Network Models

CS 579 Online Social Network Analysis

Dr. Cindy Hood
11/20/25

Remaining Exam and Deliverables

- Final project progress report video posted
  - Due 11/21
- HW #5 posted
  - Problems that will help you prepare for Exam 2
  - Due 11/24 (No late days)
  - Social Media Mining book posted on Canvas
- Exam 2
  - Cumulative but will focus on material covered after Exam 1
  - 12/2
- Final project poster presentation/video (online students)
  - 12/4
  - Section 2 students please fill out survey if you haven't done so
- Final project report
  - Due by midnight on 12/10

Final Project Progress Report

- You will create a presentation about your progress on the final project to date and then submit a video of you/your team presenting the slides. Each student should speak and the speaker should be shown in the video while they are presenting. The video should be 2-4 minutes long.
- The presentation should include:
- Intro slide with the title of the project and student name(s)
- A summary of the project components clearly illustrating pieces that are being reused from HW 4 along with other components and how the components fit together.
- A plan for completing the project highlighting the starting point (what was done for HW4) and the steps to be completed along with discussion on progress made to date.
- You will submit a link to your video presentation. Please be sure that it is accessible to Prof Hood and the TAs. One submission per team.

Final Project Presentation

- Poster session
  - Outside of SB 104 during class time 11:25am – 12:40pm 12/4
    - Will start before class and continue after
    - I will send out an email for you to schedule a timeslot
    - Plan to be present for the entire class period
      - For you to complete your reviews
      - For others to complete their review of your poster
  - I have mounting putty/tape to stick your posters to the wall there
  - Pizza will be served
- Online video
  - Must be posted to google folder (I will send link) by 11:00am 12/4

Poster/Presentation Requirements

- Must use printed poster (on campus) or slides (online)
- Each team member must present and be visible for online presentation
- No more than 5 minutes
- Content
  - Title and names of all team members
  - Intro – What did you do
  - Design of project
  - Execution of project
    - Data utilized
  - Results
  - Discussion of results and conclusion
    - What worked/what didn’t work?
    - How can you evaluate your project?
      - What did you do?
      - What could you do?
    - What surprised you?
    - What would you do if you had more time?

Focus on the most interesting parts of the project

Poster Creation

- Your poster should be 24” x 36”
- Powerpoint
  - https://designshack.net/articles/business-articles/how-to-make-a-poster-in-powerpoint/

Poster Printing

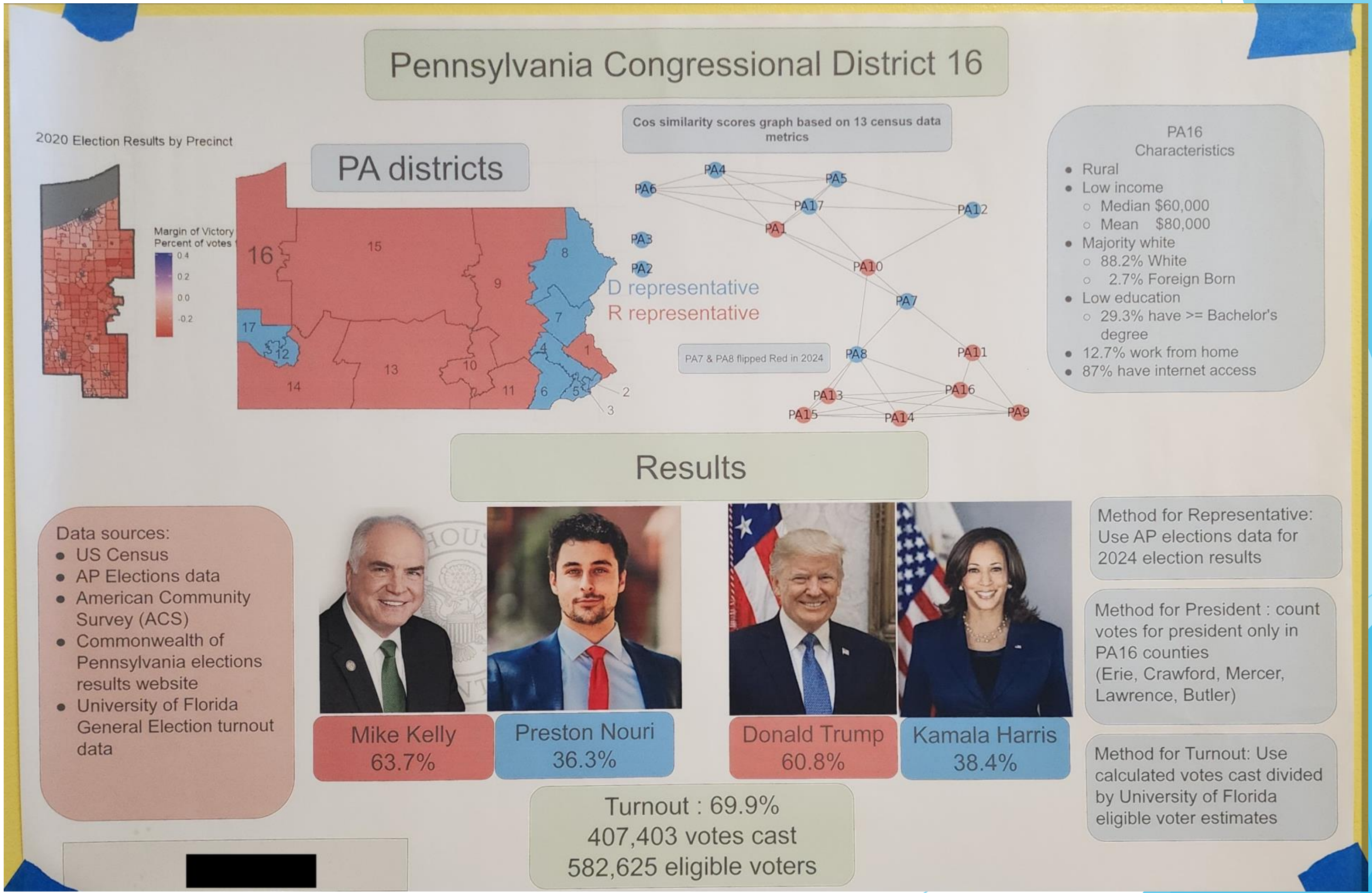
How to Print to a Plotter

Using one of the HP plotters on the Mies Campus (located in Crown and Tech North) takes some additional preparation to get a quality plot and avoid waisting your printing funds. The OTS staff have produced a series of four videos to help you learn how to plot properly. Please review each of these videos before attempting to use a plotter.

1. Plotter Introduction – Preparing Your File
2. Z6200 Plotter Instructional Video

https://www.iit.edu/ots/services/student-printing

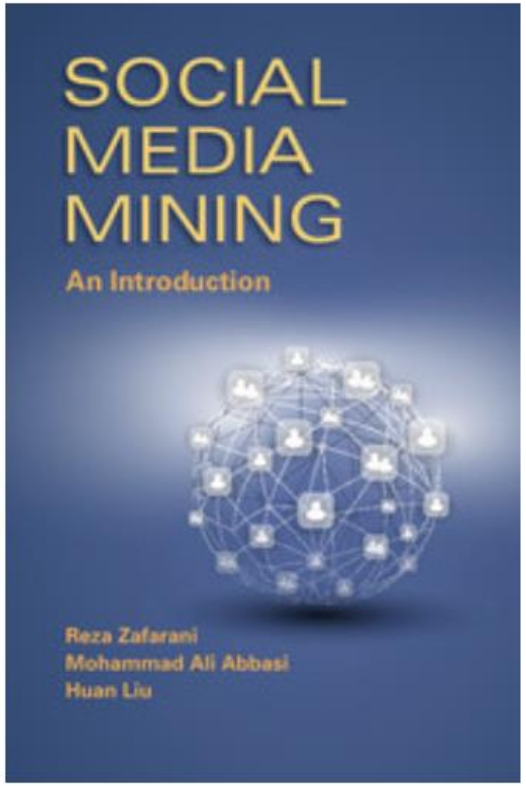
Example from last fall



Project presentation review assignments

- All reviews must be completed on 12/4
  - On campus during class time
  - Online

References



http://www.socialmediamining.info

Some additional resources

- Myatt and Johnson (2014), *Making Sense of Data I*, 2nd Edition, Wiley, ISBN: 978-1-118-40741-7
  - https://onlinelibrary.wiley.com/doi/epdf/10.1002/9781118422007
- *Speech and Language Processing*, Dan Jurafsky and James H. Martin, https://web.stanford.edu/~jurafsky/slp3/
- *An Introduction to Statistical Learning*, Gareth James, Daniela Witten, Trevor Hastie, Robert Tibshirani, Jonathan Taylor (Python version only)
  - https://www.statlearning.com
- Good YouTube channel for getting intuition
  - https://www.youtube.com/@statquest/videos
- Networks, Crowds, and Markets: Reasoning About a Highly Connected World by David Easley and Jon Kleinberg.
  - http://www.cs.cornell.edu/home/kleinber/networks-book/

Unimodal networks

- The network metrics and analysis techniques covered are for unimodal networks
  - Nodes are all the same type of object
    - Person
    - Event
    - Place
    - Etc.
  - Edges are all the same type of relationship
- How to analyze networks with
  - Multiple types of relationships
  - Multiple types of nodes

Multiple relationships

- Social relationships are commonly more complex than a single kind of relationship
  - There can be multiple types of connections simultaneously
- Assumption is that a person's ”behavior is strongly shaped by the complex interaction of many simultaneous constraints and opportunities arising from how the individual is embedded in multiple kinds of relationships”
- “The characteristics and behavior of whole populations may depend on multiple dimensions of integration/cleavage. Solidarity may be established by economic exchange, shared information, kinship, and other ties operating simultaneously.”

https://faculty.ucr.edu/~hanneman/nettext/C16_Multi_Plex.html

Multiple relationships

- Relationships can be directed or not
- Relationships can have different types of labels

Approach for multiple relationships

- Count relationships
- Combine relationships into single metric
- Separate graphs for each relationship
  - Analyze separately
  - Multilevel graphs
    - Connections

https://faculty.ucr.edu/~hanneman/nettext/C16_Multi_Plex.html

Two-mode networks

- Typically social media data measures relationships at the micro level and use analysis techniques to infer presence of social structure at the macro level
- Data with two different types of nodes can provide insight at both micro and macro level
- Ex/ People attending events
  - At micro level, modeling behavior of people
  - At macro level, seeing the structure of the events
- Bipartite graphs
  - Often Affiliation networks

https://faculty.ucr.edu/~hanneman/nettext/C17_Two_mode.html

There are many different ways to create models

- How do you test your models?
  - Verification
    - Is your implementation correct?
  - Validation
    - Is your model doing what it needs to for your problem?

- It is sometimes said that validation can be expressed by the query "Are you building the right thing?"[12] and verification by "Are you building it right?".[12]

  12. ^ a b Barry Boehm, *Software Engineering Economics*, 1981

https://en.wikipedia.org/wiki/Verification_and_validation

How can you evaluate your model(s) for the final project?

Network Models

- Model-Driven Models!

Random graphs

Small-World Model

Preferential Attachment

Random Graphs

Random Graphs

- We have to assume how friendships are formed
  - The most basic form:

Random Graph assumption:
Edges (i.e., friendships) between nodes (i.e., individuals) are formed randomly.

We discuss two random graph models $G(n, p)$ and $G(n, m)$

Random Graph Model - $G(n,p)$

- Consider a graph with a fixed number of nodes n
- Any of the $\binom{n}{2}$ edges can be formed independently, with probability p
- The graph is called a $G(n,p)$ *random graph*

Proposed independently by Edgar Gilbert and by Solomonoff and Rapoport.

Random Graph Model - $G(n, m)$

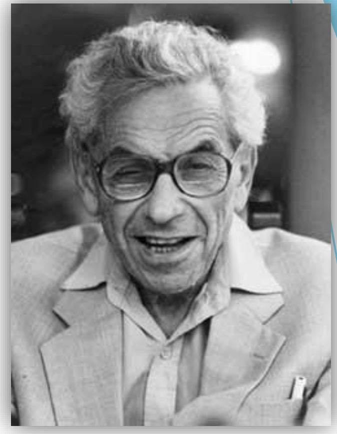
- Assume both number of nodes n and number of edges m are fixed.

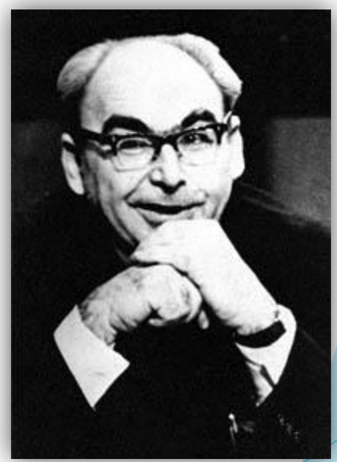
- Determine which m edges are selected from the set of possible edges

- Let Ω denote the set of graphs with n nodes and m edges
  - There are $|\Omega|$ different graphs with n nodes and m edges

$$|\Omega| = \binom{\binom{n}{2}}{m}$$

- To generate a random graph, we uniformly select one of the $|\Omega|$ graphs (the selection probability is $1/|\Omega|$)





This model was first proposed by *Paul Erdös* and *Alfred Rényi*

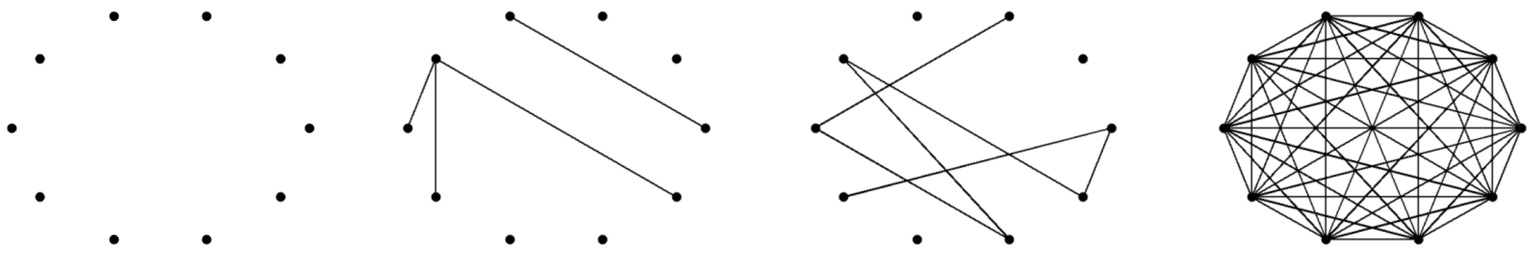
Evolution of Random Graphs

- Create your own Random Graph Evolution demo: https://github.com/dgleich/erdosrenyi-demo

The Giant Component

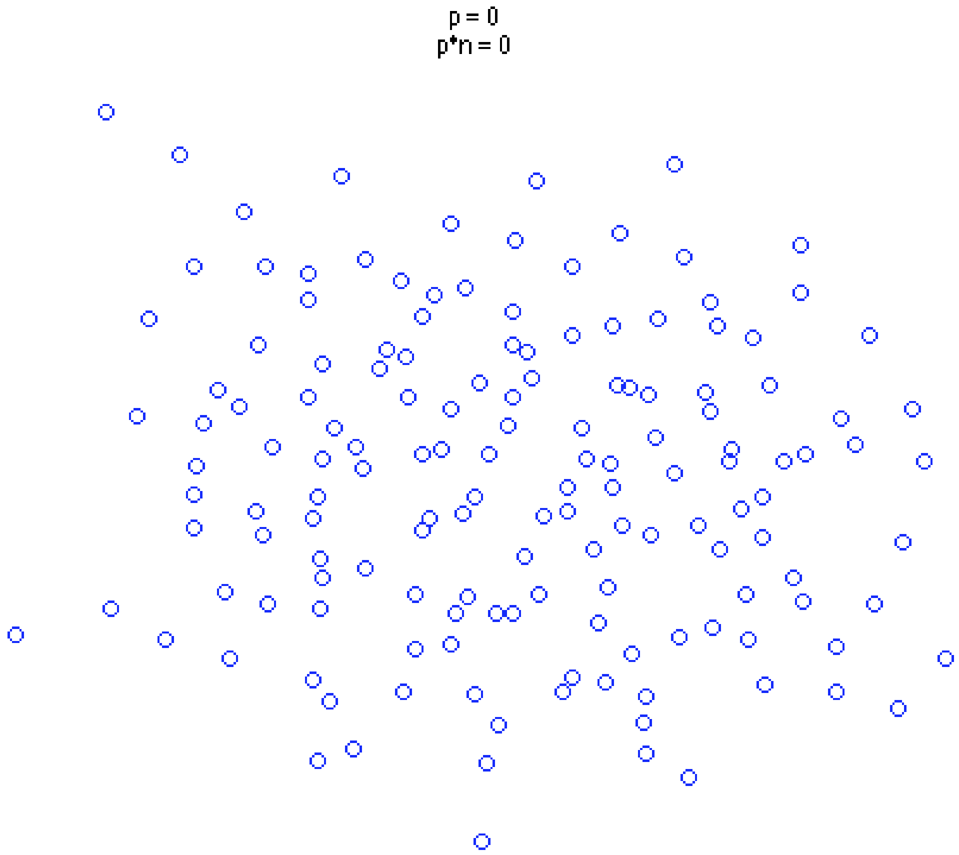
- In random graphs, as we increase p, a large fraction of nodes start getting connected
  - i.e., we have a path between any pair
- This large fraction forms a connected component:
  - **Largest connected component**, also known as the **Giant component**
- In random graphs:
  - $p = 0$
    - the size of the giant component is 0
  - $p = 1$
    - the size of the giant component is n

The Giant Component



Probability (p)	0.0	0.055	0.11	1.0
Average Node Degree (c)	0.0	0.8	≈1	n-1=9
Diameter	0	2	6	1
Giant Component Size	0	4	7	10
Average Path Length	0.0	1.5	2.66	1.0

Demo ($n = 150$)



From *David Gleich*

1st Phase Transition (Rise of the Giant Component)

- **Phase Transition**: the point where diameter value starts to shrink in a random graph
  - We have other phase transitions in random graphs
    - E.g., when the graph becomes connected
- The phase transition we focus on happens when
  - average node degree $c = 1$ (or when $p = 1/(n-1)$)
- At this Phase Transition:
  1. The giant component, which just started to appear, starts to grow, and
  2. The diameter, which *just* reached its maximum value, starts decreasing.

Random Graphs

If $c < 1$:

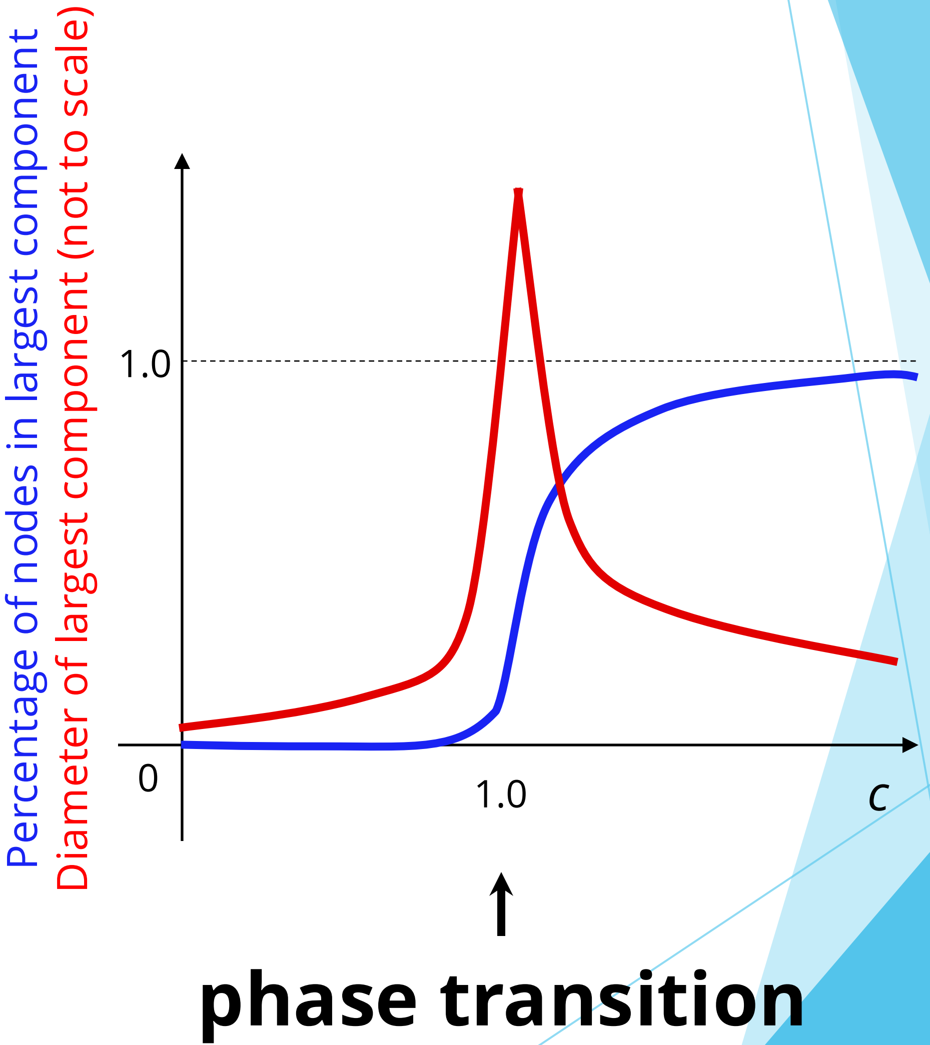
- **small**, isolated clusters
- **small** diameters
- **short** path lengths

At $c = 1$:

- a ***giant component*** appears
- diameter **peaks**
- path lengths are **long**

For $c > 1$:

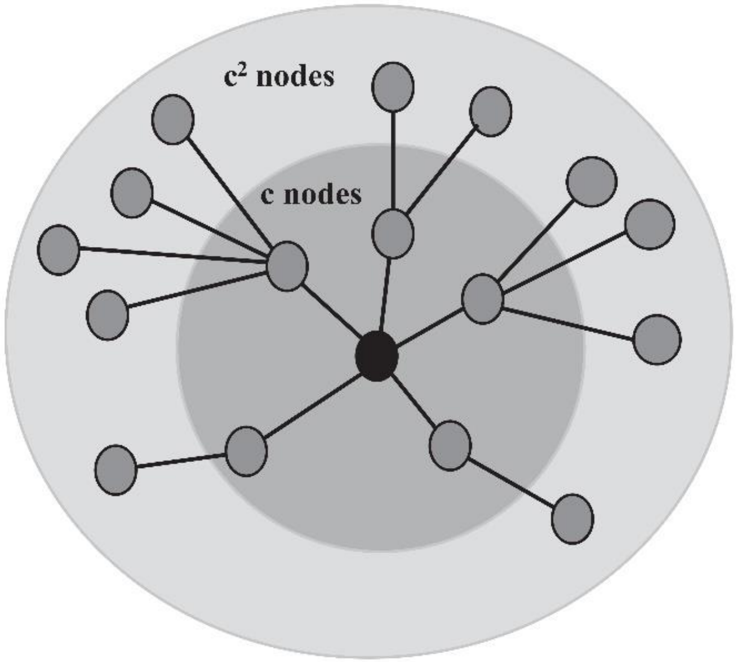
- almost **all** nodes **connected**
- diameter **shrinks**
- path lengths **shorten**



Why $c = 1$? [Rough Idea]

Consider a random graph with expected node degree c

- In this graph,
  - Consider any **connected** set of nodes S;
  - Let $S' = V - S$ denote the complement set; and
  - Assume $|S| << |S'|$.
- For any node v in S,
  - If we move one hop away from v, we visit approximately c nodes.
- If we move one hop away from nodes in S,
  - we visit approximately $|S|c$ nodes.



- If S is small, the nodes in S only visit nodes in S' and when moving one hop away from S, the set of nodes *guaranteed to be connected* gets larger by a factor c.
- In the limit, if we want this connected component to become the largest component, then after traveling n hops, its size must grow and we must have

$$c^n \geq 1 \text{ or equivalently } c \geq 1$$

Properties of Random Graphs

Real-World Networks / Simulated Random Graphs

	Original Network				Simulated Random Graph	
Network	*Size*	*Average Degree*	*Average Path Length*	C	*Average Path Length*	C
Film Actors	225,226	61	3.65	0.79	2.99	0.00027
Medline Coauthorship	1,520,251	18.1	4.6	0.56	4.91	1.8×10^{-4}
E.Coli	282	7.35	2.9	0.32	3.04	0.026
C.Elegans	282	14	2.65	0.28	2.25	0.05

Small-World Model

Small-world Model

- Small-world model
  - or the **Watts-Strogatz (WS)** model
  - A special type of random graph
  - Exhibits small-world properties:
    - Short average path length
    - High clustering coefficient

- It was proposed by Duncan J. Watts and Steven Strogatz in their joint 1998 Nature paper

Watts, Duncan J., and Steven H. Strogatz. "Collective dynamics of 'small-world'networks." *nature* 393.6684 (1998): 440-442.

Small-world Model

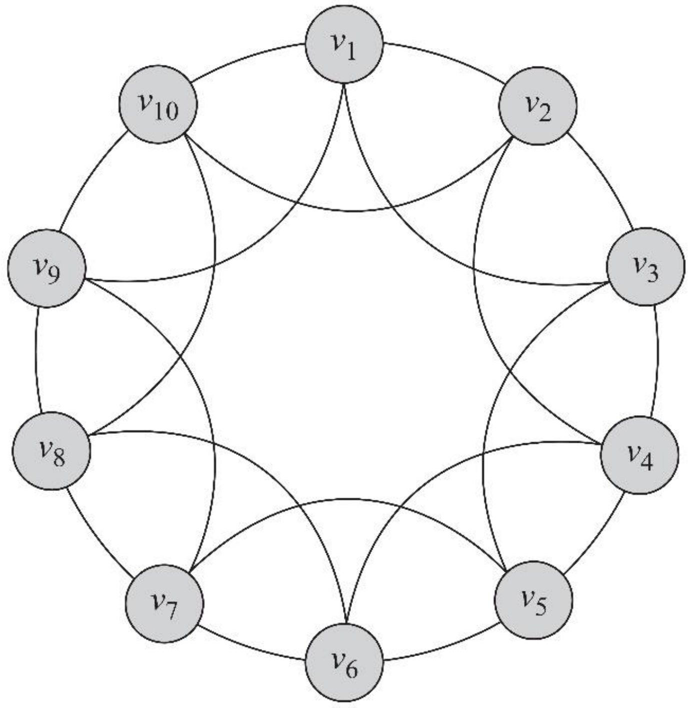
- In real-world interactions, many individuals have a limited and often at least, a fixed number of connections

- In graph theory terms, this assumption is equivalent to embedding users in a regular network

- A regular (ring) lattice is a special case of regular networks where there exists a certain pattern on how ordered nodes are connected to one another

- In a regular lattice of degree c, nodes are connected to their previous $c/2$ and following $c/2$ neighbors

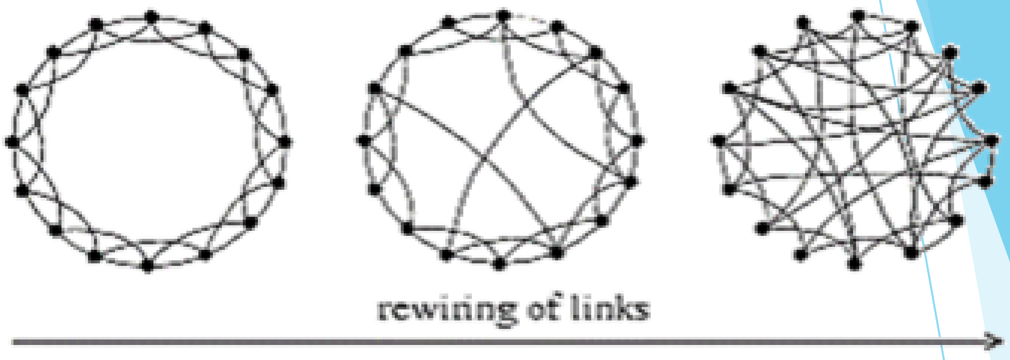
- Formally, for node set V={v_1, v_2, v_3,…, v_n}, an edge exists between node i and j if and only if



$$0 \leq \min(n - |i-j|, |i-j|) \leq c/2$$

Generating a Small-World Graph

- The lattice has a **high**, but **fixed**, clustering coefficient
- The lattice has a **high** average path length



- In the small-world model, a parameter $0 \le \beta \le 1$ controls randomness in the model
  - When β is 0, the model is basically a regular lattice
  - When $\beta = 1$, the model becomes a random graph
- The model starts with a regular lattice and starts adding random edges [through **rewiring**]
  - **Rewiring:** take an edge, change one of its end-points randomly

Constructing Small World Networks

Algorithm 4.1 Small-World Generation Algorithm

Require: Number of nodes $|V|$, mean degree c, parameter β

return A small-world graph $G(V, E)$
G = A regular ring lattice with $|V|$ nodes and degree c
for node v_i (starting from v_1), and all edges $e(v_i, v_j)$, $i < j$ **do**
    v_k = Select a node from V uniformly at random.
    if rewiring $e(v_i, v_j)$ to $e(v_i, v_k)$ does not create loops in the graph or multiple edges between v_i and v_k **then**
        rewire $e(v_i, v_j)$ with probability β: $E = E - \{e(v_i, v_j)\}$, $E = E \cup \{e(v_i, v_k)\}$;
    end if
end for
Return $G(V, E)$

As in many network generating algorithms

- Disallow self-edges
- Disallow multiple edges

Small-World Model Properties

Degree Distribution

- The degree distribution for the small-world model is

$$P(d_v = d) = \sum_{n=0}^{\min(d-c/2,c/2)} \binom{c/2}{n} (1-\beta)^n \beta^{c/2-n} \frac{(\beta c/2)^{d-c/2-n}}{(d-c/2-n)} e^{-\beta c/2}$$

- In practice, in the graph generated by the small world model, most nodes have similar degrees due to the underlying lattice.

Regular Lattice vs. Random Graph

- Regular Lattice:
  - Clustering Coefficient (**high**):

$$\frac{3(c-2)}{4(c-1)} \approx \frac{3}{4}$$

  - Average Path Length (**high**): $n/2c$

- Random Graph:
  - Clustering Coefficient (**low**): p
  - Average Path Length (**ok!**) : $\ln|V|/\ln c$

What happens in Between?

- Does smaller average path length mean smaller clustering coefficient?

- Does larger average path length mean larger clustering coefficient?

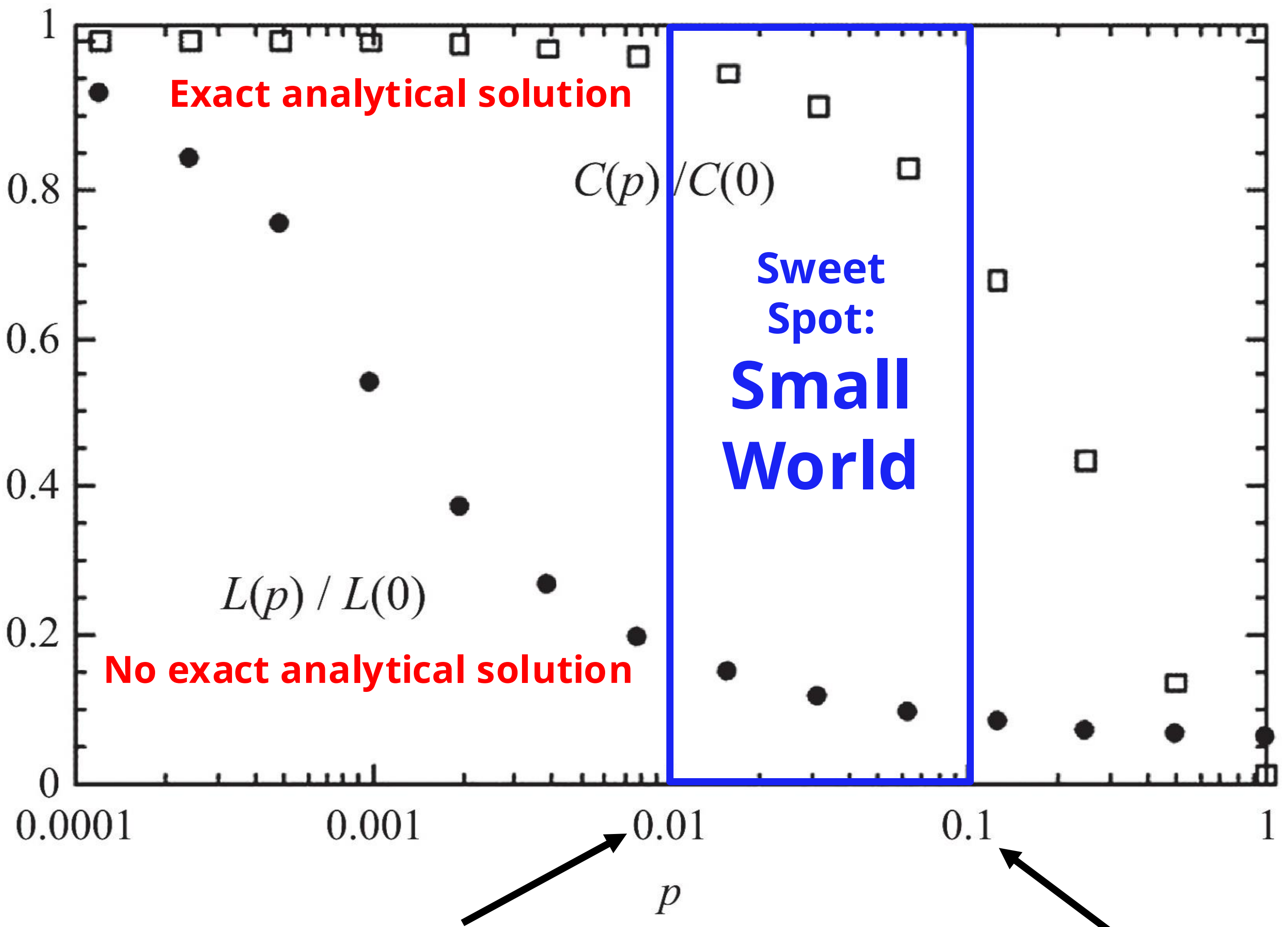
Numerical simulation:

- We increase p (i.e., β) from 0 to 1
- Assume
  - $L(0)$ is the average path length of the regular lattice
  - $C(0)$ is the clustering coefficient of the regular lattice
  - For any $p, L(p)$ denotes the average path length of the small-world graph and $C(p)$ denotes its clustering coefficient

Observations:

- Fast decrease of average distance $L(p)$
- Slow decrease in clustering coefficient $C(p)$

Change in Clustering Coefficient /Avg. Path Length



1% of links rewired

10% of links rewired

Clustering Coefficient for Small-world model

- The probability that a connected triple stays connected after rewiring consists of

  1. The probability that none of the 3 edges were rewired is $(1-p)^3$
  2. The probability that other edges were rewired back to form a connected triple
     - Very small and can be ignored

- Clustering coefficient

$$C(p) \approx (1-p)^3 C(0)$$

Modeling with the Small-World Model

- Given a real-world network in which average degree is c and clustering coefficient C is given,
  - we set $C(p) = C$ and determine β $(= p)$ using equation

$$C(p) \approx (1-p)^3 C(0)$$

- Given β, c, and n (size of the real-world network), we can simulate the small-world model

Real-World Network and Simulated Graphs

	Original Network				Simulated Graph	
Network	*Size*	*Average Degree*	*Average Path Length*	*C*	*Average Path Length*	*C*
Film Actors	225,226	61	3.65	0.79	4.2	0.73
Medline Coauthorship	1,520,251	18.1	4.6	0.56	5.1	0.52
E.Coli	282	7.35	2.9	0.32	4.46	0.31
C.Elegans	282	14	2.65	0.28	3.49	0.37

Preferential Attachment Model

Preferential Attachment Model

- **Main assumption:**
  - When a new user joins the network, the probability of connecting to existing nodes is proportional to existing nodes' degrees
  - For the new node v
    - Connect v to a random node v_i with probability

$$P(v_i) = \frac{d_i}{\sum_j d_j}$$

  - Proposed by Albert-László Barabási and Réka Albert
    - A special case of the Yule process

Distribution of wealth in the society:

The rich get richer

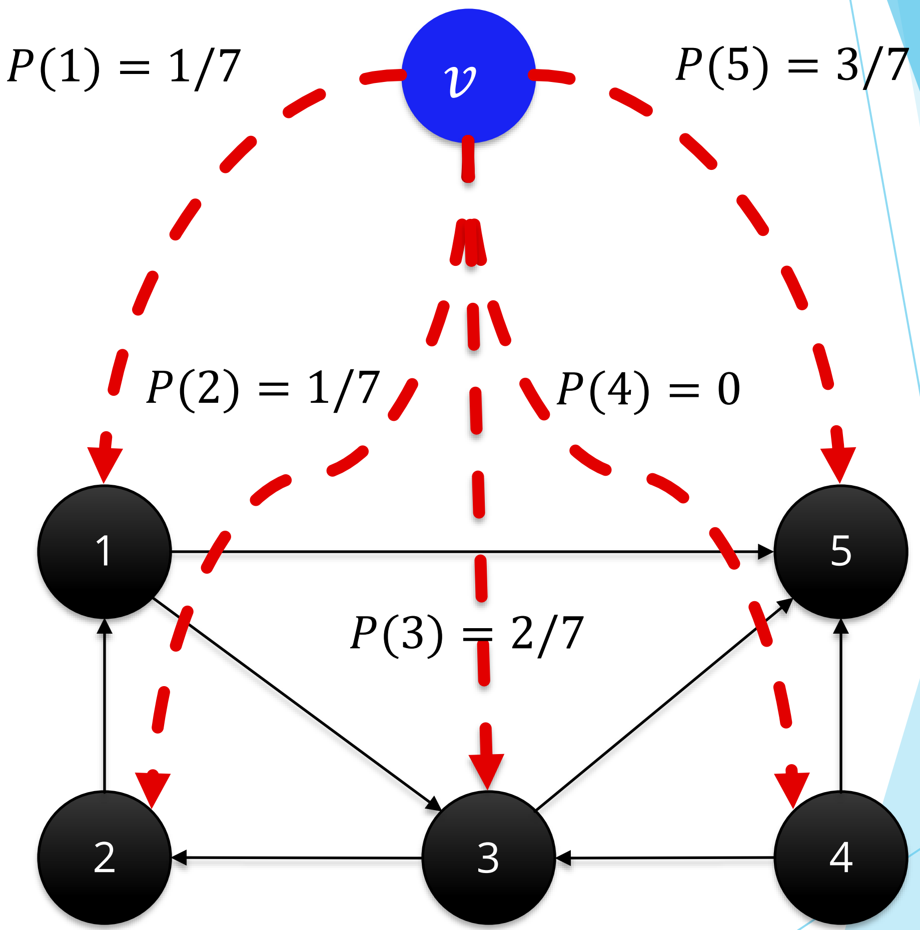
Barabási, Albert-László, and Réka Albert. "Emergence of scaling in random networks." *science* 286.5439 (1999): 509-512.

Preferential Attachment: Example

- Node v arrives

$$P(v_i) = \frac{d_i}{\sum_j d_j}$$

- $P(1) = 1/7$
- $P(2) = 1/7$
- $P(3) = 2/7$
- $P(4) = 0$
- $P(5) = 3/7$



Constructing Scale-free Networks

Algorithm 4.2 Preferential Attachment

Require: Graph $G(V_0, E_0)$, where $|V_0| = m_0$ and $d_v \geq 1 \; \forall \; v \in V_0$, number of expected connections $m \leq m_0$, time to run the algorithm t

return A scale-free network
`//Initial graph with` m_0 `nodes with degrees at least 1`
$G(V, E) = G(V_0, E_0)$;
for 1 to t **do**
    $V = V \cup \{v_i\}$; `// add new node` v_i
    while $d_i \neq m$ **do**
        Connect v_i to a random node $v_j \in V$, $i \neq j$ (i.e., $E = E \cup \{e(v_i, v_j)\}$) with probability $P(v_j) = \frac{d_j}{\sum_k d_k}$.
    end while
end for
Return $G(V, E)$

Properties of the Preferential Attachment Model

Properties

- **Degree Distribution:** $$P(d) = \frac{2m^2}{d^3}$$
- **Clustering Coefficient:** $$C = \frac{m_0 - 1}{8} \frac{(\ln t)^2}{t}$$
- **Average Path Length:** $$l \sim \frac{\ln |V|}{\ln(\ln |V|)}$$

Modeling with the Preferential Attachment Model

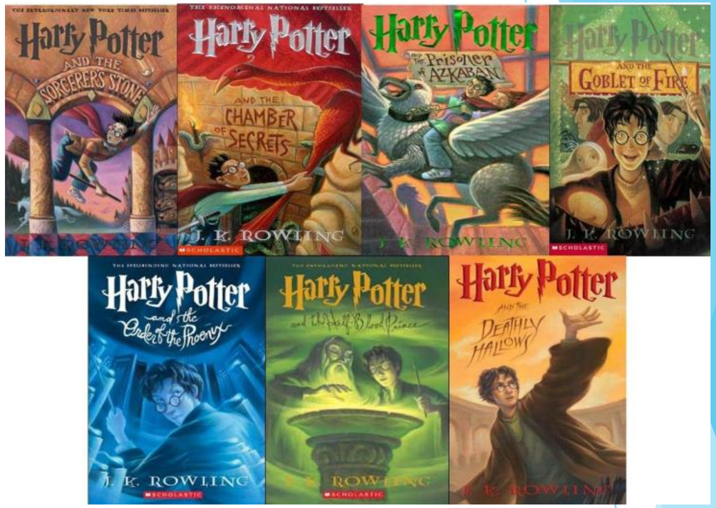
- Similar to random graphs, we can simulate real-world networks by generating a preferential attachment model by setting the expected degree m
  - See Algorithm 4.2 in the book

Real-World Networks and Simulated Graphs

	Original Network				Simulated Graph	
Network	*Size*	*Average Degree*	*Average Path Length*	C	*Average Path Length*	C
Film Actors	225,226	61	3.65	0.79	4.90	≈ 0.005
Medline Coauthorship	1,520,251	18.1	4.6	0.56	5.36	≈ 0.0002
E.Coli	282	7.35	2.9	0.32	2.37	0.03
C.Elegans	282	14	2.65	0.28	1.99	0.05

Unpredictability of the Rich-Get-Richer Effects

- The initial stages of one’s rise to popularity are fragile
- Once a user is well established, the rich-get-richer dynamics of popularity is likely to push the user even higher
- **But** getting the rich-get-richer process started in the first place is full of potential accidents and near-misses



If we could roll time back to 1997, and then run history forward again, would the Harry Potter books again sell hundreds of millions of copies?

See more: Salganik, Matthew J., Peter Sheridan Dodds, and Duncan J. Watts. "Experimental study of inequality and unpredictability in an artificial cultural market." *science* 311.5762 (2006): 854-856.