CS 124/LINGUIST 180
From Languages to Information

Dan Jurafsky
Stanford University

Social Networks: Small Worlds, Weak Ties, and Power Laws

Slides from Jure Leskovec, Lada Adamic, James Moody, Bing Liu,

Networks

- Information in networks, not just text!
- Pagerank: the structure of a network tells you something
- What are the properties of networks and what can we learn from them?

Social network analysis

- Social network analysis is the study of entities (people in an organization), and their interactions and relationships.
- The interactions and relationships can be represented with a network or graph,
  - each vertex (or node) represents an actor and
  - each link represents a relationship. May be directed or not.

Various measures of centrality

- A central actor: involved in many ties.
- **Degree centrality**: number of direct connections a node has
- **Prestige centrality**: everyone points to this actor:
  - Number of in-links
  - *Pagerank* is based on prestige

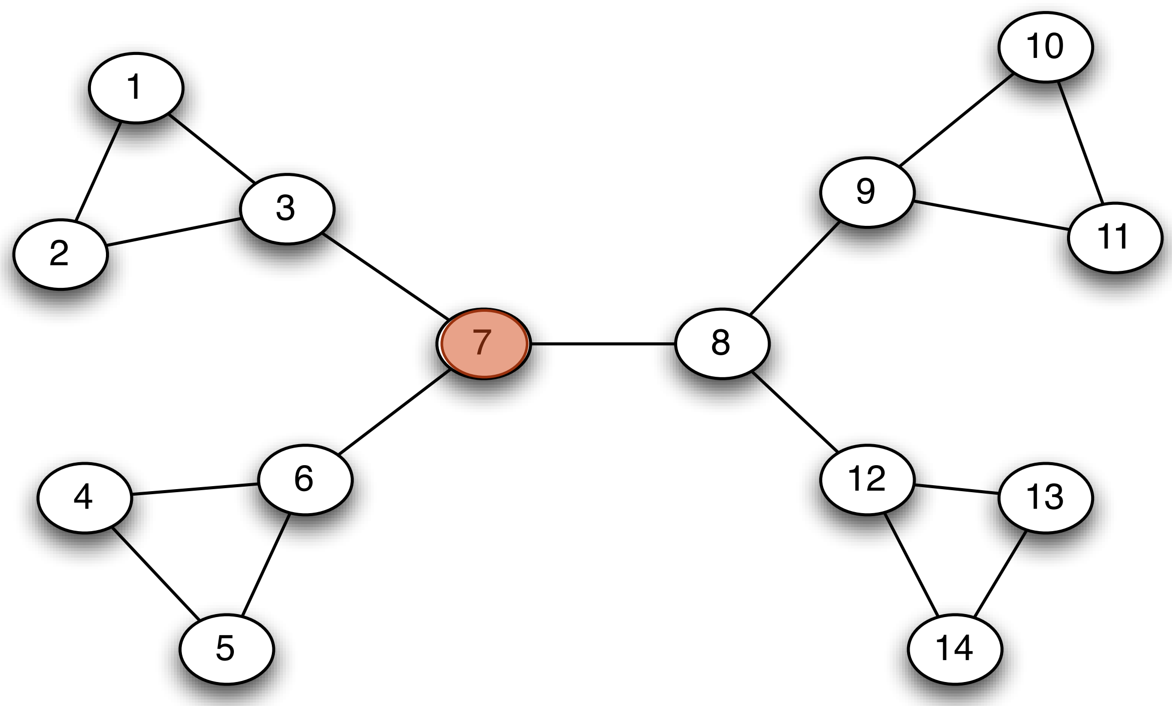
Modified from Bing Liu

Betweenness Centrality

A node with high **betweenness**

- lots of paths have to pass through it
- influences network, choke-point for information
- failure is a problem

Betweenness of node 7 should be high



Betweenness Centrality

- The **betweenness** of a node A (or an edge A-B)=

$$\frac{\text{number of shortest paths that go through A (or A-B)}}{\text{total number of shortest paths that exist between all pairs of nodes}}$$

Betweenness

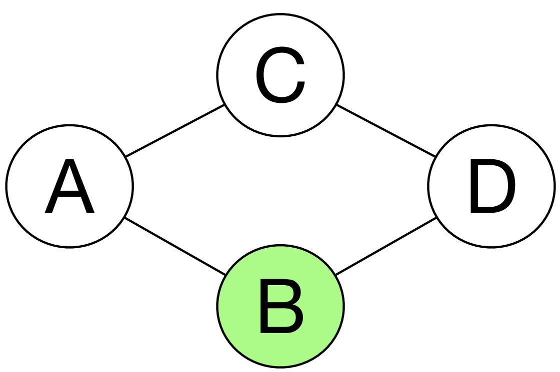
$$\frac{\text{number of shortest paths that go through A}}{\text{total number of shortest paths between all pairs of nodes}}$$

More formally:

$$\frac{\sum_{s \neq v \neq t} \sigma_{st}(v)}{\sum_{s \neq v \neq t} \sigma_{st}}$$

where $\sigma_{st} = \textit{the number of shortest paths between s and t, and}$
$\sigma_{st}(v) = \textit{the number of shortest paths between s and t that go through v}$

Betweenness of B?

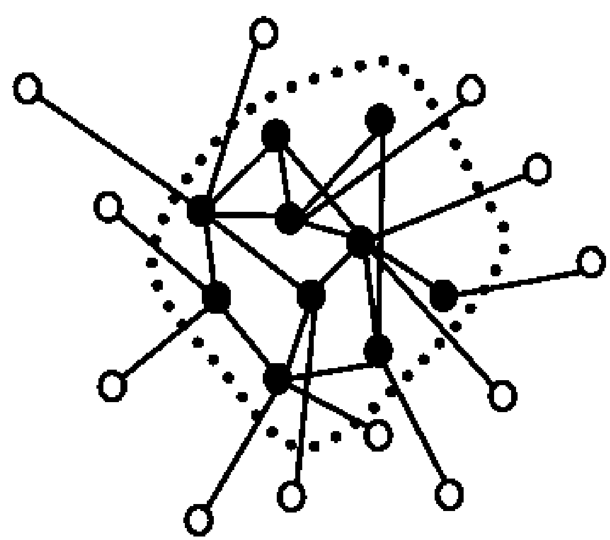


	# sh. paths through B	# shortest paths
AC	0	1
AD	1	2
CD	0	1

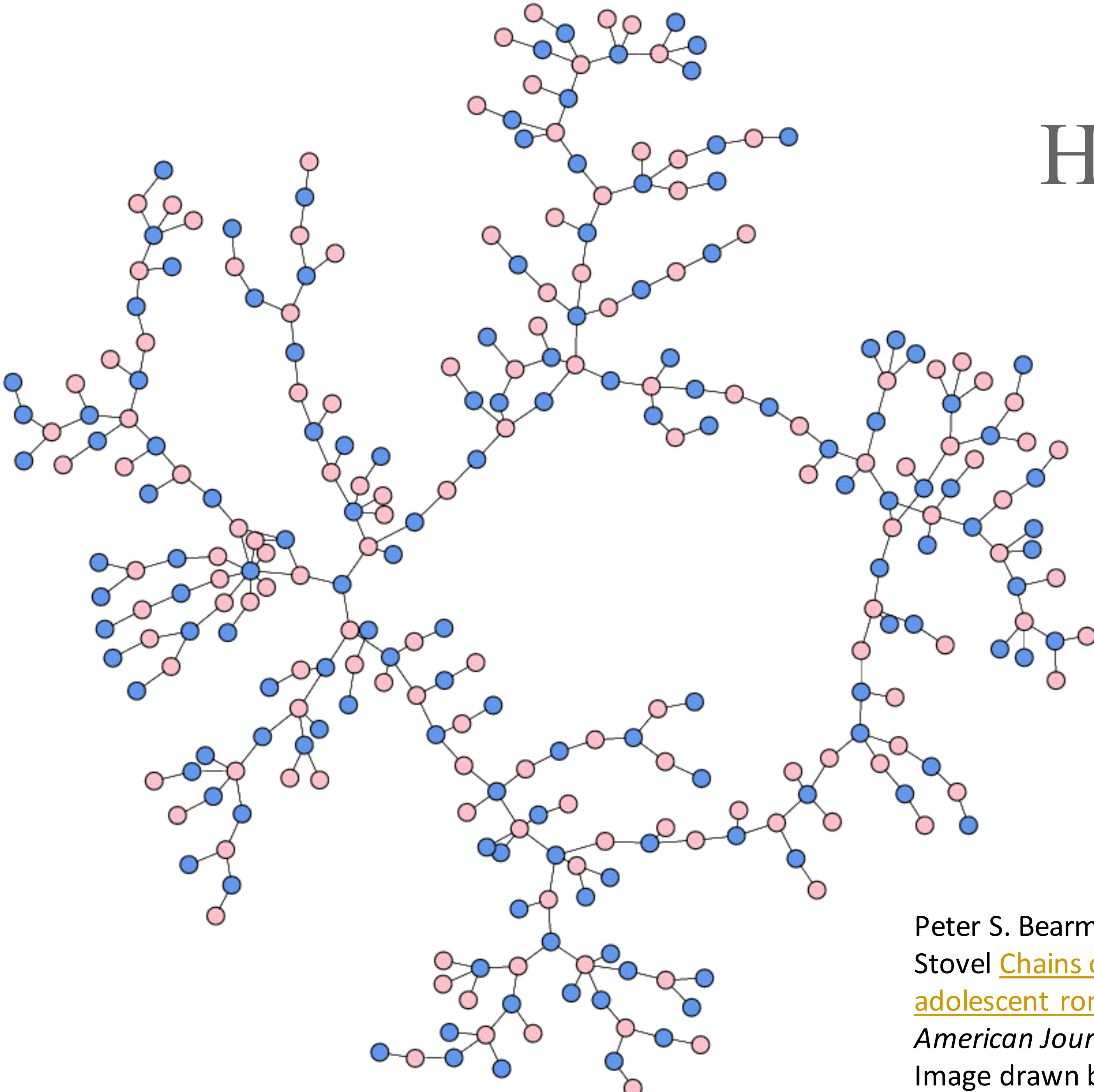
Centrality = 1/4

An example network

- Network of which students have had sex with each other in a high school.
  - important for studying disease spread, etc.
- What do you think its shape is?
- For example: is it core-periphery (like the web)?



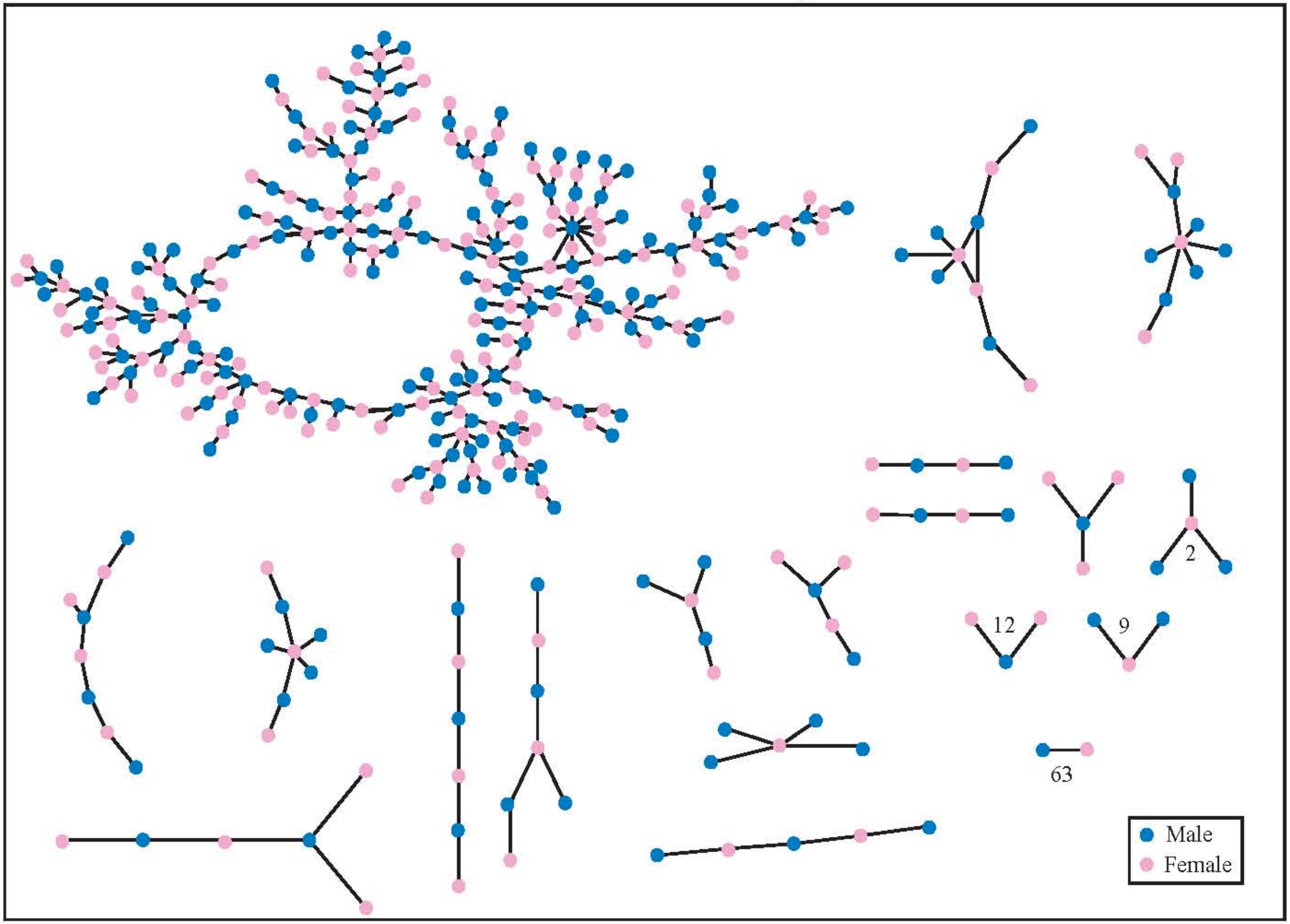
High school dating



Peter S. Bearman, James Moody and Katherine Stovel [Chains of affection: The structure of adolescent romantic and sexual networks]
American Journal of Sociology 110 44-91 (2004)
Image drawn by Mark Newman

Slide from Drago Radev

The Structure of Romantic and Sexual Relations at "Jefferson High School"



Each circle represents a student and lines connecting students represent romantic relations occuring within the 6 months preceding the interview. Numbers under the figure count the number of times that pattern was observed (i.e. we found 63 pairs unconnected to anyone else).

Why does the graph have this shape?

- Teens probably don’t say:
  - “By selecting this partner, I maximize the probability of inducing a spanning tree.”

- The “microtaboo” Bearman and Moody propose
  - don’t date your ex-girlfriend’s boyfriend’s ex-girlfriend
  - (or the reverse)
  - a simulation shows this constraint results in spanning tree

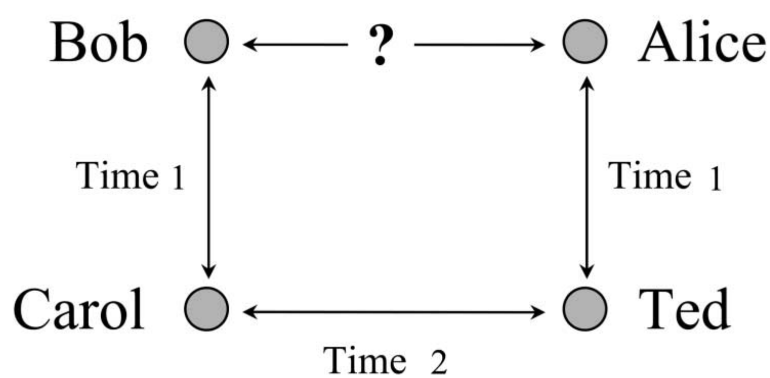


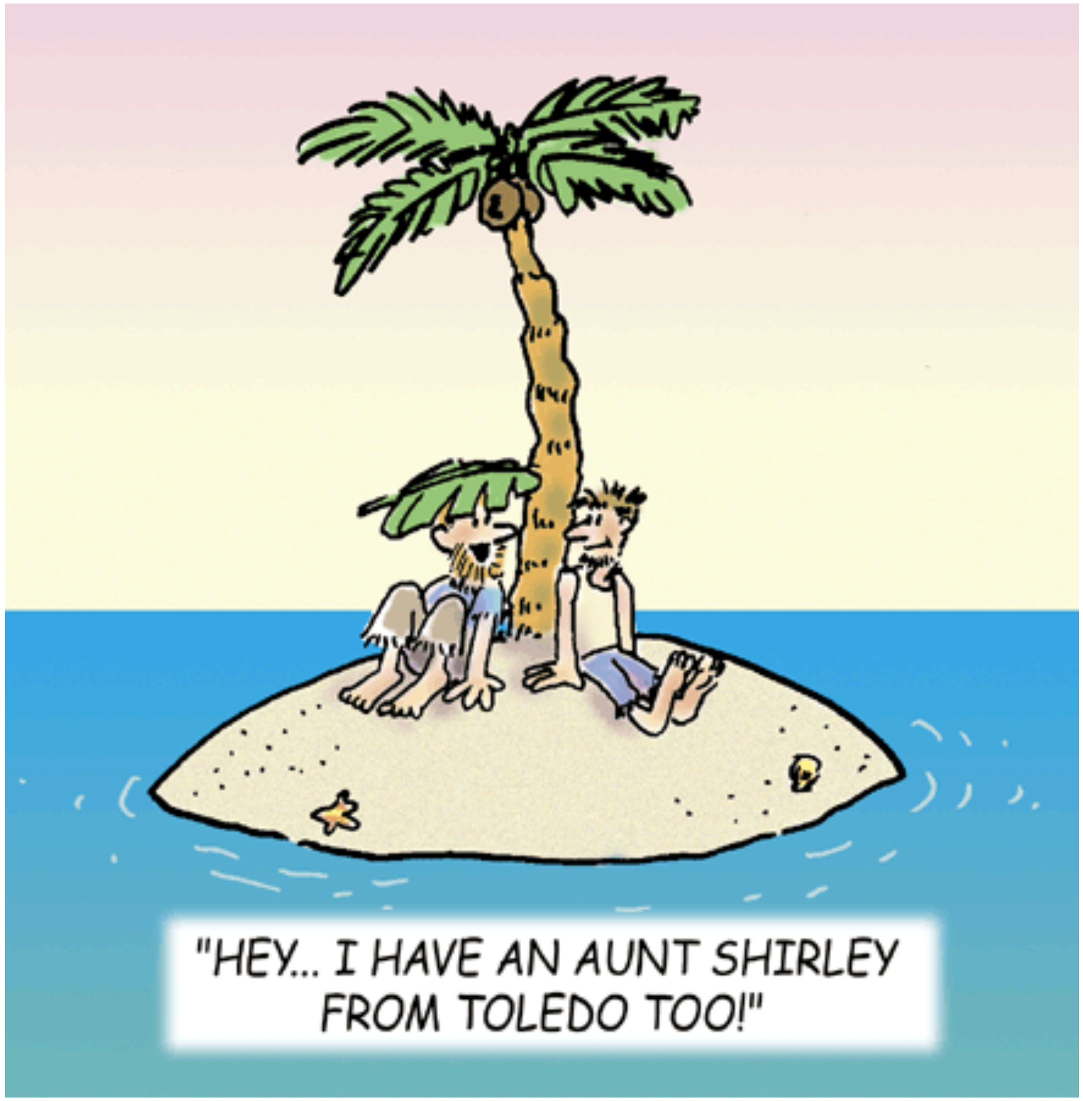
Fig. 8.—Hypothetical cycle of length 4

CS 124/LINGUIST 180
From Languages to Information

Dan Jurafsky
Stanford University

Small Worlds

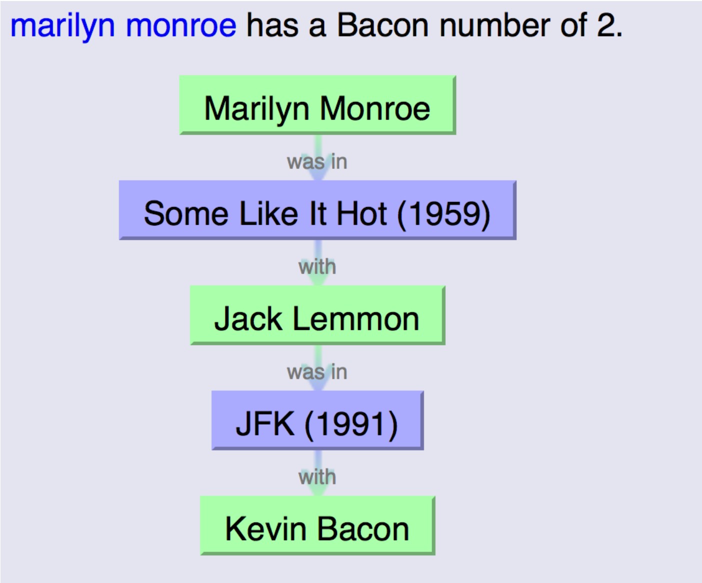
Small worlds



Slide from Lada Adamic

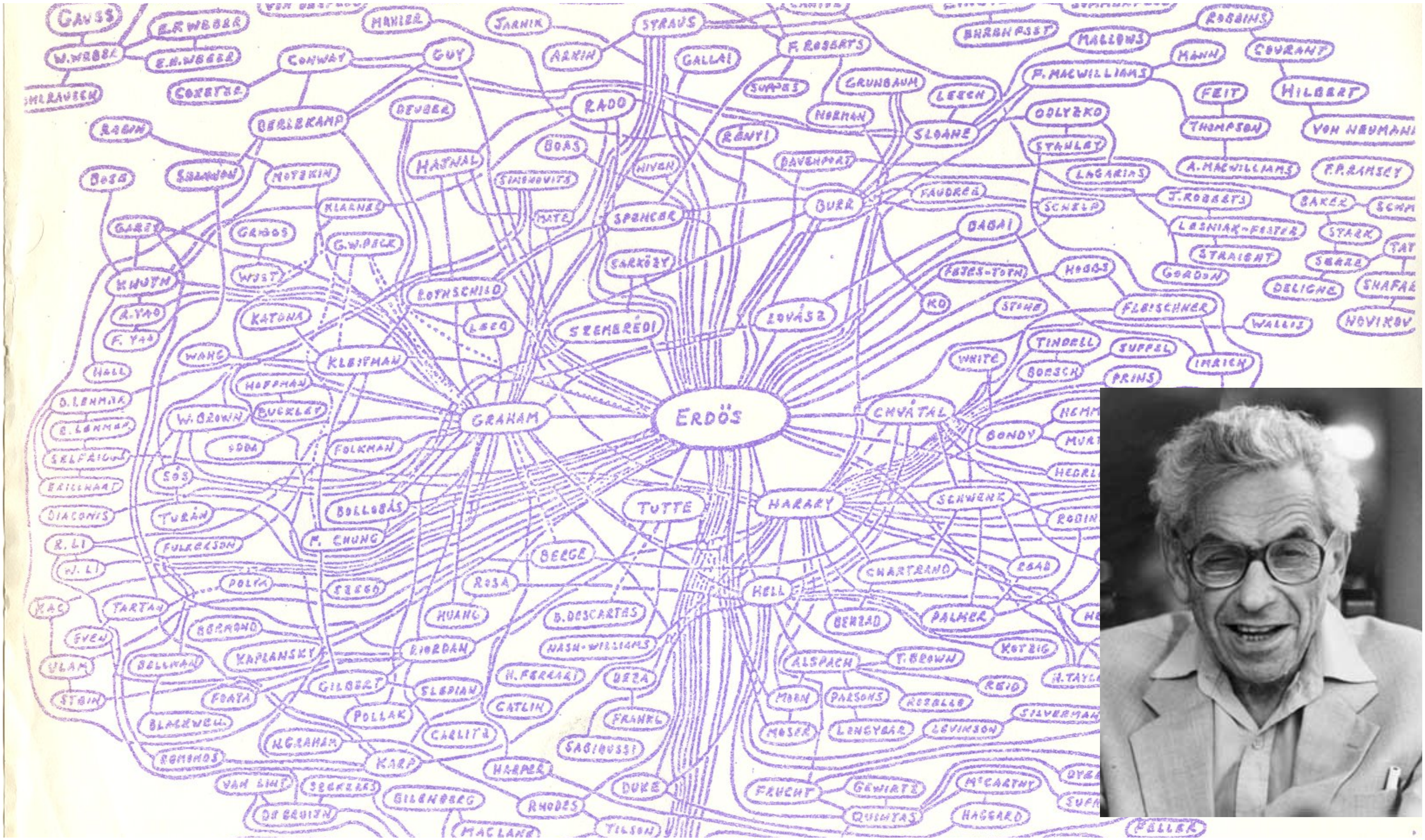
Six Degrees of Kevin Bacon

- **Popularization of a small-world idea:**
- **The Bacon number:**
  - Create a network of Hollywood actors
  - Connect two actors if they co-appeared in the movie
- **Bacon number:** number of steps to Kevin Bacon
  - As of 2013, the highest (finite) Bacon number reported is 11
  - Only approx. 12% of all actors cannot be linked to Bacon



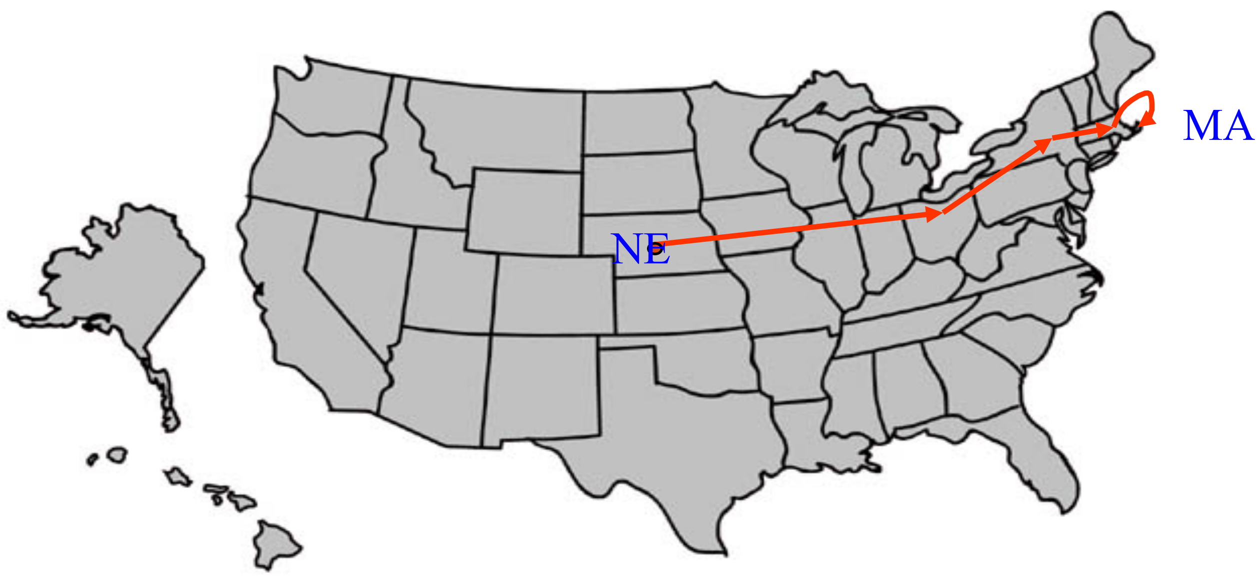
Slide adapted from Jure Leskovec

Erdös numbers are small too



The Small World Experiment

What is the typical shortest path between any two people?



Stanley Milgram (1967)

- Chose 300 people in Omaha, NE and Wichita, KA
- Ask them to get a letter to a stock-broker in Boston by passing it through friends
- **How many steps did it take?**

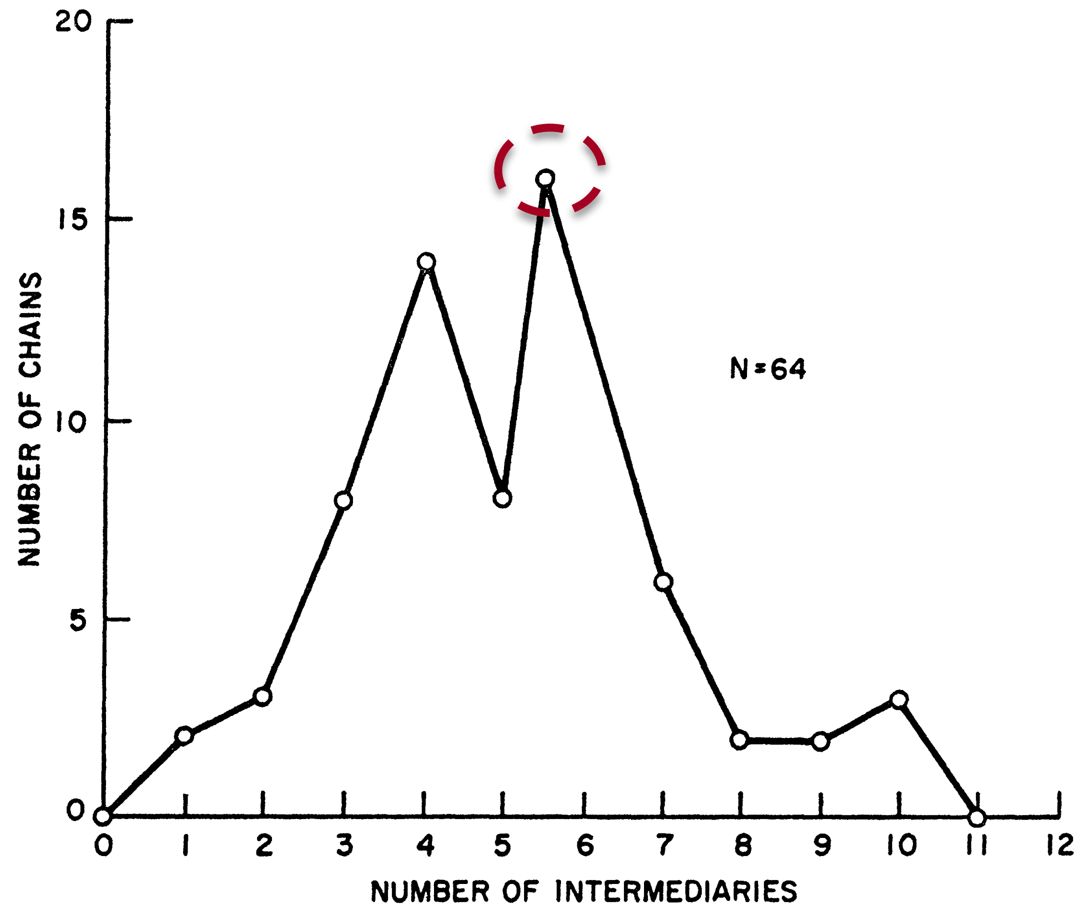
Slide from Lada Adamic

Milgram's small world experiment

It took 6.2 steps on average

"Six degrees of separation"

Can we check this computationally?



Facebook

Backstrom Boldi Rosa Ugander and Vigna, 2012
“Four Degrees of Separation"

99.6% of all pairs of users connected by paths of 5 degrees (6 hops)

92% are connected by only four degrees (5 hops).

721 million users
69 billion friendship links

Fun facts: Origins of the “6 degress” hypothesis

- Hungarian writer Karinthy’s 1929 play “Chains” (Láncszemek)
  - https://djjr-courses.wdfiles.com/local--files/soc180%3Akarinthy-chain-links/Karinthy-Chain-Links_1929.pdf

Duncan Watts: Networks, Dynamics and the Small-World Phenomenon

- Why do we see the small world pattern?
- What implications does it has for the dynamical properties of social systems?

Slide from James Moody

Duncan Watts: Networks, Dynamics and the Small-World Phenomenon

Watts says there are 4 conditions that make the small world phenomenon interesting:

1) The network is **large** - *O*(Billions)
2) The network is **sparse** - people are connected to a small fraction of the total network
3) The network is **decentralized** -- no single (or small #) of stars
4) The network is highly **clustered** -- most friendship circles are overlapping

Slide from James Moody

Duncan Watts: Networks, Dynamics and the Small-World Phenomenon

Formally, we can characterize a graph through 2 statistics.

1) The **characteristic path length, *L***

The average length of the shortest paths connecting any two nodes.

*(Note: this is not quite the same as the **diameter** of the graph, which is the **maximum** shortest path connecting any two nodes)*

2) The **clustering coefficient, C**

The average local density.

A *small world graph* is any graph with a relatively small *L* and a relatively large *C*.

Slide from James Moody

Local clustering coefficient (Watts&Strogatz 1998)

- For a vertex i

  C = The fraction of pairs of neighbors of the node that are connected

  “What percentage of your friends know each other?”

  - Let n_i be the number of neighbors of vertex *i*

$$C_i = \frac{\text{number of connections between i's neighbors}}{\text{maximum number of possible connections between i's neighbors}}$$

$$C_{i\ \text{directed}} = \frac{\text{\# directed connections between i's neighbors}}{n_i * (n_i - 1)}$$

$$C_{i\ \text{undirected}} = \frac{\text{\# undirected connections between i's neighbors}}{n_i * (n_i - 1)/2}$$

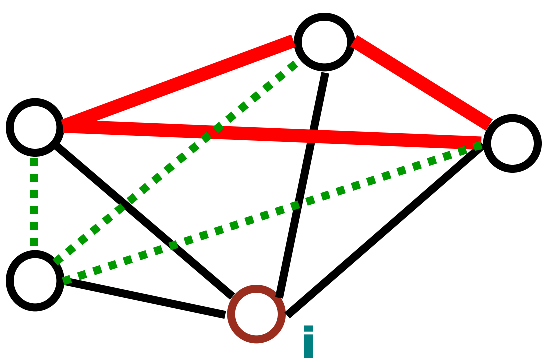
Slide from Lada Adamic

Local clustering coefficient

(Watts & Strogatz 1998)

- Average C_i over all n vertices

$$C = \frac{1}{n}\sum_i C_i$$



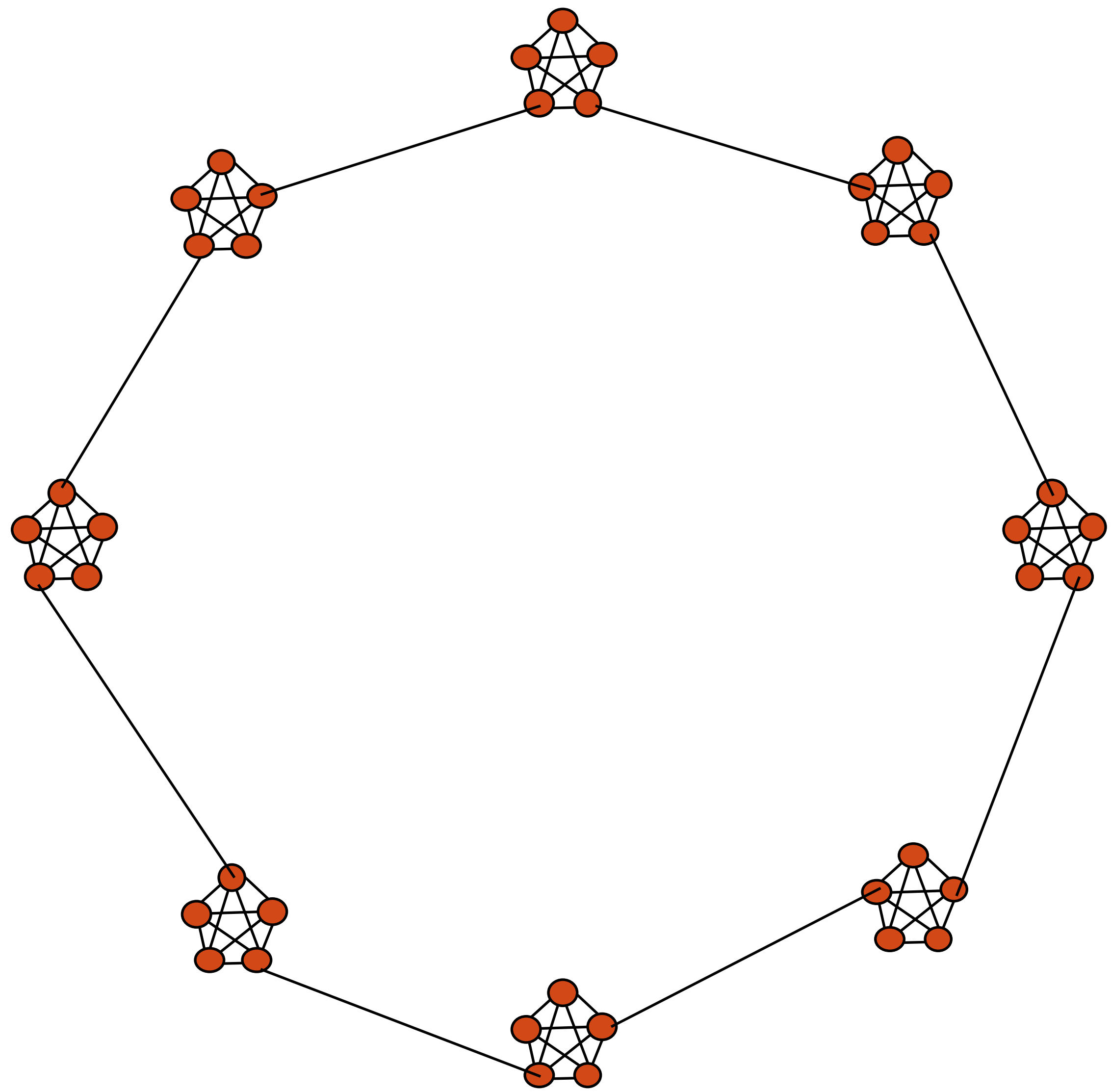
$n_i = 4$
max number of connections:
$4*3/2 = 6$
3 connections present
$C_i = 3/6 = 0.5$

link present
link absent

Slide from Lada Adamic

Watts and Strogatz "Caveman network"

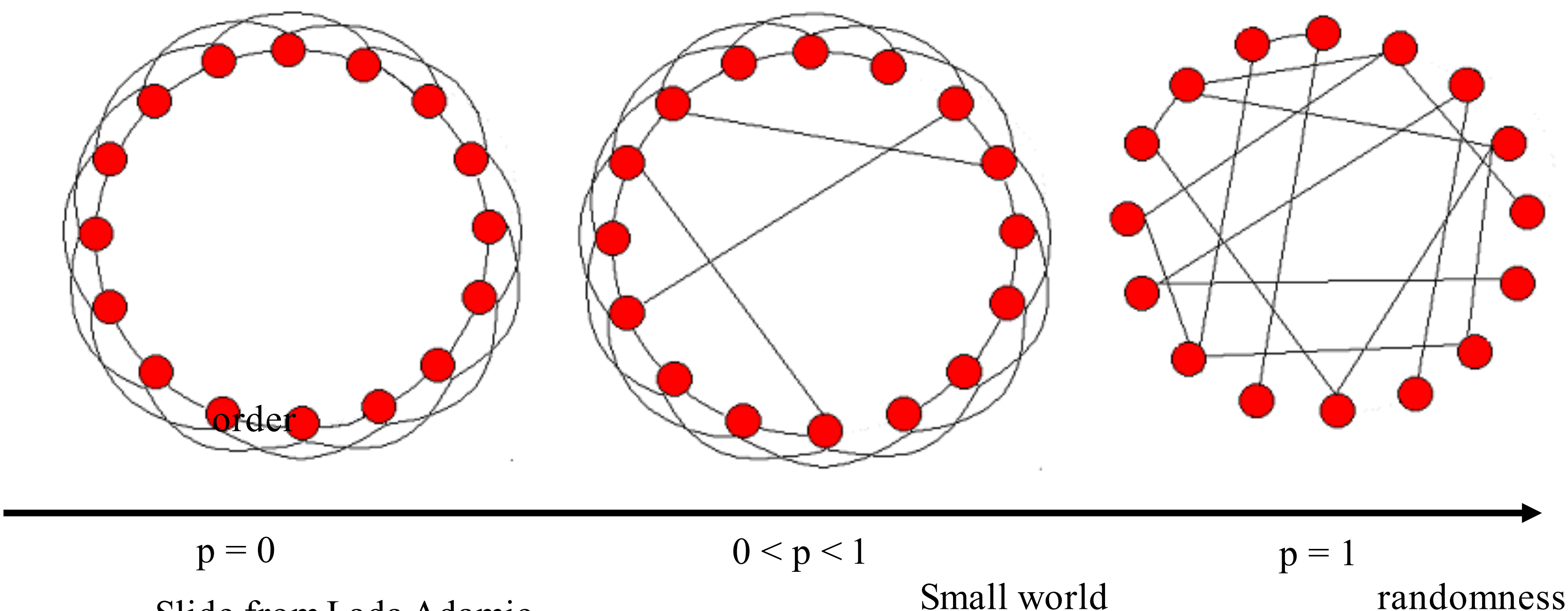
- Everyone in a cave knows each other
- A few people make connections
- Are C and L high or low?
- C high, L high



Slide from Lada Adamic

Watts and Strogatz model [WS98]

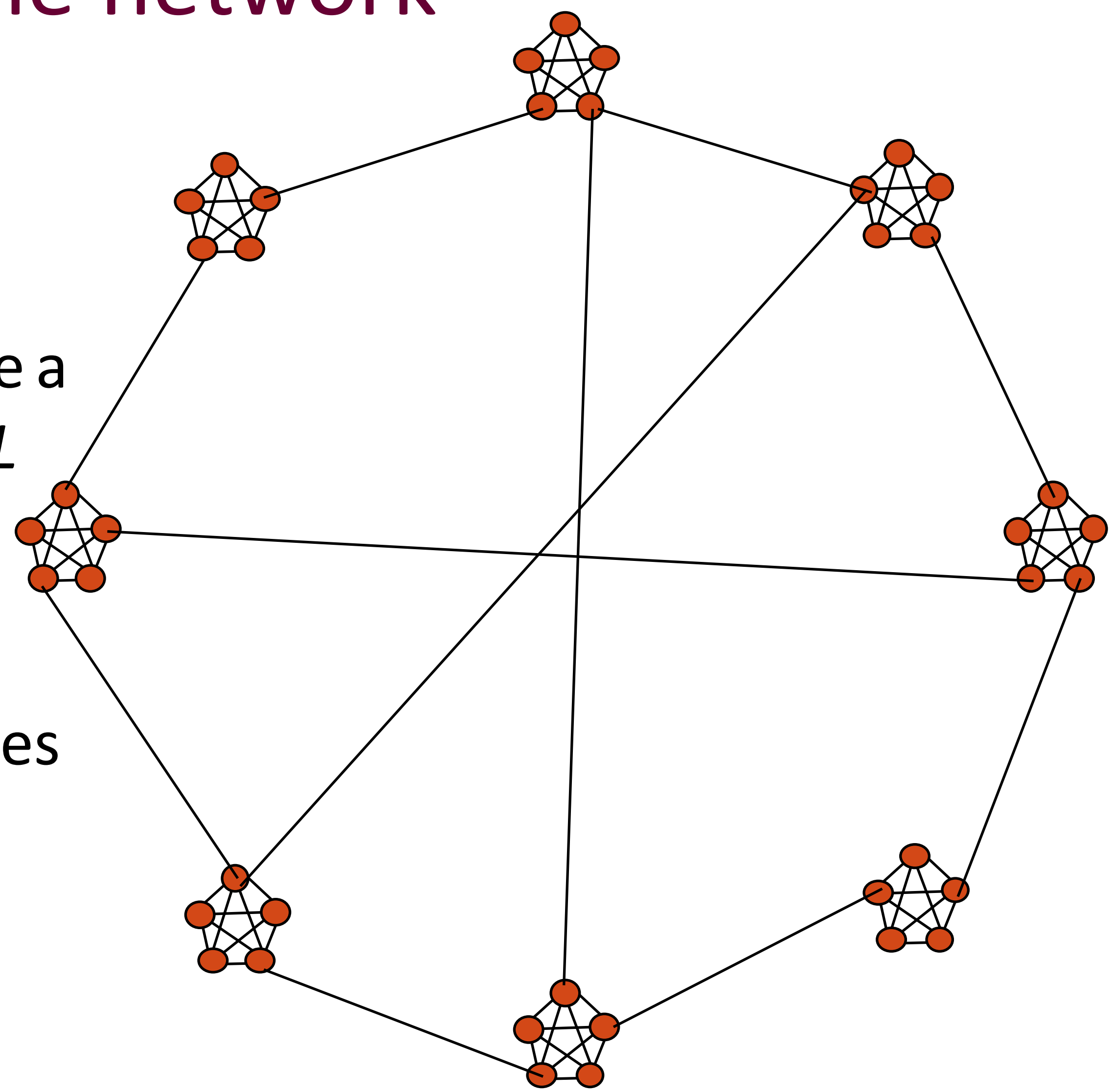
- Start with a ring, where every node is connected to the next z nodes (a regular lattice)
- With probability p, rewire every edge (or, add a shortcut) to a uniformly chosen destination.



Slide from Lada Adamic

Why does this work? Key is fraction of shortcuts in the network

In a highly clustered, ordered network, a single random connection will create a shortcut that lowers *L* dramatically



Small world properties can be created by a small number of shortcuts

Slide from Lada Adamic

Clustering and Path Length

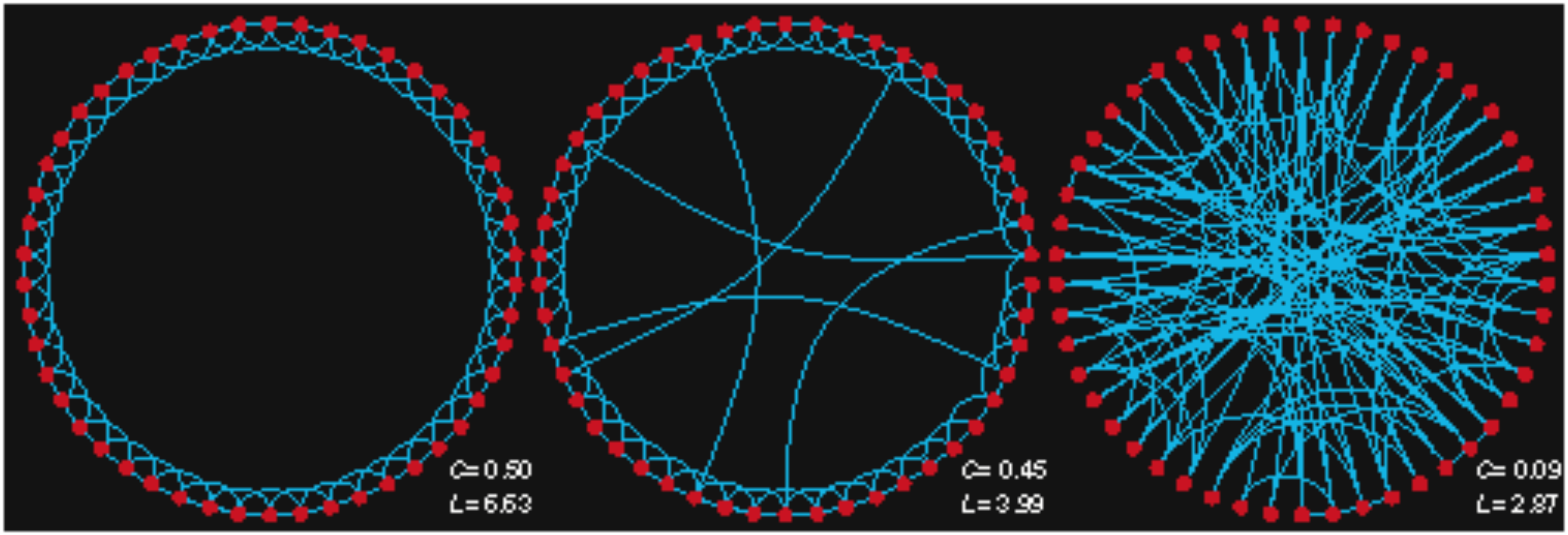


Figure 1. Watts–Strogatz model interpolates between a regular lattice *(left)* and a random graph *(right)*. Randomly rewiring just a few edges *(center)* reduces the average distance between nodes, *L*, but has little effect on the clustering coefficient, *C*. The result is a "small-world" graph.

Regular Graphs have a high clustering coefficient but also a high L

Slide from Lada Adamic

Random Graphs have a low clustering coefficient but a low L

Small World: Summary

- **Could a network with high clustering be at the same time a small world?**
  - Yes! You don't need more than a few random links

The Watts Strogatz Model:

- Provides insight on the interplay between clustering and the small-world
- Captures the structure of many realistic networks
- Accounts for the high clustering of real networks

Slide from Jure Leskovec

CS 124/LINGUIST 180
From Languages to Information

Dan Jurafsky
Stanford University

Weak links

Weak links

- Mark Granovetter (1960s) studied how people find jobs. He found out that most job referrals were through personal contacts
- But more by acquaintances and not close friends.

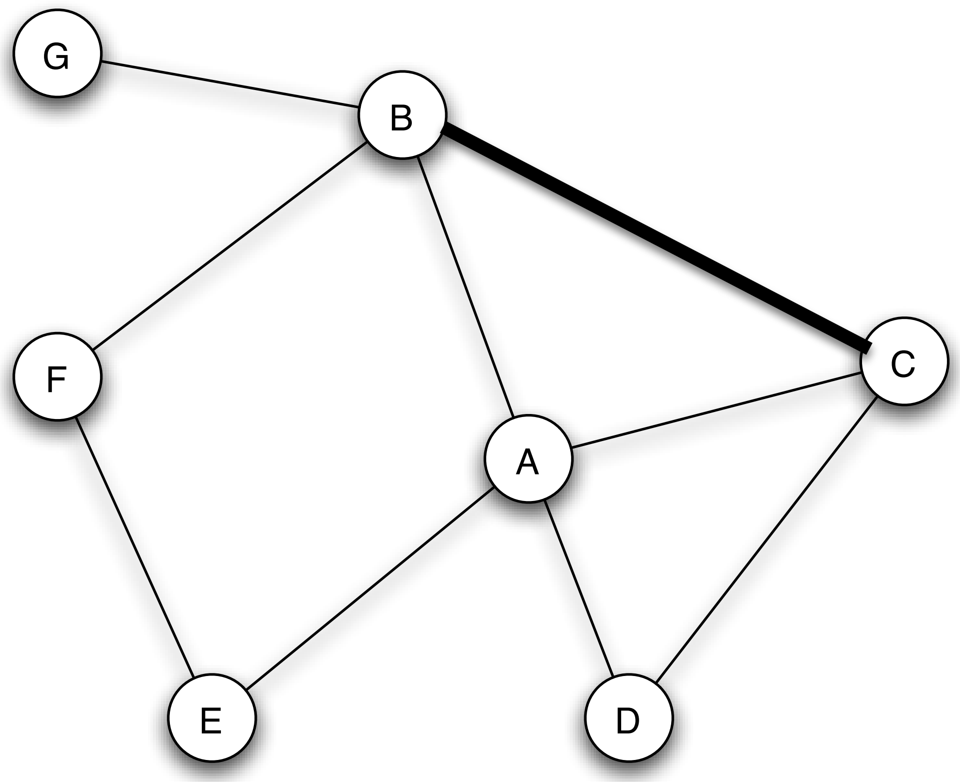
- Aside:
  - Accepted by the American Journal of Sociology after 4 years of unsuccessful attempts elsewhere.
  - One of the most cited papers in sociology.

- Mystery: Why didn't jobs come from close friends?

Adapted from Drago Radev

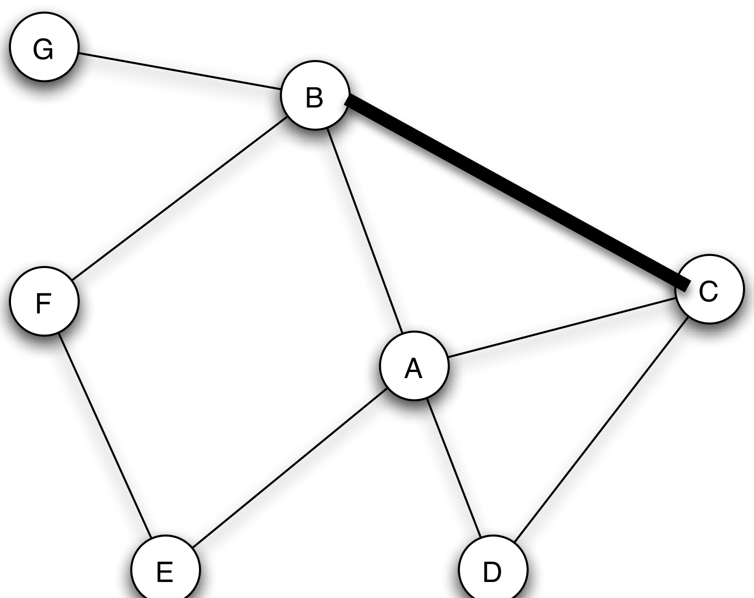
Triadic Closure

"If two people in a social network have a friend in common, then there is an increased likelihood that they will become friends themselves at some point in the future." (Anatole Rapoport 1953)



Reminder: clustering coefficient C

- C of a node A is the probability that two randomly selected friends of A are friends themselves
- A before new edge = 1/6
- (of B-C, B-D, B-E, **C-D**, C-E, C-E)
- After new edge?   2/6
- Triadic closure leads to higher clustering coefficients

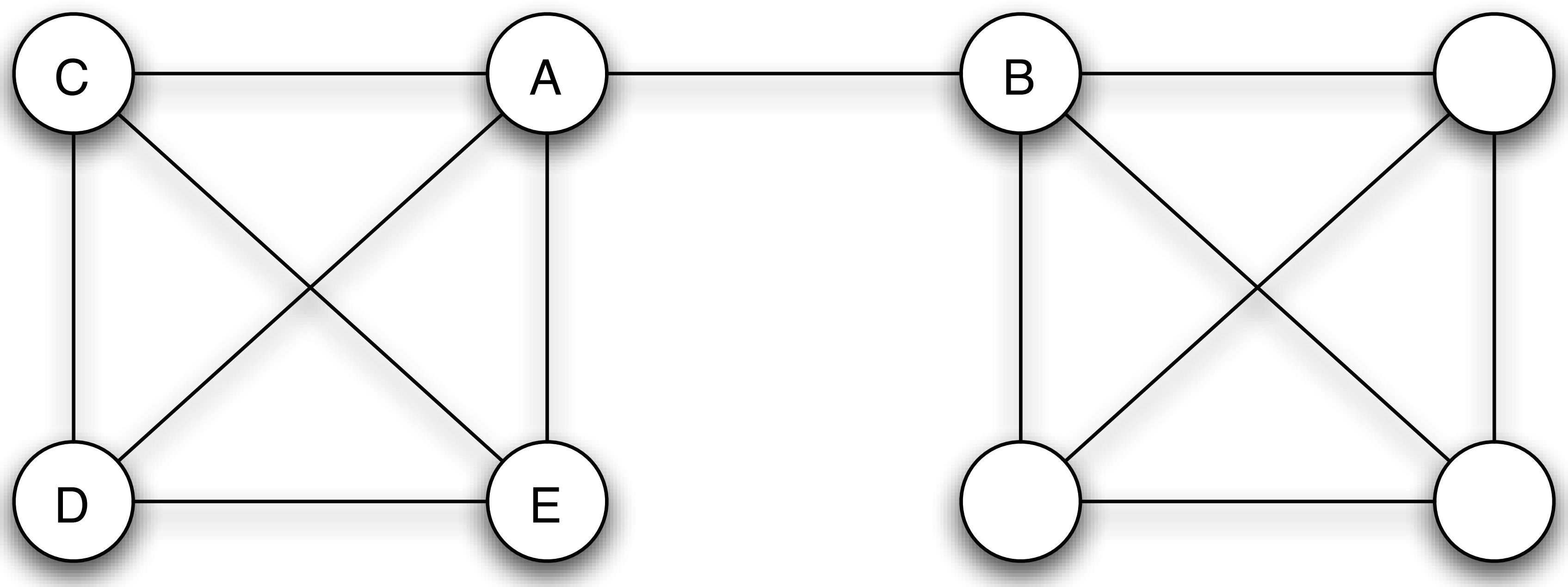


Why Triadic Closure?

1. We meet our friends through other friends
   - B and C have **opportunity** to meet through A
2. B and C's mutual friendship with A gives them a reason to **trust** A
3. A has incentive to bring B and C together to avoid **stress**:
   - if A is friends with two people who don't like each other it causes stress
   - **Bearman and Moody: teenage girls with low clustering coefficients in their network of friends much more likely to consider suicide**

Bridges

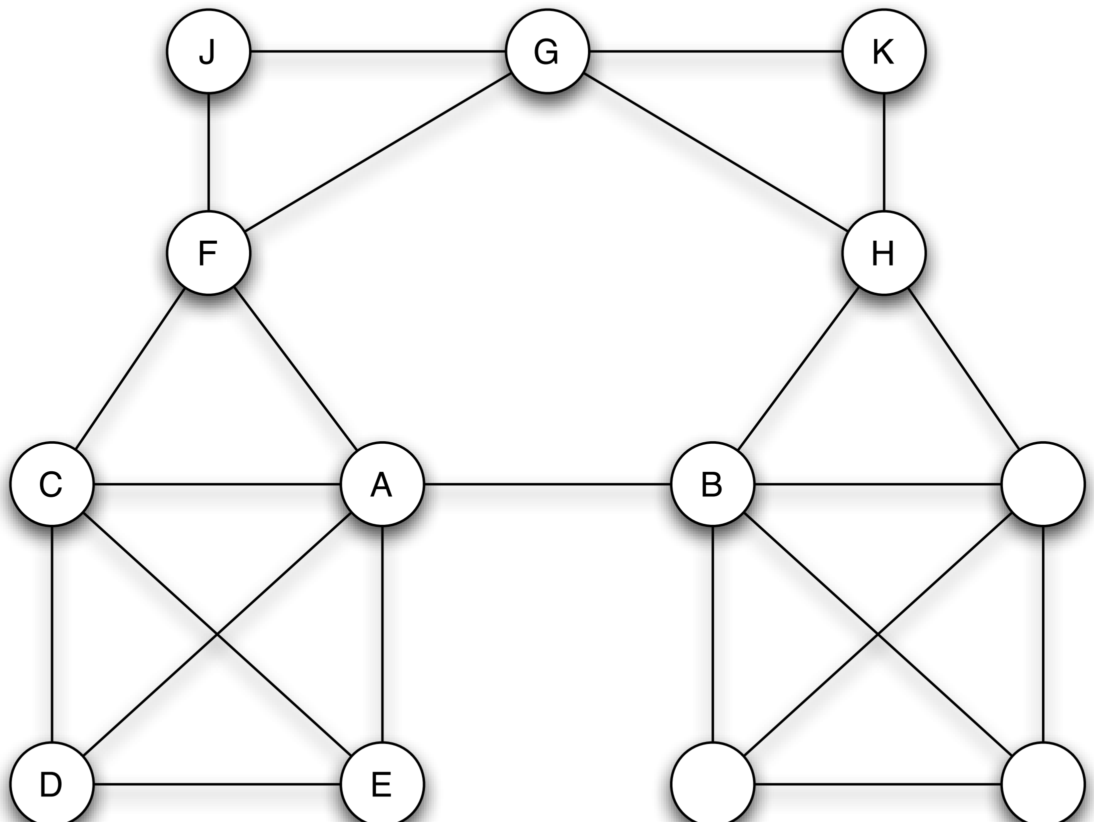
A bridge is an edge whose removal places A and B in different components



If A is going to get new information (like a job) that she doesn't already know about, it might come from B

Local Bridge

A local bridge is an edge whose endpoints A and B have no friends in common (so a local bridge does not form the side of any triangle)



If A is going to get new information (like a job) that she doesn't already know about, it might come from B

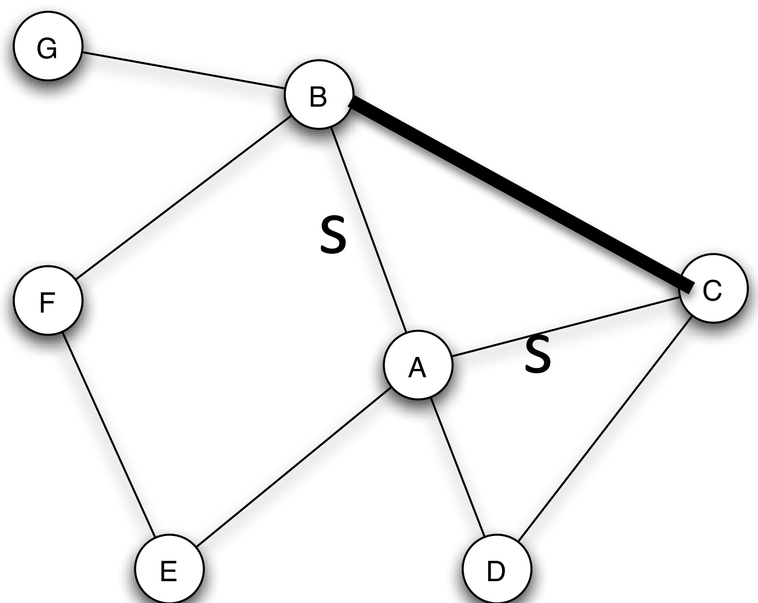
Strong and Weak Ties

- Strength of ties
  - amount of time spent together
  - emotional intensity
  - intimacy (mutual confiding)
  - reciprocal services
- Simplifying assumption:
  - Ties are either strong (s) or weak (w)

Adapted from James Moody

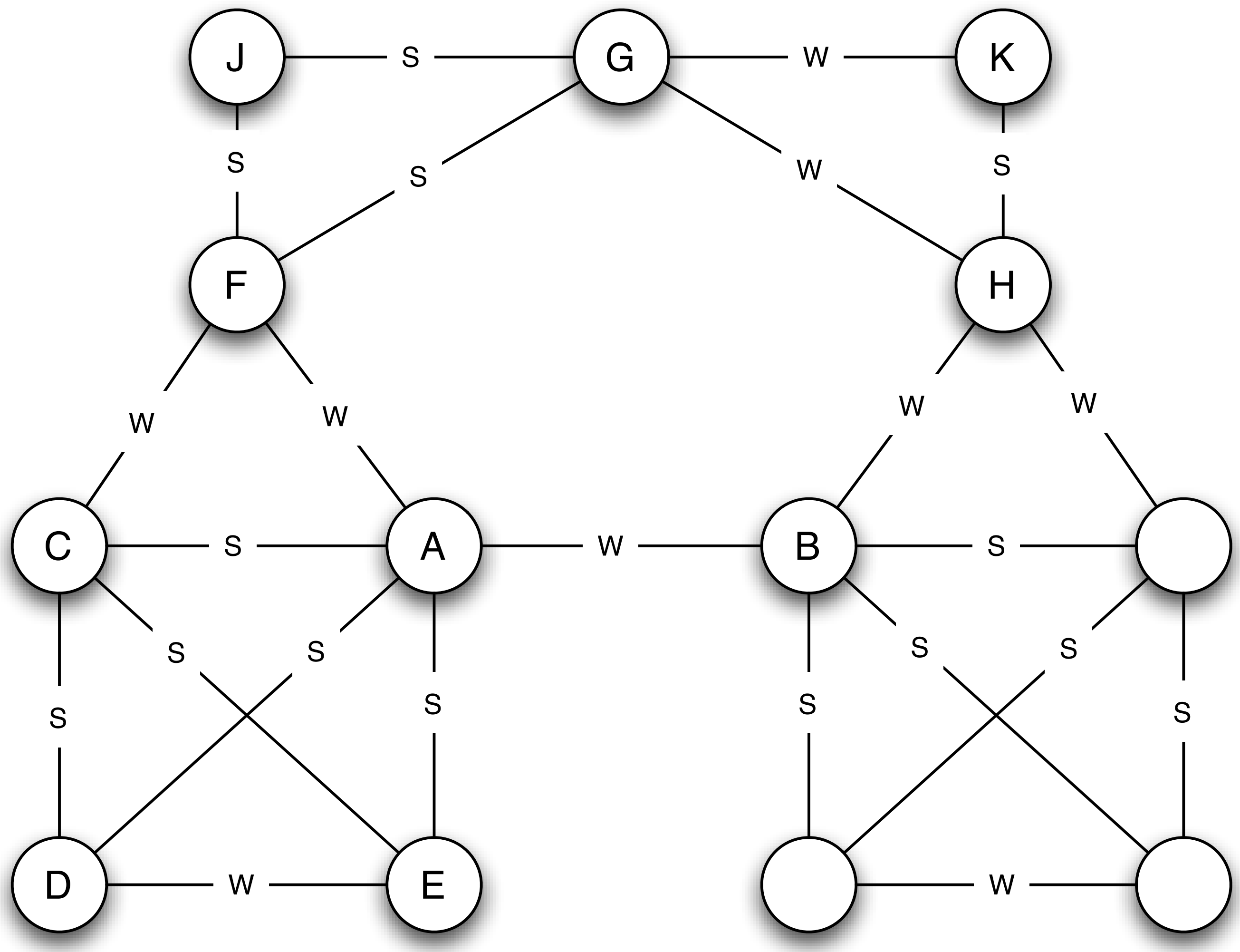
Strong ties and triadic closure

- The new B-C edge more likely to form if A-B and A-C are **strong** ties
- More extreme: if A has strong ties to B and to C, there must be an edge B-C



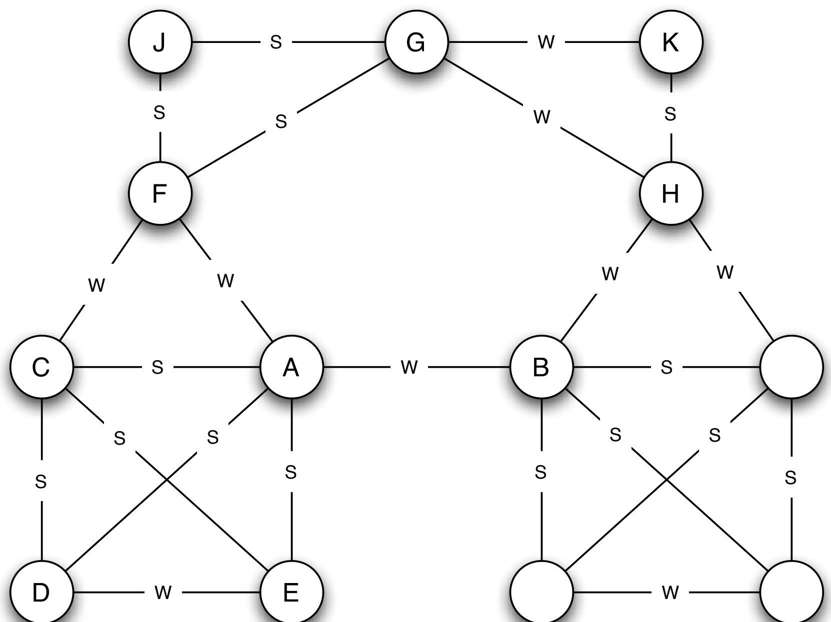
Strong triadic closure

If a node Q has two strong ties to nodes Y and Z, there is an edge between Y and Z



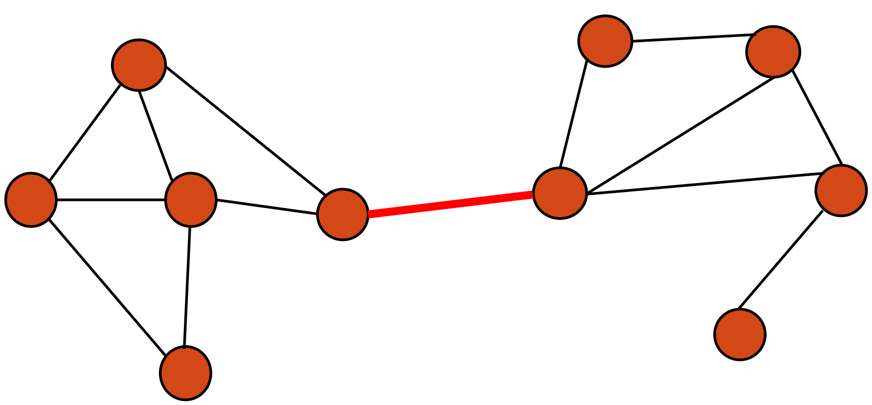
Closure and bridges

- If a node A in a network satisfies the Strong Triadic Closure Property and is involved in at least two strong ties, then any local bridge it is involved in must be a weak tie.
- So local bridges are likely to be weak ties
- Explaining why jobs came from weak ties



Strength of weak ties

- Weak ties can occur between cohesive groups
  - old college friend
  - former colleague from work



weak ties will tend to have low transitivity

Slide from James Moody

Strength of weak ties – how to get a job

- Granovetter: How often did you see the contact that helped you find the job prior to the job search
  - 16.7% often (at least once a week)
  - 55.6% occasionally (more than once a year but less than twice a week)
  - 27.8% rarely – once a year or less
- Weak ties will tend to have different *information* than we and our close contacts do
- Long paths rare
  - 39.1 % info came directly from employer
  - 45.3 % one intermediary
  - 3.1 % > 2 (more frequent with younger, inexperienced job seekers)
- Compatible with Watts/Strogatz small world model: short average shortest paths thanks to 'shortcuts' that are non-transitive

Slide from James Moody

More evidence for strength of weak ties

In the Milgram small world experiments, acquaintanceship ties were more effective than family, close friends at passing information

Summary

- Triangles (triadic closure) lead to higher clustering coefficients
  - Your friends will tend to become friends
- Local bridges will often be weak ties
- Information comes over weak ties

CS 124/LINGUIST 180
From Languages to Information

Dan Jurafsky
Stanford University

Power Laws

Degree of nodes

- Many nodes on the internet have low degree
  - One or two connections
- A few (hubs) have very high degree
- The number P(k) of nodes with degree k follows a power law:

$$P(k) \propto k^{-\alpha}$$

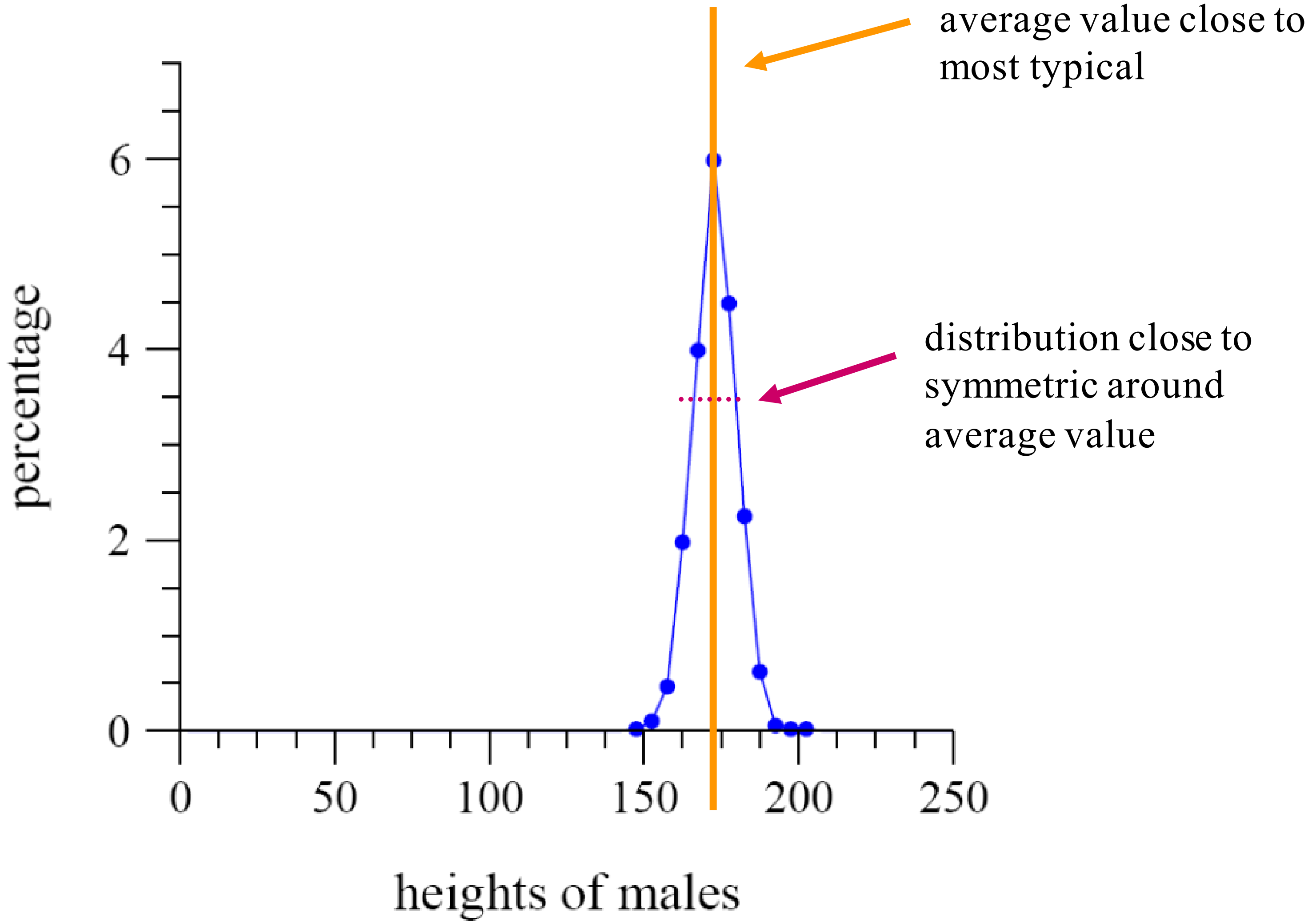
- Where alpha for the internet is about 2.1
- I.e., the fraction of web pages with k in-links is proportional to $1/k^2$

Power-law distributions

- Right skew
  - normal distribution is centered on mean
  - power-law or Zipf distribution is not
- High ratio of max to min
  - human heights (max and min not that different)
  - city sizes
- Power-law distributions have no "scale" (unlike a normal distribution)

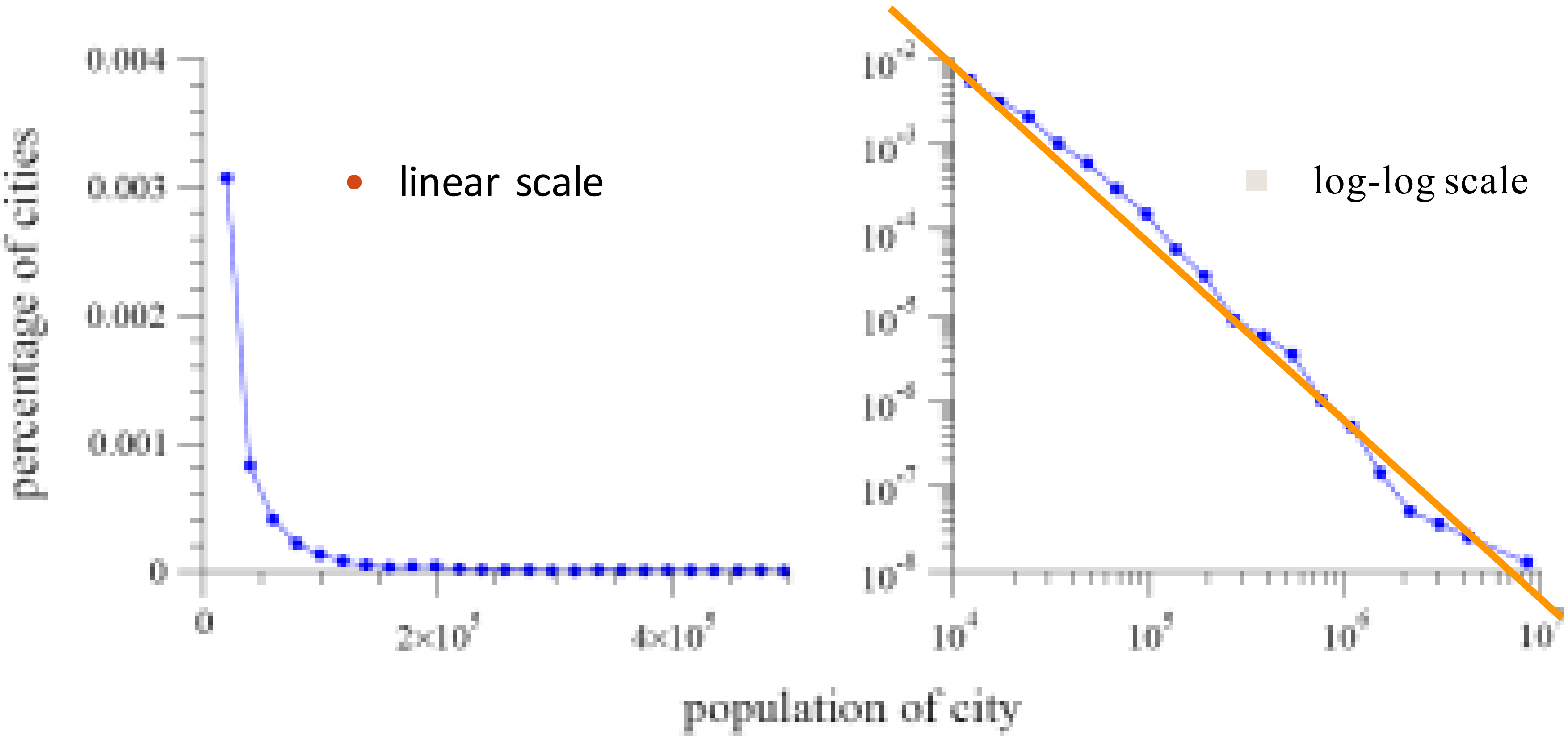
Slide from Lada Adamic

Normal (Gaussian) distribution of human heights



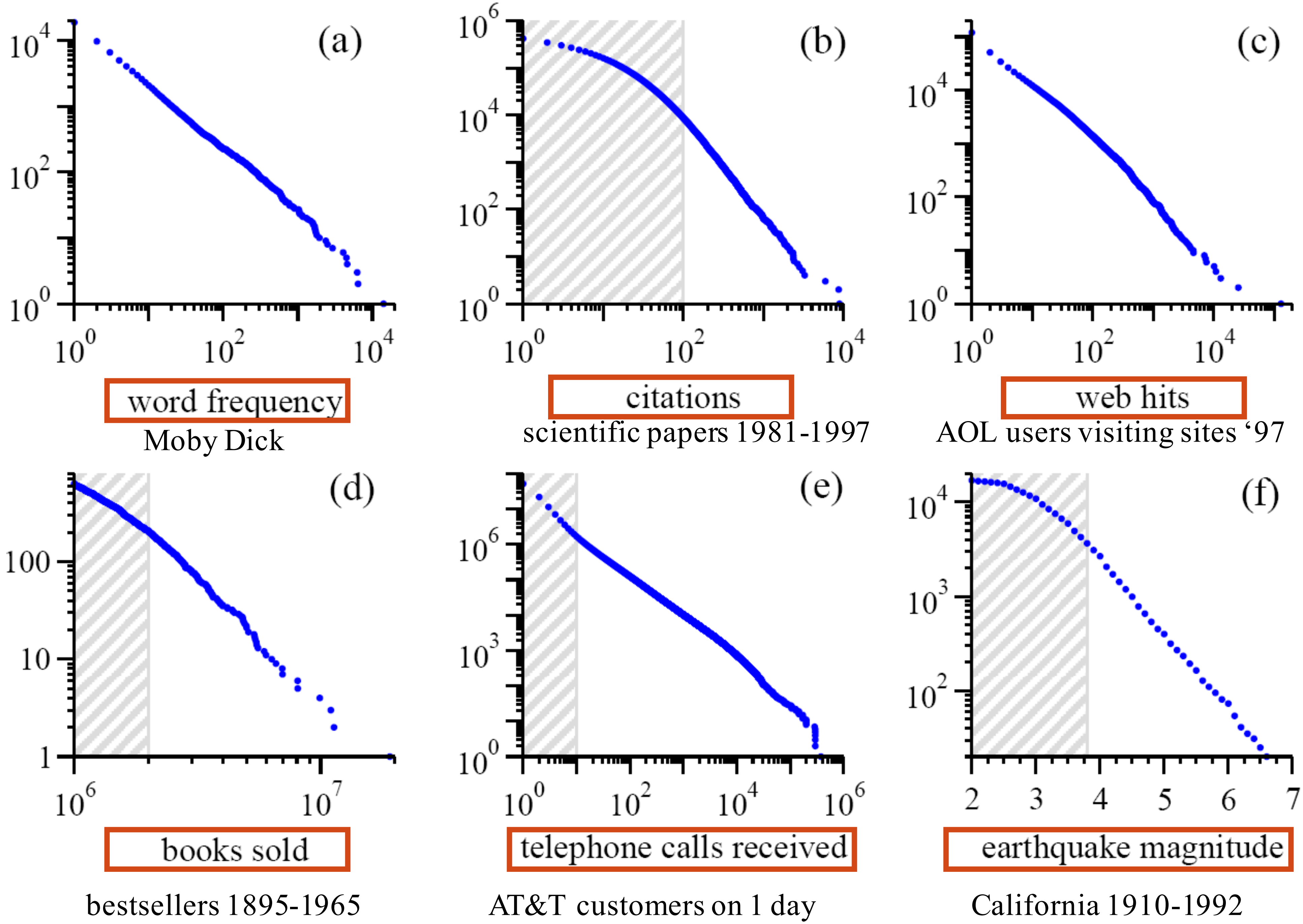
Sli

Power-law distribution

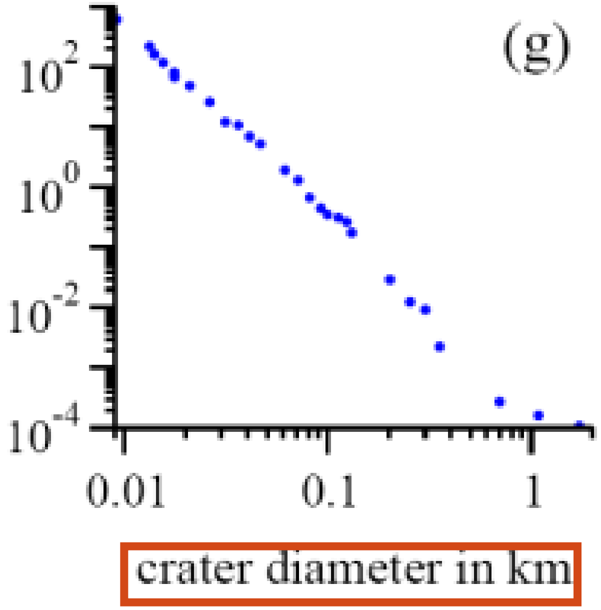


- high skew (asymmetry)
- straight line on a log-log plot

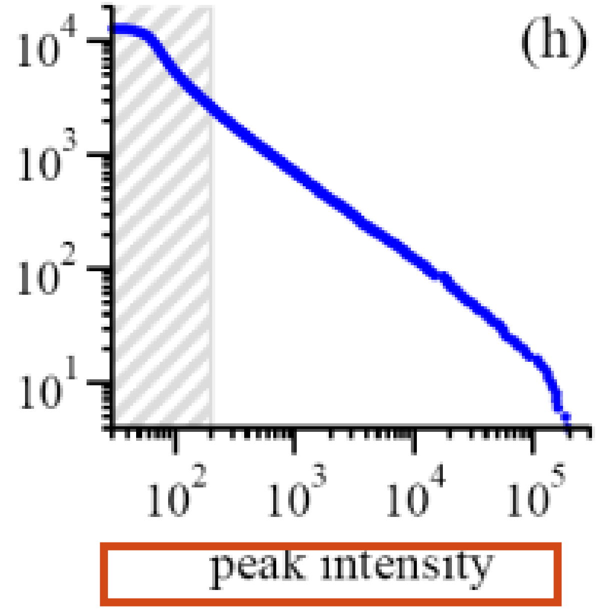
Slide from Lada Adamic


(a)
(b)
(c)
(d)
(e)
(f)
word frequency
Moby Dick
citations
scientific papers 1981-1997
web hits
AOL users visiting sites '97
books sold
bestsellers 1895-1965
telephone calls received
AT&T customers on 1 day
earthquake magnitude
California 1910-1992

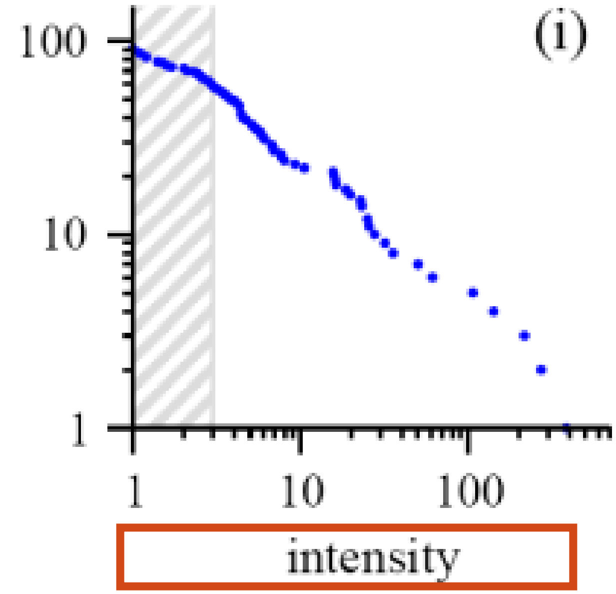
Yet more power laws



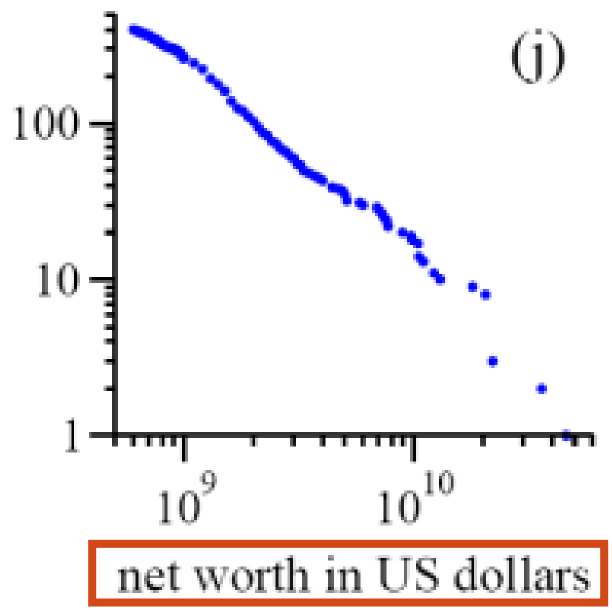
Moon



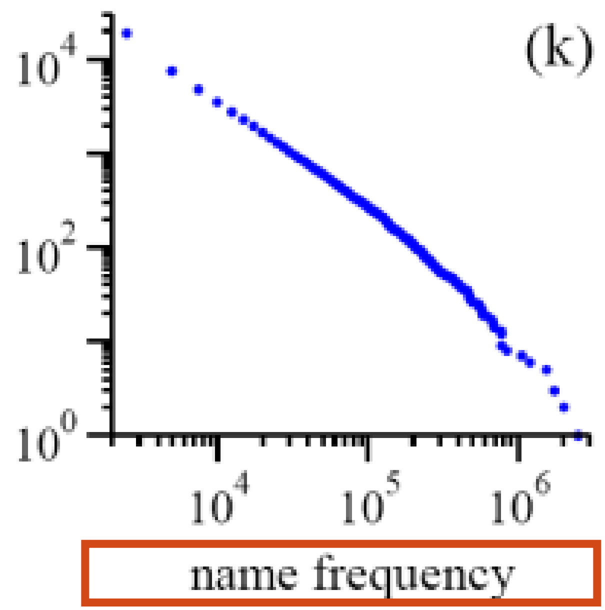
Solar flares



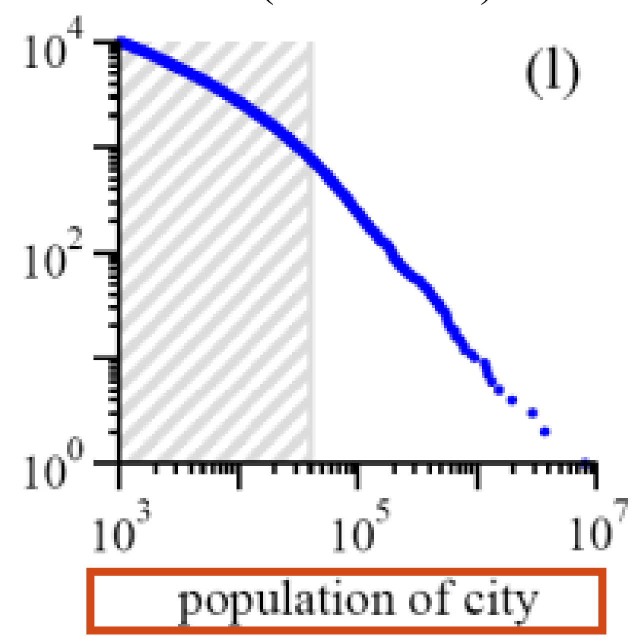
wars (1816-1980)



richest individuals 2003



US family names 1990



US cities 2003

Power law distribution

- Straight line on a log-log plot

$$\ln(p(x)) = c - \alpha \ln(x)$$

- Exponentiate both sides to get that *p(x)*, the probability of observing an item of size ‘x’ is given by

$$p(x) = Cx^{-\alpha}$$

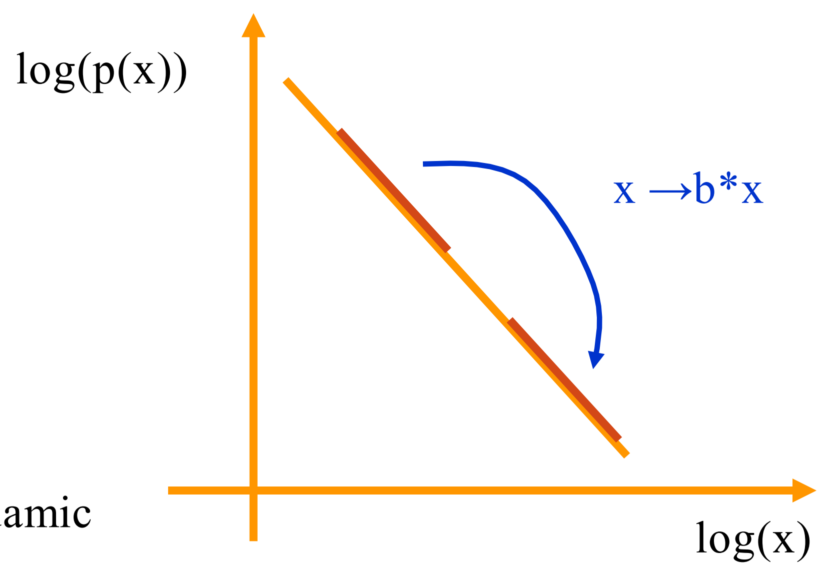
normalization constant (probabilities over all x must sum to 1)

power law exponent α

Slide from Lada Adamic

What does it mean to be scale free?

- A power law looks the same no mater what scale we look at it on (2 to 50 or 200 to 5000)
- Only true of a power-law distribution!
- $p(bx) = g(b)\ p(x)$ – shape of the distribution is unchanged except for a multiplicative constant
- $p(bx) = (bx)^{-\alpha} = b^{-\alpha}\ x^{-\alpha}$



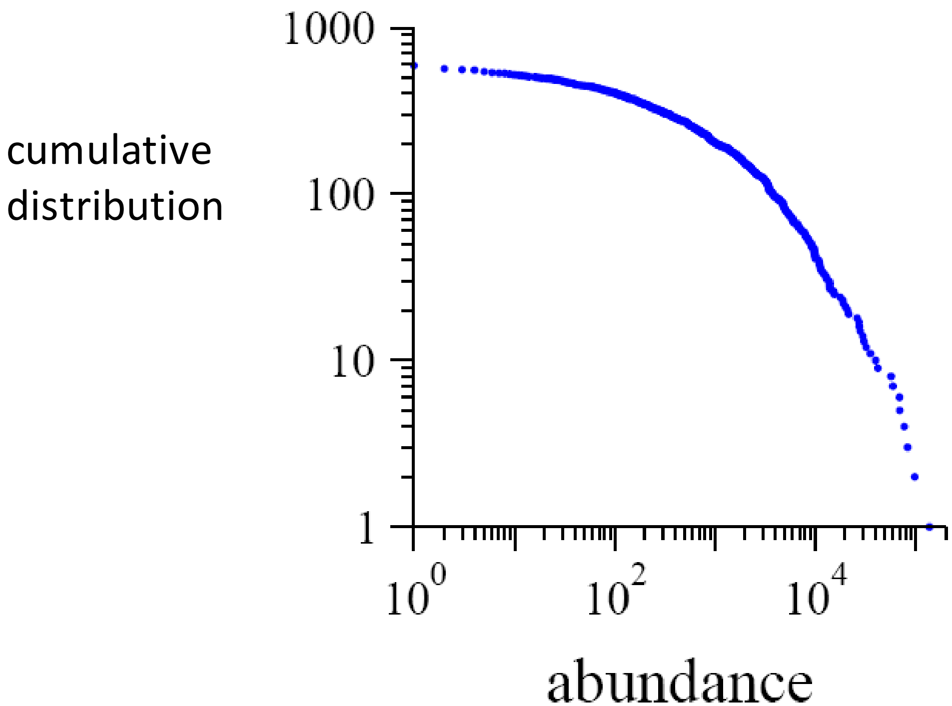
Slide from Lada Adamic

Many real world networks are power law	exponent α (in/out degree)
film actors co-appearance	2.3
telephone call graph	2.1
email networks	1.5/2.0
sexual contacts	3.2
WWW	2.3/2.7
internet	2.5
peer-to-peer	2.1
metabolic network	2.2
protein interactions	2.4

Slide from Lada Adamic

Hey, not everything is a power law

- number of sightings of 591 bird species in the North American Bird survey in 2003.



- another examples:
  - size of wildfires (in acres)

Slide from Lada Adamic

Zipf's law is a power-law

- Zipf
  - George Kingsley Zipf
    - how frequent is the 3rd or 8th or 100th most common word?
    - Intuition: small number of very frequent words (“the”, “of”)
    - lots and lots of rare words (“expressive”, “Jurafsky”)
  - **Zipf's law:** the frequency of the r'th most frequent word is inversely proportional to its rank:

    $\boldsymbol{y \sim r^{-\beta}}$, with β close to unity.

Pareto's law and power-laws

- Pareto
  - The Italian economist Vilfredo Pareto was interested in the distribution of income.
  - Pareto's law is expressed in terms of the cumulative distribution (the probability that a person earns X or more).

$$P[X > x] \sim x^{-k}$$

Slide from Lada Adamic

Income

- The fraction I of the income going to the richest P of the population is given by

Income fraction= $(100/P)^{k-1}$

- if $k = 0.5$

  top 1 percent gets $100^{-0.5} = .10$

- currently $k = 0.6$ [Jones, 2015 “Pareto and Piketty]

  top 1 percent gets $100^{-0.4} = .16$

- (higher k = more inequality)

Where do power laws come from?

- Many different processes can lead to power laws
- There is no one unique mechanism that explains it all

Slide from Lada Adamic

Preferential attachment

- Price (1965)
  - **Citation networks**
  - **new citations to a paper are proportional to the number it already has**
  - each new paper is generated with m citations
  - new papers cite previous papers with probability proportional to their in-degree (citations)

Slide from Lada Adamic

This is a "Rich get Richer" Model

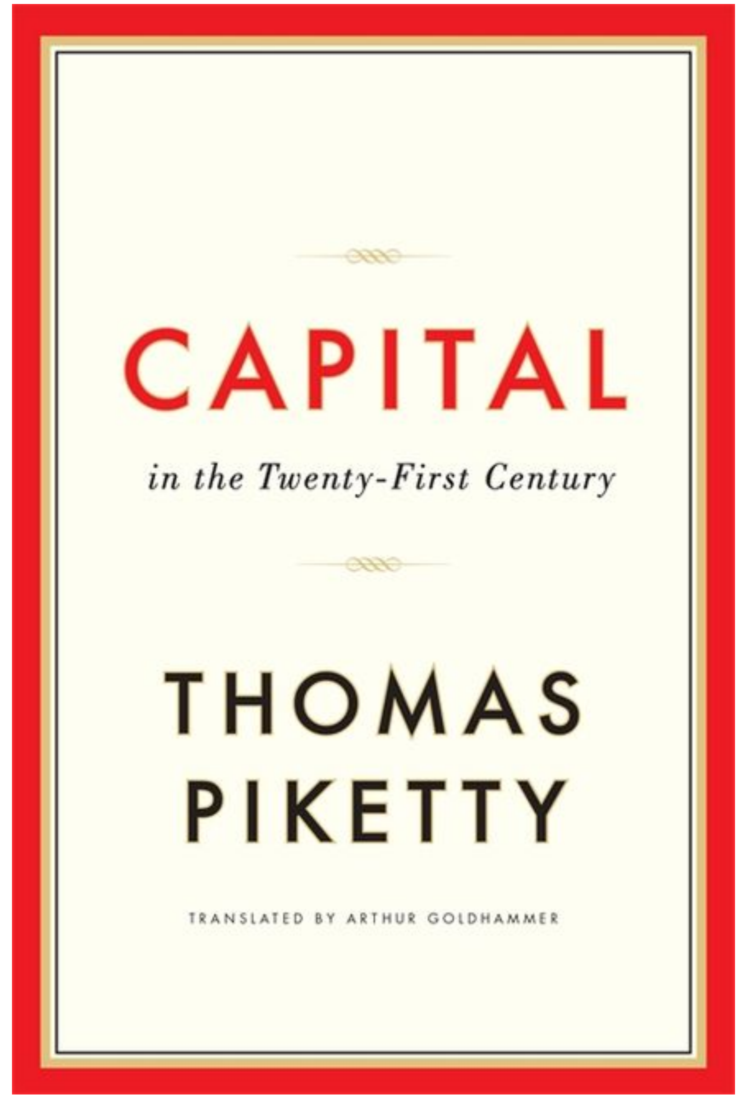
Explanation for various power law effects

1. **Citations**
2. Assume **cities** are formed at different times, and that, once formed, a city grows in proportion to its current size simply as a result of people having children
3. **Words**: people are more likely to use a word that is frequent (perhaps it comes to mind more easily or faster)

Implications: Wealth

- Thomas Piketty's book, #1 on NY Times best seller list in 2014
- Focuses on rise of inequality in wealth
- That same power law
- An equation from a Stanford economist, wealth is a power law on η:

$$\eta_{\text{wealth}} = \frac{r - g - \tau - \alpha}{n + d}$$



Power laws

- Many processes are distributed as power laws
  - Word frequencies, citations, web hits
- Power law distributions have interesting properties
  - scale free, skew, high max/min ratios
- Various mechanisms explain their prevalence
  - rich-get-richer, etc
- Explain lots of phenomena we have been dealing with
  - the use of stop words lists (a small fraction of word types cover most tokens in running text)

CS 124/LINGUIST 180
From Languages to Information

Dan Jurafsky
Stanford University

Power Laws

What classes should I take to follow up on this class?

Follow-up CS courses

Spring 2016

- CS224U: Natural Language Understanding
- CS276: Information Retrieval and Web Search
- CS224D: Deep Learning for Natural Language Processing

Fall 2016 (probably)

- CS147: Introduction to HCI Design
- CS221: Artificial Intelligence
- CS229: Machine Learning
- CS224W: Social and Information Network Analysis

Winter 2016

- CS224N: Natural Language Processing
- CS246: Mining Massive Datasets

Follow-up Linguistics courses

General:

Ling 1: Intro to Linguistics

Ling 140 Language Acquisition

Social meaning:

(Spring 2016) Ling 65: African-American Vernacular English

(Spring 2016) Ling 150: Language and Society

Ling 156: Language and Gender

Ling 1XX: The Linguistics of Advertising

Meaning/Understanding:

Ling 130a: Semantics and Pragmatics

Ling 141: Language and Gesture

Others:

Ling 105 Phonetics

LING 121a: The Syntax of English

LING 121b: Crosslinguistic Syntax

Ling 192: Language Testing