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 is in networks, not just text!
- We've seen 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 is one involved in many ties.

- **Degree centrality**: number of immediate connections a node has (undirected)
- **Prestige centrality**: everyone points to this actor (directed)
  - 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 \textit{betweenness} of a node A (or an edge A-B) =
  
  \[
  \text{number of shortest paths that go through A (or A-B)} \div \text{total number of shortest paths that exist between all pairs of nodes}
  \]
Betweenness

number of shortest paths that go through A

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

Betweenness of B?

<table>
<thead>
<tr>
<th></th>
<th># sh. paths through B</th>
<th># shortest paths</th>
</tr>
</thead>
<tbody>
<tr>
<td>AC</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>AD</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>CD</td>
<td>0</td>
<td>1</td>
</tr>
</tbody>
</table>

Centrality = \( 1/4 \)
An example network

- Network of which students have had romantic relationships with each other in a high school.
- Important for studying disease spread, very relevant right now!
- 1993 study considering only heterosexual relationships
- What do you think its shape is?
- For example: is it core-periphery (like the web)?
High school dating

Image drawn by Mark Newman


Slide from Drago Radev

Image drawn by Mark Newman
The Structure of Romantic and Sexual Relations at "Jefferson High School"

Each circle represents a student and lines connecting students represent romantic relations occurring 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 (spanning tree, no cycles)?

- Teens probably don’t say:
  - “By selecting this partner, I maximize the probability of inducing a spanning tree with no short cycles.”
- The “microtaboo” Bearman and Moody propose
  - don’t date your ex-girlfriend’s boyfriend’s ex-girlfriend
    - (or the reverse)
    - (or a generalization with homosexual/non-binary nodes)
  - a simulation shows this constraint results in spanning tree that avoids short cycles
CS 124/LINGUIST 180
From Languages to Information

Dan Jurafsky
Stanford University

Small Worlds
Small worlds

Slide from Lada Adamic

"HEY... I HAVE AN AUNT SHIRLEY FROM TOLEDO TOO!"
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
Erdős numbers are small too
The Small World Experiment

What is the typical shortest path between any two people?

- 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

Stanley Milgram (1967)
Milgram’s small world experiment

It took 6.2 steps on average

“Six degrees of separation”

Can we check this computationally?
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 degrees” hypothesis

- Hungarian writer Karinthy’s 1929 play “Chains” (Láncszemek)
Duncan Watts: Networks, Dynamics and the Small-World Phenomenon

- Why do we see the small world pattern?
- What implications does it have 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** - \( O(\text{Billions}) \)
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
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 \( C \).
Local clustering coefficient (Watts & Strogatz 1998)

• For a vertex $i$

$C = \text{The fraction of pairs of neighbors of the node that are connected}\n\text{“What percentage of your friends know each other?”}$

• Let $n_i$ be the number of neighbors of vertex $i$

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

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

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

Slide from Lada Adamic
Local clustering coefficient

(Watts & Strogatz 1998)

• Average $C_i$ over all $n$ nodes

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

• $C_i$ for one node $i$

Node $i$ has four neighbors ($n_i = 4$)
Max # of connections: $4 \times 3/2 = 6$
# of connections present = 3
$C_i = 3/6 = 0.5$

Slide adapted 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.
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
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

**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.
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
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, D-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.

![Diagram showing triangle formation due to triadic closure.](image)
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.
Closure and bridges

- 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

Slide from James Moody

weak ties will tend to have low transitivity
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

Slide from James Moody
More evidence for strength of weak ties

In the Milgram small world experiments, acquaintance 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 befriend each other
- Local bridges will often be weak ties
- Information comes over weak ties
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
  • Populations of cities are power-law distributed
  • Contrast: human heights are not (max and min not that different)

Slide from Lada Adamic
Normal (Gaussian) distribution of human heights

average value close to most typical

distribution close to symmetric around average value

Slide from Lada Adamic
Power-law distribution

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

Slide from Lada Adamic
Power laws are seemingly everywhere.

- Moby Dick
- Scientific papers 1981-1997
- AOL users visiting sites '97
- Bestsellers 1895-1965
- Telephone calls received
- California 1910-1992
Yet more power laws

- Crater diameter in km (Moon)
- Peak intensity (Solar flares)
- Intensity of wars (1816-1980)
- Net worth in US dollars (richest individuals 2003)
- Name frequency (US family names 1990)
- Population of city (US cities 2003)
Power law distribution are straight lines on log-log plots

- Let $p(x)$ be the probability of observing an item with value $x$
- Power law:
  \[
  p(x) = \frac{c}{x^\alpha} = cx^{-\alpha}
  \]
- Take the log of both sides
  \[
  \log(p(x)) = \log(c) - \alpha \log(x)
  \]
- A line with $-\alpha$ as slope
Power laws are "scale free"

- A power law looks the same no matter 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
- \( p(bx) = (bx)^{-\alpha} = b^{-\alpha} x^{-\alpha} \)
Many real world networks are power law

<table>
<thead>
<tr>
<th>Network</th>
<th>Exponent $\alpha$ (In/Out degree)</th>
</tr>
</thead>
<tbody>
<tr>
<td>film actors co-appearance</td>
<td>2.3</td>
</tr>
<tr>
<td>telephone call graph</td>
<td>2.1</td>
</tr>
<tr>
<td>email networks</td>
<td>1.5/2.0</td>
</tr>
<tr>
<td>sexual contacts</td>
<td>3.2</td>
</tr>
<tr>
<td>WWW</td>
<td>2.3/2.7</td>
</tr>
<tr>
<td>internet</td>
<td>2.5</td>
</tr>
<tr>
<td>peer-to-peer</td>
<td>2.1</td>
</tr>
<tr>
<td>metabolic network</td>
<td>2.2</td>
</tr>
<tr>
<td>protein interactions</td>
<td>2.4</td>
</tr>
</tbody>
</table>
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:

\[ y \sim r^{-\beta} \], with \( \beta \) 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).

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

• The fraction I of the income going to the richest P of the population:

\[ I = \frac{100}{P}\eta - 1 \]

• if \( \eta = 0.5 \)
  top 1 percent gets \( 100^{-0.5} = 0.10 \)

• Currently \( \eta = 0.6 \)  [Jones, 2015 “Pareto and Piketty”]
  top 1 percent gets \( 100^{-0.4} = 0.16 \)

• (higher \( \eta \) = more inequality)

• Thomas Piketty’s book, #1 on NY Times best seller list in 2014, studies how \( \eta \) relates to rise of wealth 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
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)
This is a “Rich get Richer” Model

Explanation for various power law effects

1. **Citations**: randomly draw citations from the reference section of papers

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)
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)
What classes should I take to follow up on this class?
Follow up CS Language courses

Spr 2020:  cs222U Natural Language Understanding
Spr 2020:  cs224S Spoken Language Processing
(Spr 2020:  cs346 Ethical and Social Issues in NLP)

Aut 2021:  cs224W Machine Learning with Graphs
Win 2021:  cs224N Natural Language Processing w/Deep Learning
Win 2021:  cs246 Mining Massive Data Sets
Spr 2021:  cs276 Information Retrieval and Web Search
Follow-up Linguistics courses

This Spring!

LINGUIST 134A: Structure of Discourse
LINGUIST 130B: Introduction to Lexical Semantics
LINGUIST 127: Linguistic Meaning and Legal Interpretation

Next year:

LINGUIST 1: Intro to Linguistics
LINGUIST 150: Language and Society
LINGUIST 156: Language and Gender
LINGUIST 130a: Semantics and Pragmatics
Rest of class

- **Tuesday March 10**: NLP For Social Good Lecture, by *video* (and probably also optionally in person)
  - I will ask questions on the final about these lectures
  - Also quiz on Networks due at 11:59pm
- **Thursday March 12**: Review session will go *virtual* due to COVID-19, in some way; details being worked out.
- Finals will also go online, details being worked out
  - **Tuesday May 17 12:15-3:15**
  - **Thursday May 19 12:15-3:15**