individuals in the crowd

Crowd's (quadratic) error = Average (quadratic) error – Crowd diversity

- *Diversity:* spread of estimates / expertise in the crowd

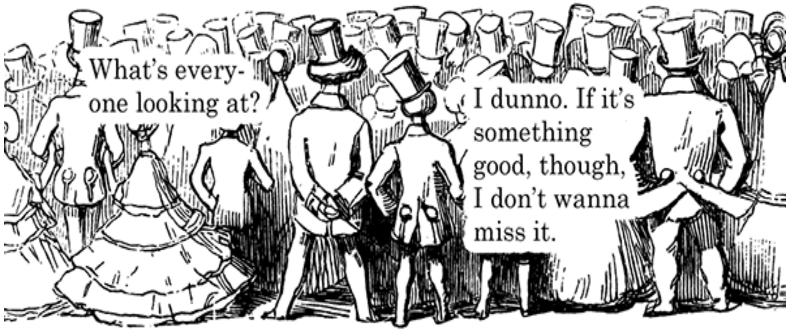
Diversity Prediction Theorem

- [Crowd’s error] = [Average error] – [Diversity]
- How do we get a small Crowd’s error?
  - A crowd of experts: [Average error] is small, [Diversity] will also be small, usually
  - A crowd of non experts: [Average error] will be fairly large, but if we have a balanced large [Diversity], we get a small error, we also need relatively large crowds to make the probabilities work

Diversity Prediction Theorem

- [Crowd's error] = [Average error] – [Diversity]

- When things can go wrong?
  - The non experts are **badly wrong** and have a (wrong) **bias** in their estimates, such that [Diversity] can't counterbalance the [Average error]
  - When the estimates are **not independent**, such that, for instance, a *wrong bias* can be established because of social interactions, driving the crowd to the wrong answer



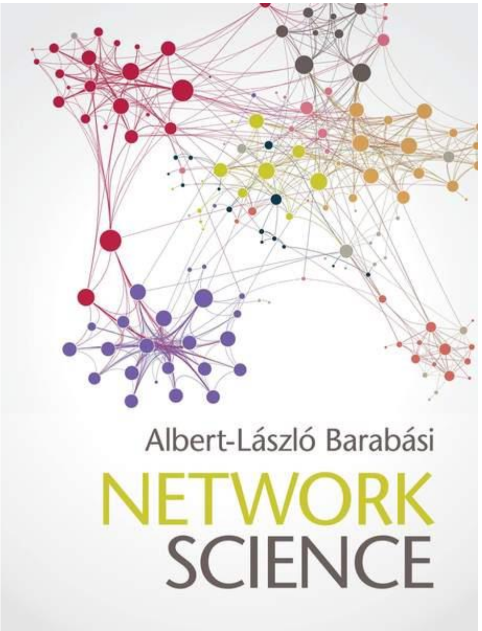
Jan Lorenz, Heiko Rauhut, Frank Schweitzer, Dirk Helbing, *How social influence can undermine the wisdom of crowd effect*. Proceedings of the National Academy of Sciences (PNAS), 108 (22) 9020-9025, May 2011

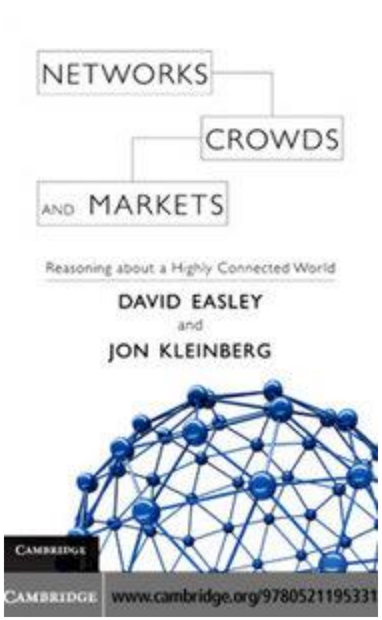
AGENT INTERACTIONS: INTERCONNECTION NETWORK

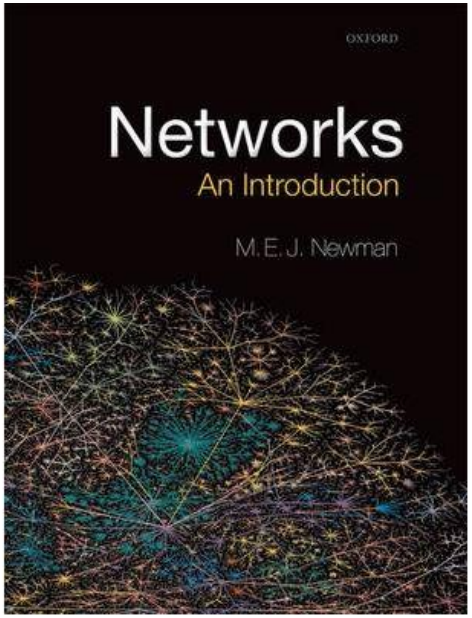
- *Social interactions* → **Network**, information sharing that propagates through a set of *interconnection channels*
- Interconnection Networks strongly affect how in a complex system information propagates, that in turn determines how individuals evolve over time
- How is an interconnection network represented **mathematically**?
- What **properties** do networks have? How are they measured?
- How do we **model** networks to understand their properties? How are real networks different from the ones produced by a simple model?
- What are useful networks for the task at hand?

Recommended readings

- Barabasi, “Network Science”
- Easley & Kleinberg, “Networks, Crowds, and Markets: Reasoning about a Highly Connected World”
- Newman, “Networks”





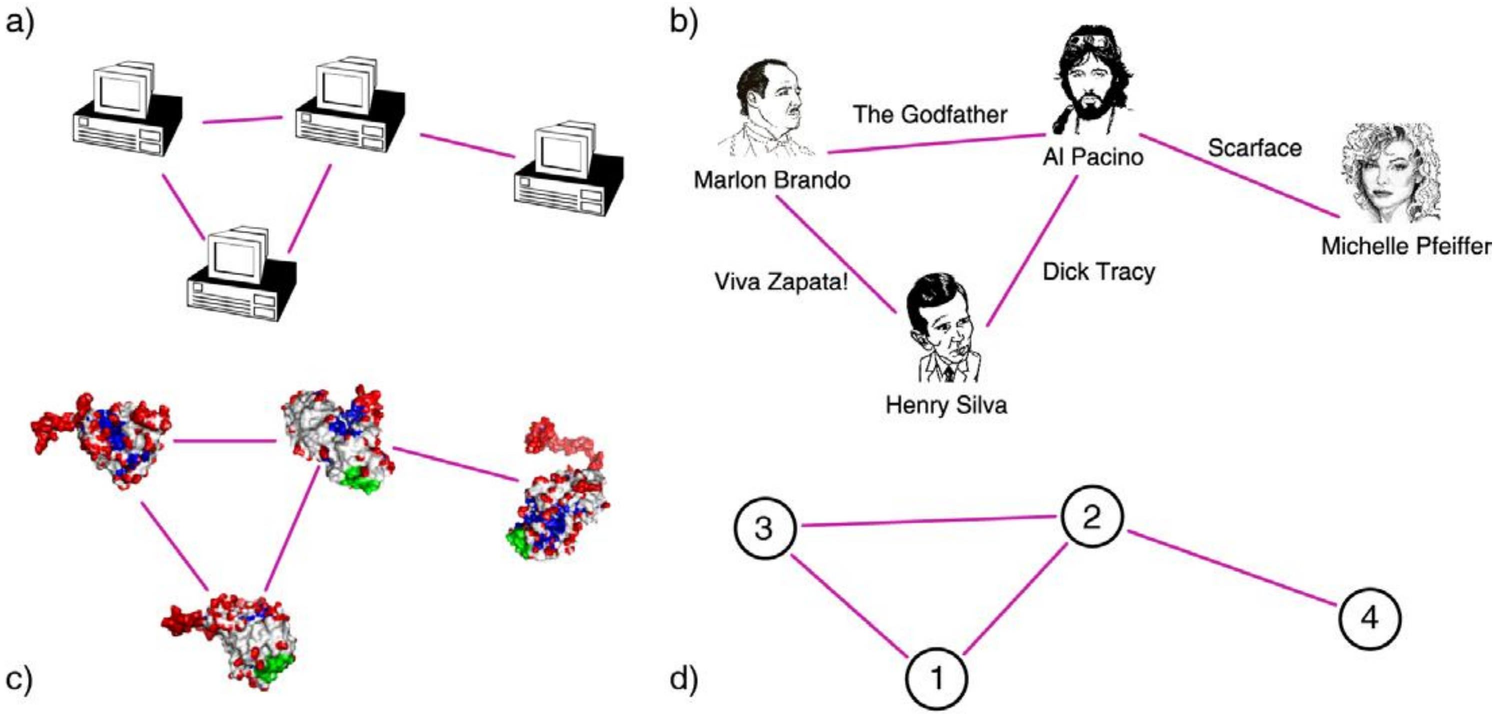


COMPLEX SYSTEMS AS NETWORKS

Many complex systems can be represented as networks

Any complex system has an associated network of communication / interaction among the components

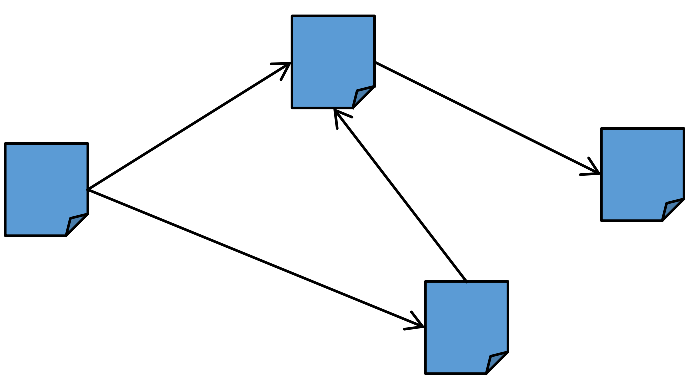
- Nodes = components of the complex system
- Links = interactions between them



Directed vs. Undirected Networks

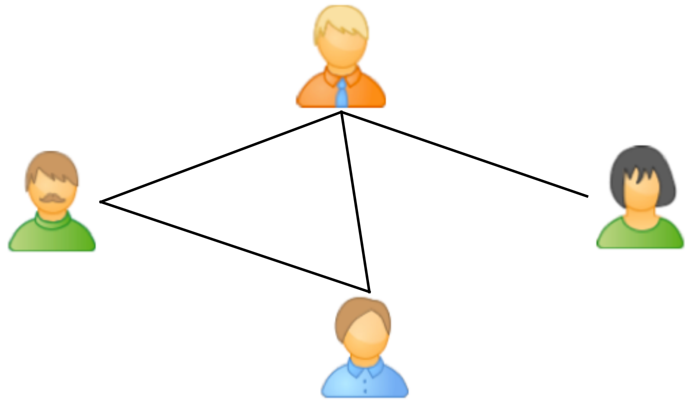
Directed

- Directed links
  - interaction flows one way
- Examples
  - WWW: web pages and hyperlinks
  - Citation networks: scientific papers and citations
  - Twitter follower graph



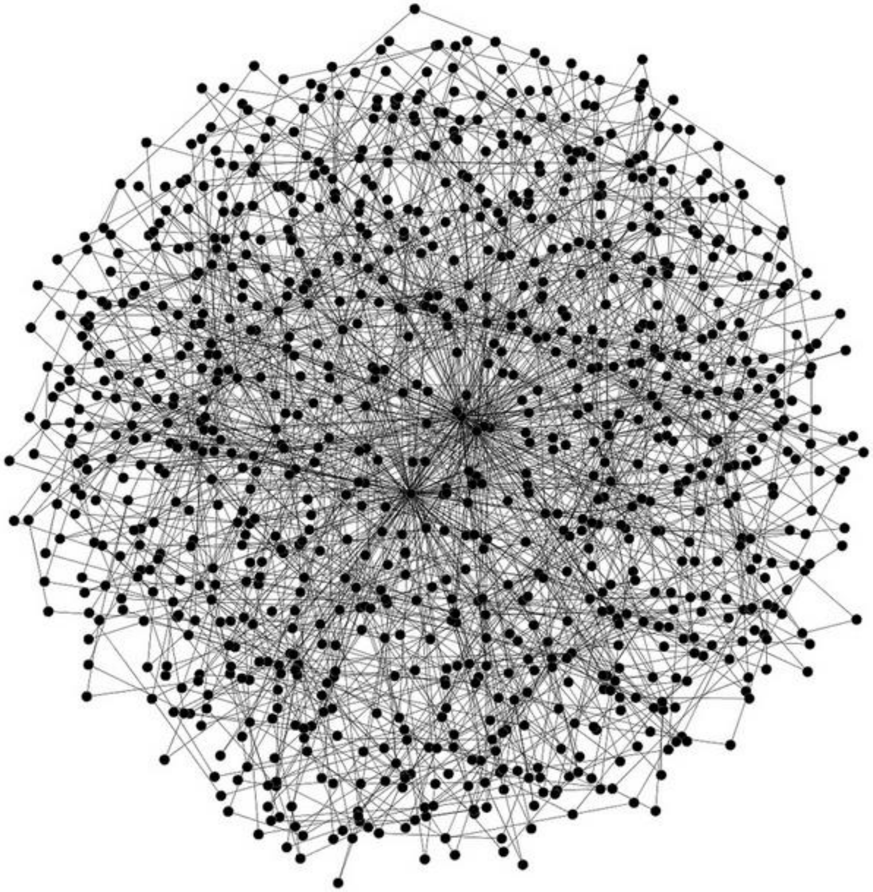
Undirected

- Undirected links
  - Interactions flow both ways
- Examples
  - Social networks: people and friendships
  - Collaboration networks: scientists and co-authored papers

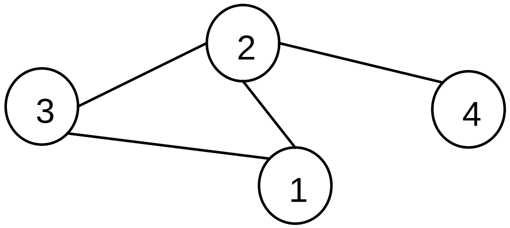


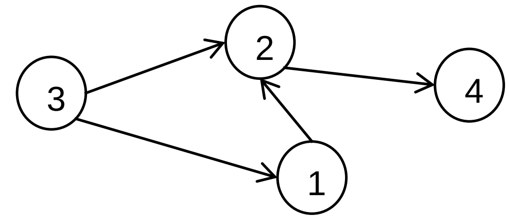
How do we characterize networks?

- Size
  - Number of nodes
  - Number of links
- Degree
  - Average degree
  - Degree distribution
- Diameter
- Clustering coefficient
- …



NODE DEGREE





<u>Undirected networks</u>

- Node degree: number of links to other nodes

$$[k_1 = 2, k_2 = 3,\ k_3 = 2,\ k_4 = 1]$$

- Number of links

$$L = \frac{1}{2}\sum_{i=1}^{N} k_i$$

- Average degree

$$\langle k \rangle = \frac{1}{N}\sum_{i=1}^{N} k_i = \frac{2L}{N}$$

<u>Directed networks</u>

- Indegree:

$$[k_1^{in} = 1\ \ k_2^{in} = 2, k_3^{in} = 0, k_4^{in} = 1]$$

- Outdegree:

$$[k_1^{out} = 1\ \ k_2^{out} = 1, k_3^{out} = 2, k_4^{out} = 0]$$

- Total degree = in + out
- Number of links

$$L = \sum_{i=1}^{N} k_i^{in} = \sum_{i=1}^{N} k_i^{out}$$

- Average degree: L/N

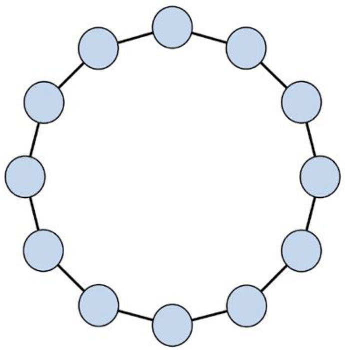
DEGREE DISTRIBUTION

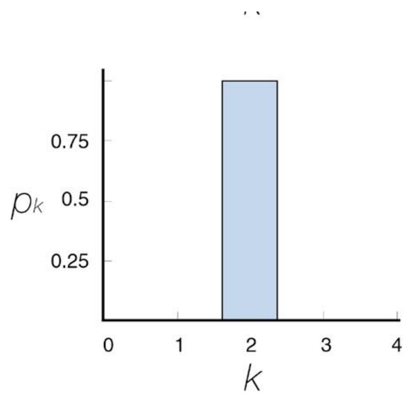
- Degree distribution p_k is the probability that a randomly selected node has degree k

$$p_k = N_k/N$$

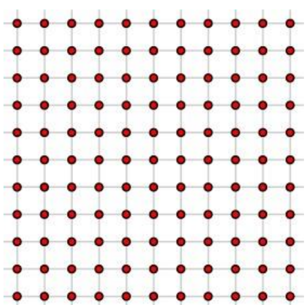
  - Where N_k is number of nodes of degree k

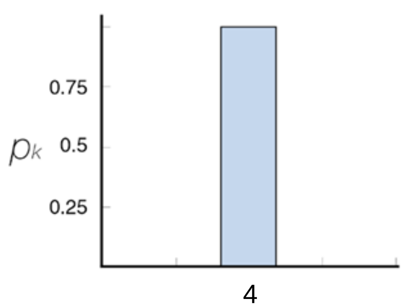
regular lattice



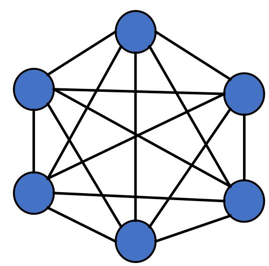


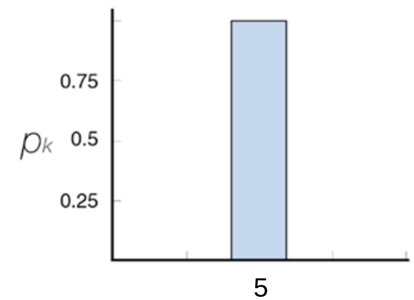
regular lattice



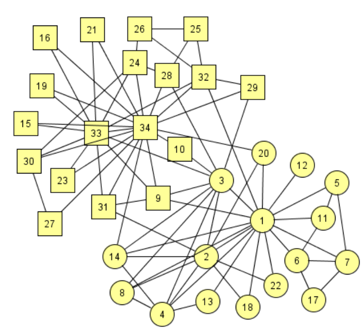


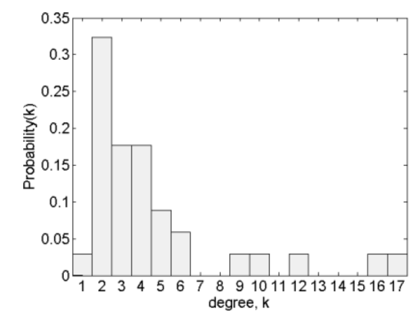
clique (fully connected graph)





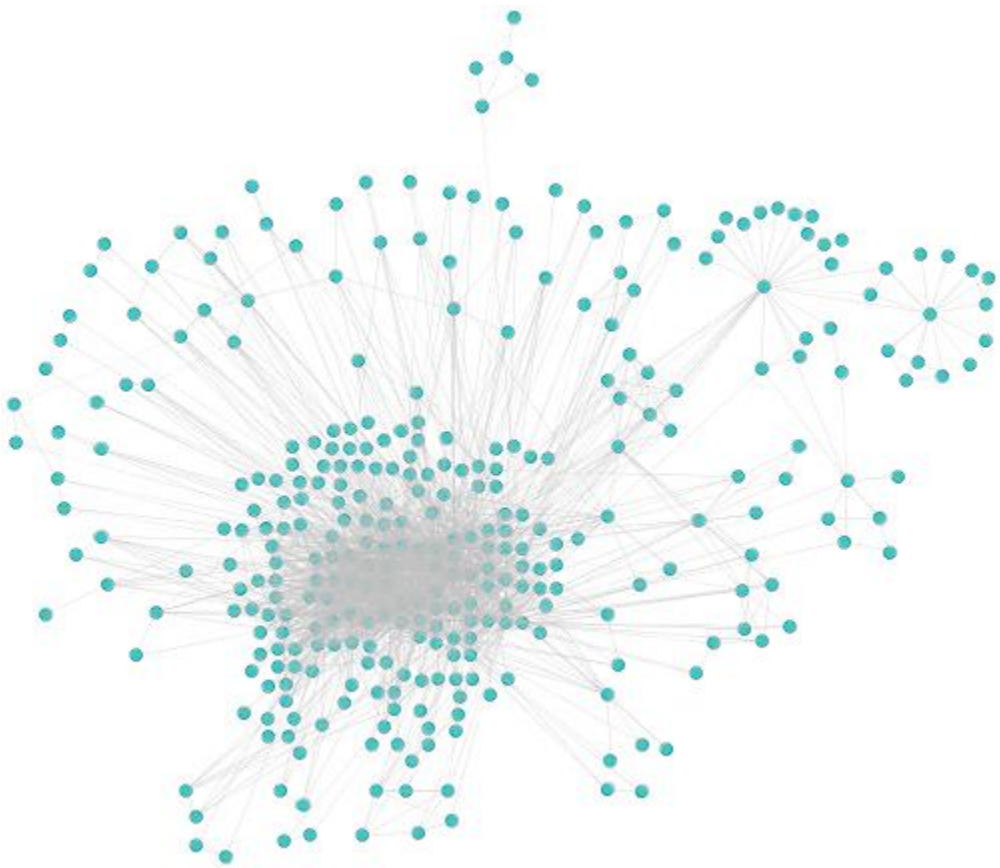
karate club friendship network

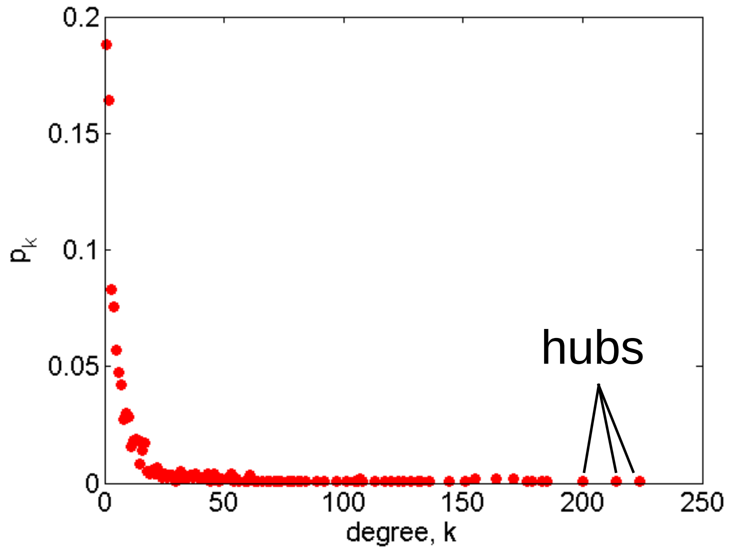


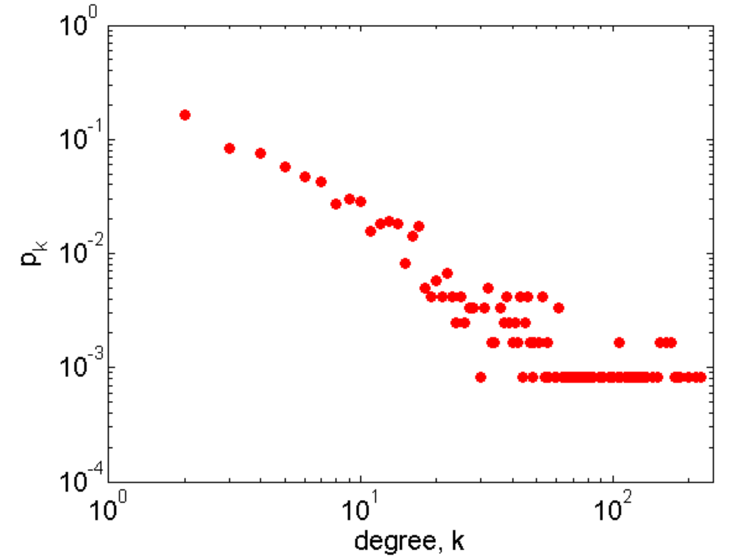


DEGREE DISTRIBUTION IN REAL NETWORKS

Degree distribution of real-world networks is highly heterogeneous, i.e., it can vary significantly

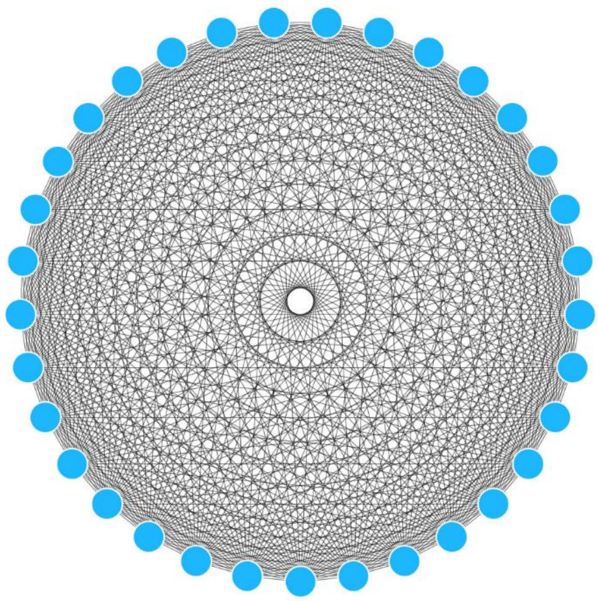




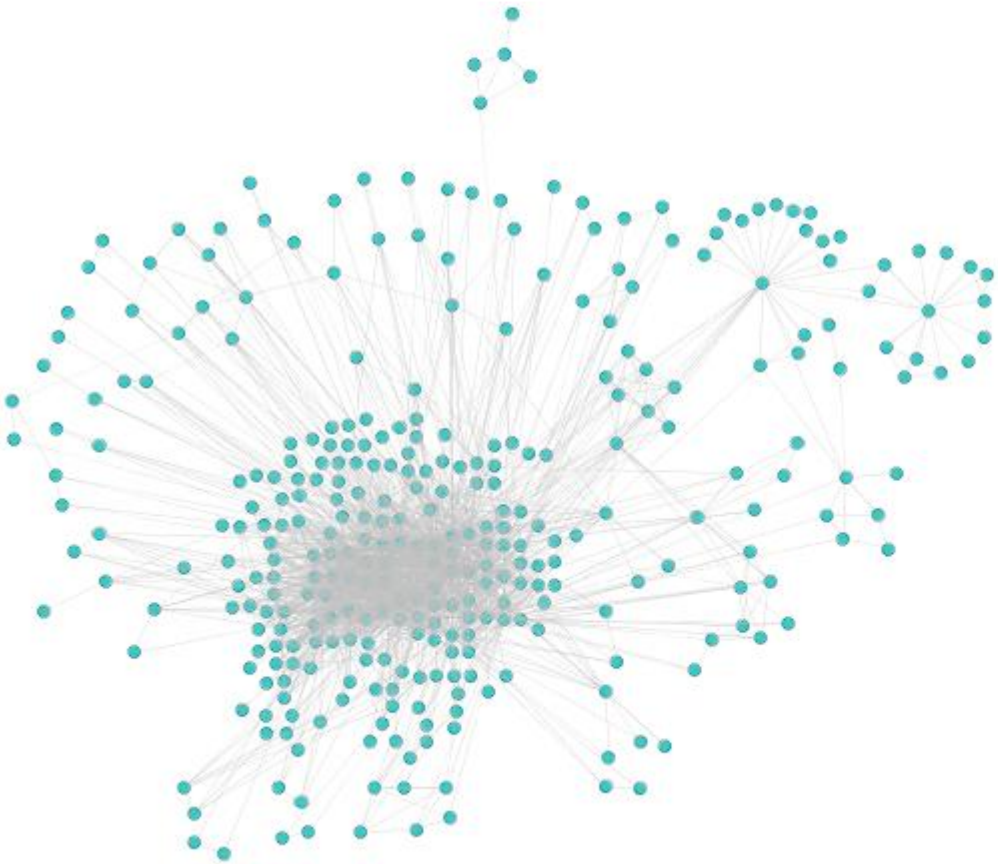


Real networks are sparse

- Complete graph



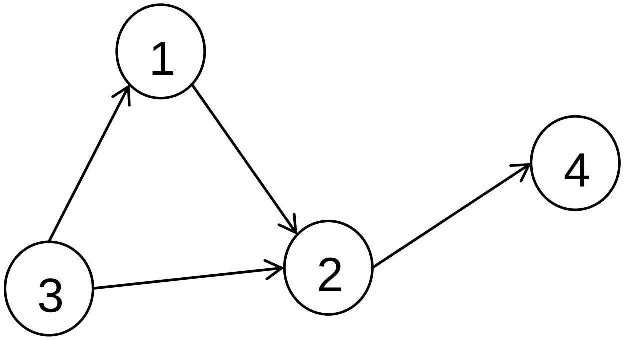
- Real network
  $L \ll N(N-1)/2$



MATHEMATICAL REPRESENTATION OF DIRECTED GRAPHS

- Adjacency list
  - List of links

    [(1,2), (2,4), (3,1), (3,2)]



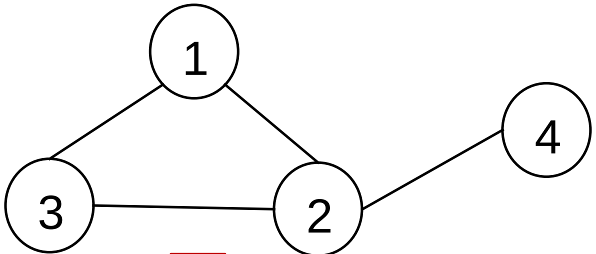
- Adjacency matrix

  $N \times N$ matrix $\boldsymbol{A}$ such that

  - $A_{ij} = 1$ if link (i, j) exists
  - $A_{ij} = 0$ if there is no link

$$A_{ij} = \begin{array}{cc} & \begin{matrix} j & & & \end{matrix} \\ \begin{matrix} i \\ \\ \\ \end{matrix} & \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} \end{array}$$

UNDIRECTED VS. DIRECTED GRAPHS

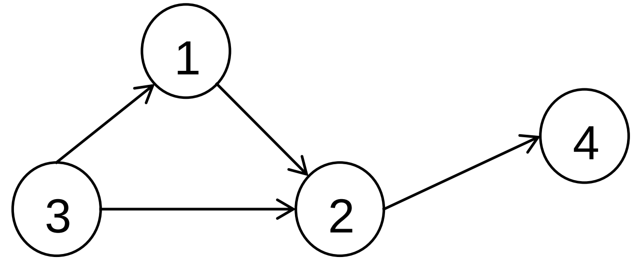


$$A_{ij} = \begin{bmatrix} 0 & 1 & 1 & 0 \\ 1 & 0 & 1 & 1 \\ 1 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix}$$

$$k_2 = \sum_{j=1}^{4} A_{2j} = \sum_{i=1}^{4} A_{i2} = 3$$

Symmetric $\quad A_{ij} = A_{ji}$

$$L = \frac{1}{2} \sum_{i,j=1}^{N} A_{ij}$$



$$A_{ij} = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

$$k_2^{\,in} = \sum_{j=1}^{4} A_{2j} = 2 \qquad k_2^{\,out} = \sum_{i=1}^{4} A_{i2} = 1$$

$$L = \sum_{i,j=1}^{N} A_{ij}$$

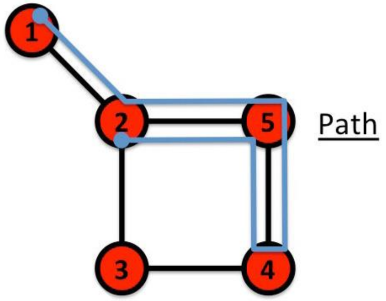
PATHS AND DISTANCES IN NETWORKS

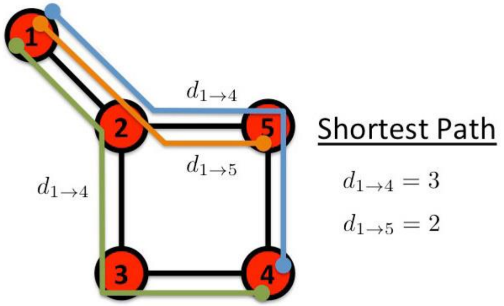
- <u>Path:</u> sequence of links (or nodes) from one node to another

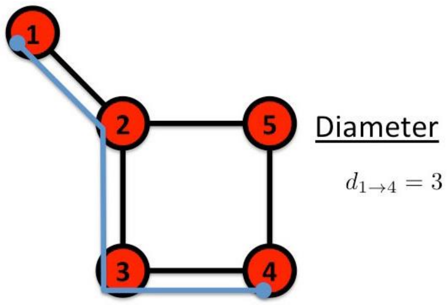
- <u>Walk:</u> a Path of length n from one node to another, that can include repeated nodes / links (e.g., [1-2-1])

- <u>Shortest Path:</u> path with the shortest distance between two nodes

- <u>Diameter:</u> Shortest paths between most distant nodes







Computing paths/distances

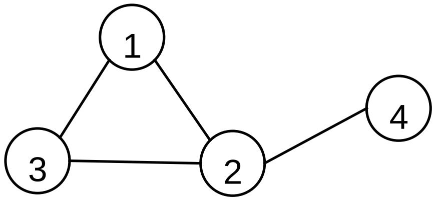
Number of *walks* N_{ij} between nodes i and j can be calculated using the adjacency matrix

- A_{ij} gives paths of length $d = 1$
- $(A^2)_{ij}$ gives #walks of length $d = 2$
- $\left(A^l\right)_{ij}$ gives #walks of length $d = l$

- The minimum l such that $\left(A^l\right)_{ij} > 0$ gives the <u>distance</u> (in hops) between i and j

$A_{ij} =$

0	1	1	0
1	0	1	1
1	1	0	0
0	1	0	0



$(A^2)_{ij} =$

2	1	1	1
1	3	1	0
1	1	2	1
1	0	1	1

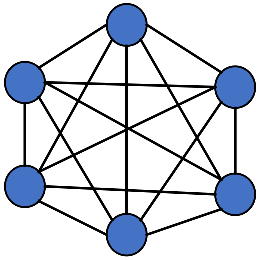
$(A^3)_{ij} =$

2	4	3	1
4	2	4	3
3	4	2	1
1	3	1	0

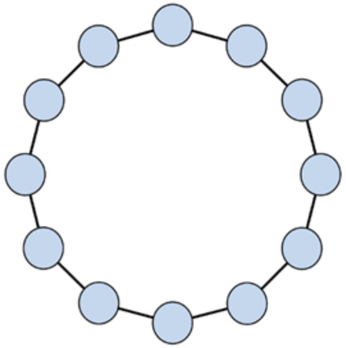
- $A_{ij} = (a_{ij})$
- $(A^2)_{34} = a_{31}a_{14} + a_{32}a_{24} + a_{33}a_{34} + a_{34}a_{44} + a_{35}a_{54} + a_{36}a_{64}$
- $a_{34}a_{44}$ is the # of walks from 3 to 1 multiplied by the # of walks from 1 to 4 → # of walks from 3 to 4 through 1
- $a_{3k}a_{k4}$ is the # of walks from 3 to k multiplied by the # of walks from k to 4 → # of walks from 3 to 4 through k
- Sum of all two-steps walks between 3 and 4

AVERAGE DISTANCE IN NETWORKS

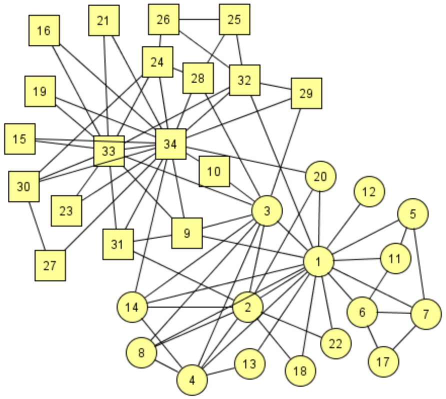
clique: d=1

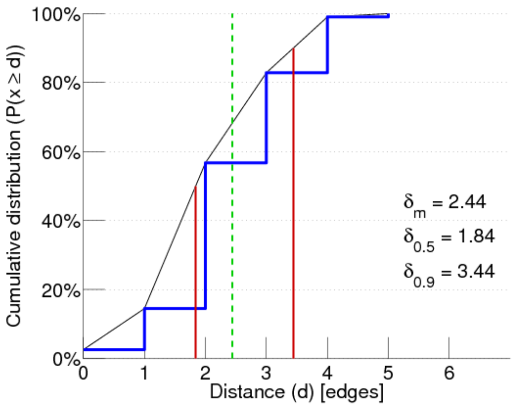


regular lattice (ring): d~N

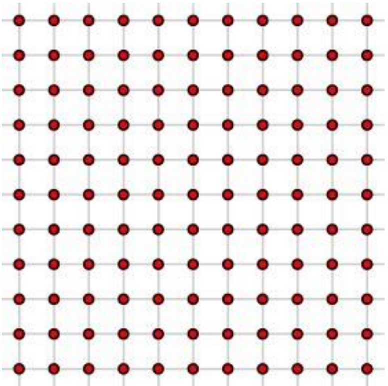


karate club friendship network: d=2.44





regular lattice (square): d~N$^{1/2}$



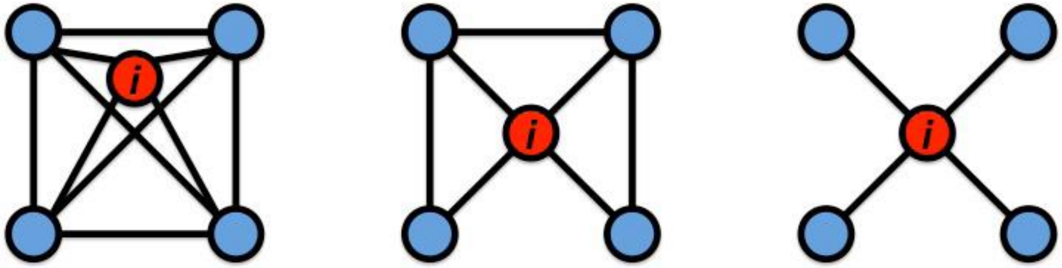
CLUSTERING

- **Clustering coefficient** captures the probability of neighbors of a given node i to be linked
- **Local clustering coefficient** of a vertex i in a graph quantifies how close its neighbors are to being a *clique*

$$C_i = \frac{2L_i}{k_i(k_i - 1)}$$

L_i is number of links between neighbors of I

In a clique: $\frac{k_i(k_i-1)}{2}$ links for k_i neighbors



$C_i = 1$ $\quad$ $C_i = 1/2$ $\quad$ $C_i = 0$

$C = 1$ $\quad$ $C = 9/14$ $\quad$ $C = 0$

Properties of real world networks

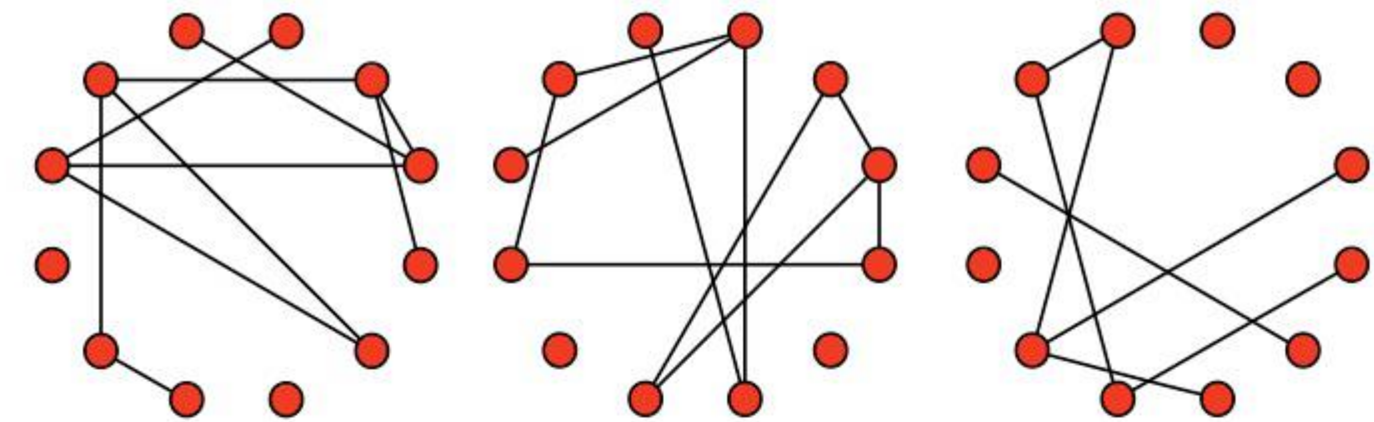
- Real networks are fundamentally different from what we'd expect
  - *Degree distribution*
    - Real networks are <u>scale-free</u>
  - *Average distance between nodes*
    - Real networks are <u>small world</u>
  - *Clustering*
    - Real networks are <u>locally dense</u>
- What do we expect?
  - Create a model of a network. Useful for calculating network properties and thinking about networks.

Random network model

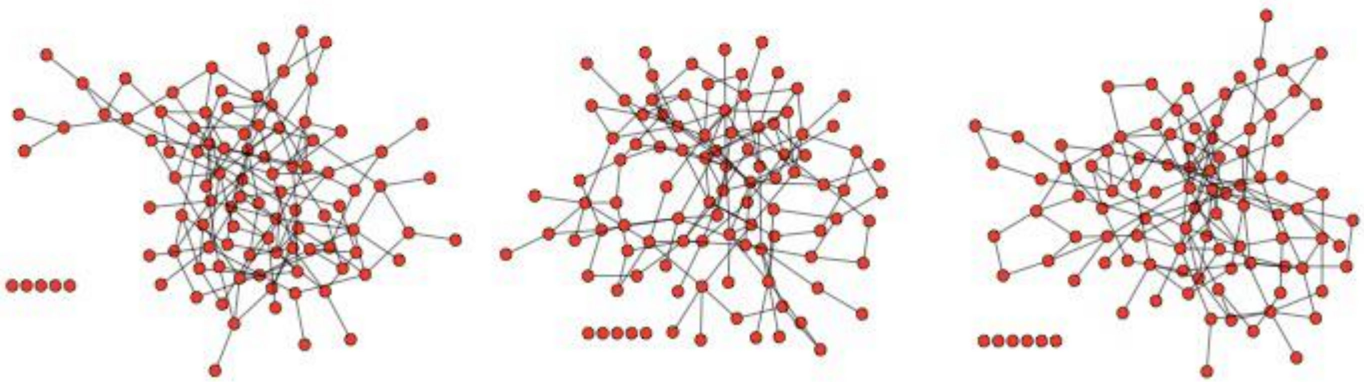
- Networks do not have a regular structure
- Given N nodes, how can we link them in a way that reproduces the observed complexity of real networks?
- Let connect nodes at random!
- *Erdos-Renyi model of a random network*
  - Given N isolated nodes
  - Select a pair of nodes. Pick a random number between 0 and 1. If the number $> p$, create a link
  - Repeat previous step for each remaining node pair
  - Average degree: $\langle k \rangle = p(N-1)$
- Easy to compute properties of random networks

RANDOM NETWORKS ARE TRULY RANDOM

N=12, p=1/6



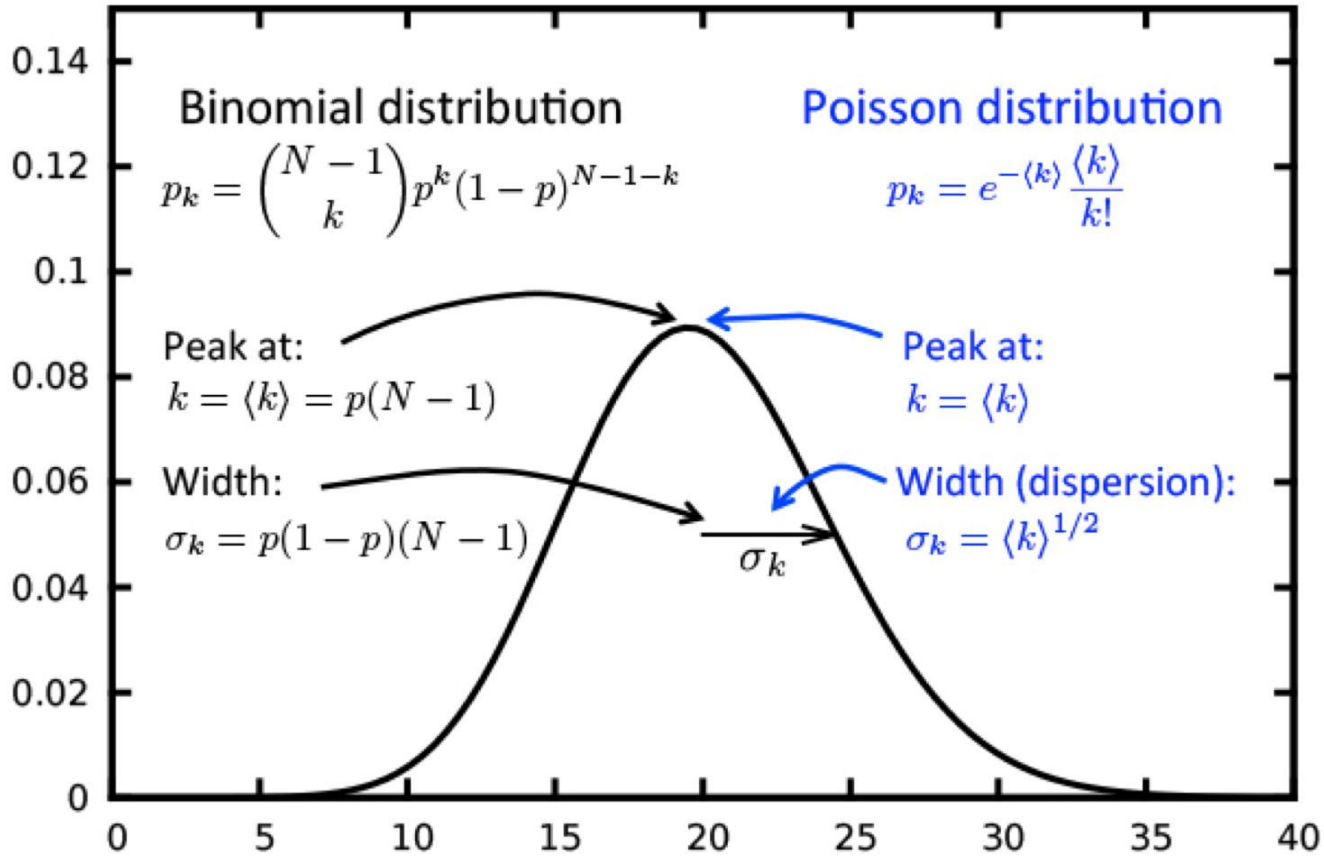
N=100, p=1/6



Average degree: $\langle k \rangle = p(N-1)$

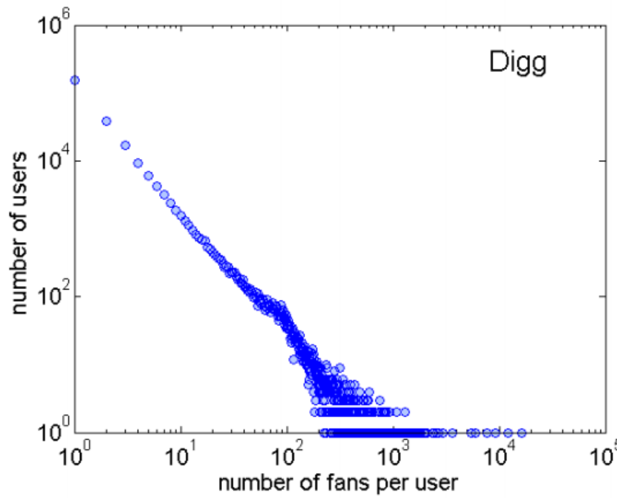
DEGREE DISTRIBUTION IN RANDOM NETWORK

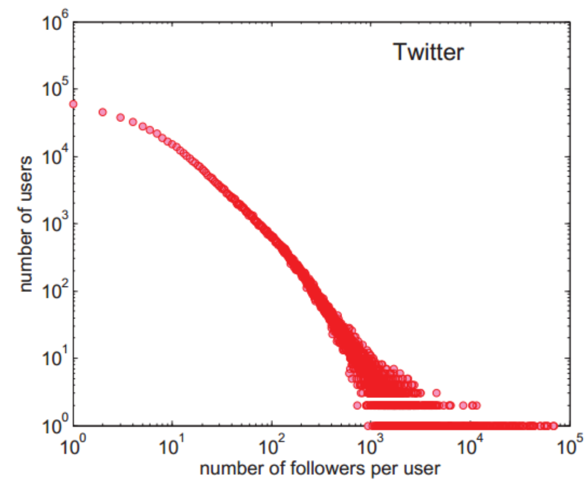
- Follows a binomial distribution
- For sparse networks, <k> << N, Poisson distribution.
  - Depends only on <k>, not network size N



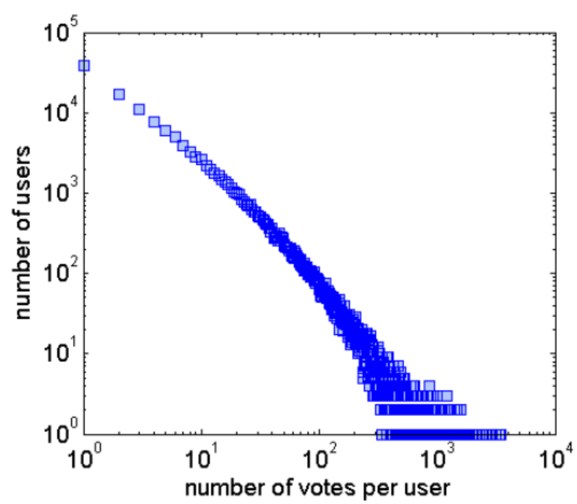
REAL NETWORKS DO NOT HAVE POISSON DEGREE DISTRIBUTION

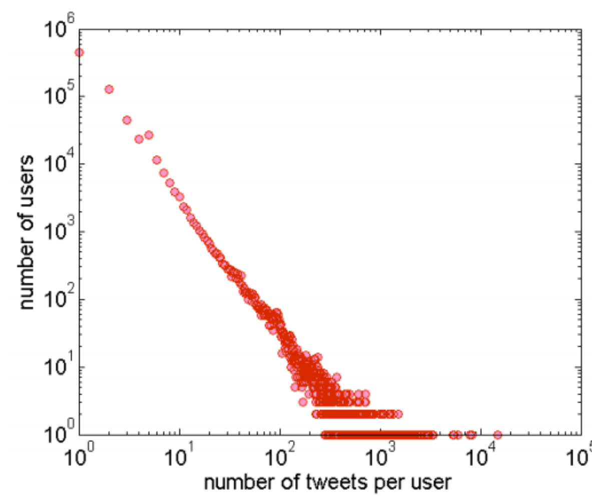
degree (followers) distribution





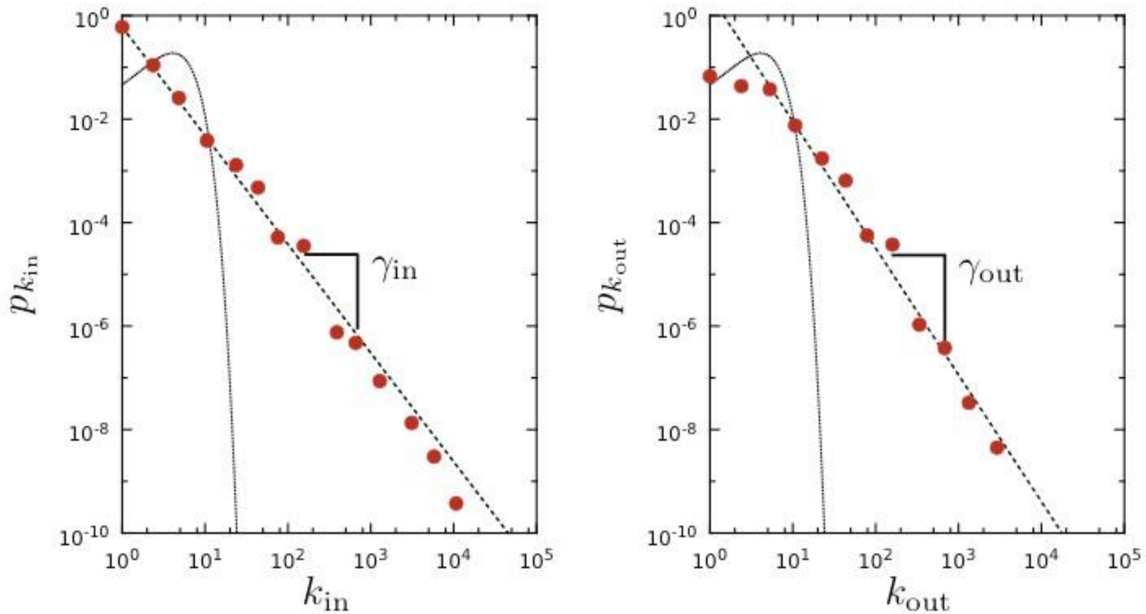
activity (num posts) distribution





Scale free property

WWW hyperlinks distribution

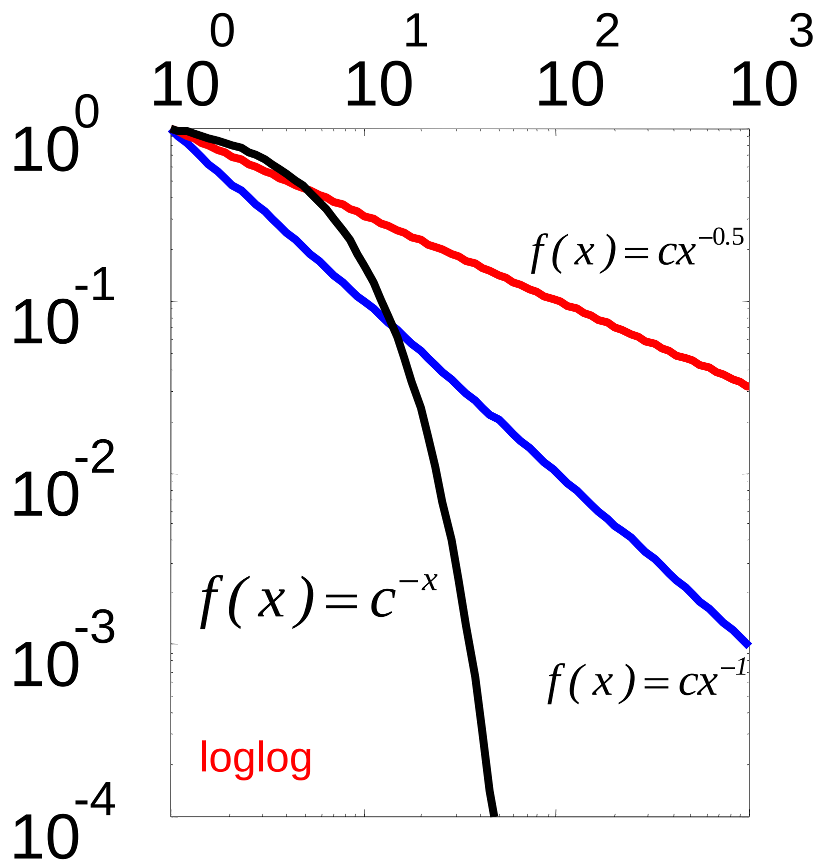


Power-law distribution: $\boldsymbol{p_k \sim k^{-\gamma}}$

- Networks whose degree distribution follows a power-law distribution are called <u>**scale free**</u> networks
- Real network have hubs

Random vs scale-free networks

Random networks and scale-free networks are very different. Differences are apparent when degree distribution is plotted on log scale.



The Milgram experiment

- In 1960's, Stanley Milgram asked 160 randomly selected people in Kansas and Nebraska to deliver a letter to a stock broker in Boston.
  - Rule: can only forward the letter to a friend who is more likely to know the target person
- How many steps would it take?

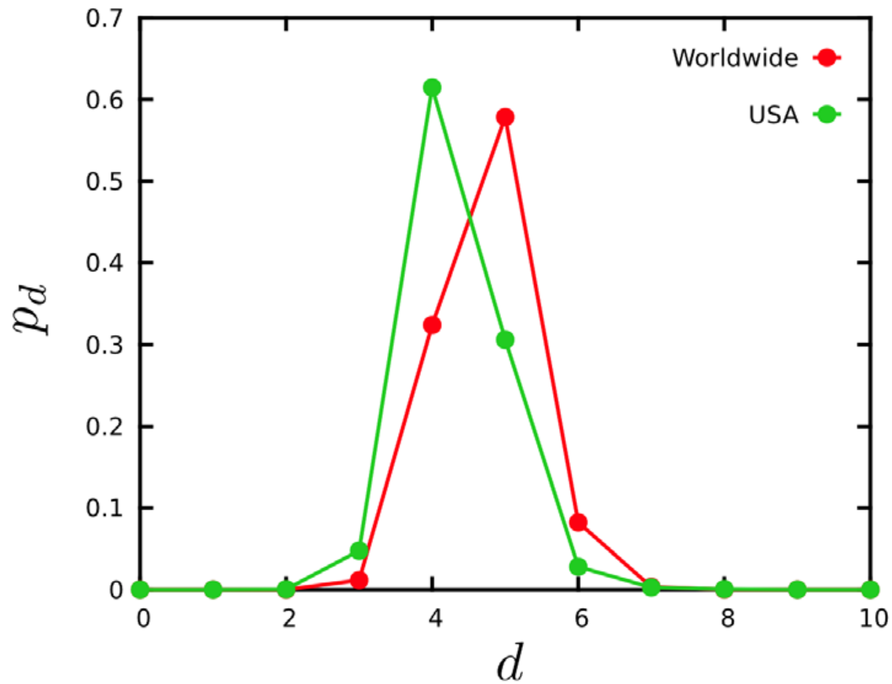
The Milgram experiment

- Within a few days the first letter arrived, passing through only two links.
- Eventually 42 of the 160 letters made it to the target, some requiring close to a dozen intermediates.
- The median number of steps in completed chains was 5.5

→"**six degrees of separation**"

FACEBOOK IS A VERY SMALL WORLD

- Ugander et al. directly measured distances between nodes in the Facebook social graph (May 2011)
  - 721 million active users
  - 68 billion symmetric friendship links
  - the average distance between the users was 4.74

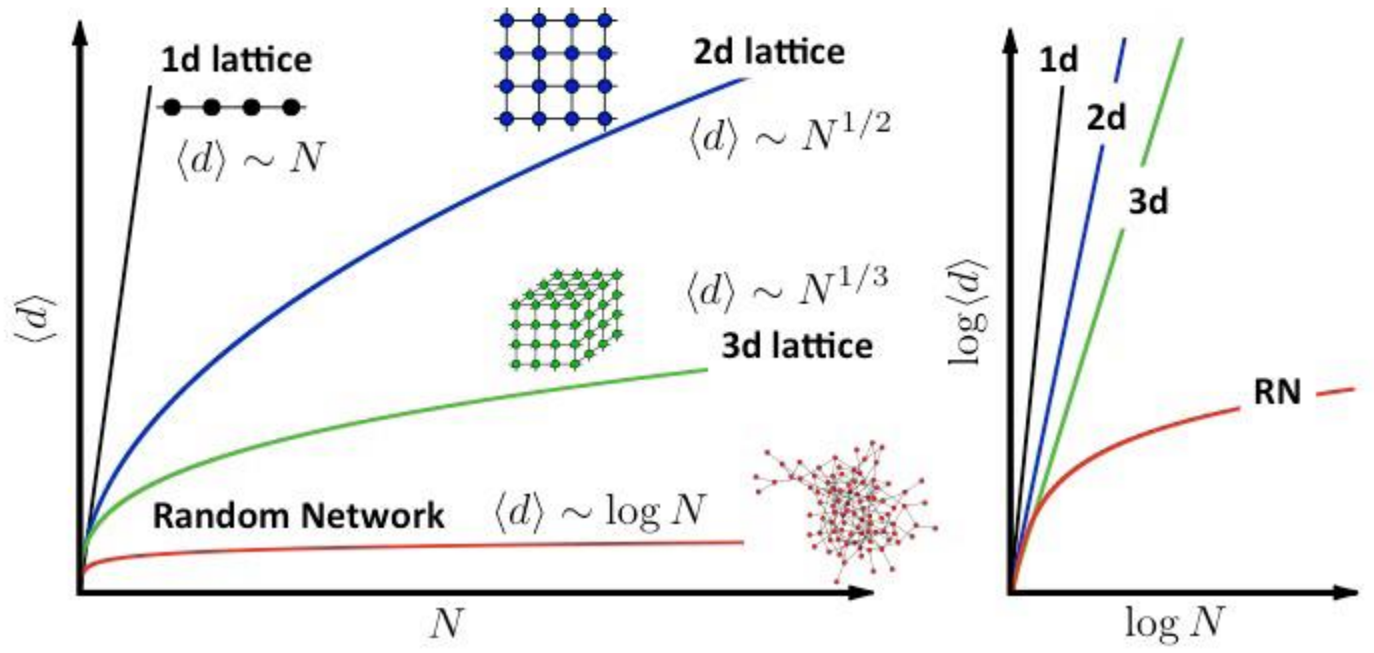


Small world property

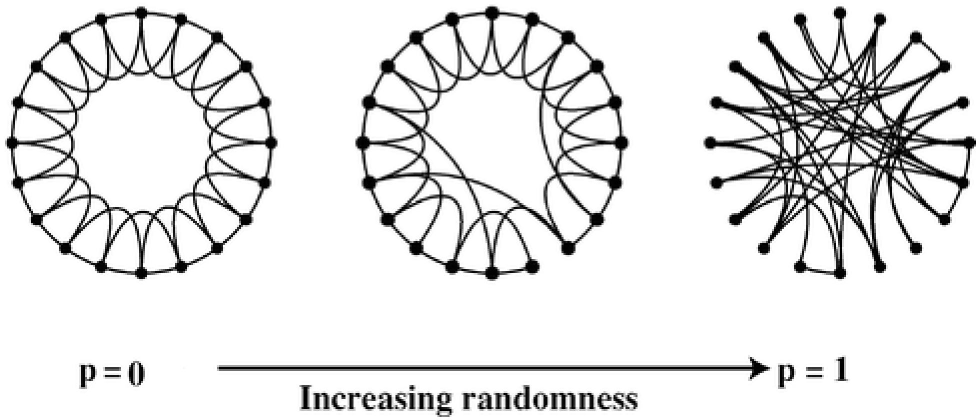
- *Distance between any two nodes in a network is surprisingly short*
  - "six degrees of separation": you can reach any other individual in the world through a short sequence of intermediaries
- What is small?
  - Consider a random network with average degree $\langle k \rangle$
  - Expected number of nodes a distance d is $N(d) \sim \langle k \rangle^d$
  - Diameter $d_{max} \sim \log N \,/\, \log\langle k \rangle$
  - <u>*Random networks are small*</u>

WHAT IS IT SURPRISING?

- **Regular lattices (e.g., physical geography) do *not* have the small world property**
  - Distances grow *polynomially* with system size
  - In networks, distances *grow logarithmically* with network size



SMALL WORLD EFFECT IN RANDOM NETWORKS

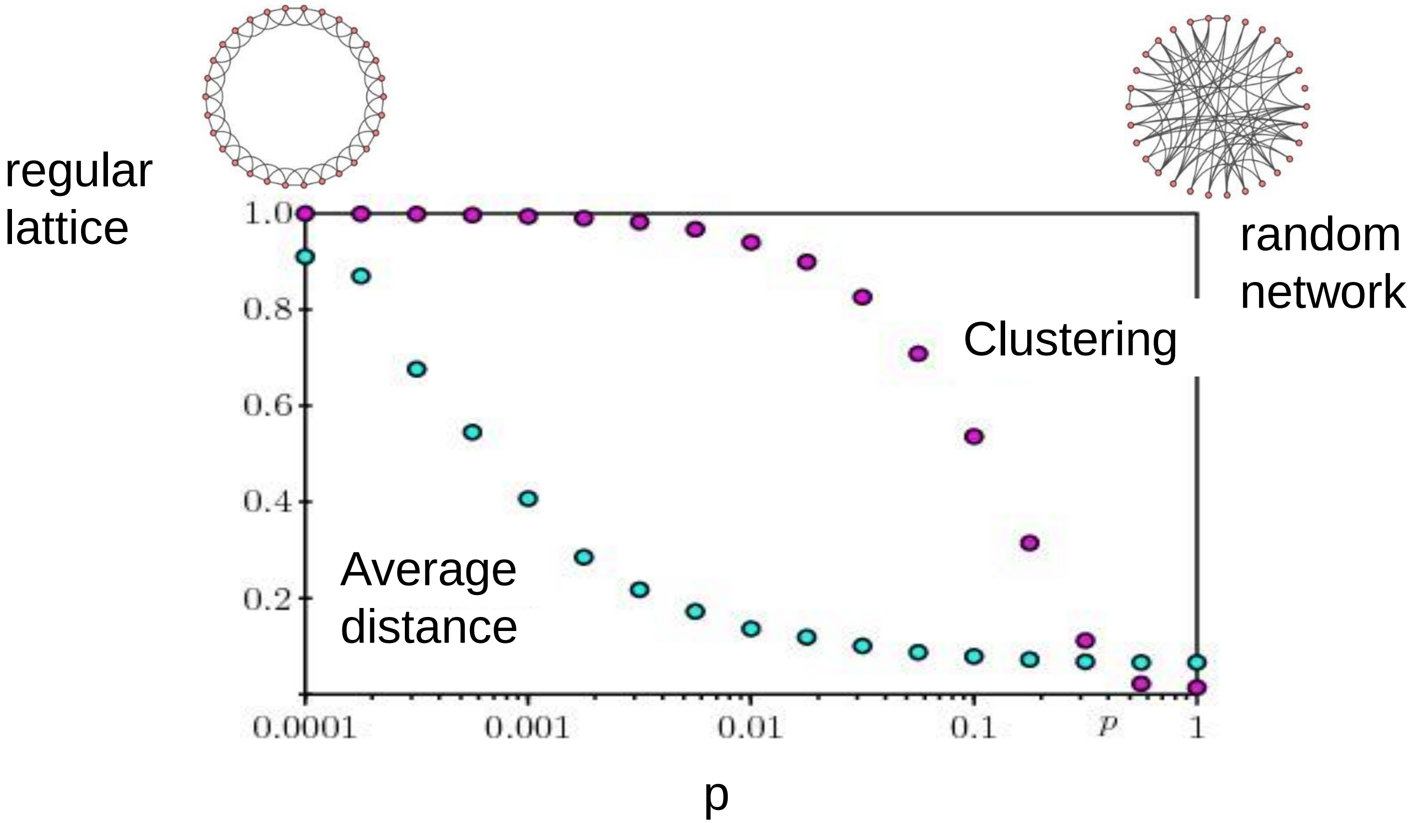


Watts-Strogatz model

- Start with a regular lattice, e.g., a ring where each node is connected to immediate and next neighbors.
  - Local clustering is $C = 3/4$
- With probability p, rewire link to a randomly chosen node
  - For small p, clustering remains high, but diameter shrinks
  - For large p, becomes random network

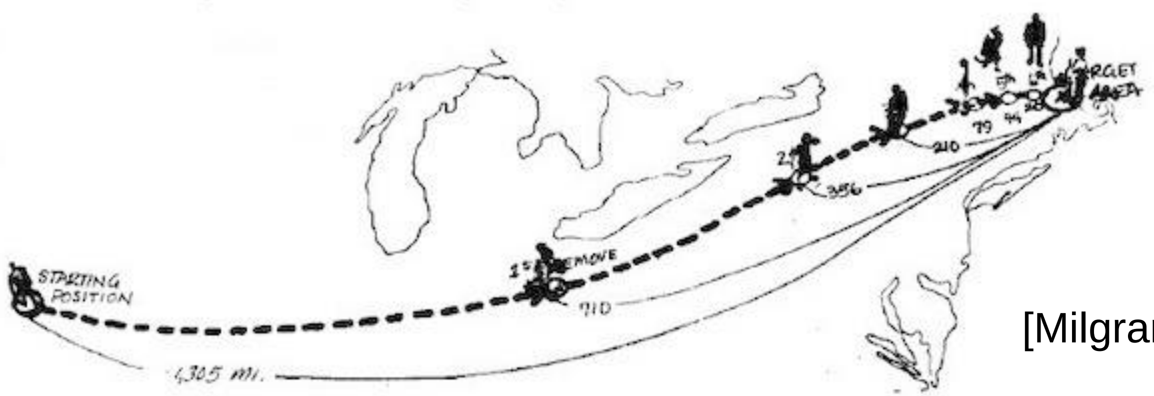
SMALL WORLD NETWORKS

- **Small world networks** constructed using Watts-Strogatz model have small average distance and high clustering, just like real networks
- *Long-distance links,* joining distant local clusters



Social networks are searchable

- Milgram experiments showed that
  - Short chains exist!
  - People can find them!
    - Using only local knowledge (who their friends are, their location and profession)
    - How are short chains discovered with this limited information?
    - Hint: geographic information?

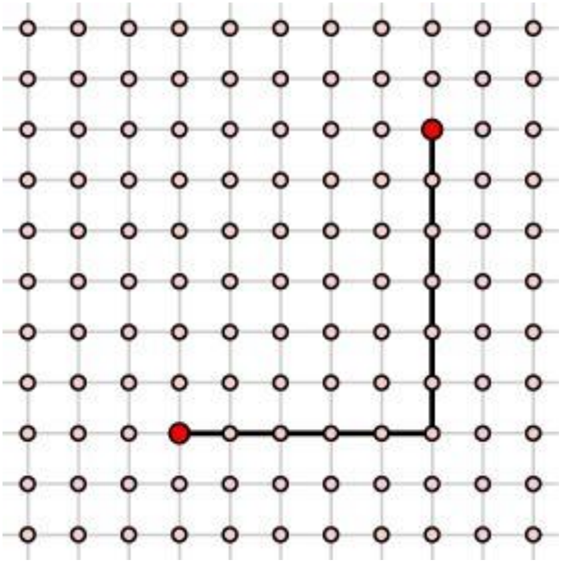


[Milgram]

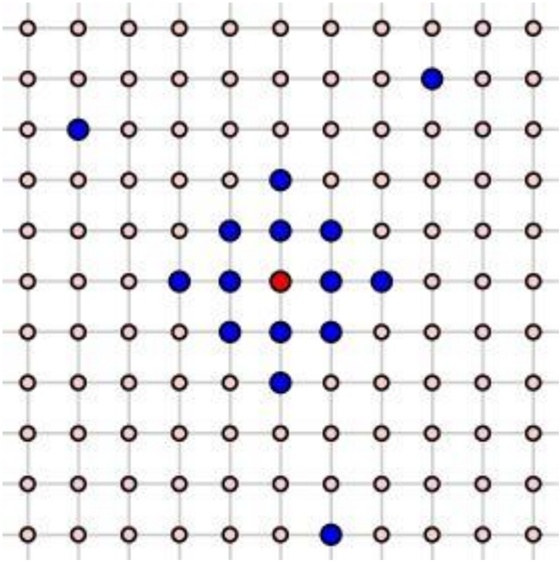
KLEINBERG MODEL OF GEOGRAPHIC LINKS

- Incorporate geographic distance in the distribution of links

Distance between nodes is d

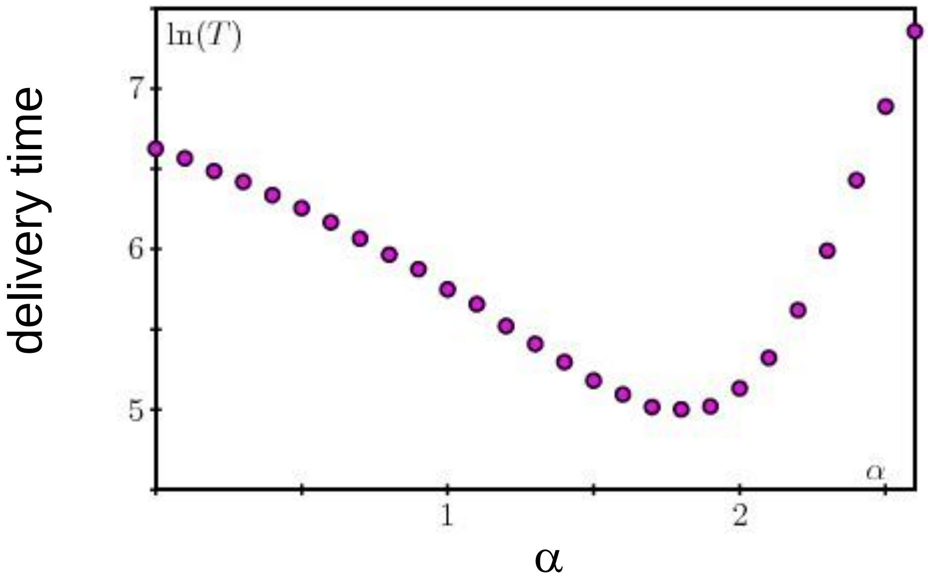


Link to all nodes within distance *r*, then add *q* long range links with probability d^{-a}



How does this affect short chains?

- Simulate Milgram experiment
  - at each time step, a node selects a friend who is closer to the target (in lattice space) and forwards the letter to it
    - Each node uses only local information about its own social network and not the entire structure of the network
  - delivery time T is the time for the letter to reach the target

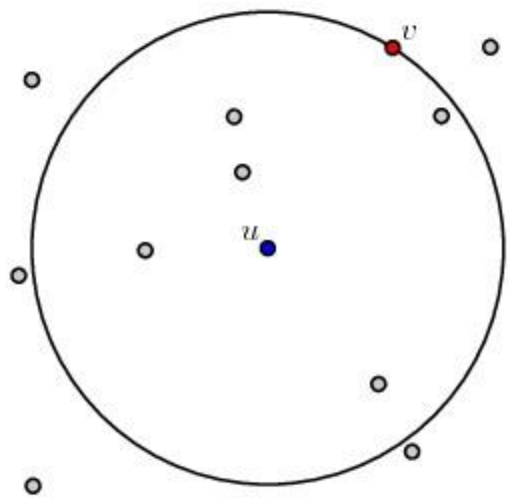


KLEINBERG'S ANALYSIS

- Network is only searchable when a=2
  - i.e., probability to form a link drops as square of distance
  - Average delivery time is at most proportional to (log N)2
- For other values of a, the average chain length produced by search algorithm is at least Nb.

DOES THIS HOLD FOR REAL NETWORKS?

- Liben-Nowell et al. tested Kleinberg's prediction for the LiveJournal network of 1M+ bloggers
  - Blogger's geographic information in profile
  - How does friendship probability in LiveJournal network depend on distance between people?
- People are not uniformly distributed spatially
  - Coasts, cities are denser



Use rank, instead of distance d(u,v)
$\text{rank}_u(v) = 6$

Since $rank_u(v) \sim d(u,v)^2$, and link probability $Pr(u \rightarrow v) \sim d(u,v)^{-2}$,
we expect that $Pr(u \rightarrow v) \sim 1/rank_u(v)$

LIVEJOURNAL IS A SEARCHABLE NETWORK

- Probability that a link exists between two people as a function of the rank between them
  - LiveJournal is a rank-based network → it is searchable