Algorithmic and Economic Aspects of Networks

Nicole Immorlica

Syllabus

1. Jan. 8th (today): Graph theory, network structure
2. Jan. 15th: Random graphs, probabilistic network formation
3. Jan. 20th: Epidemics
4. Feb. 3rd: Search
5. Feb. 5th: Game theory, strategic network formation
6. Feb. 12th: Diffusion
7. Feb. 19th: Learning
8. Feb. 26th: Markets
9. Mar. 5th: TBA
10. Mar. 12th: Final project presentations

Assignments

- Readings, weekly
  - From *Social and Economic Networks*, by Jackson
  - Research papers, one per week
- Reaction papers, weekly
- Class presentations
- Two problem sets (?)
- Final projects, end of term
  - Survey paper
  - Theoretical/empiracle analysis

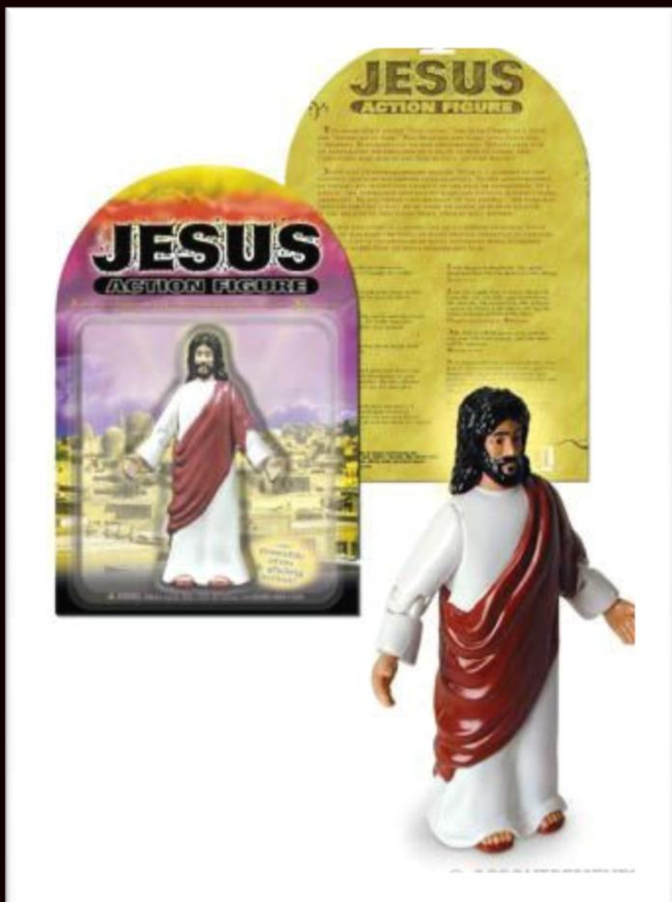
Grading (Approximate)

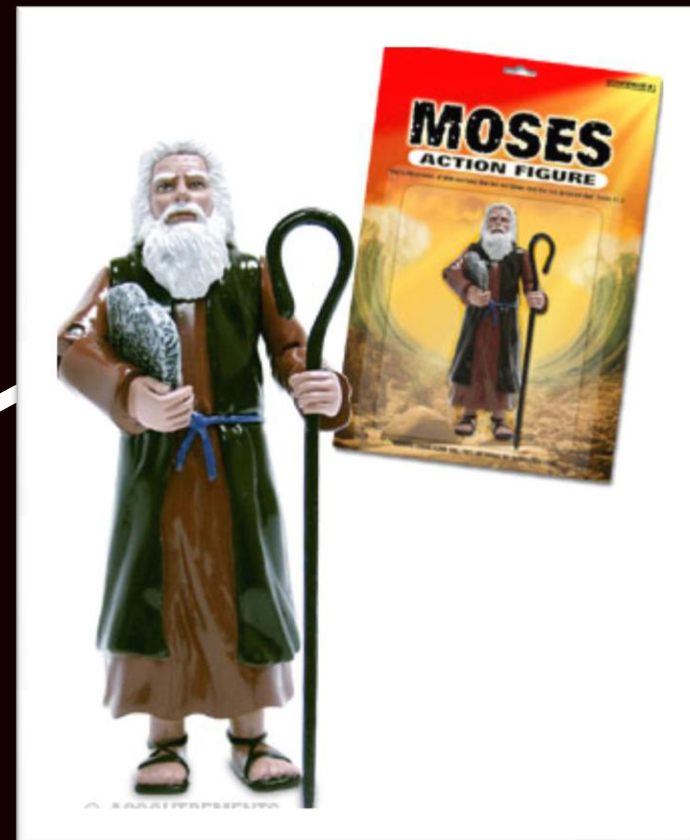
- Participation, class presentations [15%]
- Reaction papers [25%]
- Problem sets [20%]
- Final project [40%]

Networks

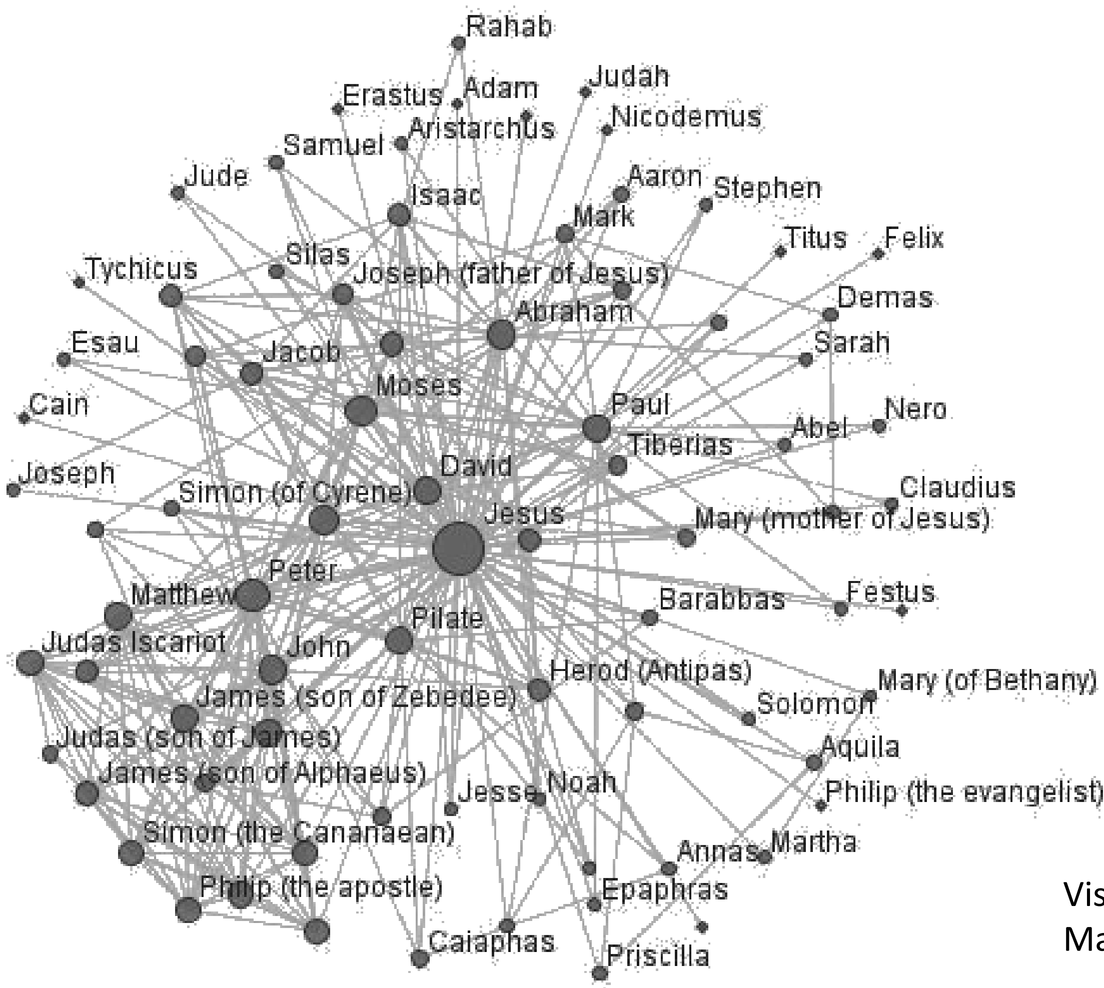
- A network is a graph that represents *relationships* between independent *entities*.
  - Graph of friendships (or in the virtual world, networks like facebook)
  - Graph of scientific collaborations
  - Web graph (links between webpages)
  - Internet: Inter/Intra-domain graph

New Testament Social Network





New Testament Social Network



Visualization from ManyEyes

United Routes Network

Seattle

Honolulu

United Routes Network

Erdos Collaboration Network

Harry Buhrman

One-Sided Versus Two-Sided Error in Probabilistic Computation

Harry Buhrman[1*] and Lance Fortnow[2**]

[1] CWI, PO Box 94079, 1090 GB Amsterdam, The Netherlands,
buhrman@cwi.nl,
http://www.cwi.nl/~buhrman

[2] University of Chicago, Department of Computer Science,
1100 E. 58th St., Chicago, IL 60637
fortnow@cs.uchicago.edu,
http://www.cs.uchicago.edu/~fortnow

Abstract. We demonstrate how to use Lautemann's proof that BPP is in Σ_2^p to exhibit that BPP is in $\mathrm{RP}^{\mathbf{PromiseRP}}$. Immediate consequences show that if PromiseRP is easy or if there exist quick hitting set generators then P = BPP. Our proof vastly simplifies the proofs of the later result due to Andreev, Clementi and Rolim and Andreev, Clementi, Rolim and Trevisan.
Clementi, Rolim and Trevisan question whether the promise is necessary for the above results, i.e., whether $\mathrm{BPP} \subseteq \mathrm{RP}^{\mathbf{RP}}$ for instance. We give a relativized world where $\mathrm{P} = \mathrm{RP} \neq \mathrm{BPP}$ and thus the promise is indeed needed.

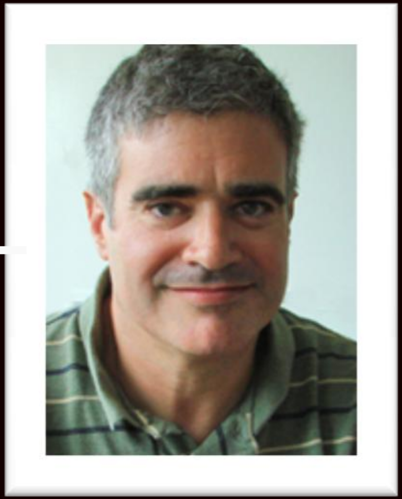
1 Introduction

Andreev, Clementi and Rolim [ACR98] show how given access to a quick hitting set generator, one can approximate the size of easily describable sets. As an immediate consequence one gets that if quick hitting set generators exist then $\mathbf{P} = \mathbf{BPP}$. Andreev, Clementi, Rolim and Trevisan [ACRT97] simplify the proof and apply the result to simulating $\mathbf{BPP}$ with weak random sources.

Much earlier, Lautemann [Lau83] gave a proof that $\mathbf{BPP} \subseteq \Sigma_2^p = \mathbf{NP}^{\mathbf{NP}}$, simplifying work of Gács and Sipser [Sip83]. Lautemann's proof uses two simple applications of the probabilistic method to get the existence results needed. As often with the case of the probabilistic method, the proof actually shows that the overwhelming number of possibilities fulfill the needed requirements. With this observation, we show that Lautemann's proof puts $\mathbf{BPP}$ in the class $\mathbf{RP}^{\mathbf{PromiseRP}}$[1]. Since quick hitting set generators derandomize $\mathbf{PromiseRP}$ problems, we get the existence of quick hitting set generators implies $\mathbf{P} = \mathbf{BPP}$. This greatly simplifies the proofs of Andreev, Clementi and Rolim [ACR98] and Andreev, Clementi, Rolim and Trevisan [ACRT97].

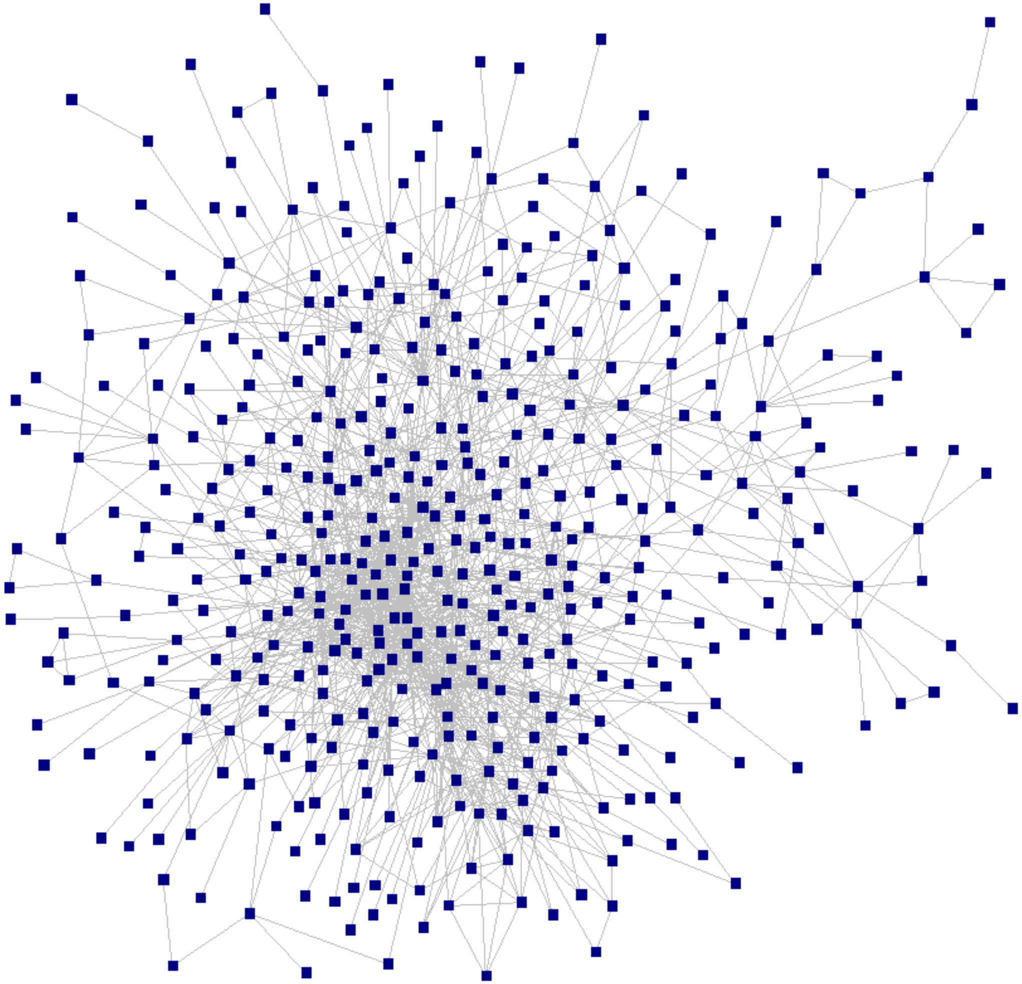
* Partially supported by the European Union through NeuroCOLT ESPRIT Working Group Nr. 8556, and HC&M grant nr. ERB4050PL93-0516.
** Supported in part by NSF grant CCR 92-53582.

C. Meinel and S. Tison (Eds.): STACS'99, LNCS 1563, pp. 100–109, 1999.
© Springer-Verlag Berlin Heidelberg 1999



Lance Fortnow

Erdos Collaboration Network



Visualization from Orgnet

Graph Theory

- Graph $G = (V,E)$
- A set V of n nodes or vertices

$$V = \{i\}$$

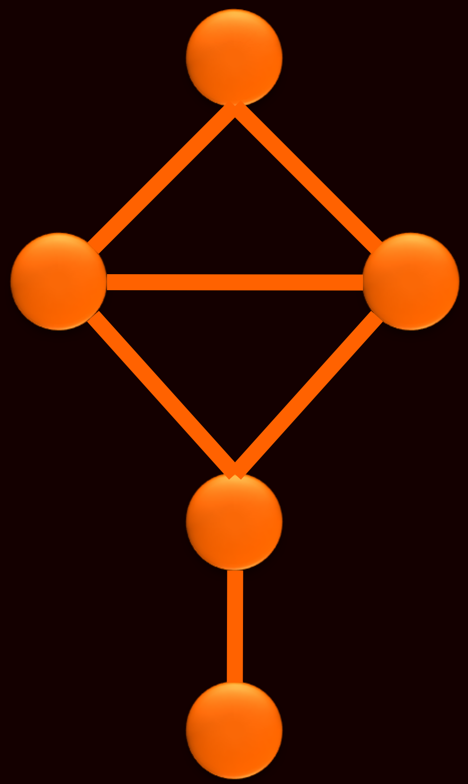
- A subset E of $V \times V$ of m pairs of vertices, called edges

$$E = \{(i,j)\}$$

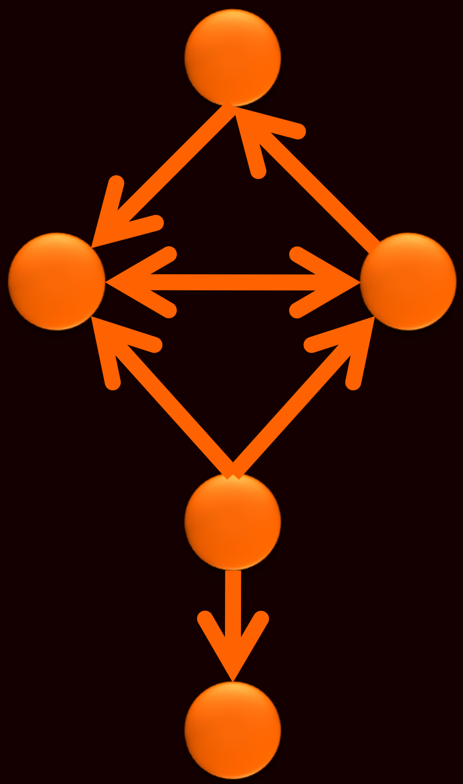
Edges can be directed (pair order matters), or undirected.

Drawing Graphs

Undirected graphs



Directed graphs



Representing Graphs

List of edges

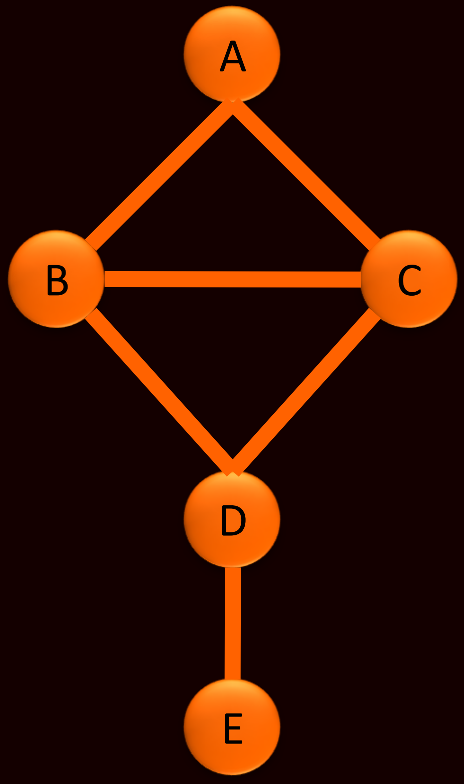
(A,B), (A,C), (B,C), (B,D), (C,D), (D,E)

Node adjacency list

A: B, C; B: A, C; C: A, B; D: B, C, E; E: D

Adjacency matrix – A_{ij} = 1 if (i,j) is an edge, else = 0

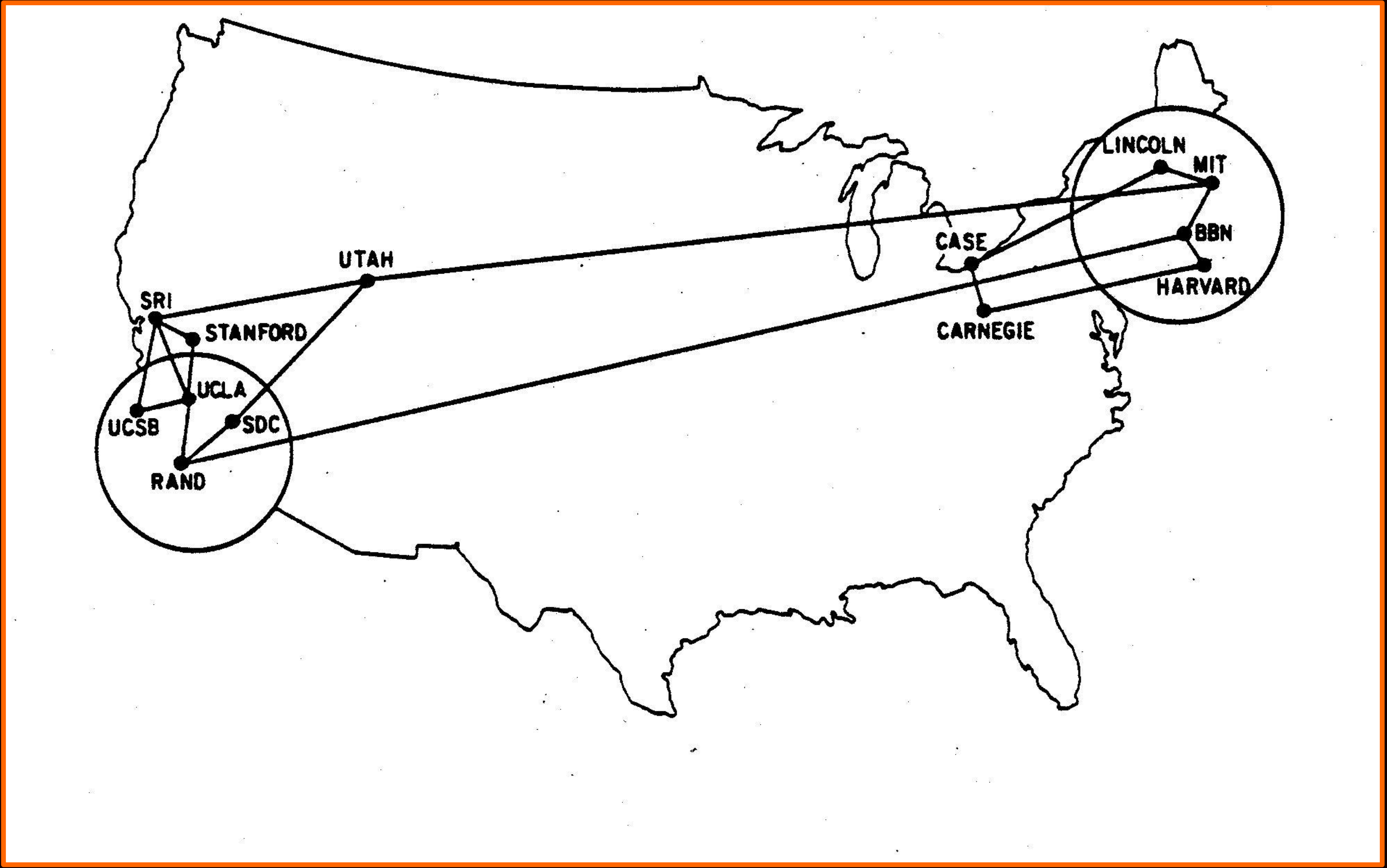
	A	B	C	D	E
A	0	1	1	0	0
B	1	0	1	1	0
C	1	1	0	1	0
D	0	1	1	0	1
E	0	0	0	1	0



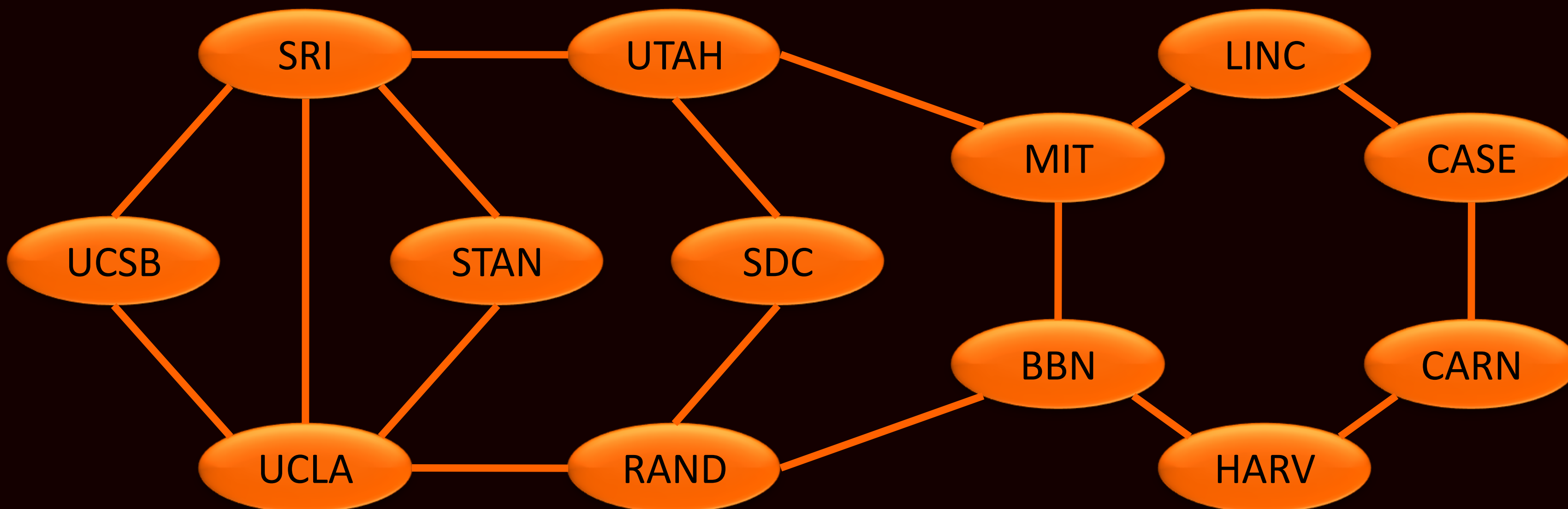
Modeling Networks

symmetric networks = undirected graphs

Example: Arpanet



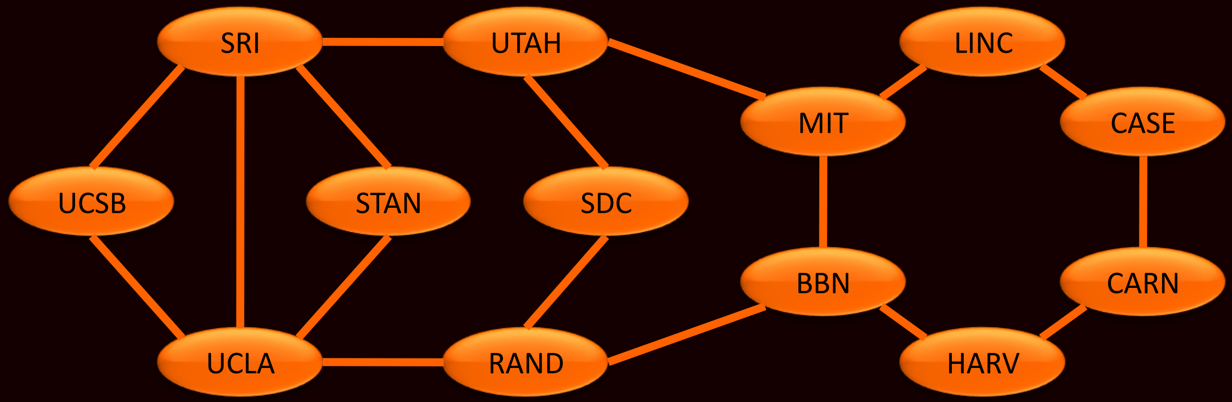
Example: Arpanet



Neighborhoods and Degrees

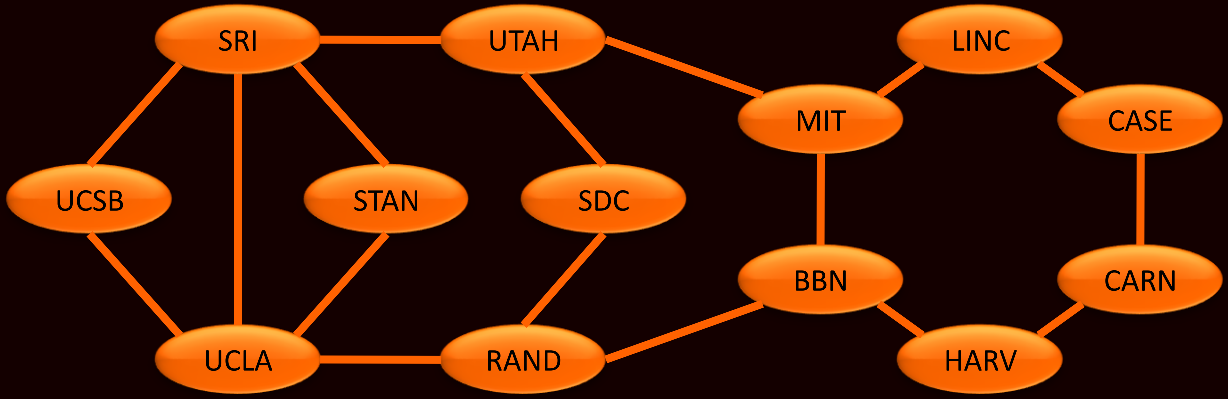
neighborhood of node i = nodes j with edge to i

degree of node i = number of neighbors



Paths and Cycles

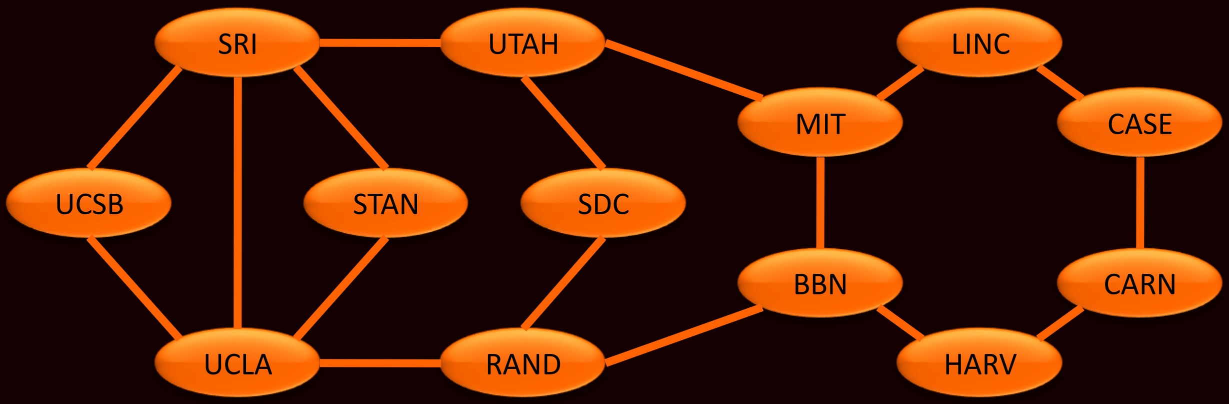
"path" = sequence of nodes such that each consecutive pair is connected



MIT – UTAH – SRI – STAN

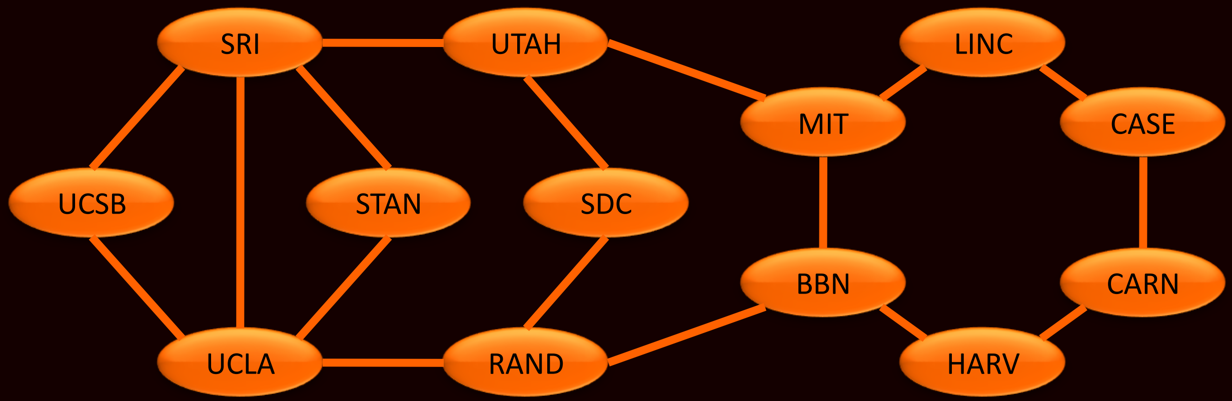
Paths and Cycles

simple path = one that does not repeat nodes



Paths and Cycles

cycle = path that starts and ends at same node

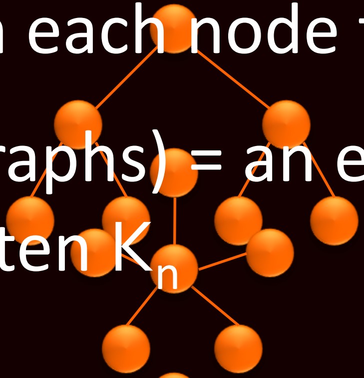


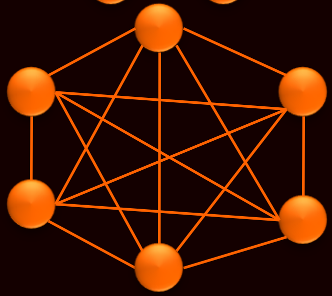
Special Graphs

Trees = a unique path between each pair of vertices

Stars = an edge from each node to center node

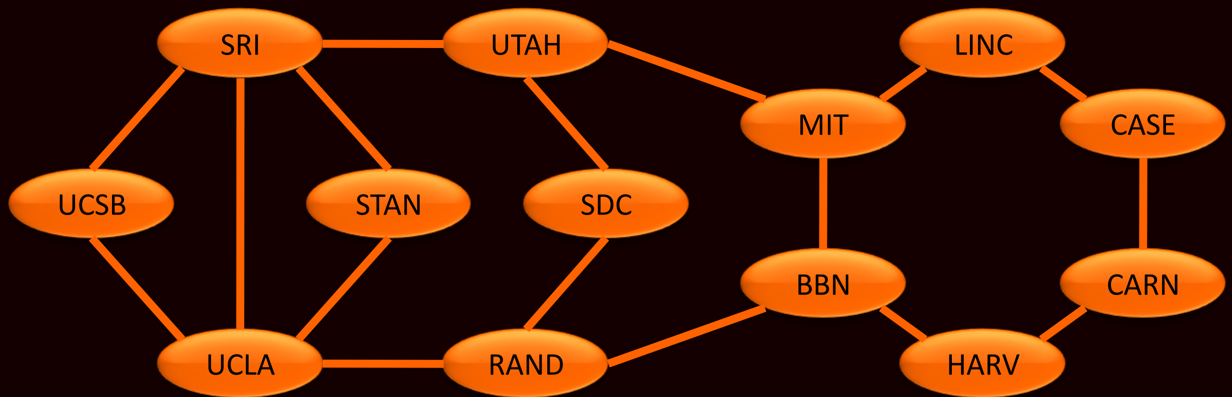
Cliques (complete graphs) = an edge between each pair of vertices, written K_n





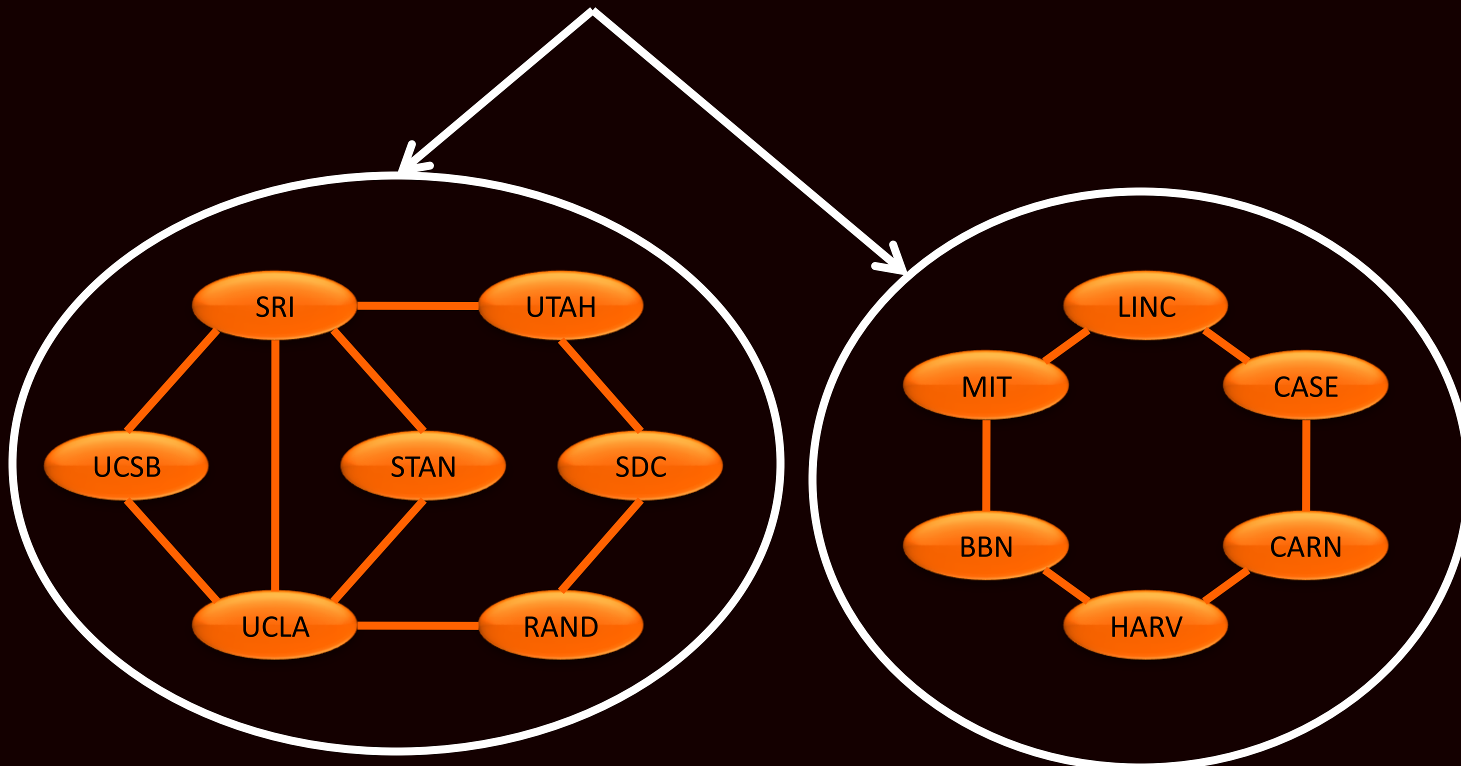
Connectivity

A graph is connected if there is a path between every pair of nodes.



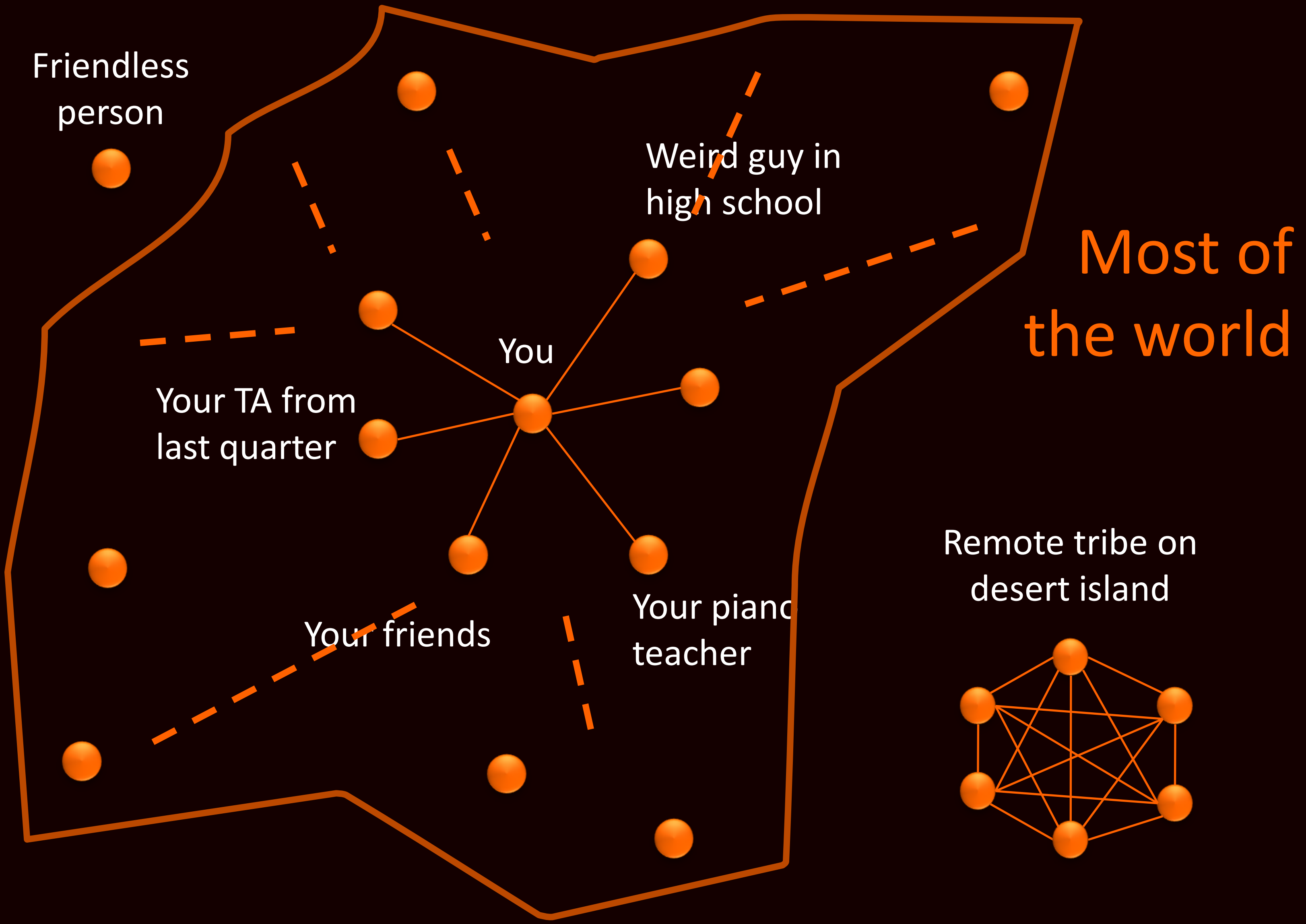
CONNECTED

Connected components



UNCONNECTED

Giant Components



Study Guide

- Graph representations
- Neighbors and degrees
- Paths and cycles
- Connectedness
- Giant component

Network Questions

ENGLISH	MATH
How popular are we?	What is degree dist.?
How connected are we?	What is diameter?
How tight-knit are we?	What is clustering coeff.?
How important am I?	What is centrality?

Degree Distributions

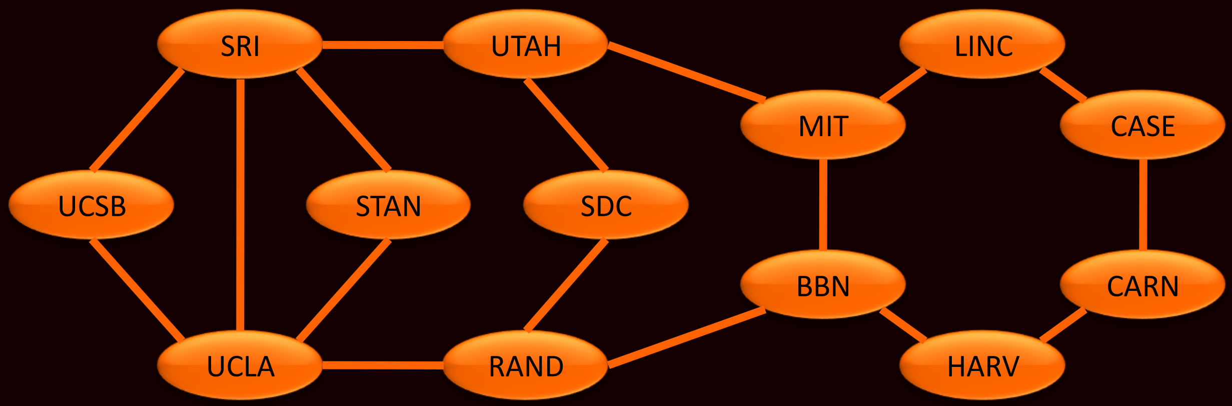
relative frequency of nodes w/different degrees

$P(d)$ = fraction of nodes with degree d

$P(d)$ = probability random node has deg. d

Degree Distributions

P(d) = fraction of nodes with degree d



P(0) = 0, P(1) = 0, P(2) = 7/13, P(3) = 4/13, P(4) = 2/13

Degree Distributions

Degree Distributions

Some special degree distributions

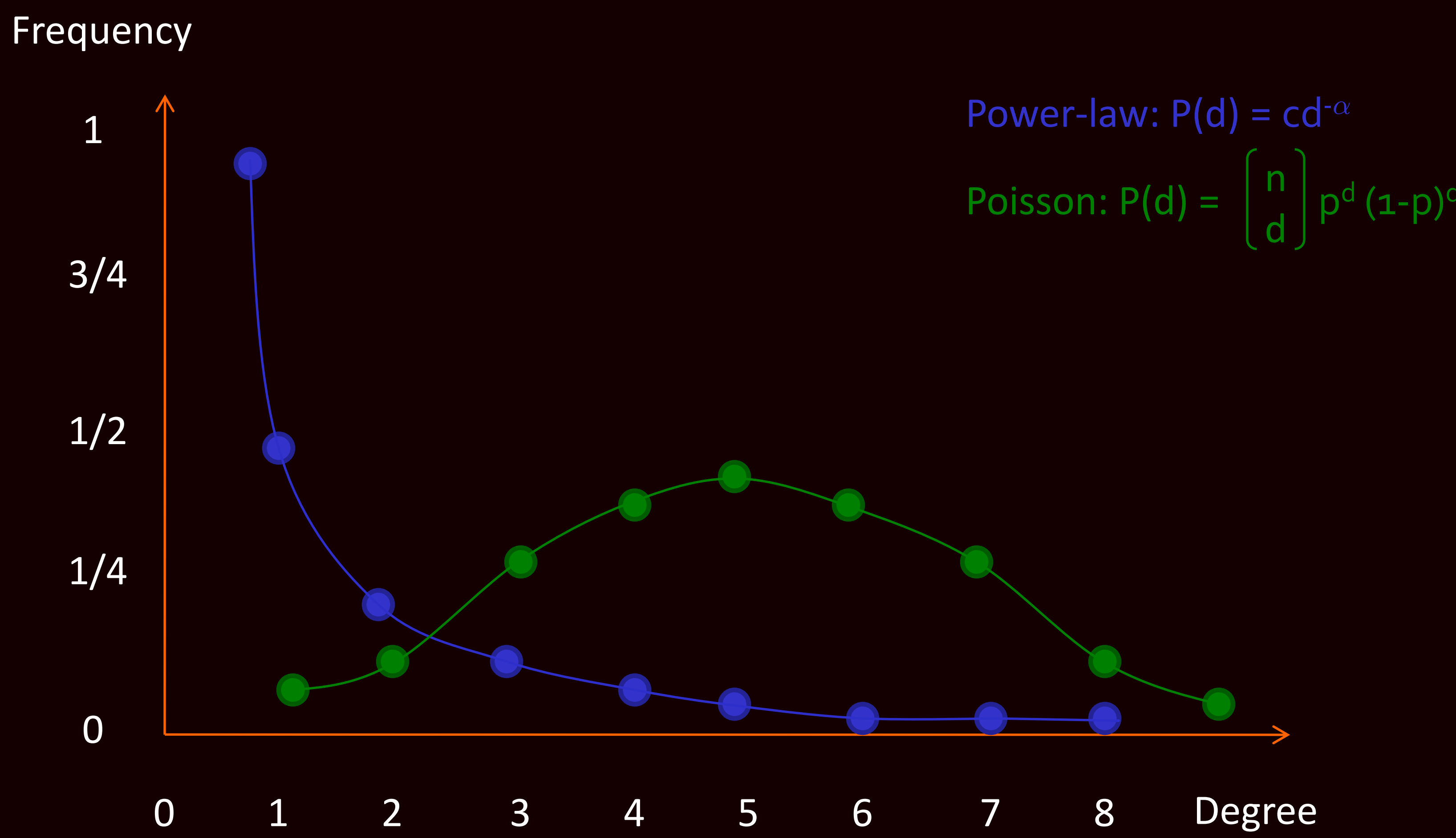
- Poisson degree distribution:

$$P(d) = \binom{n}{d} p^d (1-p)^d$$

for some $0 < p < 1$.

- Scale-free (power-law) degree dist.: $P(d) = cd^{-\alpha}$

Poisson vs Power-law



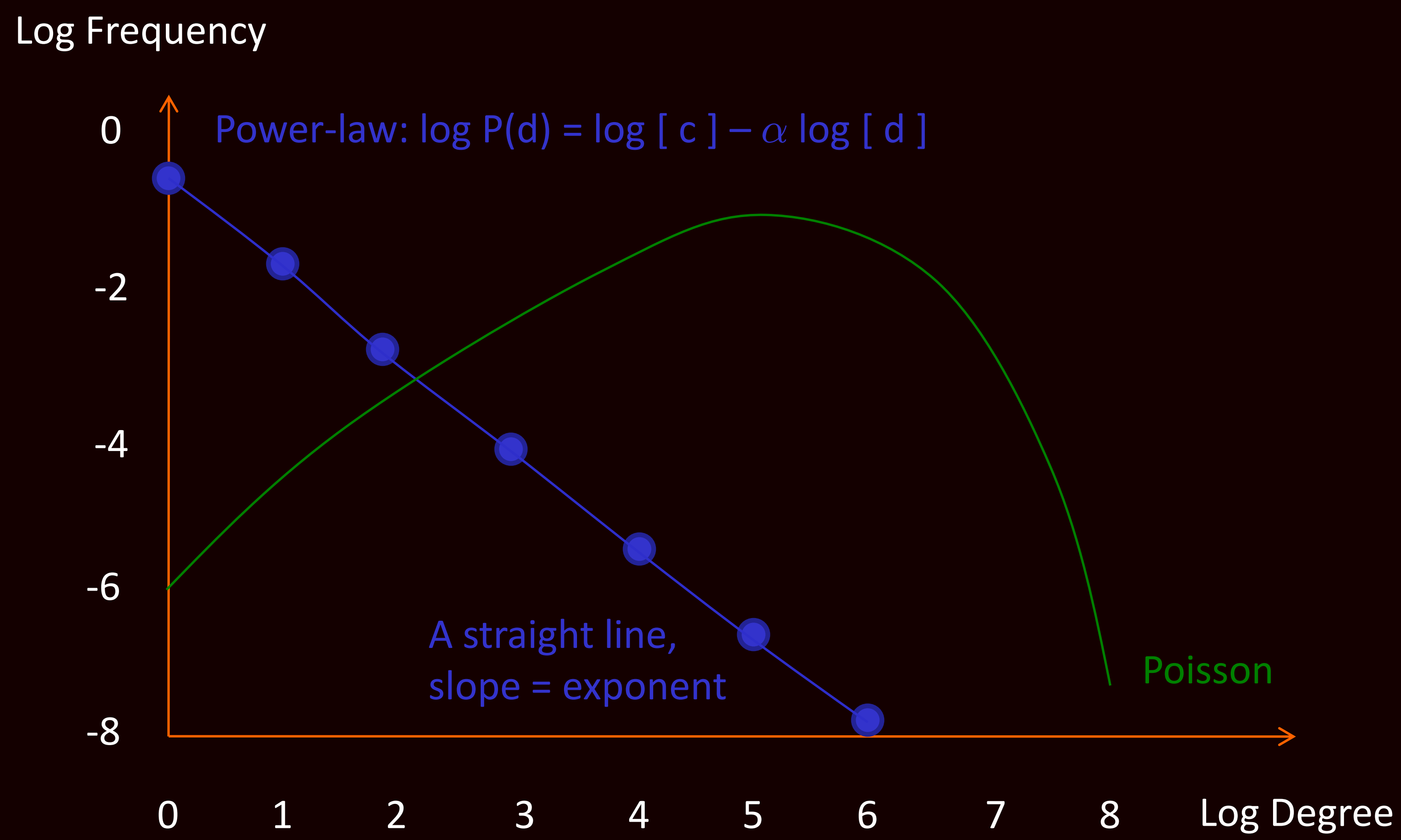
Log-log plots

Power-law: $P(d) = cd^{-\alpha}$

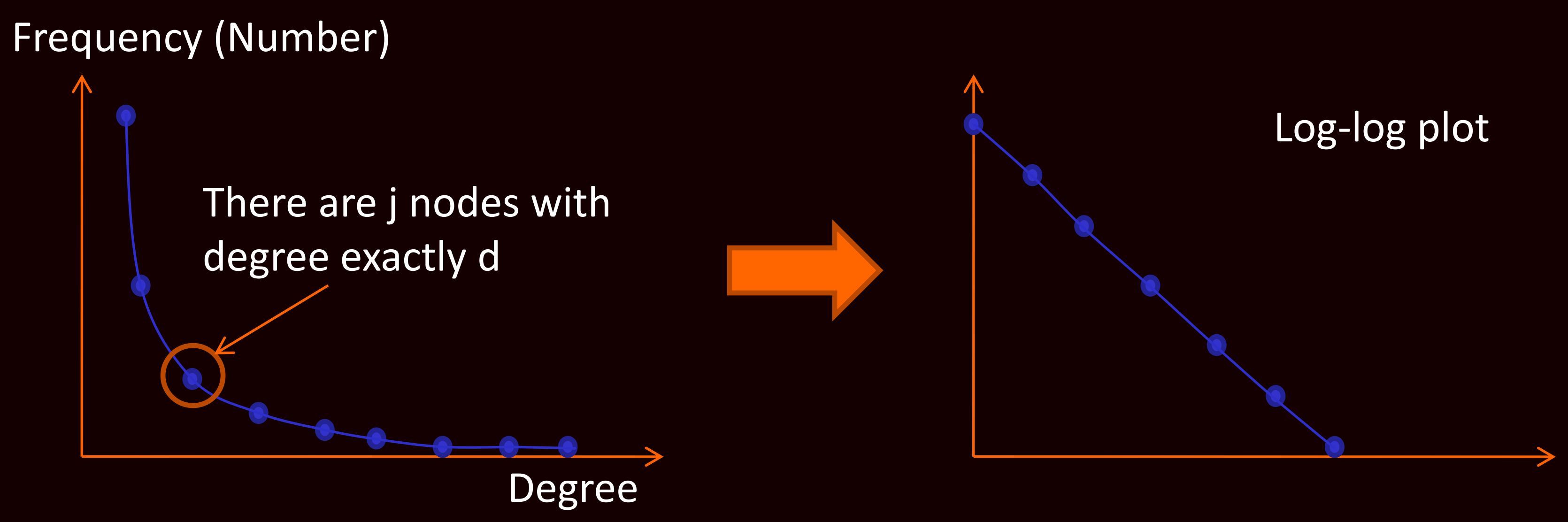
$$\log [\, P(d) \,] = \log [\, cd^{-\alpha} \,]$$
$$= \log [\, c \,] - \alpha \log [\, d \,]$$

So a straight line on a log-log plot.

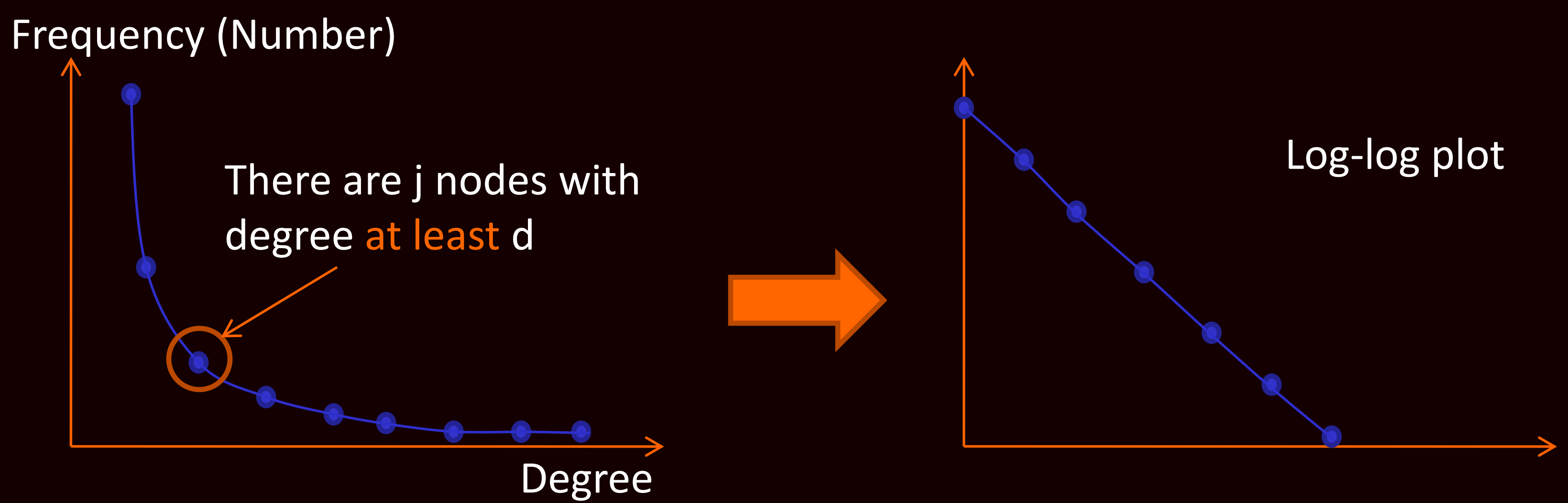
Log-log plots



Alternate views

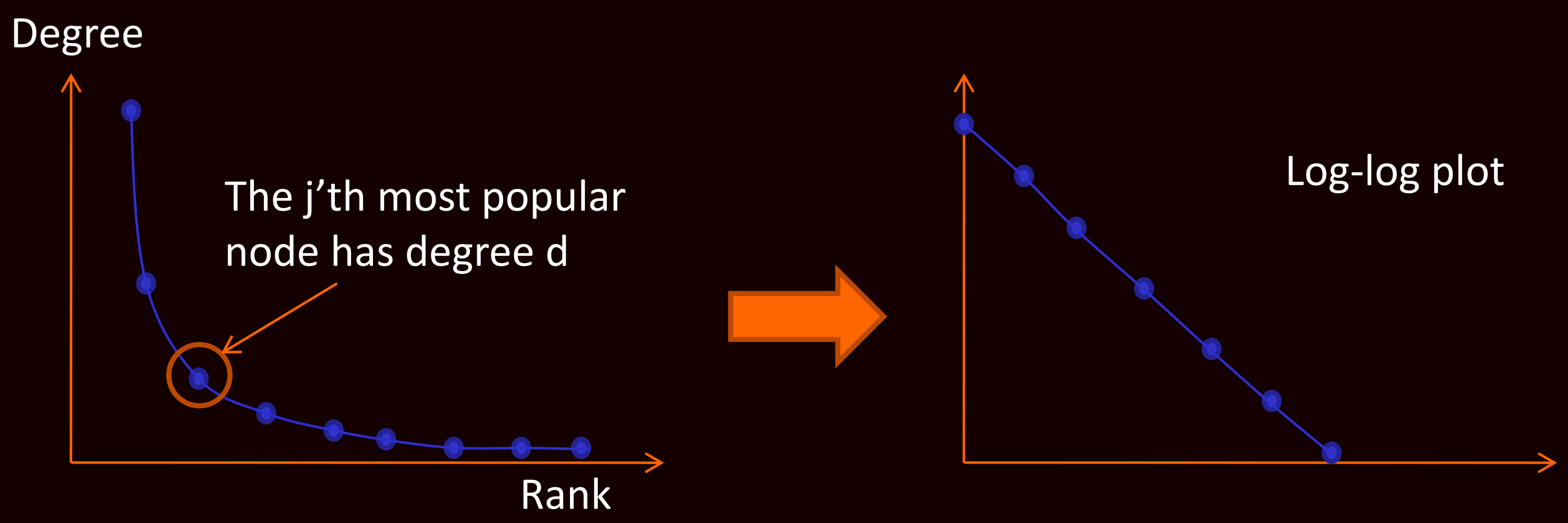


Alternate views



Cumulative Distribution

Alternate views



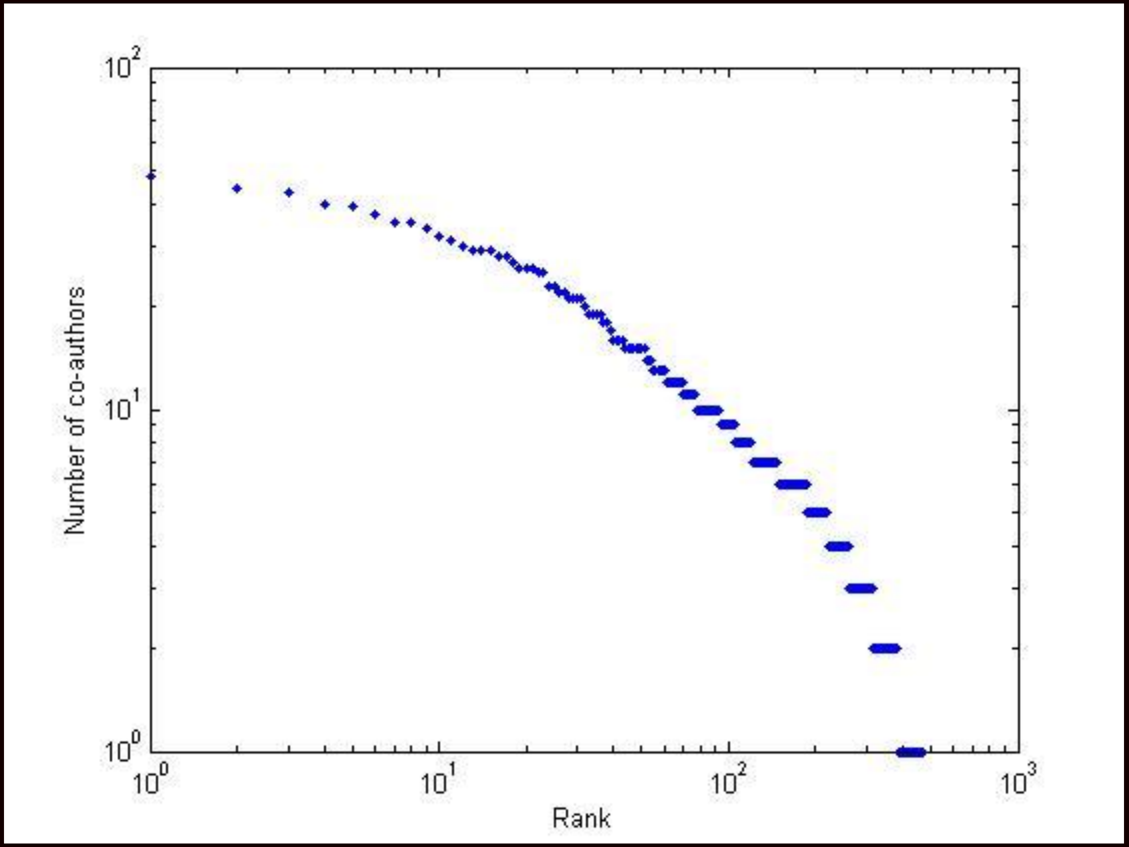
Example: Collaboration Graph

- Power law exp:

  c = 2.97

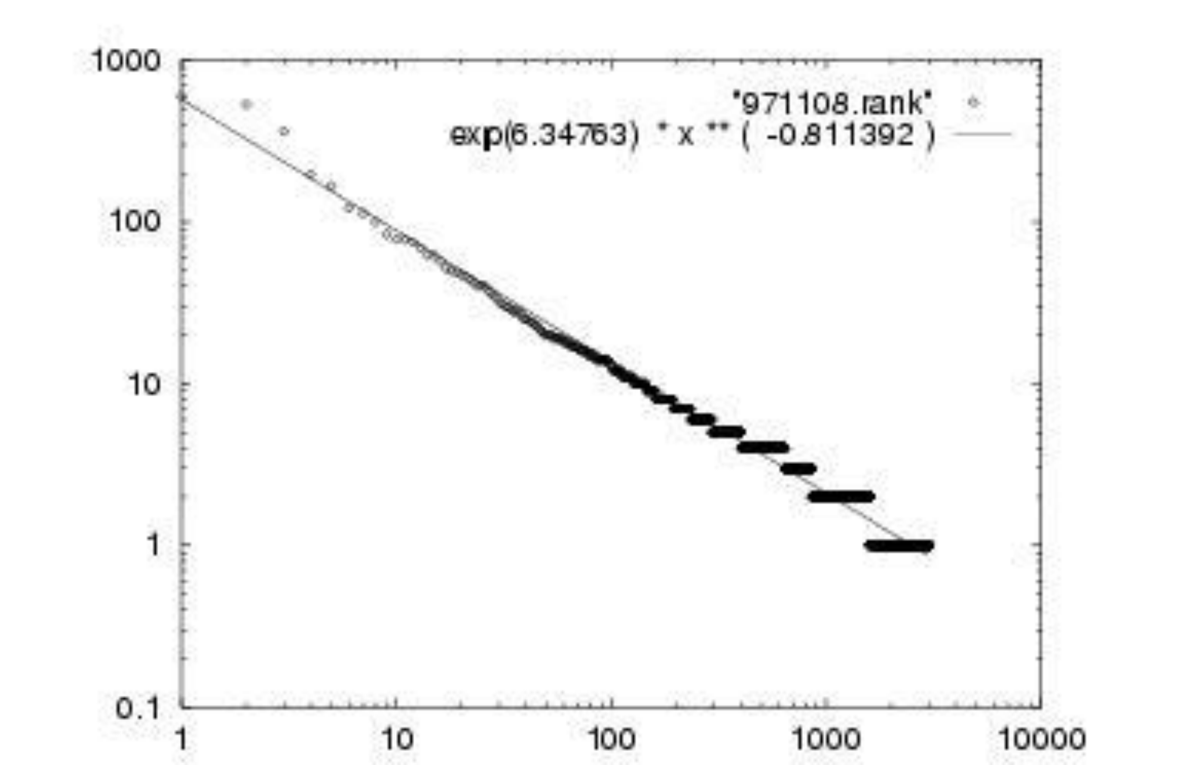
- With exponential decay factor,

  c = 2.46



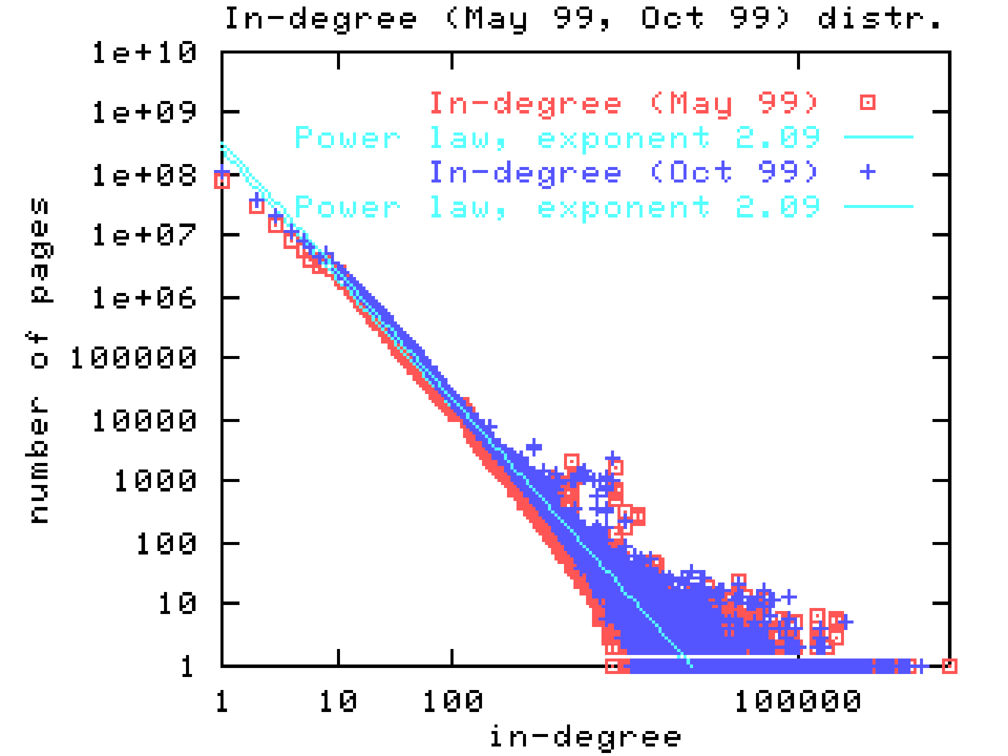
Example: Inter-Domain Internet

- Power law exponent: 2.15 < c < 2.2



Example: Web Graph In-Degree

- Power law exponent: c = 2.09



Q1. How popular are we?

Many social networks have
power-law degree distributions.

A few very popular people,
many many unpopular people.

Diameter

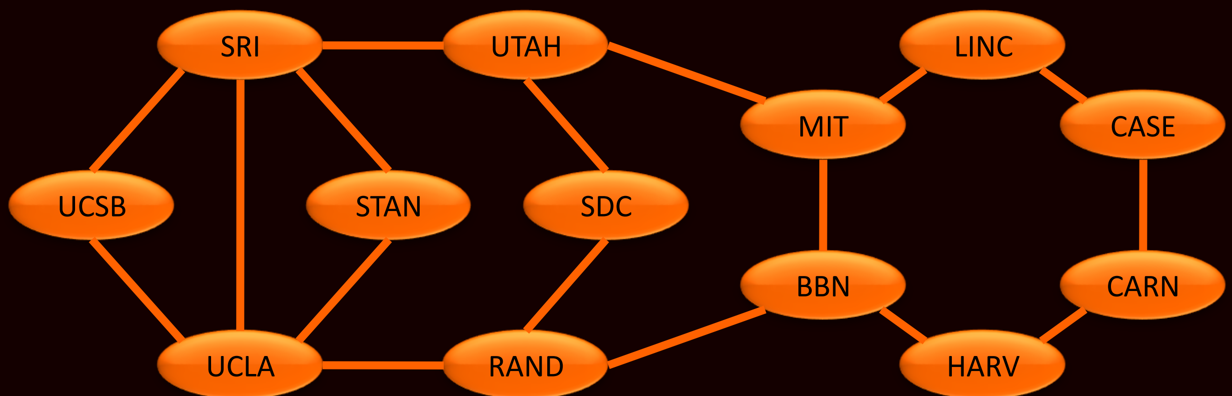
How far apart are the nodes of a graph?

How far apart are nodes i and j?

What is the length of a path from i to j?

Length of a Path

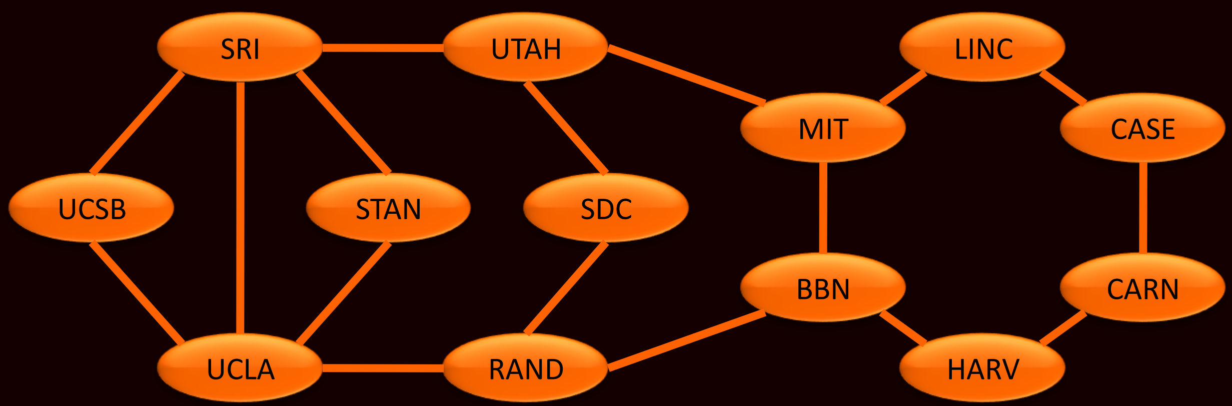
The length of a path is # of edges it contains.



LINC – MIT – BBN – RAND – SDC has length 4.

Distance

The distance between two nodes is the length of the shortest path between them.

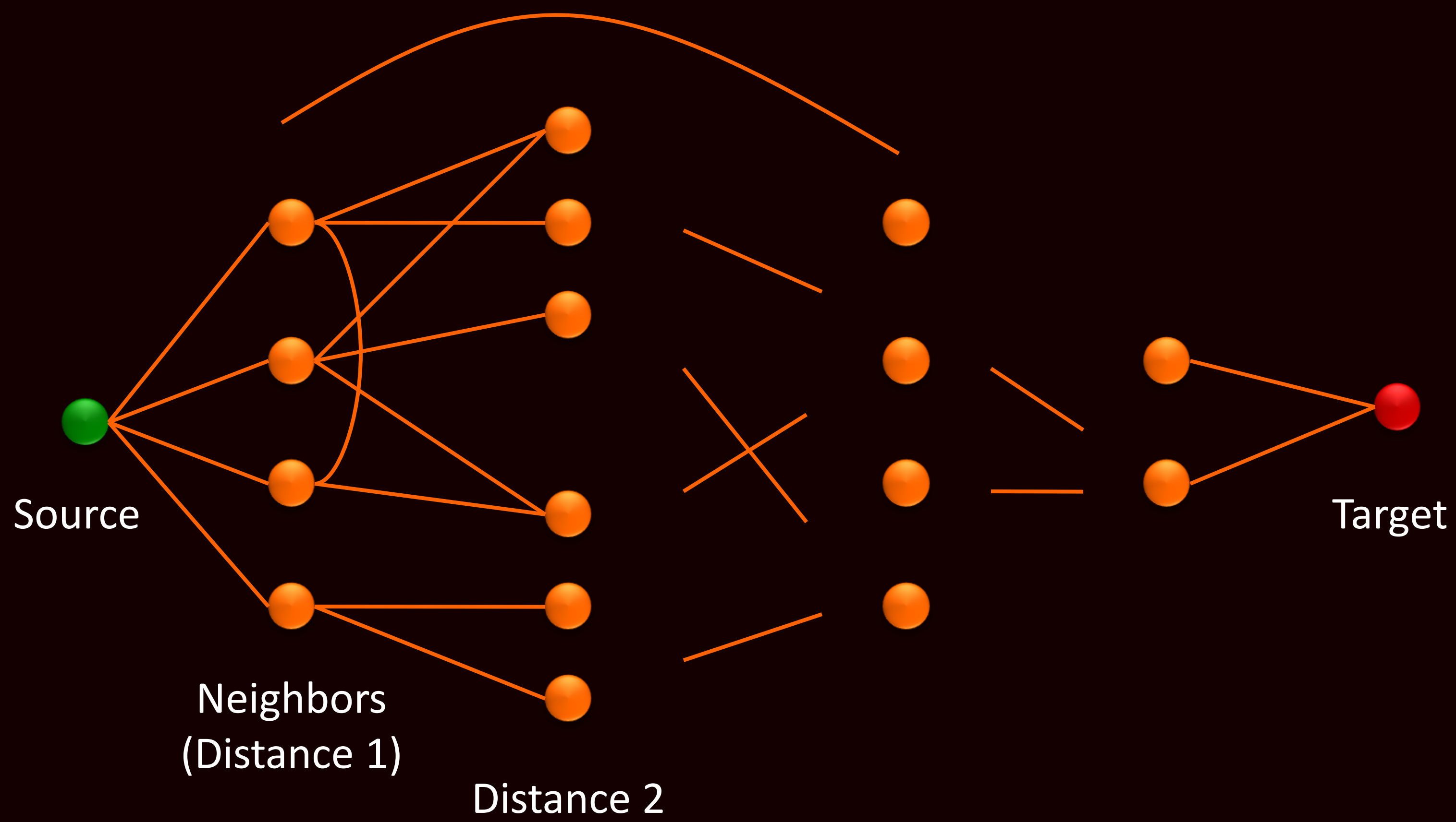


Distance between LINC and SDC is 3.

Computing Distance

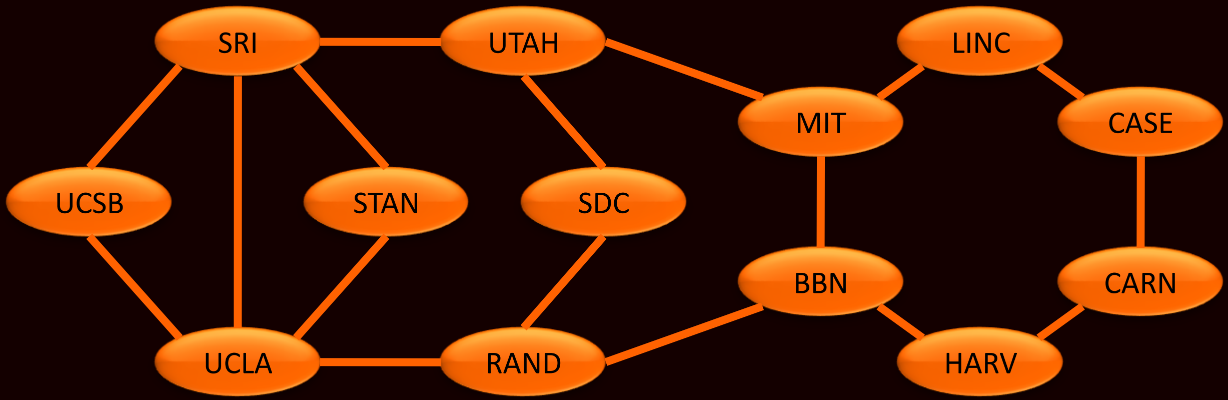
Grow a breadth-first search tree.

Computing Distance



Diameter

The diameter of a network is the maximum distance between any two nodes.



Diameter is 5.

Diameter

Of a ….

clique?

star?

tree?

tree of height k?

binary tree?

Computing Diameter

Squaring adjacency matrix.

Height of arbitrary BFS tree (2-approx).

Computing Diameter

Squaring Adjacency matrix.

$A_{ij} = 1$ if (i,j) is an edge.

$(A^2)_{ij} > 0$ if there is a k s.t. (i,k) is an edge and (k,j) is an edge.

A^p represents paths of length exactly p.

$(A+\text{Identity})^p$ represents paths of length $\leq$ p.

Computing Diameter

Height of arbitrary BFS tree (2-approx).

Consider max shortest path (i_1, ..., i_k).

Height of tree when reaching i_1 is at most k.

Number of remaining levels is at most k.

→ Height of BFS tree is at most twice the diameter k (the length of the maximum shortest path).

Final Project #1

Efficient algorithms for approximating diameters (and other statistics) of big graphs.

Examples

- Collaboration graph
  - 401,000 nodes, 676,000 edges (average degree 3.37)
  - Diameter: 23, Average distance: 7.64
- Cross-post graph, giant component
  - 30,000 nodes, 800,000 edges (average degree 53.3)
  - Diameter: 13, Average distance: 3.8
- Web graph
  - 200 million nodes, 1.5 billion edges (average degree 15)
  - Average connected distance: 16
- Inter-domain Internet
  - 3500 nodes, 6500 edges (average degree 3.71)
  - 95% of pairs of nodes within distance 5

Q2. How connected are we?

Many social networks have small diameter.

||

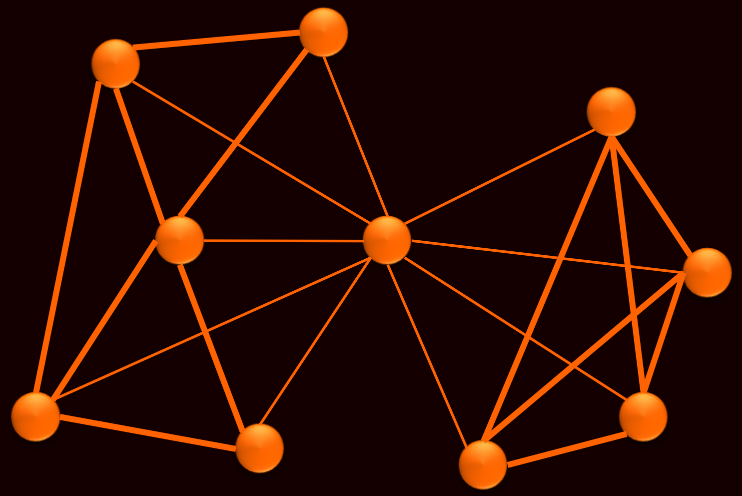
There are short connections between most people (6 degrees of separation).

Clustering Coefficient

How many of your friends are also friends?

Clustering Coefficient

The clustering coeff. of a node is the fraction of its neighbors that are connected.



Clustering coeff. = 13/72

Clustering Coefficient

The clustering coefficient of a graph is the average clustering coefficient of its nodes,

Or the fraction of triangles among all connected triples of nodes.

Examples

- Collaboration graph
  - Clustering coefficient is 0.14
  - Density of edges is 0.000008

- Cross-post graph
  - Clustering coefficient is 0.4492
  - Density of edges is 0.0016

Q 3. How tight-knit are we?

Social networks have high clustering coeff.

||

Many of our friends are friends.

Centrality

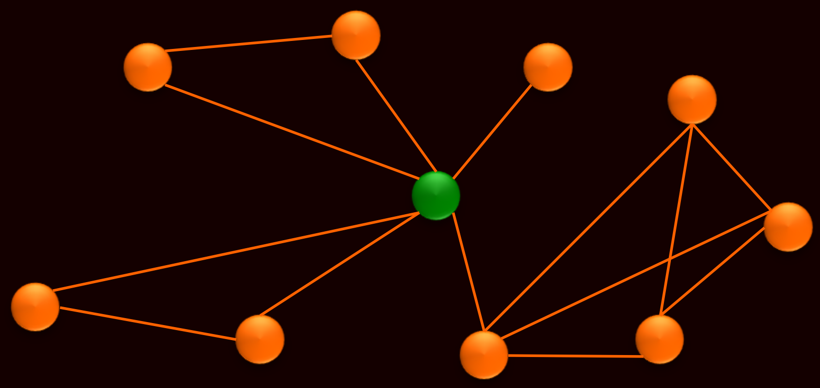
- Measures of centrality
  - Degree-based: how connected is a node
  - Closeness: how easy can a node reach others
  - Betweenness: how important is a node in connecting other nodes
  - Neighbor’s characteristics: how important, central, or influential a nodes neighbors are

Degree Centrality

The degree centrality of a node is

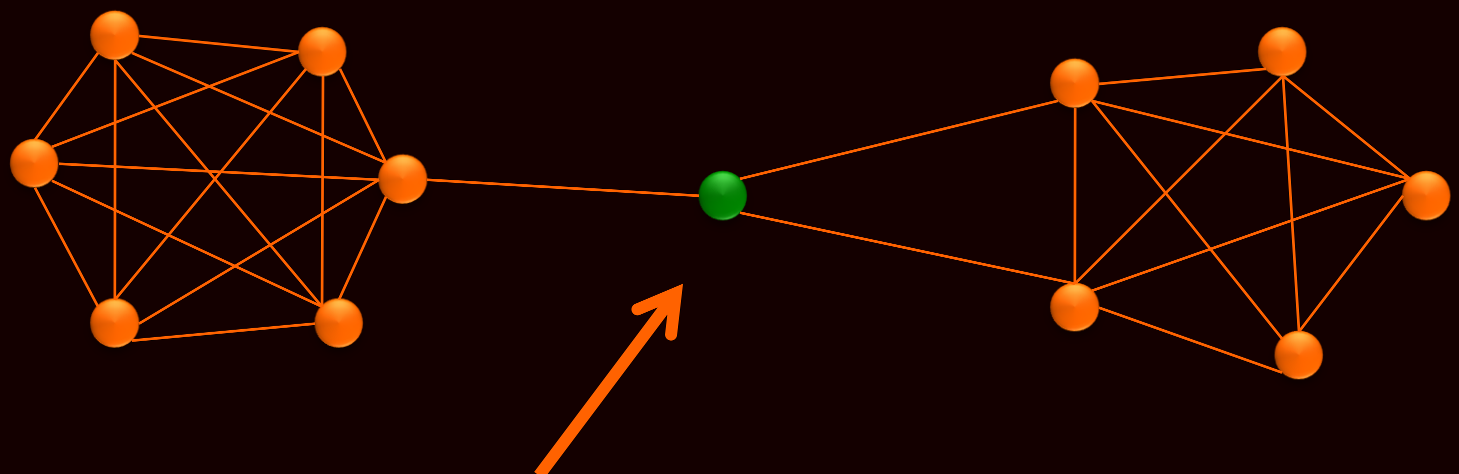
$d(i) / (n-1)$

where d(i) is the degree of node i.



Degree centrality = 6/9

Degree Centrality



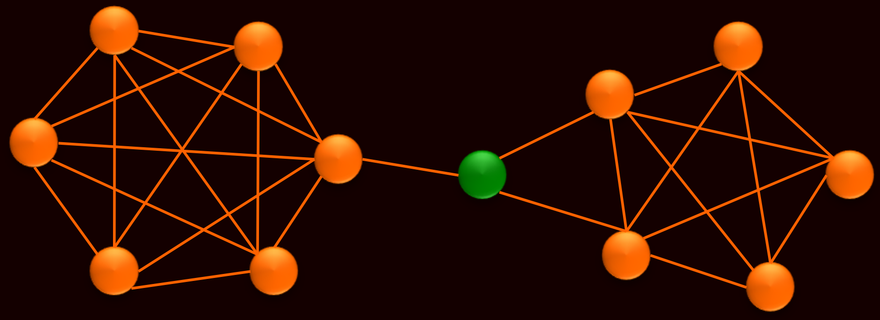
Low degree centrality, but
close to all nodes.

Closeness Centrality

The closeness centrality of a node i is

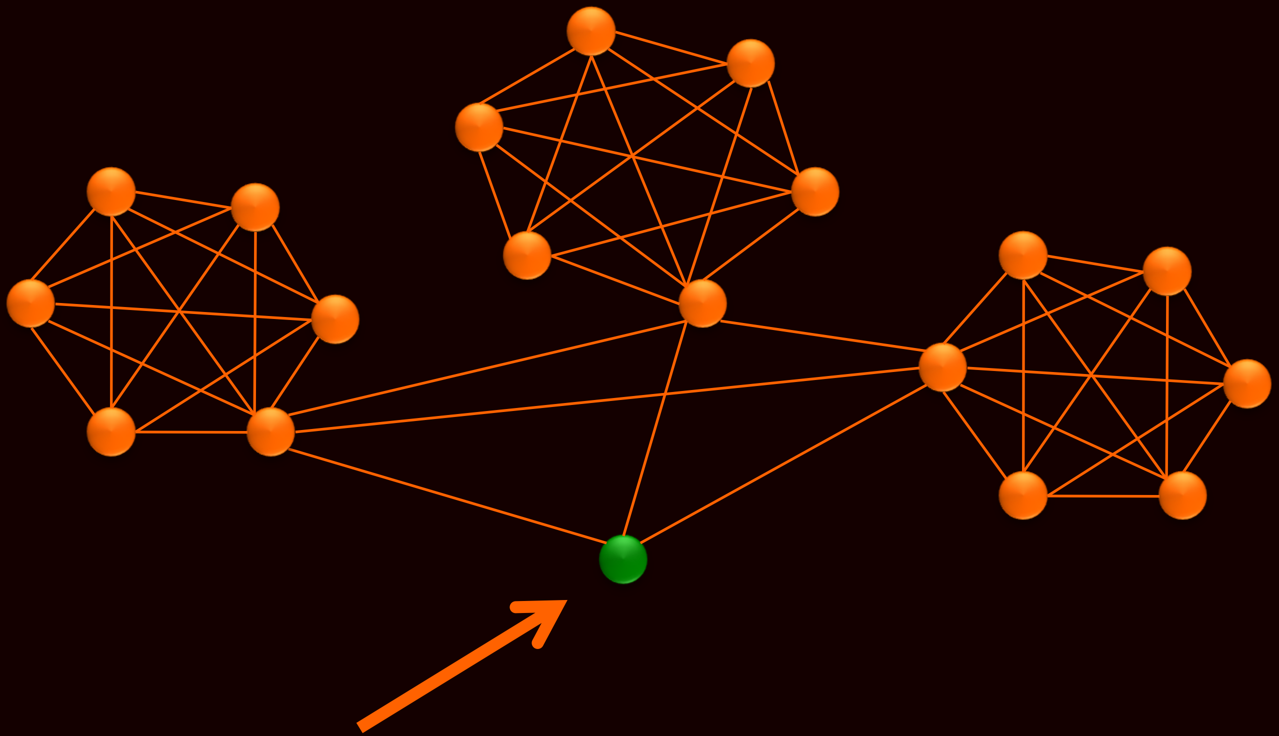
$$\sum \delta^{d(i,j)}$$

where δ is a discounting factor in [0,1] and d(i,j) is length of shortest path from i to j.



Closeness centrality
$= 3\delta + 8\delta^2$

Closeness Centrality



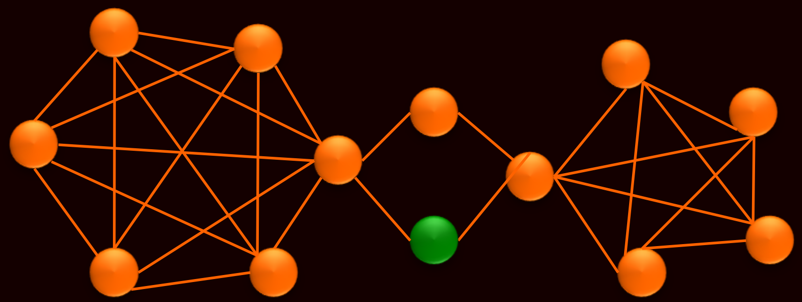
High closeness centrality,
but peripheral.

Betweenness Centrality

Betweennes centrality of node i is fraction of shortest paths passing through i:

$$\sum_{k,j} P_i\,(k,j)/P(k,j)\ /\ (n-1)(n-2)/2$$

where P_i (k,j) is # of shortest paths from j to k through i; P(k,j) is total # of shortest paths.



Betweennes centrality = [30 *(1/2)] / [12*11/2] = 15/66 = 5/22

Final Project #2

Investigate how these properties change as social network evolves.

Assignment:

- Readings:
  - Social and Economic Networks, Part I
  - *Graph Structure in the Web*, Broder et al
  - *The Strength of Weak Ties*, Granovetter
- Reactions:
  - Reaction paper to one of research papers, or a research paper of your choice
- Presentation volunteer?