

Computer modeling of complex systems



Lecture III Everything touches everything: Networks

We are all connected. Our biological existence, social world, economy, and religious traditions tell a compelling story of interrelatedness. As the great Argentinean author Jorge Luis Borges put it,

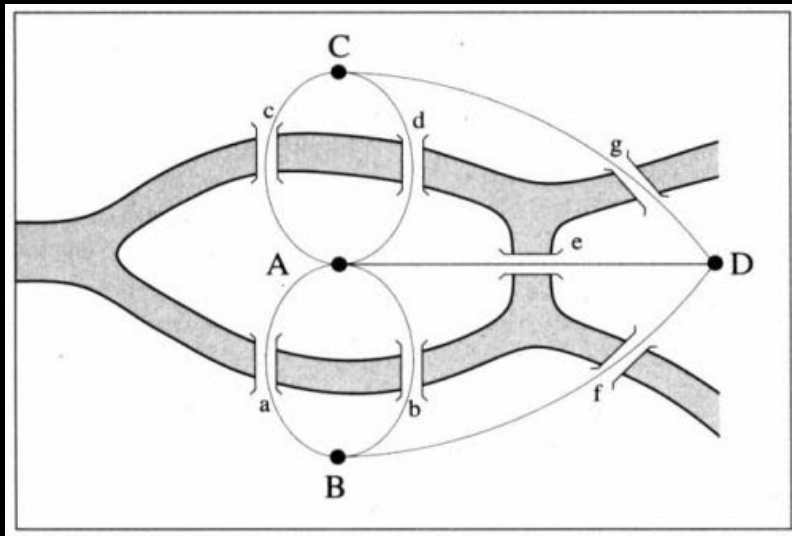
"Everything touches everything"

from: Albert-László Barabási, „Linked”



Bridges of Königsberg

Can one walk across the seven bridges and never cross the same one twice?



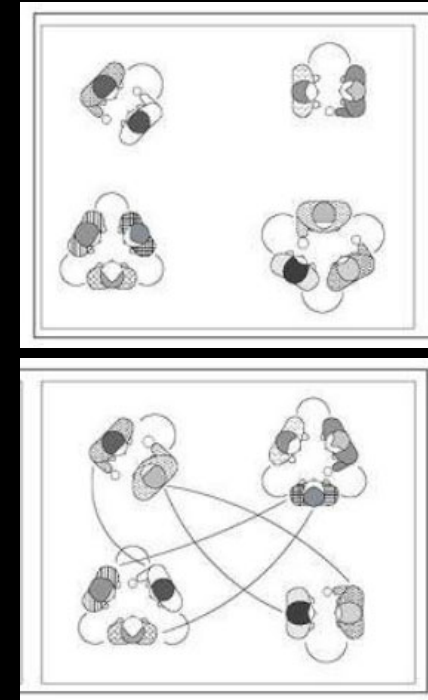
In 1736, Leonhard Euler gave birth to graph theory by replacing each of the four land areas with nodes (A to D) and each bridge with a link (a to g), obtaining a graph with four nodes and seven links. He then proved that on the Königsberg graph, a route crossing each link only once does not exist.

Erdos - Renyi model



A graph is constructed by connecting n nodes randomly. Each edge is included in the graph with probability p independent from every other edge.

Cocktail party problem



Will the rumour spread? How many connections does it take?

- one connection per node (on average) turns out to be enough for the appearance of the percolating cluster

Emergence of a giant component

- consider Erdős–Rényi graph, in which each possible edge connecting pairs of a given set of n vertices is present, independently of the other edges, with probability p
- In this model, if $p \leq (1-\epsilon)/n$ for any constant $\epsilon > 0$ then all connected components of the graph are of the order $O(\log n)$, and there is no giant component.
- however, for $p > (1+\epsilon)/n$ there appears a single giant component, with all other components having size $O(\log n)$.
- For $p=1/n$, the number of vertices in the largest component of the graph is proportional to $n^{2/3}$

Emergence of a giant component



Random Graphs

Emergence of the Giant Component

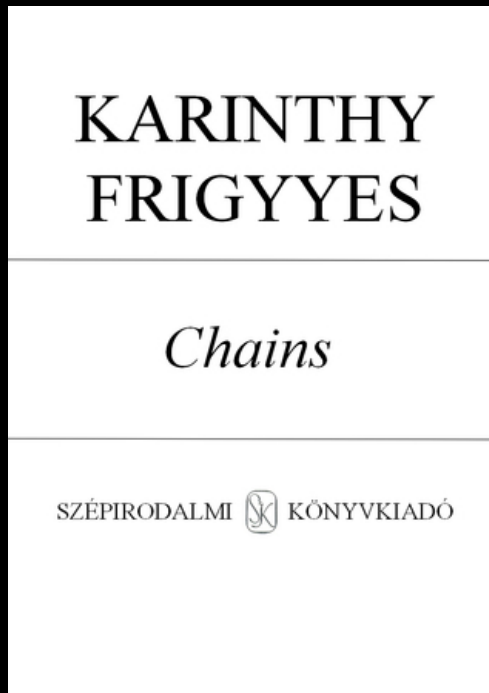
One link per node limit

- Nature exceeds the one link minimum
- Sociologists estimate that we know between 200 and 5,000 people by name
- An average neuron is connected to dozens of others

As the average number of links per node increases beyond the critical one, the number of nodes left out of the giant cluster decreases exponentially.

(The more links we add, the harder it is to find a node that remains isolated)

Small world



"To demonstrate that people on Earth today are much closer than ever, a member of the group suggested a test. He offered a bet that we could name any person among earth's one and a half billion inhabitants and through at *most five* acquaintances, one of which he knew personally, he could link to the chosen one,"

Linking with Ford

How to link the worker in Ford factory with the hero of Karinthy story?

- The worker knows the manager in the shop, who knows Ford; Ford is on friendly terms with the general director of Hearst Publications, who last year became good friends with Arpad Pasztor, someone I not only know, but is to the best of my knowledge a good friend of mine-so I could easily ask him to send a telegram via the general director telling Ford that he should talk to the manager and have the worker in the shop quickly hammer together a car for me, as I happen to need one."

Milgram's experiment (1967)

HOW TO TAKE PART IN THIS STUDY

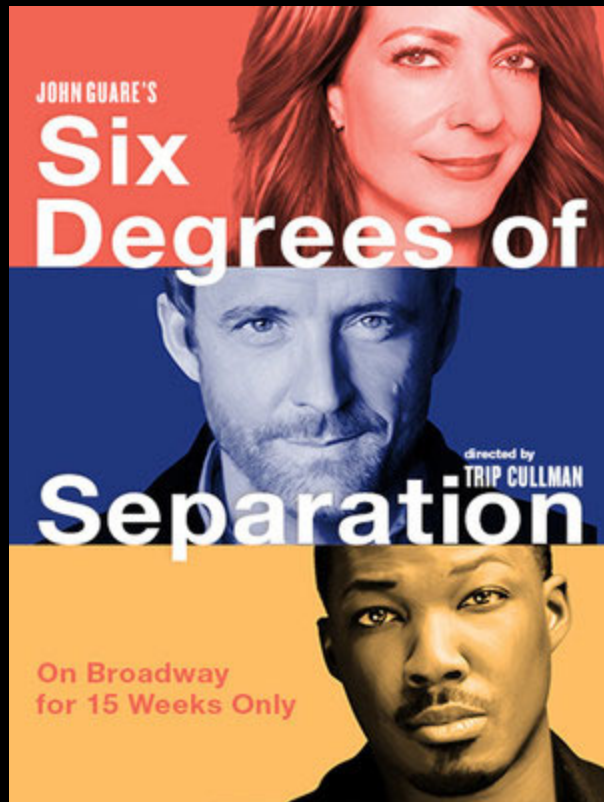
1. ADD YOUR NAME TO THE ROSTER AT THE BOTTOM OF THIS SHEET, so that the next person who receives this letter will know who it came from.
2. DETACH ONE POSTCARD. FILL IT OUT AND RETURN IT TO HARVARD UNIVERSITY. No stamp is needed. The postcard is very important. It allows us to keep track of the progress of the folder as it moves toward the target person.
3. IF YOU KNOW THE TARGET PERSON ON A PERSONAL BASIS, MAIL THIS FOLDER DIRECTLY TO HIM (HER). Do this only if you have previously met the target person and know each other on a first name basis.
4. IF YOU DO NOT KNOW THE TARGET PERSON ON A PERSONAL BASIS, DO NOT TRY TO CONTACT HIM DIRECTLY. INSTEAD, MAIL THIS FOLDER (POSTCARDS AND ALL) TO A PERSONAL ACQUAINTANCE WHO IS MORE LIKELY THAN YOU TO KNOW THE TARGET PERSON. You may send the folder to a friend, relative or acquaintance, but it must be someone you know on a first name basis.



Stanley Milgram 1933-1984

- It takes on average 5.5 links to get between person A & B

Six degrees of separation



"Everybody on this planet is separated by only six other people. Six degrees of separation. Between us and everybody else on this planet. The president of the United States. A gondolier in Venice It's not just the big names. It's anyone. A native in a rain forest. A Tierra del Fuegan. An Eskimo. I am bound to everyone on this planet by a trail of six people. It's a profound thought How every person is a new door opening up into other worlds"

We live in a well-connected ("small") world

Diameter of the world-wide-web

- How many clicks do we need to reach randomly chosen document on the Web?
- Albert, Jeong, Barabási (1999) estimated that the average path length of the WWW increases with the size of the network as $\langle d \rangle \approx 0.35 + 0.89 \ln N$
- In 1999 the WWW was estimated to have about 800 million documents, in which case the above equation predicts $\langle d \rangle \approx 18.69$.
- Currently the WWW is estimated to have about trillion nodes ($N \sim 10^{12}$), in which case the formula predicts $\langle d \rangle \approx 25$

Other networks

Network	N	L	$\langle k \rangle$	$\langle d \rangle$	d_{\max}
Power Grid	4,941	6,594	2.67	18.99	46
Mobile-Phone Calls	36,595	91,826	2.51	11.72	39
Email	57,194	103,731	1.81	5.88	18
Science Collaboration	23,133	93,437	8.08	5.35	15
Actor Network	702,3882	9,397,908	83.71	3.91	14
Citation Network	449,673	4,707,958	10.43	11.21	42
E. Coli Metabolism	1,039	5,802	5.58	2.98	8
Protein Interactions	2,018	2,930	2.90	5.61	14

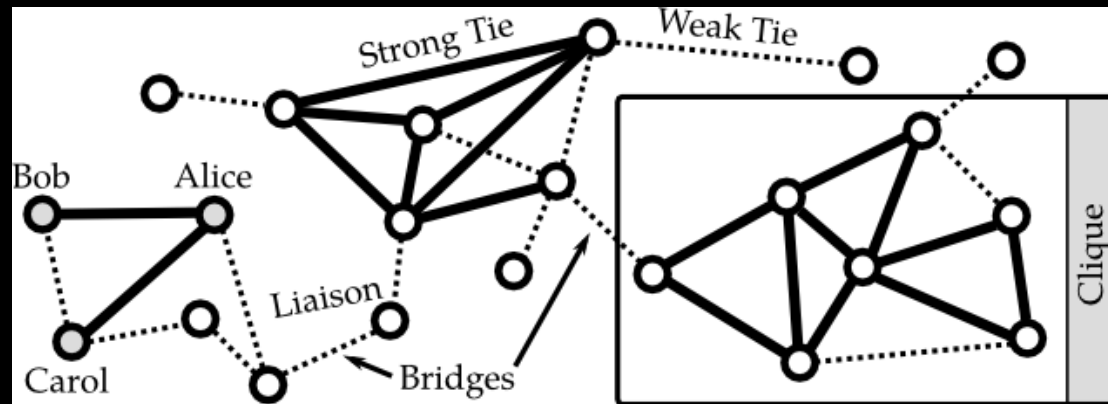


Strength of weak ties

Mark Granovetter, "The Strength of Weak Ties". *American Journal of Sociology* 78: 1360-1380, 1973

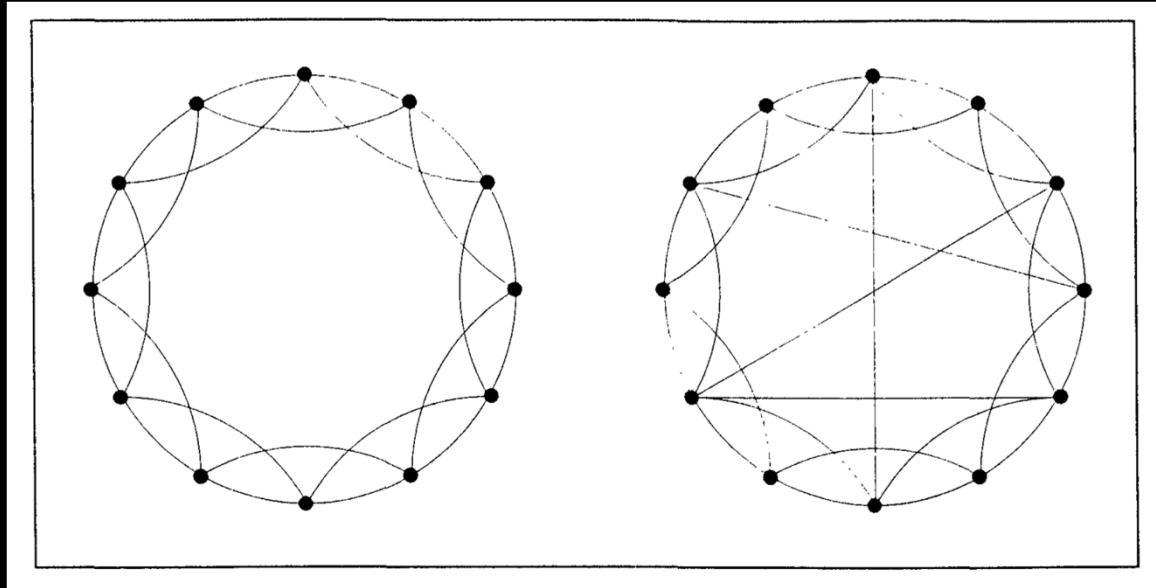
- initially sent to *American Sociological Review* (1969) with the report of the referee stating that "'the manuscript should not be published for an endless series of reasons that immediately come to mind."
- now recognized as one of the most influential sociology papers ever written (~40K citations)

Strength of weak ties



- Mark Granovetter study: over 80% of subjects found a job through a contact with whom they did not have a close relationship, more jobs were located through "friends of friends" than through close friends
- Granovetter's world is a collection of complete graphs, tiny clusters in which each node is connected to all other nodes within the cluster. These clusters are linked to each other by a few weak ties between acquaintances belonging to different circles of friends

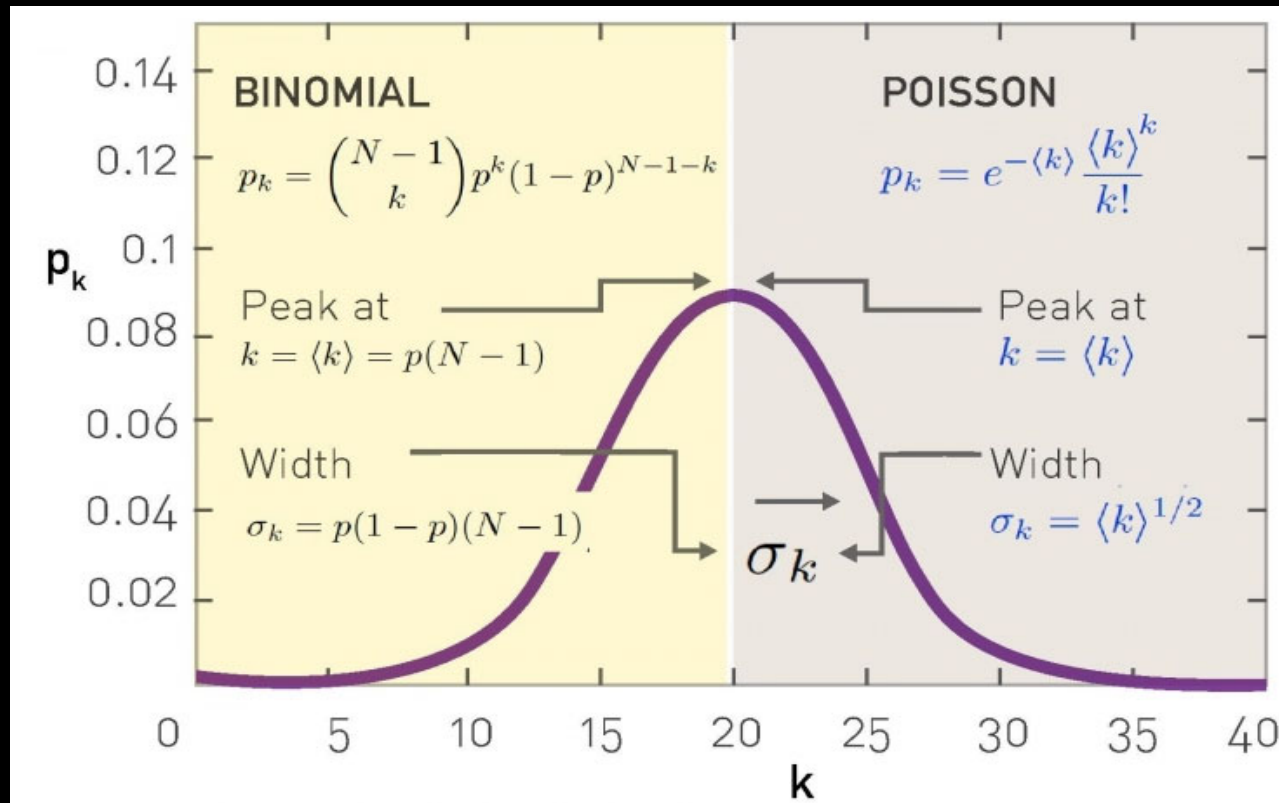
Watts-Strogatz model



- To model networks with a high degree of clustering, Watts and Strogatz started from a circle of nodes, where each node is connected to its immediate and next-nearest neighbors
- To make this world a small one, a few extra links were added, connecting randomly selected nodes. These long-range links offer the crucial shortcuts between distant nodes, drastically shortening the average separation between the nodes

Node degree distribution

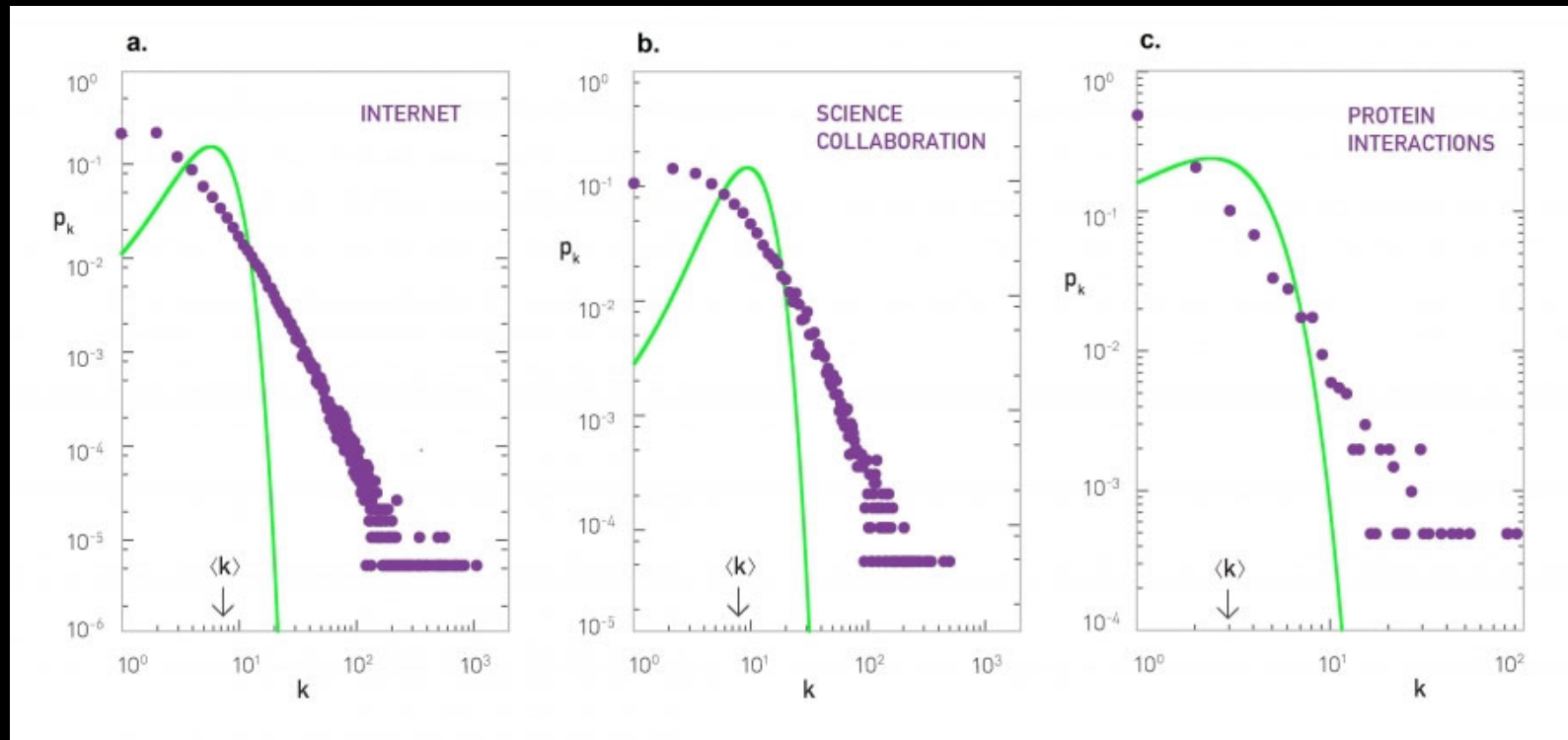
- Poisson distribution: *in a large random network the degree of most nodes is in the narrow vicinity of $\langle k \rangle$*



Real networks are not Poissonian

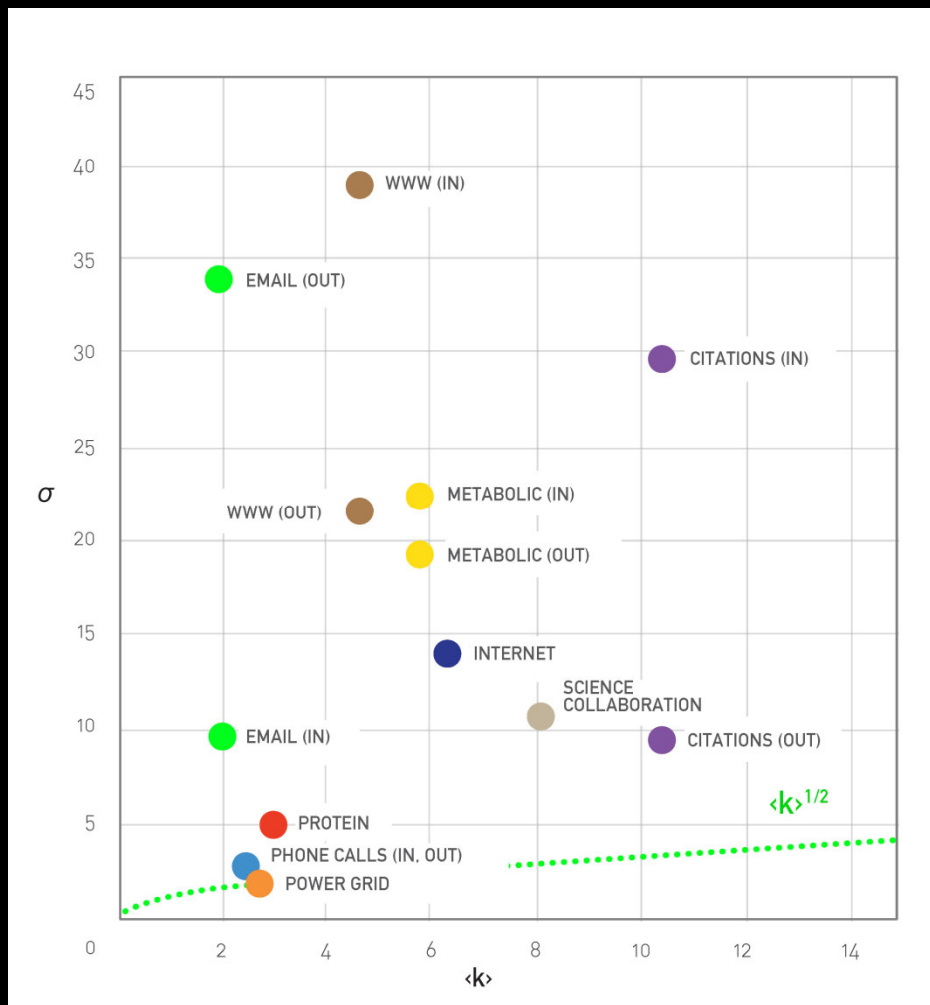
- Sociologists estimate that a typical person knows about 1,000 individuals on a first name basis,
- For a random network with $N \approx 7 \times 10^9$ nodes the most connected individual is expected to have $k_{max} = 1,185$ acquaintances
- The degree of the least connected individual is $k_{min} = 816$
- The dispersion of a random network is $\sigma_k = \langle k \rangle^{1/2}$, which for $\langle k \rangle = 1,000$ is $\sigma_k = 31.62$
- However, there is extensive evidence of individuals who have considerably more than 1,185 acquaintances. For example, US president Franklin Delano Roosevelt's appointment book has about 22,000 names, individuals he met personally

Real networks are not Poissonian



- Many real networks seem to follow power-law node degree distribution
- This implies existence of hubs: extremely well connected nodes

Other networks

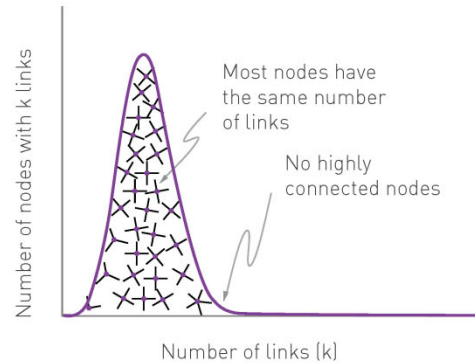


Other networks

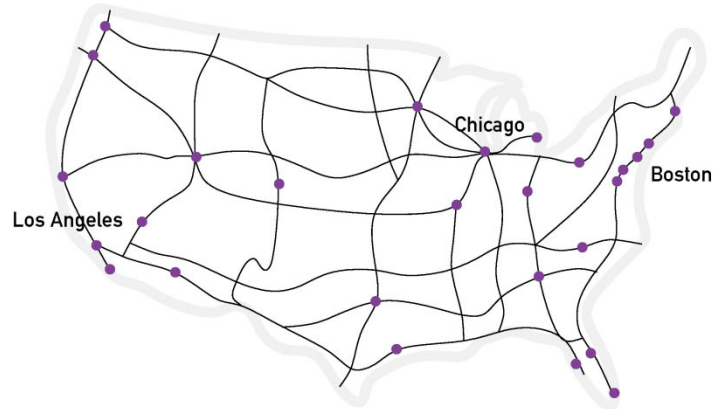
NETWORK	N	L	$\langle k \rangle$	$\langle k_{in}^2 \rangle$	$\langle k_{out}^2 \rangle$	$\langle k^2 \rangle$	γ_{in}	γ_{out}	γ
Internet	192,244	609,066	6.34	-	-	240.1	-	-	3.42*
WWW	325,729	1,497,134	4.60	1546.0	482.4	-	2.00	2.31	-
Power Grid	4,941	6,594	2.67	-	-	10.3	-	-	Exp.
Mobile Phone Calls	36,595	91,826	2.51	12.0	11.7	-	4.69*	5.01*	-
Email	57,194	103,731	1.81	94.7	1163.9	-	3.43*	2.03*	-
Science Collaboration	23,133	93,439	8.08	-	-	178.2	-	-	3.35*
Actor Network	702,388	29,397,908	83.71	-	-	47,353.7	-	-	2.12*
Citation Network	449,673	4,689,479	10.43	971.5	198.8	-	3.03**	4.00*	-
E. Coli Metabolism	1,039	5,802	5.58	535.7	396.7	-	2.43*	2.9 0*	-
Protein Interactions	2,018	2,930	2.9 0	-	-	32.3	-	-	2.89*

Poisson vs scale-free

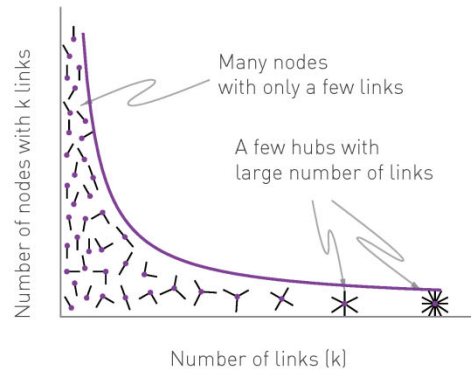
Poisson distribution



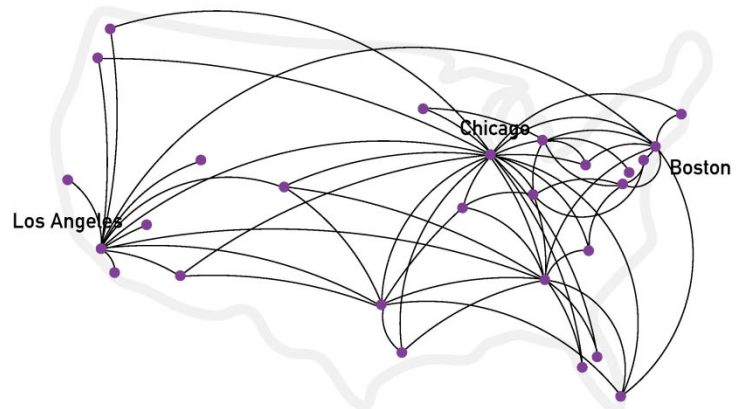
(b)



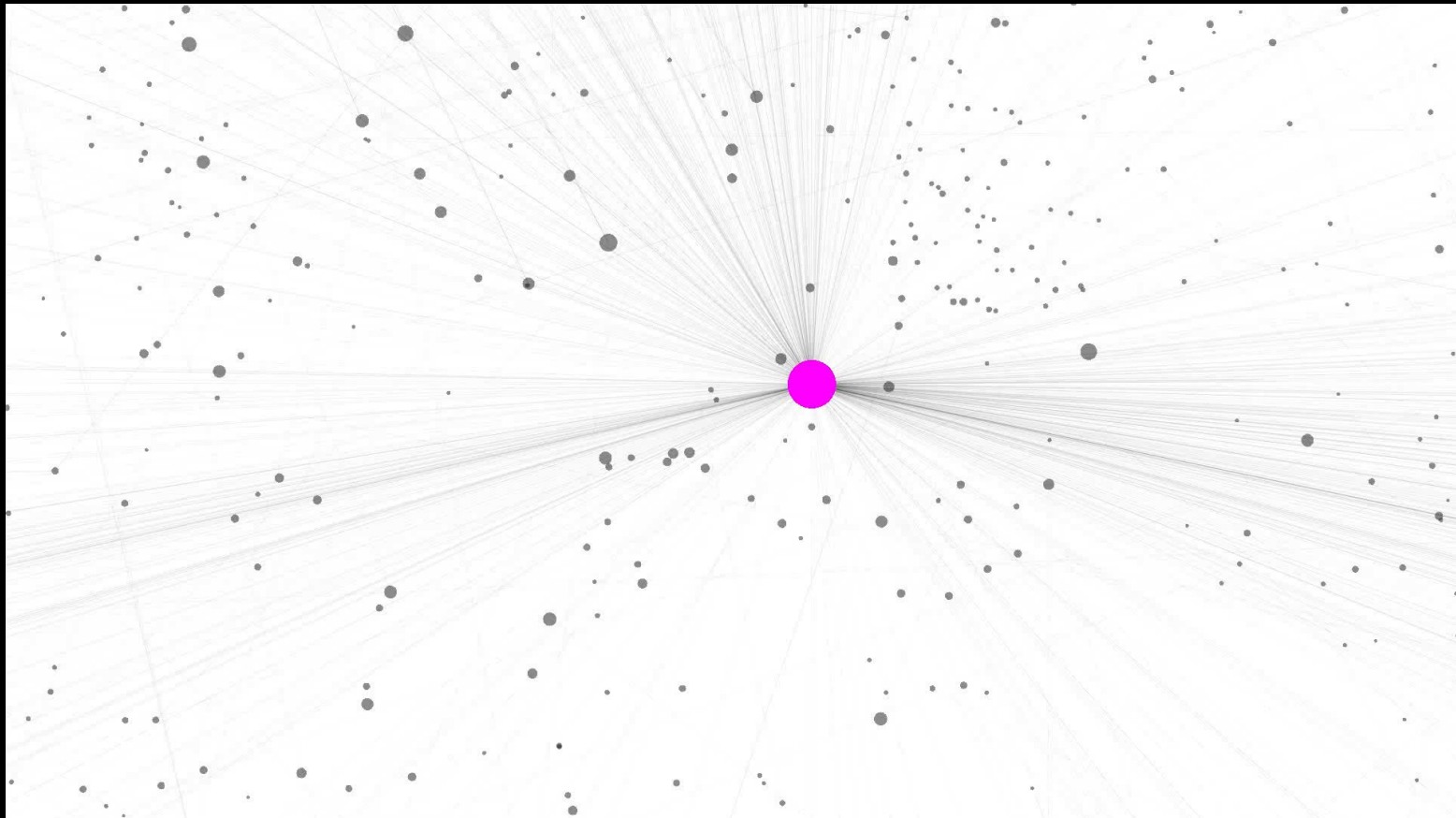
Power-law



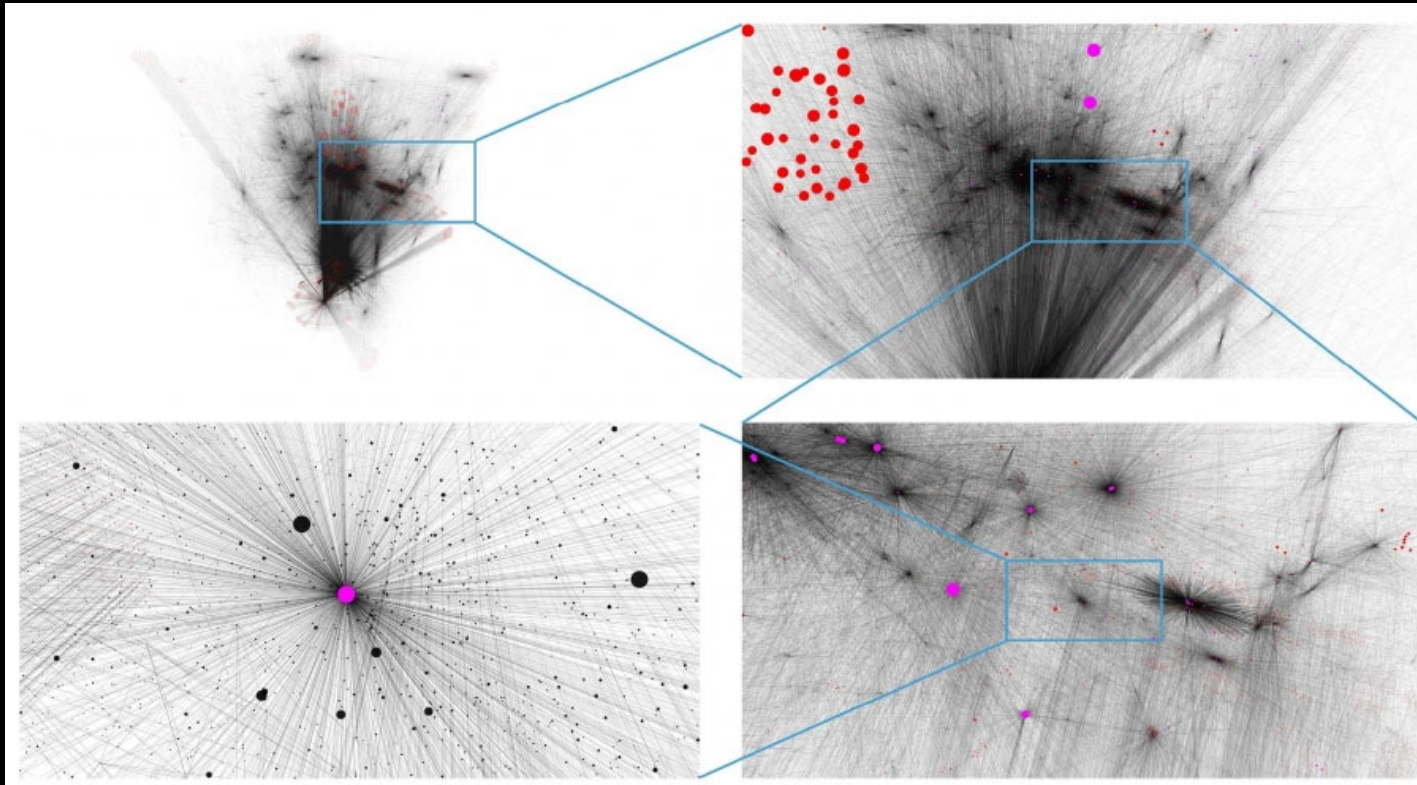
(d)



Hubs and connectors: the world-wide-web

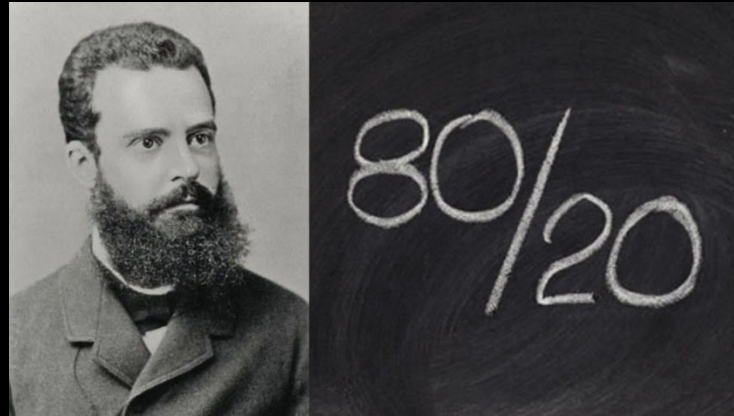


Hubs and connectors: the world-wide-web



Snapshots of the World Wide Web sample mapped out by Jeong et al (1998), showing an increasingly magnified local region of the network. The first panel displays all 325,729 nodes, offering a global view of the full dataset. Nodes with more than 50 links are shown in red and nodes with more than 500 links in purple. The closeups reveal the presence of a few highly connected nodes, called *hubs*, that accompany scale-free networks.

Hubs in economic and social networks: 80/20 rule



- Vilfredo Pareto, a 19th century economist, noticed that in Italy a few wealthy individuals earned most of the money, while the majority earned small amounts
- He concluded that incomes follow a power law
- It was popularized as the *80/20 rule*:: Roughly 80 percent of money is earned by 20 percent of the population.
- Similarly, 80 percent of profits are produced by 20 percent of the employees, 80 percent of decisions are made during 20 percent of meeting time etc.

What leads to the appearance of power-law in networks?

Rich get richer mechanism - as a new node is added to the network it connects preferentially to nodes which already have many connections. This results in a power-law distribution of degrees, where the central vertices have very large degree (hubs) - and the peripheral nodes with the minimum amount.

AL Barabási, R Albert, "Emergence of scaling in random networks"
Science 286 (5439), 509-512

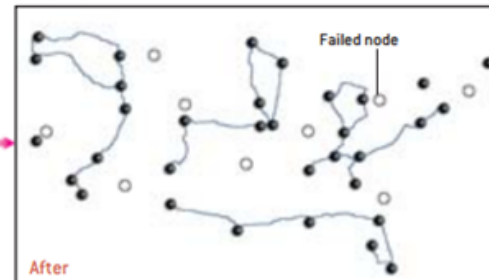
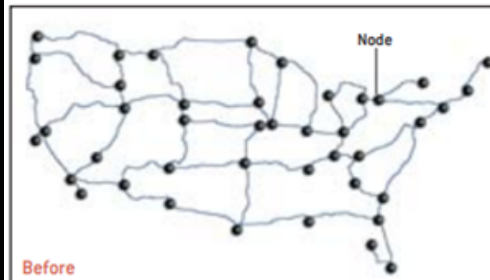


The birth of a scale-free network

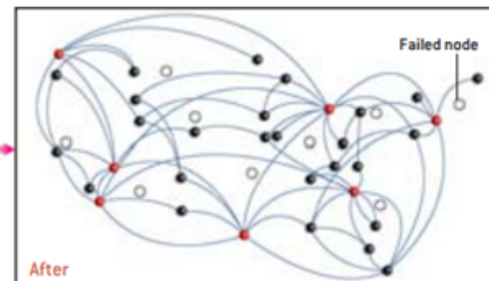
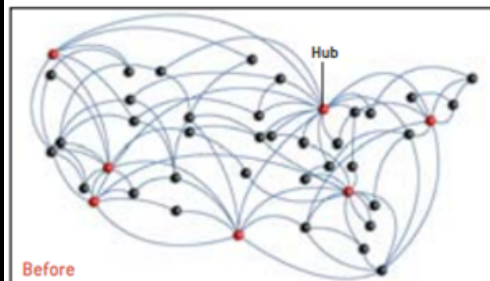


Robustness of the network

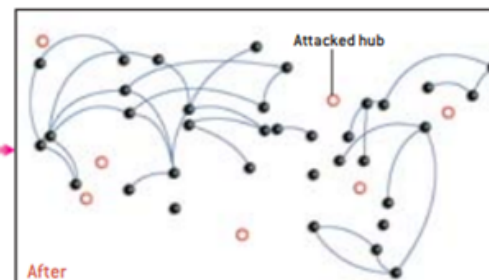
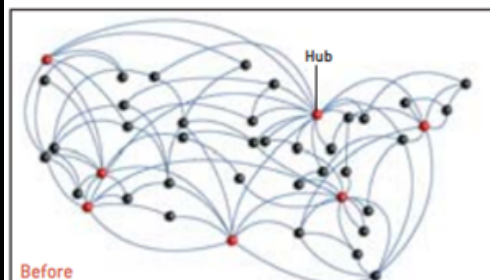
Random Network, Accidental Node Failure



Scale-Free Network, Accidental Node Failure



Scale-Free Network, Attack on Hubs



References:

- Albert-László Barabási *Network Science*
barabasi.com/networksciencebook/
- Albert-laszlo Barabasi, Jennifer Frangos *Linked: The New Science Of Networks*

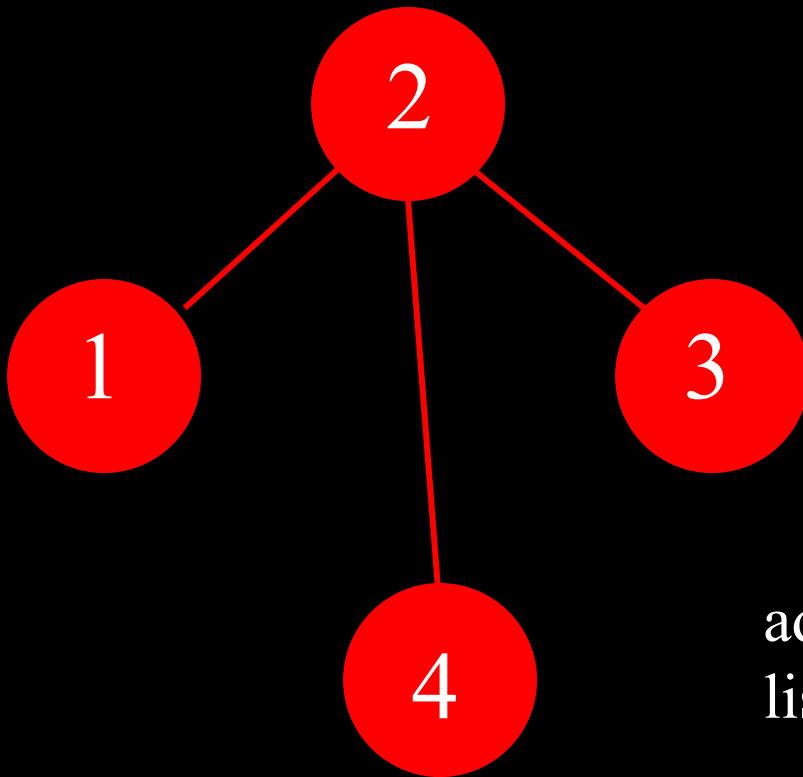
Crash course on networkx



Some definitions:

- A network (or graph) consists of a set of nodes (or vertices, actors) and a set of edges (or links, ties) that connect those nodes
 - math: “graph/vertex/edge,”
 - physics “network/node/edge,”
 - computer scientists “network/node/link,” social scientists use “network/actor/tie,”
- Node j is called a **neighbor** of node i if (and only if) node i is connected to node j
- **Adjacency matrix:** a matrix with rows and columns labeled by nodes, whose i -th row, j -th column component a_{ij} is 1 if node i is a neighbor of node j , or 0 otherwise
- **Adjacency list:** a list of lists of nodes whose i -th component is the list of node i 's neighbors

Connections



adjacency
matrix:

$$\begin{bmatrix} 0 & 1 & 0 & 0 \\ 1 & 0 & 1 & 1 \\ 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix}$$

adjacency
list:

$1 \rightarrow \{2\}$

$2 \rightarrow \{1, 3, 4\}$

$3 \rightarrow \{2\}$

$4 \rightarrow \{2\}$

Degree - the number of edges
connected to a node

A crash course

```
import networkx as nx
# creating a new empty Graph object
g = nx.Graph()
# adding a node named 'John'
g.add_node('John')
# adding a bunch of nodes at once
g.add_nodes_from(['Josh', 'Jane', 'Jess', 'Jack'])
# adding an edge between 'John' and 'Jane'
g.add_edge('John', 'Jane')
# adding a bunch of edges at once
g.add_edges_from([('Jess', 'Josh'), ('John', 'Jack'), ('Jack', 'Jane')])
# adding more edges
# undefined nodes will be created automatically
g.add_edges_from([('Jess', 'Jill'), ('Jill', 'Jeff'), ('Jeff', 'Jane')])
# removing the edge between 'John' and 'Jane'
g.remove_edge('John', 'Jane')
# removing the node 'John'
# all edges connected to that node will be removed too
g.remove_node('John')
```

Modifying network

```
>>> g
>>> g.nodes()
>>> g.edges()
>>> g.node
```

add the attributes of the node:

```
>>> g.node['Jeff']['job'] = 'student'
>>> g.node['Jeff']['age'] = 20
>>> g.node['Jeff']
```

and edges

```
>>> g.edge['Jeff']['Jane']['trust'] = 1.0
>>> g.edge['Josh']['Jess']['love'] = True
>>> g.edge['Jess']['Josh']['love']
```

Visualizing networks

```
from pylab import *
import networkx as nx
g = nx.karate_club_graph()
subplot(2, 2, 1)
nx.draw(g,node_size=50)
title("spring layout")
subplot(2, 2, 2)
nx.draw_random(g,node_size=50)
ax = plt.gca() # to get the current axis
ax.collections[0].set_edgecolor("#000000")
title("random layout")
subplot(2, 2, 3)
nx.draw_circular(g,node_size=50)
title("circular layout")
```

```
subplot(2, 2, 4)
shells = [[0, 1, 2, 32, 33],[3, 5, 6, 7, 8, 13, 23, 27, 29,
30, 31],[4, 9, 10, 11, 12, 14, 15, 16, 17, 18,19, 20, 21,
22, 24, 25, 26, 28]]
nx.draw_shell(g, nlist =
shells,node_size=50,linewidths=1,edge_color='black'
,node_color='r',
node_shape='o')
title("shell layout")
show()
```

Drawing options

```
positions = nx.spring_layout(g)
subplot(3, 2, 1)
nx.draw(g, positions, with_labels = True,node_size=150,font_size=10)
title("showing node names")
subplot(3, 2, 2)
nx.draw(g, positions, node_shape = ">",node_size=50)
title("using different node shape")
subplot(3, 2, 3)
nx.draw(g, positions,
node_size = [g.degree(i) * 10 for i in g.nodes()])
title("changing node sizes")
subplot(3, 2, 4)
nx.draw(g, positions, edge_color = "pink",node_color = ["yellow" if i < 17 else "green" for i in g.nodes()])
title("coloring nodes and edges")
subplot(3, 2, 5)
nx.draw_networkx_nodes(g, positions,node_size=50)
title("nodes only")
show()
```

Random graphs

```
subplot(2, 2, 1)
nx.draw(nx.gnm_random_graph(10, 20))
title('random graph with\n10 nodes, 20 edges')
subplot(2, 2, 2)
nx.draw_spring(nx.gnp_random_graph(15, 0.8))
title('random graph with\n 20 nodes, 10% edge probability')
subplot(2, 2, 3)
nx.draw(nx.random_regular_graph(3, 10))
title('random regular graph with\n10 nodes of degree 3')
subplot(2, 2, 4)
nx.draw(nx.random_degree_sequence_graph([3,3,3,3,4,4,4,4,5,5]))
title('random graph with\n degree sequence\n[3,3,3,3,4,4,4,4,5,5]')
show()
```

Some classical networks

```
from pylab import *
import networkx as nx
n = 20
er = nx.erdos_renyi_graph(n, 0.2)
ws = nx.newman_watts_strogatz_graph(n, 4, 0.2)
ba = nx.barabasi_albert_graph(n, 3)
figure(figsize=(10,10))
subplot(2, 2, 1)
nx.draw_circular(er,node_size=10)
subplot(2, 2, 2)
nx.draw_circular(ws,node_size=10)
subplot(2, 2, 3)
nx.draw_circular(ba,node_size=10)
show()
```


and their options:

`barabasi_albert_graph(n, m)` - returns a random graph by attaching new nodes each with m edges with a probability that proportional to the number of links the existing nodes have. The initialization is a graph with m nodes and no edges, the final graph has n nodes.

`newman_watts_strogatz_graph(n, k, p)` - returns a Watts-Strogatz small-world graph. First creates a ring of n nodes, each connected with its k nearest. Then shortcuts are added as follows: for each edge $u-v$ in the ring with probability p add a new edge $u-w$ with uniformly random choice of node w

`erdos_renyi_graph(n, p)` - returns a random graph with n nodes. Chooses each of the possible edges with probability p .

Try generating other networks:

<https://networkx.github.io/documentation/latest/reference/generators.html>

