Block 7: Network models

ELEC 573: Network Science and Analytics

Santiago Segarra

Electrical and Computer Engineering
Rice University
segarra@rice.edu

Fall 2021

You are here

Wk.	Date	Topic	HW	Project
1	23-Aug	Introduction to course	HW0 out	
2	30-Aug	Graph theory	HW0 solutions posted	
3	6-Sep	LABOR DAY (no class)	HW1 out	
4	13-Sep	Centrality measures / Community detection		
5	20-Sep	Community detection		
6	27-Sep	Signal Processing and Deep learning for graphs	HW1 due	
7	4-Oct	Signal Processing and Deep learning for graphs	HW2 out	
8	11-Oct	FALL BREAK (no class)		
9	18-Oct	Network models	HW2 due	
10	25-Oct	Network models	HW3 out	Project proposal due
11	1-Nov	Epidemics		
12	8-Nov	Inference of network topologies, features, and processes	HW3 due	
13	15-Nov	Inference of network topologies, features, and processes		
14	22-Nov	Inference of network topologies, features, and processes		Project progress report
15	29-Nov	Inference of network topologies, features, and processes		
	13-Dec	Project presentation (video recording) and final report due		

You are here

Wk.	Date	Topic	HW	Project
1	23-Aug	Introduction to course	HW0 out	
2	30-Aug	Graph theory	HW0 solutions posted	
3	6-Sep	LABOR DAY (no class)	HW1 out	
4	13-Sep	Centrality measures / Community detection		
5	20-Sep	Community detection		
6	27-Sep	Signal Processing and Deep learning for graphs	HW1 due	
7	4-Oct	Signal Processing and Deep learning for graphs	HW2 out	
8	11-Oct	FALL BREAK (no class)		
9	18-Oct	Network models	HW2 due	
10	25-Oct	Network models	HW3 out	Project proposal due
11	1-Nov	Epidemics		
12	8-Nov	Inference of network topologies, features, and processes	HW3 due	
13	15-Nov	Inference of network topologies, features, and processes		
14	22-Nov	Inference of network topologies, features, and processes		Project progress report
15	29-Nov	Inference of network topologies, features, and processes		
	13-Dec	Project presentation (video recording) and final report due		

- Comment on extra-credit in HW

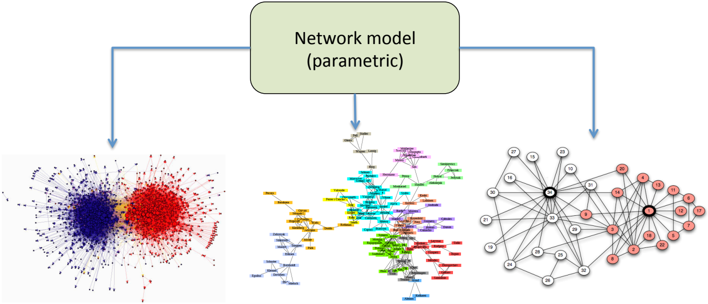
Why models for networks?

- Usual trade-off between losing details in an idealized representation while gaining insights into the simplified problem
- Simple representations of complex networks
- Derive properties mathematically
- Predict properties and outcomes
- Common features of different real networks

Why models for networks?

- Usual trade-off between losing details in an idealized representation while gaining insights into the simplified problem

- Simple representations of complex networks
- Derive properties mathematically
- Predict properties and outcomes
- Common features of different real networks



Modeling networks

- So far the focus has been on network analysis methods to:
  - $\Rightarrow$ Collect relational data and construct network graphs
  - $\Rightarrow$ Characterize and summarize their structural properties

Modeling networks

- So far the focus has been on network analysis methods to:
  - $\Rightarrow$ Collect relational data and construct network graphs
  - $\Rightarrow$ Characterize and summarize their structural properties
- Emphasis now on construction and use of models for network data

Modeling networks

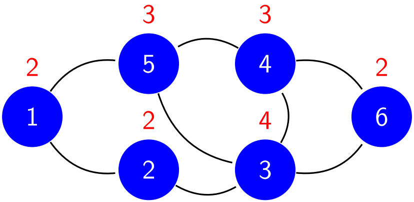
- So far the focus has been on network analysis methods to:
  - $\Rightarrow$ Collect relational data and construct network graphs
  - $\Rightarrow$ Characterize and summarize their structural properties
- Emphasis now on construction and use of models for network data
- **Def:** A model for a graph is a collection

$$\{\mathsf{P}_\theta(G), G \in \mathcal{G} : \theta \in \Theta\}$$

  - $\mathcal{G}$ is an ensemble of possible graphs
  - $\mathsf{P}_\theta(\cdot)$ is a probability distribution on $\mathcal{G}$ (often write $\mathsf{P}[\cdot]$)
  - Parameters θ ranging over values in parameter space Θ

Degree

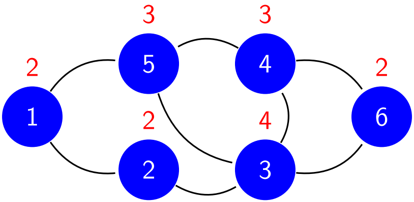
- **Def:** The degree d_v of vertex v is its number of incident edges
  $\Rightarrow$ Degree sequence arranges degrees in non-decreasing order



- In figure $\Rightarrow$ Vertex degrees shown in red, e.g., $d_1 = 2$ and $d_5 = 3$
  $\Rightarrow$ Graph's degree sequence is 2,2,2,3,3,4

Degree

- **Def:** The degree d_v of vertex v is its number of incident edges
  $\Rightarrow$ Degree sequence arranges degrees in non-decreasing order



- In figure $\Rightarrow$ Vertex degrees shown in red, e.g., $d_1 = 2$ and $d_5 = 3$
  $\Rightarrow$ Graph's degree sequence is 2,2,2,3,3,4
- In general, the degree sequence does *not uniquely* specify the graph
- High-degree vertices are likely to be influential, central, prominent

Degree distribution

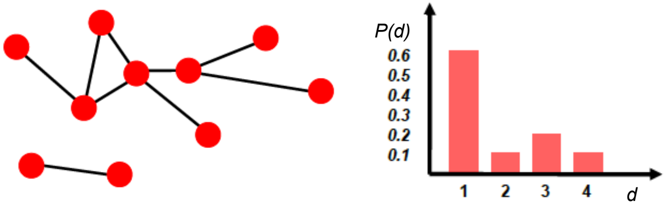
- Let $N(d)$ denote the number of vertices with degree d

  $\Rightarrow$ Fraction of vertices with degree d is $\mathrm{P}[d] := \frac{N(d)}{N_v}$

- Let $N(d)$ denote the number of vertices with degree d

  $\Rightarrow$ Fraction of vertices with degree d is $\mathrm{P}[d] := \frac{N(d)}{N_v}$

- **Def:** The collection $\{\mathrm{P}[d]\}_{d \geq 0}$ is the degree distribution of G
  - Histogram formed from the degree sequence (bins of size one)

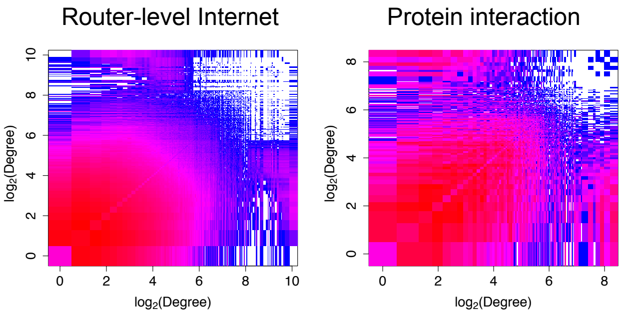


- $\mathrm{P}[d] =$ probability that randomly chosen node has degree d

  $\Rightarrow$ Summarizes the local connectivity in the network graph

Joint degree distribution

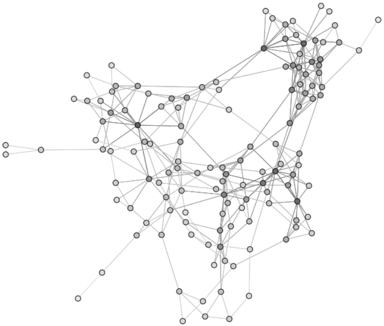
- Q: What about patterns of association among nodes of given degrees?
- A: Define the two-dimensional analogue of a degree distribution



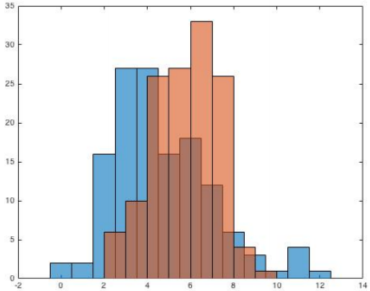
- Prob. of random edge having incident vertices with degrees (d_1, d_2)

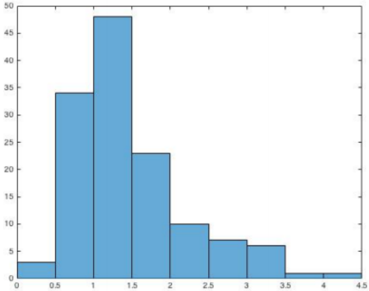
Friendship paradox

- Observed by Feld (1991)
  - ⇒ "Why Your Friends Have More Friends Than You Do"
  - ⇒ American Journal of Sociology
- Example: Network of 135 households from a rural Indian village
  - ⇒ Banerjee, Chandrasekhar, Duflo, and Jackson (2013)



Friendship paradox





- Left: empirical distribution of households' degrees (blue) and the distribution of average neighbors' degrees (red)
- Right: empirical distribution of ratio of average neighbors' degree over own degree

Friendship paradox

- Network of firm-level input-output linkages in Japan
- Carvalho, Nirei, Saito, and Tahbaz-Salehi (2016)

	Disaster Area Firms	Firms in the Rest of Japan
Log Sales	11.54	11.74
	(1.52)	(1.64)
Customers' Log Sales	14.83	14.51
	(2.37)	(2.45)
Suppliers' Log Sales	14.30	14.60
	(2.21)	(2.49)

- Firms' customers and suppliers are on average larger than the average firm

Friendship paradox

Theorem [Jackson, 2016] In any given undirected network, the average degree of neighbors is at least as high as the average degree:

$$\frac{1}{n}\sum_{i:d_i>0}\frac{\sum_{j\in N_i} d_j}{d_i} \geq \frac{1}{n}\sum_{i=1}^{n} d_i.$$

Furthermore, the inequality is strict if and only if at least two linked agents have different degrees.

Friendship paradox

Theorem [Jackson, 2016] In any given undirected network, the average degree of neighbors is at least as high as the average degree:

$$\frac{1}{n}\sum_{i:d_i>0}\frac{\sum_{j\in N_i} d_j}{d_i} \geq \frac{1}{n}\sum_{i=1}^{n} d_i.$$

Furthermore, the inequality is strict if and only if at least two linked agents have different degrees.

- In any given network,

$$\sum_{i:d_i>0}\frac{\sum_{j\in N_i} d_j}{d_i} = \sum_{i<j:i\in N_j}\left(\frac{d_i}{d_j}+\frac{d_j}{d_i}\right) = \sum_{i<j:i\in N_j}\left(\frac{(d_i-d_j)^2}{d_i d_j}+2\right).$$

- And as a result,

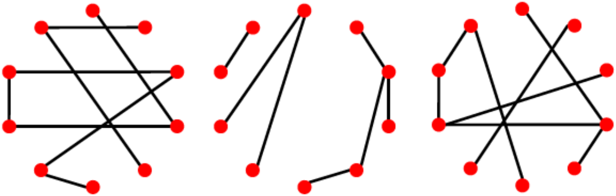
$$\sum_{i:d_i>0}\frac{\sum_{j\in N_i} d_j}{d_i} \geq \sum_{i<j:i\in N_j} 2 = \sum_{i=1}^{n} d_i.$$

A simple random graph model

- **Def:** The Erdős-Rényi random graph model $G_{n,p}$
  - Undirected graph with n vertices, i.e., of order $N_v = n$
  - Edge (u, v) present with probability p, independent of other edges

- Simulation is easy: draw $\binom{n}{2}$ i.i.d. Bernoulli(p) RVs

Example



- Three realizations of $G_{10,\frac{1}{6}}$. The size N_e is a random variable

Degree distribution of $G_{n,p}$

- Q: Degree distribution $\mathrm{P}[d]$ of the Erdős-Rényi graph $G_{n,p}$?

Degree distribution of $G_{n,p}$

- Q: Degree distribution $\mathrm{P}[d]$ of the Erdős-Rényi graph $G_{n,p}$?
- Define $\mathbb{I}\{(v,u)\} = 1$ if $(v,u) \in E$, and $\mathbb{I}\{(v,u)\} = 0$ otherwise.
  $\Rightarrow$ Fix v. For all $u \neq v$, the indicator RVs are i.i.d. Bernoulli(p)

Degree distribution of $G_{n,p}$

- Q: Degree distribution $\mathrm{P}[d]$ of the Erdős-Rényi graph $G_{n,p}$?
- Define $\mathbb{I}\{(v,u)\} = 1$ if $(v,u) \in E$, and $\mathbb{I}\{(v,u)\} = 0$ otherwise.
  $\Rightarrow$ Fix v. For all $u \neq v$, the indicator RVs are i.i.d. Bernoulli(p)
- Let D_v be the (random) degree of vertex v. Hence,

$$D_v = \sum_{u \neq v} \mathbb{I}\{(v,u)\}$$

Degree distribution of $G_{n,p}$

- Q: Degree distribution $\mathrm{P}[d]$ of the Erdős-Rényi graph $G_{n,p}$?
- Define $\mathbb{I}\{(v,u)\} = 1$ if $(v,u) \in E$, and $\mathbb{I}\{(v,u)\} = 0$ otherwise.
  $\Rightarrow$ Fix v. For all $u \neq v$, the indicator RVs are i.i.d. Bernoulli(p)
- Let D_v be the (random) degree of vertex v. Hence,

$$D_v = \sum_{u \neq v} \mathbb{I}\{(v,u)\}$$

  $\Rightarrow$ D_v is binomial with parameters $(n-1, p)$ and

$$\mathrm{P}[d] = \mathrm{P}[D_v = d] = \binom{n-1}{d} p^d (1-p)^{(n-1)-d}$$

- In words, the probability of having exactly d edges incident to v
  $\Rightarrow$ Same for all $v \in V$, by independence of the $G_{n,p}$ model

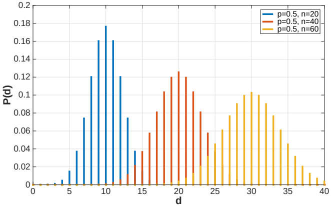
Behavior for large n

- Q: How does the degree distribution look like for a large network?

Behavior for large n

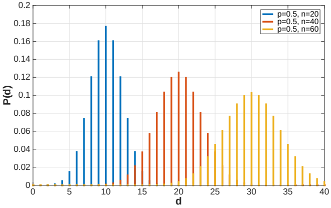
- Q: How does the degree distribution look like for a large network?
- Recall D_v is a sum of $n-1$ i.i.d. Bernoulli(p) RVs

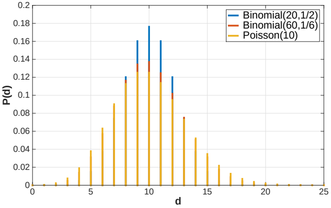
  $\Rightarrow$ Central Limit Theorem: $D_v \sim \mathcal{N}(np, np(1-p))$ for large n



Behavior for large n

- Q: How does the degree distribution look like for a large network?
- Recall D_v is a sum of $n-1$ i.i.d. Bernoulli(p) RVs
  $\Rightarrow$ Central Limit Theorem: $D_v \sim \mathcal{N}(np, np(1-p))$ for large n





- Makes most sense to increase n with fixed $\mathbb{E}[D_v] = (n-1)p = \mu$
  $\Rightarrow$ Law of rare events: $D_v \sim \text{Poisson}(\mu)$ for large n

- Substituting $p = \mu/n$ in the binomial PMF yields

$$\begin{aligned} \mathrm{P}_n(d) &= \frac{n!}{(n-d)!d!}\left(\frac{\mu}{n}\right)^d\left(1-\frac{\mu}{n}\right)^{n-d} \\ &= \frac{n(n-1)\dots(n-d+1)}{n^d}\,\frac{\mu^d}{d!}\,\frac{(1-\mu/n)^n}{(1-\mu/n)^d} \end{aligned}$$

Law of rare events

- Substituting $p = \mu/n$ in the binomial PMF yields

$$\begin{aligned} \mathrm{P}_n(d) &= \frac{n!}{(n-d)!d!}\left(\frac{\mu}{n}\right)^d\left(1-\frac{\mu}{n}\right)^{n-d} \\ &= \frac{n(n-1)\ldots(n-d+1)}{n^d}\,\frac{\mu^d}{d!}\,\frac{(1-\mu/n)^n}{(1-\mu/n)^d} \end{aligned}$$

- In the limit, red term is $\lim_{n\to\infty}(1-\mu/n)^n = e^{-\mu}$

Law of rare events

- Substituting $p = \mu/n$ in the binomial PMF yields

$$\begin{aligned} \mathrm{P}_n(d) &= \frac{n!}{(n-d)!d!}\left(\frac{\mu}{n}\right)^d\left(1-\frac{\mu}{n}\right)^{n-d} \\ &= \frac{n(n-1)\ldots(n-d+1)}{n^d}\,\frac{\mu^d}{d!}\,\frac{(1-\mu/n)^n}{(1-\mu/n)^d} \end{aligned}$$

- In the limit, red term is $\lim_{n\to\infty}(1-\mu/n)^n = e^{-\mu}$

- Black and blue terms converge to 1. Limit is the *Poisson* PMF

$$\lim_{n\to\infty}\mathrm{P}_n(d) = 1\,\frac{\mu^d}{d!}\,\frac{e^{-\mu}}{1} = e^{-\mu}\frac{\mu^d}{d!}$$

- Approximation usually called "law of rare events"
  - Individual edges happen with small probability $p = \mu/n$
  - The aggregate (degree, number of edges), though, need not be rare

Behavior of $G_{n,p}$ for increasing p

- ER graphs exhibit phase transitions

  $\Rightarrow$ Sharp transitions between behaviors as $n \to \infty$

ER connectivity theorem

- A threshold function for the connectivity of $G_{n,p(n)}$ is $p(n) = \frac{\ln(n)}{n}$
- Let $p(n) = \lambda \frac{\ln(n)}{n}$ then

  $\Rightarrow$ If $\lambda < 1 \;\Rightarrow \mathbb{P}(connected) \to 0$ as $n \to \infty$

  $\Rightarrow$ If $\lambda > 1 \;\Rightarrow \mathbb{P}(connected) \to 1$ as $n \to \infty$

Behavior of $G_{n,p}$ for increasing p

- ER graphs exhibit phase transitions

  $\Rightarrow$ Sharp transitions between behaviors as $n \to \infty$

ER connectivity theorem

- A threshold function for the connectivity of $G_{n,p(n)}$ is $p(n) = \frac{\ln(n)}{n}$
- Let $p(n) = \lambda \frac{\ln(n)}{n}$ then

  $\Rightarrow$ If $\lambda < 1 \;\Rightarrow\; \mathbb{P}(\textit{connected}) \to 0$ as $n \to \infty$

  $\Rightarrow$ If $\lambda > 1 \;\Rightarrow\; \mathbb{P}(\textit{connected}) \to 1$ as $n \to \infty$

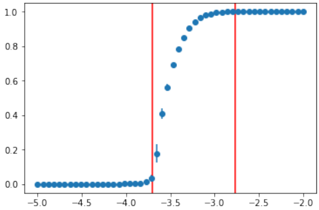
ER giant component theorem

- A threshold function for the emergence of a giant component in $G_{n,p(n)}$ is $p(n) = \frac{1}{n}$
- Let $p(n) = \frac{\lambda}{n}$ then

  $\Rightarrow$ If $\lambda < 1 \;\Rightarrow$ Size of largest component $\sim \ln(n)$ as $n \to \infty$

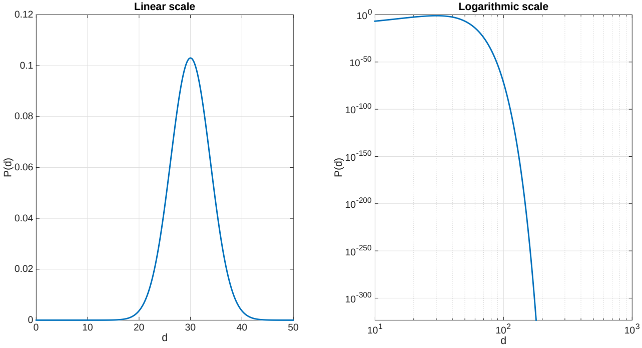
  $\Rightarrow$ If $\lambda > 1 \;\Rightarrow$ Size of largest component $\sim n$ as $n \to \infty$

- You will test this in the homework
  - $\Rightarrow$ Plot the relative size of largest component
  - $\Rightarrow$ As a function of $log(p)$



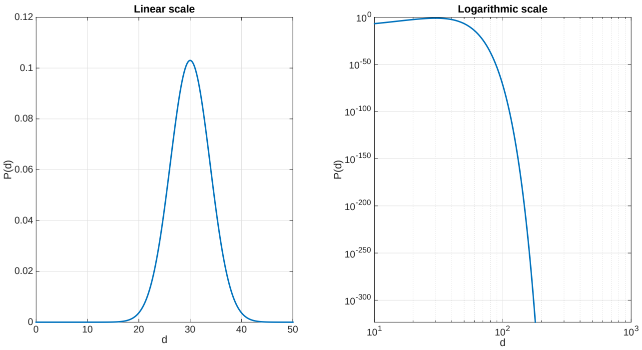
- Movie: $G_{n,p}$ for increasing p
- `https://www.youtube.com/watch?v=mpe44sTSoF8`

- For large graphs, $G_{n,p}$ suggests $\mathrm{P}[d]$ with an exponential tail
  - $\Rightarrow$ Unlikely to see degrees spanning several orders of magnitude



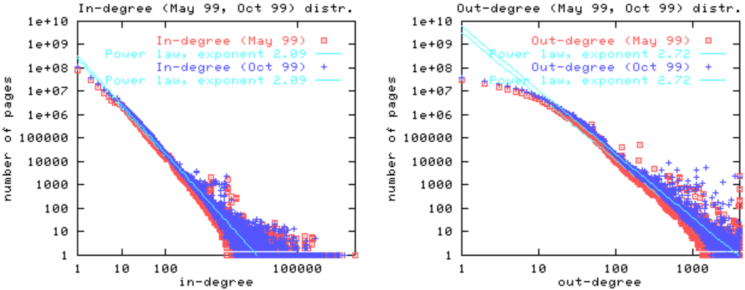
- Concentrated distribution around the mean $\mathbb{E}[D_v] = (n-1)p$

- For large graphs, $G_{n,p}$ suggests $\mathrm{P}[d]$ with an exponential tail
  - $\Rightarrow$ Unlikely to see degrees spanning several orders of magnitude



- Concentrated distribution around the mean $\mathbb{E}[D_v] = (n-1)p$
- Q: Is this in agreement with real-world networks?

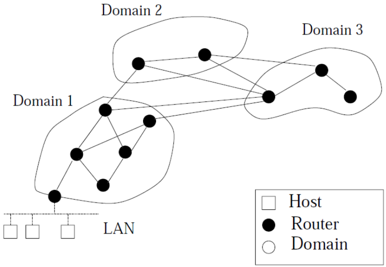
- Degree distributions of the WWW analyzed in [Broder et al '00]
  - $\Rightarrow$ Web a digraph, study both in- and out-degree distributions

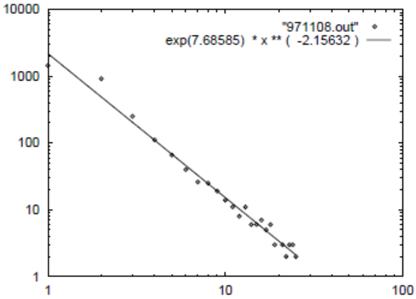


- Majority of vertices naturally have small degrees
  - $\Rightarrow$ Nontrivial amount with orders of magnitude higher degrees

Internet autonomous systems

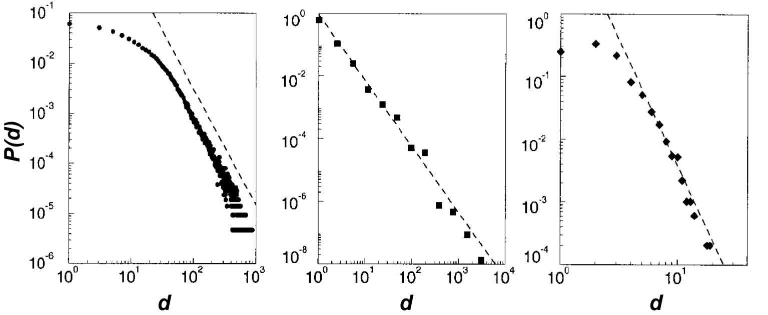
- The topology of the AS-level Internet studied in [Faloutsos[3] '99]





- Right-skewed degree distributions also found for router-level Internet

- More heavy-tailed degree distributions found in [Barabasi-Albert '99]

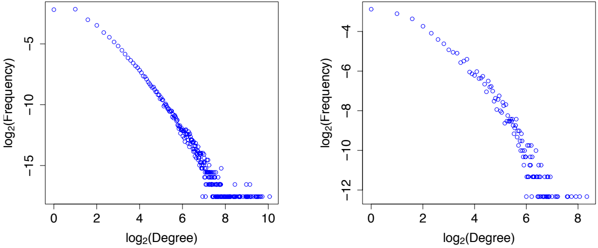


Author collaboration

Web graph

Power grid

Power-law degree distributions



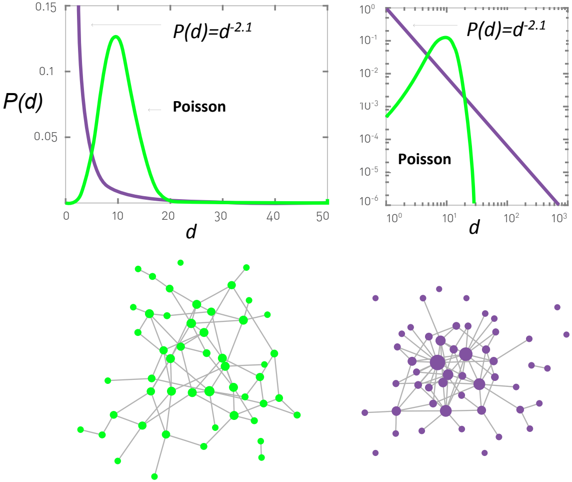
- Log-log plots show roughly a linear decay, suggesting the power law

$$\mathrm{P}[d] \propto d^{-\alpha} \Rightarrow \log \mathrm{P}[d] = C - \alpha \log d$$

  - Power-law exponent (negative slope) is typically $\alpha \in [2, 3]$
  - Normalization constant C is mostly uninteresting

- Power laws often best followed in the tail, i.e., for $d \geq d_{\text{min}}$

Power law and exponential degree distributions



- Erdős-Rényi's Poisson degree distribution exhibits a sharp cutoff

  $\Rightarrow$ Power laws upper bound exponential tails for large enough d

Scale-free networks

- Scale-free network: degree distribution with power-law tail
  - Name motivated for the scale-invariance property of power laws

Scale-free networks

- Scale-free network: degree distribution with power-law tail
  - Name motivated for the scale-invariance property of power laws
- **Def:** A scale-free function $f(x)$ satisfies $f(ax) = bf(x)$, for $a, b \in \mathbb{R}$

Example

- Power-law functions $f(x) = x^{-\alpha}$ are scale-free since

$$f(ax) = (ax)^{-\alpha} = a^{-\alpha} f(x) = bf(x), \text{ where } b := a^{-\alpha}$$

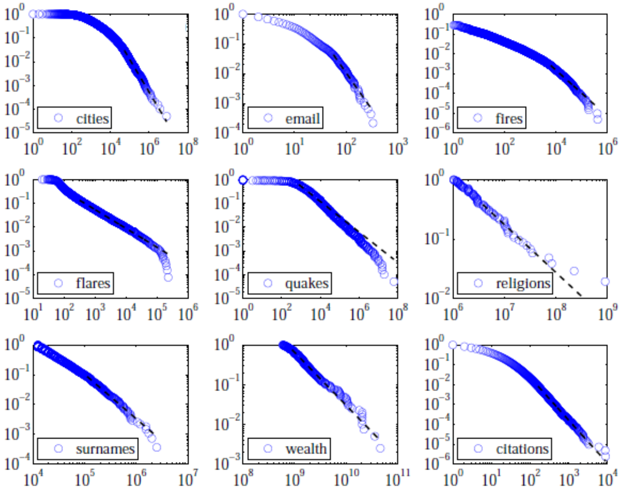
- Exponential functions $f(x) = c^x$ are not scale-free because

$$f(ax) = c^{ax} = (c^x)^a = (f(x))^a \neq bf(x), \text{ except when } a = b = 1$$

- No 'characteristic scale' for the degrees. More soon

$\Rightarrow$ Functional form of the distribution is invariant to scale

Power-law distributions are ubiquitous



- Power-law distributions widespread beyond networks [Clauset et al '07]

Normalization

- The power-law degree distribution $\mathrm{P}[d] = Cd^{-\alpha}$ is a PMF, hence

$$1 = \sum_{d=0}^{\infty} \mathrm{P}[d] = \sum_{d=0}^{\infty} Cd^{-\alpha} \Rightarrow C = \frac{1}{\sum_{d=0}^{\infty} d^{-\alpha}}$$

Normalization

- The power-law degree distribution $\mathrm{P}[d] = Cd^{-\alpha}$ is a PMF, hence

$$1 = \sum_{d=0}^{\infty} \mathrm{P}[d] = \sum_{d=0}^{\infty} Cd^{-\alpha} \Rightarrow C = \frac{1}{\sum_{d=0}^{\infty} d^{-\alpha}}$$

- Often a power law is only valid for the tail $d \geq d_{\text{min}}$, hence

$$C = \frac{1}{\sum_{d=d_{\text{min}}}^{\infty} d^{-\alpha}} \approx \frac{1}{\int_{d_{\text{min}}}^{\infty} x^{-\alpha} dx} = (\alpha - 1) d_{\text{min}}^{\alpha-1}$$

$\Rightarrow$ Sound approximation since $\mathrm{P}[d]$ varies slowly for large d

- The normalized power-law degree distribution is

$$\mathrm{P}[d] = \frac{\alpha - 1}{d_{\text{min}}} \left(\frac{d}{d_{\text{min}}}\right)^{-\alpha}, \quad d \geq d_{\text{min}}$$

Power-law probability density function

- Often convenient to treat degrees as real valued, i.e., $d \in \mathbb{R}_+$
- Define a power-law PDF for the tail of the degree distribution as

$$p(d) = \frac{\alpha - 1}{d_{\min}} \left(\frac{d}{d_{\min}} \right)^{-\alpha}, \quad d \geq d_{\min}$$

$\Rightarrow$ A valid PDF, already showed that $\int_{d_{\min}}^{\infty} p(x)dx = 1$

$\Rightarrow$ Convergence of the integral requires $\alpha > 1$

Power-law probability density function

- Often convenient to treat degrees as real valued, i.e., $d \in \mathbb{R}_+$
- Define a power-law PDF for the tail of the degree distribution as

$$p(d) = \frac{\alpha - 1}{d_{\min}} \left(\frac{d}{d_{\min}} \right)^{-\alpha}, \quad d \geq d_{\min}$$

$\Rightarrow$ A valid PDF, already showed that $\int_{d_{\min}}^{\infty} p(x)dx = 1$

$\Rightarrow$ Convergence of the integral requires $\alpha > 1$

- Ex: Probability that a random node has degree exceeding 100 is

$$\mathrm{P}\left[D_v > 100\right] = \int_{100}^{\infty} \frac{\alpha - 1}{d_{\min}} \left(\frac{x}{d_{\min}} \right)^{-\alpha} dx = \left(\frac{100}{d_{\min}} \right)^{1-\alpha}$$

Divergent moments

- Q: What is the m-th moment of a power-law distributed RV?
- From the definition of moment and the power-law PDF one has

$$\mathbb{E}\left[D_v^m\right] = \int_{d_{\min}}^{\infty} x^m p(x) dx = \frac{\alpha - 1}{d_{\min}^{1-\alpha}} \left[\frac{x^{m+1-\alpha}}{m+1-\alpha}\right]_{d_{\min}}^{\infty}$$

$\Rightarrow$ Convergence of the integral requires $m + 1 < \alpha$

Divergent moments

- Q: What is the m-th moment of a power-law distributed RV?
- From the definition of moment and the power-law PDF one has

$$\mathbb{E}\left[D_v^m\right] = \int_{d_{\min}}^{\infty} x^m p(x) dx = \frac{\alpha - 1}{d_{\min}^{1-\alpha}} \left[\frac{x^{m+1-\alpha}}{m+1-\alpha}\right]_{d_{\min}}^{\infty}$$

$\Rightarrow$ Convergence of the integral requires $m + 1 < \alpha$

- For real-world networks, typically $\alpha \in (2, 3)$ so

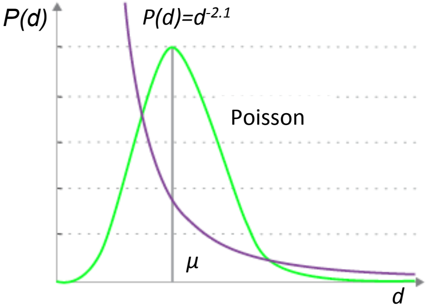
$$\mathbb{E}\left[D_v\right] = \left(\frac{\alpha - 1}{\alpha - 2}\right) d_{\min} < \infty \quad \text{and} \quad \mathbb{E}\left[D_v^m\right] = \infty, \; m \geq 2$$

- In particular, the second moment and variance are infinite

$\Rightarrow$ Consistent with variability and heterogeneity of degrees

Revisiting the scale-free property

- A measure of scale of a RV is its standard deviation σ



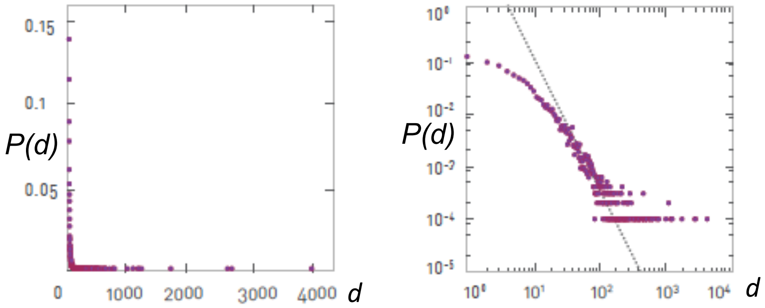
Large random network $G_{n,p}$

- Randomly chosen node has degree $d = \mu \pm \sqrt{\mu}$. The scale is μ

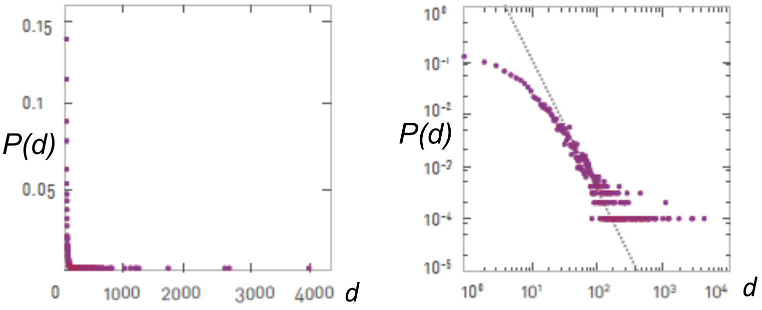
Scale-free network

- Randomly chosen node has degree $d = \mu \pm \infty$. There is no scale

- A simple histogram may be problematic for visualizing $\mathrm{P}[d]$
  - $\Rightarrow$ Use log-log scale to warp probabilities and widespread degrees



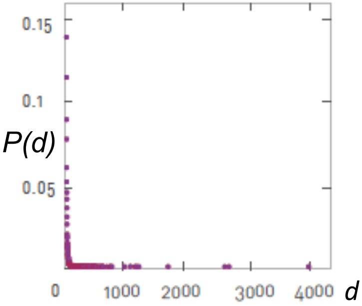
- A simple histogram may be problematic for visualizing $\mathrm{P}[d]$
  - $\Rightarrow$ Use log-log scale to warp probabilities and widespread degrees

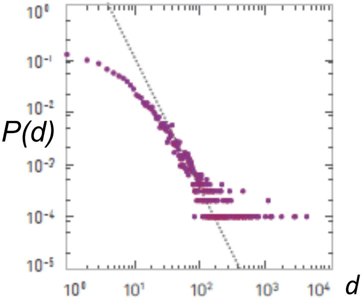


- Large statistical fluctuations ('noise') in the tail for large d
  - $\Rightarrow$ With bins of size one, high-degree counts are small

Visualizing power-law degree distributions

- A simple histogram may be problematic for visualizing $\mathrm{P}[d]$
  - $\Rightarrow$ Use log-log scale to warp probabilities and widespread degrees





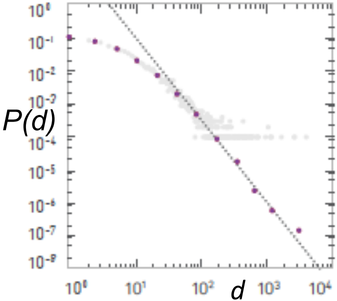
- Large statistical fluctuations ('noise') in the tail for large d
  - $\Rightarrow$ With bins of size one, high-degree counts are small
  - $\Rightarrow$ Makes sense to increase the bin size

Logarithmic binning

- Uniformly widening bins sacrifices resolution for small degrees
  $\Rightarrow$ Use bins of different sizes in different parts of the histogram

Logarithmic binning

- Uniformly widening bins sacrifices resolution for small degrees
  - $\Rightarrow$ Use bins of different sizes in different parts of the histogram



- Logarithmic binning is widely used. The n-th bin is

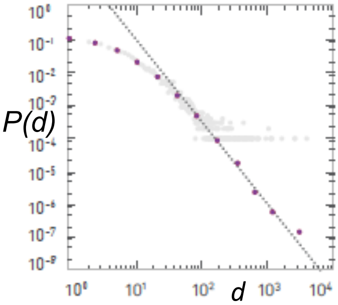
$$a^{n-1} \leq d < a^n, \quad n = 1, 2, \ldots$$

Ex: Common choice is $a = 2$, n-th bin has width $2^n - 2^{n-1} = 2^{n-1}$

Logarithmic binning

- Uniformly widening bins sacrifices resolution for small degrees

  $\Rightarrow$ Use bins of different sizes in different parts of the histogram



- Logarithmic binning is widely used. The n-th bin is

$$a^{n-1} \leq d < a^n, \quad n = 1, 2, \ldots$$

Ex: Common choice is $a = 2$, n-th bin has width $2^n - 2^{n-1} = 2^{n-1}$

- Normalize by the bin width. Wider bins will accrue higher counts

- **Def:** The complementary cumulative distribution function (CCDF) is

$$\bar{F}(d) = \mathrm{P}\left[D_v \geq d\right]$$

$\Rightarrow$ Function $\bar{F}(d)$ is the fraction of vertices with degree at least d

Complementary cumulative distribution function

- **Def:** The complementary cumulative distribution function (CCDF) is

$$\bar{F}(d) = \mathrm{P}\left[D_v \geq d\right]$$

$\Rightarrow$ Function $\bar{F}(d)$ is the fraction of vertices with degree at least d

- For a power-law PDF, the CCDF also obeys a power law since

$$\mathrm{P}\left[D_v \geq d\right] = \int_d^\infty \frac{\alpha - 1}{d_{\min}} \left(\frac{x}{d_{\min}}\right)^{-\alpha} dx = \left(\frac{d}{d_{\min}}\right)^{-(\alpha-1)}$$

Complementary cumulative distribution function

- **Def:** The complementary cumulative distribution function (CCDF) is

$$\bar{F}(d) = \mathrm{P}\left[D_v \geq d\right]$$

$\Rightarrow$ Function $\bar{F}(d)$ is the fraction of vertices with degree at least d

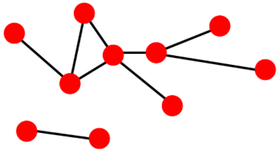
- For a power-law PDF, the CCDF also obeys a power law since

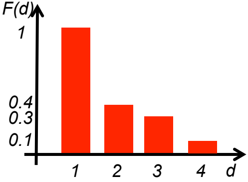
$$\mathrm{P}\left[D_v \geq d\right] = \int_d^\infty \frac{\alpha - 1}{d_{\min}} \left(\frac{x}{d_{\min}}\right)^{-\alpha} dx = \left(\frac{d}{d_{\min}}\right)^{-(\alpha-1)}$$

- If the PDF has exponent α, then CCDF $\bar{F}(d)$ has exponent $\alpha - 1$

Computing the CCDF

Step 1: List the degrees d_v in descending order
Step 2: Assign ranks r_v (from 1 to N_v) to vertices in that order
Step 3: The CCDF is the plot of r_v/N_v versus degree d_v



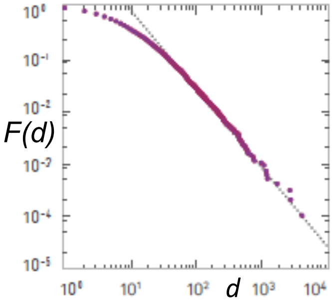


d_v	r_v	r_v/N_v	$F(d)$
4	1	0.1	0.1
3	2	0.2	
3	3	0.3	0.3
2	4	0.4	0.4
1	5	0.5	
1	6	0.6	
1	7	0.7	
1	8	0.8	
1	9	0.9	
1	10	1.0	1.0

- If degrees are repeated, CCDF is the largest value of r_v/N_v
- If d not observed, $\bar{F}(d)$ = value for next (larger) observed degree

Visualizing power laws with the CCDF

- Plot the CCDF in a log-log scale and look for a straight-line behavior



- Mitigates noise using cumulative frequencies (cf. raw frequencies)
- No binning needed $\Rightarrow$ Avoids information loss as bins widen

Fitting power-law distributions

- Basic, yet nontrivial task is to estimate the exponent α from data
- A power law implies the linear model $\log \mathrm{P}[d] = C - \alpha \log d + \epsilon$

  $\Rightarrow$ Natural to form the linear least-squares (LS) estimator

$$\{\hat{\alpha}, \hat{C}\} = \arg\min_{\alpha, C} \sum_i \left(\log \mathrm{P}[d_i] - C + \alpha \log d_i\right)^2$$

Fitting power-law distributions

- Basic, yet nontrivial task is to estimate the exponent α from data
- A power law implies the linear model $\log \mathrm{P}[d] = C - \alpha \log d + \epsilon$

  $\Rightarrow$ Natural to form the linear least-squares (LS) estimator

$$\{\hat{\alpha}, \hat{C}\} = \arg\min_{\alpha, C} \sum_i \left(\log \mathrm{P}[d_i] - C + \alpha \log d_i\right)^2$$

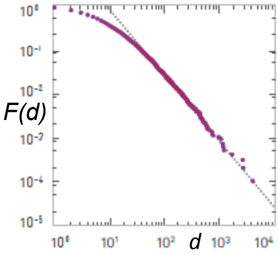
- Simple, very popular, but not advisable for at least three reasons:
  1) Extremely noisy high-degree data, where the counts are the lowest
  2) Estimates are biased. The log transform distorts unevenly the errors
  3) If the power law is only valid in the tail, need to hand pick $d_{\min}$

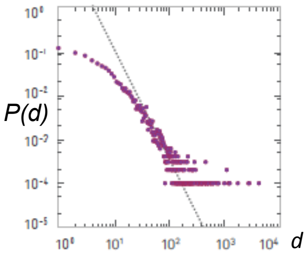
- A solution to the noise problem is to use the CCDF $\bar{F}(d)$

  $\Rightarrow$ Cumulative frequencies smoothen the noise

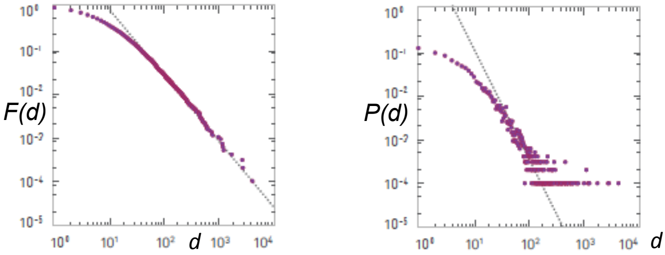
- Recall the CCDF follows a power law with exponent $\alpha - 1$

  $\Rightarrow$ Can use a linear regression-based approach to find $\hat{\alpha}$, but ...





- A solution to the noise problem is to use the CCDF $\bar{F}(d)$
  - $\Rightarrow$ Cumulative frequencies smoothen the noise
- Recall the CCDF follows a power law with exponent $\alpha - 1$
  - $\Rightarrow$ Can use a linear regression-based approach to find $\hat{\alpha}$, but ...



- Successive points in the CCDF plot are not mutually independent
  - $\Rightarrow$ (Ordinary) LS is not optimal for correlated errors

Maximum-likelihood estimator

- Suppose $\{d_i\}_{i=1}^{N_v}$ are independent and follow a power law. MLE of α?

  $\Rightarrow$ The data PDF is $f(d;\alpha) = \frac{\alpha-1}{d_{\min}}\left(\frac{d}{d_{\min}}\right)^{-\alpha}$, $d \geq d_{\min}$

Maximum-likelihood estimator

- Suppose $\{d_i\}_{i=1}^{N_v}$ are independent and follow a power law. MLE of α?

  $\Rightarrow$ The data PDF is $f(d;\alpha) = \frac{\alpha-1}{d_{\min}}\left(\frac{d}{d_{\min}}\right)^{-\alpha}$, $d \geq d_{\min}$

- The log-likelihood function is (up to constants independent of α)

$$\ell_{N_v}(\alpha) = \sum_{i=1}^{N_v} \log f(d_i;\alpha) \propto N_v \log(\alpha - 1) - \alpha \sum_{i=1}^{N_v} \log\left(\frac{d_i}{d_{\min}}\right)$$

Maximum-likelihood estimator

- Suppose $\{d_i\}_{i=1}^{N_v}$ are independent and follow a power law. MLE of α?

  $\Rightarrow$ The data PDF is $f(d;\alpha) = \frac{\alpha-1}{d_{\min}}\left(\frac{d}{d_{\min}}\right)^{-\alpha}$, $d \geq d_{\min}$

- The log-likelihood function is (up to constants independent of α)

$$\ell_{N_v}(\alpha) = \sum_{i=1}^{N_v} \log f(d_i;\alpha) \propto N_v \log(\alpha - 1) - \alpha \sum_{i=1}^{N_v} \log\left(\frac{d_i}{d_{\min}}\right)$$

- The MLE $\hat{\alpha}$ (a.k.a. Hill estimator) solves the equation

$$\left.\frac{\partial \ell_{N_v}(\alpha)}{\partial \alpha}\right|_{\alpha=\hat{\alpha}} = \frac{N_v}{\hat{\alpha}-1} - \sum_{i=1}^{N_v} \log\left(\frac{d_i}{d_{\min}}\right) = 0$$

Maximum-likelihood estimator

- Suppose $\{d_i\}_{i=1}^{N_v}$ are independent and follow a power law. MLE of α?

  $\Rightarrow$ The data PDF is $f(d;\alpha) = \frac{\alpha-1}{d_{\min}}\left(\frac{d}{d_{\min}}\right)^{-\alpha}$, $d \geq d_{\min}$

- The log-likelihood function is (up to constants independent of α)

$$\ell_{N_v}(\alpha) = \sum_{i=1}^{N_v} \log f(d_i;\alpha) \propto N_v \log(\alpha - 1) - \alpha \sum_{i=1}^{N_v} \log\left(\frac{d_i}{d_{\min}}\right)$$

- The MLE $\hat{\alpha}$ (a.k.a. Hill estimator) solves the equation

$$\left.\frac{\partial \ell_{N_v}(\alpha)}{\partial \alpha}\right|_{\alpha=\hat{\alpha}} = \frac{N_v}{\hat{\alpha}-1} - \sum_{i=1}^{N_v} \log\left(\frac{d_i}{d_{\min}}\right) = 0$$

- The solution is

$$\hat{\alpha} = 1 + \left[\frac{1}{N_v}\sum_{i=1}^{N_v} \log\left(\frac{d_i}{d_{\min}}\right)\right]^{-1}$$

- Q: How can we go around hand-picking the value of $d_{\min}$?

1) Rank-order degrees to obtain the sequence $d_{(1)} \leq \ldots \leq d_{(N_v)}$

Hill plot of ML estimates

- Q: How can we go around hand-picking the value of d_{min}?

1) Rank-order degrees to obtain the sequence $d_{(1)} \leq \ldots \leq d_{(N_v)}$
2) For each $k \in \{1, \ldots, N_v - 1\}$ let $d_{\text{min}} = d_{(N_v - k)}$. The MLEs are

$$\hat{\alpha}(k) = 1 + \left[\frac{1}{k} \sum_{i=0}^{k-1} \log \left(\frac{d_{(N_v - i)}}{d_{(N_v - k)}} \right) \right]^{-1}$$

Hill plot of ML estimates

- Q: How can we go around hand-picking the value of $d_{\min}$?

1) Rank-order degrees to obtain the sequence $d_{(1)} \leq \ldots \leq d_{(N_v)}$
2) For each $k \in \{1, \ldots, N_v - 1\}$ let $d_{\min} = d_{(N_v - k)}$. The MLEs are

$$\hat{\alpha}(k) = 1 + \left[\frac{1}{k} \sum_{i=0}^{k-1} \log \left(\frac{d_{(N_v - i)}}{d_{(N_v - k)}} \right) \right]^{-1}$$

3) Draw and examine the Hill plot of $\hat{\alpha}(k)$ versus k

- If a power law is credible, the Hill plot should 'settle down'

  $\Rightarrow$ Identify stable $\hat{\alpha}$ for a wide range of (intermediate) k values

Hill plot of ML estimates

- Q: How can we go around hand-picking the value of $d_{\min}$?

1) Rank-order degrees to obtain the sequence $d_{(1)} \leq \ldots \leq d_{(N_v)}$
2) For each $k \in \{1, \ldots, N_v - 1\}$ let $d_{\min} = d_{(N_v - k)}$. The MLEs are

$$\hat{\alpha}(k) = 1 + \left[\frac{1}{k} \sum_{i=0}^{k-1} \log \left(\frac{d_{(N_v - i)}}{d_{(N_v - k)}} \right) \right]^{-1}$$

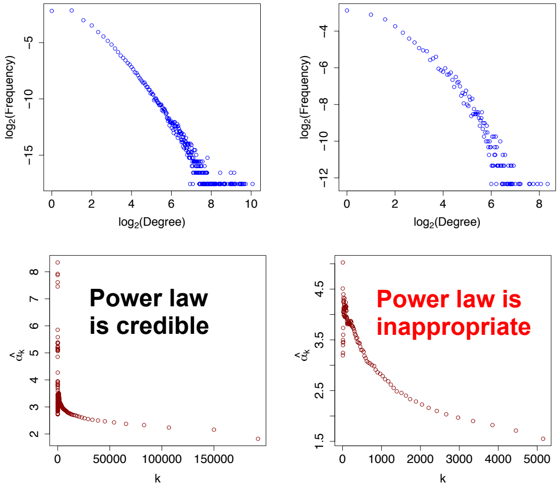
3) Draw and examine the Hill plot of $\hat{\alpha}(k)$ versus k

- If a power law is credible, the Hill plot should 'settle down'

  $\Rightarrow$ Identify stable $\hat{\alpha}$ for a wide range of (intermediate) k values

- Q: Why focus on values on the intermediate range?
  - Small k: Inaccurate estimation due to limited data
  - Large k: Bias if power law is only valid in the tail

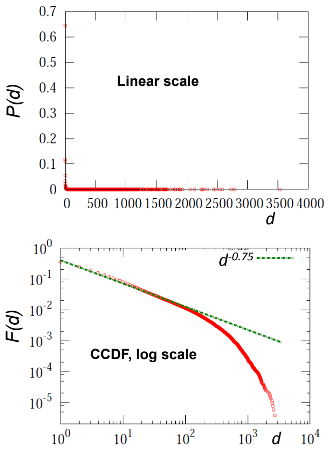
Example: Internet and protein interaction data



- Sharp decay in $\hat{\alpha}$ suggests a simple power-law model is inappropriate

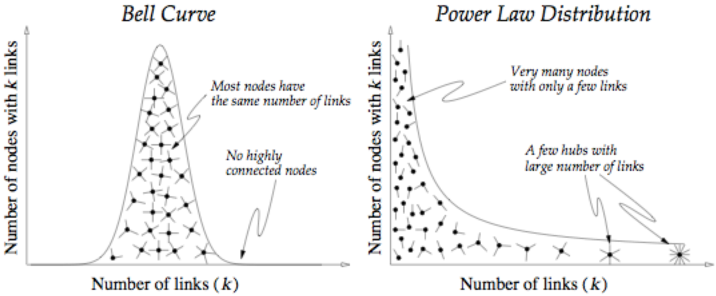
Example: Flickr data

- Flickr social network: $N_v \approx 0.6M$, $N_e \approx 3.5M$ [Leskovec et al '08]



- Good fit to a power law with exponential cutoff $\bar{F}(d) \propto d^{-\alpha}e^{-\beta d}$

- Popularity is a phenomenon characterized by extreme imbalances
  - How can we quantify these imbalances? Why do they arise?



- Basic models of network behavior can be very insightful

  $\Rightarrow$ Result of coupled decisions, correlated behavior in a population

Preferential attachment model

- Simple model for the creation of e.g., links among Web pages
- Vertices are created one at a time, denoted $1, \ldots, N_v$

Preferential attachment model

- Simple model for the creation of e.g., links among Web pages
- Vertices are created one at a time, denoted $1, \ldots, N_v$
- When node j is created, it makes a single arc to i, $1 \leq i < j$

Preferential attachment model

- Simple model for the creation of e.g., links among Web pages
- Vertices are created one at a time, denoted $1, \ldots, N_v$
- When node j is created, it makes a single arc to i, $1 \leq i < j$
- Creation of (j, i) governed by a probabilistic rule:
  - With probability p, j links to i chosen uniformly at random
  - With probability $1 - p$, j links to i with probability $\propto d_i^{in}$

Preferential attachment model

- Simple model for the creation of e.g., links among Web pages

- Vertices are created one at a time, denoted $1, \ldots, N_v$
- When node j is created, it makes a single arc to i, $1 \leq i < j$
- Creation of (j, i) governed by a probabilistic rule:
  - With probability p, j links to i chosen uniformly at random
  - With probability $1 - p$, j links to i with probability $\propto d_i^{in}$

- The resulting graph is directed, each vertex has $d_v^{out} = 1$

Preferential attachment model

- Simple model for the creation of e.g., links among Web pages
- Vertices are created one at a time, denoted $1, \ldots, N_v$
- When node j is created, it makes a single arc to i, $1 \leq i < j$
- Creation of (j, i) governed by a probabilistic rule:
  - With probability p, j links to i chosen uniformly at random
  - With probability $1 - p$, j links to i with probability $\propto d_i^{in}$
- The resulting graph is directed, each vertex has $d_v^{out} = 1$
- Preferential attachment model leads to "rich-gets-richer" dynamics

  $\Rightarrow$ Arcs formed preferentially to (currently) most popular nodes

  $\Rightarrow$ Prob. that i increases its popularity $\propto$ i's current popularity

Preferential attachment yields power laws

Theorem

The preferential attachment model gives rise to a power-law in-degree distribution with exponent $\alpha = 1 + \frac{1}{1-p}$, *i.e.,*

$$\mathrm{P}\left[d^{in} = d\right] \propto d^{-\left(1+\frac{1}{1-p}\right)}$$

Preferential attachment yields power laws

Theorem

The preferential attachment model gives rise to a power-law in-degree distribution with exponent $\alpha = 1 + \frac{1}{1-p}$, *i.e.,*

$$\mathrm{P}\left[d^{in} = d\right] \propto d^{-\left(1+\frac{1}{1-p}\right)}$$

- Key: "j links to i with probability $\propto d_i^{in}$" equivalent to copying, i.e., "j chooses k uniformly at random, and links to i if $(k, i) \in E$"
- Reflect: Copy other's decision vs. independent decisions in $G_{n,p}$

Preferential attachment yields power laws

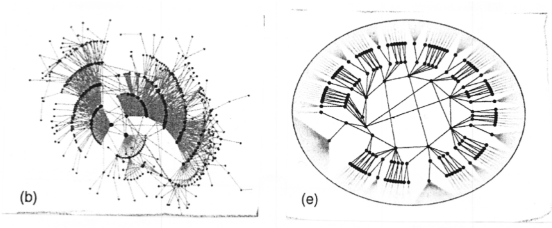
Theorem

The preferential attachment model gives rise to a power-law in-degree distribution with exponent $\alpha = 1 + \frac{1}{1-p}$, i.e.,

$$\mathrm{P}\left[d^{in} = d\right] \propto d^{-\left(1+\frac{1}{1-p}\right)}$$

- Key: "j links to i with probability $\propto d_i^{in}$" equivalent to copying, i.e., "j chooses k uniformly at random, and links to i if $(k, i) \in E$"
- Reflect: Copy other's decision vs. independent decisions in $G_{n,p}$
- As $p \to 0$ $\Rightarrow$ Copying more frequent $\Rightarrow$ Smaller $\alpha \to 2$
  - Intuitive: more likely to see extremely popular pages (heavier tail)

Network of the week



- The network of routers connecting the internet is scale-free (with caveats)
  - ⇒ The degree distribution follows a power law
- Potentially, by attacking popular nodes we can disconnect the internet?
  - ⇒ NO (fortunately)
- Preferential attachment implies power-law degree distribution
- However, the converse is NOT true! [Chiang 12]
- Power law can arise from constrained optimization of network performance

Network of the week

- Power laws and preferential attachment are big deals
  - $\Rightarrow$ "Emergence of Scaling in Random Networks" (over 40k cits.)
- Their emergence in real networks is still a debate
- *"We would even squint at the computer screen from an angle to get a better idea if a curve was straight or not" - Petter Holme*
- *"There must be a thousand papers, in which people plot the degree distribution, put a line through it and say it's scale-free without really doing the careful statistical work" - Aaron Clauset*

Classical random graph models

- Assign equal probability on all undirected graphs of given order and size

- Assign equal probability on all undirected graphs of given order and size
  - Specify collection $\mathcal{G}_{N_v,N_e}$ of graphs $G(V,E)$ with $|V| = N_v$, $|E| = N_e$
  - Assign $\mathrm{P}[G] = \binom{N}{N_e}^{-1}$ to each $G \in \mathcal{G}_{N_v,N_e}$, where $N = |V^{(2)}| = \binom{N_v}{2}$

Classical random graph models

- Assign equal probability on all undirected graphs of given order and size
  - Specify collection $\mathcal{G}_{N_v,N_e}$ of graphs $G(V,E)$ with $|V| = N_v$, $|E| = N_e$
  - Assign $\mathrm{P}[G] = \binom{N}{N_e}^{-1}$ to each $G \in \mathcal{G}_{N_v,N_e}$, where $N = |V^{(2)}| = \binom{N_v}{2}$
- Most common variant is the Erdős-Rényi random graph model $G_{n,p}$

  $\Rightarrow$ Undirected graph on $N_v = n$ vertices

  $\Rightarrow$ Edge (u,v) present w.p. p, independent of other edges

Classical random graph models

- Assign equal probability on all undirected graphs of given order and size
  - Specify collection $\mathcal{G}_{N_v,N_e}$ of graphs $G(V,E)$ with $|V| = N_v$, $|E| = N_e$
  - Assign $\mathrm{P}[G] = \binom{N}{N_e}^{-1}$ to each $G \in \mathcal{G}_{N_v,N_e}$, where $N = |V^{(2)}| = \binom{N_v}{2}$
- Most common variant is the Erdős-Rényi random graph model $G_{n,p}$

  $\Rightarrow$ Undirected graph on $N_v = n$ vertices

  $\Rightarrow$ Edge (u,v) present w.p. p, independent of other edges
- Simulation: simply draw $N = \binom{N_v}{2} \approx N_v^2/2$ i.i.d. $\mathrm{Ber}(p)$ RVs
  - Inefficient when $p \sim N_v^{-1} \Rightarrow$ sparse graph, most draws are 0
  - Skip non-edges drawing $\mathrm{Geo}(p)$ i.i.d. RVs, runs in $O(N_v + N_e)$ time

Properties of $G_{n,p}$

- $G_{n,p}$ is well-studied and tractable. Noteworthy properties:

Properties of $G_{n,p}$

- $G_{n,p}$ is well-studied and tractable. Noteworthy properties:

P1) Degree distribution $\mathrm{P}[d]$ is binomial with parameters $(n-1, p)$

- Large graphs have concentrated $\mathrm{P}[d]$ with exponentially-decaying tails

Properties of $G_{n,p}$

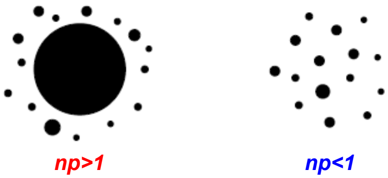
- $G_{n,p}$ is well-studied and tractable. Noteworthy properties:

P1) Degree distribution $\mathrm{P}[d]$ is binomial with parameters $(n-1, p)$

- Large graphs have concentrated $\mathrm{P}[d]$ with exponentially-decaying tails

P2) Phase transition on the emergence of a giant component

- If $np > 1$, $G_{n,p}$ has a giant component of size $O(n)$ w.h.p.
- If $np < 1$, $G_{n,p}$ has components of size only $O(\log n)$ w.h.p.



Properties of $G_{n,p}$

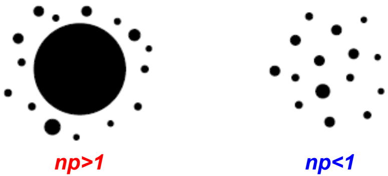
- $G_{n,p}$ is well-studied and tractable. Noteworthy properties:

P1) Degree distribution $\mathrm{P}[d]$ is binomial with parameters $(n-1, p)$

- Large graphs have concentrated $\mathrm{P}[d]$ with exponentially-decaying tails

P2) Phase transition on the emergence of a giant component

- If $np > 1$, $G_{n,p}$ has a giant component of size $O(n)$ w.h.p.
- If $np < 1$, $G_{n,p}$ has components of size only $O(\log n)$ w.h.p.



P3) Small clustering coefficient $O(n^{-1})$ and short diameter $O(\log n)$ w.h.p.

Generalized random graph models

- Recipe for generalization of Erdős-Rényi models
  - $\Rightarrow$ Specify $\mathcal{G}$ of fixed order N_v, possessing a desired characteristic
  - $\Rightarrow$ Assign equal probability to each graph $G \in \mathcal{G}$

Generalized random graph models

- Recipe for generalization of Erdős-Rényi models
  - $\Rightarrow$ Specify $\mathcal{G}$ of fixed order N_v, possessing a desired characteristic
  - $\Rightarrow$ Assign equal probability to each graph $G \in \mathcal{G}$
- Configuration model: fixed degree sequence $\{d_{(1)}, \ldots, d_{(N_v)}\}$
  - Size fixed under this model, since $N_e = \bar{d} N_v / 2 \Rightarrow \mathcal{G} \subset \mathcal{G}_{N_v, N_e}$
  - Equivalent to specifying model via conditional distribution on $\mathcal{G}_{N_v, N_e}$

Generalized random graph models

- Recipe for generalization of Erdős-Rényi models
  - $\Rightarrow$ Specify $\mathcal{G}$ of fixed order N_v, possessing a desired characteristic
  - $\Rightarrow$ Assign equal probability to each graph $G \in \mathcal{G}$
- Configuration model: fixed degree sequence $\{d_{(1)}, \ldots, d_{(N_v)}\}$
  - Size fixed under this model, since $N_e = \bar{d}N_v/2 \Rightarrow \mathcal{G} \subset \mathcal{G}_{N_v,N_e}$
  - Equivalent to specifying model via conditional distribution on $\mathcal{G}_{N_v,N_e}$
- Configuration models useful as reference, i.e., 'null' models

  Ex: compare observed G with $G' \in \mathcal{G}$ having power law $\mathrm{P}[d]$

  Ex: expected group-wise edge counts in modularity measure

P1) Phase transition on the emergence of a giant component

- Condition depends on first two moments of given $\mathrm{P}[d]$
- Giant component has size $O(N_v)$ as in $G_{N_v,p}$

- M. Molloy and B. Reed, "A critical point for random graphs with a given degree sequence," *Random Struct. and Alg.*, vol. 6, pp. 161-180, 1995

Results on the configuration model

P1) Phase transition on the emergence of a giant component

- Condition depends on first two moments of given $\mathrm{P}[d]$
- Giant component has size $O(N_v)$ as in $G_{N_v,p}$

- M. Molloy and B. Reed, "A critical point for random graphs with a given degree sequence," *Random Struct. and Alg.*, vol. 6, pp. 161-180, 1995

P2) Clustering coefficient vanishes slower than in $G_{N_v,p}$

- M. Newman et al, "Random graphs with arbitrary degree distributions and their applications", *Physical Rev. E*, vol. 64, p. 26,118, 2001

Results on the configuration model

P1) Phase transition on the emergence of a giant component

- Condition depends on first two moments of given $\mathrm{P}[d]$
- Giant component has size $O(N_v)$ as in $G_{N_v,p}$

- M. Molloy and B. Reed, "A critical point for random graphs with a given degree sequence," *Random Struct. and Alg.*, vol. 6, pp. 161-180, 1995

P2) Clustering coefficient vanishes slower than in $G_{N_v,p}$

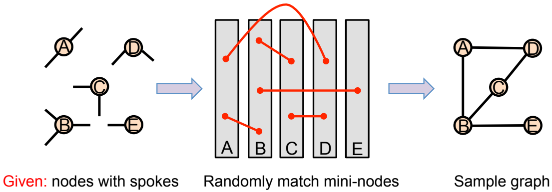
- M. Newman et al, "Random graphs with arbitrary degree distributions and their applications", *Physical Rev. E*, vol. 64, p. 26,118, 2001

P3) Special case of given power-law degree distribution $\mathrm{P}[d] \sim Cd^{-\alpha}$

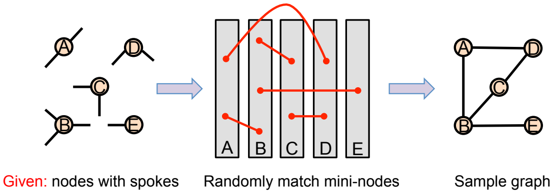
- For $\alpha \in (2,3)$, short diameter $O(\log N_v)$ as in $G_{N_v,p}$

- F. Chung and L. Lu, "The average distances in random graphs with given expected degrees," *PNAS*, vol. 99, pp. 15,879-15,882, 2002

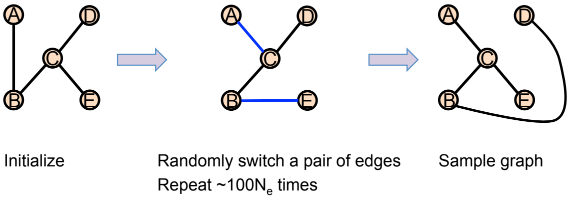
- Matching algorithm



- Matching algorithm



- Switching algorithm



Task: Assessing significance in network graphs

- Consider a graph G^{obs} derived from observations
- Q: Is a structural characteristic $\eta(G^{obs})$ significant, i.e., unusual?

  $\Rightarrow$ Assessing significance requires a frame of reference, or null model

Task: Assessing significance in network graphs

- Consider a graph G^{obs} derived from observations
- Q: Is a structural characteristic $\eta(G^{obs})$ significant, i.e., unusual?
  - $\Rightarrow$ Assessing significance requires a frame of reference, or null model
  - $\Rightarrow$ Random graph models often used in setting up such comparisons

Task: Assessing significance in network graphs

- Consider a graph G^{obs} derived from observations
- Q: Is a structural characteristic $\eta(G^{obs})$ significant, i.e., unusual?
  - $\Rightarrow$ Assessing significance requires a frame of reference, or null model
  - $\Rightarrow$ Random graph models often used in setting up such comparisons
- Define collection $\mathcal{G}$, and compare $\eta(G^{obs})$ with values $\{\eta(G) : G \in \mathcal{G}\}$
  - $\Rightarrow$ Formally, construct the reference distribution

$$\mathbb{P}_{\eta,\mathcal{G}}(t) = \frac{|\{G \in \mathcal{G} : \eta(G) \leq t\}|}{|\mathcal{G}|}$$

- If $\eta(G^{obs})$ found to be sufficiently unlikely under $\mathbb{P}_{\eta,\mathcal{G}}(t)$
  - $\Rightarrow$ Evidence against the null H_0: G^{obs} is a uniform draw from $\mathcal{G}$

- Zachary's karate club has clustering coefficient $\text{cl}(G^{obs}) = 0.2257$

  $\Rightarrow$ Random graph models to assess whether the value is unusual

Example: Zachary's karate club

- Zachary's karate club has clustering coefficient $\text{cl}(G^{obs}) = 0.2257$
  - $\Rightarrow$ Random graph models to assess whether the value is unusual
- Construct two 'comparable' abstract frames of reference
  1) Collection $\mathcal{G}_1$ of random graphs with same $N_v = 34$ and $N_e = 78$
  2) Add the constraint that $\mathcal{G}_2$ has the same degree distribution as G^{obs}

Example: Zachary's karate club

- Zachary's karate club has clustering coefficient $\text{cl}(G^{obs}) = 0.2257$

  $\Rightarrow$ Random graph models to assess whether the value is unusual

- Construct two 'comparable' abstract frames of reference

  1) Collection $\mathcal{G}_1$ of random graphs with same $N_v = 34$ and $N_e = 78$
  2) Add the constraint that $\mathcal{G}_2$ has the same degree distribution as G^{obs}

- $|\mathcal{G}_1| \approx 8.4 \times 10^{96}$ and $|\mathcal{G}_2|$ much smaller, but still large

  $\Rightarrow$ Enumerating $\mathcal{G}_1$ intractable to obtain $\mathrm{P}_{\eta,\mathcal{G}_1}(t)$ exactly

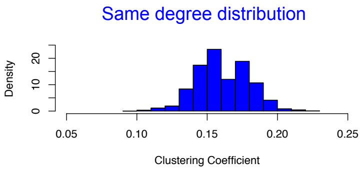
- Instead use simulations to approximate both distributions

  $\Rightarrow$ Draw 10,000 uniform samples G from each $\mathcal{G}_1$ and $\mathcal{G}_2$

  $\Rightarrow$ Calculate $\eta(G) = \text{cl}(G)$ for each sample, plot histograms

Example: Zachary's karate club (cont.)

- Plot histograms to approximate the distributions



- Unlikely to see a value $\text{cl}(G^{obs}) = 0.2257$ under both graph models

  Ex: only 3 out of 10,000 samples from $\mathcal{G}_1$ had $\text{cl}(G) > 0.2257$

- Strong evidence to reject G^{obs} obtained as sample from $\mathcal{G}_1$ or $\mathcal{G}_2$

Small-world models

You are here

Wk.	Date	Topic	HW	Project
1	23-Aug	Introduction to course	HW0 out	
2	30-Aug	Graph theory	HW0 solutions posted	
3	6-Sep	LABOR DAY (no class)	HW1 out	
4	13-Sep	Centrality measures / Community detection		
5	20-Sep	Community detection		
6	27-Sep	Signal Processing and Deep learning for graphs	HW1 due	
7	4-Oct	Signal Processing and Deep learning for graphs	HW2 out	
8	11-Oct	FALL BREAK (no class)		
9	18-Oct	Network models	HW2 due	
10	25-Oct	Network models	HW3 out	Project proposal due
11	1-Nov	Epidemics		
12	8-Nov	Inference of network topologies, features, and processes	HW3 due	
13	15-Nov	Inference of network topologies, features, and processes		
14	22-Nov	Inference of network topologies, features, and processes		Project progress report
15	29-Nov	Inference of network topologies, features, and processes		
	13-Dec	Project presentation (video recording) and final report due		

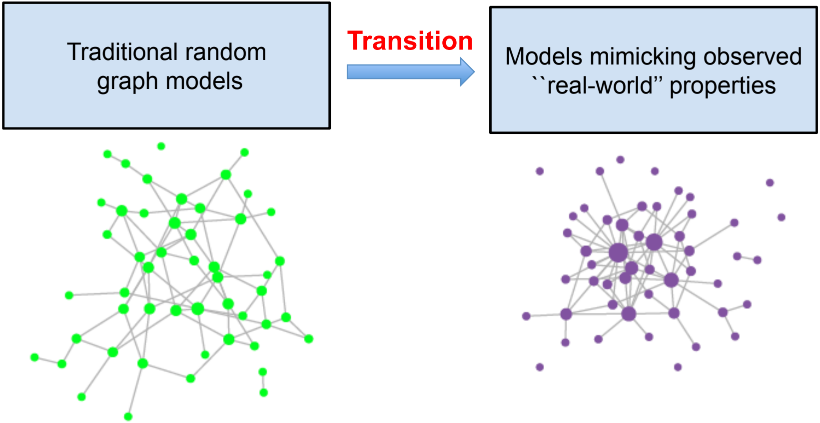
You are here

Wk.	Date	Topic	HW	Project
1	23-Aug	Introduction to course	HW0 out	
2	30-Aug	Graph theory	HW0 solutions posted	
3	6-Sep	LABOR DAY (no class)	HW1 out	
4	13-Sep	Centrality measures / Community detection		
5	20-Sep	Community detection		
6	27-Sep	Signal Processing and Deep learning for graphs	HW1 due	
7	4-Oct	Signal Processing and Deep learning for graphs	HW2 out	
8	11-Oct	FALL BREAK (no class)		
9	18-Oct	Network models	HW2 due	
10	25-Oct	Network models	HW3 out	Project proposal due
11	1-Nov	Epidemics		
12	8-Nov	Inference of network topologies, features, and processes	HW3 due	
13	15-Nov	Inference of network topologies, features, and processes		
14	22-Nov	Inference of network topologies, features, and processes		Project progress report
15	29-Nov	Inference of network topologies, features, and processes		
	13-Dec	Project presentation (video recording) and final report due		

- Comment on HW grading and project proposal

Models for real-world networks

- Arguably the most important innovation in modern graph modeling



A "small" world?

- Six degrees of separation popularized by a play [Guare'90]
  - ⇒ Short paths between us and everyone else on the planet
  - ⇒ Term relatively new, the concept has a long history

A “small” world?

- Six degrees of separation popularized by a play [Guare’90]
  - $\Rightarrow$ Short paths between us and everyone else on the planet
  - $\Rightarrow$ Term relatively new, the concept has a long history
- Traced back to F. Karinthy in the 1920s
  - $\Rightarrow$ ‘Shrinking’ modern world due to increased human connectedness
  - $\Rightarrow$ Challenge: find someone whose distance from you is > 5
  - $\Rightarrow$ Inspired by G. Marconi’s Nobel prize speech in 1909

A "small" world?

- Six degrees of separation popularized by a play [Guare'90]
  - ⇒ Short paths between us and everyone else on the planet
  - ⇒ Term relatively new, the concept has a long history
- Traced back to F. Karinthy in the 1920s
  - ⇒ 'Shrinking' modern world due to increased human connectedness
  - ⇒ Challenge: find someone whose distance from you is > 5
  - ⇒ Inspired by G. Marconi's Nobel prize speech in 1909
- First mathematical treatment [Kochen-Pool'50]
  - ⇒ Formally modeled the mechanics of social networks
  - ⇒ But left 'degrees of separation' question unanswered

A "small" world?

- Six degrees of separation popularized by a play [Guare'90]
  - $\Rightarrow$ Short paths between us and everyone else on the planet
  - $\Rightarrow$ Term relatively new, the concept has a long history
- Traced back to F. Karinthy in the 1920s
  - $\Rightarrow$ 'Shrinking' modern world due to increased human connectedness
  - $\Rightarrow$ Challenge: find someone whose distance from you is > 5
  - $\Rightarrow$ Inspired by G. Marconi's Nobel prize speech in 1909
- First mathematical treatment [Kochen-Pool'50]
  - $\Rightarrow$ Formally modeled the mechanics of social networks
  - $\Rightarrow$ But left 'degrees of separation' question unanswered
- Chain of events led to a groundbreaking experiment [Milgram'67]

Milgram's experiment

- Q1: What is the typical geodesic distance between two people?
  - $\Rightarrow$ Experiment on the global friendship (social) network
  - $\Rightarrow$ Cannot measure in full, so need to probe explicitly

Milgram's experiment

- Q1: What is the typical geodesic distance between two people?

  $\Rightarrow$ Experiment on the global friendship (social) network

  $\Rightarrow$ Cannot measure in full, so need to probe explicitly

- S. Milgram's ingenious small-world experiment in 1967
  - 296 letters sent to people in Wichita, KS and Omaha, NE
  - Letters indicated a (unique) contact person in Boston, MA
  - Asked them to forward the letter to the contact, following rules

- **Def:** friend is someone known on a first-name basis

  Rule 1: If contact is a friend then send her the letter; else

  Rule 2: Relay to friend most-likely to be a contact's friend

Milgram's experiment

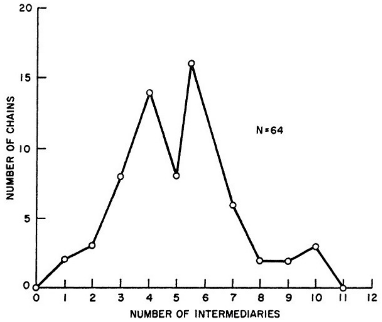
- Q1: What is the typical geodesic distance between two people?
  - $\Rightarrow$ Experiment on the global friendship (social) network
  - $\Rightarrow$ Cannot measure in full, so need to probe explicitly
- S. Milgram's ingenious small-world experiment in 1967
  - 296 letters sent to people in Wichita, KS and Omaha, NE
  - Letters indicated a (unique) contact person in Boston, MA
  - Asked them to forward the letter to the contact, following rules
- **Def:** friend is someone known on a first-name basis

  Rule 1: If contact is a friend then send her the letter; else

  Rule 2: Relay to friend most-likely to be a contact's friend
- Q2: How many letters arrived? How long did they take?

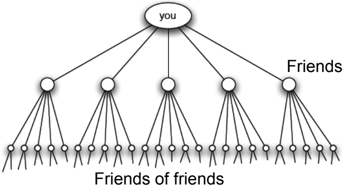
Milgram's experimental results

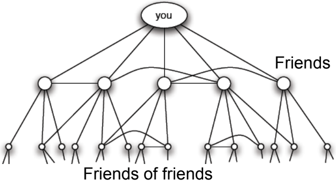
- 64 of 296 letter reached the destination, average path length $\bar{\ell} = 6.2$
  - $\Rightarrow$ Inspiring Guare's '6 degrees of separation'
- Conclusion: short paths connect arbitrary pairs of people



- S. Milgram, "The small-world problem," *Psychology Today*, vol. 2, pp. 60-67, 1967

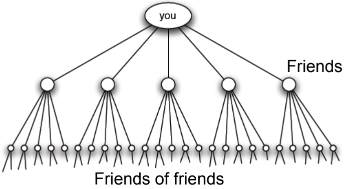
- Milgram demonstrated that short paths are in abundance
- Q: Is the small-world theory reasonable? Sure, e.g., assumes:
  - We have 100 friends, each of them has 100 other friends, . . .
  - After 5 degrees we get 10^{10} friends $\sim$ the Earth's population

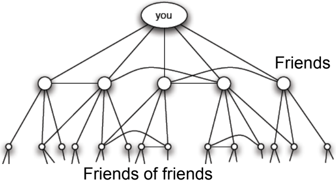




Moment to reflect

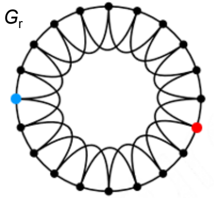
- Milgram demonstrated that short paths are in abundance
- Q: Is the small-world theory reasonable? Sure, e.g., assumes:
  - We have 100 friends, each of them has 100 other friends, . . .
  - After 5 degrees we get 10^{10} friends $\sim$ the Earth's population



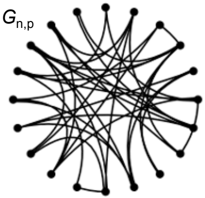


- Not a realistic model of social networks exhibiting:
  - $\Rightarrow$ Homophily [Lazarzfeld'54]
  - $\Rightarrow$ Triadic closure [Rapoport'53]
- Q: How can networks be highly-structured locally **and** globally small?

Structure and randomness as extremes



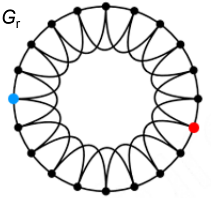
High clustering and diameter



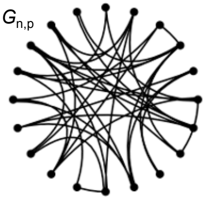
Low clustering and diameter

- One-dimensional regular lattice G_r on N_v vertices
  - Each node is connected to its $2r$ closest neighbors (r to each side)

Structure and randomness as extremes



High clustering and diameter



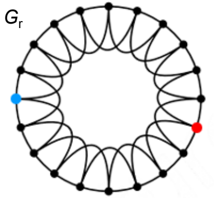
Low clustering and diameter

- One-dimensional regular lattice G_r on N_v vertices
  - Each node is connected to its $2r$ closest neighbors (r to each side)

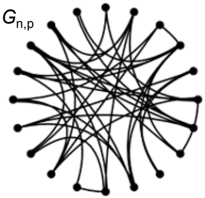
Structure yields high clustering and high diameter

$$\mathrm{cl}(G_r) = \frac{3r-3}{4r-2} \text{ and } \mathrm{diam}(G_r) = \frac{N_v}{2r}$$

Structure and randomness as extremes



High clustering and diameter



Low clustering and diameter

- One-dimensional regular lattice G_r on N_v vertices
  - Each node is connected to its $2r$ closest neighbors (r to each side)

  Structure yields high clustering and high diameter

$$\mathrm{cl}(G_r) = \frac{3r-3}{4r-2} \text{ and } \mathrm{diam}(G_r) = \frac{N_v}{2r}$$

- Other extreme is a $G_{N_v,p}$ random graph with $p = O(N_v^{-1})$

  Randomness yields low clustering and low diameter

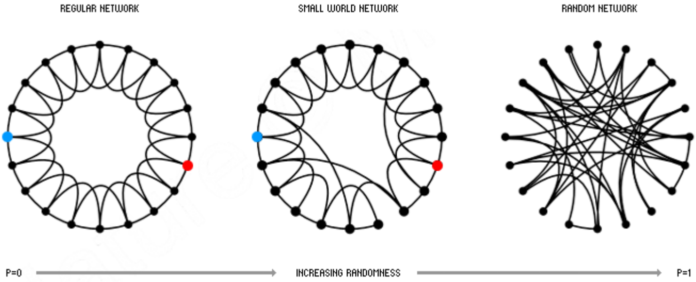
$$\mathrm{cl}(G_{N_v,p}) = O(N_v^{-1}) \text{ and } \mathrm{diam}(G_{N_v,p}) = O(\log N_v)$$

- Small-world model: blend of structure with little randomness

  S1: Start with regular lattice that has desired clustering

  S2: Introduce randomness to generate shortcuts in the graph

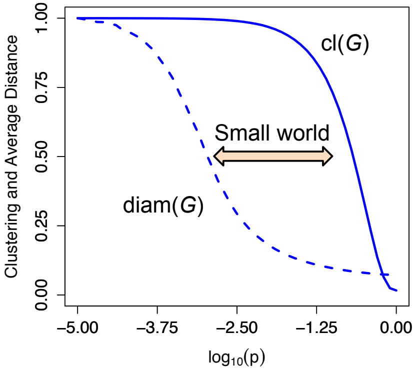
  $\Rightarrow$ Each edge is randomly rewired with (small) probability p



- Rewiring interpolates between the regular and random extremes

Numerical results

- Simulate Watts-Strogatz model with $N_v = 1,000$ and $r = 6$
  - Rewiring probability p varied from 0 (lattice G_r) to 1 (random $G_{N_v,p}$)
  - Normalized $\text{cl}(G)$ and $\text{diam}(G)$ to maximum values ($p = 0$)



- Broad range of $p \in [10^{-3}, 10^{-1}]$ yields small $\text{diam}(G)$ and high $\text{cl}(G)$

Closing remarks

- Structural properties of Watts-Strogatz model [Barrat-Weigt'00]

P1: Large N_v analysis of clustering coefficient

$$\mathrm{cl}(G) \approx \frac{3r-3}{4r-2}(1-p^3) = \mathrm{cl}(G_r)(1-p^3)$$

P2: Degree distribution concentrated around $2r$

Closing remarks

- Structural properties of Watts-Strogatz model [Barrat-Weigt'00]

  P1: Large N_v analysis of clustering coefficient

$$\mathrm{cl}(G) \approx \frac{3r-3}{4r-2}(1-p^3) = \mathrm{cl}(G_r)(1-p^3)$$

  P2: Degree distribution concentrated around $2r$

- Small-world graph models of interest across disciplines

- Particularly relevant to 'communication' in a broad sense

  $\Rightarrow$ Spread of news, gossip, rumors

  $\Rightarrow$ Spread of natural diseases and epidemics

  $\Rightarrow$ Search of content in peer-to-peer networks

Time-evolving networks

- Many networks grow or otherwise evolve in time

  Ex: Web, scientific citations, Twitter, ...

- General approach to model construction mimicking network growth
  - Specify simple mechanisms for network dynamics
  - Study emergent structural characteristics as time $t \to \infty$

- Q: Do these properties match observed ones in real-world networks?

Time-evolving networks

- Many networks grow or otherwise evolve in time

  Ex: Web, scientific citations, Twitter, ...

- General approach to model construction mimicking network growth
  - Specify simple mechanisms for network dynamics
  - Study emergent structural characteristics as time $t \to \infty$

- Q: Do these properties match observed ones in real-world networks?

- Two fundamental and popular classes of growth processes

  $\Rightarrow$ Preferential attachment models

  $\Rightarrow$ Copying models

- Tenable mechanisms for popularity and gene duplication, respectively

Preferential attachment model

- Simple model for the creation of e.g., links among Web pages
- Vertices are created one at a time, denoted $1, \ldots, N_v$

Preferential attachment model

- Simple model for the creation of e.g., links among Web pages
- Vertices are created one at a time, denoted $1, \ldots, N_v$
- When node j is created, it makes a single arc to i, $1 \leq i < j$

Preferential attachment model

- Simple model for the creation of e.g., links among Web pages

- Vertices are created one at a time, denoted $1, \ldots, N_v$
- When node j is created, it makes a single arc to i, $1 \leq i < j$
- Creation of (j, i) governed by a probabilistic rule:
  - With probability p, j links to i chosen uniformly at random
  - With probability $1 - p$, j links to i with probability $\propto d_i^{in}$

Preferential attachment model

- Simple model for the creation of e.g., links among Web pages

- Vertices are created one at a time, denoted $1, \ldots, N_v$
- When node j is created, it makes a single arc to i, $1 \leq i < j$
- Creation of (j, i) governed by a probabilistic rule:
  - With probability p, j links to i chosen uniformly at random
  - With probability $1 - p$, j links to i with probability $\propto d_i^{in}$

- The resulting graph is directed, each vertex has $d_v^{out} = 1$

Preferential attachment model

- Simple model for the creation of e.g., links among Web pages

- Vertices are created one at a time, denoted $1, \ldots, N_v$
- When node j is created, it makes a single arc to i, $1 \leq i < j$
- Creation of (j, i) governed by a probabilistic rule:
  - With probability p, j links to i chosen uniformly at random
  - With probability $1 - p$, j links to i with probability $\propto d_i^{in}$

- The resulting graph is directed, each vertex has $d_v^{out} = 1$

- Preferential attachment model leads to "rich-gets-richer" dynamics

  $\Rightarrow$ Arcs formed preferentially to (currently) most popular nodes

  $\Rightarrow$ Prob. that i increases its popularity $\propto$ i's current popularity

Preferential attachment yields power laws

Theorem

The preferential attachment model gives rise to a power-law in-degree distribution with exponent $\alpha = 1 + \frac{1}{1-p}$, *i.e.,*

$$\mathsf{P}\left[d^{in} = d\right] \propto d^{-\left(1+\frac{1}{1-p}\right)}$$

Preferential attachment yields power laws

Theorem

The preferential attachment model gives rise to a power-law in-degree distribution with exponent $\alpha = 1 + \frac{1}{1-p}$, *i.e.,*

$$\mathsf{P}\left[d^{in} = d\right] \propto d^{-\left(1+\frac{1}{1-p}\right)}$$

- Key: "j links to i with probability $\propto d_i^{in}$" equivalent to copying, i.e., "j chooses k uniformly at random, and links to i if $(k, i) \in E$"
- Reflect: Copy other's decision vs. independent decisions in $G_{n,p}$

Preferential attachment yields power laws

Theorem

The preferential attachment model gives rise to a power-law in-degree distribution with exponent $\alpha = 1 + \frac{1}{1-p}$, *i.e.,*

$$\mathrm{P}\left[d^{in} = d\right] \propto d^{-\left(1+\frac{1}{1-p}\right)}$$

- Key: "j links to i with probability $\propto d_i^{in}$" equivalent to copying, i.e., "j chooses k uniformly at random, and links to i if $(k, i) \in E$"
- Reflect: Copy other's decision vs. independent decisions in $G_{n,p}$
- As $p \to 0$ $\Rightarrow$ Copying more frequent $\Rightarrow$ Smaller $\alpha \to 2$
  - Intuitive: more likely to see extremely popular pages (heavier tail)

Continuous approximation

- In-degree $d_i^{in}(t)$ of node i at time $t \geq i$ is a RV. Two facts
  F1) Initial condition: $d_i^{in}(i) = 0$ since node i just created at time $t = i$
  F2) Dynamics of $d_i^{in}(t)$: Probability that new node $t+1 > i$ links to i is

$$\mathrm{P}\left[(t+1, i) \in E\right] = p \times \frac{1}{t} + (1-p) \times \frac{d_i^{in}(t)}{t}$$

Continuous approximation

- In-degree $d_i^{in}(t)$ of node i at time $t \geq i$ is a RV. Two facts
  F1) Initial condition: $d_i^{in}(i) = 0$ since node i just created at time $t = i$
  F2) Dynamics of $d_i^{in}(t)$: Probability that new node $t+1 > i$ links to i is

$$\mathrm{P}\left[(t+1, i) \in E\right] = p \times \frac{1}{t} + (1-p) \times \frac{d_i^{in}(t)}{t}$$

- Will study a deterministic, continuous approximation to the model
  - Continuous time $t \in [0, N_v]$
  - Continuous degrees $x_i^{in}(t) : [i, N_v] \mapsto \mathbb{R}_+$ are deterministic

Continuous approximation

- In-degree $d_i^{in}(t)$ of node i at time $t \geq i$ is a RV. Two facts
  F1) Initial condition: $d_i^{in}(i) = 0$ since node i just created at time $t = i$
  F2) Dynamics of $d_i^{in}(t)$: Probability that new node $t+1 > i$ links to i is

$$\mathrm{P}\left[(t+1, i) \in E\right] = p \times \frac{1}{t} + (1-p) \times \frac{d_i^{in}(t)}{t}$$

- Will study a deterministic, continuous approximation to the model
  - Continuous time $t \in [0, N_v]$
  - Continuous degrees $x_i^{in}(t) : [i, N_v] \mapsto \mathbb{R}_+$ are deterministic

- Require in-degrees to satisfy the following growth equation

$$\frac{dx_i^{in}(t)}{dt} = \frac{p}{t} + \frac{(1-p)x_i^{in}(t)}{t}, \quad x_i^{in}(i) = 0$$

Solving the differential equation

- Solve the first-order differential equation for $x_i^{in}(t)$ (let $q = 1 - p$)

$$\frac{dx_i^{in}}{dt} = \frac{p + qx_i^{in}}{t}$$

Solving the differential equation

- Solve the first-order differential equation for $x_i^{in}(t)$ (let $q = 1 - p$)

$$\frac{dx_i^{in}}{dt} = \frac{p + qx_i^{in}}{t}$$

- Divide both sides by $p + qx_i^{in}(t)$ and integrate over t

$$\int \frac{1}{p + qx_i^{in}} \frac{dx_i^{in}}{dt} dt = \int \frac{1}{t} dt$$

Solving the differential equation

- Solve the first-order differential equation for $x_i^{in}(t)$ (let $q = 1 - p$)

$$\frac{dx_i^{in}}{dt} = \frac{p + qx_i^{in}}{t}$$

- Divide both sides by $p + qx_i^{in}(t)$ and integrate over t

$$\int \frac{1}{p + qx_i^{in}} \frac{dx_i^{in}}{dt} dt = \int \frac{1}{t} dt$$

- Solving the integrals, we obtain (c is a constant)

$$\ln\left(p + qx_i^{in}\right) = q \ln(t) + c$$

Solving the differential equation (cont.)

- Exponentiating and letting $K = e^c$ we find

$$\ln\left(p + q x_i^{in}(t)\right) = q \ln(t) + c \;\Rightarrow\; x_i^{in}(t) = \frac{1}{q}\left(K t^q - p\right)$$

Solving the differential equation (cont.)

- Exponentiating and letting $K = e^c$ we find

$$\ln\left(p + q x_i^{in}(t)\right) = q \ln(t) + c \;\Rightarrow\; x_i^{in}(t) = \frac{1}{q}\left(K t^q - p\right)$$

- To determine the unknown constant K, use the initial condition

$$0 = x_i^{in}(i) = \frac{1}{q}\left(K i^q - p\right) \;\Rightarrow\; K = \frac{p}{i^q}$$

Solving the differential equation (cont.)

- Exponentiating and letting $K = e^c$ we find

$$\ln\left(p + q x_i^{in}(t)\right) = q \ln(t) + c \;\Rightarrow\; x_i^{in}(t) = \frac{1}{q}\left(K t^q - p\right)$$

- To determine the unknown constant K, use the initial condition

$$0 = x_i^{in}(i) = \frac{1}{q}\left(K i^q - p\right) \;\Rightarrow\; K = \frac{p}{i^q}$$

- Hence, the deterministic approximation of $d_i^{in}(t)$ evolves as

$$x_i^{in}(t) = \frac{1}{q}\left(\frac{p}{i^q} \times t^q - p\right) = \frac{p}{q}\left[\left(\frac{t}{i}\right)^q - 1\right]$$

Obtaining the degree distribution

- Q: At time t, what fraction $\bar{F}(d)$ of all nodes have in-degree $\geq d$?

  Approximation: What fraction of all functions $x_i^{in}(t) \geq d$ by time t?

$$x_i^{in}(t) = \frac{p}{q}\left[\left(\frac{t}{i}\right)^q - 1\right] \geq d$$

Obtaining the degree distribution

- Q: At time t, what fraction $\bar{F}(d)$ of all nodes have in-degree $\geq d$?
  Approximation: What fraction of all functions $x_i^{in}(t) \geq d$ by time t?

$$x_i^{in}(t) = \frac{p}{q}\left[\left(\frac{t}{i}\right)^q - 1\right] \geq d$$

- Can be rewritten in terms of i as

$$i \leq t\left[\left(\frac{q}{p}\right)d + 1\right]^{-1/q}$$

Obtaining the degree distribution

- Q: At time t, what fraction $\bar{F}(d)$ of all nodes have in-degree $\geq d$?
  Approximation: What fraction of all functions $x_i^{in}(t) \geq d$ by time t?

$$x_i^{in}(t) = \frac{p}{q}\left[\left(\frac{t}{i}\right)^q - 1\right] \geq d$$

- Can be rewritten in terms of i as

$$i \leq t\left[\left(\frac{q}{p}\right)d + 1\right]^{-1/q}$$

- By time t there are exactly t nodes in the graph, so the fraction is

$$\bar{F}(d) = \left[\left(\frac{q}{p}\right)d + 1\right]^{-1/q} = 1 - F(d)$$

Identifying the power law

- The degree distribution is given by the PDF $p(d)$
- Recall that the PDF, CDF and CCDF are related, namely

$$p(x) = \frac{dF(x)}{dx} = -\frac{d\bar{F}(x)}{dx}$$

Identifying the power law

- The degree distribution is given by the PDF $p(d)$
- Recall that the PDF, CDF and CCDF are related, namely

$$p(x) = \frac{dF(x)}{dx} = -\frac{d\bar{F}(x)}{dx}$$

- Differentiating $\bar{F}(d) = \left[\left(\frac{q}{p}\right) d + 1\right]^{-1/q}$ yields

$$p(d) = \frac{1}{p}\left[\left(\frac{q}{p}\right) d + 1\right]^{-(1+\frac{1}{q})}$$

- Showed $p(d) \propto d^{-(1+1/q)}$, a power law with exponent $\alpha = 1 + \frac{1}{1-p}$
  - $\Rightarrow$ Disclaimer: Relied on heuristic arguments
  - $\Rightarrow$ Rigorous, probabilistic analysis possible

The Barabási-Albert model

- Barabási-Albert (BA) model is for undirected graphs
- Initial graph $G_{BA}(0)$ of $N_v(0)$ vertices and $N_e(0)$ edges $(t = 0)$

The Barabási-Albert model

- Barabási-Albert (BA) model is for undirected graphs

- Initial graph $G_{BA}(0)$ of $N_v(0)$ vertices and $N_e(0)$ edges $(t=0)$
- For $t = 1, 2, \ldots$ current graph $G_{BA}(t-1)$ grows to $G_{BA}(t)$ by:
  - Adding a new vertex u of degree $d_u(t) = m \geq 1$
  - The m new edges are incident to m different vertices in $G_{BA}(t-1)$
  - New vertex u is connected to $v \in V(t-1)$ w.p.

$$\mathrm{P}\left[(u,v) \in E(t)\right] = \frac{d_v(t-1)}{\sum_{v'} d_{v'}(t-1)}$$

- Vertices connected to u preferentially towards higher degrees

  $\Rightarrow G_{BA}(t)$ has $N_v(t) = N_v(0) + t$ and $N_e(t) = N_e(0) + tm$

- A. Barabási and R. Albert, "Emergence of scaling in random networks," *Science*, vol. 286, pp. 509-512, 1999

The Barabási-Albert model

- Barabási-Albert (BA) model is for undirected graphs
- Initial graph $G_{BA}(0)$ of $N_v(0)$ vertices and $N_e(0)$ edges ($t = 0$)
- For $t = 1, 2, \ldots$ current graph $G_{BA}(t-1)$ grows to $G_{BA}(t)$ by:
  - Adding a new vertex u of degree $d_u(t) = m \geq 1$
  - The m new edges are incident to m different vertices in $G_{BA}(t-1)$
  - New vertex u is connected to $v \in V(t-1)$ w.p.

$$\mathrm{P}\left[(u,v) \in E(t)\right] = \frac{d_v(t-1)}{\sum_{v'} d_{v'}(t-1)}$$

- Vertices connected to u preferentially towards higher degrees

  $\Rightarrow$ $G_{BA}(t)$ has $N_v(t) = N_v(0) + t$ and $N_e(t) = N_e(0) + tm$

- A. Barabási and R. Albert, "Emergence of scaling in random networks," *Science*, vol. 286, pp. 509-512, 1999

http://estebanmoro.org/post/2015-12-21-temporal-networks-with-r-and-igraph-updated/

Linearized chord diagram

- BA model ambiguous in how to select m vertices $\propto$ to their degree
  - $\Rightarrow$ Joint distribution not specified by marginal on each vertex
- Linearized chord diagram (LCD) model removes ambiguities

Linearized chord diagram

- BA model ambiguous in how to select m vertices $\propto$ to their degree
  $\Rightarrow$ Joint distribution not specified by marginal on each vertex
- Linearized chord diagram (LCD) model removes ambiguities
- For $m = 1$, start with $G_{LCD}(0)$ consisting of a vertex with a self-loop

Linearized chord diagram

- BA model ambiguous in how to select m vertices $\propto$ to their degree

  $\Rightarrow$ Joint distribution not specified by marginal on each vertex

- Linearized chord diagram (LCD) model removes ambiguities

- For $m = 1$, start with $G_{LCD}(0)$ consisting of a vertex with a self-loop
- For $t = 1, 2, \ldots$ current graph $G_{LCD}(t-1)$ grows to $G_{LCD}(t)$ by:
  - Adding a new vertex v_t with an edge to $v_s \in V(t)$
  - Vertex v_s, $1 \leq s \leq t$ is chosen w.p.

$$\mathrm{P}\left[s = j\right] = \begin{cases} \frac{d_{v_j}(t-1)}{2t-1}, & \text{if } 1 \leq j \leq t-1, \\ \frac{1}{2t-1}, & \text{if } j = t \end{cases}$$

Linearized chord diagram

- BA model ambiguous in how to select m vertices $\propto$ to their degree

  $\Rightarrow$ Joint distribution not specified by marginal on each vertex

- Linearized chord diagram (LCD) model removes ambiguities

- For $m = 1$, start with $G_{LCD}(0)$ consisting of a vertex with a self-loop
- For $t = 1, 2, \ldots$ current graph $G_{LCD}(t-1)$ grows to $G_{LCD}(t)$ by:
  - Adding a new vertex v_t with an edge to $v_s \in V(t)$
  - Vertex v_s, $1 \leq s \leq t$ is chosen w.p.

$$\mathrm{P}\left[s = j\right] = \begin{cases} \frac{d_{v_j}(t-1)}{2t-1}, & \text{if } 1 \leq j \leq t-1, \\ \frac{1}{2t-1}, & \text{if } j = t \end{cases}$$

- For $m > 1$ simply run the above process m times for each t
  - Collapse all created vertices into a single one, retaining edges

- A. Bollobás et al, "The degree sequence of a scale-free random graph process," *Random Struct. and Alg.*, vol. 18, pp. 279-290, 2001

Properties of the LCD model

P1) The LCD model allows for loops and multi-edges, occurring rarely

Properties of the LCD model

P1) The LCD model allows for loops and multi-edges, occurring rarely

P2) $G_{LCD}(t)$ has power-law degree distribution with $\alpha = 3$, as $t \to \infty$

Properties of the LCD model

P1) The LCD model allows for loops and multi-edges, occurring rarely

P2) $G_{LCD}(t)$ has power-law degree distribution with $\alpha = 3$, as $t \to \infty$

P3) The BA model yields connected graphs if $G_{BA}(0)$ connected

$\Rightarrow$ Not true for the LCD model, but $G_{LCD}(t)$ connected w.h.p.

Properties of the LCD model

P1) The LCD model allows for loops and multi-edges, occurring rarely

P2) $G_{LCD}(t)$ has power-law degree distribution with $\alpha = 3$, as $t \to \infty$

P3) The BA model yields connected graphs if $G_{BA}(0)$ connected

$\Rightarrow$ Not true for the LCD model, but $G_{LCD}(t)$ connected w.h.p.

P4) Small-world behavior

$$\text{diam}(G_{LCD}(t)) = \begin{cases} O(\log N_v(t)), & m = 1 \\ O(\frac{\log N_v(t)}{\log \log N_v(t)}), & m > 1 \end{cases}$$

Properties of the LCD model

P1) The LCD model allows for loops and multi-edges, occurring rarely

P2) $G_{LCD}(t)$ has power-law degree distribution with $\alpha = 3$, as $t \to \infty$

P3) The BA model yields connected graphs if $G_{BA}(0)$ connected

$\Rightarrow$ Not true for the LCD model, but $G_{LCD}(t)$ connected w.h.p.

P4) Small-world behavior

$$\text{diam}(G_{LCD}(t)) = \begin{cases} O(\log N_v(t)), & m = 1 \\ O(\frac{\log N_v(t)}{\log \log N_v(t)}), & m > 1 \end{cases}$$

P5) Unsatisfactory clustering, since small for $m > 1$

$$\mathbb{E}\left[\text{cl}(G_{LCD}(t))\right] \approx \frac{m-1}{8} \frac{(\log N_v(t))^2}{N_v(t)}$$

$\Rightarrow$ Marginally better than $O(N_v^{-1})$ in classical random graphs

Copying models

- Copying is another mechanism of fundamental interest
  Ex: gene duplication to re-use information in organism's evolution
- Different from preferential attachment, but still results in power laws

Copying models

- Copying is another mechanism of fundamental interest

  Ex: gene duplication to re-use information in organism's evolution

- Different from preferential attachment, but still results in power laws

- Initialize with a graph $G_C(0)$ $(t = 0)$

Copying models

- Copying is another mechanism of fundamental interest

  Ex: gene duplication to re-use information in organism's evolution

- Different from preferential attachment, but still results in power laws

- Initialize with a graph $G_C(0)$ $(t = 0)$
- For $t = 1, 2, \ldots$ current graph $G_C(t-1)$ grows to $G_C(t)$ by:
  - Adding a new vertex u
  - Choosing vertex $v \in V(t-1)$ with uniform probability $\frac{1}{N_v(t-1)}$
  - Joining vertex u with v's neighbors independently w.p. p

Copying models

- Copying is another mechanism of fundamental interest
  Ex: gene duplication to re-use information in organism's evolution
- Different from preferential attachment, but still results in power laws
- Initialize with a graph $G_C(0)$ $(t = 0)$
- For $t = 1, 2, \ldots$ current graph $G_C(t-1)$ grows to $G_C(t)$ by:
  - Adding a new vertex u
  - Choosing vertex $v \in V(t-1)$ with uniform probability $\frac{1}{N_v(t-1)}$
  - Joining vertex u with v's neighbors independently w.p. p
- Case $p = 1$ leads to full duplication of edges from an existing node
- F. Chung et al, "Duplication models for biological networks," *Journal of Computational Biology*, vol. 10, pp. 677-687, 2003

Power laws in biology

TABLE 1. POWER LAW EXPONENTS FOR SOME NETWORKS[a]

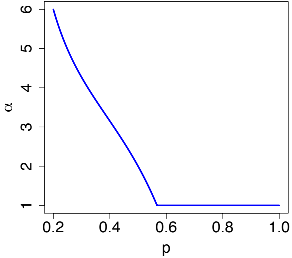
Network	*Approx. exponent β*	*References*
Nonbiological		
Internet	2.1 (in), 2.5 (out)	1–6
Citations	3	6
Actors	2.3	6
Power-grid	4	1, 6
Phone calls	2.1–2.3	6
Biological		
Yeast protein–protein net	1.5, 1.6, 1.7, 2.5	7, 9, 24, 25
E. coli metabolic net	1.7, 2.2	1, 10
Yeast gene expression net	1.4–1.7	9
Gene functional interactions	1.6	11

[a] Some examples of power-law distributions that have been examined and the exponent estimated for each of them. The references are indicated on the right. This is intended to be a representative sample of power-law behaviors.

Asymptotic degree distribution

- Degree distribution tends to a power law w.h.p. [Chung et al'03]
  - $\Rightarrow$ Exponent α is the plotted solution to the equation

$$p(\alpha - 1) = 1 - p^{\alpha - 1}$$



- Full duplication does not lead to power-law behavior; but does if
  - $\Rightarrow$ Partial duplication performed a fraction $q \in (0, 1)$ of times

Fitting network growth models

- Most common practical usage of network growth models is predictive
  Goal: compare characteristics of G^{obs} and $G(t)$ from the models
- Little attempt to date to fit network growth models to data
  - $\Rightarrow$ Expected due to simplicity of such models
  - $\Rightarrow$ Still useful to estimate e.g., the duplication probability p

Fitting network growth models

- Most common practical usage of network growth models is predictive
  Goal: compare characteristics of G^{obs} and $G(t)$ from the models
- Little attempt to date to fit network growth models to data
  - $\Rightarrow$ Expected due to simplicity of such models
  - $\Rightarrow$ Still useful to estimate e.g., the duplication probability p
- To fit a model ideally would like to observe a sequence $\{G^{obs}(\tau)\}_{\tau=1}^{t}$
  - $\Rightarrow$ Unfortunately, such dynamic network data is still fairly elusive
- Q: Can we fit a network growth model to a single snap-shot G^{obs}?

Fitting network growth models

- Most common practical usage of network growth models is predictive

  Goal: compare characteristics of G^{obs} and $G(t)$ from the models

- Little attempt to date to fit network growth models to data

  $\Rightarrow$ Expected due to simplicity of such models

  $\Rightarrow$ Still useful to estimate e.g., the duplication probability p

- To fit a model ideally would like to observe a sequence $\{G^{obs}(\tau)\}_{\tau=1}^{t}$

  $\Rightarrow$ Unfortunately, such dynamic network data is still fairly elusive

- Q: Can we fit a network growth model to a single snap-shot G^{obs}?
- A: Yes, if we leverage the Markovianity of the growth process

Duplication-attachment models

- Similar to all network growth models described so far, suppose:

  As1: A single vertex is added to $G(t-1)$ to create $G(t)$; and

  As2: The manner in which it is added depends only on $G(t-1)$

- In other words, we assume $\{G(t)\}_{t=0}^{\infty}$ is a Markov chain

Duplication-attachment models

- Similar to all network growth models described so far, suppose:
  - **As1:** A single vertex is added to $G(t-1)$ to create $G(t)$; and
  - **As2:** The manner in which it is added depends only on $G(t-1)$
- In other words, we assume $\{G(t)\}_{t=0}^{\infty}$ is a Markov chain
- Let graph $\delta(G(t), v)$ be obtained by deleting v and its edges from $G(t)$
- **Def:** vertex v is removable if $G(t)$ can be obtained from $\delta(G(t), v)$ via copying. If $G(t)$ has no removable vertices, we call it irreducible

Duplication-attachment models

- Similar to all network growth models described so far, suppose:

  As1: A single vertex is added to $G(t-1)$ to create $G(t)$; and
  As2: The manner in which it is added depends only on $G(t-1)$

- In other words, we assume $\{G(t)\}_{t=0}^{\infty}$ is a Markov chain

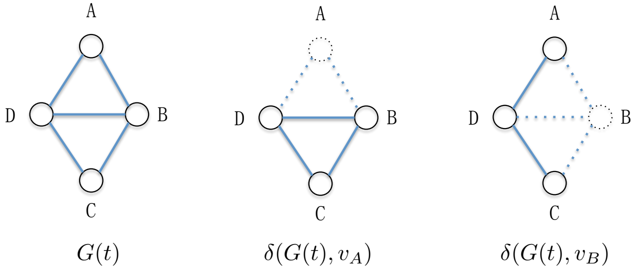
- Let graph $\delta(G(t), v)$ be obtained by deleting v and its edges from $G(t)$

- **Def:** vertex v is removable if $G(t)$ can be obtained from $\delta(G(t), v)$ via copying. If $G(t)$ has no removable vertices, we call it irreducible

- The class of duplication-attachment (DA) models satisfies:
  (i) The initial graph $G(0)$ is irreducible; and
  (ii) $\mathrm{P}_\theta(G(t) \,|\, G(t-1)) > 0 \Leftrightarrow G(t)$ obtained by copying a vertex in $G(t-1)$

- C. Wiuf et al, "A likelihood approach to analysis of network data," *PNAS*, vol. 103, pp. 7566-7570, 2006

Example: reducible graph



- Vertex v_A is removable (likewise v_c by symmetry)

  $\Rightarrow$ Obtain $G(t)$ from $\delta(G(t, v_a))$ by copying v_c

- This implies that $G(t)$ is reducible

  $\Rightarrow$ Notice though that v_B or v_D are not removable

MLE for DA model parameters

- Suppose that $G^{obs} = G(t)$ represents the observed network graph
- The likelihood for the parameter θ is recursively given by

$$\mathcal{L}\left(\theta; G(t)\right) = \frac{1}{t} \sum_{v \in \mathcal{R}_{G(t)}} \mathrm{P}_\theta\left(G(t) \,\middle|\, \delta(G(t), v)\right) \mathcal{L}\left(\theta; \delta(G(t), v)\right)$$

$\Rightarrow$ $\mathcal{R}_{G(t)}$ is the set of all removable nodes in $G(t)$

MLE for DA model parameters

- Suppose that $G^{obs} = G(t)$ represents the observed network graph
- The likelihood for the parameter θ is recursively given by

$$\mathcal{L}(\theta; G(t)) = \frac{1}{t} \sum_{v \in \mathcal{R}_{G(t)}} \mathrm{P}_\theta \left(G(t) \,\middle|\, \delta(G(t), v) \right) \mathcal{L}(\theta; \delta(G(t), v))$$

$\Rightarrow$ $\mathcal{R}_{G(t)}$ is the set of all removable nodes in $G(t)$

- The MLE for θ is thus defined as

$$\hat{\theta} = \arg\max_{\theta} \mathcal{L}(\theta; G(t))$$

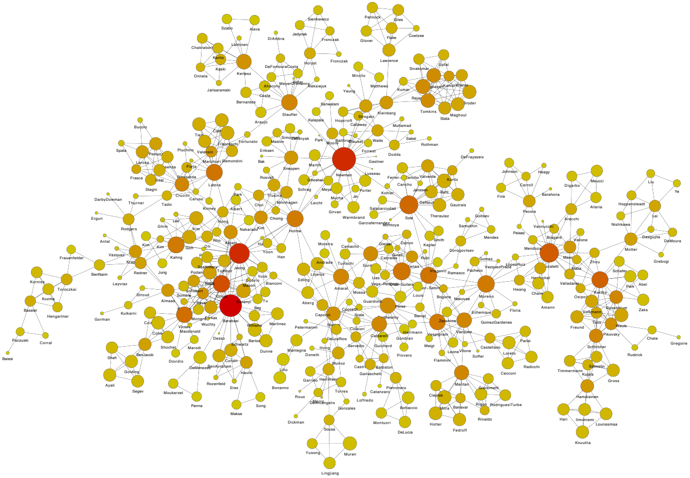
$\Rightarrow$ Computing $\mathcal{L}(\theta; G(t))$ non-trivial, even for modest-size graphs

- Monte Carlo methods to approximate $\mathcal{L}(\theta; G(t))$ [Wiuf et al'06]

$\Rightarrow$ Open issues: vector $\boldsymbol{\theta}$, other growth models, scalability

Network of the week

- Collaborations between network scientists
- "Finding community structure in networks using the eigenvectors of matrices" Newman



- Good statistical graph models should be [Robbins-Morris'07]:
  - $\Rightarrow$ Estimable from and reasonably representative of the data
  - $\Rightarrow$ Theoretically plausible about the underlying network effects
  - $\Rightarrow$ Discriminative among competing effects to best explain the data

Statistical graph models

- Good statistical graph models should be [Robbins-Morris'07]:
  - ⇒ Estimable from and reasonably representative of the data
  - ⇒ Theoretically plausible about the underlying network effects
  - ⇒ Discriminative among competing effects to best explain the data
- ERGMs were designed with these requirements in mind
  - ⇒ also known as p^* models
- G. Robbins et al., "An introduction to exponential random graph (p^*) models for social networks," *Social Networks*, vol. 29, pp. 173-191, 2007

Exponential family

- **Def:** discrete random vector $\mathbf{Z} \in \mathcal{Z}$ belongs to an exponential family if

$$\mathsf{P}_\theta(\mathbf{Z} = \mathbf{z}) = \exp\left\{\boldsymbol{\theta}^\top \mathbf{g}(\mathbf{z}) - \psi(\theta)\right\}$$

  - $\boldsymbol{\theta} \in \mathbb{R}^p$ is a vector of parameters and $\mathbf{g} : \mathcal{Z} \mapsto \mathbb{R}^p$ is a function
  - $\psi(\theta)$ is a normalization term, ensuring $\sum_{\mathbf{z} \in \mathcal{Z}} \mathsf{P}_\theta(\mathbf{z}) = 1$
  - Ex: Bernoulli, binomial, Poisson, geometric distributions

Exponential family

- **Def:** discrete random vector $\mathbf{Z} \in \mathcal{Z}$ belongs to an exponential family if

$$\mathsf{P}_{\theta}(\mathbf{Z} = \mathbf{z}) = \exp\left\{\boldsymbol{\theta}^{\top}\mathbf{g}(\mathbf{z}) - \psi(\theta)\right\}$$

  - $\boldsymbol{\theta} \in \mathbb{R}^p$ is a vector of parameters and $\mathbf{g} : \mathcal{Z} \mapsto \mathbb{R}^p$ is a function
  - $\psi(\theta)$ is a normalization term, ensuring $\sum_{\mathbf{z}\in\mathcal{Z}} \mathsf{P}_{\theta}(\mathbf{z}) = 1$
  - Ex: Bernoulli, binomial, Poisson, geometric distributions

- For continuous exponential families, the pdf has an analogous form

  Ex: Gaussian, Pareto, chi-square distributions

- Exponential families share useful algebraic and geometric properties

  $\Rightarrow$ Mathematically convenient for inference and simulation

Exponential random graph model

- Let $G(V, E)$ be a random undirected graph, with $Y_{ij} := \mathbb{I}\{(i,j) \in E\}$
  - Matrix $\mathbf{Y} = [Y_{ij}]$ is the random adjacency matrix, $\mathbf{y} = [y_{ij}]$ a realization

Exponential random graph model

- Let $G(V, E)$ be a random undirected graph, with $Y_{ij} := \mathbb{I}\{(i,j) \in E\}$
  - Matrix $\mathbf{Y} = [Y_{ij}]$ is the random adjacency matrix, $\mathbf{y} = [y_{ij}]$ a realization
- An ERGM specifies in exponential family form the distribution of $\mathbf{Y}$, i.e.,

$$\mathrm{P}_\theta(\mathbf{Y} = \mathbf{y}) = \left(\frac{1}{\kappa(\boldsymbol{\theta})}\right) \exp\left\{\sum_H \theta_H g_H(\mathbf{y})\right\}, \quad \text{where}$$

(i) each H is a configuration, meaning a set of possible edges in G;

(ii) $g_H(\mathbf{y})$ is the network statistic corresponding to configuration H

$$g_H(\mathbf{y}) = \prod_{y_{ij} \in H} y_{ij} = \mathbb{I}\{H \text{ occurs in } \mathbf{y}\}$$

(iii) $\theta_H \neq 0$ only if all edges in H are conditionally dependent; and

(iv) $\kappa(\boldsymbol{\theta})$ is a normalization constant ensuring $\sum_{\mathbf{y}} \mathrm{P}_\theta(\mathbf{y}) = 1$

Discussion

- Graph order N_v is fixed and given, only edges are random
  - $\Rightarrow$ Assumed unweighted, undirected edges. Extensions possible

Discussion

- Graph order N_v is fixed and given, only edges are random

  $\Rightarrow$ Assumed unweighted, undirected edges. Extensions possible

- ERGMs describe random graphs ‘built-on’ localized patterns
  - These configurations are the structural characteristics of interest
  - Ex: Are there reciprocity effects? Add mutual arcs as configurations
  - Ex: Are there transitivity effects? Consider triangles

Discussion

- Graph order N_v is fixed and given, only edges are random

  $\Rightarrow$ Assumed unweighted, undirected edges. Extensions possible

- ERGMs describe random graphs 'built-on' localized patterns
  - These configurations are the structural characteristics of interest
  - Ex: Are there reciprocity effects? Add mutual arcs as configurations
  - Ex: Are there transitivity effects? Consider triangles

- (In)dependence is conditional on all other variables (edges) in G

  $\Rightarrow$ Control configurations relevant (i.e., $\theta_H \neq 0$) to the model

- Well-specified dependence assumptions imply particular model classes

A general framework for model construction

- In positing an ERGM for a network, one implicitly follows five steps
  - $\Rightarrow$ Explicit choices connecting hypothesized theory to data analysis

Step 1: Each edge (relational tie) is regarded as a random variable

Step 2: A dependence hypothesis is proposed

Step 3: Dependence hypothesis implies a particular form to the model

Step 4: Simplification of parameters through e.g., homogeneity

Step 5: Estimate and interpret model parameters

Example: Bernoulli random graphs

- Assume edges present independently of all other edges (e.g., in $G_{n,p}$)

  $\Rightarrow$ Simplest possible (and unrealistic) dependence assumption

Example: Bernoulli random graphs

- Assume edges present independently of all other edges (e.g., in $G_{n,p}$)

  $\Rightarrow$ Simplest possible (and unrealistic) dependence assumption

- For each (i, j), we assume Y_{ij} independent of Y_{uv}, for all $(u, v) \neq (i, j)$

  $\Rightarrow \theta_H = 0$ for all H involving two or more edges

- Edge configurations i.e., $g_H(\mathbf{y}) = y_{ij}$ relevant, and the ERGM becomes

$$\mathrm{P}_\theta(\mathbf{Y} = \mathbf{y}) = \left(\frac{1}{\kappa(\boldsymbol{\theta})}\right) \exp\left\{\sum_{i,j} \theta_{ij} y_{ij}\right\}$$

Example: Bernoulli random graphs

- Assume edges present independently of all other edges (e.g., in $G_{n,p}$)
  $\Rightarrow$ Simplest possible (and unrealistic) dependence assumption
- For each (i, j), we assume Y_{ij} independent of Y_{uv}, for all $(u, v) \neq (i, j)$
  $\Rightarrow$ $\theta_H = 0$ for all H involving two or more edges
- Edge configurations i.e., $g_H(\mathbf{y}) = y_{ij}$ relevant, and the ERGM becomes

$$\mathrm{P}_\theta(\mathbf{Y} = \mathbf{y}) = \left(\frac{1}{\kappa(\boldsymbol{\theta})}\right) \exp\left\{\sum_{i,j} \theta_{ij} y_{ij}\right\}$$

- Specifies that edge (i, j) present independently, with probability

$$p_{ij} = \frac{\exp(\theta_{ij})}{1 + \exp(\theta_{ij})}$$

Constraints on parameters: homogeneity

- Too many parameters makes estimation infeasible from single $\mathbf{y}$
  - $\Rightarrow$ Under independence have N_v^2 parameters $\{\theta_{ij}\}$. Reduction?

Constraints on parameters: homogeneity

- Too many parameters makes estimation infeasible from single $\mathbf{y}$

  $\Rightarrow$ Under independence have N_v^2 parameters $\{\theta_{ij}\}$. Reduction?

- Homogeneity across all G, i.e., $\theta_{ij} = \theta$ for all (i, j) yields

$$\mathrm{P}_\theta(\mathbf{Y} = \mathbf{y}) = \left(\frac{1}{\kappa(\boldsymbol{\theta})}\right) \exp\{\theta L(\mathbf{y})\}$$

  - Relevant statistic is the number of edges observed $L(\mathbf{y}) = \sum_{i,j} y_{ij}$
  - ERGM identical to $G_{n,p}$, where $p = \dfrac{\exp\theta}{1 + \exp\theta}$

Constraints on parameters: homogeneity

- Too many parameters makes estimation infeasible from single $\mathbf{y}$
  $\Rightarrow$ Under independence have N_v^2 parameters $\{\theta_{ij}\}$. Reduction?
- Homogeneity across all G, i.e., $\theta_{ij} = \theta$ for all (i, j) yields

$$\mathrm{P}_\theta(\mathbf{Y} = \mathbf{y}) = \left(\frac{1}{\kappa(\boldsymbol{\theta})}\right) \exp\{\theta L(\mathbf{y})\}$$

  - Relevant statistic is the number of edges observed $L(\mathbf{y}) = \sum_{i,j} y_{ij}$
  - ERGM identical to $G_{n,p}$, where $p = \dfrac{\exp\theta}{1 + \exp\theta}$

Ex: suppose we know a priori that vertices fall in two sets

- Can impose homogeneity on edges within and between sets, i.e.,

$$\mathrm{P}_\theta(\mathbf{Y} = \mathbf{y}) = \left(\frac{1}{\kappa(\boldsymbol{\theta})}\right) \exp\{\theta_1 L_1(\mathbf{y}) + \theta_{12} L_{12}(\mathbf{y}) + \theta_2 L_2(\mathbf{y})\}$$

Example: Markov random graphs

- Markov dependence notion for graphs [Frank-Strauss'86]
  - Assumes two ties are dependent if they share a common node
  - Edge status Y_{ij} dependent on any other edge involving i or j

Example: Markov random graphs

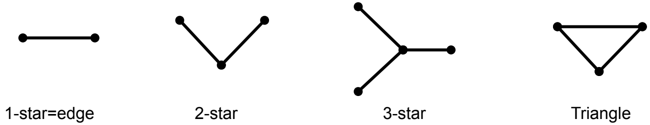
- Markov dependence notion for graphs [Frank-Strauss'86]
  - Assumes two ties are dependent if they share a common node
  - Edge status Y_{ij} dependent on any other edge involving i or j

Theorem

Under homogeneity, G is a Markov random graph if and only if

$$P_\theta(\mathbf{Y} = \mathbf{y}) = \left(\frac{1}{\kappa(\boldsymbol{\theta})}\right) \exp\left\{\sum_{k=1}^{N_v-1} \theta_k S_k(\mathbf{y}) + \theta_\tau T(\mathbf{y})\right\}, \quad \textit{where}$$

$S_k(\mathbf{y})$ is the number of k-stars, and $T(\mathbf{y})$ the number of triangles



Alternative statistics

- Including many higher-order terms challenges estimation
  - $\Rightarrow$ High-order star effects often omitted, e.g., $\theta_k = 0$, $k \geq 4$
  - $\Rightarrow$ But these models tend to fit real data poorly. Dilemma?

Alternative statistics

- Including many higher-order terms challenges estimation
  - $\Rightarrow$ High-order star effects often omitted, e.g., $\theta_k = 0$, $k \geq 4$
  - $\Rightarrow$ But these models tend to fit real data poorly. Dilemma?
- Idea: Impose parametric form $\theta_k \propto (-1)^k \lambda^{2-k}$ [Snijders et al'06]
- Combine $S_k(\mathbf{y})$, $k \geq 2$ into a single alternating k-star statistic, i.e.,

$$\mathrm{AKS}_{\lambda}(\mathbf{y}) = \sum_{k=2}^{N_v - 1} (-1)^k \frac{S_k(\mathbf{y})}{\lambda^{k-2}}, \quad \lambda > 1$$

Alternative statistics

- Including many higher-order terms challenges estimation
  - $\Rightarrow$ High-order star effects often omitted, e.g., $\theta_k = 0$, $k \geq 4$
  - $\Rightarrow$ But these models tend to fit real data poorly. Dilemma?
- Idea: Impose parametric form $\theta_k \propto (-1)^k \lambda^{2-k}$ [Snijders et al'06]
- Combine $S_k(\mathbf{y})$, $k \geq 2$ into a single alternating k-star statistic, i.e.,

$$\text{AKS}_\lambda(\mathbf{y}) = \sum_{k=2}^{N_v-1} (-1)^k \frac{S_k(\mathbf{y})}{\lambda^{k-2}}, \quad \lambda > 1$$

- Can show $\text{AKS}_\lambda(\mathbf{y}) \propto$ the geometrically-weighted degree count

$$\text{GWD}_\gamma(\mathbf{y}) = \sum_{d=0}^{N_v-1} e^{-\gamma d} N_d(\mathbf{y}), \quad \gamma > 0$$

  - $\Rightarrow$ $N_d(\mathbf{y})$ is the number of vertices with degree d

Incorporating vertex attributes

- Straightforward to incorporate vertex attributes to ERGMs
  Ex: gender, seniority in organization, protein function

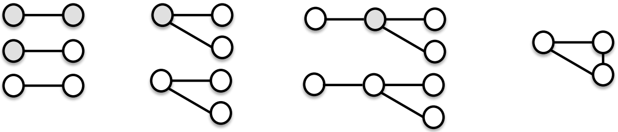
Incorporating vertex attributes

- Straightforward to incorporate vertex attributes to ERGMs
  Ex: gender, seniority in organization, protein function
- Consider a realization $\mathbf{x}$ of a random vector $\mathbf{X} \in \mathbb{R}^{N_v}$ defined on V
- Specify an exponential family form for the conditional distribution

$$\mathrm{P}_\theta(\mathbf{Y} = \mathbf{y} \,\big|\, \mathbf{X} = \mathbf{x})$$

$\Rightarrow$ Will include additional statistics $g(\cdot)$ of $\mathbf{y}$ and $\mathbf{x}$

- Ex: configurations for Markov, binary vertex attributes



Estimating ERGM parameters

- MLE for the parameter vector $\boldsymbol{\theta}$ in an ERGM is

$$\hat{\boldsymbol{\theta}} = \arg\max_{\boldsymbol{\theta}} \left\{ \boldsymbol{\theta}^\top \mathbf{g}(\mathbf{y}) - \psi(\boldsymbol{\theta}) \right\}, \quad \text{where } \psi(\boldsymbol{\theta}) := \log \kappa(\boldsymbol{\theta})$$

Estimating ERGM parameters

- MLE for the parameter vector $\boldsymbol{\theta}$ in an ERGM is

$$\hat{\boldsymbol{\theta}} = \arg\max_{\boldsymbol{\theta}} \left\{ \boldsymbol{\theta}^\top \mathbf{g}(\mathbf{y}) - \psi(\boldsymbol{\theta}) \right\}, \quad \text{where } \psi(\boldsymbol{\theta}) := \log \kappa(\boldsymbol{\theta})$$

- Optimality condition yields

$$\mathbf{g}(\mathbf{y}) = \nabla\psi(\boldsymbol{\theta})\big|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}}$$

- Using also that $\mathbb{E}_\theta[\mathbf{g}(\mathbf{Y})] = \nabla\psi(\boldsymbol{\theta})$, the MLE solves

$$\mathbb{E}_{\hat{\theta}}[\mathbf{g}(\mathbf{Y})] = \mathbf{g}(\mathbf{y})$$

Estimating ERGM parameters

- MLE for the parameter vector $\boldsymbol{\theta}$ in an ERGM is

$$\hat{\boldsymbol{\theta}} = \arg\max_{\boldsymbol{\theta}} \left\{ \boldsymbol{\theta}^\top \mathbf{g}(\mathbf{y}) - \psi(\boldsymbol{\theta}) \right\}, \quad \text{where } \psi(\boldsymbol{\theta}) := \log \kappa(\boldsymbol{\theta})$$

- Optimality condition yields

$$\mathbf{g}(\mathbf{y}) = \nabla \psi(\boldsymbol{\theta})\big|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}}$$

- Using also that $\mathbb{E}_\theta[\mathbf{g}(\mathbf{Y})] = \nabla\psi(\boldsymbol{\theta})$, the MLE solves

$$\mathbb{E}_{\hat{\theta}}[\mathbf{g}(\mathbf{Y})] = \mathbf{g}(\mathbf{y})$$

- Unfortunately $\psi(\boldsymbol{\theta})$ cannot be computed except for small graphs
  - $\Rightarrow$ Involves a summation over $2^{\binom{N_v}{2}}$ values of $\mathbf{y}$ for each $\boldsymbol{\theta}$
  - $\Rightarrow$ Numerical methods needed to obtain approximate values of $\hat{\boldsymbol{\theta}}$

Proof of $\mathbb{E}\left[g(\mathbf{Y})\right] = \nabla\psi(\theta)$

- The pmf of $\mathbf{Y}$ is $\mathsf{P}_\theta(\mathbf{Y} = \mathbf{y}) = \exp\left\{\boldsymbol{\theta}^\top \mathbf{g}(\mathbf{y}) - \psi(\theta)\right\}$, hence

$$
\begin{aligned}
\mathbb{E}_\theta[g(\mathbf{Y})] &= \sum_{\mathbf{y}} g(\mathbf{y}) \mathsf{P}_\theta(\mathbf{Y} = \mathbf{y}) \\
&= \sum_{\mathbf{y}} g(\mathbf{y}) \exp\left\{\boldsymbol{\theta}^\top \mathbf{g}(\mathbf{y}) - \psi(\theta)\right\}
\end{aligned}
$$

Proof of $\mathbb{E}\left[g(\mathbf{Y})\right] = \nabla\psi(\theta)$

- The pmf of $\mathbf{Y}$ is $\mathsf{P}_\theta(\mathbf{Y}=\mathbf{y}) = \exp\left\{\boldsymbol{\theta}^\top \mathbf{g}(\mathbf{y}) - \psi(\theta)\right\}$, hence

$$\mathbb{E}_\theta[g(\mathbf{Y})] = \sum_{\mathbf{y}} g(\mathbf{y})\mathsf{P}_\theta(\mathbf{Y}=\mathbf{y})$$
$$= \sum_{\mathbf{y}} g(\mathbf{y}) \exp\left\{\boldsymbol{\theta}^\top \mathbf{g}(\mathbf{y}) - \psi(\theta)\right\}$$

- Recall $\psi(\boldsymbol{\theta}) = \log \sum_{\mathbf{y}} \exp\left\{\boldsymbol{\theta}^\top \mathbf{g}(\mathbf{y})\right\}$ and use the chain rule

$$\nabla\psi(\boldsymbol{\theta}) = \frac{\sum_{\mathbf{y}} g(\mathbf{y}) \exp\left\{\boldsymbol{\theta}^\top \mathbf{g}(\mathbf{y})\right\}}{\sum_{\mathbf{y}} \exp\left\{\boldsymbol{\theta}^\top \mathbf{g}(\mathbf{y})\right\}} = \frac{\sum_{\mathbf{y}} g(\mathbf{y}) \exp\left\{\boldsymbol{\theta}^\top \mathbf{g}(\mathbf{y})\right\}}{\exp \psi(\theta)}$$
$$= \sum_{\mathbf{y}} g(\mathbf{y}) \exp\left\{\boldsymbol{\theta}^\top \mathbf{g}(\mathbf{y}) - \psi(\theta)\right\}$$

Proof of $\mathbb{E}\left[g(\mathbf{Y})\right] = \nabla\psi(\theta)$

- The pmf of $\mathbf{Y}$ is $\mathsf{P}_\theta(\mathbf{Y}=\mathbf{y}) = \exp\left\{\boldsymbol{\theta}^\top \mathbf{g}(\mathbf{y}) - \psi(\theta)\right\}$, hence

$$\mathbb{E}_\theta[g(\mathbf{Y})] = \sum_{\mathbf{y}} g(\mathbf{y})\mathsf{P}_\theta(\mathbf{Y}=\mathbf{y})$$
$$= \sum_{\mathbf{y}} g(\mathbf{y}) \exp\left\{\boldsymbol{\theta}^\top \mathbf{g}(\mathbf{y}) - \psi(\theta)\right\}$$

- Recall $\psi(\boldsymbol{\theta}) = \log \sum_{\mathbf{y}} \exp\left\{\boldsymbol{\theta}^\top \mathbf{g}(\mathbf{y})\right\}$ and use the chain rule

$$\nabla\psi(\boldsymbol{\theta}) = \frac{\sum_{\mathbf{y}} g(\mathbf{y}) \exp\left\{\boldsymbol{\theta}^\top \mathbf{g}(\mathbf{y})\right\}}{\sum_{\mathbf{y}} \exp\left\{\boldsymbol{\theta}^\top \mathbf{g}(\mathbf{y})\right\}} = \frac{\sum_{\mathbf{y}} g(\mathbf{y}) \exp\left\{\boldsymbol{\theta}^\top \mathbf{g}(\mathbf{y})\right\}}{\exp\psi(\theta)}$$
$$= \sum_{\mathbf{y}} g(\mathbf{y}) \exp\left\{\boldsymbol{\theta}^\top \mathbf{g}(\mathbf{y}) - \psi(\theta)\right\}$$

- The red and blue sums are identical $\Rightarrow \mathbb{E}_\theta[g(\mathbf{Y})] = \nabla\psi(\boldsymbol{\theta})$ follows

- Idea: for fixed $\boldsymbol{\theta}_0$, maximize instead the log-likelihood ratio

$$r(\boldsymbol{\theta}, \boldsymbol{\theta}_0) = \ell(\boldsymbol{\theta}) - \ell(\boldsymbol{\theta}_0) = (\boldsymbol{\theta} - \boldsymbol{\theta}_0)^\top \mathbf{g}(\mathbf{y}) - [\psi(\boldsymbol{\theta}) - \psi(\boldsymbol{\theta}_0)]$$

- Key identity: will show that

$$\exp\{\psi(\boldsymbol{\theta}) - \psi(\boldsymbol{\theta}_0)\} = \mathbb{E}_{\theta_0}\left[\exp\left\{(\boldsymbol{\theta} - \boldsymbol{\theta}_0)^\top \mathbf{g}(\mathbf{Y})\right\}\right]$$

Markov chain Monte Carlo MLE

- Idea: for fixed $\boldsymbol{\theta}_0$, maximize instead the log-likelihood ratio

$$r(\boldsymbol{\theta}, \boldsymbol{\theta}_0) = \ell(\boldsymbol{\theta}) - \ell(\boldsymbol{\theta}_0) = (\boldsymbol{\theta} - \boldsymbol{\theta}_0)^\top \mathbf{g}(\mathbf{y}) - [\psi(\boldsymbol{\theta}) - \psi(\boldsymbol{\theta}_0)]$$

- Key identity: will show that

$$\exp\{\psi(\boldsymbol{\theta}) - \psi(\boldsymbol{\theta}_0)\} = \mathbb{E}_{\theta_0}\left[\exp\left\{(\boldsymbol{\theta} - \boldsymbol{\theta}_0)^\top \mathbf{g}(\mathbf{Y})\right\}\right]$$

- Markov chain Monte Carlo MLE algorithm to search over $\boldsymbol{\theta}$

Step 1: draw samples $\mathbf{Y}_1, \ldots, \mathbf{Y}_n$ from the ERGM under $\boldsymbol{\theta}_0$

Step 2: approximate the above $\mathbb{E}_{\theta_0}[\cdot]$ via sample averaging

Step 3: the logarithm of the result approximates $\psi(\boldsymbol{\theta}) - \psi(\boldsymbol{\theta}_0)$

Step 4: evaluate an $\approx$ log-likelihood ratio $r(\boldsymbol{\theta}, \boldsymbol{\theta}_0)$

- For large n, the maximum value found approximates the MLE $\hat{\boldsymbol{\theta}}$

- Recall $\exp \psi(\boldsymbol{\theta}) = \sum_{\mathbf{y}} \exp\left\{\boldsymbol{\theta}^\top \mathbf{g}(\mathbf{y})\right\}$ to write

$$\exp\left\{\psi(\boldsymbol{\theta}) - \psi(\boldsymbol{\theta}_0)\right\} = \frac{\sum_{\mathbf{y}} \exp\left\{\boldsymbol{\theta}^\top \mathbf{g}(\mathbf{y})\right\}}{\exp \psi(\boldsymbol{\theta}_0)}$$

Derivation of key identity

- Recall $\exp \psi(\boldsymbol{\theta}) = \sum_{\mathbf{y}} \exp \left\{ \boldsymbol{\theta}^\top \mathbf{g}(\mathbf{y}) \right\}$ to write

$$\exp \{\psi(\boldsymbol{\theta}) - \psi(\boldsymbol{\theta}_0)\} = \frac{\sum_{\mathbf{y}} \exp \left\{ \boldsymbol{\theta}^\top \mathbf{g}(\mathbf{y}) \right\}}{\exp \psi(\boldsymbol{\theta}_0)}$$

- Multiplying and dividing by $\exp \left\{ \boldsymbol{\theta}_0^\top \mathbf{g}(\mathbf{y}) \right\} > 0$ yields

$$\begin{aligned}\exp \{\psi(\boldsymbol{\theta}) - \psi(\boldsymbol{\theta}_0)\} &= \sum_{\mathbf{y}} \exp \left\{ (\boldsymbol{\theta} - \boldsymbol{\theta}_0)^\top \mathbf{g}(\mathbf{y}) \right\} \times \frac{\exp \left\{ \boldsymbol{\theta}_0^\top \mathbf{g}(\mathbf{y}) \right\}}{\exp \psi(\boldsymbol{\theta}_0)} \\ &= \sum_{\mathbf{y}} \exp \left\{ (\boldsymbol{\theta} - \boldsymbol{\theta}_0)^\top \mathbf{g}(\mathbf{y}) \right\} \mathrm{P}_{\theta_0}(\mathbf{Y} = \mathbf{y}) \\ &= \mathbb{E}_{\theta_0} \left[\exp \left\{ (\boldsymbol{\theta} - \boldsymbol{\theta}_0)^\top \mathbf{g}(\mathbf{Y}) \right\} \right]\end{aligned}$$

- Used $\exp \left\{ \boldsymbol{\theta}_0^\top \mathbf{g}(\mathbf{y}) - \psi(\boldsymbol{\theta}_0) \right\}$ is the exponential family pmf $\mathrm{P}_{\theta_0}(\mathbf{Y} = \mathbf{y})$

- Best fit chosen from a given class of models . . .

- Best fit chosen from a given class of models ...
  may not be a good fit to the data if model class not rich enough

Model goodness-of-fit

- Best fit chosen from a given class of models ...
  may not be a good fit to the data if model class not rich enough
- Assessing goodness-of-fit for ERGMs

  Step 1: simulate numerous random graphs from the fitted model
  Step 2: compare high-level characteristics with those of G^{obs}
  Ex: distributions of degree, centrality, diameter
- If significant differences found in G^{obs}, conclude
  $\Rightarrow$ Systematic gap between specified model class and data
  $\Rightarrow$ Lack of goodness-of-fit

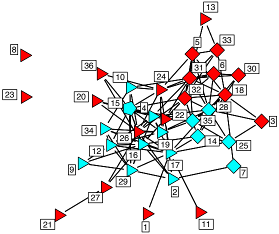
Model goodness-of-fit

- Best fit chosen from a given class of models ...
  may not be a good fit to the data if model class not rich enough
- Assessing goodness-of-fit for ERGMs

  Step 1: simulate numerous random graphs from the fitted model
  Step 2: compare high-level characteristics with those of G^{obs}
  Ex: distributions of degree, centrality, diameter
- If significant differences found in G^{obs}, conclude
  - $\Rightarrow$ Systematic gap between specified model class and data
  - $\Rightarrow$ Lack of goodness-of-fit
- **Take home:** model specification for ERGMs highly nontrivial
  - $\Rightarrow$ Goodness-of-fit diagnostics can play key facilitating role

Lawyer collaboration network

- Network G^{obs} of working relationships among lawyers [Lazega'01]
  - Nodes are $N_v = 36$ partners, edges indicate partners worked together



- Data includes various node-level attributes:
  - Seniority (node labels indicate rank ordering)
  - Office location (triangle, square or pentagon)
  - Type of practice, i.e., litigation (red) and corporate (cyan)
  - Gender (three partners are female labeled 27, 29 and 34)

- Goal: study cooperation among social actors in an organization

Modeling lawyer collaborations

- Assess network effects $S_1(\mathbf{y}) = N_e$ and alternating k-triangles statistic

$$\mathrm{AKT}_\lambda(\mathbf{y}) = 3T_1(\mathbf{y}) + \sum_{k=2}^{N_v-2} (-1)^{k+1} \frac{T_k(\mathbf{y})}{\lambda^{k-1}}$$

$\Rightarrow$ $T_k(\mathbf{y})$ counts sets of k individual triangles sharing a common base

Modeling lawyer collaborations

- Assess network effects $S_1(\mathbf{y}) = N_e$ and alternating k-triangles statistic

$$\text{AKT}_\lambda(\mathbf{y}) = 3T_1(\mathbf{y}) + \sum_{k=2}^{N_v-2} (-1)^{k+1} \frac{T_k(\mathbf{y})}{\lambda^{k-1}}$$

$\Rightarrow$ $T_k(\mathbf{y})$ counts sets of k individual triangles sharing a common base

- Test the following set of exogenous effects:

$$h^{(1)}(\mathbf{x}_i, \mathbf{x}_j) = \text{seniority}_i + \text{seniority}_j, \quad h^{(2)}(\mathbf{x}_i, \mathbf{x}_j) = \text{practice}_i + \text{practice}_j$$
$$h^{(3)}(\mathbf{x}_i, \mathbf{x}_j) = \mathbb{I}\left\{\text{practice}_i = \text{practice}_j\right\}, \quad h^{(4)}(\mathbf{x}_i, \mathbf{x}_j) = \mathbb{I}\left\{\text{gender}_i = \text{gender}_j\right\}$$
$$h^{(5)}(\mathbf{x}_i, \mathbf{x}_j) = \mathbb{I}\left\{\text{office}_i = \text{office}_j\right\}, \quad \mathbf{h}(\mathbf{x}_i, \mathbf{x}_j) := [h^{(1)}(\mathbf{x}_i, \mathbf{x}_j), \ldots, h^{(5)}(\mathbf{x}_i, \mathbf{x}_j)]^T$$

Modeling lawyer collaborations

- Assess network effects $S_1(\mathbf{y}) = N_e$ and alternating k-triangles statistic

$$\mathrm{AKT}_\lambda(\mathbf{y}) = 3T_1(\mathbf{y}) + \sum_{k=2}^{N_v-2} (-1)^{k+1} \frac{T_k(\mathbf{y})}{\lambda^{k-1}}$$

$\Rightarrow$ $T_k(\mathbf{y})$ counts sets of k individual triangles sharing a common base

- Test the following set of exogenous effects:

$$h^{(1)}(\mathbf{x}_i, \mathbf{x}_j) = \text{seniority}_i + \text{seniority}_j, \quad h^{(2)}(\mathbf{x}_i, \mathbf{x}_j) = \text{practice}_i + \text{practice}_j$$
$$h^{(3)}(\mathbf{x}_i, \mathbf{x}_j) = \mathbb{I}\left\{\text{practice}_i = \text{practice}_j\right\}, \quad h^{(4)}(\mathbf{x}_i, \mathbf{x}_j) = \mathbb{I}\left\{\text{gender}_i = \text{gender}_j\right\}$$
$$h^{(5)}(\mathbf{x}_i, \mathbf{x}_j) = \mathbb{I}\left\{\text{office}_i = \text{office}_j\right\}, \quad \mathbf{h}(\mathbf{x}_i, \mathbf{x}_j) := [h^{(1)}(\mathbf{x}_i, \mathbf{x}_j), \ldots, h^{(5)}(\mathbf{x}_i, \mathbf{x}_j)]^T$$

- Resulting ERGM

$$\mathbb{P}_{\boldsymbol{\theta},\boldsymbol{\beta}}(\mathbf{Y} = \mathbf{y} | \mathbf{X} = \mathbf{x}) = \frac{1}{\kappa(\boldsymbol{\theta}, \boldsymbol{\beta})} \exp\left\{\theta_1 S_1(\mathbf{y}) + \theta_2 \mathrm{AKT}_\lambda(\mathbf{y}) + \boldsymbol{\beta}^T \mathbf{g}(\mathbf{y}, \mathbf{x})\right\}$$
$$\mathbf{g}(\mathbf{y}, \mathbf{x}) = \sum_{i,j} y_{ij} \mathbf{h}(\mathbf{x}_i, \mathbf{x}_j)$$

Model fitting result

- Fitting results using the MCMC MLE approach

Parameter	Estimate	'Standard Error'
Density (θ_1)	-6.2073	0.5697
Alternating k-triangles (θ_2)	0.5909	0.0882
Seniority Main Effect (β_1)	0.0245	0.0064
Practice Main Effect (β_2)	0.3945	0.1103
Same Practice (β_3)	0.7721	0.1973
Same Gender (β_4)	0.7302	0.2495
Same Office (β_5)	1.1614	0.1952

- Identified factors that may increase odds of cooperation

  Ex: same practice, gender and office location double odds

- Strong evidence for transitivity effects since $\hat{\theta}_2 \gg \mathrm{se}(\hat{\theta}_2)$

  $\Rightarrow$ Something beyond basic homophily explaining such effects

- Assess goodness-of-fit to G^{obs}
  - Sample from fitted ERGM
- Compared distributions of
  - Degree
  - Edge-wise shared partners
  - Geodesic distance
- Plots show good fit overall