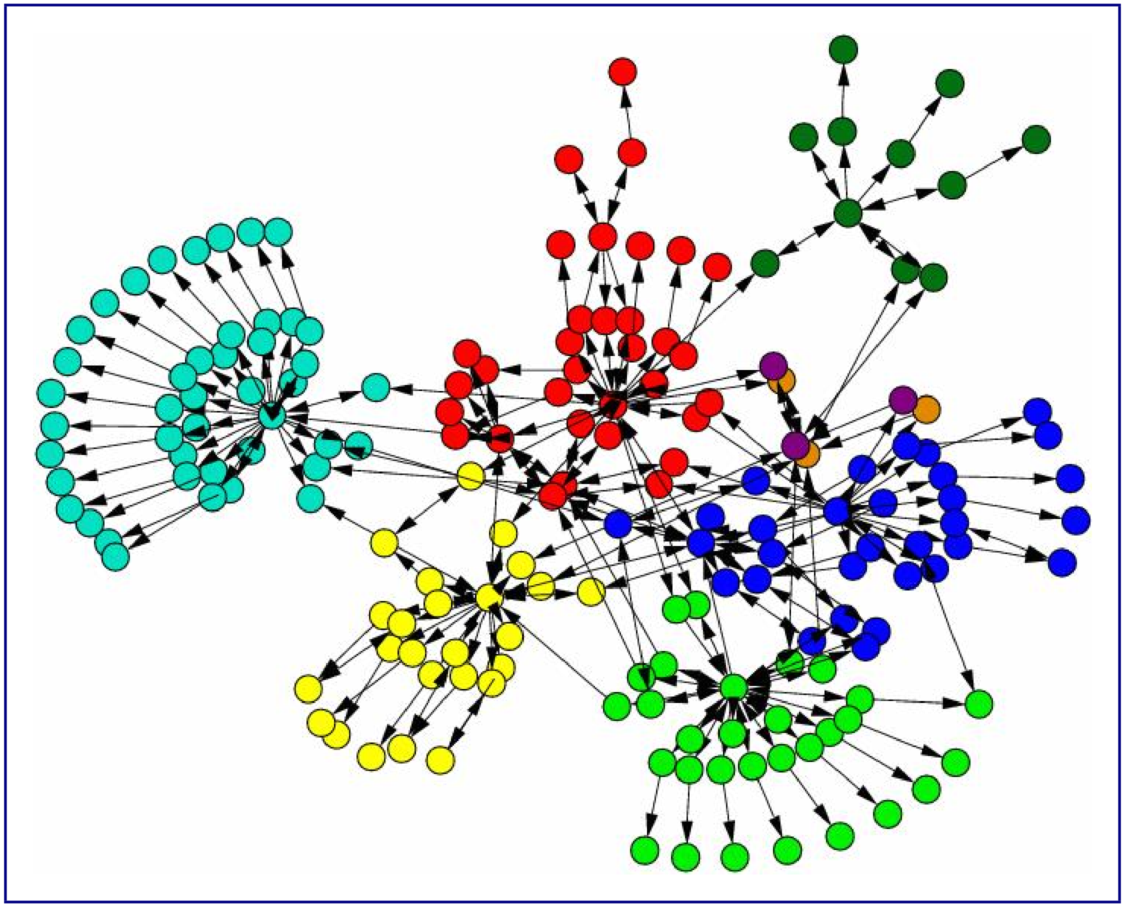
ECS 253 / MAE 253, Lecture 10

May 3, 2018



"Web search and decentralized search on small-world networks"

Announcements

- First example of the problems introduced by using software routines rather than working with the raw data/writing code. (Matlab histogram normalization.)
- HW3, HW3a, HW3b are posted. Due May 15th.
- Office hours today 3:30-4:20pm in 3057 Kemper.

Search for information

Assume some resource of interest is stored at the vertices of a network. Need effective search algorithms for:

- Web pages (static network, highly profitable)
- Files in a file-sharing network (dynamic network)

Would like to determine rapidly where in the network a particular item of interest can be found.

$\rightarrow$ Worst-case scenario, search all N nodes.

More efficient search strategies

- Can the nodes be ordered in any way?
  Then may be able to use algorithms like binary search for an ordered list.
- Can nodes learn information about their neighbors?
- Can nodes store information about their neighbors?

Search on power-law random graphs

Adamic, R. M. Lukose, A. R. Puniyani, and B. A. Huberman, "Search in power-law networks", Phys. Rev. E, 64 2001.)

- Modeled after file-sharing networks like GNUTELLA.
- Nodes can store information on neighbors (and even next-nearest neighbors).
- The article is a very nice example of the use of generating functions for random graphs (the topic of future lectures).
- Breadth-first passing always to highest-degree node possible find between $O(N^{2/3})$ and $O(N^{1/2})$.

Decentralized search on small-world networks

What is a small world?

- Dates back at least to 1929, Frigyes Karinthy, "Everything is different"
  – The modern world was "shrinking"
  – Due to technological advances in communications and travel, friendship networks could grow larger and span greater distances

- Michael Gurevich, 1961 MIT PhD dissertation

- Manfred Kochen, 1978: For the US size population, "it is practically certain that any two individuals can contact one another by means of at most two intermediaries"

- Stanley Milgram, "The Small World Problem" in *Psychology Today*, 1967.

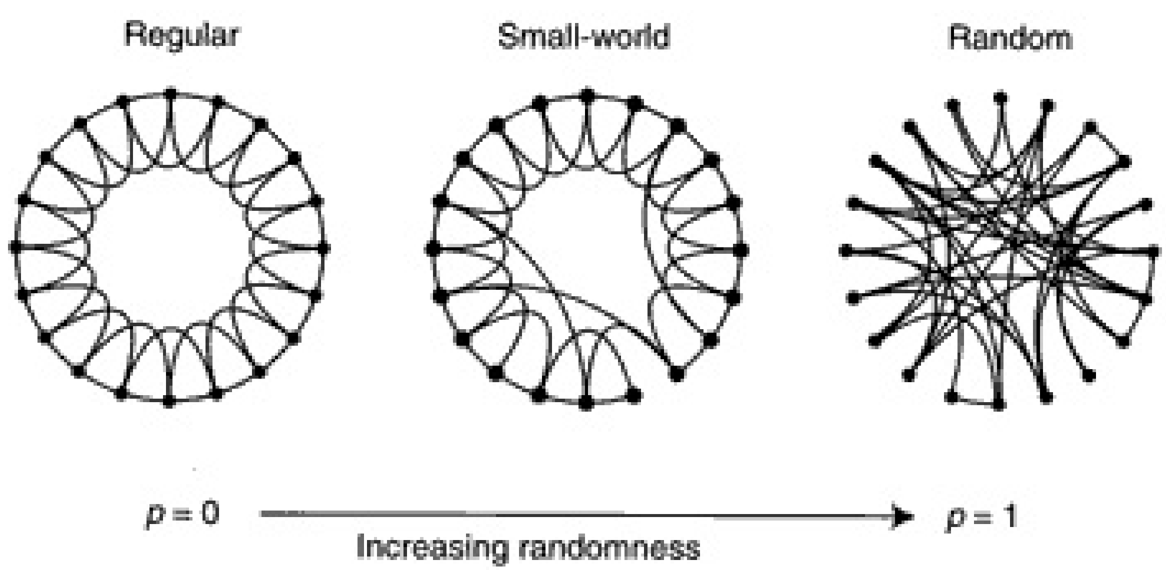
- John Guare, "Six Degrees of Separation" a play in 1990 and film in 1993.

- Watts and Strogatz, *Nature* 1998

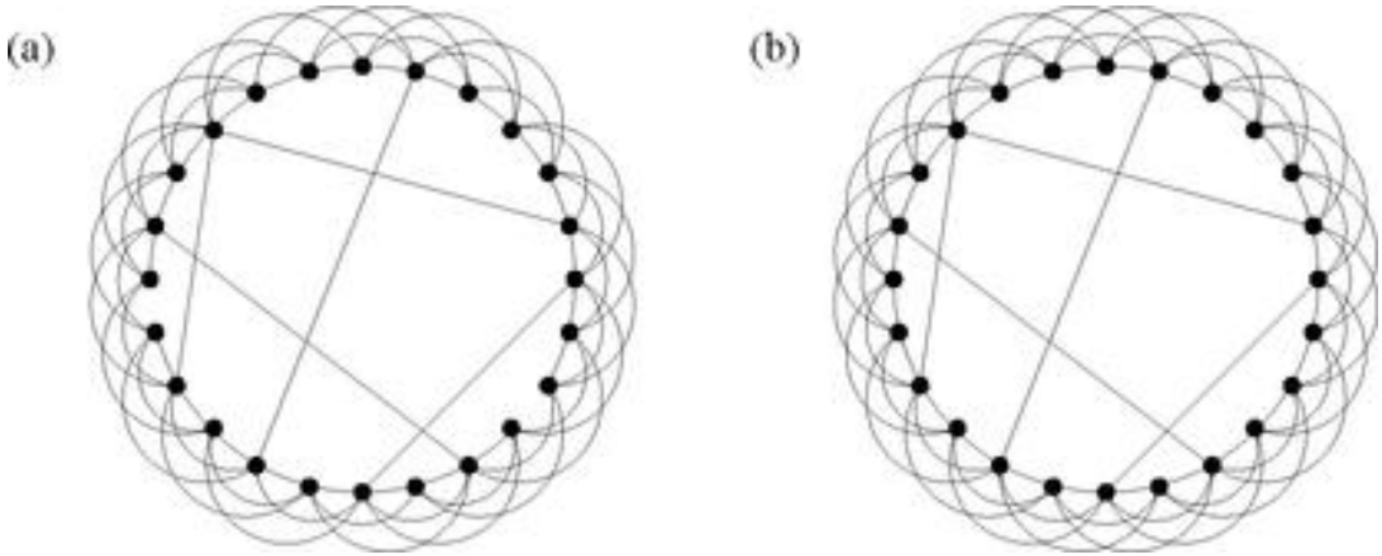
Mathematically defining a small-world

[Watts and Strogatz, *Nature*, 393 (1998)]

- Start with 1D ring of N nodes, with each node connected to it's c nearest neighbors. (All nodes start with degree c; in the picture below $c = 4$.)
- Randomly re-wire each link independently with probability p.
- Number of edges = nc/2. Expected number of rewired edges = ncp/2.



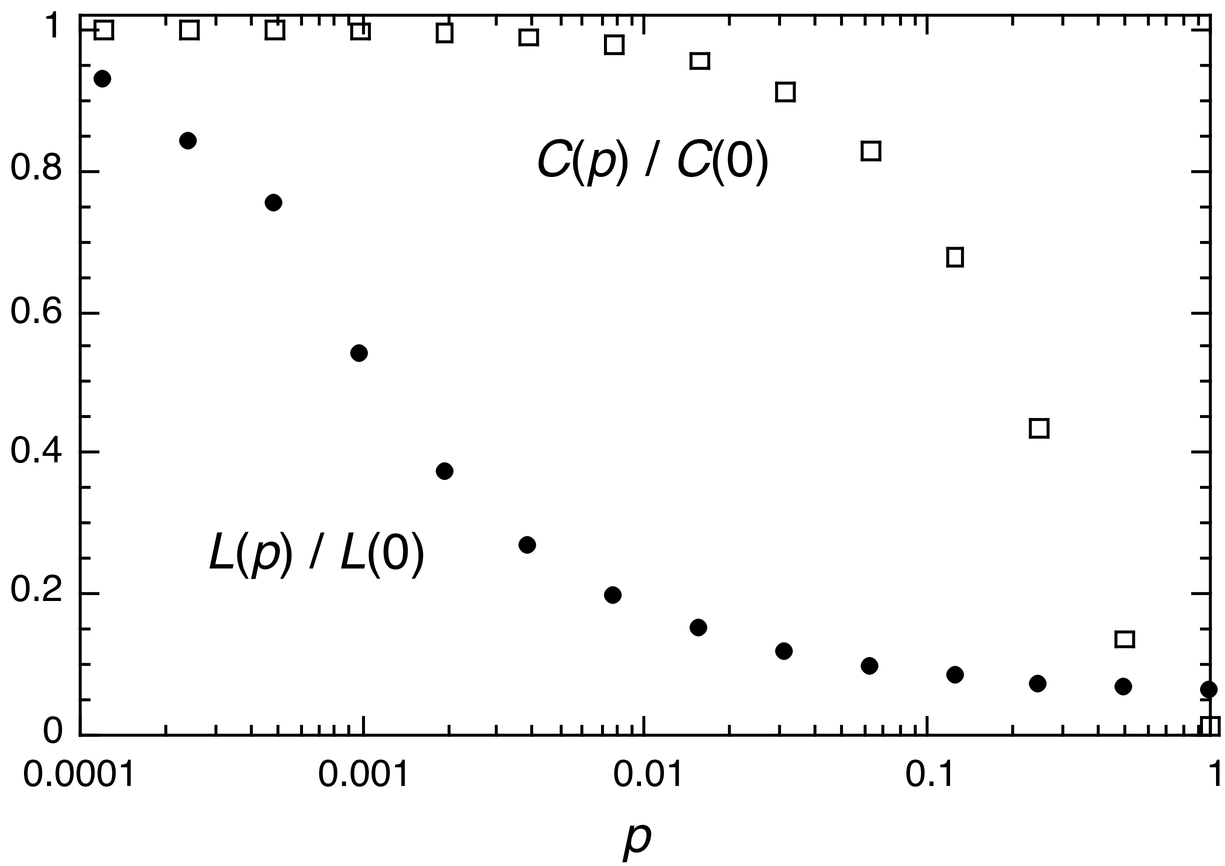
Two definitions of the SW model are essentially equivalent:



- (a) Original model, rewire edges with probability p.
  Average final node degree is c.

- (b) Add shortcuts with probability p, but do not remove edges.
  Average final node degree is $c + cp$.

Ave shortest path $L(p)$ and clustering coefficient $C(p)$

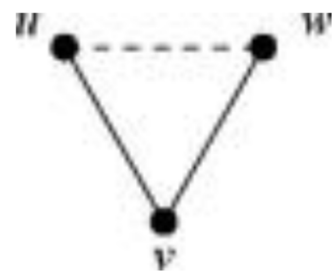


- Small-worlds have small diameter and large clustering coefficient.
- They are remarkably easy to generate (just a tiny p required).

Clustering coefficient, C_i

- C_i for a vertex i is the number of pairs of neighbors of i that are connected divided by the numbers of pairs of neighbors of i.

  C_i = (number of links between neighbors of i, excluding i) / (total number of links that could exist between neighbors).



- In more formal terms: $C_i = 2|e_{jk}|/k_i(k_i - 1)$,

  where $|e_{jk}|$ is the total number of links between all nodes j and k that are connected to node i (NOT including i), and k_i is the degree of node i.

- Note the factor of 2 comes from the fact we are considering undirected edges, so the total number of edges that could exist between neighbors of i is $k_i(k_i-1)/2$. If a node is disconnected (*i.e.*, $k_i = 0$), then assume $C_i = 0$ for that node.

“Six degrees” (i.e., a small world)

- Six Degrees of Separation — 1993 Dramatic Film
- Six Degrees: The Science of a Connected Age. — 2003 book by Duncan Watts

Fun related websites

- The Oracle of Bacon – The path to Kevin Bacon
  http://oracleofbacon.org/

- SixDegrees.org – connecting causes and celebrities
  http://www.sixdegrees.org/

- Six degrees of Kevin Garnett
  http://www.slate.com/articles/sports/slate_labs/2013/10/six_degrees_of_kevin_garnett_connect_any_two_athletes_who_ve_ever_played.html

- Six degrees of NBA seperation
  http://harvardsportsanalysis.wordpress.com/2011/03/04/six-degrees-of-nba-separation/
  (Blog post explaining use of Dijkstra's algorithm)

Navigation

Clearly if central coordination, can use short paths to deliver info quickly.

But, can someone living in a small world actually make use of this info and do efficient decentralized routing?

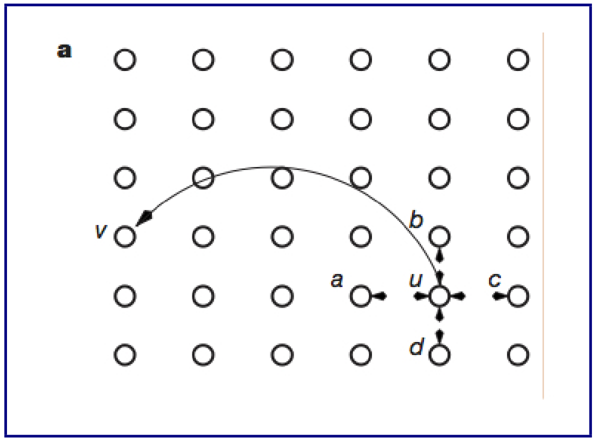
- Instead of designing search algorithms, given a local greedy algorithm, are there any topologies that enable $O(\log N)$ delivery times?

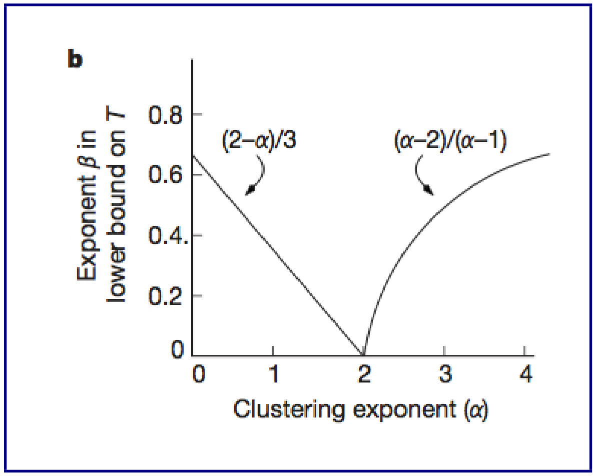
Precise topologies required

[J. M. Kleinberg, "Navigation in a small world", *Nature*, 406 (2000)]

- Find mean delivery time $t \sim N^{\beta}$, unless $\alpha = 2$.

- Only for $\alpha = 2$ will decentralized routing work, and packet can go from source to destination in $O(\log N)$ steps.





- For d-dimensional lattice need $\alpha = d$.

But we know greedy decentralized routing works for human networks (c.f. Milgram's experiments "six-degrees of separation" [S. Milgram, "The small world problem", *Psych. Today*, 2, 1967.]

So how do we get beyond a lattice model?

Navigating social networks

[Watts, P. S. Dodds, and M. E. J. Newman, “Identity and search in social networks”, Science, 296 (2002)]

[Kleinberg, “Small world phenomena and the dynamics of information”, in Proceedings of NIPS 2001].

- Premise: people navigate social networks by looking for common features between there acquaintances and the targets (occupation, city inhabited, age, ...)
- Brings in DATA!

Hierarchical “social distance” tree

- Individuals are grouped into categories along many attributes.
- One tree for each attribute.
- Trees are not the network, but complementary mental constructs believed to be at work.
- Assume likelyhood of acquaintance falls of exponentially with “social distance”.

Building P2P architectures

- "Chord A Scalable Peer-to- peer Lookup Service for Internet Applications

  I. Stoica, R. Morris, D. Karger, F. Kaashoek, H. Balakrishnan, ACM SIGCOMM, 2001.
  (Cited 4825 times)

- Gnutella

  "Peer-to-Peer Architecture Case Study: Gnutella Network", M Ripeanu, Proceedings of International Conference on Peer-to-peer Computing, 2001.

Summary

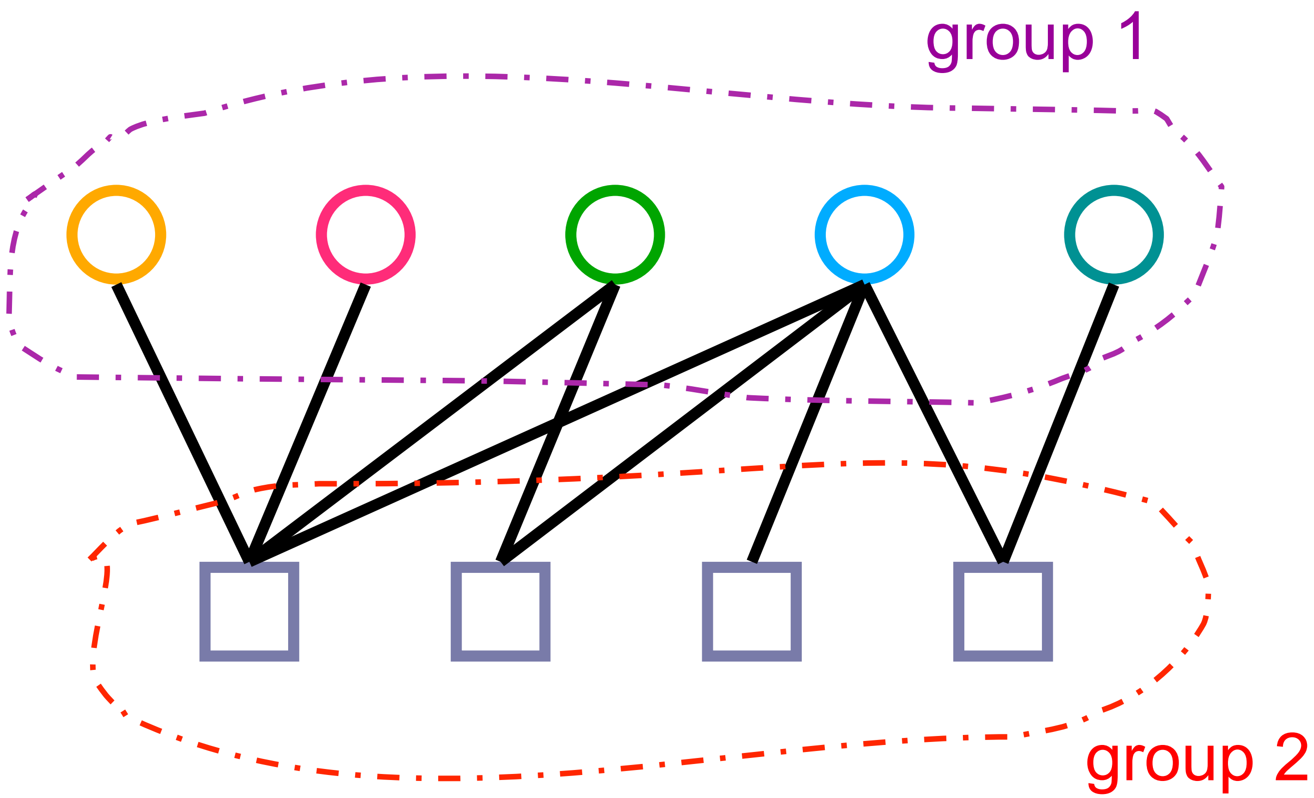
Web search

- Centralized
- Make use of link structure (topology)

Decentralized search

- Efficiency/Speed depends on underlying topology
- Gossip algorithms (D. Kempe and J. Kleinberg): spreading shared information quickly through local exchanges (e.g., sums, local averages/consensus).
- Applications to P2P, sensor networks .... communications ... satellites

Other important network paradigms:
Bipartite networks, trees, cliques, and cores



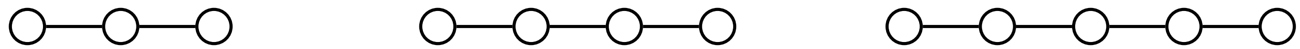
Other important network paradigms

- Bipartite networks
- Hypergraphs
- Trees
- Planar graphs
- Cliques

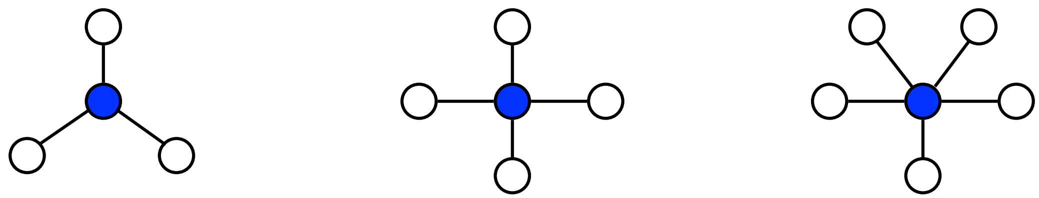
This content largely from Adamic's lectures

Some Basic Types of Graphs

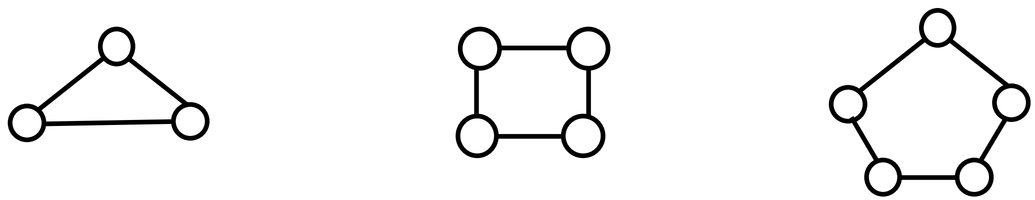
Paths



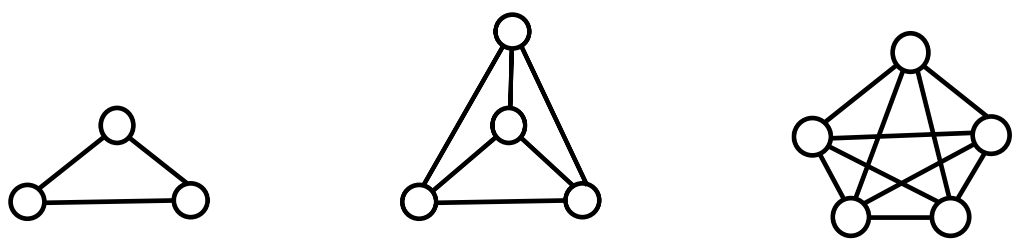
Stars



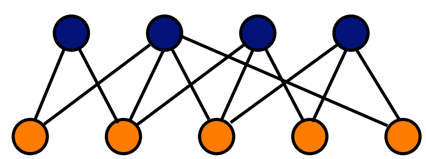
Cycles



Complete Graphs



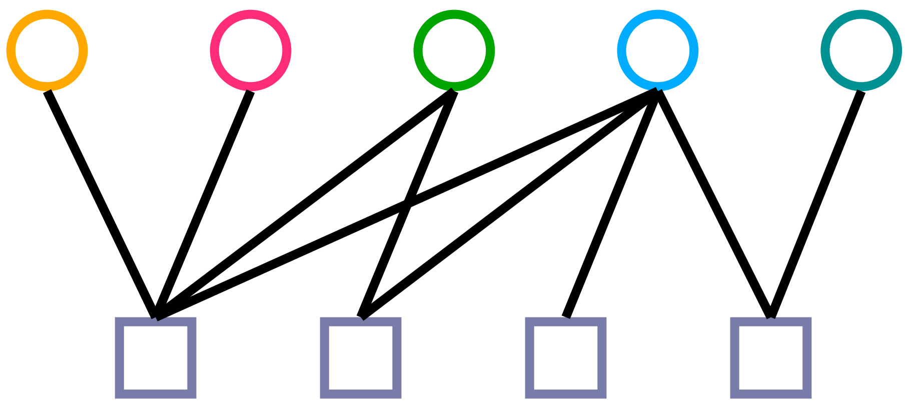
Bipartite Graphs



Bipartite (two-mode) networks

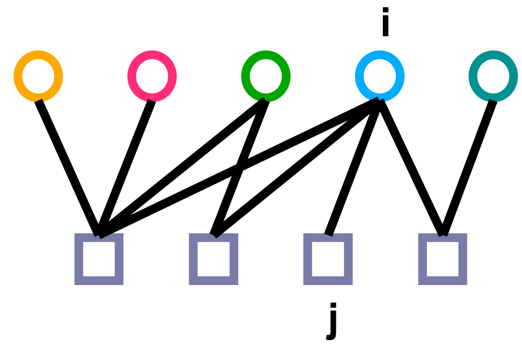
- edges occur only between two groups of nodes, not within those groups

- for example, we may have individuals and *events*
  - directors and boards of directors
  - customers and the items they purchase
  - metabolites and the reactions they participate in



in matrix notation

- B_{ij}
  - = 1 if node i from the first group
    links to node j from the second group
  - = 0 otherwise

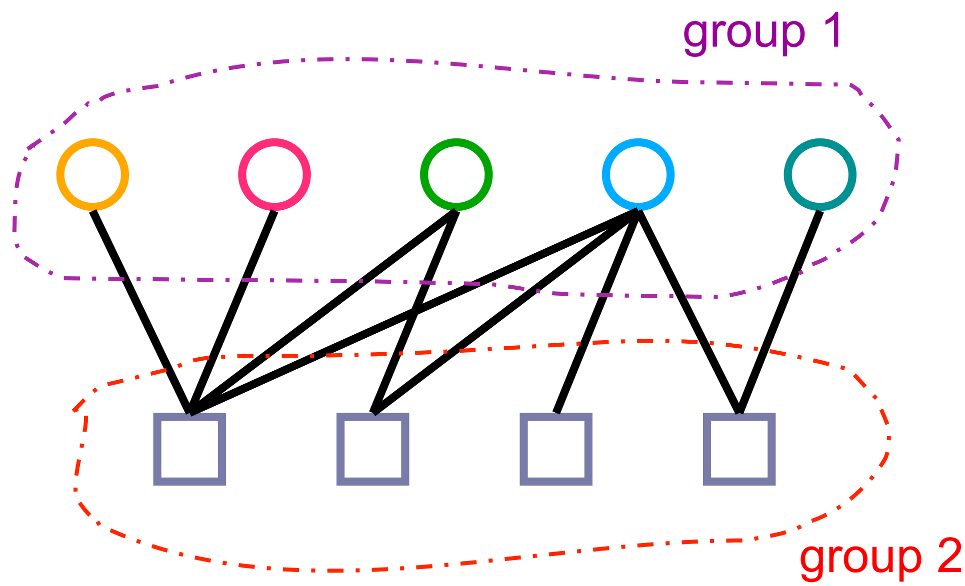


- B is usually not a square matrix!
  - for example: we have n customers and m products

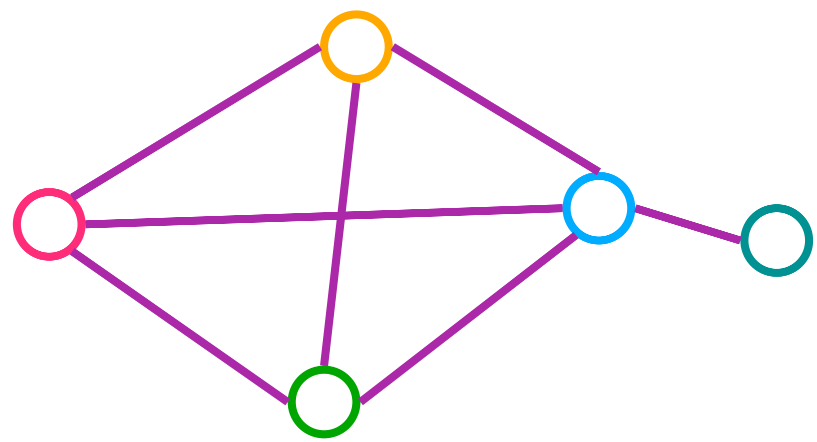
$$B = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 1 & 1 & 1 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

going from a bipartite to a one-mode graph

- Two-mode network



- One mode projection
  - two nodes from the first group are connected if they link to the same node in the second group
  - naturally high occurrence of cliques
  - some loss of information
  - Can use weighted edges to preserve group occurrences



Our SWN examples are from bipartite collapse

- The Oracle of Bacon – The path to Kevin Bacon
  http://oracleofbacon.org/

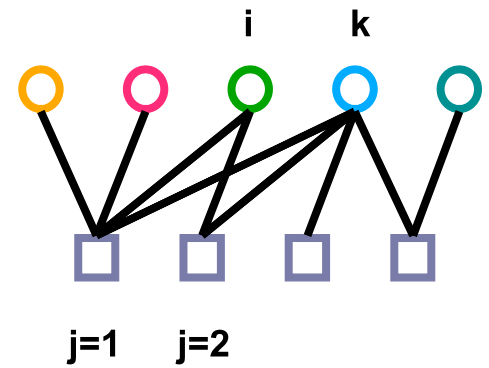
- SixDegrees.org – connecting causes and celebrities
  http://www.sixdegrees.org/

- Six degrees of Kevin Garnett
  http://www.slate.com/articles/sports/slate_labs/2013/10/six_degrees_of_kevin_garnett_connect_any_two_athletes_who_ve_ever_played.html

- Six degrees of NBA seperation
  http://harvardsportsanalysis.wordpress.com/2011/03/04/six-degrees-of-nba-separation/
  (Blog post explaining use of Dijkstra's algorithm)

Collapsing to a one-mode network

- i and k are linked if they both link to j
- $P_{ij} = \sum_k B_{ki} B_{kj}$
- $P' = B B^T$
  - the transpose of a matrix swaps B_{xy} and B_{yx}
  - if B is an *nxm* matrix, B^T is an *mxn* matrix



$$B = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 1 & 1 & 1 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

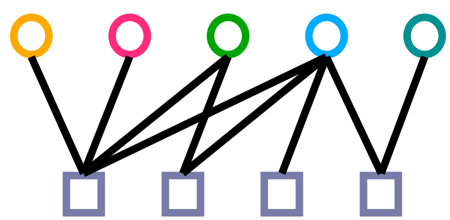
$$B^T = \begin{pmatrix} 1 & 1 & 1 & 1 & 0 \\ 0 & 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 & 1 \end{pmatrix}$$

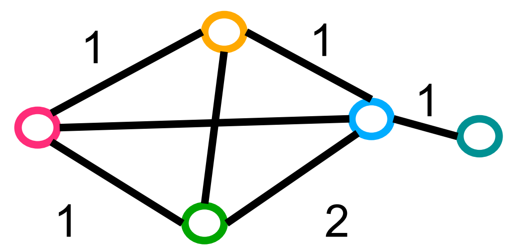
Matrix multiplication

- general formula for matrix multiplication $Z_{ij} = \sum_k X_{ik} Y_{kj}$
- let $Z = P'$, $X = B$, $Y = B^T$

$$P' = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 1 & 1 & 1 \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 1 & 1 & 1 & 0 \\ 0 & 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 & 1 \end{pmatrix} = \begin{pmatrix} 1 & 1 & 1 & 1 & 0 \\ 1 & 1 & 1 & 1 & 0 \\ 1 & 1 & 2 & 2 & 0 \\ 1 & 1 & 2 & 4 & 1 \\ 0 & 0 & 0 & 1 & 1 \end{pmatrix}$$

$$\begin{bmatrix} 1 & 1 & 1 & 1 \end{bmatrix} \begin{pmatrix} 1 \\ 1 \\ 0 \\ 0 \end{pmatrix} = 1*1 + 1*1 + 1*0 + 1*0 = 2$$

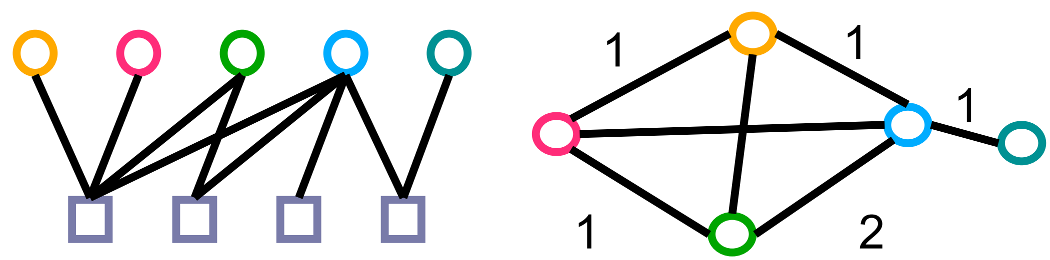




Collapsing a two-mode network to a one mode-network

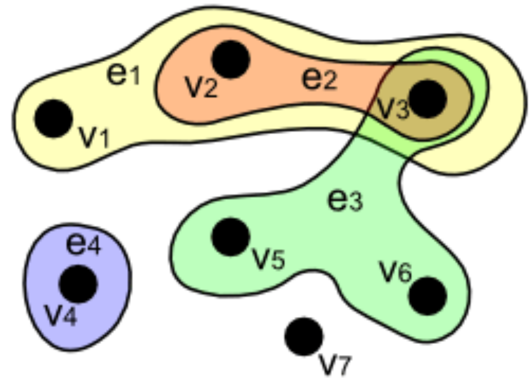
- Assume the nodes in group 1 are people and the nodes in group 2 are movies
- P' is symmetric
- The diagonal entries of P' give the number of movies each person has seen
- The off-diagonal elements of P' give the number of movies that both people have seen

$$P' = \begin{pmatrix} 1 & 1 & 1 & 1 & 0 \\ 1 & 1 & 1 & 1 & 0 \\ 1 & 1 & 2 & 2 & 0 \\ 1 & 1 & 2 & 4 & 1 \\ 0 & 0 & 0 & 1 & 1 \end{pmatrix}$$

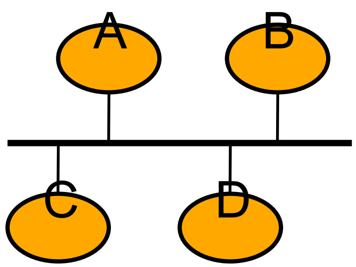


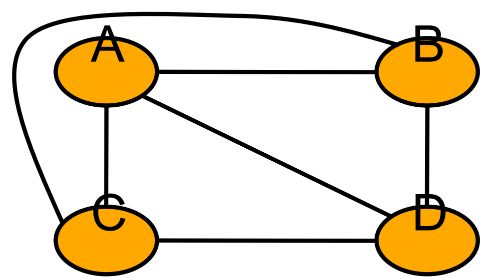
HyperGraphs

- Edges join more than two nodes at a time (*hyperEdge*)
- Affliation networks



- Examples
  - Families
  - Subnetworks





Hypergraphs — beyond dyadic iteractions

In ways, good models of social networks.

- “Complex Networks as Hypergraphs”
  Ernesto Estrada, Juan A. Rodriguez-Velazquez
  arXiv:physics/0505137, 2005.

- “Random hypergraphs and their applications”,
  G Ghoshal, V Zlatić, G Caldarelli, MEJ Newman,
  Physical Review E 79 (6), 2009.

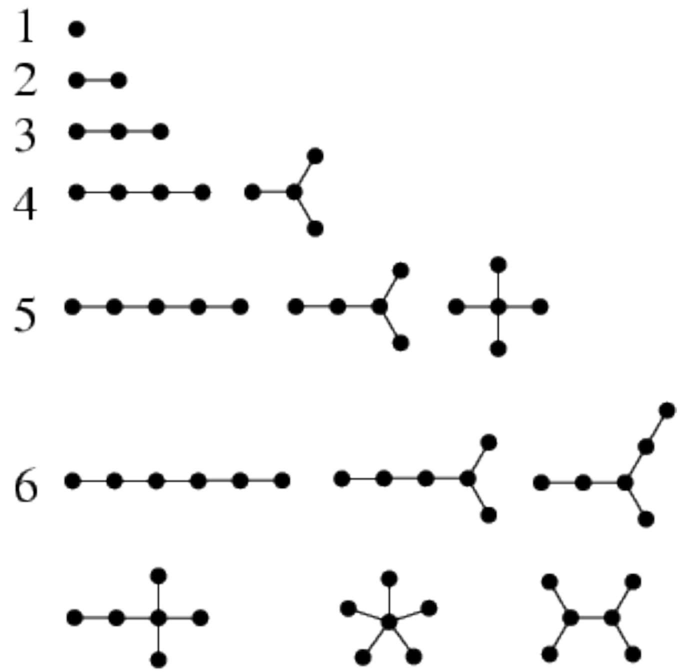
- Ramanathan, R., et al. “Beyond graphs: Capturing groups in networks.”
  NetSciCom, 2011 IEEE Conference, 2011.

- “Information Flows: A Critique of Transfer Entropies”
  Ryan G. James, Nix Barnett, James P. Crutchfield
  Accepted to Physical Review Letters, April 12, 2016.

- See also hypergraph mining, hypergraph learning algorithms, overlapping communities, link prediction...

Trees

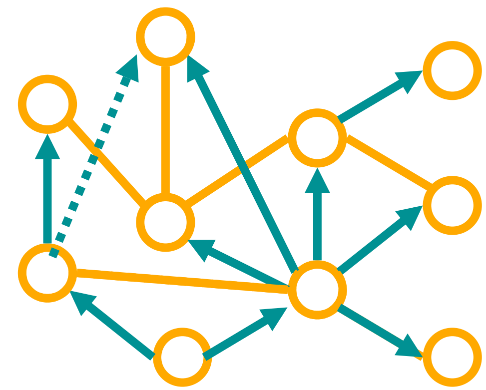
- Trees are undirected graphs that contain no cycles



- For n nodes, number of edges m = n-1
- Any node can be dedicated as the root

examples of trees

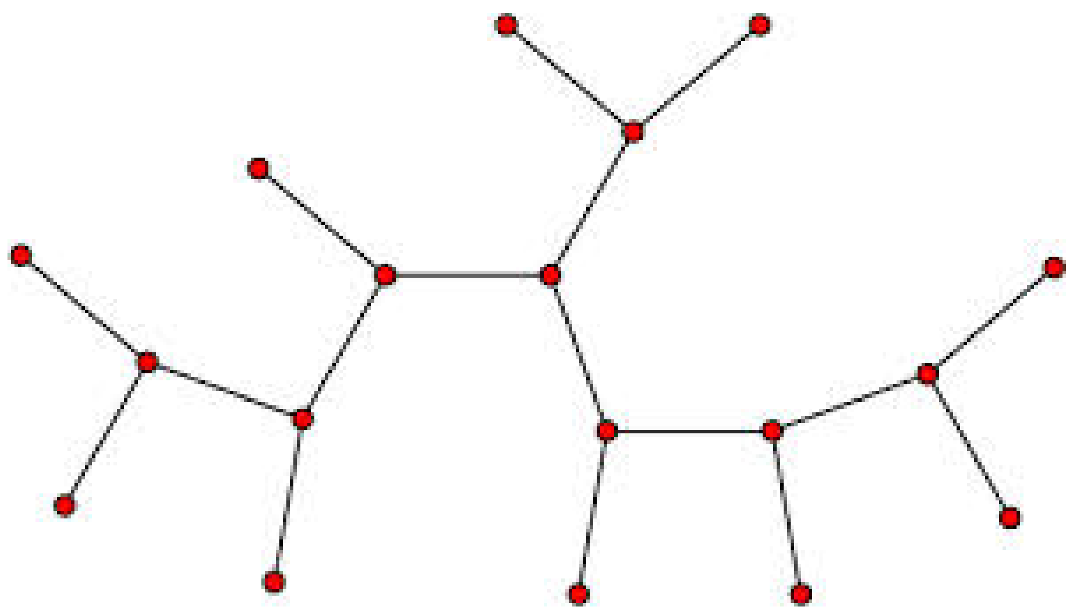
- In nature
  - trees
  - river networks
  - arteries (or veins, but not both)
- Man made
  - sewer system
- Computer science
  - binary search trees
  - decision trees (AI)
- Network analysis
  - minimum spanning trees
    - from one node – how to reach all other nodes most quickly
    - may not be unique, because shortest paths are not always unique
    - depends on weight of edges



Searching on a tree

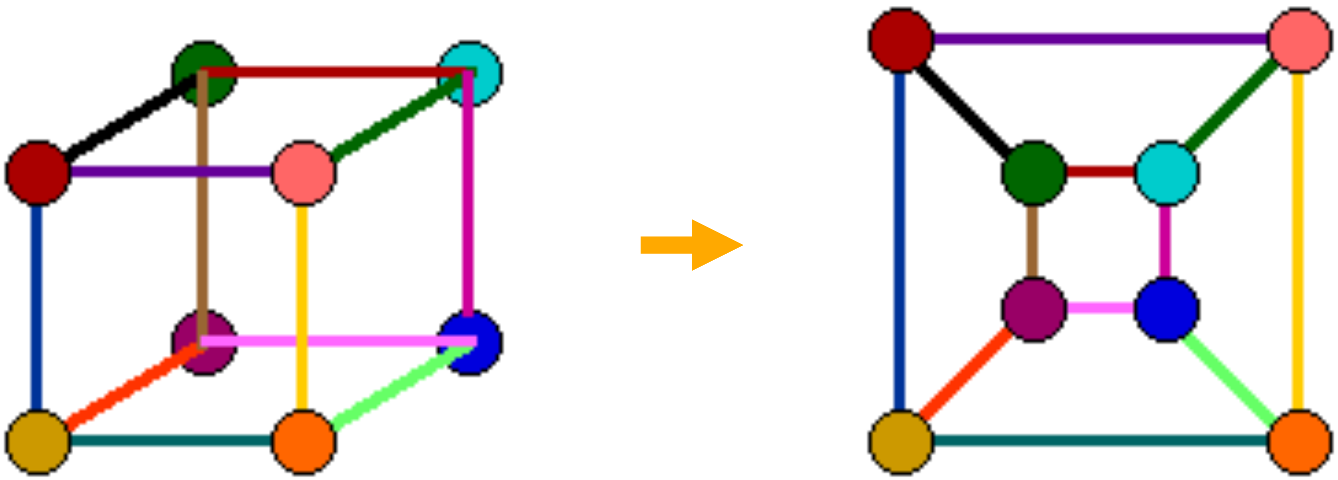
- Breadth first search :
  explore all the neighbors first

- Depth first search :
  take a step out in hop-count each iteration

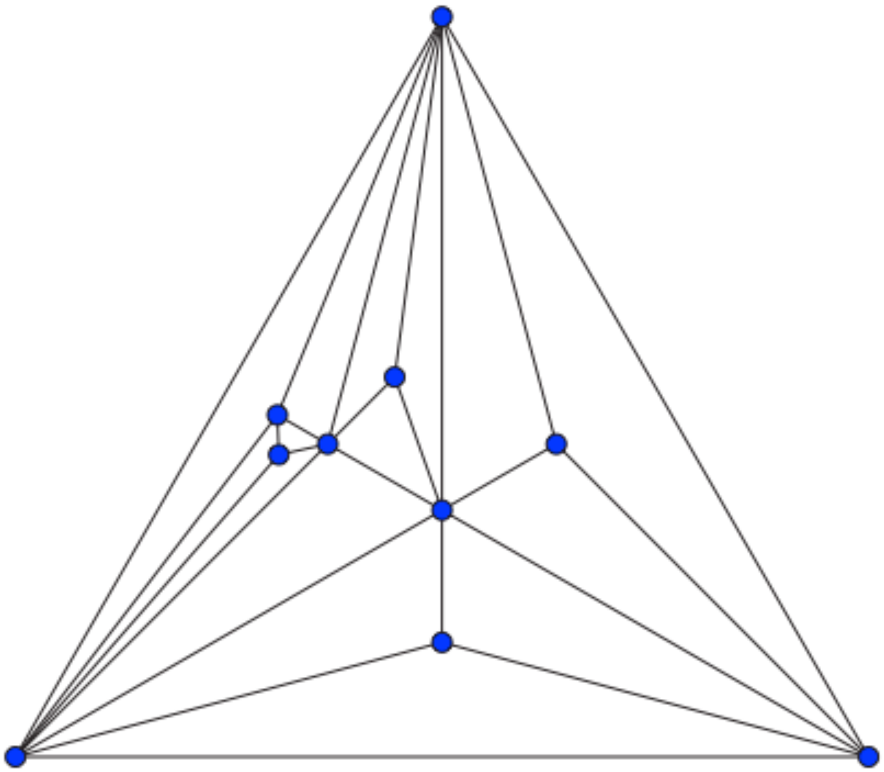


Planar graphs

- A graph is planar if it can be drawn on a plane without any edges crossing



Apollonian network

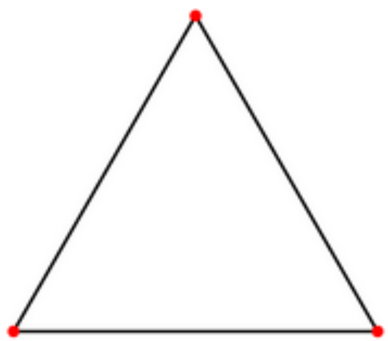


- An undirected graph formed by a process of recursively subdividing a randomly selected triangle into three smaller triangles.

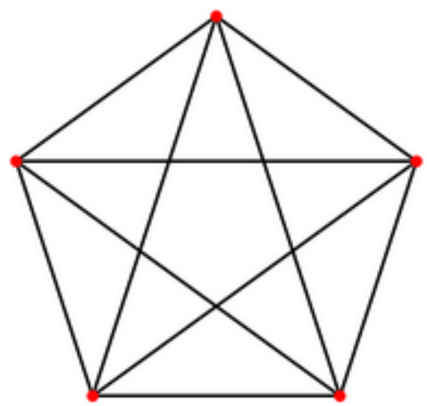
- A planar graph with power law degree distribution, and small world property.

Cliques and complete graphs

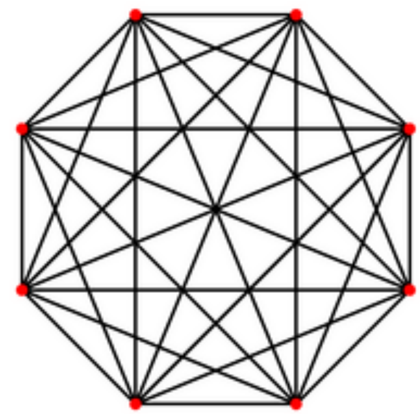
- K_n is the complete graph (clique) with K vertices
  - each vertex is connected to every other vertex
  - there are n*(n-1)/2 undirected edges



K_3

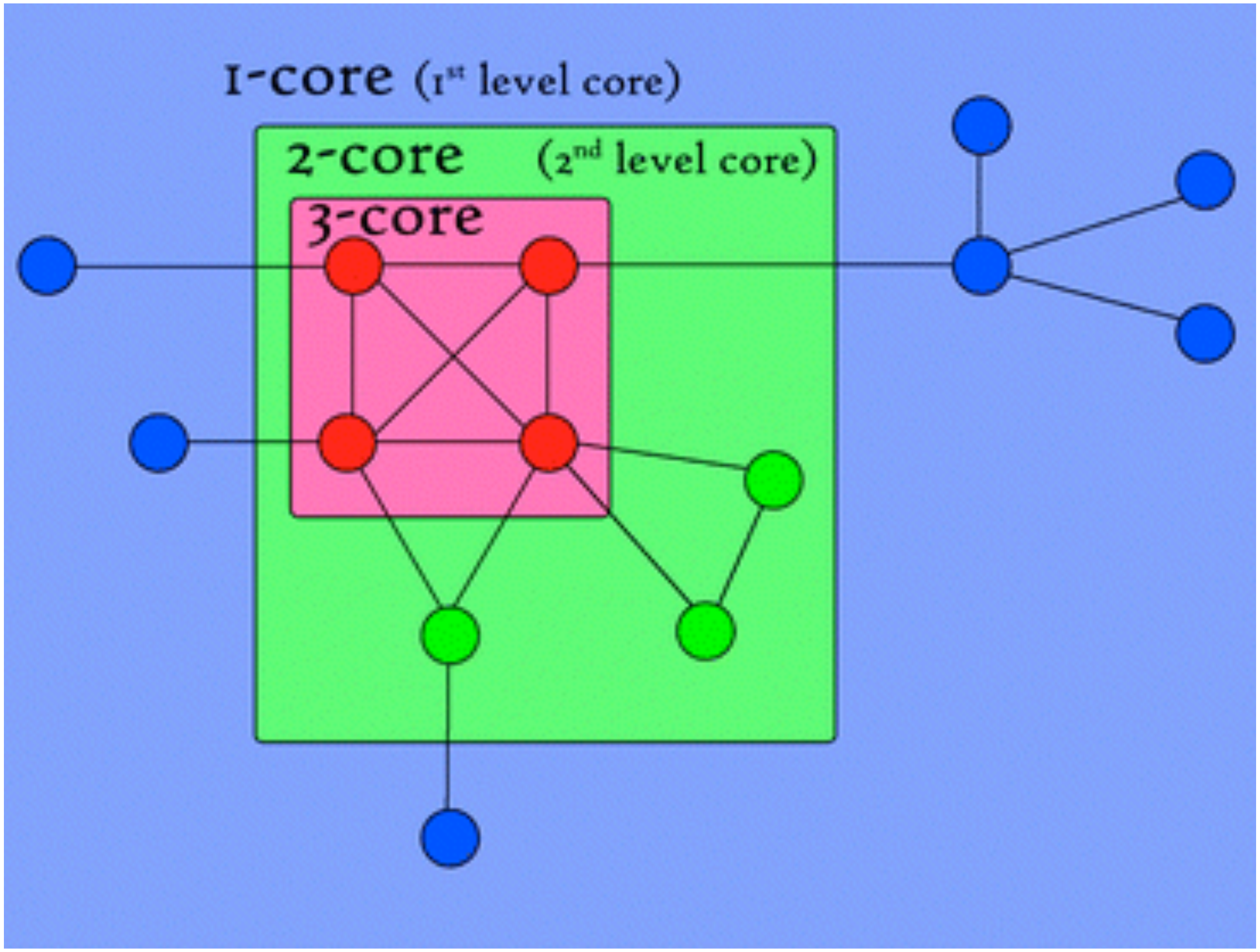


K_5

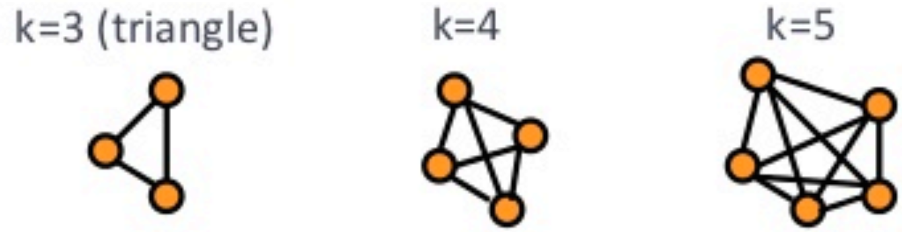


K_8

The k-core and k-shell

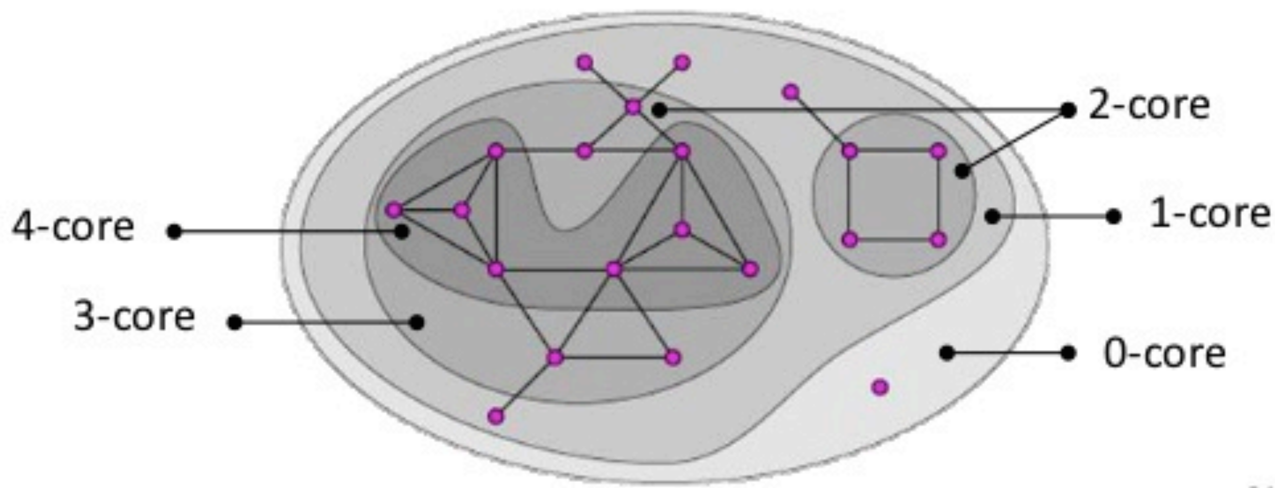


- k-clique



- N-clique

- k-core



k-core decomposition

- For visualization

- k-core decomposition of the Internet
  - router level
  - AS level
  - e.g. Carmi et. al. PNAS 2007.

  "A model of Internet topology using k-shell decomposition"
  A nucleus, a fractal layer, and tendrils.

- in random graphs and statistical physics:

  "K-core organization of complex networks"
  SN Dorogovtsev, AV Goltsev, JFF Mendes
  Physical review letters, 2006.