Probabilistic Methods for Complex Graphs

Linyuan Lu

lu@math.sc.edu

University of South Carolina

Center for Combinatorics at Nankai University
June 3, 2008

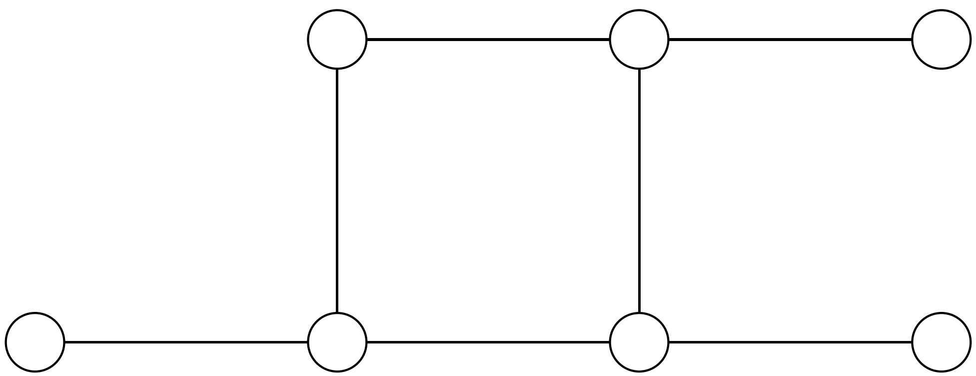
Outline

- Complex networks
- Evolution models
- Static models
- Coupling methods

Preliminary

A graph consists of two sets V and E.

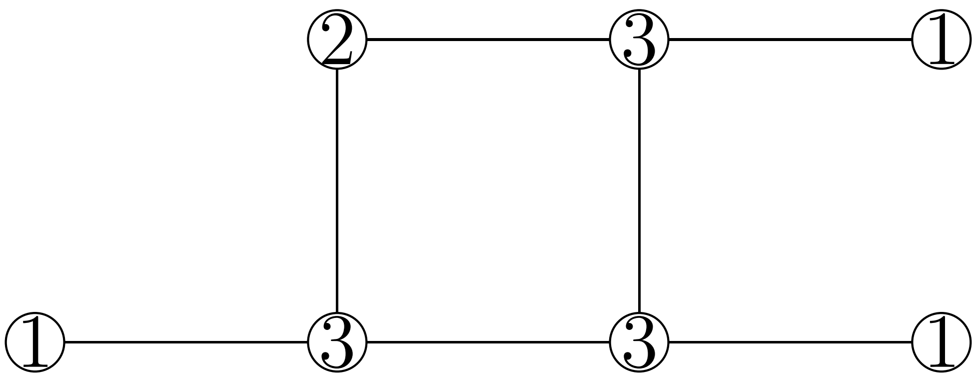
- V is the set of vertices (or nodes).
- E is the set of edges, where each edge is a pair of vertices.



Preliminary

A graph consists of two sets V and E.

- V is the set of vertices (or nodes).
- E is the set of edges, where each edge is a pair of vertices.



The degree of a vertex is the number of edges, which are incident to that vertex.

Examples of complex graphs

WWW Graphs

Call Graphs

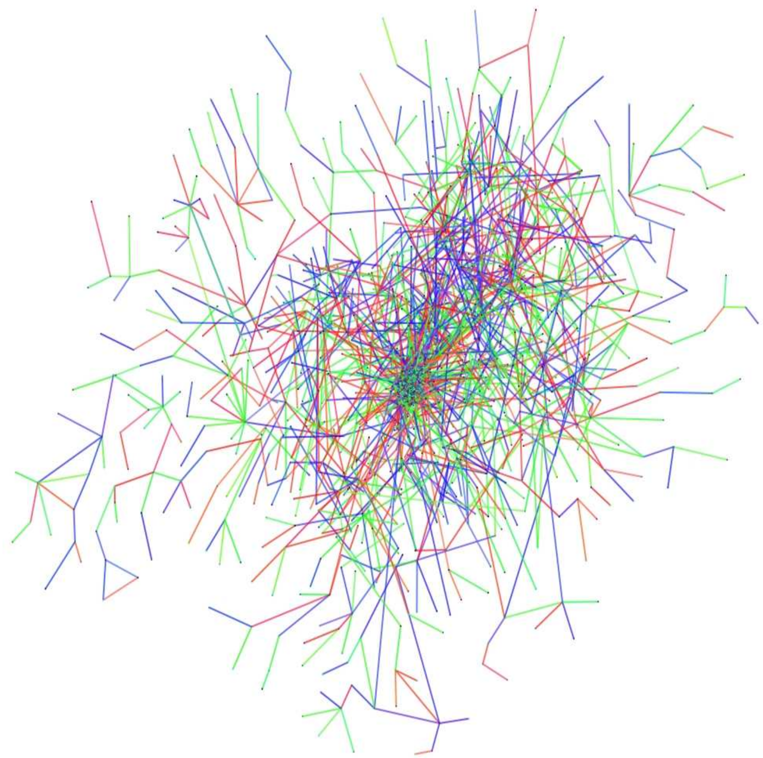
Collaboration Graphs

Gene Regulatory Graphs

Graph of U.S. Power Grid

Costars Graph of Actors

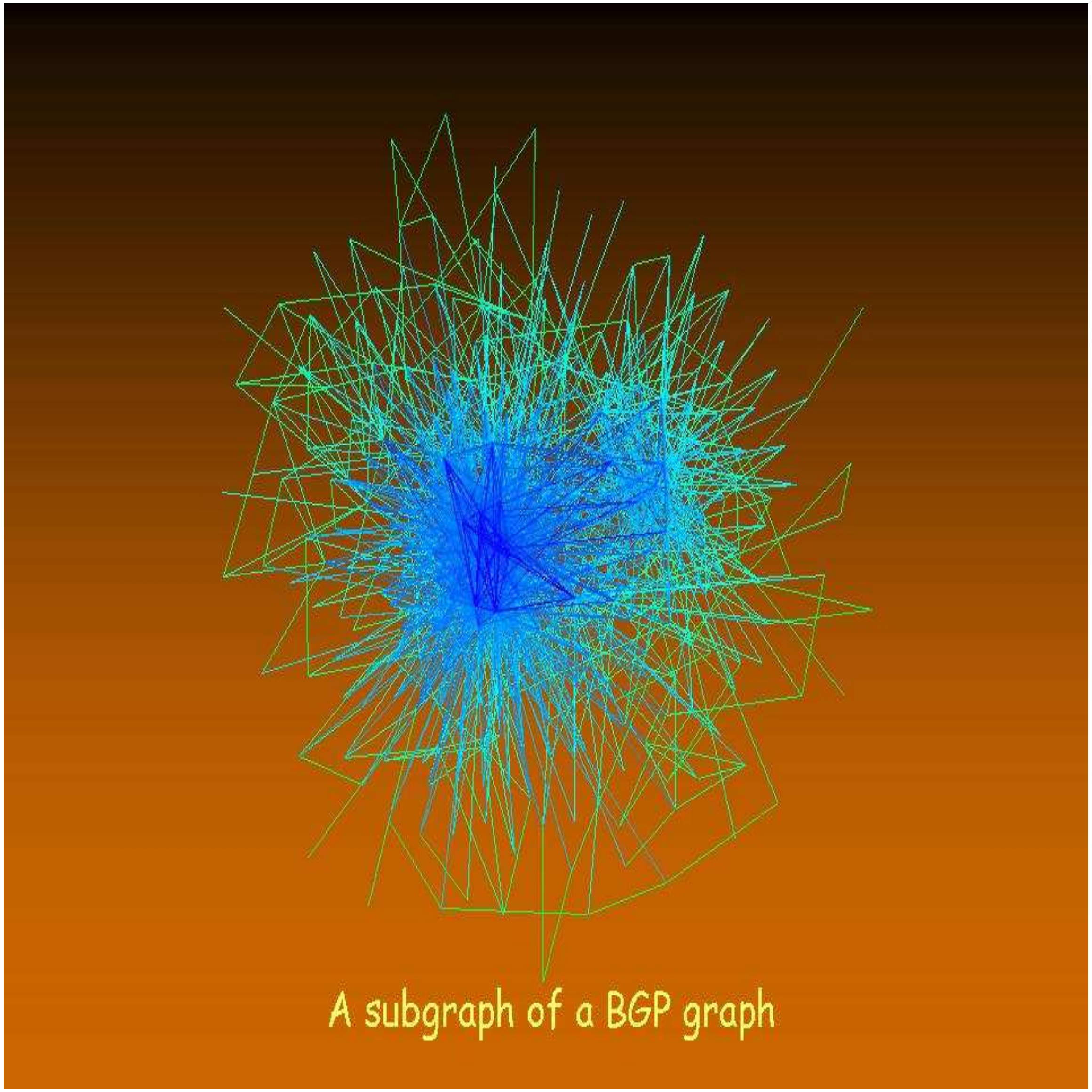
⋮



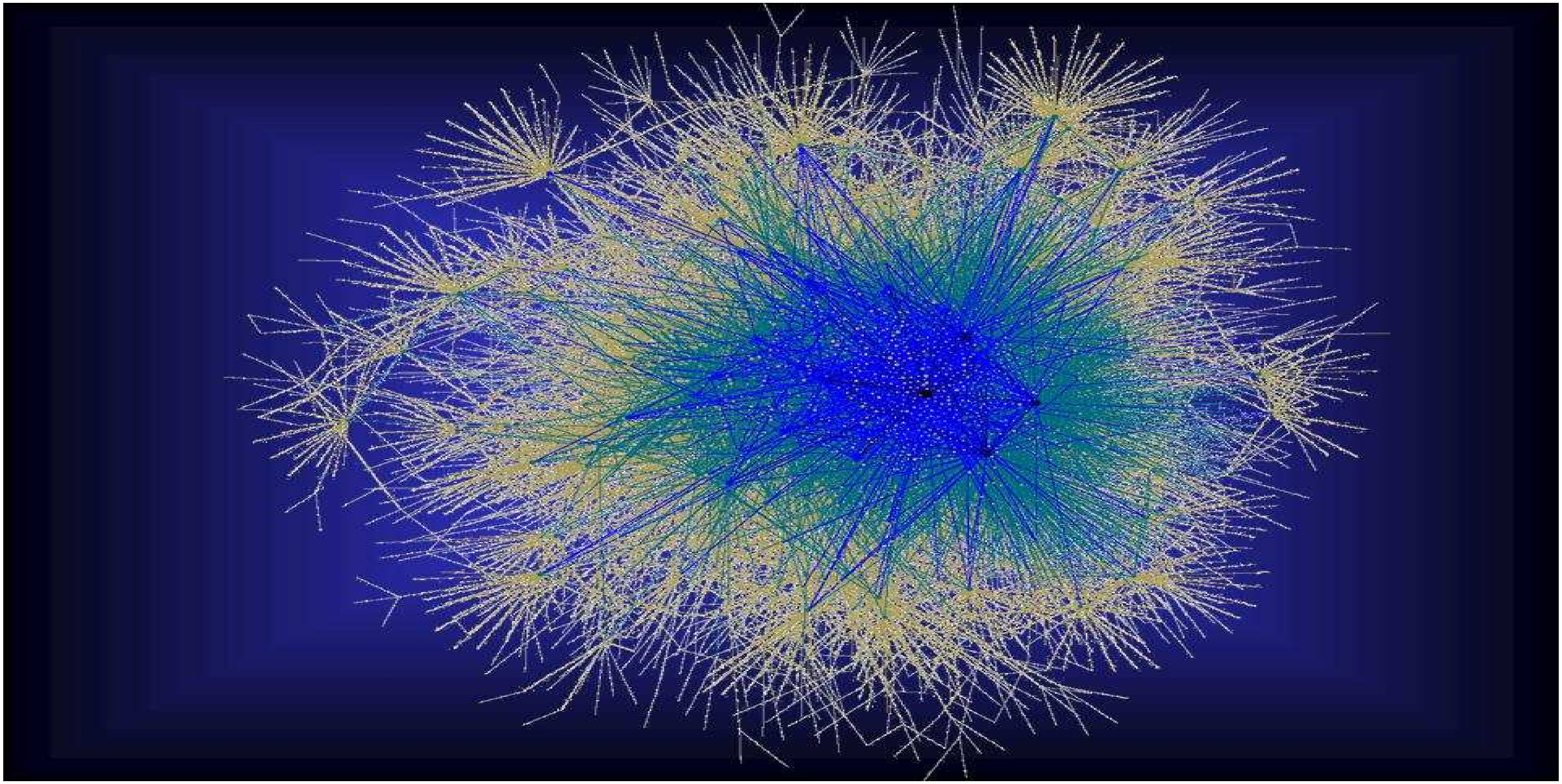
BGP Graph

Vertex: AS (autonomous system)

Edges: AS pairs in BGP routing table.



Large BGP subgraph



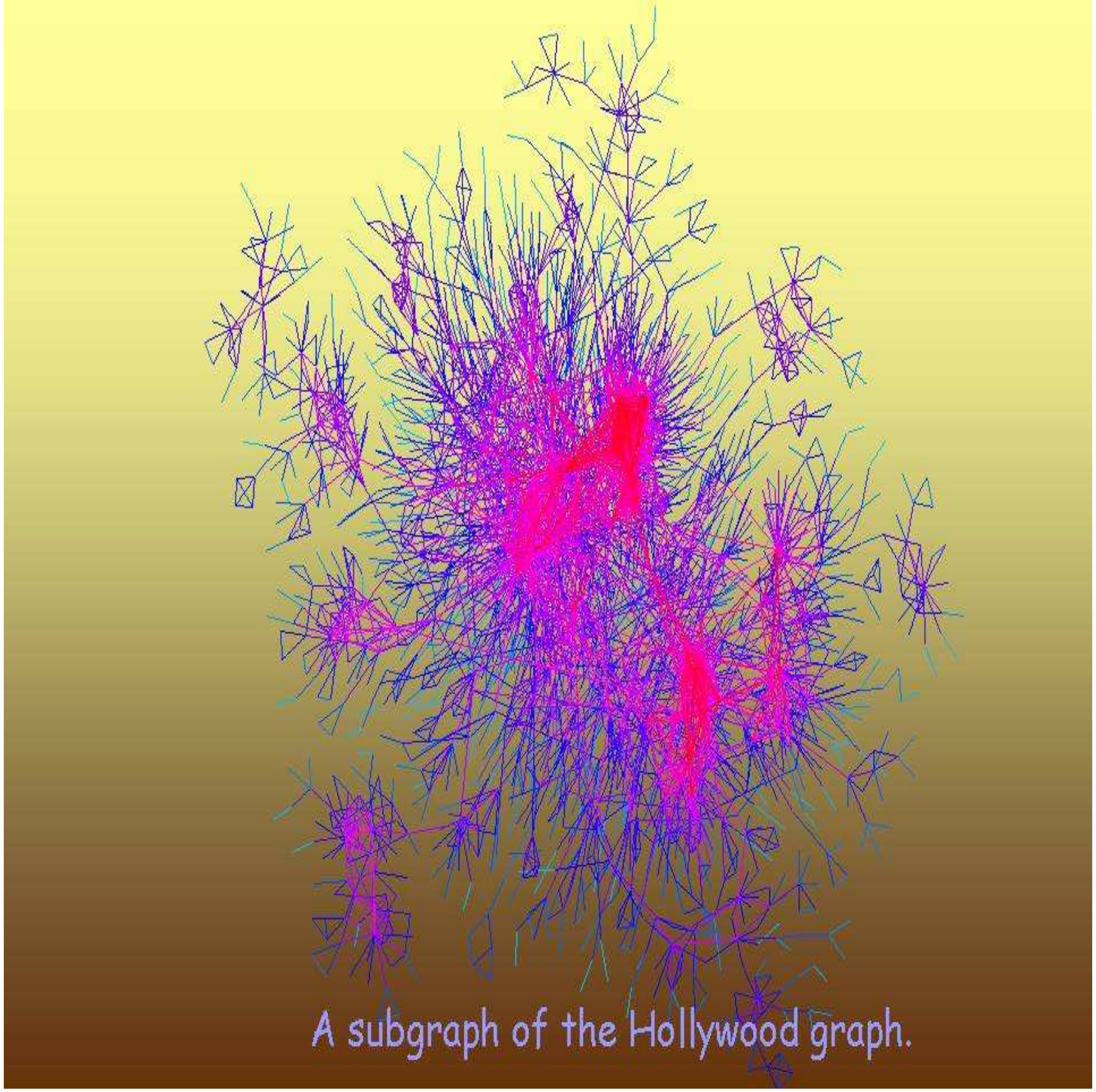
Only a portion of 6400 vertices and 13000 edges is drawn.

Hollywood Graph

Vertex: actors and actress

Edges: co-playing in the same movie

Only 10,000 out of 225,000 are shown.



Folklore of Erdős numbers

- Erdős has Erdős number 0.
- Erdős' coauthor has Erdős number 1.
- Erdős' coauthor's coauthor has Erdős number 2.

$\vdots$

Folklore of Erdős numbers

- Erdős has Erdős number 0.
- Erdős' coauthor has Erdős number 1.
- Erdős' coauthor's coauthor has Erdős number 2.

$\vdots$

My Erdős number is 2.

Folklore of Erdős numbers

- Erdős has Erdős number 0.
- Erdős' coauthor has Erdős number 1.
- Erdős' coauthor's coauthor has Erdős number 2.

⋮

My Erdős number is 2.

Erdős number is the graph distance to Erdős in the Collaboration graph.

Folklore of Erdős numbers

- Erdős has Erdős number 0.
- Erdős' coauthor has Erdős number 1.
- Erdős' coauthor's coauthor has Erdős number 2.

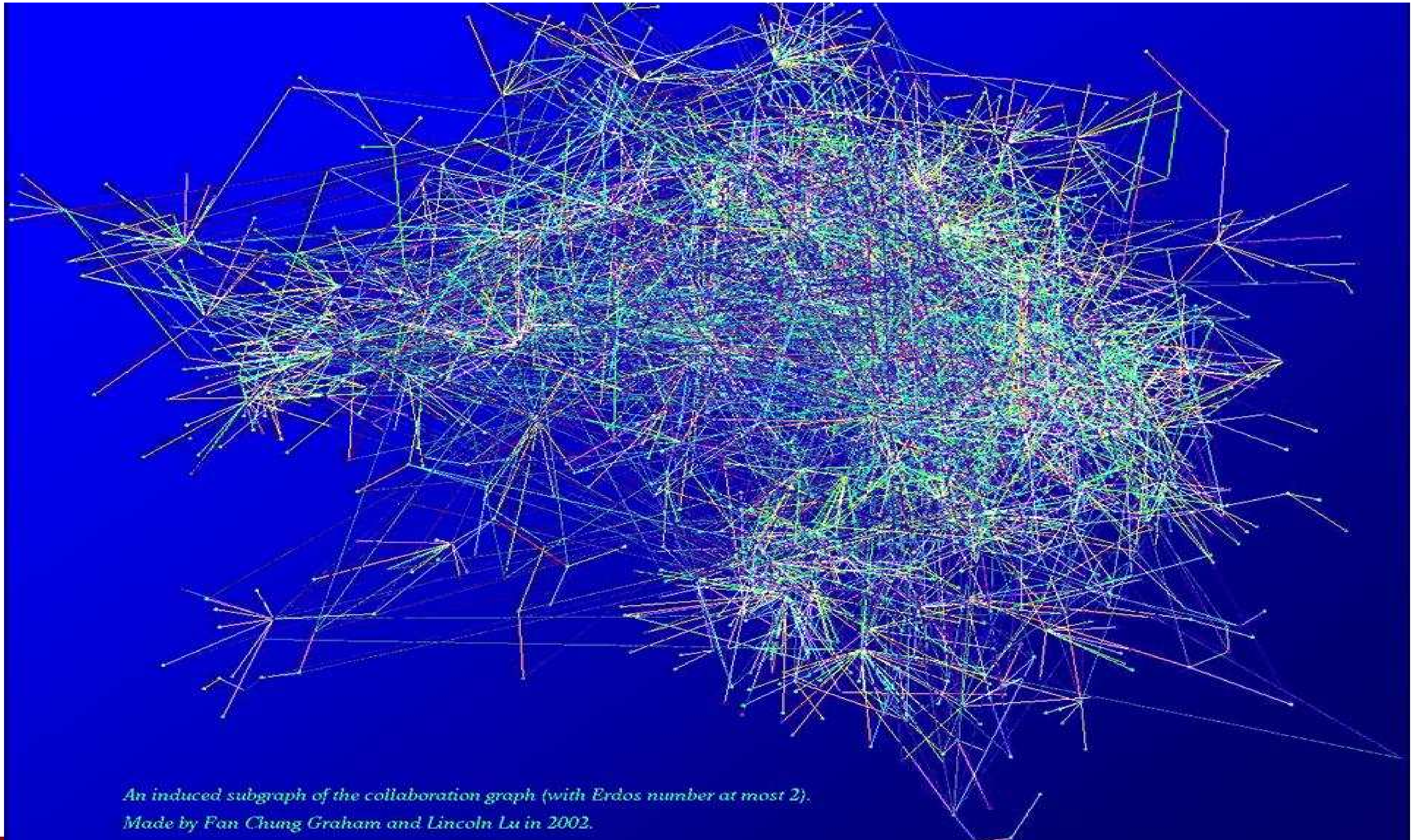
$\vdots$

My Erdős number is 2.

Erdős number is the graph distance to Erdős in the Collaboration graph.

My Chen number is 3.

Collaboration Graph



Characteristics

- Large

Characteristics

- Large
- Sparse

Characteristics

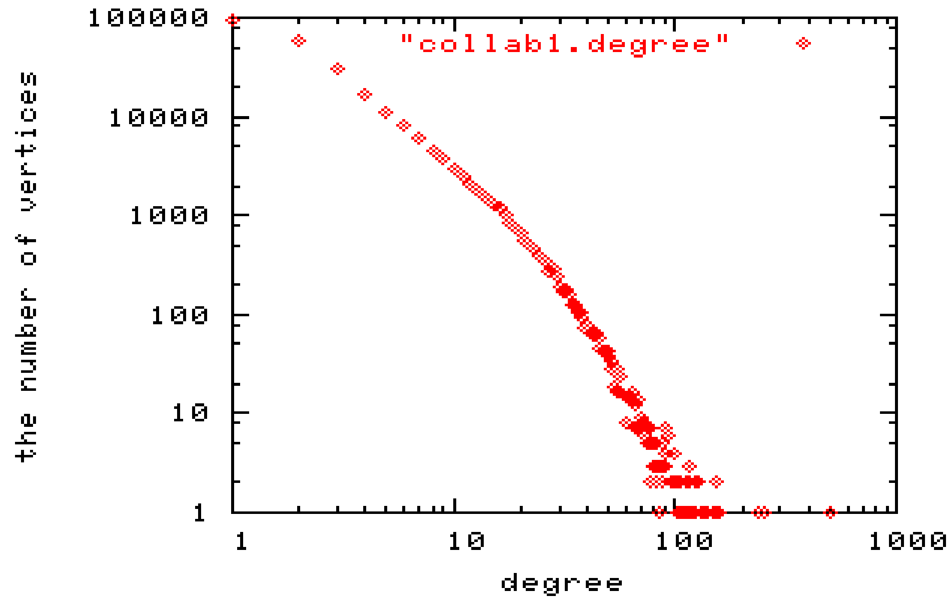
- Large
- Sparse
- Power law degree distribution

Characteristics

- Large
- Sparse
- Power law degree distribution
- Small world phenomenon

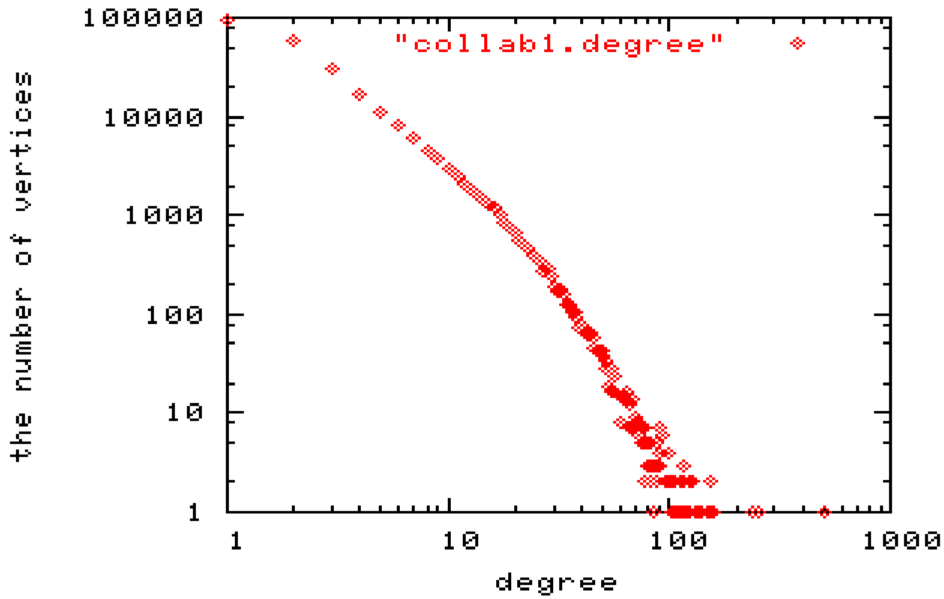
The power law

The number of vertices of degree k is approximately proportional to $k^{-\beta}$ for some positive β.



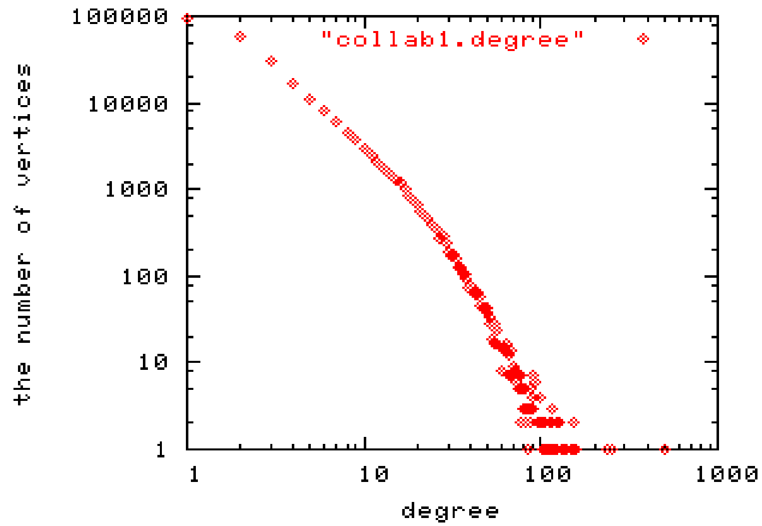
The power law

The number of vertices of degree k is approximately proportional to $k^{-\beta}$ for some positive β.



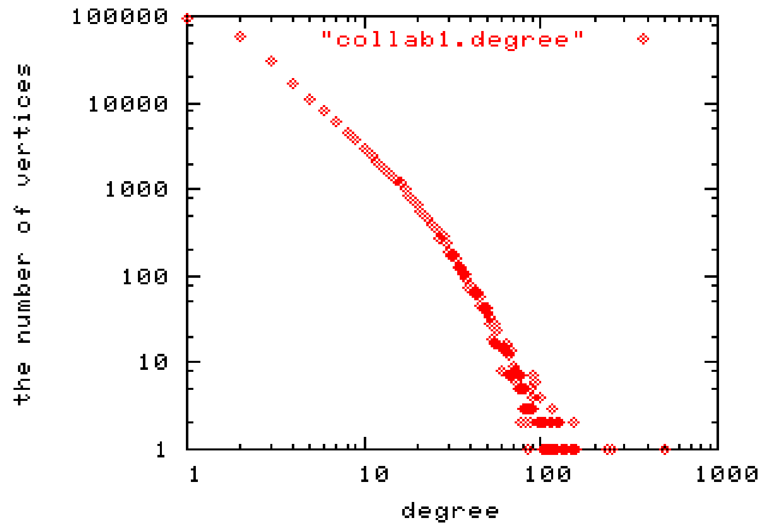
A power law graph is a graph whose degree sequence satisfies the power law.

Power law distribution



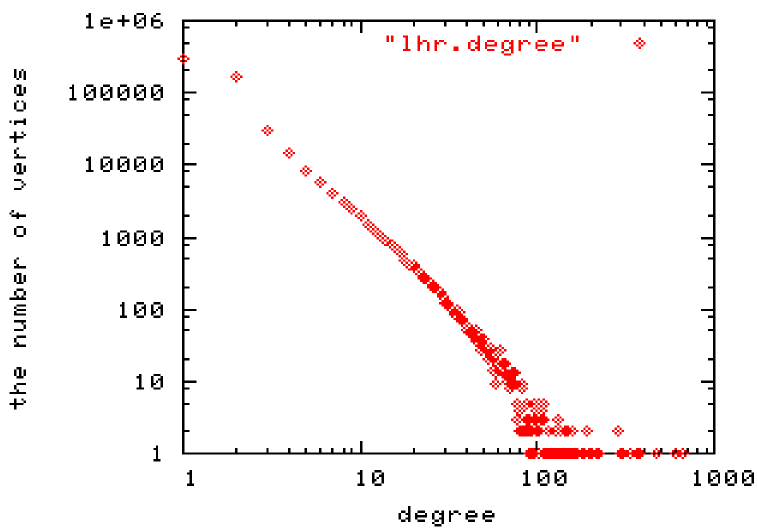
Left: The collaboration graph follows the power law degree distribution with exponent $\beta \approx 3.0$

Power law distribution

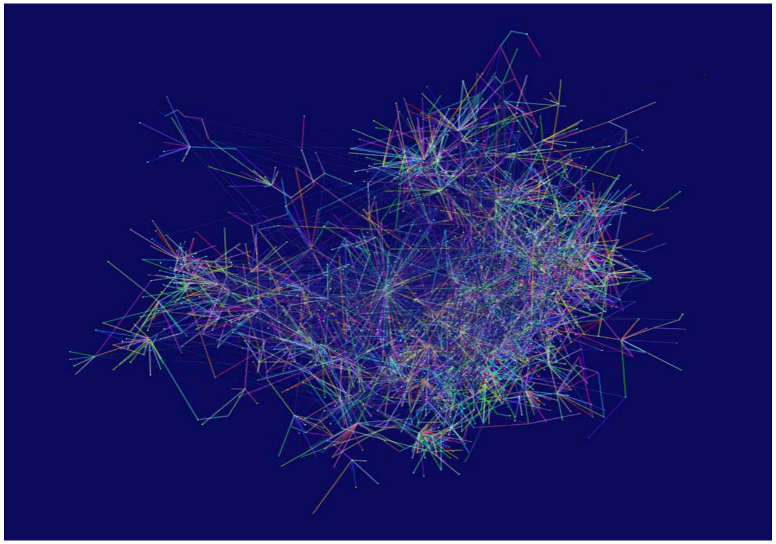


Left: The collaboration graph follows the power law degree distribution with exponent $\beta \approx 3.0$

Right: An IP graph follows the power law degree distribution with exponent $\beta \approx 2.4$

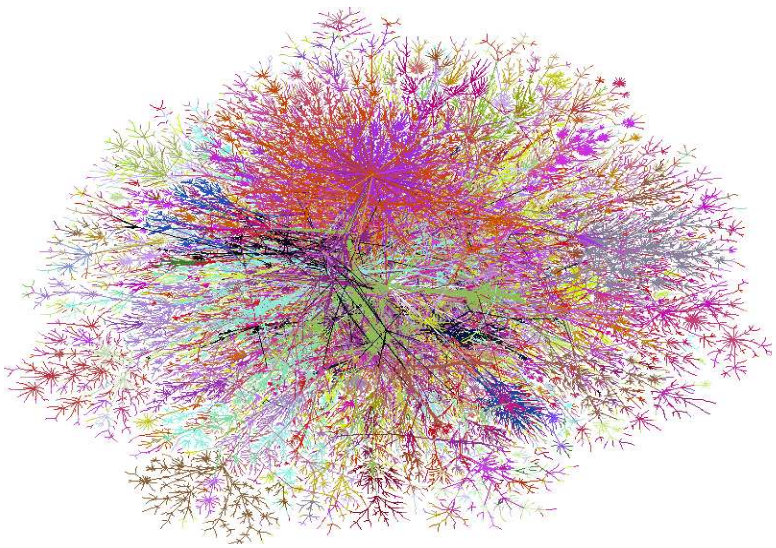


Power law graphs

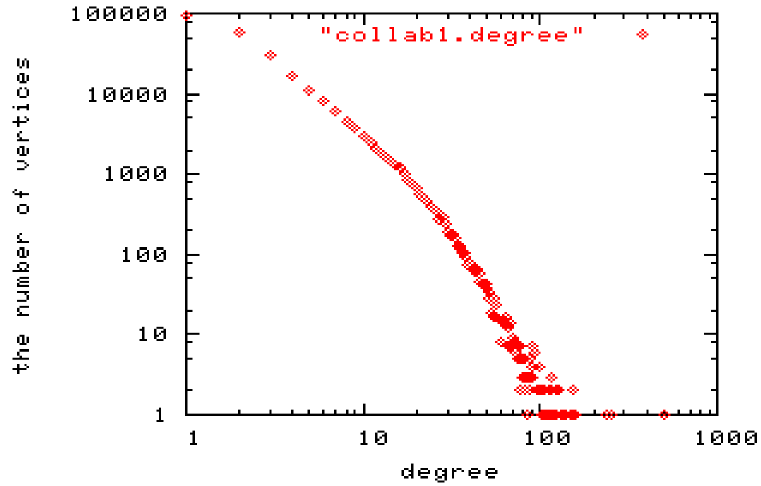
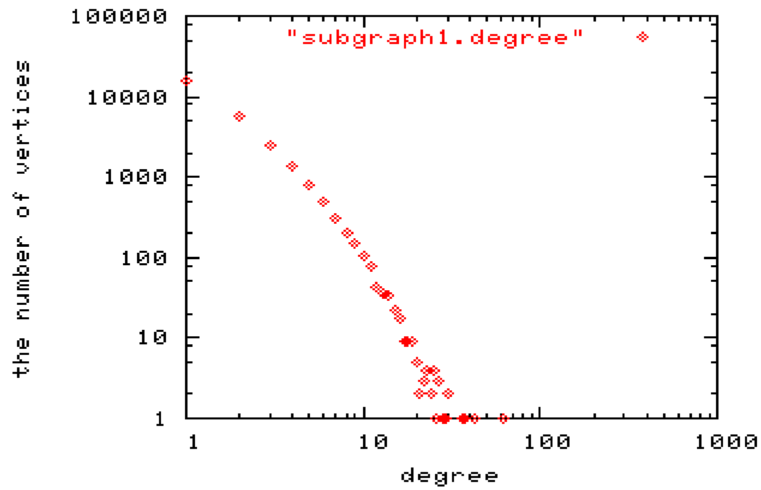


Left: Part of the collaboration graph (authors with Erdős number 2)

Right: An IP graph (by Bill Cheswick)



Robustness of Power Law

size	degree distribution
25,3339	
52,186	

Basic questions

- How to model power law graphs?

Basic questions

- How to model power law graphs?

- What graph properties can be derived from the model?

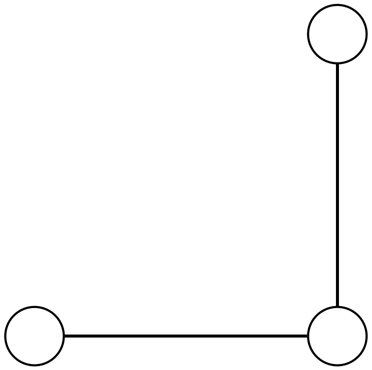
Random graphs

A random graph is a set of graphs together with a probability distribution on that set.

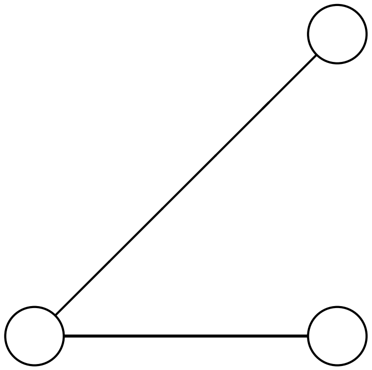
Random graphs

A random graph is a set of graphs together with a probability distribution on that set.

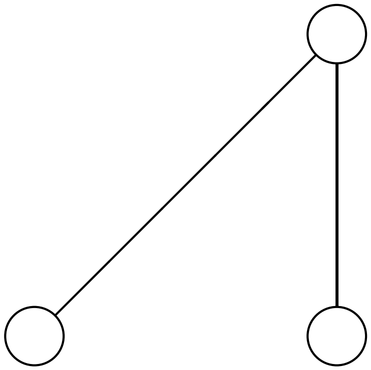
Example: A random graph on 3 vertices and 2 edges with the uniform distribution on it.



Probability $\frac{1}{3}$



Probability $\frac{1}{3}$

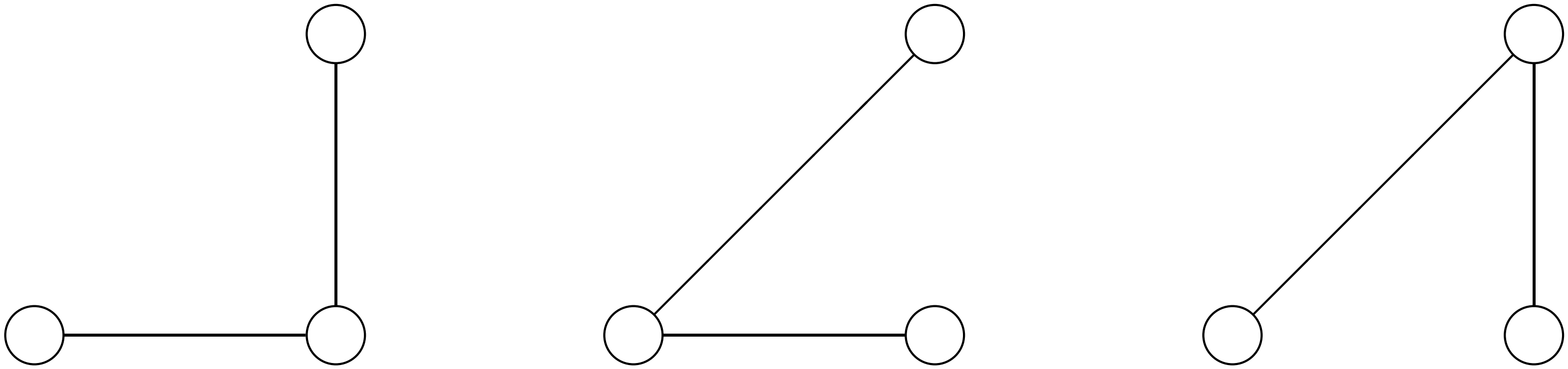


Probability $\frac{1}{3}$

Random graphs

A random graph is a set of graphs together with a probability distribution on that set.

Example: A random graph on 3 vertices and 2 edges with the uniform distribution on it.



Probability $\frac{1}{3}$ Probability $\frac{1}{3}$ Probability $\frac{1}{3}$

A random graph G *almost surely* satisfies a property P, if

$$Pr(G \text{ satisfies } P) = 1 - o_n(1).$$

Evolution models

Graph evolution

$$\cdots \subset G_{t-1} \subset G_t \subset G_{t+1} \subset \cdots$$

- Preferential attachment models
  - Barabási, Albert, etc.
  - Kleinberg, Kumar, Raghavan, etc.
  - Aiello, Chung, Lu

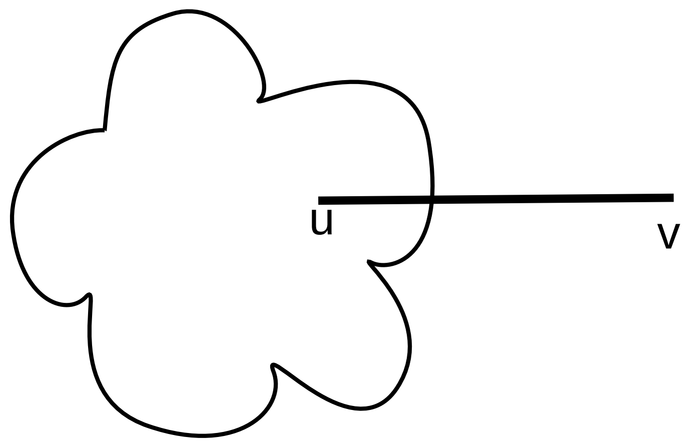
Evolution models

Graph evolution

$$\cdots \subset G_{t-1} \subset G_t \subset G_{t+1} \subset \cdots$$

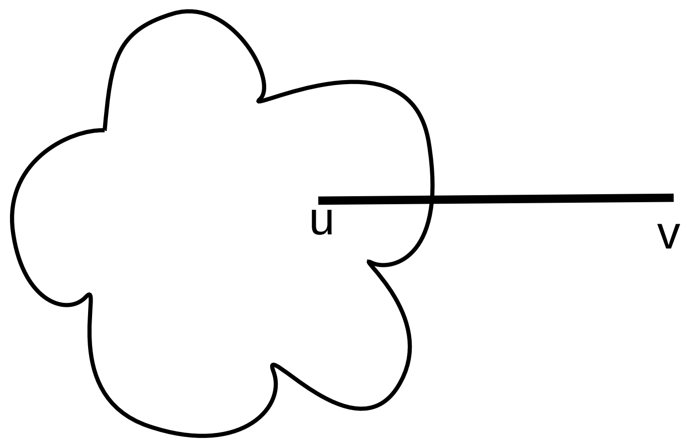
- Preferential attachment models
  - Barabási, Albert, etc.
  - Kleinberg, Kumar, Raghavan, etc.
  - Aiello, Chung, Lu
- Partial duplication models (Chung, Dewey, Galas, Lu)

Preferential attachment



At time t, add a new vertex v to the existed network and attach v to a vertex u, which is selected with probability proportional to its current degree.

Preferential attachment



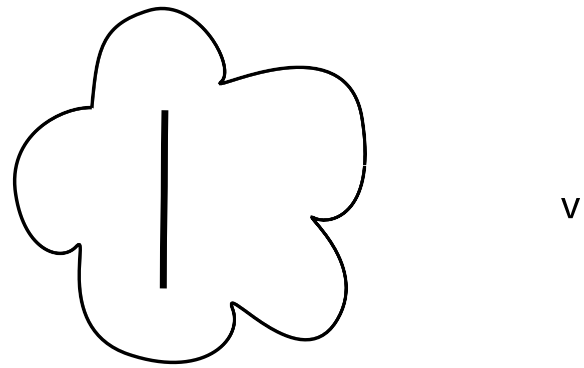
At time t, add a new vertex v to the existed network and attach v to a vertex u, which is selected with probability proportional to its current degree.

Barabási, Albert (1999) The preferential attachment model almost surely generates a power low graph with exponent $\beta = 3$.

A general model

At time t,

- add expected $\mu^{e,e}$ random random edges to existed network.
- add expected $\mu^{n,e}$ random edges between new vertex and existed network.
- add expected $\mu^{n,n}$ loops to the new vertex.

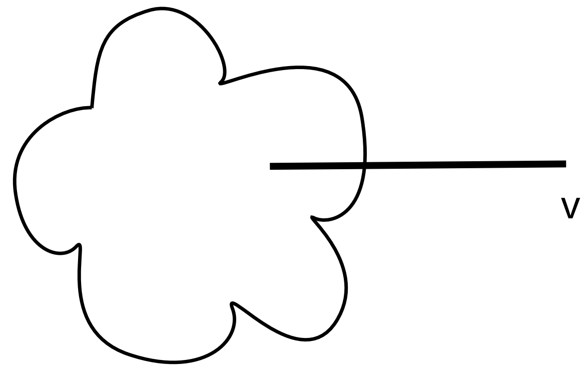


Aiello, Chung, Lu (2001): This general preferential attachment model almost surely generates a power low graph with exponent $\beta = 2 + \frac{2\mu^{n,n}+\mu^{n,e}}{\mu^{n,e}+2\mu^{e,e}}$

A general model

At time t,

- add expected $\mu^{e,e}$ random random edges to existed network.
- add expected $\mu^{n,e}$ random edges between new vertex and existed network.
- add expected $\mu^{n,n}$ loops to the new vertex.

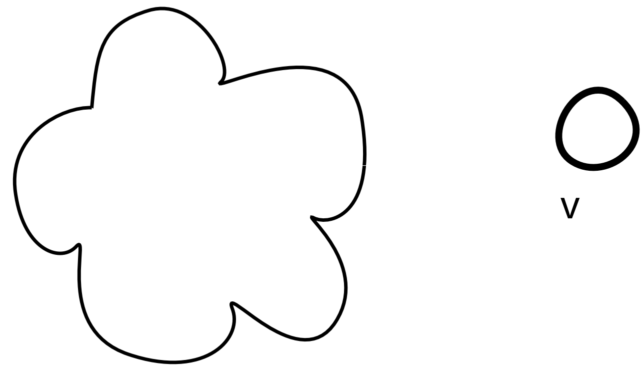


Aiello, Chung, Lu (2001): This general preferential attachment model almost surely generates a power low graph with exponent $\beta = 2 + \frac{2\mu^{n,n}+\mu^{n,e}}{\mu^{n,e}+2\mu^{e,e}}$

A general model

At time t,

- add expected $\mu^{e,e}$ random random edges to existed network.
- add expected $\mu^{n,e}$ random edges between new vertex and existed network.
- add expected $\mu^{n,n}$ loops to the new vertex.

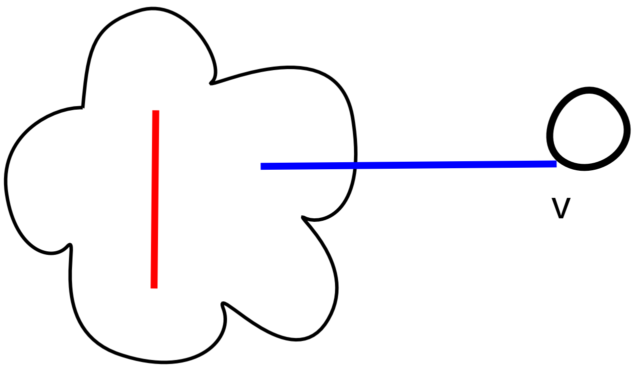


Aiello, Chung, Lu (2001): This general preferential attachment model almost surely generates a power low graph with exponent $\beta = 2 + \frac{2\mu^{n,n}+\mu^{n,e}}{\mu^{n,e}+2\mu^{e,e}}$

A general model

At time t,

- add expected $\mu^{e,e}$ random random edges to existed network.
- add expected $\mu^{n,e}$ random edges between new vertex and existed network.
- add expected $\mu^{n,n}$ loops to the new vertex.

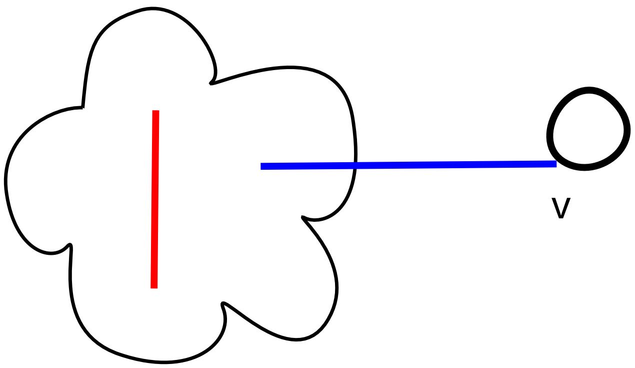


Aiello, Chung, Lu (2001): This general preferential attachment model almost surely generates a power low graph with exponent $\beta = 2 + \frac{2\mu^{n,n}+\mu^{n,e}}{\mu^{n,e}+2\mu^{e,e}}$

A general model

At time t,

- add expected $\mu^{e,e}$ random random edges to existed network.
- add expected $\mu^{n,e}$ random edges between new vertex and existed network.
- add expected $\mu^{n,n}$ loops to the new vertex.



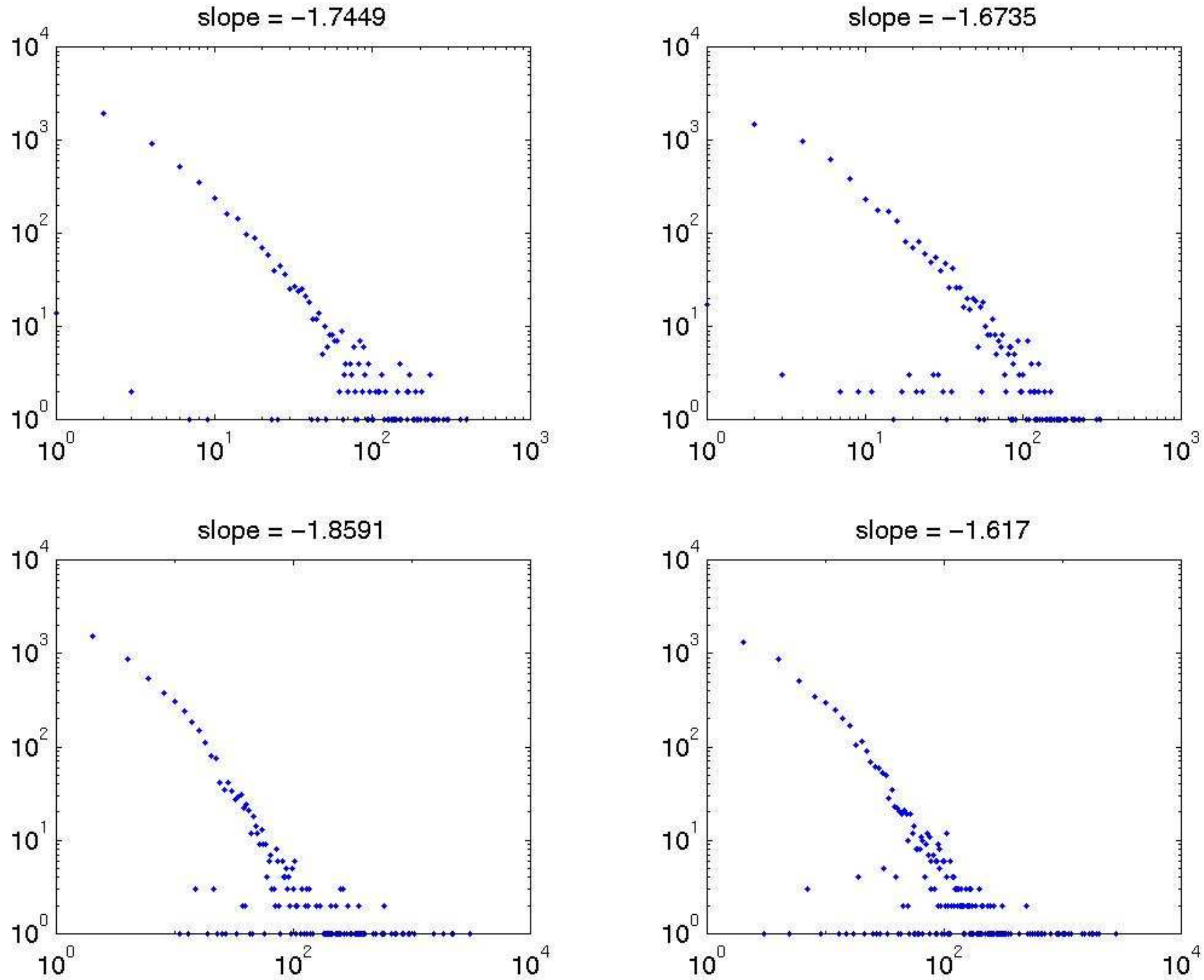
Aiello, Chung, Lu (2001): This general preferential attachment model almost surely generates a power low graph with exponent $\beta = 2 + \frac{2\mu^{n,n}+\mu^{n,e}}{\mu^{n,e}+2\mu^{e,e}}$

Similar results hold for directed graph model.

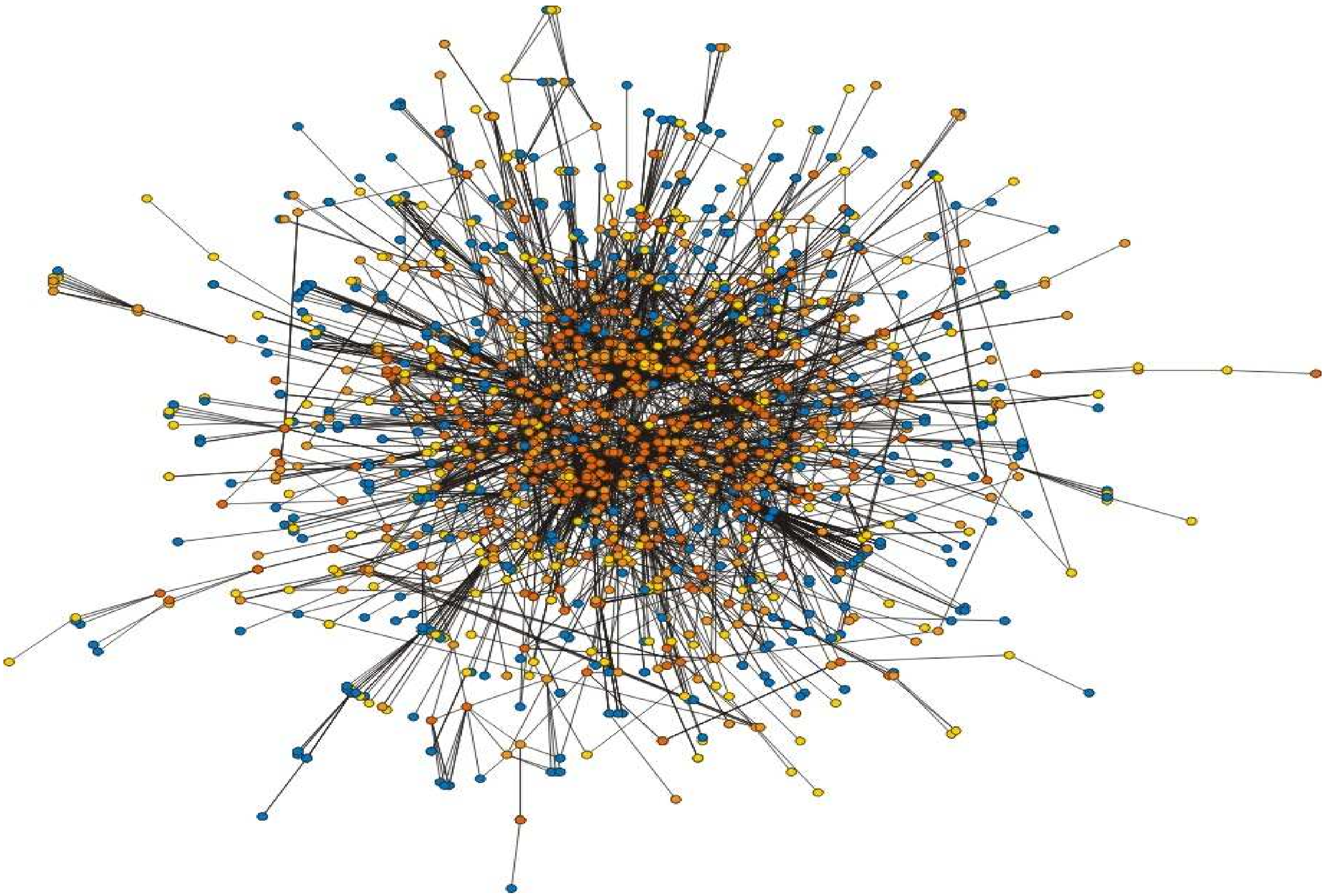
A question

Are there power law graphs with exponent $\beta < 2$?

Ecological networks



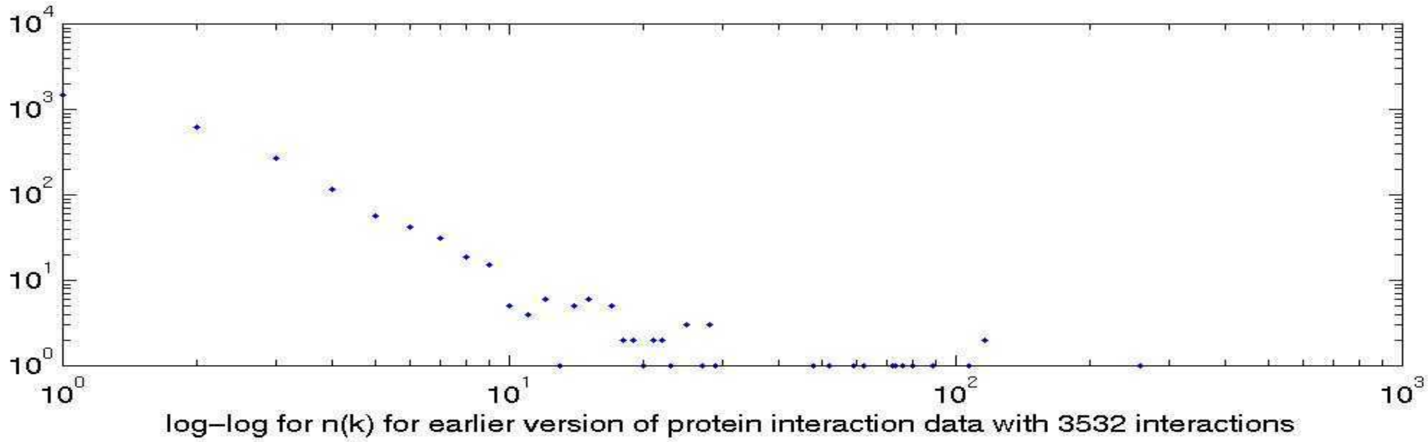
Protein-interaction network

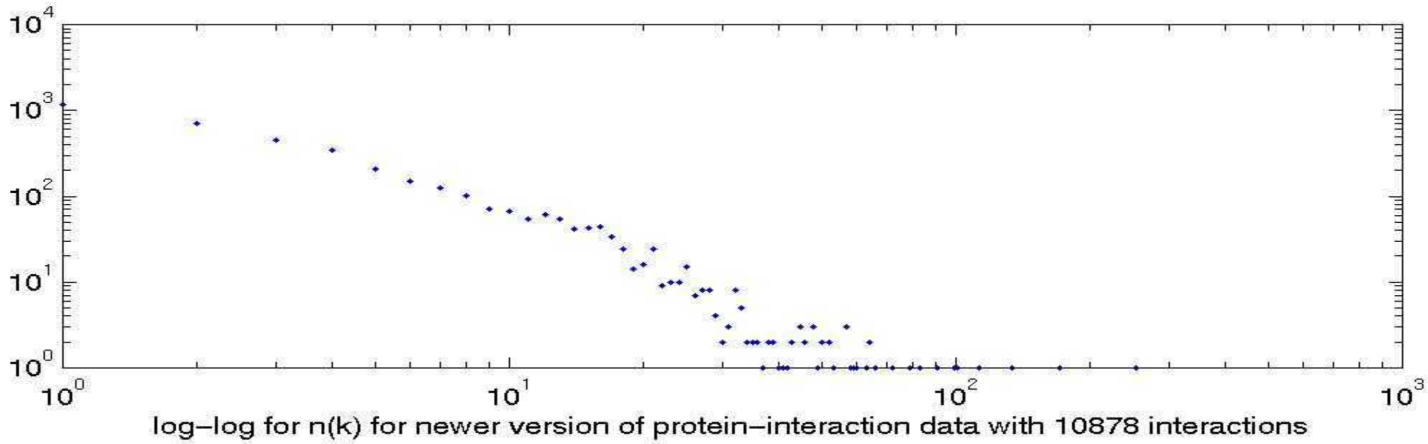


Snel, Bork & Huynen, PNAS 99, 5890 (2002)

Degree distribution

The protein-interaction networks have $\beta \approx 1.7$





A critical threshold $\beta = 2$

Range	$1 < \beta < 2$	$2 < \beta$
Average degree	Unbounded	Bounded
Examples	Biological networks	Non-biological networks
Models	**Partial Duplication model**	Preferential attachment models

Partial-duplication model

Evolution of graphs

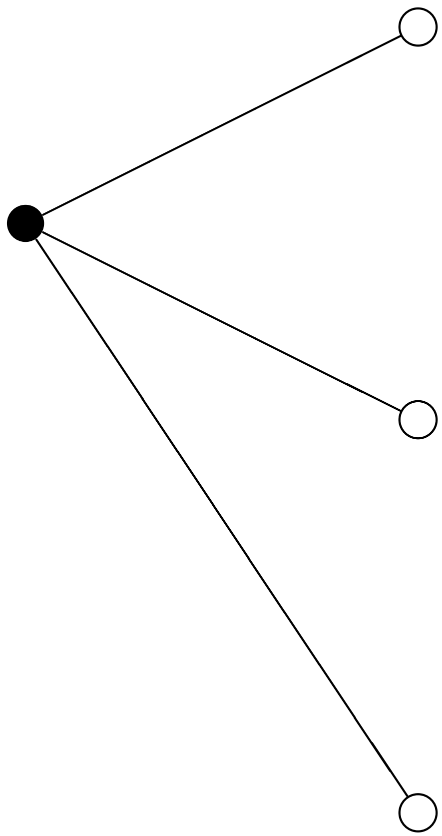
$$\cdots \subset G_{t-1} \subset G_t \subset G_{t+1} \subset \cdots$$

Construct G_{t+1} from G_t,

- Select a random vertex u of G_t uniformly.
- Add a new vertex v.
- For each neighbor w of u, with probability p, add an edge wv independently.

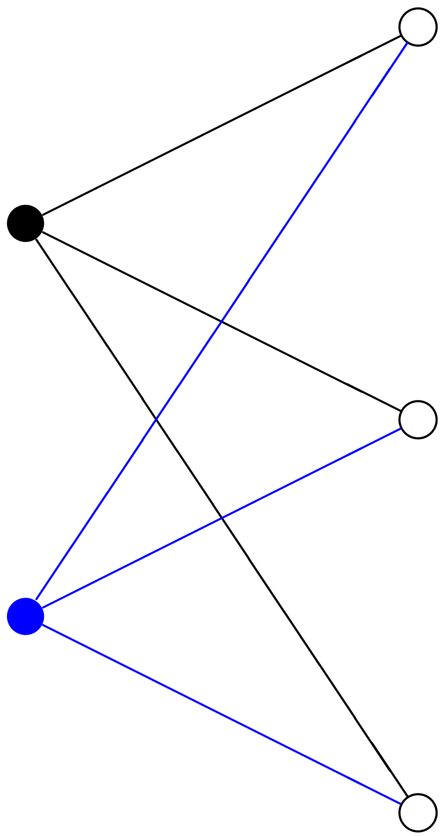
Partial-duplication

Full duplication



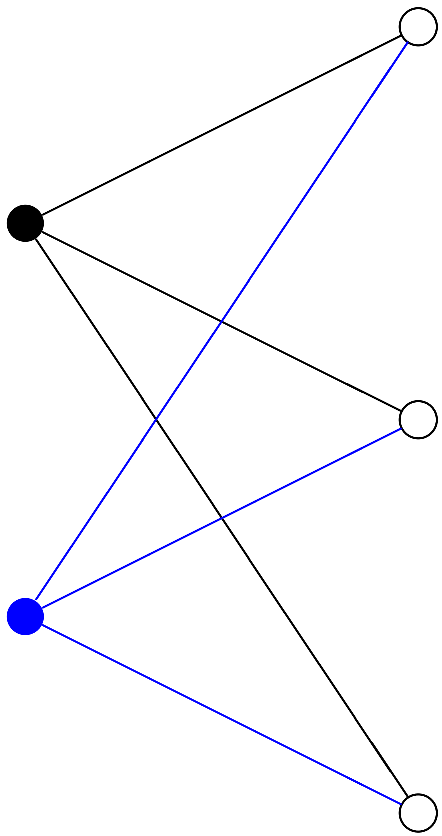
Partial-duplication

Full duplication

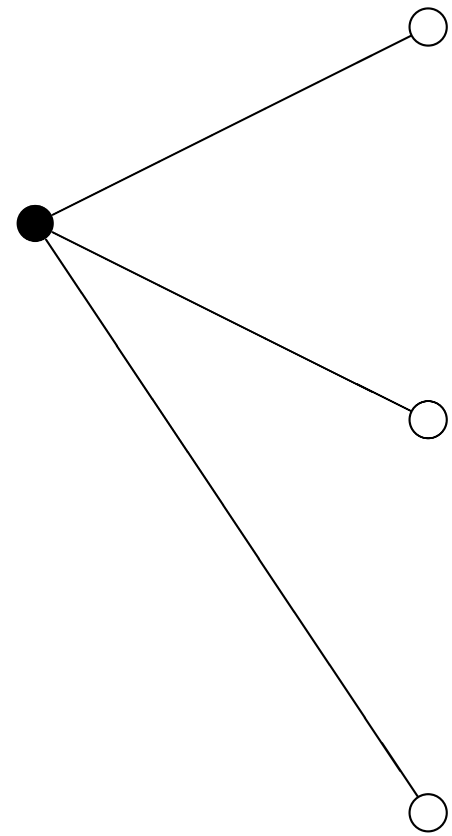


Partial-duplication

Full duplication

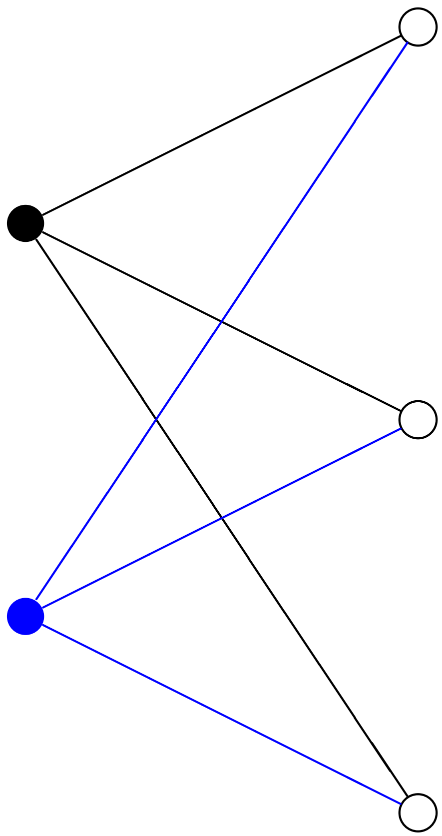


Partial duplication

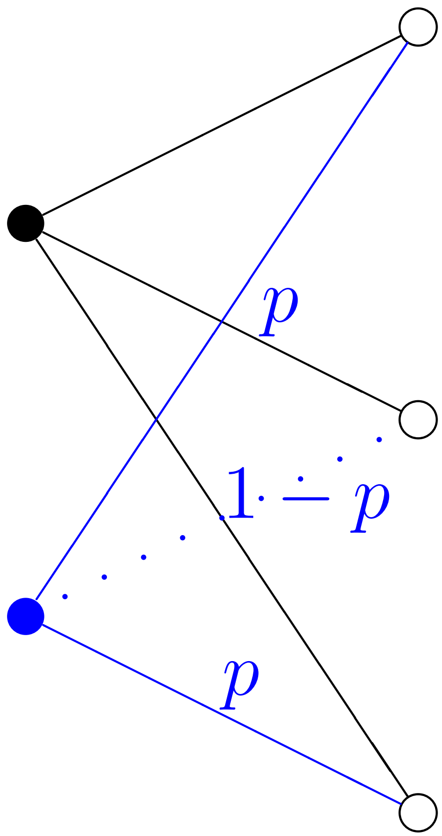


Partial-duplication

Full duplication



Partial duplication

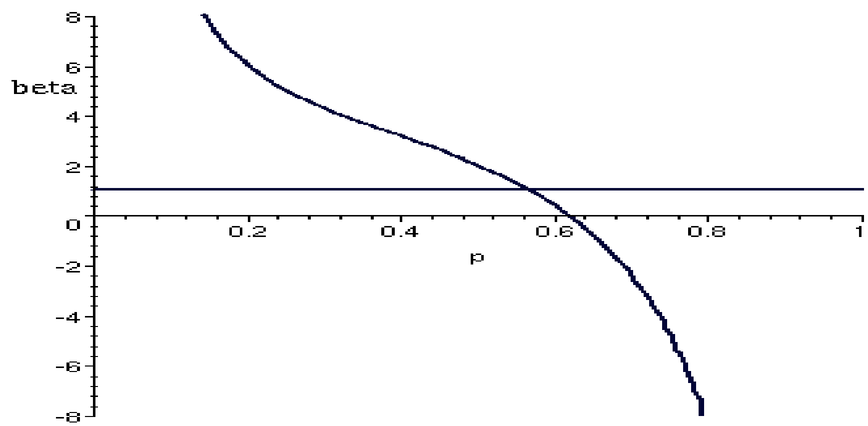


Results

Chung, Dewey, Galas, Lu (2002) Almost surely, the partial duplication model with selection probability p generates power law graphs with the exponent β satisfying

$$p(\beta - 1) = 1 - p^{\beta - 1}.$$

In particular, if $\frac{1}{2} < p < 1$ then $\beta < 2$.



Static models

- Erdős-Rényi model $G(n, p)$
- Random Graphs with given expected degree sequences.
- Configuration model with given degree sequences.

Erdős-Rényi model $G(n, p)$

- n nodes

Erdős-Rényi model $G(n, p)$

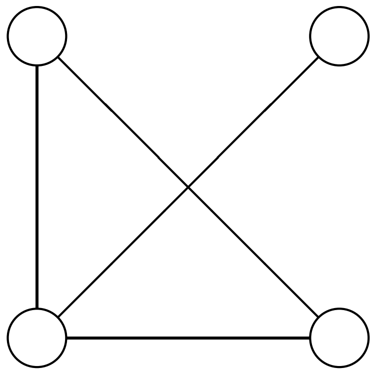
- n nodes
- For each pair of vertices, create an edge independently with probability p.

Erdős-Rényi model $G(n, p)$

- n nodes
- For each pair of vertices, create an edge independently with probability p.
- The graph with e edges has the probability $p^e(1-p)^{\binom{n}{2}-e}$.

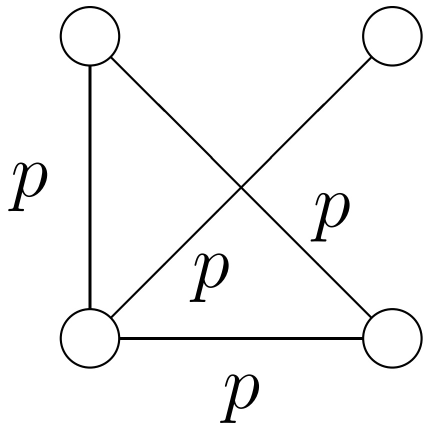
Erdős-Rényi model $G(n,p)$

- n nodes
- For each pair of vertices, create an edge independently with probability p.
- The graph with e edges has the probability $p^e(1-p)^{\binom{n}{2}-e}$.



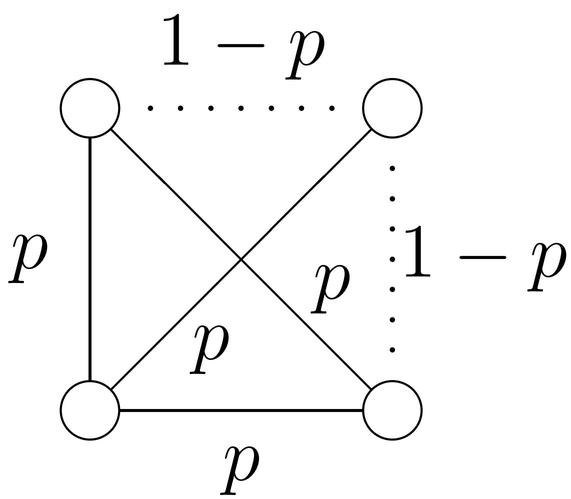
Erdős-Rényi model $G(n, p)$

- n nodes
- For each pair of vertices, create an edge independently with probability p.
- The graph with e edges has the probability $p^e(1-p)^{\binom{n}{2}-e}$.



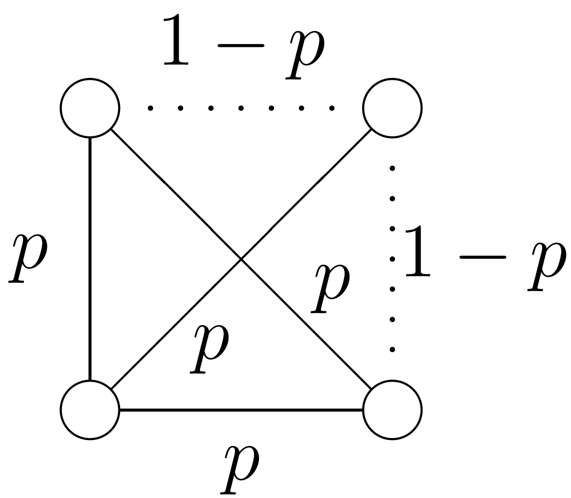
Erdős-Rényi model $G(n, p)$

- n nodes
- For each pair of vertices, create an edge independently with probability p.
- The graph with e edges has the probability $p^e(1-p)^{\binom{n}{2}-e}$.



Erdős-Rényi model $G(n, p)$

- n nodes
- For each pair of vertices, create an edge independently with probability p.
- The graph with e edges has the probability $p^e(1-p)^{\binom{n}{2}-e}$.



The probability of this graph is

$$p^4(1-p)^2.$$

Evolution of $G(n, p)$

Erdős-Rényi 1960s:

- $p \sim c/n$ for $0 < c < 1$: The largest connected component of $G_{n,p}$ is a tree and has about $\frac{1}{\alpha}(\log n - \frac{5}{2} \log \log n)$ vertices, where $\alpha = c - 1 - \log c$.

Evolution of $G(n, p)$

Erdős-Rényi 1960s:

- $p \sim c/n$ for $0 < c < 1$: The largest connected component of $G_{n,p}$ is a tree and has about $\frac{1}{\alpha}(\log n - \frac{5}{2}\log\log n)$ vertices, where $\alpha = c - 1 - \log c$.
- $p \sim 1/n + \mu/n$, the double jump.

Evolution of $G(n,p)$

Erdős-Rényi 1960s:

- $p \sim c/n$ for $0 < c < 1$: The largest connected component of $G_{n,p}$ is a tree and has about $\frac{1}{\alpha}(\log n - \frac{5}{2}\log\log n)$ vertices, where $\alpha = c - 1 - \log c$.
- $p \sim 1/n + \mu/n$, the double jump.
- $p \sim c/n$ for $c > 1$: Except for one "giant" component, all the other components are relatively small. The giant component has approximately $f(c)n$ vertices, where

$$f(c) = 1 - \frac{1}{c}\sum_{k=1}^{\infty} \frac{k^{k-1}}{k!}(ce^{-c})^k.$$

Model $G(w_1, w_2, \ldots, w_n)$

Random graph model with given expected degree sequence

- n nodes with weights $w_1, w_2, \ldots, w_n$.

Model $G(w_1, w_2, \ldots, w_n)$

Random graph model with given expected degree sequence

- n nodes with weights $w_1, w_2, \ldots, w_n$.
- For each pair (i, j), create an edge independently with probability $p_{ij} = w_i w_j \rho$, where $\rho = \frac{1}{\sum_{i=1}^{n} w_i}$.

Model $G(w_1, w_2, \ldots, w_n)$

Random graph model with given expected degree sequence

- n nodes with weights $w_1, w_2, \ldots, w_n$.
- For each pair (i, j), create an edge independently with probability $p_{ij} = w_i w_j \rho$, where $\rho = \frac{1}{\sum_{i=1}^n w_i}$.
- The graph H has probability

$$\prod_{ij \in E(H)} p_{ij} \prod_{ij \notin E(H)} (1 - p_{ij}).$$

Model $G(w_1, w_2, \ldots, w_n)$

Random graph model with given expected degree sequence

- n nodes with weights $w_1, w_2, \ldots, w_n$.
- For each pair (i, j), create an edge independently with probability $p_{ij} = w_i w_j \rho$, where $\rho = \frac{1}{\sum_{i=1}^n w_i}$.
- The graph H has probability

$$\prod_{ij \in E(H)} p_{ij} \prod_{ij \notin E(H)} (1 - p_{ij}).$$

- The expected degree of vertex i is w_i.

Model $G(w_1, w_2, \ldots, w_n)$

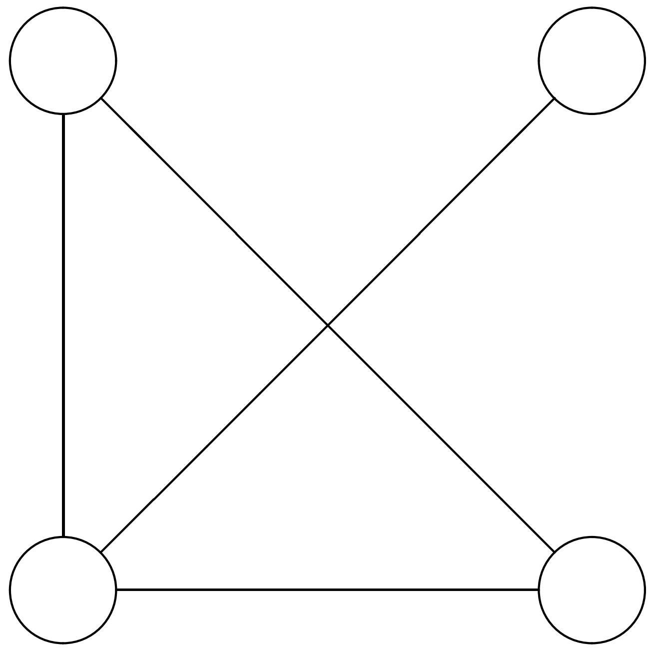
Random graph model with given expected degree sequence

- n nodes with weights $w_1, w_2, \ldots, w_n$.
- For each pair (i, j), create an edge independently with probability $p_{ij} = w_i w_j \rho$, where $\rho = \frac{1}{\sum_{i=1}^n w_i}$.
- The graph H has probability

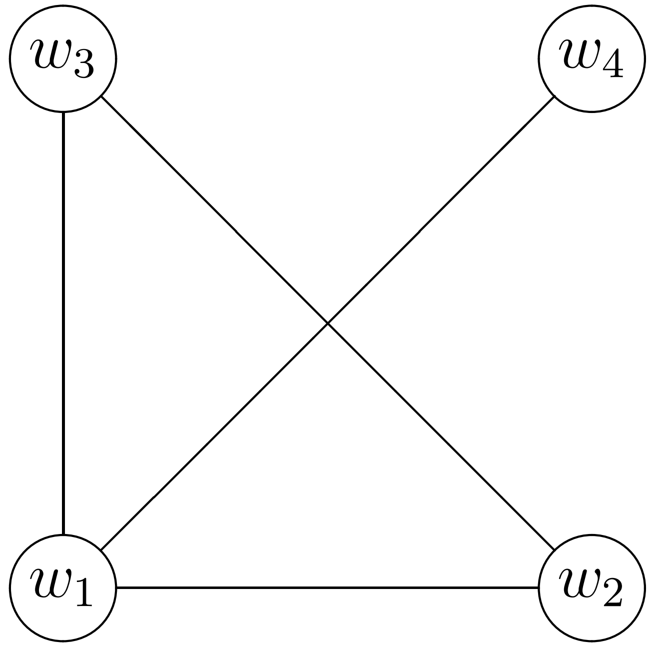
$$\prod_{ij \in E(H)} p_{ij} \prod_{ij \notin E(H)} (1 - p_{ij}).$$

- The expected degree of vertex i is w_i.

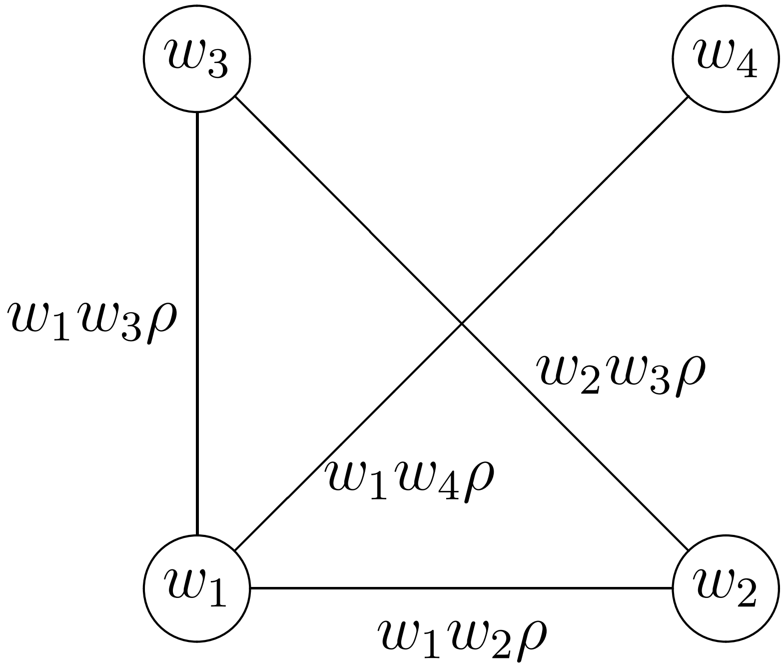
An example: $G(w_1, w_2, w_3, w_4)$



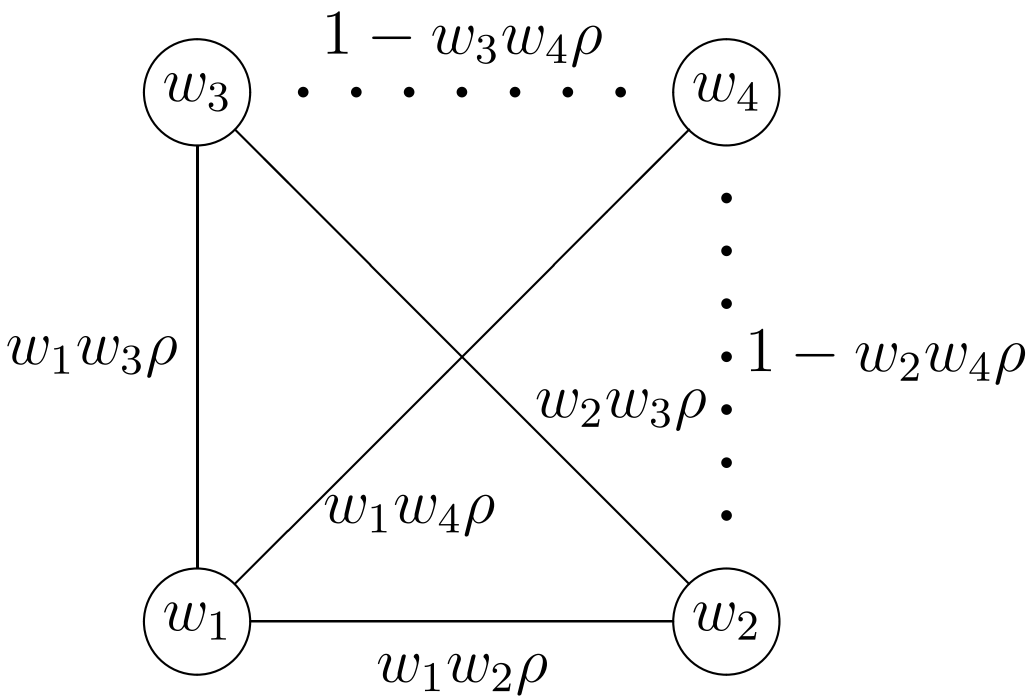
An example: $G(w_1, w_2, w_3, w_4)$



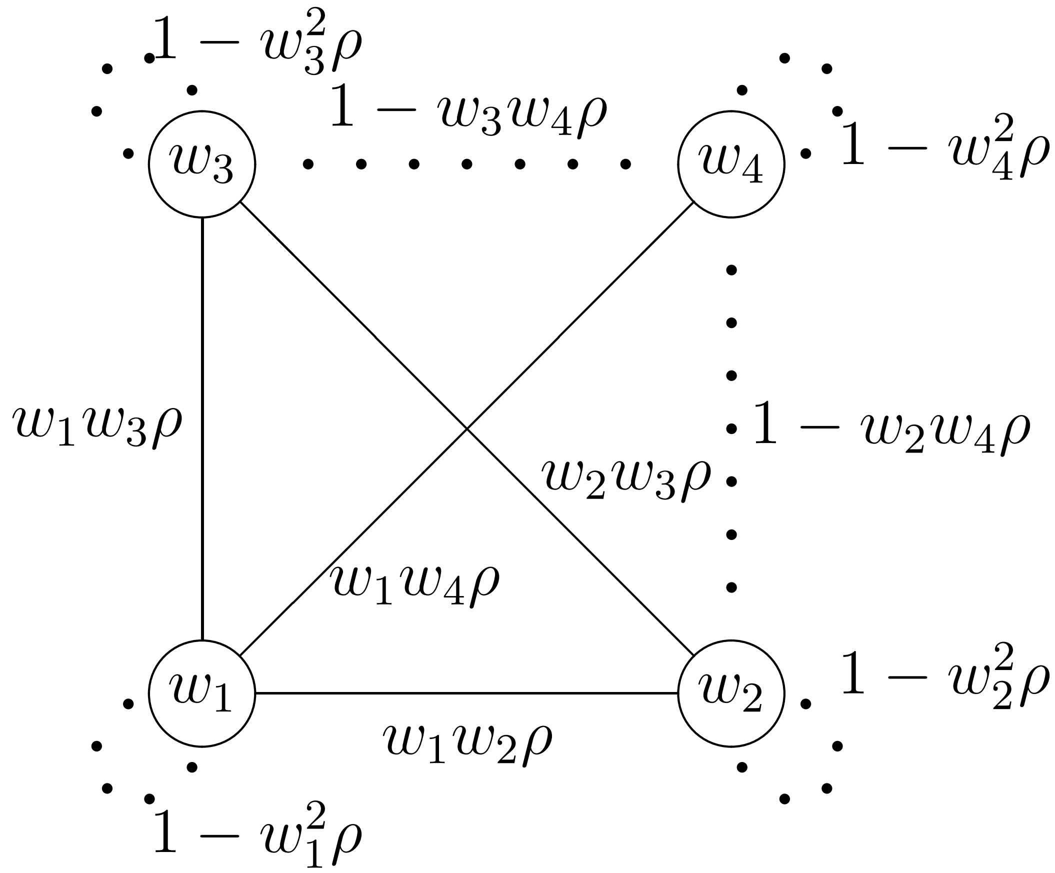
An example: $G(w_1, w_2, w_3, w_4)$



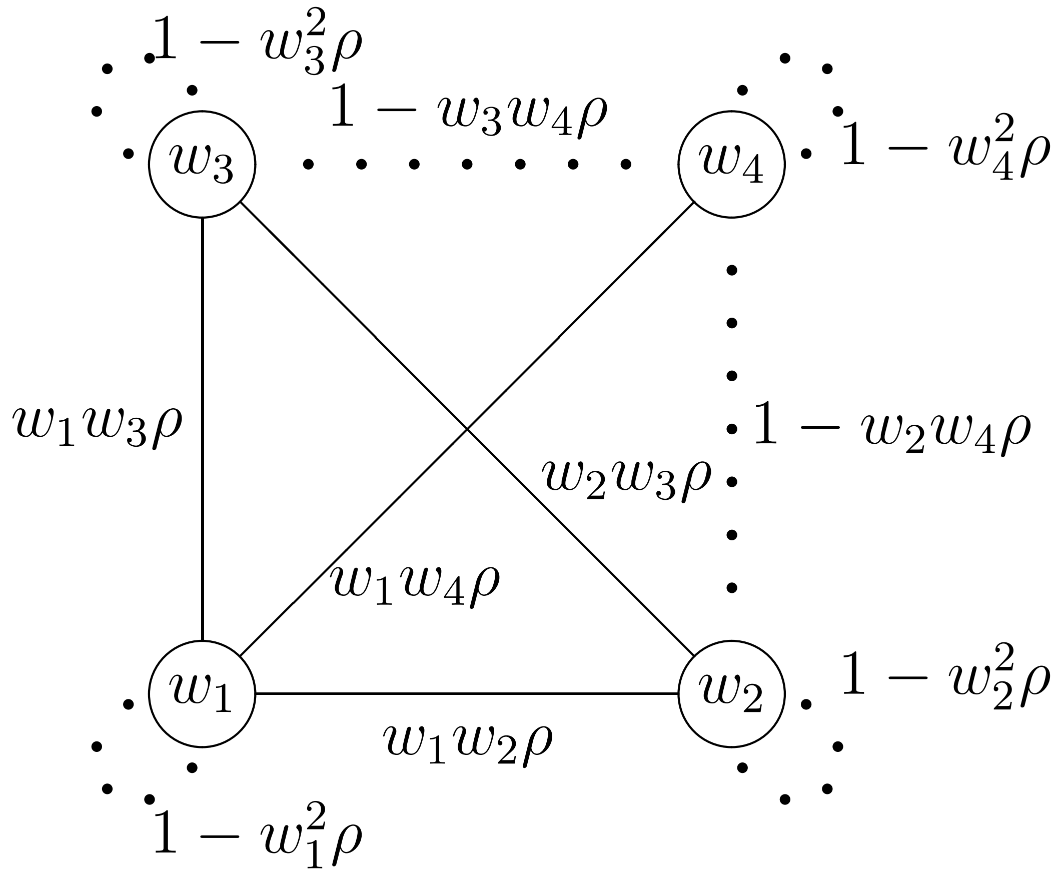
An example: $G(w_1, w_2, w_3, w_4)$



An example: $G(w_1, w_2, w_3, w_4)$



An example: $G(w_1, w_2, w_3, w_4)$



The probability of the graph is

$$w_1^3w_2^2w_3^2w_4\rho^4(1 - w_2w_4\rho) \times (1 - w_3w_4\rho)\prod_{i=1}^{4}(1 - w_i^2\rho).$$

Notations

For $G = G(w_1, \ldots, w_n)$, let

- $d = \frac{1}{n} \sum_{i=1}^{n} w_i$
- $\tilde{d} = \frac{\sum_{i=1}^{n} w_i^2}{\sum_{i=1}^{n} w_i}$.
- The volume of S: $\mathrm{Vol}(S) = \sum_{i \in S} w_i$.

Notations

For $G = G(w_1, \ldots, w_n)$, let

- $d = \frac{1}{n}\sum_{i=1}^n w_i$
- $\tilde{d} = \frac{\sum_{i=1}^n w_i^2}{\sum_{i=1}^n w_i}$.
- The volume of S: $\mathrm{Vol}(S) = \sum_{i \in S} w_i$.

We have

$$\tilde{d} \geq d$$

"$=$" holds if and only if $w_1 = \cdots = w_n$.

Notations

For $G = G(w_1, \ldots, w_n)$, let

- $d = \frac{1}{n}\sum_{i=1}^{n} w_i$
- $\tilde{d} = \frac{\sum_{i=1}^{n} w_i^2}{\sum_{i=1}^{n} w_i}$.
- The volume of S: $\mathrm{Vol}(S) = \sum_{i \in S} w_i$.

We have

$$\tilde{d} \geq d$$

"$=$" holds if and only if $w_1 = \cdots = w_n$.

A connected component S is called a giant component if

$$\mathrm{Vol}(S) = \Theta(\mathrm{Vol}(G)).$$

Connected components

Chung and Lu (2001) For $G = G(w_1, \ldots, w_n)$,

- If $\tilde{d} < 1 - \epsilon$, then almost surely, all components have volume at most $O(\sqrt{n} \log n)$.

Connected components

Chung and Lu (2001) For $G = G(w_1, \ldots, w_n)$,

- If $\tilde{d} < 1 - \epsilon$, then almost surely, all components have volume at most $O(\sqrt{n} \log n)$.
- If $d > 1 + \epsilon$, then almost surely there is a unique giant component of volume $\Theta(\mathrm{Vol}(G))$. All other components have size at most

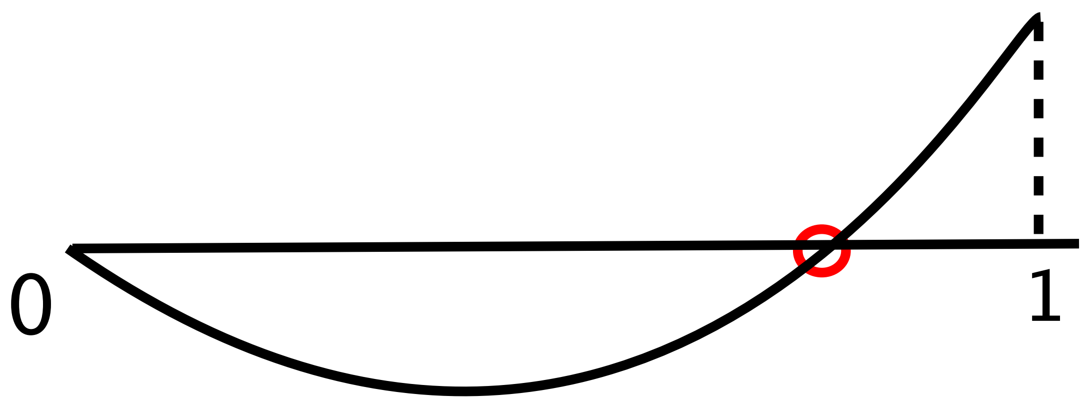
$$\begin{cases} \frac{\log n}{d-1-\log d - \epsilon d} & \text{if } \frac{1}{1-\epsilon} < d < \frac{2}{1-\epsilon} \\ \frac{\log n}{1+\log d - \log 4 + 2\log(1-\epsilon)} & \text{if } d > \frac{4}{e(1-\epsilon)^2}. \end{cases}$$

Volume of Giant Component

Chung and Lu (2004)

If the average degree is strictly greater than 1, then almost surely the giant component in a graph G in $G(\mathbf{w})$ has volume $(\lambda_0 + O(\sqrt{n}\frac{\log^{3.5} n}{\mathrm{Vol}(G)}))\mathrm{Vol}(G)$, where λ_0 is the unique positive root of the following equation:

$$\sum_{i=1}^{n} w_i e^{-w_i \lambda} = (1-\lambda) \sum_{i=1}^{n} w_i .$$



$G(n,p)$ **verse** $G(w_1,\ldots,w_n)$

Question: Does the random graph with equal expected degrees generates the smallest giant component among all possible degree distribution with the same volume?

$G(n,p)$ verse $G(w_1,\ldots,w_n)$

Question: Does the random graph with equal expected degrees generates the smallest giant component among all possible degree distribution with the same volume?

Chung Lu (2004)

- Yes, for $1 < d \leq \frac{e}{e-1}$.

$G(n, p)$ verse $G(w_1, \ldots, w_n)$

Question: Does the random graph with equal expected degrees generates the smallest giant component among all possible degree distribution with the same volume?

Chung Lu (2004)

- Yes, for $1 < d \le \frac{e}{e-1}$.
- No, for sufficiently large d.

$G(n,p)$ verse $G(w_1,\ldots,w_n)$

Question: Does the random graph with equal expected degrees generates the smallest giant component among all possible degree distribution with the same volume?
Chung Lu (2004)

- Yes, for $1 < d \leq \frac{e}{e-1}$.
- No, for sufficiently large d.
- When $d \geq \frac{4}{e}$, almost surely the giant component of $G(w_1,\ldots,w_n)$ has volume at least

$$\Big(\frac{1}{2}(1+\sqrt{1-\frac{4}{de}}) + o(1)\Big)\mathrm{Vol}(G).$$

This is asymptotically best possible.

"Six degree separation"

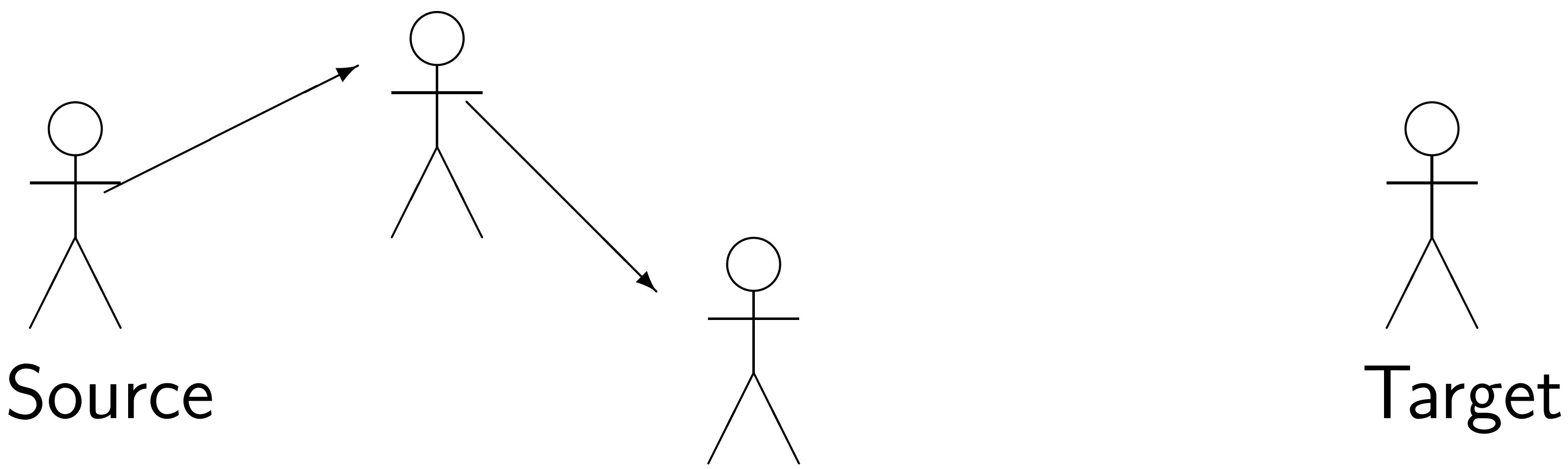
Experiments of Stanley Milgram (1967)

“Six degree separation”

Experiments of Stanley Milgram (1967)

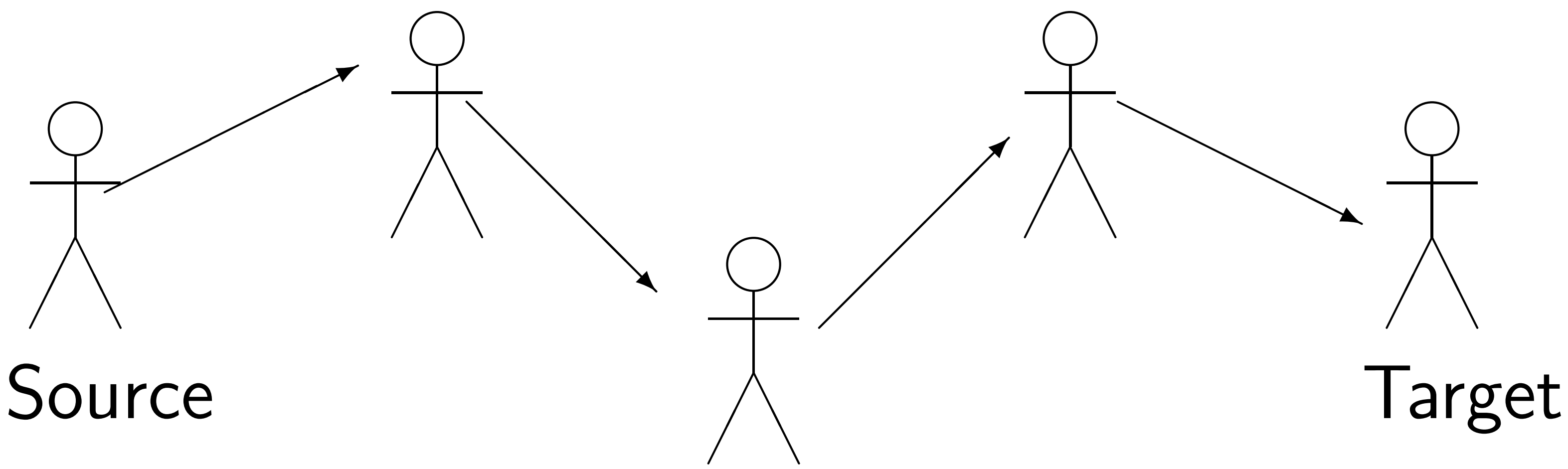
"Six degree separation"

Experiments of Stanley Milgram (1967)



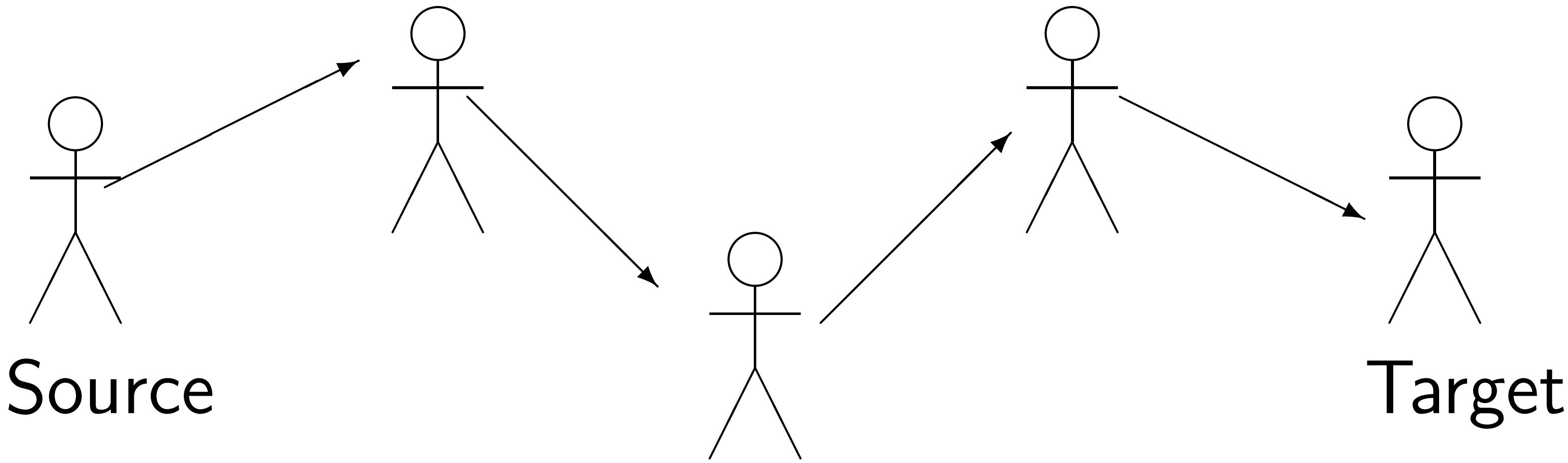
“Six degree separation”

Experiments of Stanley Milgram (1967)



“Six degree separation”

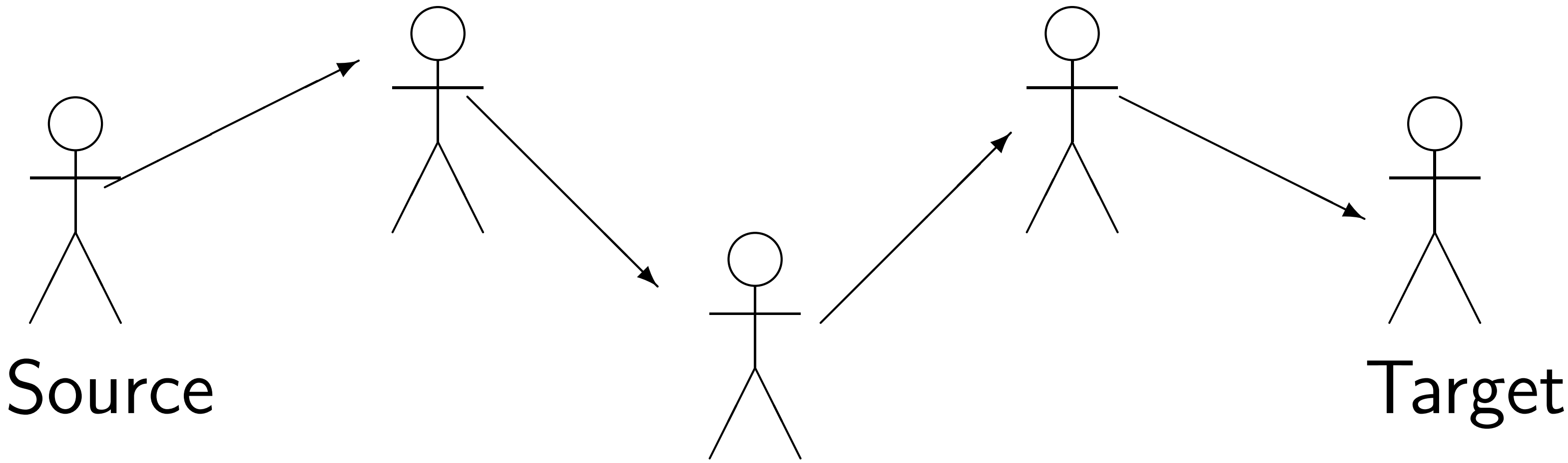
Experiments of Stanley Milgram (1967)



Diameter: the maximum distance $d(u, v)$, where u and v are in the same connected component.

"Six degree separation"

Experiments of Stanley Milgram (1967)



Diameter: the maximum distance $d(u,v)$, where u and v are in the same connected component.
Average distance: the average among all distance $d(u,v)$ for pairs of u and v in the same connected component.

Diameter of $G(n, p)$

Bollobás (1985): (denser graph)

$$diam(G(n,p)) = \lfloor \frac{\log n}{\log np} \rfloor \text{ or } \lceil \frac{\log n}{\log np} \rceil \text{ if } np \gg \log n.$$

Diameter of $G(n, p)$

Bollobás (1985): (denser graph)

$$diam(G(n,p)) = \lfloor \frac{\log n}{\log np} \rfloor \text{ or } \lceil \frac{\log n}{\log np} \rceil \text{ if } np \gg \log n.$$

Chung Lu, (2000) (Sparser graph)

$$diam(G(n,p)) = \begin{cases} (1+o(1))\frac{\log n}{\log np} & \text{if } np \to \infty \\ \Theta(\frac{\log n}{\log np}) & \text{if } \infty > np > 1. \end{cases}$$

Diameter of $G(w_1, \ldots, w_n)$

Chung Lu (2002)

- For a random graph G with admissible expected degree sequence $(w_1, \ldots, w_n)$, the average distance is almost surely $(1 + o(1))\frac{\log n}{\log \tilde{d}}$.

Diameter of $G(w_1, \ldots, w_n)$

Chung Lu (2002)

- For a random graph G with admissible expected degree sequence $(w_1, \ldots, w_n)$, the average distance is almost surely $(1 + o(1))\frac{\log n}{\log \tilde{d}}$.
- For a random graph G with strongly admissible expected degree sequence $(w_1, \ldots, w_n)$, the diameter is almost surely $\Theta(\frac{\log n}{\log \tilde{d}})$.

Diameter of $G(w_1, \ldots, w_n)$

Chung Lu (2002)

- For a random graph G with admissible expected degree sequence $(w_1, \ldots, w_n)$, the average distance is almost surely $(1 + o(1))\frac{\log n}{\log \tilde{d}}$.
- For a random graph G with strongly admissible expected degree sequence $(w_1, \ldots, w_n)$, the diameter is almost surely $\Theta(\frac{\log n}{\log \tilde{d}})$.

These results apply to $G(n, p)$ and random power law graph with $\beta > 3$.

Admissible condition

(i) $\log \tilde{d} \ll \log n$.
(ii) $d > 1 + \epsilon$. $w_i > \epsilon$ for all but $o(n)$ vertices.
(iii) $\exists$ a subset U:

$$\mathrm{Vol}_2(U) = (1 + o(1))\mathrm{Vol}_2(G) \gg \mathrm{Vol}_3(U) \frac{\log \tilde{d} \log\log n}{\tilde{d} \log n}.$$

Here $\mathrm{Vol}_k(U) = \sum_{i \in U} w_i^k$.

Admissible condition

(i) $\log \tilde{d} \ll \log n$.
(ii) $d > 1 + \epsilon$. $w_i > \epsilon$ for all but $o(n)$ vertices.
(iii) $\exists$ a subset U:

$$\mathrm{Vol}_2(U) = (1 + o(1))\mathrm{Vol}_2(G) \gg \mathrm{Vol}_3(U)\frac{\log \tilde{d} \operatorname{loglog} n}{\tilde{d} \log n}.$$

Here $\mathrm{Vol}_k(U) = \sum_{i \in U} w_i^k$.
Roughly speaking, G is close to $G(n, p)$. No dense subgraphs.

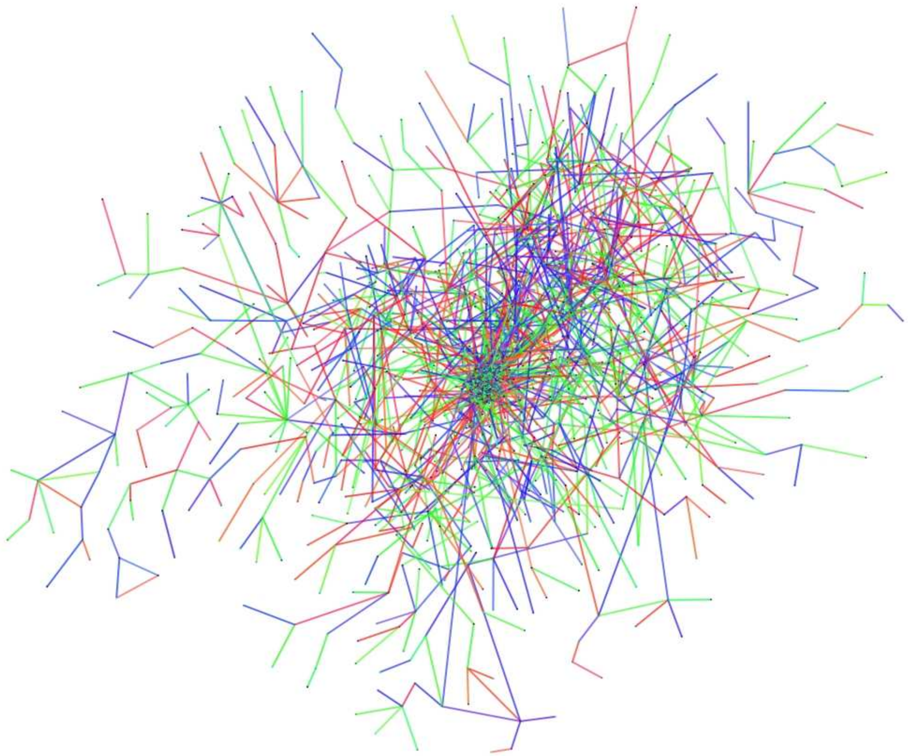
Admissible condition

(i) $\log \tilde{d} \ll \log n$.
(ii) $d > 1 + \epsilon$. $w_i > \epsilon$ for all but $o(n)$ vertices.
(iii) $\exists$ a subset U:

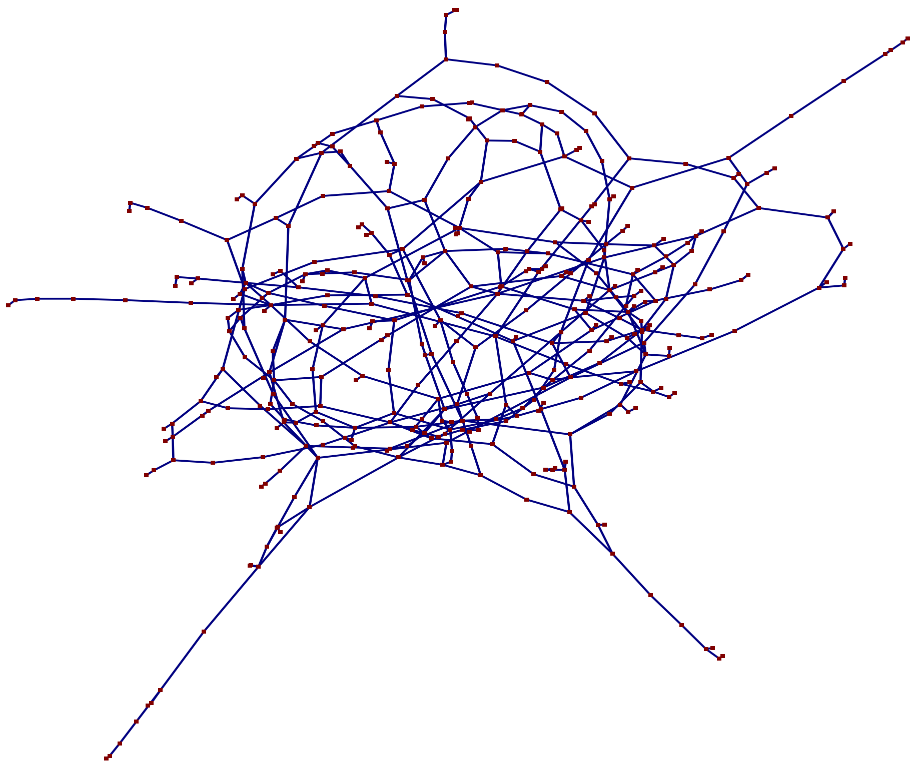
$$\mathrm{Vol}_2(U) = (1 + o(1))\mathrm{Vol}_2(G) \gg \mathrm{Vol}_3(U)\frac{\log \tilde{d} \log\log n}{\tilde{d} \log n}.$$

Here $\mathrm{Vol}_k(U) = \sum_{i \in U} w_i^k$.
Roughly speaking, G is close to $G(n, p)$. No dense subgraphs.
Example: Power law graphs with $\beta > 3$ and $G(n, p)$.

Non-admissible graph versus admissible graph

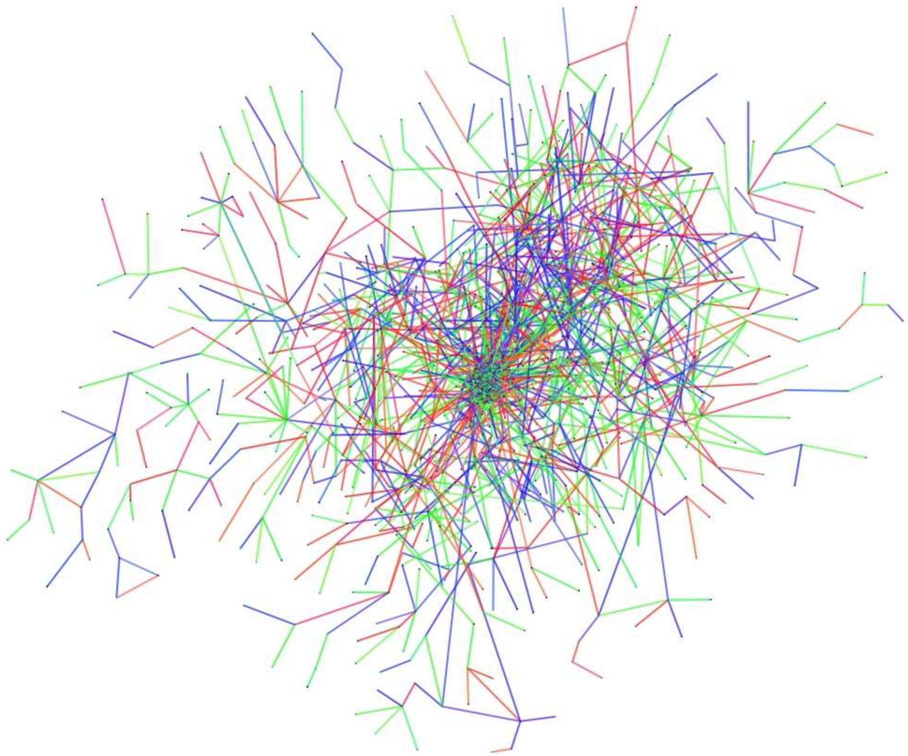


A random subgraph of the Collaboration Graph.

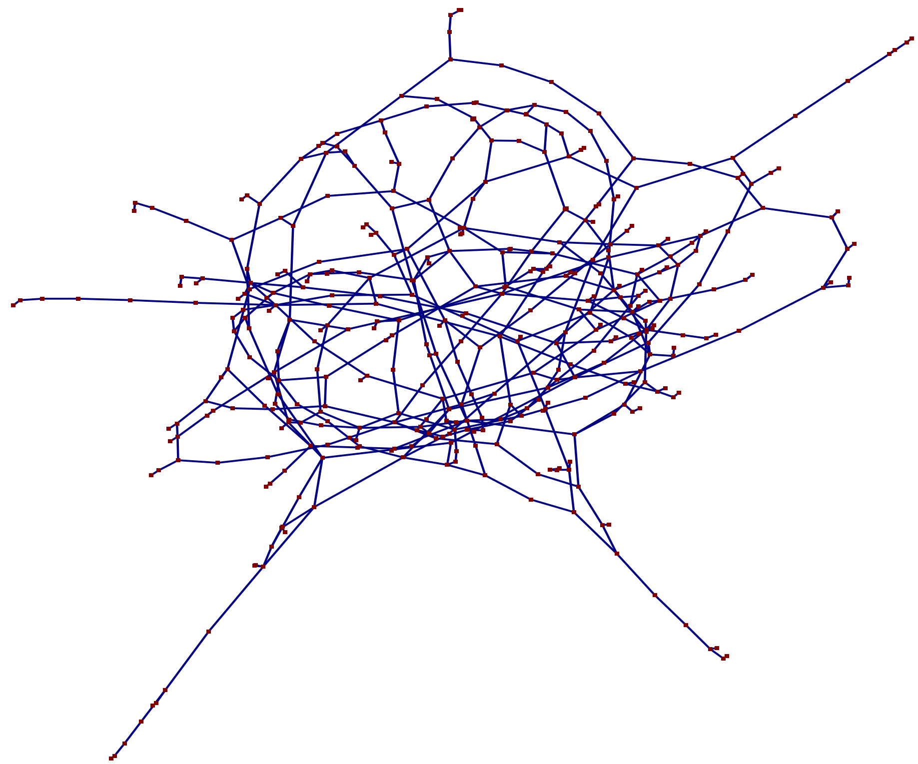


A Connected component of $G(n, p)$ with $n = 500$ and $p = 0.002$.

Non-admissible graph versus admissible graph



A random subgraph of the Collaboration Graph.



A Connected component of $G(n, p)$ with $n = 500$ and $p = 0.002$.

- Dense core for non-admissible graphs.
- No dense core for admissible graphs.

Power law graphs with $\beta \in (2, 3)$

Chung, Lu (2002)

- Examples: the WWW graph, Collaboration graph, etc.

Power law graphs with $\beta \in (2, 3)$

Chung, Lu (2002)

- Examples: the WWW graph, Collaboration graph, etc.
- Non-admissible.

Power law graphs with $\beta \in (2, 3)$

Chung, Lu (2002)

- Examples: the WWW graph, Collaboration graph, etc.
- Non-admissible.
- Containing a dense core, with diameter $\log \log n$.

Power law graphs with $\beta \in (2,3)$

Chung, Lu (2002)

- Examples: the WWW graph, Collaboration graph, etc.
- Non-admissible.
- Containing a dense core, with diameter $\log \log n$.
- Mostly vertices are within the distance of $O(\log \log n)$ from the core.

Power law graphs with $\beta \in (2, 3)$

Chung, Lu (2002)

- Examples: the WWW graph, Collaboration graph, etc.
- Non-admissible.
- Containing a dense core, with diameter $\log\log n$.
- Mostly vertices are within the distance of $O(\log\log n)$ from the core.
- There are some vertices at the distance of $O(\log n)$.

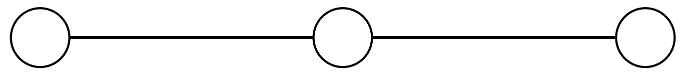
Power law graphs with $\beta \in (2,3)$

Chung, Lu (2002)

- Examples: the WWW graph, Collaboration graph, etc.
- Non-admissible.
- Containing a dense core, with diameter $\log \log n$.
- Mostly vertices are within the distance of $O(\log \log n)$ from the core.
- There are some vertices at the distance of $O(\log n)$.

The diameter is $\Theta(\log n)$, while the average distance is $O(\log \log n)$.

Eigenvalues of a graph

A graph G: 
Adjacency matrix:

$$A = \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix}$$

Eigenvalues are

$$2, 0, 0.$$

Wigner's semicircle law

Wigner (1958)

- A is a real symmetric $n \times n$ matrix.
- Entries a_{ij} are independent random variables.
- $E(a_{ij}^{2k+1}) = 0$.
- $E(a_{ij}^2) = m^2$.
- $E(a_{ij}^{2k}) < M$.

The distribution of eigenvalues of A converges into a semicircle distribution of radius $2m\sqrt{n}$.

Wigner's semicircle law

Wigner (1958)

- A is a real symmetric $n \times n$ matrix.
- Entries a_{ij} are independent random variables.
- $E(a_{ij}^{2k+1}) = 0$.
- $E(a_{ij}^{2}) = m^2$.
- $E(a_{ij}^{2k}) < M$.

The distribution of eigenvalues of A converges into a semicircle distribution of radius $2m\sqrt{n}$.

Füredi and Komlós (1981): The eigenvalues of $G(n, p)$ follows Wigner's semicircle law.

Experimental results

- **Faloutsos et al. (1999)** The eigenvalues of the Internet graph do not follow the semicircle law.
- **Farkas et. al. (2001), Goh et. al. (2001)** The spectrum of a power law graph follows a "triangular-like" distribution.
- **Mihail and Papadimitriou (2002)** They showed that the large eigenvalues are determined by the large degrees. Thus, the significant part of the spectrum of a power law graph follows the power law.

$$\mu_i \approx \sqrt{d_i}.$$

Eigenvalues of $G(w_1, \ldots, w_n)$

Chung, Vu, and Lu (2003)
Suppose $w_1 \geq w_2 \geq \ldots \geq w_n$. Let μ_i be i-th largest eigenvalue of $G(w_1, w_2, \ldots, w_n)$. Let $m = w_1$ and $\tilde{d} = \sum_{i=1}^n w_i^2 \rho$. Almost surely we have:

- $(1–o(1)) \max\{\sqrt{m}, \tilde{d}\} \leq \mu_1 \leq 7\sqrt{\log n} \cdot \max\{\sqrt{m}, \tilde{d}\}$.

Eigenvalues of $G(w_1, \ldots, w_n)$

Chung, Vu, and Lu (2003)
Suppose $w_1 \geq w_2 \geq \ldots \geq w_n$. Let μ_i be i-th largest eigenvalue of $G(w_1, w_2, \ldots, w_n)$. Let $m = w_1$ and $\tilde{d} = \sum_{i=1}^{n} w_i^2 \rho$. Almost surely we have:

- $(1 - o(1)) \max\{\sqrt{m}, \tilde{d}\} \leq \mu_1 \leq 7\sqrt{\log n} \cdot \max\{\sqrt{m}, \tilde{d}\}$.
- $\mu_1 = (1 + o(1))\tilde{d}$, if $\tilde{d} > \sqrt{m} \log n$.

Eigenvalues of $G(w_1, \ldots, w_n)$

Chung, Vu, and Lu (2003)
Suppose $w_1 \geq w_2 \geq \ldots \geq w_n$. Let μ_i be i-th largest eigenvalue of $G(w_1, w_2, \ldots, w_n)$. Let $m = w_1$ and $\tilde{d} = \sum_{i=1}^n w_i^2 \rho$. Almost surely we have:

- $(1 - o(1)) \max\{\sqrt{m}, \tilde{d}\} \leq \mu_1 \leq 7\sqrt{\log n} \cdot \max\{\sqrt{m}, \tilde{d}\}$.
- $\mu_1 = (1 + o(1))\tilde{d}$, if $\tilde{d} > \sqrt{m} \log n$.
- $\mu_1 = (1 + o(1))\sqrt{m}$, if $\sqrt{m} > \tilde{d} \log^2 n$.

Eigenvalues of $G(w_1, \ldots, w_n)$

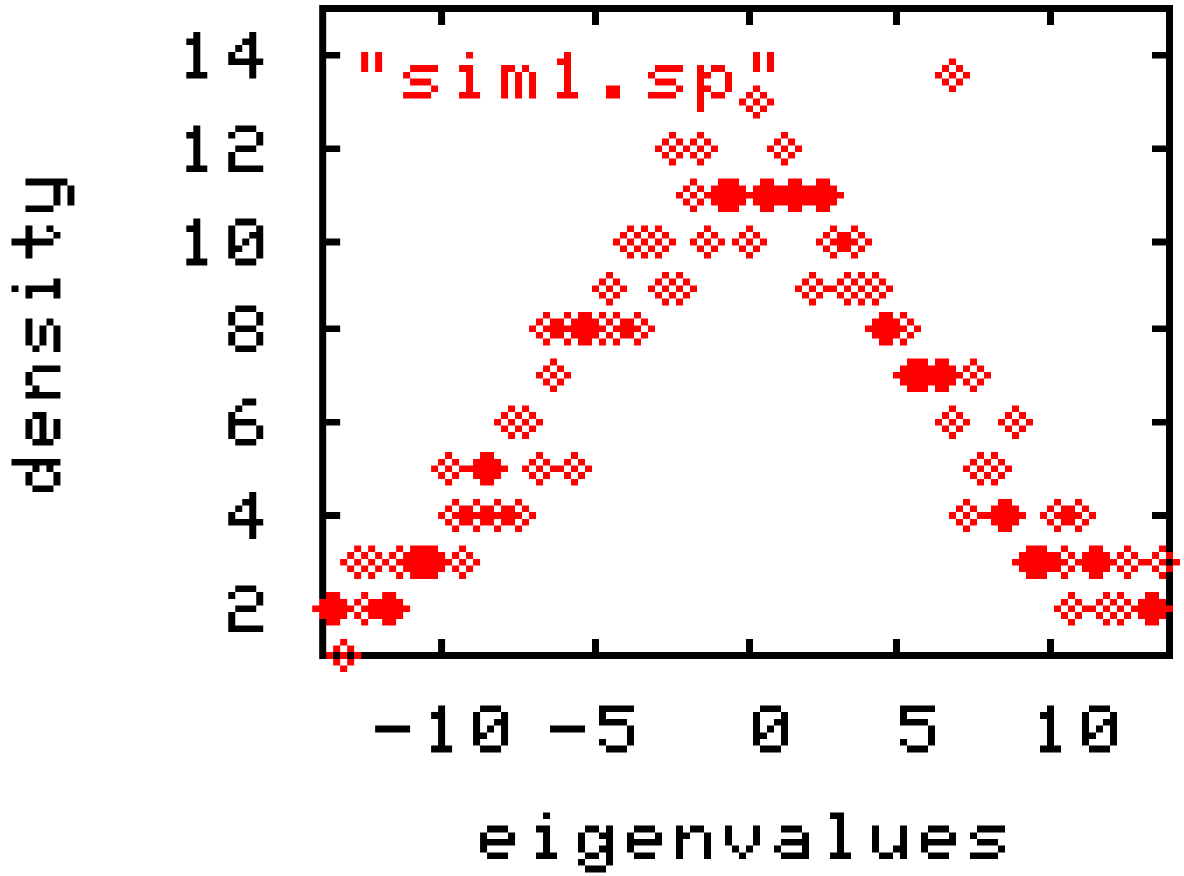
Chung, Vu, and Lu (2003)

Suppose $w_1 \geq w_2 \geq \ldots \geq w_n$. Let μ_i be i-th largest eigenvalue of $G(w_1, w_2, \ldots, w_n)$. Let $m = w_1$ and $\tilde{d} = \sum_{i=1}^{n} w_i^2 \rho$. Almost surely we have:

- $(1-o(1)) \max\{\sqrt{m}, \tilde{d}\} \leq \mu_1 \leq 7\sqrt{\log n} \cdot \max\{\sqrt{m}, \tilde{d}\}$.
- $\mu_1 = (1 + o(1))\tilde{d}$, if $\tilde{d} > \sqrt{m} \log n$.
- $\mu_1 = (1 + o(1))\sqrt{m}$, if $\sqrt{m} > \tilde{d} \log^2 n$.
- $\mu_k \approx \sqrt{w_k}$ and $\mu_{n+1-k} \approx -\sqrt{w_k}$, if $\sqrt{w_k} > \tilde{d} \log^2 n$.

Random power law graphs

The first k and last k eigenvalues of the random power law graph with $\beta > 2.5$ follows the power law distribution with exponent $2\beta - 1$. It results a "triangular-like" shape.

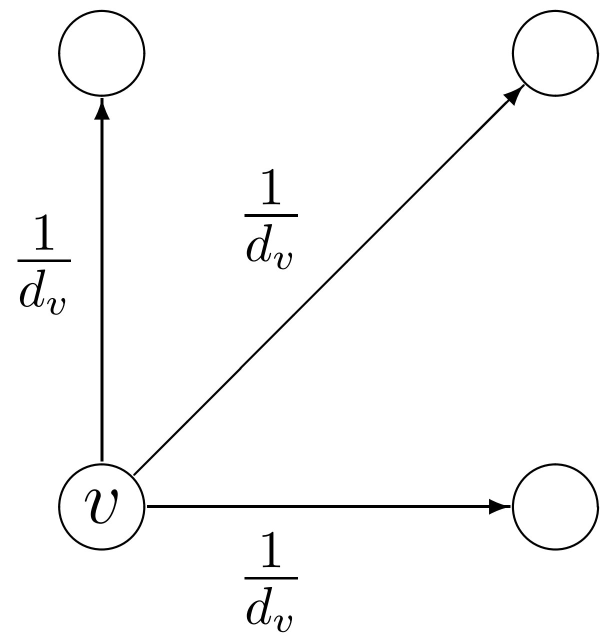


Laplacian spectrum

Random walks on a graph G:

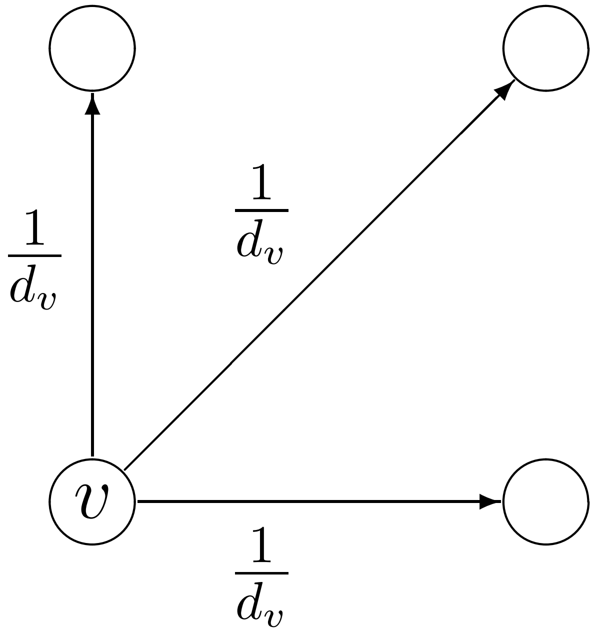
$$\pi_{k+1} = AD^{-1}\pi_k.$$

$$AD^{-1} \sim D^{-1/2}AD^{-1/2}.$$



Laplacian spectrum

Random walks on a graph G:



$$\pi_{k+1} = AD^{-1}\pi_k.$$

$$AD^{-1} \sim D^{-1/2}AD^{-1/2}.$$

Laplacian spectrum

$$0 = \lambda_0 \le \lambda_1 \le \cdots \le \lambda_{n-1} \le 2$$

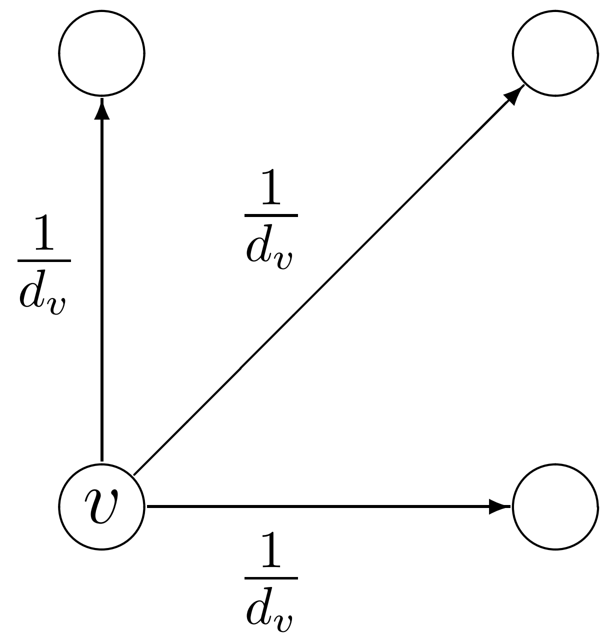
are the eigenvalues of $L = I - D^{-1/2}AD^{-1/2}$.

Laplacian spectrum

Random walks on a graph G:

$$\pi_{k+1} = AD^{-1}\pi_k.$$

$$AD^{-1} \sim D^{-1/2}AD^{-1/2}.$$



Laplacian spectrum

$$0 = \lambda_0 \leq \lambda_1 \leq \cdots \leq \lambda_{n-1} \leq 2$$

are the eigenvalues of $L = I - D^{-1/2}AD^{-1/2}$.
The eigenvalues of AD^{-1} are $1, 1-\lambda_1, \ldots, 1-\lambda_{n-1}$.

Spectral Radius

Let

- $w_{min} = \min\{w_1, \ldots, w_n\}$
- $d = \frac{1}{n}\sum_{i=1}^n w_i$
- $g(n)$ — a function tending to infinity arbitrarily slowly.

Chung, Vu, and Lu (2003)

If $w_{\min} \gg \log^2 n$, then almost surely the Laplacian spectrum λ_i's of $G(w_1, \ldots, w_n)$ satisfy

$$\max_{i \neq 0} |1 - \lambda_i| \leq (1 + o(1))\frac{4}{\sqrt{d}} + \frac{g(n)\log^2 n}{w_{\min}}.$$

Approximation

$$M = D^{-1/2} A D^{-1/2} - \phi_0^* \phi_0$$

where

$$\phi_0 = \frac{1}{\sqrt{\sum_{i=1}^n d_i}} (\sqrt{d_1}, \ldots, \sqrt{d_n})^*.$$

$$C = W^{-1/2} A W^{-1/2} - \chi^* \chi$$

where

$$\chi = \frac{1}{\sqrt{\sum_{i=1}^n w_i}} (\sqrt{w_1}, \ldots, \sqrt{w_n})^*.$$

Approximation

$$M = D^{-1/2} A D^{-1/2} - \phi_0^* \phi_0$$

where

$$\phi_0 = \frac{1}{\sqrt{\sum_{i=1}^n d_i}} (\sqrt{d_1}, \ldots, \sqrt{d_n})^*.$$

$$C = W^{-1/2} A W^{-1/2} - \chi^* \chi$$

where

$$\chi = \frac{1}{\sqrt{\sum_{i=1}^n w_i}} (\sqrt{w_1}, \ldots, \sqrt{w_n})^*.$$

- C can be viewed as the “expectation” of M.

Approximation

$$M = D^{-1/2} A D^{-1/2} - \phi_0^* \phi_0$$

where

$$\phi_0 = \frac{1}{\sqrt{\sum_{i=1}^n d_i}} (\sqrt{d_1}, \ldots, \sqrt{d_n})^*.$$

$$C = W^{-1/2} A W^{-1/2} - \chi^* \chi$$

where

$$\chi = \frac{1}{\sqrt{\sum_{i=1}^n w_i}} (\sqrt{w_1}, \ldots, \sqrt{w_n})^*.$$

- C can be viewed as the "expectation" of M. We have

$$\|M - C\| \leq (1 + o(1)) \frac{2}{\sqrt{d}}.$$

Approximation

$$M = D^{-1/2}AD^{-1/2} - \phi_0^*\phi_0$$

where

$$\phi_0 = \frac{1}{\sqrt{\sum_{i=1}^n d_i}}(\sqrt{d_1}, \ldots, \sqrt{d_n})^*.$$

$$C = W^{-1/2}AW^{-1/2} - \chi^*\chi$$

where

$$\chi = \frac{1}{\sqrt{\sum_{i=1}^n w_i}}(\sqrt{w_1}, \ldots, \sqrt{w_n})^*.$$

- C can be viewed as the “expectation” of M. We have

$$\|M - C\| \leq (1 + o(1))\frac{2}{\sqrt{d}}.$$

- M has eigenvalues $0, 1-\lambda_1, \ldots, 1-\lambda_{n-1}$, since $M = I - L - \phi_0^*\phi_0$ and $L\phi_0 = 0$.

Results on spectrum of C

Chung, Vu, and Lu (2003)
We have

- If $w_{\min} \gg \sqrt{d}\log^2 n$, then
$$\|C\| = (1 + o(1))\frac{2}{\sqrt{d}}.$$

Results on spectrum of C

Chung, Vu, and Lu (2003)
We have

- If $w_{\min} \gg \sqrt{d}\log^2 n$, then
$$\|C\| = (1 + o(1))\frac{2}{\sqrt{d}}.$$
- If $w_{\min} \gg \sqrt{d}$, the eigenvalues of C follow the semi-circle distribution with radius $r \approx \frac{2}{\sqrt{d}}$.

Coupling methods

- G_1 and G_2: two random graphs on n vertices.

Coupling methods

- G_1 and G_2: two random graphs on n vertices.
- Almost surely $G_1 \succeq G_2$: for any monotone property A

$$\Pr(G_1 \text{ satisfies } A) \geq \Pr(G_2 \text{ satisfies } A) - o_n(1).$$

Coupling methods

- G_1 and G_2: two random graphs on n vertices.
- Almost surely $G_1 \succeq G_2$: for any monotone property A

$$\Pr(G_1 \text{ satisfies } A) \geq \Pr(G_2 \text{ satisfies } A) - o_n(1).$$

- Almost surely $G_1 \preceq G_2$: for any monotone property A

$$\Pr(G_1 \text{ satisfies } A) \leq \Pr(G_2 \text{ satisfies } A) + o_n(1).$$

Coupling methods

- G_1 and G_2: two random graphs on n vertices.
- Almost surely $G_1 \succeq G_2$: for any monotone property A

$$\Pr(G_1 \text{ satisfies } A) \geq \Pr(G_2 \text{ satisfies } A) - o_n(1).$$

- Almost surely $G_1 \preceq G_2$: for any monotone property A

$$\Pr(G_1 \text{ satisfies } A) \leq \Pr(G_2 \text{ satisfies } A) + o_n(1).$$

A monotone property is closed under edge-addition.

- "G is Hamiltonian."
- "G contains a subgraph H."
- "The diameter of G is at most k."

Example of coupling

- $F(n,m)$: uniform random graphs on n vertices and m edges.
- $G(n,p)$: Erdős-Rényi random graphs.

With $p = \frac{m}{\binom{n}{2}}$, for any $\delta > 0$, almost surely we have

$$G(n,(1-\delta)p) \preceq F(n,m) \preceq G(n,(1+\delta)p).$$

Example of coupling

- $F(n,m)$: uniform random graphs on n vertices and m edges.
- $G(n,p)$: Erdős-Rényi random graphs.

With $p = \frac{m}{\binom{n}{2}}$, for any $\delta > 0$, almost surely we have

$$G(n,(1-\delta)p) \preceq F(n,m) \preceq G(n,(1+\delta)p).$$

Can we couple evolution models with static models?

$G(p_1, p_2, p_3, p_4, m)$

At each time t,

- with probability p_1, take a vertex-growth step; add a new vertex v and form m new edges from v to existing vertices u chosen with probability proportional to d_u.

$G(p_1, p_2, p_3, p_4, m)$

At each time t,

- with probability p_1, take a vertex-growth step; add a new vertex v and form m new edges from v to existing vertices u chosen with probability proportional to d_u.
- with probability p_2, take a m edge-growth steps;

$G(p_1, p_2, p_3, p_4, m)$

At each time t,

- with probability p_1, take a vertex-growth step; add a new vertex v and form m new edges from v to existing vertices u chosen with probability proportional to d_u.
- with probability p_2, take a m edge-growth steps;
- with probability p_3, take a vertex-deletion step;

$G(p_1, p_2, p_3, p_4, m)$

At each time t,

- with probability p_1, take a vertex-growth step; add a new vertex v and form m new edges from v to existing vertices u chosen with probability proportional to d_u.
- with probability p_2, take a m edge-growth steps;
- with probability p_3, take a vertex-deletion step;
- with probability $p_4 = 1 - p_1 - p_2 - p_3$, take m edge-deletion steps.

Degree distribution

Chung-Lu (2004), Frieze-Cooper-Vera (2004)
For $p_1 > p_3$ and $p_2 > p_4$, $G(p_1, p_2, p_3, p_4, m)$ almost surely generates a power law graphs with exponent

$$\beta = 2 + \frac{p_1 + p_3}{p_1 + 2p_2 - p_3 - 2p_4}.$$

Coupling result

Suppose $p_3 < p_1$, $p_4 < p_2$, and $\log n \ll m < t^{\frac{p_1}{2(p_1+p_2)}}$. Then $G(p_1, p_2, p_{3,4}, m)$ dominates and is dominated by an edge-independent graph with probability $p_{ij}^{(t)}$ of having an edge between vertices i and j, $i < j$, at time t, with $p_{ij}^{(t)}$ satisfying:

$$\begin{cases} \frac{p_2 m}{2p_4\tau(2p_2-p_4)} \frac{l^{2\alpha-1}}{i^\alpha j^\alpha}\left(1 + \left(1 - \frac{p_4}{p_2}\right)\left(\frac{j}{t}\right)^{\frac{1}{2\tau}+2\alpha-1}\right) & \text{if } i^\alpha j^\alpha \gg \frac{p_2 m t^{2\alpha-1}}{4\tau^2 p_4} \\ 1 - (1 + o(1))\frac{2p_4\tau}{p_2 m} i^\alpha j^\alpha t^{1-2\alpha} & \text{if } i^\alpha j^\alpha \ll \frac{p_2 m t^{2\alpha-1}}{4\tau^2 p_4} \end{cases}$$

where $\alpha = \frac{p_1(p_1+2p_2-p_3-2p_4)}{2(p_1+p_2-p_4)(p_1-p_3)}$ and $\tau = \frac{(p_1+p_2-p_4)(p_1-p_3)}{p_1+p_3}$.

Corollary

Suppose $m > \log^{1+\epsilon} n$.

- $G(p1, p2, p3, p4, m)$ follows the power law distribution with exponent $\beta = 2 + (p1 + p3)/(p1 + 2p2 - p3 - 2p4)$.

Corollary

Suppose $m > \log^{1+\epsilon} n$.

- $G(p1, p2, p3, p4, m)$ follows the power law distribution with exponent $\beta = 2 + (p1 + p3)/(p1 + 2p2 - p3 - 2p4)$.
- For $p_2 > p_3 + p_4$, we have $2 < \beta < 3$. Almost surely a random graph in $G(p_1, p_2, p_3, p_4, m)$ has diameter $\Theta(\log n)$ and average distance $O(\frac{\log \log n}{\log(1/(\beta - 2))})$.

Corollary

Suppose $m > \log^{1+\epsilon} n$.

- $G(p1, p2, p3, p4, m)$ follows the power law distribution with exponent $\beta = 2 + (p1 + p3)/(p1 + 2p2 - p3 - 2p4)$.
- For $p_2 > p_3 + p_4$, we have $2 < \beta < 3$. Almost surely a random graph in $G(p_1, p_2, p_3, p_4, m)$ has diameter $\Theta(\log n)$ and average distance $O(\frac{\log\log n}{\log(1/(\beta-2)})$.
- For $p_2 < p_3 + p_4$, we have $\beta > 3$. Almost surely a random graph in $G(p_1, p_2, p_3, p_4, m)$ has diameter $\Theta(\log n)$ and average distance $O(\frac{\log n}{\log d})$ where d is the average degree.

Corollary

Suppose $m > \log^{1+\epsilon} n$.

- $G(p1, p2, p3, p4, m)$ follows the power law distribution with exponent $\beta = 2 + (p1 + p3)/(p1 + 2p2 - p3 - 2p4)$.
- For $p_2 > p_3 + p_4$, we have $2 < \beta < 3$. Almost surely a random graph in $G(p_1, p_2, p_3, p_4, m)$ has diameter $\Theta(\log n)$ and average distance $O(\frac{\log \log n}{\log(1/(\beta-2)})$.
- For $p_2 < p_3 + p_4$, we have $\beta > 3$. Almost surely a random graph in $G(p_1, p_2, p_3, p_4, m)$ has diameter $\Theta(\log n)$ and average distance $O(\frac{\log n}{\log d})$ where d is the average degree.
- Almost surely a random graph in $G(p_1, p_2, p_3, p_4, m)$ has spectral gap λ at least $1/8 + o(1)$.

Summary

Topics we have covered:

- Examples of complex networks
- Evolution models
- Static models
- Coupling methods

Summary

Topics we have covered:

- Examples of complex networks
- Evolution models
- Static models
- Coupling methods

Topics we have not covered but important:

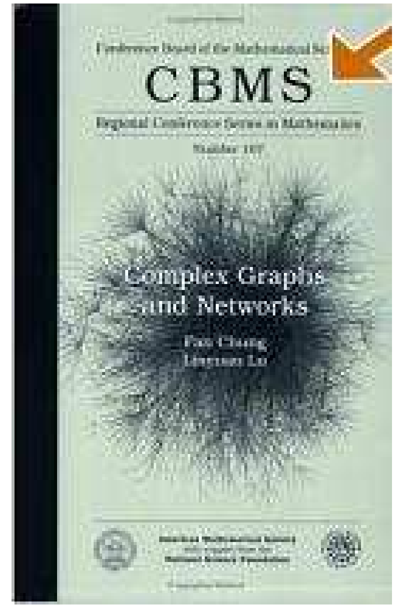
- Random graphs with (exact) degree sequence
- Geometric graphs and hybrid random graphs
- Quasi-randomness and spectral analysis
- Algorithms

References

1. Fan Chung and Linyuan Lu, Concentration inequalities and martingale inequalities — a survey, *Internet Mathematics*, **3** (2006), No. 1, 79–127.
2. Fan Chung and Linyuan Lu, The volume of the giant component for a random graph with given expected degrees, *SIAM J. Discrete Math.*, **20** (2006), No. 2, 395–411.
3. Reid Andersen, Fan Chung, and Linyuan Lu, Modeling the small-world phenomenon with local network flow, *Internet Mathematics*, **2** No. 3, (2005), 359–385.
4. Fan Chung and Linyuan Lu, Coupling on-line and off-line analyses for random power law graphs, *Internet Mathematics*, **1** No. 4, (2004), 409–461.
5. Fan Chung and Linyuan Lu, The small world phenomenon in hybrid power law graphs, *Lect. Notes Phys.* **650** (2004), 89-104.
6. Fan Chung and Linyuan Lu, The average distances in random graphs with given expected degrees, *Internet Mathematics* **1**, No. 1, (2003), 91–114.
7. Fan Chung, Linyuan Lu and Van Vu, The spectra of random graphs with given expected degrees, *Proceedings of National Academy of Sciences*, **100**, No. 11, (2003), 6313-6318.
8. Fan Chung and Linyuan Lu, T. Gregory Dewey, and David J. Galas. Duplication models for biological networks, *Journal of Computational Biology*, **10**, No. 5, (2003), 677-688.
9. Fan Chung, Linyuan Lu, and Van Vu, Eigenvalues of random power law graphs, *Annals of Combinatorics*, **7** (2003), 21–33.
10. Fan Chung and Linyuan Lu. The average distances in random graphs with given expected degrees, *Proc. Natl. Acad. Sci.* **99** (2002), 15879–15882.
11. Fan Chung and Linyuan Lu. Connected components in a random graph with given degree sequences, *Annals of Combinatorics*, **6** (2002), 125–145.

Further reading

Fan Chung and Linyuan Lu
Complex graphs and networks
CBMS Regional Conference Series in Mathematics; number 107, (2006), 264+vii pages.
ISBN-10: 0-8218-3657-9,
ISBN-13: 978-0-8218-3657-6.



http://www.math.sc.edu/∼lu/

Open questions

Any questions asked in (random) graph theory should be asked again with restriction on degree distribution.

Open questions

Any questions asked in (random) graph theory should be asked again with restriction on degree distribution.

But good questions come with motivations.

Open questions

Any questions asked in (random) graph theory should be asked again with restriction on degree distribution.

But good questions come with motivations.

Acknowledgment

- Center for Combinatorics at Nankai University
- NSF Grant DMS 0701111
- Thank you, audience!