# Power-Law and Maximum Entropy Optimization of Large Networks

Roman V. Belavkin    Panos Pardalos

Faculty of Science and Technology
Middlesex University, London NW4 4BT, UK

Department of Industrial and Systems Engineering
University of Florida
Gainesville, FL 32611-6595, USA

July 21, 2018
ACDL 2018

Analysis of large networks

The power-law degree sequence

Power-law and maximum entropy

Power-law exponent (inverse temperature)

Free energy and phase transition

## Analysis of large networks

The power-law degree sequence

Power-law and maximum entropy

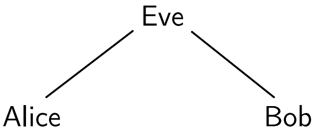
Power-law exponent (inverse temperature)

Free energy and phase transition

# Graphs and networks

## Definition (Graph)

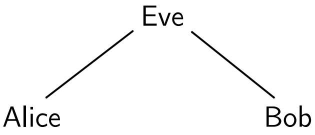
$G = (V, E)$, where $V = \{v_1, \ldots, v_N\}$ — *vertexes*, $E \subseteq V \times V$ — *edges*.



# Graphs and networks

## Definition (Graph)

$G = (V, E)$, where $V = \{v_1, \ldots, v_N\}$ — *vertexes*, $E \subseteq V \times V$ — *edges*.

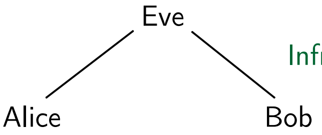



Social networks : edges represent friendships (e.g. Facebook, YouTube, LiveJournal).

# Graphs and networks

## Definition (Graph)

$G = (V, E)$, where $V = \{v_1, \ldots, v_N\}$ — *vertexes*, $E \subseteq V \times V$ — *edges*.



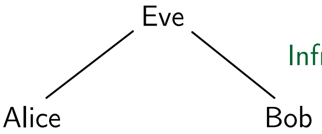

Social networks : edges represent friendships (e.g. Facebook, YouTube, LiveJournal).

Infrastructure : vertexes represent physical locations, edges represent links (e.g. Euroroad, US power grid).

# Graphs and networks

## Definition (Graph)

$G = (V, E)$, where $V = \{v_1, \ldots, v_N\}$ — *vertexes*, $E \subseteq V \times V$ — *edges*.




Social networks : edges represent friendships (e.g. Facebook, YouTube, LiveJournal).

Infrastructure : vertexes represent physical locations, edges represent links (e.g. Euroroad, US power grid).

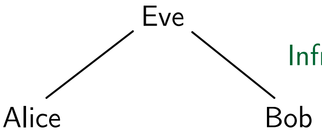
## Questions

- Is the graph connected?
- If not, is there a *giant component*?
- Which nodes are the most important (*centrality*)?
- What is the average degree (number of links of a node)?

# Graphs and networks

## Definition (Graph)

$G = (V, E)$, where $V = \{v_1, \dots, v_N\}$ — *vertexes*, $E \subseteq V \times V$ — *edges*.




**Social networks** : edges represent friendships (e.g. Facebook, YouTube, LiveJournal).

**Infrastructure** : vertexes represent physical locations, edges represent links (e.g. Euroroad, US power grid).

## Questions

- Is the graph connected?
- If not, is there a *giant component*?
- Which nodes are the most important (*centrality*)?
- What is the average degree (number of links of a node)?
- What is the average distance between nodes (shortest path length)?

# Degree Sequence

## Definition (Degree sequence)

Function $N : \mathbb{N} \to \mathbb{N} \cup \{0\}$ representing the nuber $N(k)$ of vertexes $v \in V$ with degree $k$ (number of edges $(v, \cdot)$ or $(\cdot, v) \in E \subseteq V \times V$):

$$N(k) := |\{v \in V : \deg(v) = k\}|$$

# Degree Sequence

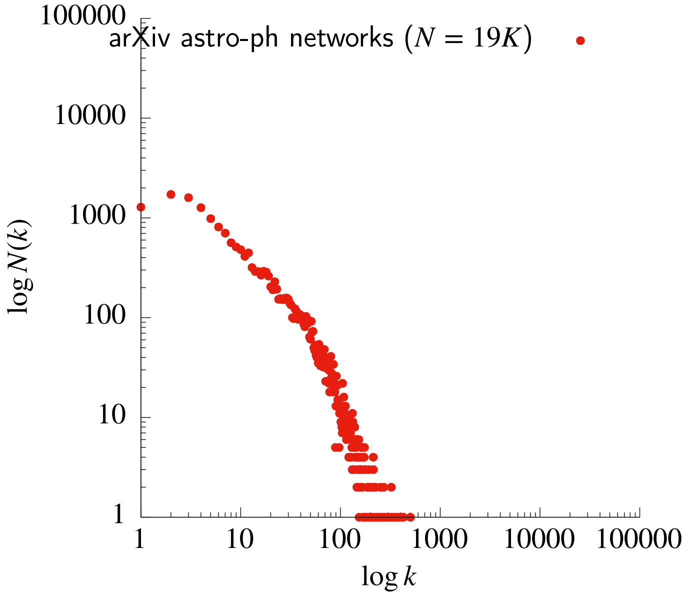
## Definition (Degree sequence)

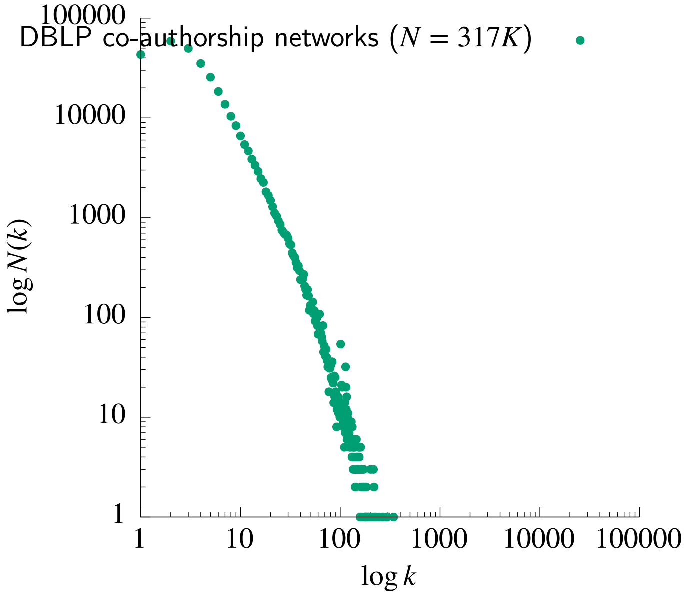
Function $N : \mathbb{N} \to \mathbb{N} \cup \{0\}$ representing the nuber $N(k)$ of vertexes $v \in V$ with degree $k$ (number of edges $(v, \cdot)$ or $(\cdot, v) \in E \subseteq V \times V$):

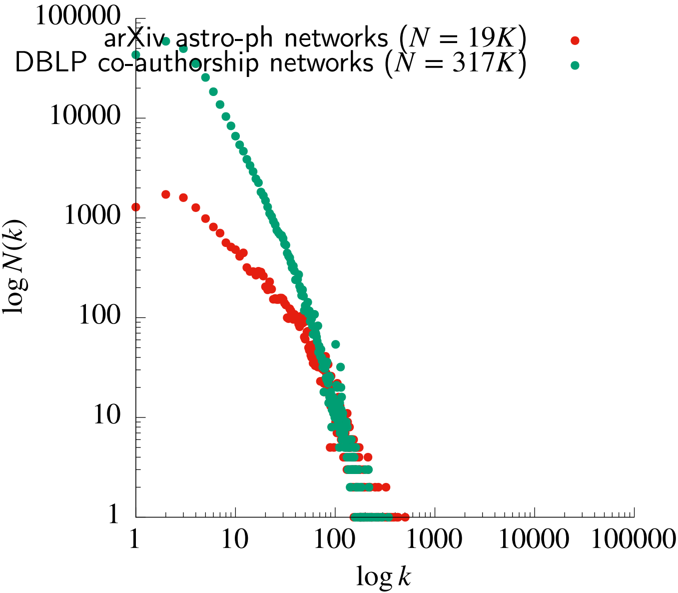
$$N(k) := |\{v \in V : \deg(v) = k\}|$$

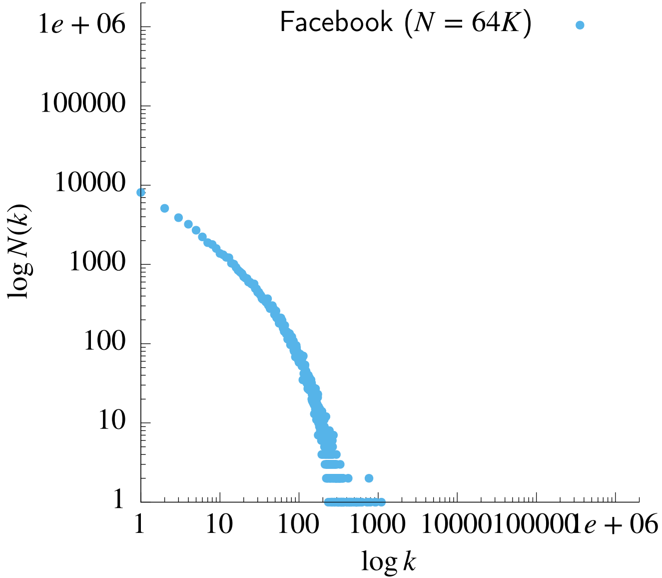
Normalized $N(k)$ is the degree *distribution*

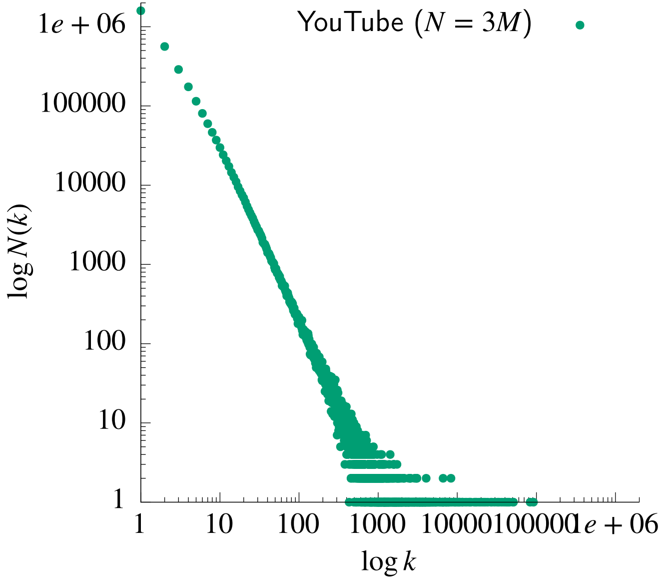
$$P(k) = \frac{N(k)}{N}$$

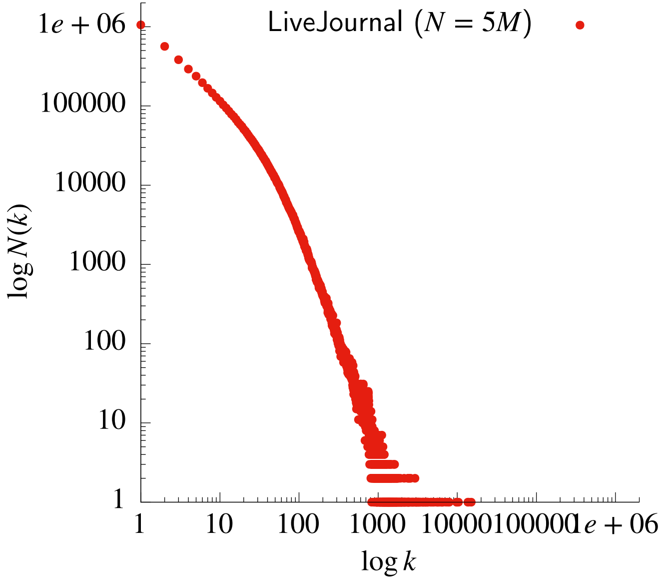

arXiv astro-ph networks ($N = 19K$)
100000
10000
1000
100
10
1
$\log N(k)$
1
10
100
1000
10000
100000
$\log k$

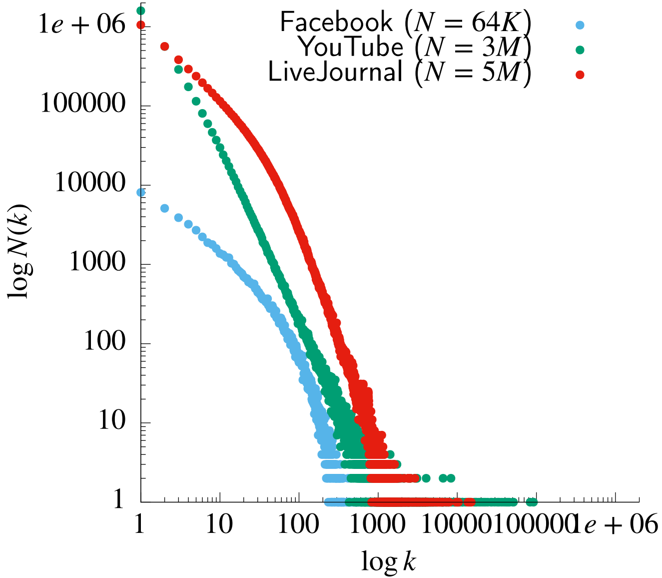

DBLP co-authorship networks ($N = 317K$)
$\log N(k)$
$\log k$
100000
10000
1000
100
10
1
1
10
100
1000
10000
100000

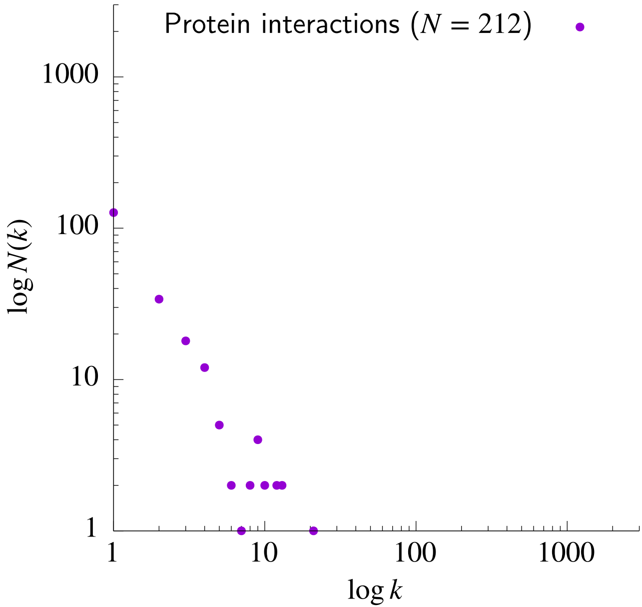

arXiv astro-ph networks ($N = 19K$)
DBLP co-authorship networks ($N = 317K$)
100000
10000
1000
100
10
1
$\log N(k)$
1
10
100
1000
10000
100000
$\log k$

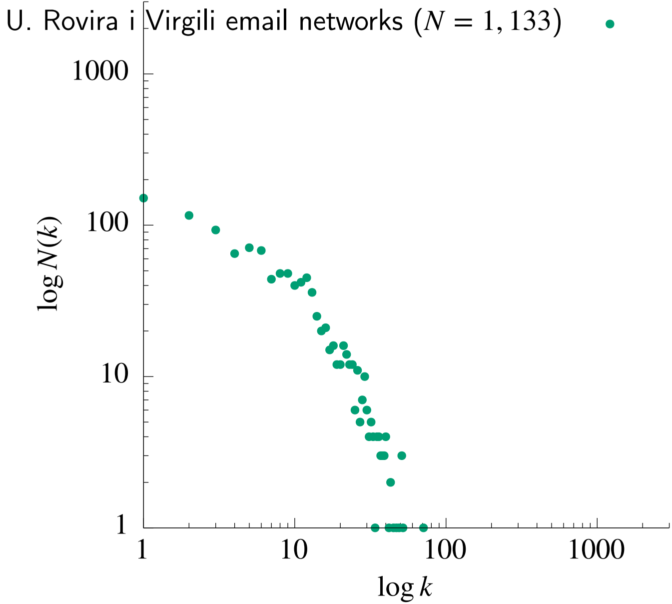

Facebook ($N = 64K$)
1e + 06
100000
10000
1000
100
10
1
log N(k)
1
10
100
1000
10000
100000
1e + 06
log k

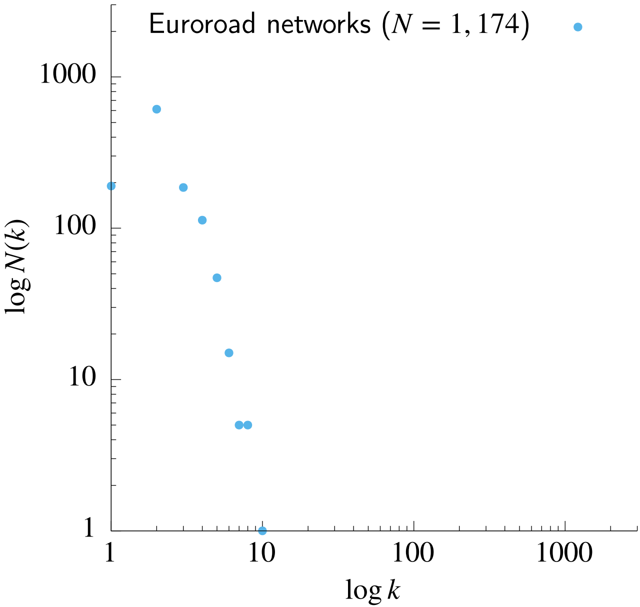

YouTube ($N = 3M$)
$1e+06$
100000
10000
1000
100
10
1
$\log N(k)$
1
10
100
1000
10000
100000
$1e+06$
$\log k$

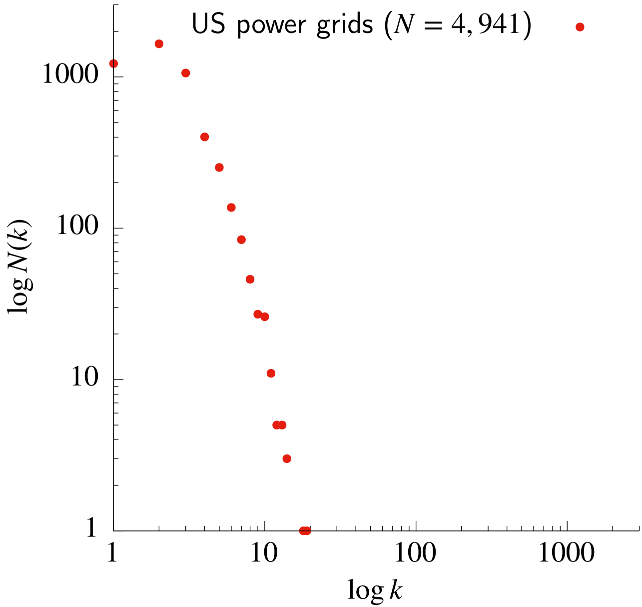

LiveJournal ($N = 5M$)
$\log N(k)$
$\log k$
1e + 06
100000
10000
1000
100
10
1
1
10
100
1000
10000
100000
1e + 06

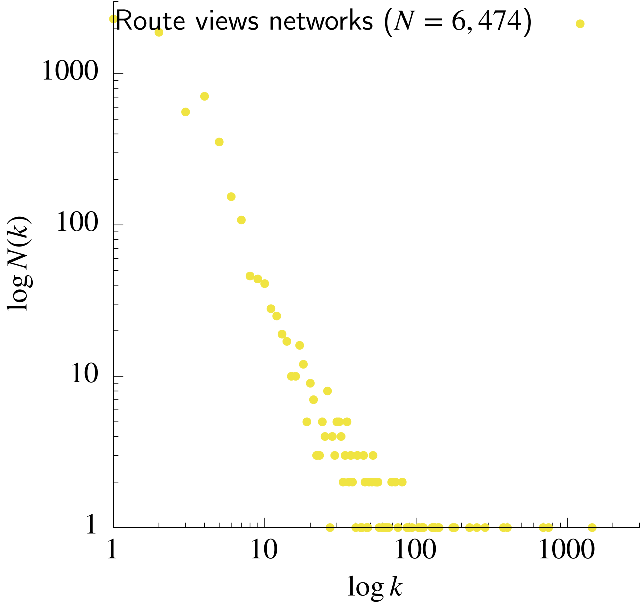

Facebook ($N = 64K$)
YouTube ($N = 3M$)
LiveJournal ($N = 5M$)
1e + 06
100000
10000
1000
100
10
1
$\log N(k)$
1
10
100
1000
10000
100000
1e + 06
$\log k$


Protein interactions ($N = 212$)
1000
100
10
1
1
10
100
1000
$\log N(k)$
$\log k$


U. Rovira i Virgili email networks ($N = 1,133$)
1000
100
10
1
$\log N(k)$
1
10
100
1000
$\log k$


Euroroad networks ($N = 1,174$)
1000
100
10
1
1
10
100
1000
$\log N(k)$
$\log k$


US power grids ($N = 4,941$)
1000
100
10
1
$\log N(k)$
1
10
100
1000
$\log k$


Route views networks ($N = 6,474$)
1000
100
10
1
1
10
100
1000
$\log N(k)$
$\log k$

## Power-Law

$$\ln N(k) = \alpha - \beta \ln k$$

$\alpha$ — intercept

$\beta$ — slope or *exponent*

$$P(k) = \frac{k^{-\beta}}{\sum_{k=1}^{N} k^{-\beta}}$$

## Power-Law

$$\ln N(k) = \alpha - \beta \ln k$$

$\alpha$ — intercept

$\beta$ — slope or *exponent*

$$P(k) = \frac{k^{-\beta}}{\sum_{k=1}^{N} k^{-\beta}}$$

- Almost surely connected for $\beta < 1$, and a.s. disconnected for $\beta > 1$ (?, ?, ?).

## Power-Law

$$\ln N(k) = \alpha - \beta \ln k$$

$\alpha$ — intercept

$\beta$ — slope or *exponent*

$$P(k) = \frac{k^{-\beta}}{\sum_{k=1}^{N} k^{-\beta}}$$

- Almost surely connected for $\beta < 1$, and a.s. disconnected for $\beta > 1$ (?, ?, ?).
- A giant componnent almost surely exists if $Q := \mathbb{E}_P\{k^2\} - 2\mathbb{E}_P\{k\} \geq 0$ (?, ?).

## Power-Law

$$\ln N(k) = \alpha - \beta \ln k$$

$\alpha$ — intercept

$\beta$ — slope or *exponent*

$$P(k) = \frac{k^{-\beta}}{\sum_{k=1}^{N} k^{-\beta}}$$

- Almost surely connected for $\beta < 1$, and a.s. disconnected for $\beta > 1$ (?, ?, ?).
- A giant componnent almost surely exists if $Q := \mathbb{E}_P\{k^2\} - 2\mathbb{E}_P\{k\} \geq 0$ (?, ?).
- For power-law with $N \to \infty$ ($\sum_{k=1}^{N} k^{-\beta} \to \zeta(\beta)$) this becomes $\zeta(\beta - 2) - 2\zeta(\beta - 1) = 0$, which guves the value $\beta_0 \approx 3.47875$ (?, ?, ?).

## Power-Law

$$\ln N(k) = \alpha - \beta \ln k$$

$\alpha$ — intercept
$\beta$ — slope or *exponent*

$$P(k) = \frac{k^{-\beta}}{\sum_{k=1}^{N} k^{-\beta}}$$

- Almost surely connected for $\beta < 1$, and a.s. disconnected for $\beta > 1$ (?, ?, ?).
- A giant componnent almost surely exists if $Q := \mathbb{E}_P\{k^2\} - 2\mathbb{E}_P\{k\} \geq 0$ (?, ?).
- For power-law with $N \to \infty$ ($\sum_{k=1}^{N} k^{-\beta} \to \zeta(\beta)$) this becomes $\zeta(\beta - 2) - 2\zeta(\beta - 1) = 0$, which guves the value $\beta_0 \approx 3.47875$ (?, ?, ?).
- Generated by preferential attachment (?, ?):

## Power-Law

$$\ln N(k) = \alpha - \beta \ln k$$

$\alpha$ — intercept

$\beta$ — slope or *exponent*

$$P(k) = \frac{k^{-\beta}}{\sum_{k=1}^{N} k^{-\beta}}$$

- Almost surely connected for $\beta < 1$, and a.s. disconnected for $\beta > 1$ (?, ?, ?).
- A giant componnent almost surely exists if $Q := \mathbb{E}_P\{k^2\} - 2\mathbb{E}_P\{k\} \geq 0$ (?, ?).
- For power-law with $N \to \infty$ $(\sum_{k=1}^{N} k^{-\beta} \to \zeta(\beta))$ this becomes $\zeta(\beta - 2) - 2\zeta(\beta - 1) = 0$, which guves the value $\beta_0 \approx 3.47875$ (?, ?, ?).
- Generated by preferential attachment (?, ?):

$$P[(i,j) \in E \mid k_i] = \frac{k_i^{\gamma}}{\sum_{k_i=1}^{N} k_i^{\gamma}}$$

# Properties of power-law graphs and networks

Small world : relatively short average path length (distance) between any pair of nodes.

# Properties of power-law graphs and networks

Small world : relatively short average path length (distance) between any pair of nodes.

High clustering : many more cliques are formed compared to a random graph (?, ?).

# Properties of power-law graphs and networks

Small world : relatively short average path length (distance) between any pair of nodes.

High clustering : many more cliques are formed compared to a random graph (?, ?).

High robustness : error and attack tolerance (e.g. random edge or node removal).

# Power-Law and maximum entropy

## Exponential family

Changing $k^{-\beta} = e^{-\beta \ln k}$ let us write

$$P(k) = \frac{k^{-\beta}}{Z(\beta)}$$

# Power-Law and maximum entropy

## Exponential family

Changing $k^{-\beta} = e^{-\beta \ln k}$ let us write

$$P(k) = \frac{e^{-\beta \ln k}}{Z(\beta)}$$

# Power-Law and maximum entropy

## Exponential family

Changing $k^{-\beta} = e^{-\beta \ln k}$ let us write

$$P(k) = \frac{e^{-\beta \ln k}}{Z(\beta)}$$

## Maximum entropy

- $P(k) = e^{-\beta \ln k}/Z(\beta)$ is the solution to:

$$\text{maximize} \quad H(P) \quad \text{subject to} \quad \mathbb{E}_P\{\ln k\} \leq \upsilon$$

# Power-Law and maximum entropy

## Exponential family

Changing $k^{-\beta} = e^{-\beta \ln k}$ let us write

$$P(k) = \frac{e^{-\beta \ln k}}{Z(\beta)}$$

## Maximum entropy

- $P(k) = e^{-\beta \ln k}/Z(\beta)$ is the solution to:

$$\text{maximize} \quad H(P) \quad \text{subject to} \quad \mathbb{E}_P\{\ln k\} \le \upsilon$$

- where $H(P)$ is the entropy of $P(k)$:

$$H(P) = -\sum_{k=1}^{N} [\ln P(k)]\, P(k)$$

# Power-Law and maximum entropy

## Exponential family

Changing $k^{-\beta} = e^{-\beta \ln k}$ let us write

$$P(k) = \frac{e^{-\beta \ln k}}{Z(\beta)}$$

## Maximum entropy

- $P(k) = e^{-\beta \ln k}/Z(\beta)$ is the solution to:

$$\text{minimize} \quad \mathbb{E}_P\{\ln k\} \quad \text{subject to} \quad H(P) \geq \ln N - \lambda \geq 0$$

- where $H(P)$ is the entropy of $P(k)$:

$$H(P) = -\sum_{k=1}^{N} [\ln P(k)]\, P(k)$$

# Solution using Lagrange multipliers

- Lagrange function

$$K(P,\beta,\gamma) = -\sum_{k=1}^{N}[\ln P(k)]\,P(k) + \beta\left[\upsilon - \sum_{k=1}^{N}(\ln k)\,P(k)\right] + \gamma\left[1 - \sum_{k=1}^{N}P(k)\right]$$

# Solution using Lagrange multipliers

- Lagrange function

$$K(P,\beta,\gamma) = -\sum_{k=1}^{N} [\ln P(k)]\, P(k) + \beta \left[ \upsilon - \sum_{k=1}^{N} (\ln k)\, P(k) \right] + \gamma \left[ 1 - \sum_{k=1}^{N} P(k) \right]$$

- Necessary and sufficient conditions:

$$\frac{\partial}{\partial P} K(P,\beta,\gamma) = -\ln P(k) - 1 - \beta \ln k + \gamma = 0 \quad \Rightarrow \quad P(k) = e^{-\beta \ln k - \Gamma(\beta)}$$

# Solution using Lagrange multipliers

- Lagrange function

$$K(P,\beta,\gamma) = -\sum_{k=1}^{N}[\ln P(k)]\,P(k) + \beta\left[\upsilon - \sum_{k=1}^{N}(\ln k)\,P(k)\right] + \gamma\left[1 - \sum_{k=1}^{N}P(k)\right]$$

- Necessary and sufficient conditions:

$$\frac{\partial}{\partial P}K(P,\beta,\gamma) = -\ln P(k) - 1 - \beta\ln k + \gamma = 0 \quad\Rightarrow\quad P(k) = e^{-\beta\,\ln k - \Gamma(\beta)}$$

$$\frac{\partial}{\partial \beta}K(P,\beta,\gamma) = \upsilon - \sum_{k=1}^{N}(\ln k)\,P(k) = 0 \quad\Rightarrow\quad \sum_{k=1}^{N}(\ln k)\,P(k) = \upsilon$$

## Solution using Lagrange multipliers

- Lagrange function

$$K(P,\beta,\gamma) = -\sum_{k=1}^{N}[\ln P(k)]\,P(k) + \beta\left[\upsilon - \sum_{k=1}^{N}(\ln k)\,P(k)\right] + \gamma\left[1 - \sum_{k=1}^{N}P(k)\right]$$

- Necessary and sufficient conditions:

$$\frac{\partial}{\partial P}K(P,\beta,\gamma) = -\ln P(k) - 1 - \beta\ln k + \gamma = 0 \quad\Rightarrow\quad P(k) = e^{-\beta\ln k - \Gamma(\beta)}$$

$$\frac{\partial}{\partial \beta}K(P,\beta,\gamma) = \upsilon - \sum_{k=1}^{N}(\ln k)\,P(k) = 0 \quad\Rightarrow\quad \sum_{k=1}^{N}(\ln k)\,P(k) = \upsilon$$

$$\frac{\partial}{\partial \gamma}K(P,\beta,\gamma) = 1 - \sum_{k=1}^{N}P(k) = 0 \quad\Rightarrow\quad \Gamma(\beta) = \ln\sum_{k=1}^{N}e^{-\beta\ln k}$$

# Average path length (distance)

- $d(i,j)$ — distance between nodes $i$, $j$:

$$\ell := \mathbb{E}_P\{d(i,j)\}$$

# Average path length (distance)

- $d(i,j)$ — distance between nodes $i$, $j$:

$$\ell := \mathbb{E}_P\{d(i,j)\}$$

- Let $z_m$ denote the average number of neighbours at $d(i,j) = m$:

$$z_0 = 1, \quad z_1 = \mathbb{E}_P\{k\}, \quad z_2 = \mathbb{E}_P\{k(k-1)\}, \quad \dots, \quad z_{m+1} = \frac{z_2}{z_1} z_m$$

# Average path length (distance)

- $d(i,j)$ — distance between nodes $i$, $j$:

$$\ell := \mathbb{E}_P\{d(i,j)\}$$

- Let $z_m$ denote the average number of neighbours at $d(i,j) = m$:

$$z_0 = 1, \quad z_1 = \mathbb{E}_P\{k\}, \quad z_2 = \mathbb{E}_P\{k(k-1)\}, \quad \dots, \quad z_{m+1} = \frac{z_2}{z_1} z_m$$

- For graph size $N \gg z_1$ and $z_2 \gg z_1$ one can assume

$$N = \sum_{m=0}^{\ell} z_m$$

# Average path length (distance)

- $d(i,j)$ — distance between nodes $i$, $j$:

$$\ell := \mathbb{E}_P\{d(i,j)\}$$

- Let $z_m$ denote the average number of neighbours at $d(i,j) = m$:

$$z_0 = 1, \quad z_1 = \mathbb{E}_P\{k\}, \quad z_2 = \mathbb{E}_P\{k(k-1)\}, \quad \dots, \quad z_{m+1} = \frac{z_2}{z_1} z_m$$

- For graph size $N \gg z_1$ and $z_2 \gg z_1$ one can assume

$$N = 1 + \sum_{m=1}^{\ell} \left[\frac{z_2}{z_1}\right]^{m-1} z_1$$

# Average path length (distance)

- $d(i,j)$ — distance between nodes $i$, $j$:

$$\ell := \mathbb{E}_P\{d(i,j)\}$$

- Let $z_m$ denote the average number of neighbours at $d(i,j) = m$:

$$z_0 = 1, \quad z_1 = \mathbb{E}_P\{k\}, \quad z_2 = \mathbb{E}_P\{k(k-1)\}, \quad \dots, \quad z_{m+1} = \frac{z_2}{z_1} z_m$$

- For graph size $N \gg z_1$ and $z_2 \gg z_1$ one can assume

$$N \approx 1 + \left[\frac{z_2}{z_1}\right]^{\ell - 1} z_1$$

# Average path length (distance)

- $d(i,j)$ — distance between nodes $i$, $j$:

$$\ell := \mathbb{E}_P\{d(i,j)\}$$

- Let $z_m$ denote the average number of neighbours at $d(i,j) = m$:

$$z_0 = 1, \quad z_1 = \mathbb{E}_P\{k\}, \quad z_2 = \mathbb{E}_P\{k(k-1)\}, \quad \dots, \quad z_{m+1} = \frac{z_2}{z_1} z_m$$

- For graph size $N \gg z_1$ and $z_2 \gg z_1$ one can assume

$$N \approx 1 + \left[\frac{z_2}{z_1}\right]^{\ell - 1} z_1$$

- This leads to

$$\ell = \frac{\ln N - \ln z_1}{\ln(z_2/z_1)} + 1$$

# Average path length (distance)

- $d(i,j)$ — distance between nodes $i$, $j$:

$$\ell := \mathbb{E}_P\{d(i,j)\}$$

- Let $z_m$ denote the average number of neighbours at $d(i,j) = m$:

$$z_0 = 1, \quad z_1 = \mathbb{E}_P\{k\}, \quad z_2 = \mathbb{E}_P\{k(k-1)\}, \quad \dots, \quad z_{m+1} = \frac{z_2}{z_1} z_m$$

- For graph size $N \gg z_1$ and $z_2 \gg z_1$ one can assume

$$N \approx 1 + \left[\frac{z_2}{z_1}\right]^{\ell-1} z_1$$

- This leads to

$$\ell(k_i) = \frac{\ln N - \ln k_i}{\ln(z_2/z_1)} + 1$$

# Average path length (distance)

- $d(i,j)$ — distance between nodes $i$, $j$:

$$\ell := \mathbb{E}_P\{d(i,j)\}$$

- Let $z_m$ denote the average number of neighbours at $d(i,j) = m$:

$$z_0 = 1, \quad z_1 = \mathbb{E}_P\{k\}, \quad z_2 = \mathbb{E}_P\{k(k-1)\}, \quad \dots, \quad z_{m+1} = \frac{z_2}{z_1} z_m$$

- For graph size $N \gg z_1$ and $z_2 \gg z_1$ one can assume

$$N \approx 1 + \left[\frac{z_2}{z_1}\right]^{\ell - 1} z_1$$

- This leads to

$$\ell(k_i) \sim -\ln k_i$$

# Optimal communication

- Let $c(x_i, y_j)$ be some cost function for $x_i, y_j \in V$:

$$\text{minimize} \quad \mathbb{E}_P\{c(x_i, y_j)\} \quad \text{subject to} \quad I(x_i, y_j) \leq \lambda$$

# Optimal communication

- Let $c(x_i, y_j)$ be some cost function for $x_i, y_j \in V$:

$$\text{minimize} \quad \mathbb{E}_P\{c(x_i, y_j)\} \quad \text{subject to} \quad I(x_i, y_j) \leq \lambda$$

- where $I$ is Shannon's mutual information:

$$I(x_i, y_j) := \sum_{(x_i, y_j) \in X \times Y} \left[ \ln \frac{P(x_i, y_j)}{P(x_i)P(y_j)} \right] P(x_i, y_j)$$

# Optimal communication

- Let $c(x_i, y_j)$ be some cost function for $x_i, y_j \in V$:

$$\text{minimize} \quad \mathbb{E}_P\{c(x_i, y_j)\} \quad \text{subject to} \quad I(x_i, y_j) \leq \lambda$$

- where $I$ is Shannon's mutual information:

$$I(x_i, y_j) := \sum_{(x_i, y_j) \in X \times Y} \left[ \ln \frac{P(x_i, y_j)}{P(x_i)P(y_j)} \right] P(x_i, y_j)$$
$$= H\{x\} - H\{x \mid y\} = H\{y\} - H\{y \mid x\}$$

# Optimal communication

- Let $c(x_i, y_j)$ be some cost function for $x_i, y_j \in V$:

$$\text{minimize} \quad \mathbb{E}_P\{c(x_i, y_j)\} \quad \text{subject to} \quad I(x_i, y_j) \leq \lambda$$

- where $I$ is Shannon's mutual information:

$$I(x_i, y_j) := \sum_{(x_i, y_j) \in X \times Y} \left[ \ln \frac{P(x_i, y_j)}{P(x_i)P(y_j)} \right] P(x_i, y_j)$$

$$= H\{x\} - H\{x \mid y\} = H\{y\} - H\{y \mid x\}$$

- Solution

$$P(y_i \mid x_i) = e^{-\theta c(x_i, y_j) - \Gamma(\theta, x_i)} P(y_j)$$

# Optimal communication

- Let $c(x_i, y_j)$ be some cost function for $x_i, y_j \in V$:

$$\text{minimize} \quad \mathbb{E}_P\{c(x_i, y_j)\} \quad \text{subject to} \quad I(x_i, y_j) \leq \lambda$$

- where $I$ is Shannon's mutual information:

$$I(x_i, y_j) := \sum_{(x_i, y_j) \in X \times Y} \left[ \ln \frac{P(x_i, y_j)}{P(x_i) P(y_j)} \right] P(x_i, y_j)$$

$$= H\{x\} - H\{x \mid y\} = H\{y\} - H\{y \mid x\}$$

- Solution

$$P(y_i \mid x_i) = e^{-\theta c(x_i, y_j) - \Gamma(\theta, x_i)} P(y_j)$$

- $c(x_i, y_j) = \ell(k_i) - 1 \sim -\ln k_i$ yields preferential attachment:

$$P[(i, j) \in E \mid k_i] = e^{-\theta(\ell(k_i) - 1) - \Gamma(\theta)}$$

# Optimal communication

- Let $c(x_i, y_j)$ be some cost function for $x_i, y_j \in V$:

$$\text{minimize} \quad \mathbb{E}_P\{c(x_i, y_j)\} \quad \text{subject to} \quad I(x_i, y_j) \le \lambda$$

- where $I$ is Shannon's mutual information:

$$I(x_i, y_j) := \sum_{(x_i,y_j)\in X\times Y} \left[\ln \frac{P(x_i, y_j)}{P(x_i)P(y_j)}\right] P(x_i, y_j)$$

$$= H\{x\} - H\{x \mid y\} = H\{y\} - H\{y \mid x\}$$

- Solution

$$P(y_i \mid x_i) = e^{-\theta c(x_i,y_j) - \Gamma(\theta, x_i)} P(y_j)$$

- $c(x_i, y_j) = \ell(k_i) - 1 \sim -\ln k_i$ yields preferential attachment:

$$P[(i,j) \in E \mid k_i] = e^{\gamma \ln k_i - \Gamma(\gamma)}$$

# Optimal communication

- Let $c(x_i, y_j)$ be some cost function for $x_i, y_j \in V$:

$$\text{minimize} \quad \mathbb{E}_P\{c(x_i, y_j)\} \quad \text{subject to} \quad I(x_i, y_j) \leq \lambda$$

- where $I$ is Shannon's mutual information:

$$I(x_i, y_j) := \sum_{(x_i, y_j) \in X \times Y} \left[ \ln \frac{P(x_i, y_j)}{P(x_i) P(y_j)} \right] P(x_i, y_j)$$

$$= H\{x\} - H\{x \mid y\} = H\{y\} - H\{y \mid x\}$$

- Solution

$$P(y_i \mid x_i) = e^{-\theta c(x_i, y_j) - \Gamma(\theta, x_i)} P(y_j)$$

- $c(x_i, y_j) = \ell(k_i) - 1 \sim -\ln k_i$ yields preferential attachment:

$$P[(i, j) \in E \mid k_i] = \frac{k_i^{\gamma}}{\sum_{k=1}^{N} k_i^{\gamma}}$$

# Duality

## Power-law

$$P(k) = \frac{k^{-\beta}}{\sum_{k=1}^{N} k^{-\beta}}$$

Solution to

$$\text{maximize} \quad H(P) \quad \text{subject to} \quad \mathbb{E}_P\{\ln k\} \leq \upsilon$$

# Duality

## Power-law

$$P(k) = \frac{k^{-\beta}}{\sum_{k=1}^{N} k^{-\beta}}$$

Solution to

$$\text{maximize} \quad H(P) \quad \text{subject to} \quad \mathbb{E}_P\{\ln k\} \leq \upsilon$$

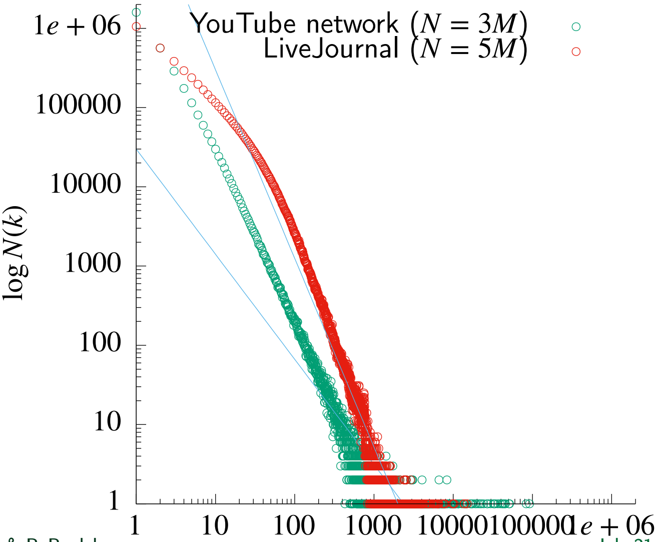
## Preferential attachment

$$P[(i,j) \in E \mid k_i)] = \frac{k_i^{\gamma}}{\sum_{k=1}^{N} k_i^{\gamma}}$$

Solution to

$$\text{minimize} \quad \mathbb{E}_P\{-\ln k\} \quad \text{subject to} \quad H\{x\} - H\{x \mid y\} \leq \lambda$$

# Exponent as slope



# Maximum likelihood estimation

- Treating $k$ as continuous, the m.l.e is (?, ?)

$$\beta = 1 + \frac{1}{\mathbb{E}_P\{\ln k\} - \ln k_0}$$

where $k_0$ is the smallest degree corresponding to power-law behaviour (i.e. $P(k_0) = \max P(k)$).

# Maximum likelihood estimation

- Treating $k$ as continuous, the m.l.e is (?, ?)

$$\beta = 1 + \frac{1}{\mathbb{E}_P\{\ln k\} - \ln k_0}$$

where $k_0$ is the smallest degree corresponding to power-law behaviour (i.e. $P(k_0) = \max P(k)$).

- Degree $k$ is discrete.

# Maximum likelihood estimation

- Treating $k$ as continuous, the m.l.e is (?, ?)

$$\beta = 1 + \frac{1}{\mathbb{E}_P\{\ln k\} - \ln k_0}$$

  where $k_0$ is the smallest degree corresponding to power-law behaviour (i.e. $P(k_0) = \max P(k)$).
- Degree $k$ is discrete.
- What about $\beta < 1$ (possible for $N < \infty$)?

# Variational approach

- Recall that $P(k) = \exp\{-\beta \ln k - \Gamma(\beta)\}$ is the solution to the maximum entropy problem, where $\beta \geq 0$ is the Lagrange multiplier such that the constraint $\mathbb{E}_P\{\ln k\} \leq \upsilon$ (or $H(P) \geq \ln N - \lambda$) is satisfied with equality:

$$\mathbb{E}_P\{\ln k\} = -\Gamma'(\beta)$$
$$H(P) = \Gamma(\beta) - \beta\,\Gamma'(\beta)$$

## Variational approach

- Recall that $P(k) = \exp\{-\beta \ln k - \Gamma(\beta)\}$ is the solution to the maximum entropy problem, where $\beta \geq 0$ is the Lagrange multiplier such that the constraint $\mathbb{E}_P\{\ln k\} \leq \upsilon$ (or $H(P) \geq \ln N - \lambda$) is satisfied with equality:

$$\mathbb{E}_P\{\ln k\} = -\Gamma'(\beta)$$
$$H(P) = \Gamma(\beta) - \beta\,\Gamma'(\beta)$$

- This leads to

$$\beta = \frac{H(P) - \Gamma(\beta)}{\mathbb{E}_P\{\ln k\}}$$

## Variational approach

- Recall that $P(k) = \exp\{-\beta \ln k - \Gamma(\beta)\}$ is the solution to the maximum entropy problem, where $\beta \geq 0$ is the Lagrange multiplier such that the constraint $\mathbb{E}_P\{\ln k\} \leq \upsilon$ (or $H(P) \geq \ln N - \lambda$) is satisfied with equality:

$$\mathbb{E}_P\{\ln k\} = -\Gamma'(\beta)$$
$$H(P) = \Gamma(\beta) - \beta\,\Gamma'(\beta)$$

- This leads to

$$\beta = \frac{H(P) - \Gamma(\beta)}{\mathbb{E}_P\{\ln k\}}$$

- We can estimate $\mathbb{E}_P\{\ln k\}$, $H(P)$.

# Variational approach

- Recall that $P(k) = \exp\{-\beta \ln k - \Gamma(\beta)\}$ is the solution to the maximum entropy problem, where $\beta \geq 0$ is the Lagrange multiplier such that the constraint $\mathbb{E}_P\{\ln k\} \leq \upsilon$ (or $H(P) \geq \ln N - \lambda$) is satisfied with equality:

$$\mathbb{E}_P\{\ln k\} = -\Gamma'(\beta)$$
$$H(P) = \Gamma(\beta) - \beta\,\Gamma'(\beta)$$

- This leads to

$$\beta = \frac{H(P) - \Gamma(\beta)}{\mathbb{E}_P\{\ln k\}}$$

- We can estimate $\mathbb{E}_P\{\ln k\}$, $H(P)$.
- Observe that $\Gamma(\beta) = -\ln P(k = 1)$.

# Variational approach

- Recall that $P(k) = \exp\{-\beta \ln k - \Gamma(\beta)\}$ is the solution to the maximum entropy problem, where $\beta \geq 0$ is the Lagrange multiplier such that the constraint $\mathbb{E}_P\{\ln k\} \leq \upsilon$ (or $H(P) \geq \ln N - \lambda$) is satisfied with equality:

$$\mathbb{E}_P\{\ln k\} = -\Gamma'(\beta)$$
$$H(P) = \Gamma(\beta) - \beta\,\Gamma'(\beta)$$

- This leads to

$$\beta = \frac{H(P) - \Gamma(\beta)}{\mathbb{E}_P\{\ln k\}}$$

- We can estimate $\mathbb{E}_P\{\ln k\}$, $H(P)$.
- Observe that $\Gamma(\beta) = -\ln P(k = 1)$.
- Making the transofrmation $k \mapsto k/k_0$ leads to

$$\beta = \frac{H(P) + \ln P(k_0)}{\mathbb{E}_P\{\ln k\} - \ln k_0}$$

# Exponent (inverse temperature)

- Recall the Lagrangian

$$K(P, \beta, \gamma) = H(P) + \beta[\upsilon - \mathbb{E}_P\{\ln k\}] + \gamma \left[1 - \sum P(k)\right]$$

# Exponent (inverse temperature)

- Recall the Lagrangian

$$K(P,\beta,\gamma) = H(P) + \beta[\upsilon - \mathbb{E}_P\{\ln k\}] + \gamma\left[1 - \sum P(k)\right] = \overline{H}(\upsilon)$$

- Its value at $P(k) = \exp\{-\beta \ln k - \Gamma(\beta)\}$ is the maximum entropy, which depends on the constraint $\upsilon = \mathbb{E}_P\{\ln k\}$.

# Exponent (inverse temperature)

- Recall the Lagrangian

$$K(P,\beta,\gamma)=H(P)+\beta[\upsilon-\mathbb{E}_P\{\ln k\}]+\gamma\left[1-\sum P(k)\right]=\overline{H}(\upsilon)$$

- Its value at $P(k)=\exp\{-\beta\,\ln k-\Gamma(\beta)\}$ is the maximum entropy, which depends on the constraint $\upsilon=\mathbb{E}_P\{\ln k\}$.
- Differentiation gives:

$$\beta=\frac{\partial\overline{H}(\upsilon)}{\partial\upsilon}$$

# Exponent (inverse temperature)

- Recall the Lagrangian

$$K(P,\beta,\gamma) = H(P) + \beta[\upsilon - \mathbb{E}_P\{\ln k\}] + \gamma\left[1 - \sum P(k)\right] = \overline{H}(\upsilon)$$

- Its value at $P(k) = \exp\{-\beta\ \ln k - \Gamma(\beta)\}$ is the maximum entropy, which depends on the constraint $\upsilon = \mathbb{E}_P\{\ln k\}$.
- Differentiation gives:

$$\beta = \frac{\partial \overline{H}(\upsilon)}{\partial \upsilon}$$

- Compare with our formula

$$\beta = \frac{H(P) - \ln P^{-1}(k_0)}{\mathbb{E}_P\{\ln k\} - \ln k_0} = \frac{\Delta H}{\Delta \upsilon}$$

Analysis of large networks

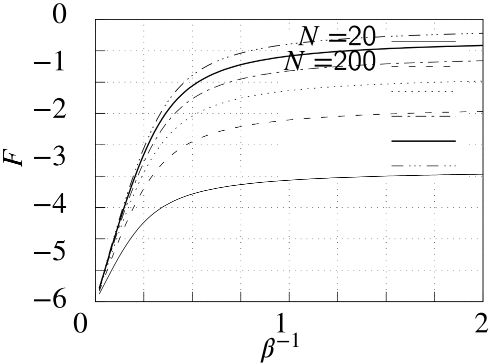
The power-law degree sequence

Power-law and maximum entropy

Power-law exponent (inverse temperature)

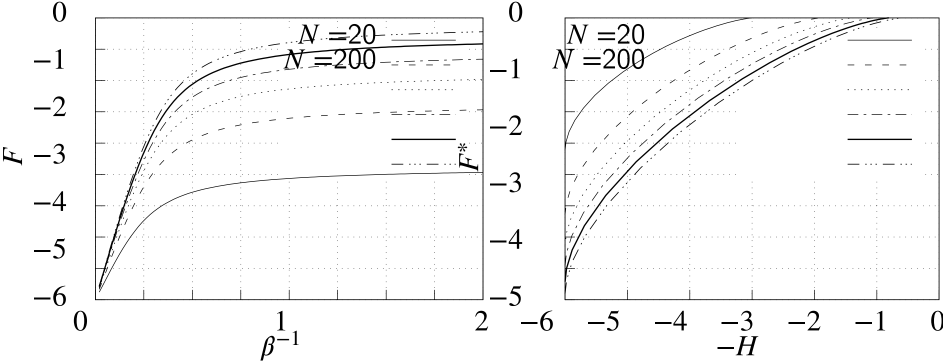
Free energy and phase transition

# Free energy



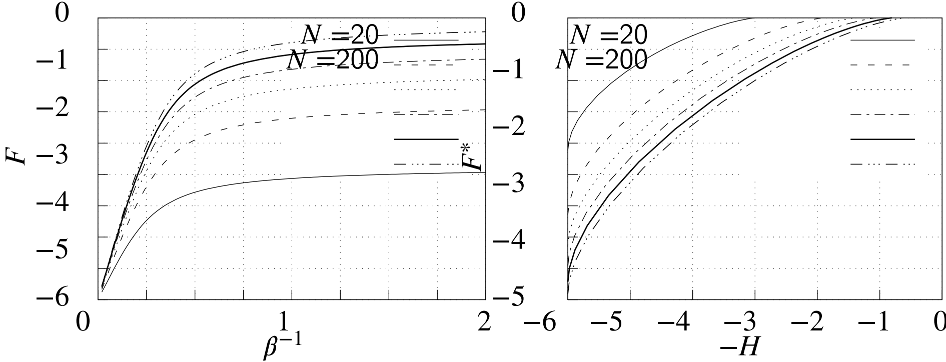

$$F(\beta^{-1}) := -\beta^{-1}\Gamma(\beta)$$

# Free energy




$F(\beta^{-1}) := -\beta^{-1}\Gamma(\beta)$ $\qquad$ $F^*(I) := \inf\{\beta^{-1}I - F(\beta^{-1})\}$

# Free energy

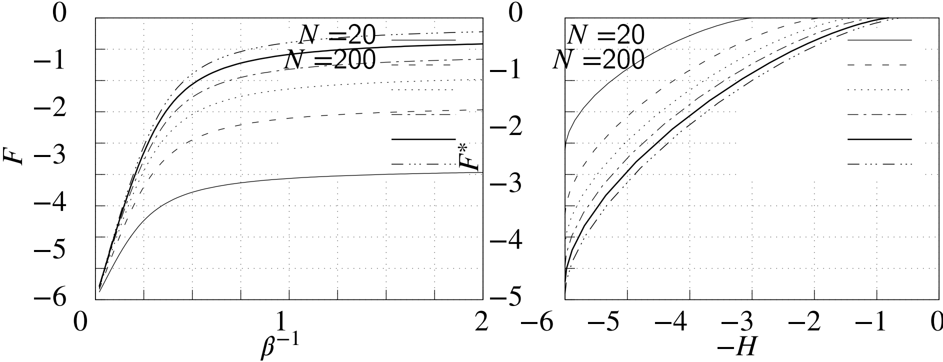



$$F(\beta^{-1}) := -\beta^{-1}\Gamma(\beta)$$

$$F^*(I) := \inf\{\beta^{-1} I - F(\beta^{-1})\}$$

$$I = F'(\beta^{-1})$$

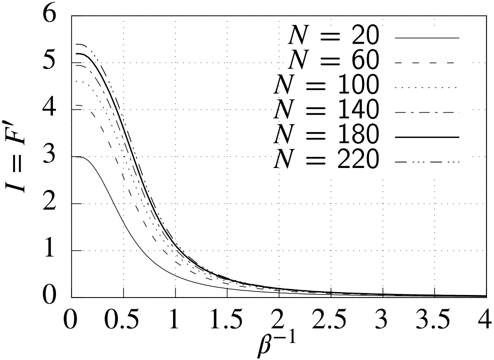
# Free energy




$$F(\beta^{-1}) := -\beta^{-1}\Gamma(\beta) \qquad F^*(I) := \inf\{\beta^{-1}I - F(\beta^{-1})\}$$

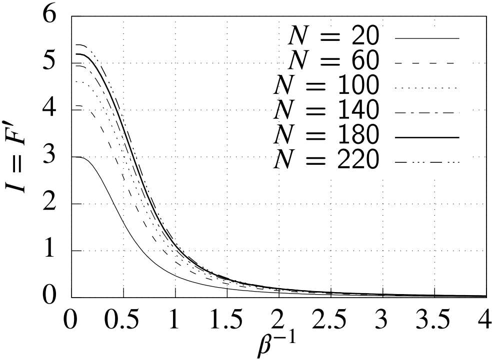
$$F^{*\prime}(I) = \beta^{-1} \qquad I = F'(\beta^{-1})$$

# Information and entropy at $\beta = 1$



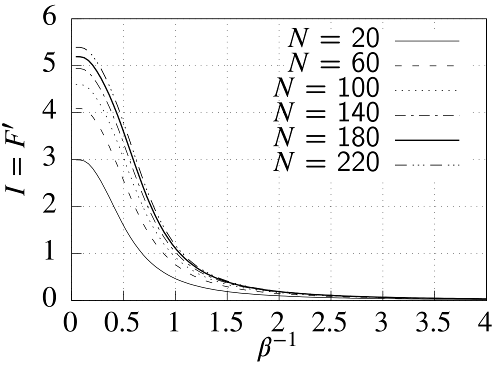

- $H(\beta) = \ln N - I(\beta)$

# Information and entropy at $\beta = 1$




- $H(\beta) = \ln N - I(\beta)$
- $I(\beta) = \ln N - H(\beta)$

# Information and entropy at $\beta = 1$

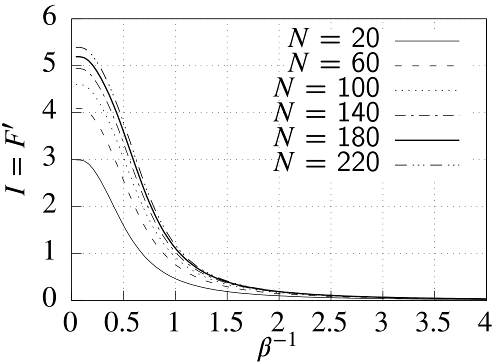



- $H(\beta) = \ln N - I(\beta)$
- $I(\beta) = \ln N - H(\beta)$

## Question

- *What happens to $F'(\beta^{-1}) = I$ in the limit $N \to \infty$ and $\beta \to 1$?*

# Information and entropy at $\beta = 1$

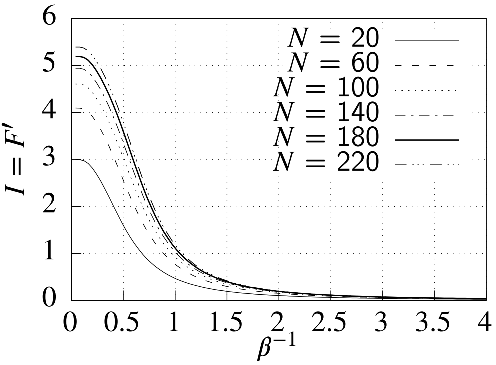



- $H(\beta) = \ln N - I(\beta)$
- $I(\beta) = \ln N - H(\beta)$

## Question

- *What happens to* $F'(\beta^{-1}) = I$ *in the limit* $N \to \infty$ *and* $\beta \to 1$*?*
- $I = \beta\,\Gamma'(\beta) - \Gamma(\beta)$

# Information and entropy at $\beta = 1$




- $H(\beta) = \ln N - I(\beta)$
- $I(\beta) = \ln N - H(\beta)$

## Question

- *What happens to* $F'(\beta^{-1}) = I$ *in the limit* $N \to \infty$ *and* $\beta \to 1$*?*
- $I = \beta\, \Gamma'(\beta) - \Gamma(\beta)$
- $\Gamma'(\beta) = -\mathbb{E}_P\{\ln k\}$, $\Gamma''(\beta) = \sigma^2(\ln k)$.

# Approximations at $\beta = 1$ and $N < \infty$

- $n$th cumulants $\Gamma^{(n)} = (\ln Z)^{(n)}$:

$$\begin{aligned}
\Gamma' &= m_1 \\
\Gamma'' &= m_2 - m_1^2 \\
\Gamma^{(3)} &= m_3 - 3m_1 m_2 + 2m_1^3 \\
\Gamma^{(4)} &= m_4 - 4m_3 m_1 - 3m_2^2 + 12m_2 m_1^2 - 6m_1^4
\end{aligned}$$

# Approximations at $\beta = 1$ and $N < \infty$

- $n$th cumulants $\Gamma^{(n)} = (\ln Z)^{(n)}$:

$$\Gamma' = m_1$$
$$\Gamma'' = m_2 - m_1^2$$
$$\Gamma^{(3)} = m_3 - 3m_1m_2 + 2m_1^3$$
$$\Gamma^{(4)} = m_4 - 4m_3m_1 - 3m_2^2 + 12m_2m_1^2 - 6m_1^4$$

- where $m_n$ are $n$th moments of $-\ln k$:

$$m_n(\beta) = \frac{Z^{(n)}(\beta)}{Z(\beta)}, \qquad Z^{(n)}(\beta) = (-1)^n \sum_{k=1}^{N} (\ln k)^n k^{-\beta}$$

# Approximations at $\beta = 1$ and $N < \infty$

- $n$th cumulants $\Gamma^{(n)} = (\ln Z)^{(n)}$:

$$\Gamma' = m_1$$
$$\Gamma'' = m_2 - m_1^2$$
$$\Gamma^{(3)} = m_3 - 3m_1 m_2 + 2m_1^3$$
$$\Gamma^{(4)} = m_4 - 4m_3 m_1 - 3m_2^2 + 12 m_2 m_1^2 - 6m_1^4$$

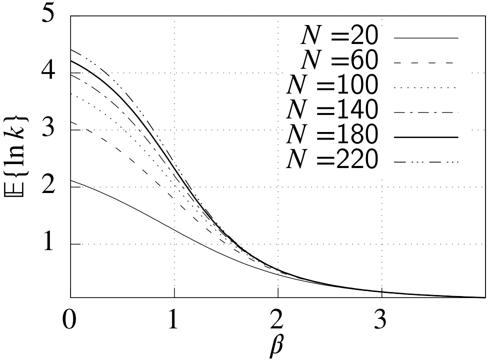
- where $m_n$ are $n$th moments of $-\ln k$:

$$m_n(\beta) = \frac{Z^{(n)}(\beta)}{Z(\beta)}, \qquad Z^{(n)}(\beta) = (-1)^n \sum_{k=1}^{N} (\ln k)^n k^{-\beta}$$

- Using $\int_1^N \frac{dx}{x} = \ln N$:

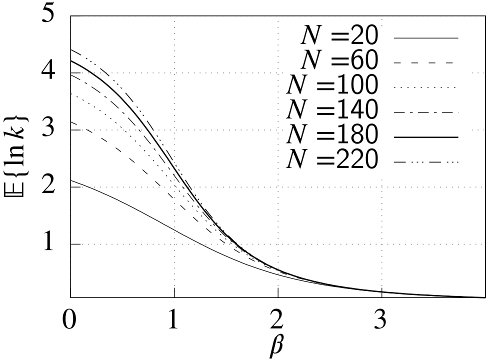
$$Z(\beta) = \sum_{k=1}^{N} \frac{1}{k^\beta}\bigg|_{\beta=1} \approx \ln N, \quad Z^{(n)}(\beta)\Big|_{\beta=1} \approx \frac{(-1)^n}{n+1} (\ln N)^{n+1}$$

# Expectation of $\ln k$ at $\beta = 1$



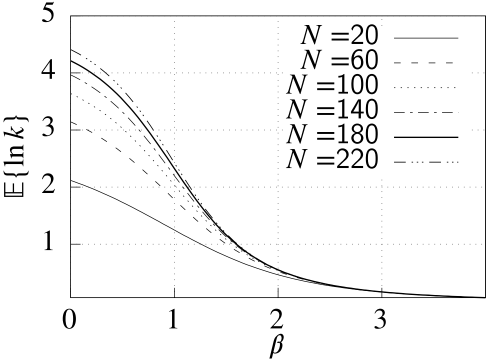

$$\mathbb{E}_P\{\ln k\} = -\Gamma'(\beta)$$

# Expectation of $\ln k$ at $\beta = 1$



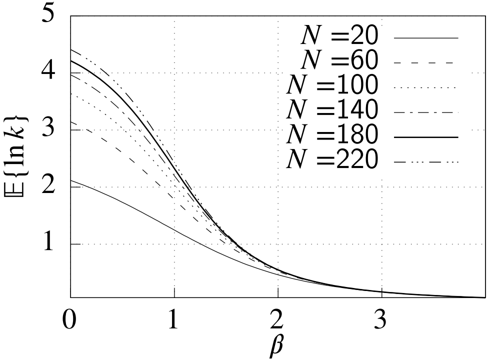

$$\mathbb{E}_P\{\ln k\} = -\,\Gamma'(\beta) = -\,\frac{Z'(\beta)}{Z(\beta)}\bigg|_{\beta=1}$$

# Expectation of $\ln k$ at $\beta = 1$




$$
\begin{aligned}
\mathbb{E}_P\{\ln k\} &= -\,\Gamma'(\beta) \\
&= -\left.\frac{Z'(\beta)}{Z(\beta)}\right|_{\beta=1} \\
&\approx \frac{1}{2}\ln N
\end{aligned}
$$

# Expectation of $\ln k$ at $\beta = 1$

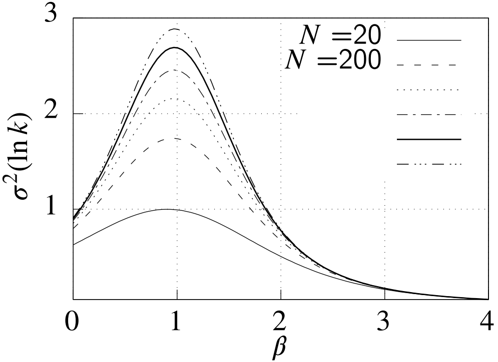



$$\mathbb{E}_P\{\ln k\} = -\Gamma'(\beta) = -\left.\frac{Z'(\beta)}{Z(\beta)}\right|_{\beta=1} \approx \frac{1}{2}\ln N$$

## Remark

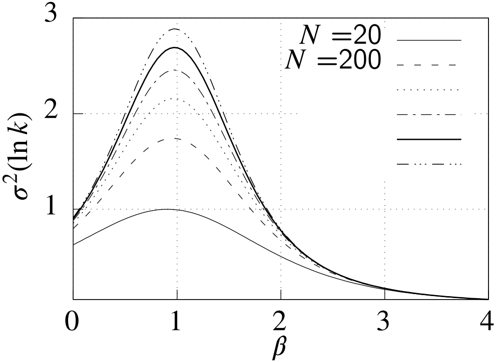
*Using Jenssen's inequality* $\ln \mathbb{E}_P\{k\} \geq \mathbb{E}_P\{\ln k\}$ *we also have* $\mathbb{E}_P\{k\} \geq \sqrt{N}$.

# Variance of $\ln k$ at $\beta = 1$




$$\sigma^2(\ln k) = \Gamma''(\beta)\big|_{\beta=1}$$
$$\approx \frac{1}{12}(\ln N)^2$$

# Variance of $\ln k$ at $\beta = 1$

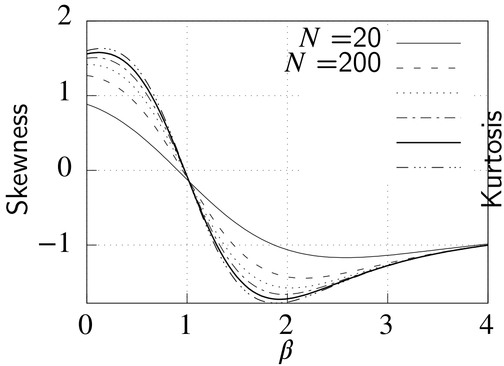



$$\sigma^2(\ln k) = \Gamma''(\beta)\big|_{\beta=1}$$
$$\approx \frac{1}{12}(\ln N)^2$$

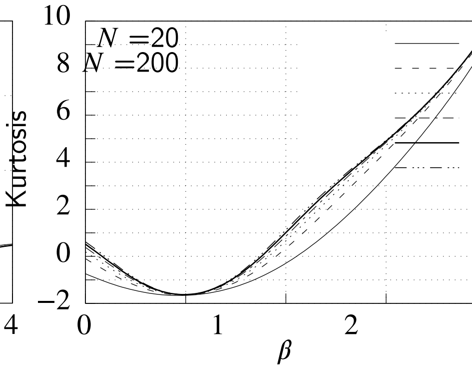
## Remark (Phase transition)

*The derivative $F'(\beta^{-1}) = I(\beta) = \beta\,\Gamma'(\beta) - \Gamma(\beta)$ is not differentiable at $\beta = 1$ in the limit $N \to \infty$, because $\Gamma''(\beta) \to \infty$.*

# Skewness and kurtosis at $\beta = 1$




$$\frac{\Gamma^{(3)}}{(\Gamma'')^{3/2}}$$




$$\frac{\Gamma^{(4)}}{(\Gamma'')^{2}}$$

# Conclusions and discussion

- The power-law topology emerges as the solutions to the maximum entropy problem with constraint $\mathbb{E}\{\ln k\} \leq \upsilon$.

## Conclusions and discussion

- The power-law topology emerges as the solutions to the maximum entropy problem with constraint $\mathbb{E}\{\ln k\} \leq \upsilon$.
- The negative log-degree $-\ln k$ is related to average path length $\ell(k)$.

# Conclusions and discussion

- The power-law topology emerges as the solutions to the maximum entropy problem with constraint $\mathbb{E}\{\ln k\} \leq \upsilon$.
- The negative log-degree $-\ln k$ is related to average path length $\ell(k)$.
- The preferential attachment emerges as the solution to the dual problem problem on minimization of the average path length with a constraint on mutual information.

# Conclusions and discussion

- The power-law topology emerges as the solutions to the maximum entropy problem with constraint $\mathbb{E}\{\ln k\} \leq \upsilon$.
- The negative log-degree $-\ln k$ is related to average path length $\ell(k)$.
- The preferential attachment emerges as the solution to the dual problem problem on minimization of the average path length with a constraint on mutual information.
- These principles can be used to optimize various types of networks:

## Conclusions and discussion

- The power-law topology emerges as the solutions to the maximum entropy problem with constraint $\mathbb{E}\{\ln k\} \leq \upsilon$.
- The negative log-degree $-\ln k$ is related to average path length $\ell(k)$.
- The preferential attachment emerges as the solution to the dual problem problem on minimization of the average path length with a constraint on mutual information.
- These principles can be used to optimize various types of networks:
    1. Define the network $G(V, E)$.

## Conclusions and discussion

- The power-law topology emerges as the solutions to the maximum entropy problem with constraint $\mathbb{E}\{\ln k\} \leq \upsilon$.
- The negative log-degree $-\ln k$ is related to average path length $\ell(k)$.
- The preferential attachment emerges as the solution to the dual problem problem on minimization of the average path length with a constraint on mutual information.
- These principles can be used to optimize various types of networks:
  1. Define the network $G(V, E)$.
  2. Define the cost function $c : V \times V \to \mathbb{R}$ (or utility $u = -c$).

# Conclusions and discussion

- The power-law topology emerges as the solutions to the maximum entropy problem with constraint $\mathbb{E}\{\ln k\} \leq \upsilon$.
- The negative log-degree $-\ln k$ is related to average path length $\ell(k)$.
- The preferential attachment emerges as the solution to the dual problem problem on minimization of the average path length with a constraint on mutual information.
- These principles can be used to optimize various types of networks:
  1. Define the network $G(V, E)$.
  2. Define the cost function $c : V \times V \to \mathbb{R}$ (or utility $u = -c$).
  3. Analyse conditional expected cost $f(k, \dots) = \mathbb{E}\{c \mid k, \dots\}$ as a function of network parameters (e.g. degree $k$).

## Conclusions and discussion

- The power-law topology emerges as the solutions to the maximum entropy problem with constraint $\mathbb{E}\{\ln k\} \leq \upsilon$.
- The negative log-degree $-\ln k$ is related to average path length $\ell(k)$.
- The preferential attachment emerges as the solution to the dual problem problem on minimization of the average path length with a constraint on mutual information.
- These principles can be used to optimize various types of networks:
  1. Define the network $G(V, E)$.
  2. Define the cost function $c : V \times V \to \mathbb{R}$ (or utility $u = -c$).
  3. Analyse conditional expected cost $f(k, \dots) = \mathbb{E}\{c \mid k, \dots\}$ as a function of network parameters (e.g. degree $k$).
  4. Generate the network using preferential attachment: $P(e \mid k) \propto e^{-\beta f(k)}$.

# Conclusions and discussion

- The power-law topology emerges as the solutions to the maximum entropy problem with constraint $\mathbb{E}\{\ln k\} \leq \upsilon$.
- The negative log-degree $-\ln k$ is related to average path length $\ell(k)$.
- The preferential attachment emerges as the solution to the dual problem problem on minimization of the average path length with a constraint on mutual information.
- These principles can be used to optimize various types of networks:
  1. Define the network $G(V, E)$.
  2. Define the cost function $c : V \times V \to \mathbb{R}$ (or utility $u = -c$).
  3. Analyse conditional expected cost $f(k, \dots) = \mathbb{E}\{c \mid k, \dots\}$ as a function of network parameters (e.g. degree $k$).
  4. Generate the network using preferential attachment: $P(e \mid k) \propto e^{-\beta f(k)}$.
  5. Tune parameter $\beta$ (inverse temperature) based on the constraints.

# Conclusions and discussion

- The power-law topology emerges as the solutions to the maximum entropy problem with constraint $\mathbb{E}\{\ln k\} \leq \upsilon$.
- The negative log-degree $-\ln k$ is related to average path length $\ell(k)$.
- The preferential attachment emerges as the solution to the dual problem problem on minimization of the average path length with a constraint on mutual information.
- These principles can be used to optimize various types of networks:
  1. Define the network $G(V, E)$.
  2. Define the cost function $c : V \times V \to \mathbb{R}$ (or utility $u = -c$).
  3. Analyse conditional expected cost $f(k, \dots) = \mathbb{E}\{c \mid k, \dots\}$ as a function of network parameters (e.g. degree $k$).
  4. Generate the network using preferential attachment: $P(e \mid k) \propto e^{-\beta f(k)}$.
  5. Tune parameter $\beta$ (inverse temperature) based on the constraints.
- This approach leads to new insights into properties of the power-law graphs (e.g. a phase transition at $\beta = 1$ characterized by maximum variance, zero skewness and minimum kurtosis of $\ln k$).

Analysis of large networks

The power-law degree sequence

Power-law and maximum entropy

Power-law exponent (inverse temperature)

Free energy and phase transition

Abello, J., Pardalos, P. M., & Resende, M. G. C. (1999). On maximum clique problems in very large graphs. In J. Abello & J. Vitter (Eds.), *External memory algorithms and visualization* (Vol. 50, pp. 119–130). American Mathematical Society.

Aiello, W., Chung, F., & Lu, L. (2000). A random graph model for massive graphs. In *Proceedings of the 32nd Annual ACM Symposium on Theory of Computing* (pp. 171–180). Portland, OR: ACM Press.

Aiello, W., Chung, F., & Lu, L. (2001). A random graph model for power law graphs. *Experimental Mathematics*, *10*(1), 53–66.

Barabási, A.-L., & Albert, R. (1999). Emergence of scaling in random networks. *Science*, *286*(5439), 509–512.

Molloy, M., & Reed, B. (1995). A critical point for random graphs with a given degree sequence. *Random Structures and Algorithms*, *6*, 161–180.

Newman, M. (2005). Power laws, Pareto distributions and Zipf's law. *Contemporary Physics*, *46*(5), 323–351.