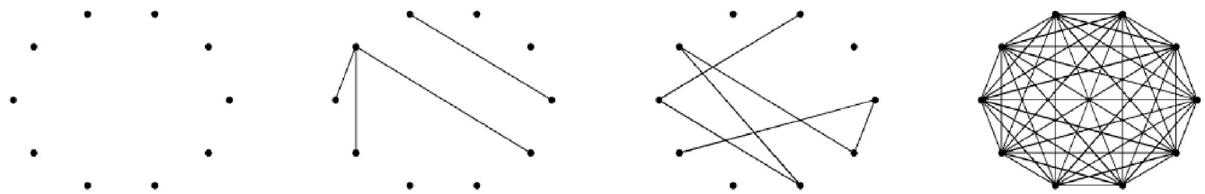
Evolution of Random Graphs

- **Create your own 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 Formation of a 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

Properties of Random Graphs

Degree Distribution

- When computing degree distribution, we estimate the probability of observing $P(d_v = d)$ for node v

- For a random graph generated by $G(n, p)$, this probability is

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

- This is a binomial degree distribution. In the limit, this will become the Poisson degree distribution

Expected Local Clustering Coefficient

The <u>expected</u> local clustering coefficient for node v of a random graph generated by $G(n, p)$ is p

Proof

$$C(v) = \frac{\text{number of connected pairs of } v\text{'s neighbors}}{\text{number of pairs of } v\text{'s neighbors}}$$

– v can have different degrees depending on the random procedure so the expected value is

$$\mathbf{E}(C(v)) = \sum_{d=0}^{n-1} \mathbf{E}(C(v)|d_v = d)\ P(d_v = d)$$

⬇

Expected Local Clustering Coefficient, Cont.

$$\mathbf{E}(C(v)) = \sum_{d=0}^{n-1} \boxed{\mathbf{E}(C(v)|d_v = d)}\, P(d_v = d)$$

$$\mathbf{E}(C(v)|d_v = d) = \frac{\text{number of connected pairs of } v\text{'s } d \text{ neighbors}}{\text{number of pairs of } v\text{'s neighbors}}$$

$$= \frac{p\binom{d}{2}}{\binom{d}{2}} = p$$

Sums up to 1

$$\mathbf{E}(C(v)) = p \boxed{\sum_{d=0}^{d=n-1} P(d_v = d)} = p$$

Global Clustering Coefficient

The global clustering coefficient of a random graph generated by $G(n, p)$ is p

Proof

- The global clustering coefficient defines the probability of two neighbors of the same node being connected.
- In a random graph, for any two nodes, this probability is the same
  - Equal to the generation probability p that determines the probability of two nodes getting connected

- Observation: Local and Global clustering coefficients are the same

Modeling with Random Graphs

- Compute the average degree c in the real-world graph
- Compute p using $c/(n-1) = p$
- Generate the random graph using p

- How representative is the generated graph?
  - **[Degree Distribution]** Random graphs do not have a power-law degree distribution
  - **[Average Path Length]** Random graphs perform well in modeling the average path lengths
  - **[Clustering Coefficient]** Random graphs drastically underestimate the clustering coefficient

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

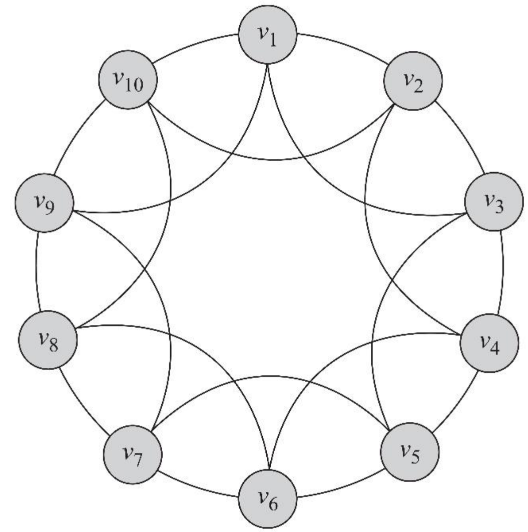
- Small-world model
  - or the **Watts-Strogatz (WS)** model
  - A special type of random graph
  - Exhibits small-world properties:
    - Short average path length
    - High clustering coefficient

- 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={$v_1, v_2, v_3, \ldots, v_n$}, an edge exists between node i and j if and only if

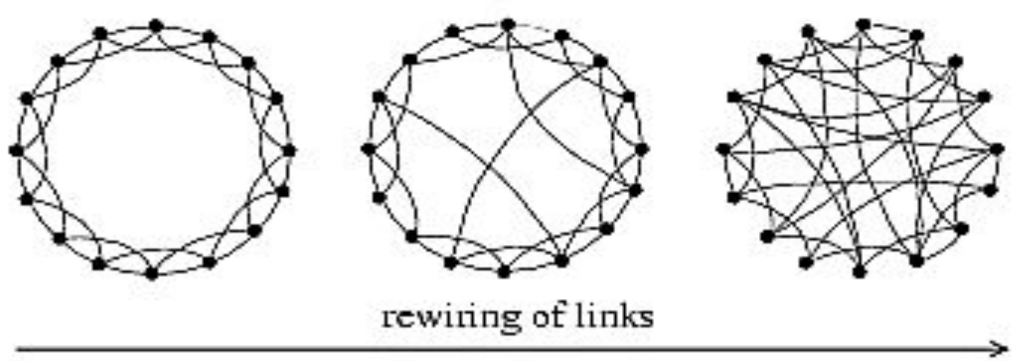


$$0 \leq \min(n - |i - j|, |i - j|) \leq c/2$$

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 $0 \leq \beta \leq 1$ 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

1: **return** A small-world graph $G(V, E)$
2: G = A regular ring lattice with $|V|$ nodes and degree c
3: **for** node v_i (starting from v_1), and all edges $e(v_i, v_j)$, $i < j$ **do**
4: $\quad v_k$ = Select a node from V uniformly at random.
5: $\quad$ **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**
6: $\quad\quad$ rewire $e(v_i, v_j)$ with probability β: $E = E - \{e(v_i, v_j)\}$, $E = E \cup \{e(v_i, v_k)\}$;
7: $\quad$ **end if**
8: **end for**
9: Return $G(V, E)$

As in many network generating algorithms, they

- 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?

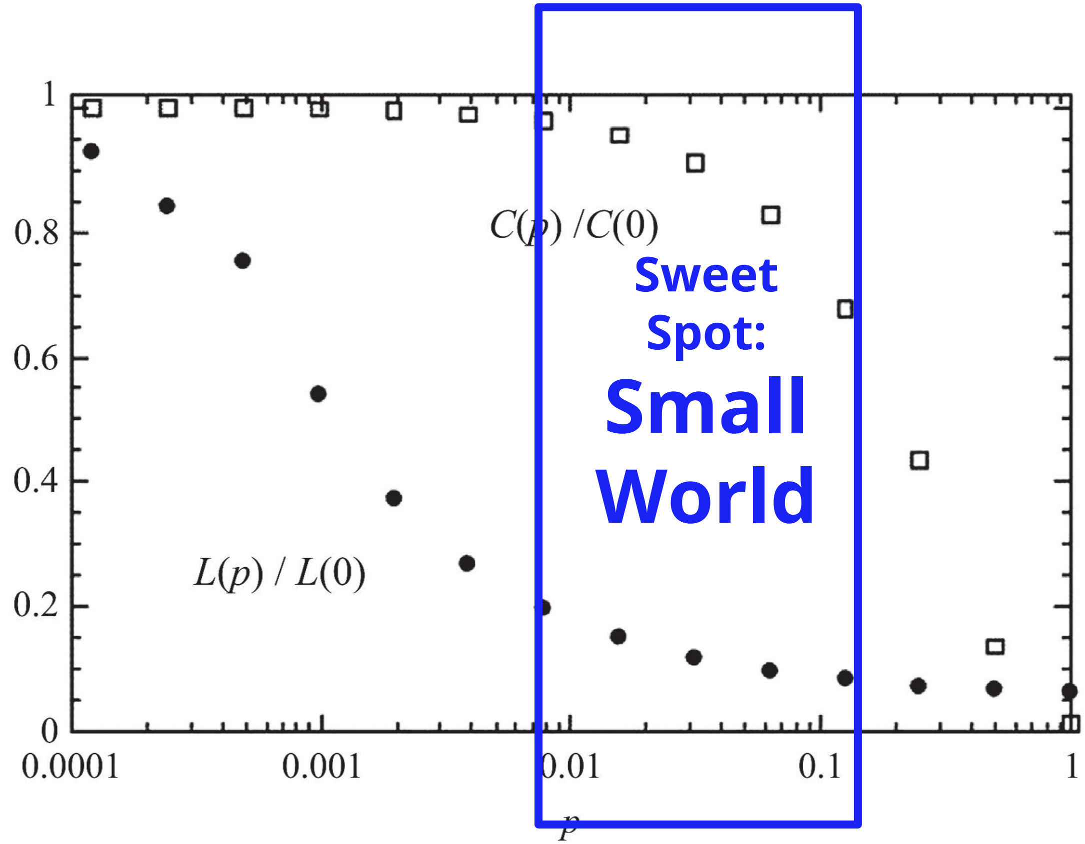
Numerical simulation:

- 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

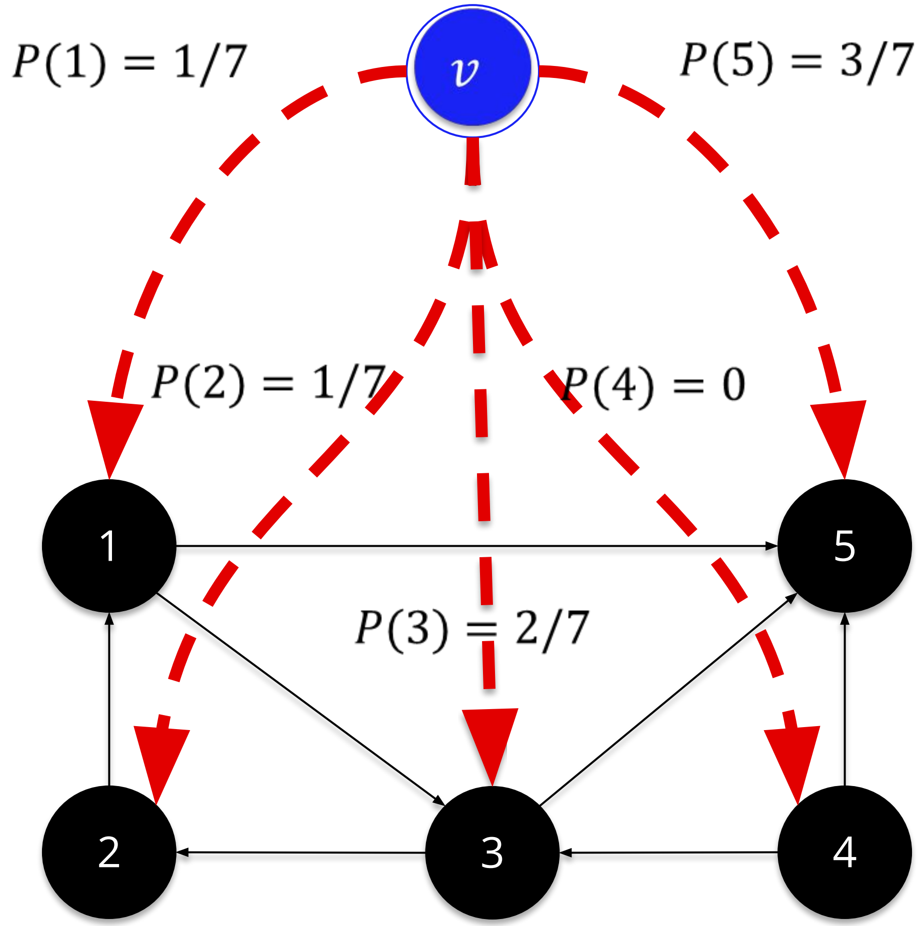
Barabási, Albert-László, and Réka Albert. "Emergence of scaling in random networks." *science* 286.5439 (1999): 509-512.

Preferential Attachment: Example

- Node v arrives

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

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



Modeling Real-World Networks w *Preferential Attachment*

- Similar to random graphs, we can simulate real-world networks by generating a preferential attachment model by setting the expected degree *m*

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:** $l \sim \frac{\ln |V|}{\ln(\ln |V|)}$

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

Real-World Networks and Simulated Graphs

Network	Original Network				Simulated Random Graph	
	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

	al Network				Simulated Graph	
		ge 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

Network	Original Network				Simulated Graph	
	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

Random
Small World
Preferential
Attachment?