Power laws

Yury Dvorkin

February 5, 2020

Table of contents

- Open-ended question, of course
- The internet-of-things comes to mind: technological, biological, social, etc network

- Continuous random variable X
- Probability density function $p(x)$ (PDF):

$$Pr(a \leq X \leq b) = \int_a^b p(x)\ dx$$

$$p(x) \geq 0$$

$$\int_{-\infty}^{\infty} p(x)\ = 1$$

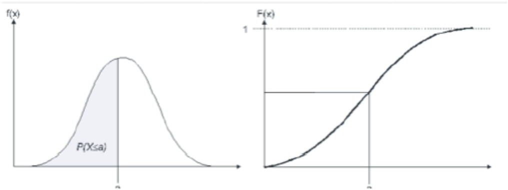
- Cumulative distribution function (CDF)

$$F(x) = Pr(X \leq x) = \int_{-\infty}^{x} p(x)\ ; \frac{d}{dx} F(x) = p(x)$$

- Complementary cumulative distribution function (cCDF)

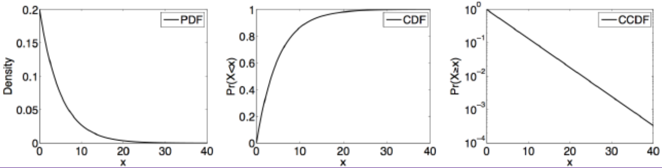
$$\bar{F}(x) = Pr(X \geq x) = 1 - F(x) = \int_x^{\infty} p(x)\, dx$$

- Gaussian: $p(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2})}$, $F(x) = \frac{1}{2}[1 + erf(\frac{x-\mu}{\sigma\sqrt{2}})]$



- Exponential $(x \geq 0)$:

$$p(x) = \lambda e^{-\lambda x},\ F(x) = 1 - e^{-\lambda x}, \bar{F}(x) = e^{-\lambda x}$$



- Discrete random variable Xi
- Probability mass function (PMF) $p(x)$:

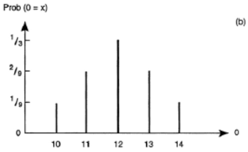
$$p(x) = Pr(Xi = x)$$

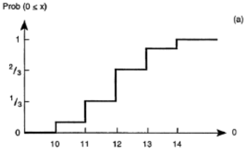
$$p(x) \geq 0$$

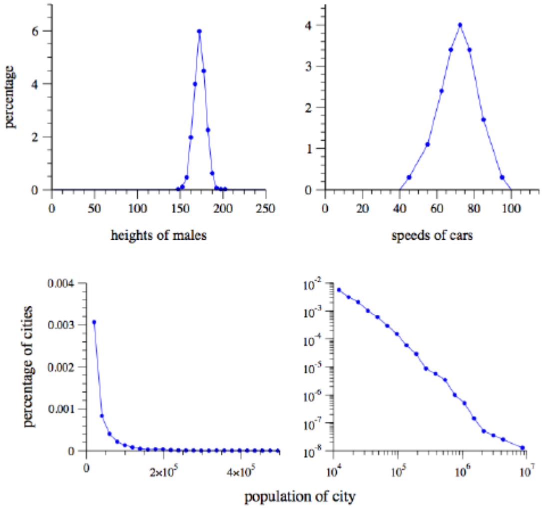
$$\sum_x p(x) = 1$$

- Cumulative distribution function (CDF)

$$F(x) = Pr(Xi \leq x) = \sum_{x' \leq x} p(x')$$






percentage
0
2
4
6
0
50
100
150
200
250
heights of males
0
1
2
3
4
0
20
40
60
80
100
speeds of cars
percentage of cities
0.004
0.003
0.002
0.001
0
0
2×10^5
4×10^5
10^-2
10^-3
10^-4
10^-5
10^-6
10^-7
10^-8
10^4
10^5
10^6
10^7
population of city

Continuous approximation

- Power law

$$p(x) = Cx^{-\alpha} = \frac{C}{x^{\alpha}}, \ \mathit{for}\ x \geq x_{min}$$

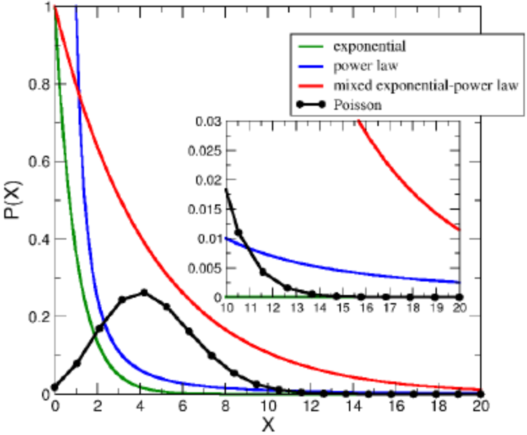
- Normalization ($\alpha > 1$)

$$1 = \int_{x_{min}}^{\infty} p(x)dx = C \int_{x_{min}}^{\infty} \frac{dx}{x^{\alpha}} = \frac{C}{\alpha - 1} x_{min}^{-\alpha+1}$$

$$C = (\alpha - 1)x_{min}^{-\alpha+1}$$

- Power law PDF

$$p(x) = \frac{\alpha - 1}{x_{min}} \left(\frac{x}{x_{min}} \right)^{-\alpha}$$



- poisson:$p(x) = \frac{\lambda^k}{k!} e^{-\lambda}$, $exponent : p(x) = Ce^{-\lambda x}$, $powerlaw :$ $p(x) = Cx^{-\alpha}$

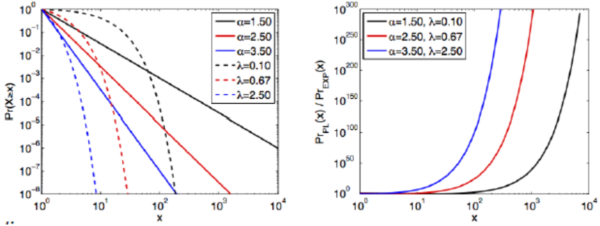
- Power law PDF

$$p(x) = Cx^{-\alpha} = \frac{\alpha - 1}{x_{min}} \left(\frac{x}{x_{min}} \right)^{-\alpha}$$

- Complimentary cumulative distribution function cCDF

$$\bar{F}(x) = Pr(X > x) = \int_x^{\infty} p(x) dx$$

$$\bar{F}(x) = \bar{C} x^{-(\alpha-1)} = \frac{C}{\alpha - 1} x^{-(\alpha-1)} = \left(\frac{x}{x_{min}} \right)^{-(\alpha-1)}$$

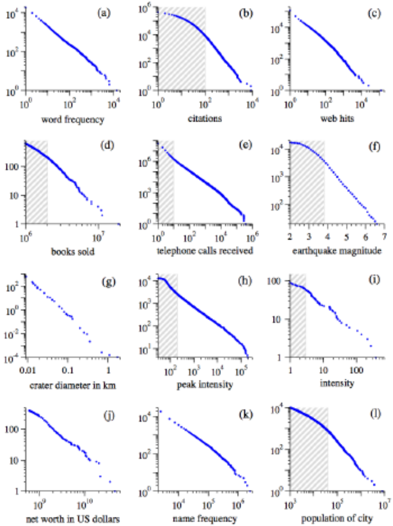


- Power law:

$$p(x) = Cx^{-\alpha}, \bar{F}(x) = \bar{C}x^{-(\alpha-1)}$$

$$\log p(x) = \log C - \alpha \log x, \log \bar{F}(x) = \log C - (\alpha - 1) \log x$$

- log-log scale



- PDF

$$p(x) = \frac{C}{x^{\alpha}}, x \geq x_{min}$$

- First moment (mean value),$\alpha \geq 2$:

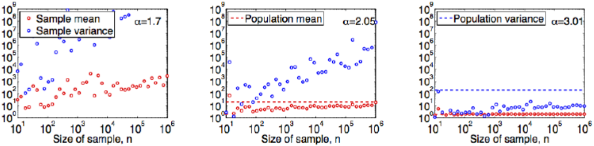
$$\langle x \rangle = \int_{x_{min}}^{\infty} xp(x)dx = C \int_{x_{min}}^{\infty} \frac{\alpha - 1}{\alpha - 2} x_{min}$$ <- typo

- Second moment, $\alpha > 3$:

$$\langle x^2 \rangle = \int_{x_{min}}^{\infty} x^2 p(x)dx = C \int_{x_{min}}^{\infty} \frac{dx}{x^{\alpha-2}} = \frac{\alpha - 1}{\alpha - 3} x_{min}^2$$

- k-th moment, $\alpha > k + 1$:

$$\langle x^k \rangle = \frac{\alpha - 1}{\alpha - 1 - k} x_{min}^k$$



- Fisrt moment (mean):

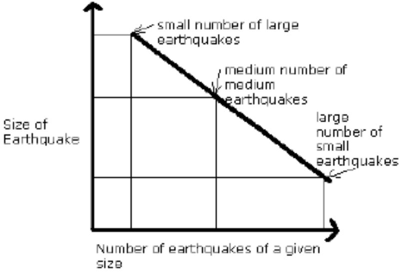
$$\langle x \rangle = C \int_{x_{min}}^{x_{max}} \frac{dx}{x^{\alpha-1}} = \frac{\alpha-1}{\alpha-2} \left(x_{min} - \frac{x_{min}^{\alpha-1}}{x_{max}^{\alpha-2}} \right)$$

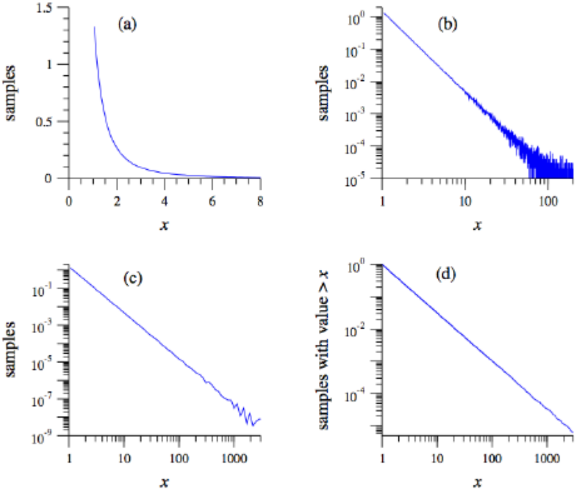
- Scaling of the density

$$x \to bx, p(bx) = C(bx)^{-\alpha} = b^{\alpha} C x^{-\alpha} \propto p(x)$$

- Scale invariance

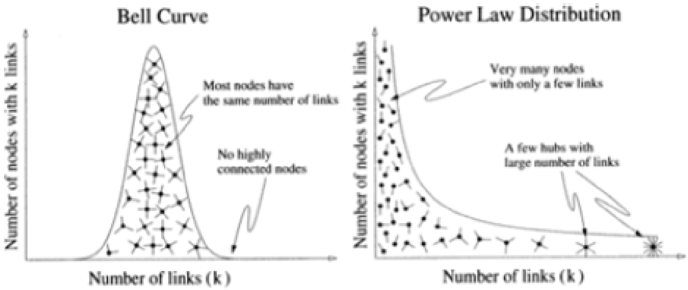
$$\frac{p(100x)}{p(10x)} = \frac{p(10x)}{p(x)}$$





Newman et.al, 2005

Scale-free networks



- ki - node degree, i.e. number of nearest neighbors, $k_i = 1, 2, \ldots k_{max}$
- n_k - number of nodes with degree k,$n_k = \sum_i \mathcal{I}(k_i == k)$
- total number of nodes $n = \sum_k n_k$
- Degree distribution $P(k_i = k) \equiv P(k)$

$$P(k) = \frac{n_k}{\sum_k n_k} = \frac{n_k}{n}$$

- CDF

$$F(k) = \sum_{k' \leq k} P(k') = \frac{1}{n} \sum_{k' \leq k} n_{k'}$$

- cCDF

$$F(k) = 1 - \sum_{k' \leq k} P(k') = \frac{1}{n} \sum_{k' \geq k} n_{k'}$$

- Power law distribution

$$p(k) = Ck^{-\gamma} = \frac{C}{k^{\gamma}}$$

- Normalization

$$\sum_{k=1}^{\infty} P(k) = C \sum_{k=1}^{\infty} k^{-\gamma} = C\zeta(\gamma) = 1; C = \frac{1}{\zeta(\gamma)}$$

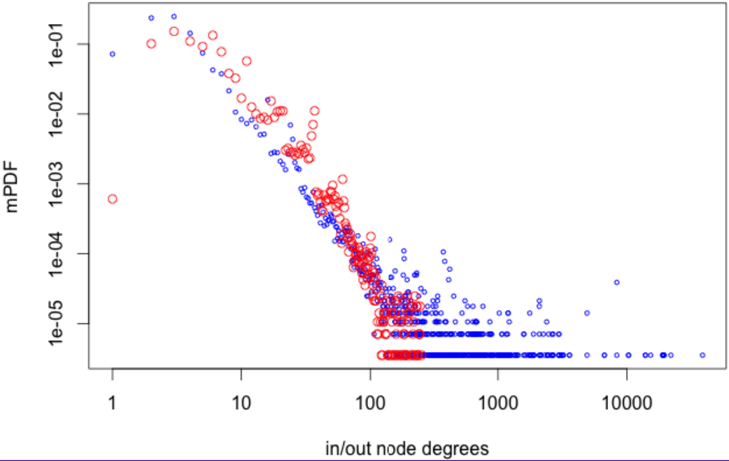
- Riemann zeta function, $\gamma > 1$

$$P(k) = \frac{k^{-\gamma}}{\zeta(\gamma)}$$

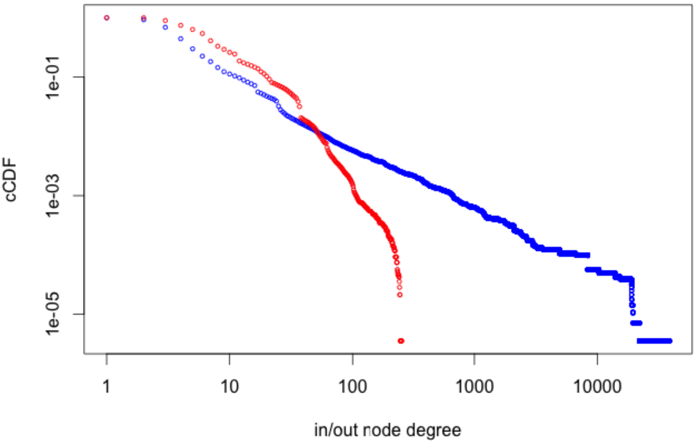
- Log-log coordinates

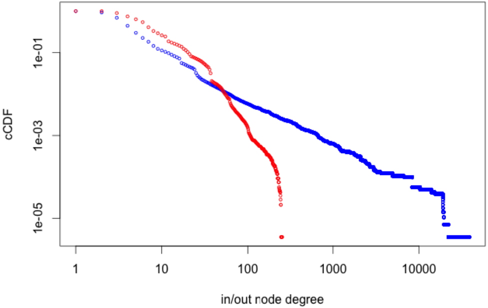
$$\log(P(k)) = -\gamma \log k + \log C$$

Probability mass function PMF/mPDF



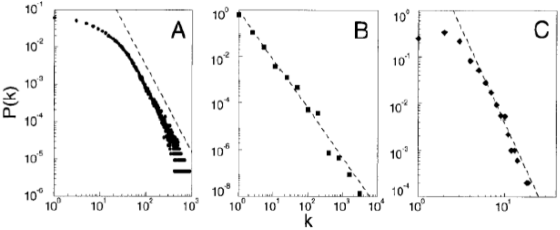
Complementary cumulative distribution function cCDF





Actor collaboration graph, N=212,250 nodes, $\langle k \rangle = 28.8, \gamma = 2.3$
WWW, N $=$ 325,729 nodes, $\langle k \rangle = 5.6, \gamma = 2.1$
Power grid data, N $=$ 4941 nodes, $\langle k \rangle = 5.5, \gamma = 4$

Barabasi et.al, 1999



In- and out- degrees of WWW crawl 1999

Broder et.al, 1999

Maximum likelihood estimation of parameter α

- Let $\{x_i\}$ be a set of n observations (points) independently sampled from the distribution

$$P(x_i) = \frac{\alpha - 1}{x_{min}} \left(\frac{x_i}{x_{min}} \right)^{-\alpha}$$

- Probability of the sample

$$P(\{x_i\}|\alpha) = \prod_i^n \frac{\alpha - 1}{x_{min}} \left(\frac{x_i}{x_{min}} \right)^{-\alpha}$$

- Bayes' theorem

$$P(\alpha|\{x_i\}) = P(\{x_i\}|\alpha) \frac{P(\alpha)}{P(\{x_i\})}$$

- log-likelihood

$$\mathcal{L} = \ln P(\alpha|\{x_i\}) = n \ln(\alpha - 1) - n \ln x_{min} - \alpha \sum_{i=1}^{n} \ln \frac{x_i}{x_{min}}$$

- maximization $\frac{\partial \mathcal{L}}{\partial \alpha} = 0$

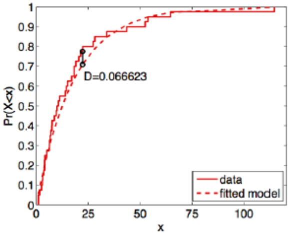
$$\alpha = 1 + n \left[\sum_{i=1}^{n} \ln \frac{x_i}{x_{min}} \right]^{-1}$$

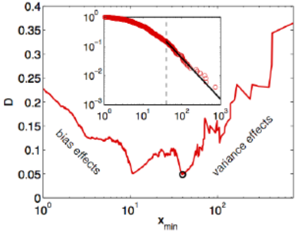
- error estimate

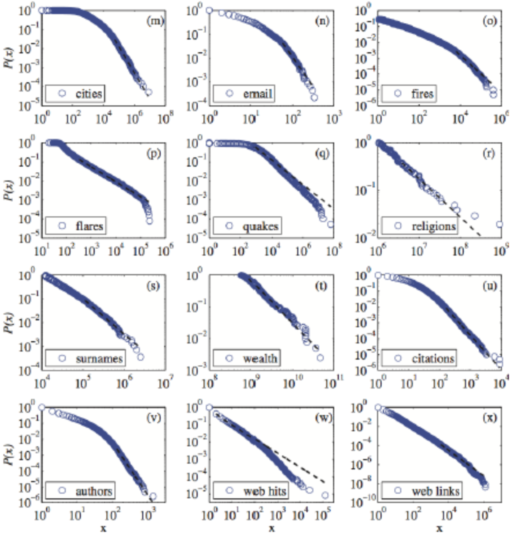
$$\sigma = \sqrt{n} \left[\sum_{i=1}^{n} \ln \frac{x_i}{x_{min}} \right]^{-1} = \frac{\alpha - 1}{\sqrt{n}}$$

- Kolmogorov-Smirnov test (compare model and experimental CDF)

$$D = \max_x |F(x|\alpha, x_{min}) - F_{exp}(x)|$$





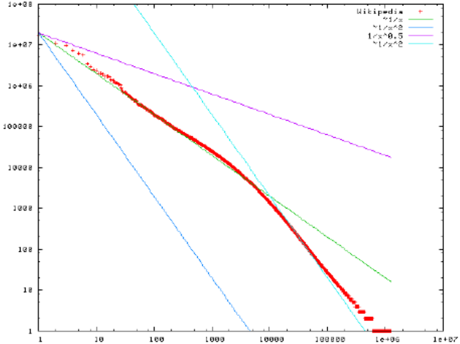

cities
email
fires
flares
quakes
religions
surnames
wealth
citations
authors
web hits
web links
P(x)
x

Word frequency table (6318 unique words, min freq 800, corpus size $>$ 85mln):

6187267 the
4239632 be
3093444 of
2687863 and
2186369 a
1924315 in
1620850 to
....
801 incredibly
801 historically
801 decision-making
800 wildly
800 reformer
800 quantum

Zipf's law - the frequency of a word in an natural language corpus is inversely proportional to its rank in frequency table $f(k) \sim 1/k$.

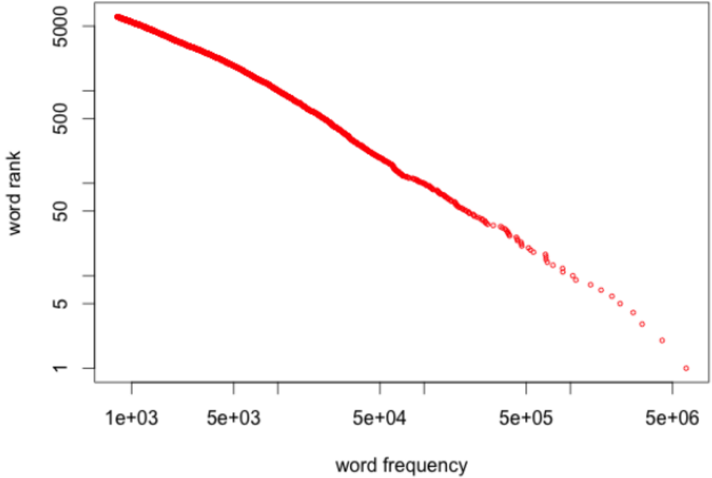
$$f(k) = \frac{1/k_s}{\sum_{k=1}^{N}(1/k^s)}$$



- Sort items by their frequency in decreasing order (frequency table)
- Fraction of the words with frequencies higher or equal to the k-th word is cCDF $\bar{F}(k) = Pr(X \geq k)$. The number of the words with frequency above k-th word is its rank k!
- Plot word rank as a function of the word frequency: rank $k - y$ axis, frequency $-x$ axis.
- Use rank-frequency plot instead of computing and plotting cumulative distribution of a quantity.

6187267 the
4239632 be
3093444 of
2687863 and
2186369 a
1924315 in

Rank-frequency plot



- What about the importance of nodes?
- Eigenvalue centrality of node i x_i) can remedy this as:

$$x_i = \kappa^{-1} \sum_{j|i} x_j \to \kappa^{-1} \sum_{j}^{n} A_{ij} x_j,$$

where κ is a given constant.

- In a matrix form, we obtain:

$$\mathbf{Ax} = \kappa \mathbf{x}$$

- Note that $\mathbf{x}$ is an eigenvector of $\mathbf{A}$
- BUT: $\mathbf{A}$ is $n \times n$, so which one from n eigenvectors to choose?
  - Hint: use the fact that $a_{ij} \geq 0$ and the Perron-Frobenious theorem
  - There is only one eigenvector with non-zero elements and its has the largest eigenvalue $\to$ set κ

- What about the importance of nodes?
- Eigenvalue centrality of node i x_i) can remedy this as:

$$x_i = \kappa^{-1} \sum_{j|i} x_j \rightarrow \kappa^{-1} \sum_{j}^{n} A_{ij} x_j,$$

where κ is a given constant.

- In a matrix form, we obtain:

$$\mathbf{A}\mathbf{x} = \kappa \mathbf{x}$$

- Note that **x** is an eigenvector of **A**
- BUT: **A** is $n \times n$, so which one from n eigenvectors to choose?
  - Hint: use the fact that $a_{ij} \geq 0$ and the Perron-Frobenious theorem
  - There is only one eigenvector with non-zero elements and its has the largest eigenvalue $\rightarrow$ set κ

- Eigenvalue centrality only works for undirected graphs!
- Problems with directed graphs:
  - $\mathbf{A}$ is asymmetric $\rightarrow$ two eigenvectors (left and right). Which one to use?
  - It is always case-specific! (Use domain judgement & ingoing rule!)
  - Also, directed graphs may have nodes, which are only outgoing or only ingoing, which will be ignored if one eigenvector is dropped
- Eigenvalue centrality only works for undirected graphs!
- But solution for directed graphs exists! Consider a modification given below:

$$x_i = \alpha \sum_{j}^{n} A_{ij} x_j + \beta,$$

where $\alpha > 0$ and $\beta > 0$.
- In a matrix form we have then $\mathbf{x} = \alpha \mathbf{A} + \beta \mathbf{1}$

- Eigenvalue centrality only works for undirected graphs!
- Problems with directed graphs:
  - $\mathbf{A}$ is asymmetric $\rightarrow$ two eigenvectors (left and right). Which one to use?
  - It is always case-specific! (Use domain judgement & ingoing rule!)
  - Also, directed graphs may have nodes, which are only outgoing or only ingoing, which will be ignored if one eigenvector is dropped
- Eigenvalue centrality only works for undirected graphs!
- But solution for directed graphs exists! Consider a modification given below:

$$x_i = \alpha \sum_{j}^{n} A_{ij} x_j + \beta,$$

where $\alpha > 0$ and $\beta > 0$.
- In a matrix form we have then $\mathbf{x} = \alpha \mathbf{A} + \beta \mathbf{1}$

- Eigenvalue centrality only works for undirected graphs!
- Problems with directed graphs:
  - $\mathbf{A}$ is asymmetric $\rightarrow$ two eigenvectors (left and right). Which one to use?
  - It is always case-specific! (Use domain judgement & ingoing rule!)
  - Also, directed graphs may have nodes, which are only outgoing or only ingoing, which will be ignored if one eigenvector is dropped
- Eigenvalue centrality only works for undirected graphs!
- But solution for directed graphs exists! Consider a modification given below:

$$x_i = \alpha \sum_{j}^{n} A_{ij} x_j + \beta,$$

where $\alpha > 0$ and $\beta > 0$.
- In a matrix form we have then $\mathbf{x} = \alpha \mathbf{A} + \beta \mathbf{1}$

- The Katz centrality follows from $\mathbf{x} = \alpha\mathbf{A} + \beta\mathbf{1}$ and from setting $\beta = 1$ (can be any):

$$\mathbf{x} = (\mathbf{I} - \alpha\mathbf{A})^{-\mathbf{1}}\mathbf{1}$$

- How to determine the value of α?
- Solving for $\alpha > 00$ leads to:

$$det(\alpha^{-1}\mathbf{I} - \mathbf{A}) = \mathbf{0} \rightarrow \alpha = \mathbf{1}/\kappa_{\mathbf{1}},$$

where $1/\kappa_1$ is the inverse of the largest eigenvalue.

- The Katz centrality follows from $\mathbf{x} = \alpha \mathbf{A} + \beta \mathbf{1}$ and from setting $\beta = 1$ (can be any):

$$\mathbf{x} = (\mathbf{I} - \alpha \mathbf{A})^{-\mathbf{1}} \mathbf{1}$$

- How to determine the value of α?
- Solving for $\alpha > 00$ leads to:

$$det(\alpha^{-1}\mathbf{I} - \mathbf{A}) = \mathbf{0} \rightarrow \alpha = \mathbf{1}/\kappa_{\mathbf{1}},$$

where $1/\kappa_1$ is the inverse of the largest eigenvalue.

- BUT: The Katz centrality can be abusive in some cases (when one a node has many outgoing connections.
- Solution(PageRank centrality): Use out-degree k^{out}:

$$x_i = \alpha \sum_j^n A_{ij} \frac{x_j}{k_j^{out}} + \beta,$$

- Which can be cast in a matrix form

$$\mathbf{x} = \alpha \mathbf{A}\mathbf{D}^{-1}\mathbf{x} + \beta \mathbf{1},$$

where $\mathbf{D} = \mathbf{diag}(\max(\mathbf{k_i^{out}}, \mathbf{1}))$

- We obtain the Pagerank centrality as

$$\mathbf{x} = (\mathbf{I} - \alpha \mathbf{A}\mathbf{D}^{-1})^{-1}\mathbf{1}$$

- BUT: The Katz centrality can be abusive in some cases (when one a node has many outgoing connections.
- Solution(PageRank centrality): Use out-degree k^{out}:

$$x_i = \alpha \sum_{j}^{n} A_{ij} \frac{x_j}{k_j^{out}} + \beta,$$

- Which can be cast in a matrix form

$$\mathbf{x} = \alpha \mathbf{A} \mathbf{D}^{-1} \mathbf{x} + \beta \mathbf{1},$$

where $\mathbf{D} = \mathbf{diag}(\max(\mathbf{k_i^{out}}, \mathbf{1}))$

- We obtain the Pagerank centrality as

$$\mathbf{x} = (\mathbf{I} - \alpha \mathbf{A} \mathbf{D}^{-1})^{-1} \mathbf{1}$$

- BUT: The Katz centrality can be abusive in some cases (when one a node has many outgoing connections.
- Solution(PageRank centrality): Use out-degree k^{out}:

$$x_i = \alpha \sum_j^n A_{ij} \frac{x_j}{k_j^{out}} + \beta,$$

- Which can be cast in a matrix form

$$\mathbf{x} = \alpha \mathbf{A}\mathbf{D}^{-1}\mathbf{x} + \beta \mathbf{1},$$

where $\mathbf{D} = \mathbf{diag}(\max(\mathbf{k_i^{out}}, \mathbf{1}))$
- We obtain the Pagerank centrality as

$$\mathbf{x} = (\mathbf{I} - \alpha \mathbf{A}\mathbf{D}^{-1})^{-1}\mathbf{1}$$

- BUT: The Katz centrality can be abusive in some cases (when one a node has many outgoing connections.
- Solution(PageRank centrality): Use out-degree k^{out}:

$$x_i = \alpha \sum_j^n A_{ij} \frac{x_j}{k_j^{out}} + \beta,$$

- Which can be cast in a matrix form

$$\mathbf{x} = \alpha \mathbf{A} \mathbf{D}^{-\mathbf{1}} \mathbf{x} + \beta \mathbf{1},$$

where $\mathbf{D} = \mathbf{diag}(\max(\mathbf{k_i^{out}}, \mathbf{1}))$
- We obtain the Pagerank centrality as

$$\mathbf{x} = (\mathbf{I} - \alpha \mathbf{A} \mathbf{D}^{-\mathbf{1}})^{-\mathbf{1}} \mathbf{1}$$

In summary, eigenvector centrality measures are very simple:

	With β	Without β
With k_i^{out}	$\mathbf{x} = (\mathbf{I} - \alpha \mathbf{A}\mathbf{D}^{-1})^{-1}\mathbf{1}$	$\mathbf{x} = \mathbf{A}\ \mathbf{D}^{-1}\mathbf{x}$
Without k_i^{out}	$\mathbf{x} = (\mathbf{I} - \alpha \mathbf{A}^{-1}\mathbf{1}$	$\mathbf{x} = \kappa^{-1}\mathbf{A}\mathbf{x}$

Degree and eigenvalue centrality do not inform on the closeness of nodes (e.g. number of hops)

- Recall the shortest distance d_{ij}
- The mean shortest distance is then:

$$l_i = \frac{1}{n}\sum_j d_{ij}$$

- The closeness centrality naturally follows then:

$$C_i = \frac{1}{l_i} = \frac{n}{\sum_j d_{ij}}$$

- Think about an application of the closeness centrality

Degree and eigenvalue centrality do not inform on the closeness of nodes (e.g. number of hops)

- Recall the shortest distance d_{ij}
- The mean shortest distance is then:

$$l_i = \frac{1}{n} \sum_j d_{ij}$$

- The closeness centrality naturally follows then:

$$C_i = \frac{1}{l_i} = \frac{n}{\sum_j d_{ij}}$$

- Think about an application of the closeness centrality

Degree, eigenvalue, and a closeness centrality do not inform on what nodes lie between a given pair of nodes

- Let $n^i_{st} \in \{0, 1\}$ attain the value 0, unless node i is included in the shortest distance d_{st}
- The betweeness centrality naturally follows then:

$$x_i = \sum_{st} n^i_{st}$$

- Normalization is possible:

$$x_i = \frac{1}{n^2} \sum_{st} \frac{n^i_{st}}{g_{st}} \text{and note that } 0 \leq x_i \leq 1$$

where $\frac{1}{n^2}$ normalizes over a total number of nodes and $\frac{1}{g_{st}}$ normalizes over a number of possible shortest paths from s to t

Degree, eigenvalue, and a closeness centrality do not inform on what nodes lie between a given pair of nodes

- Let $n^i_{st} \in \{0, 1\}$ attain the value 0, unless node i is included in the shortest distance d_{st}
- The betweeness centrality naturally follows then:

$$x_i = \sum_{st} n^i_{st}$$

- Normalization is possible:

$$x_i = \frac{1}{n^2} \sum_{st} \frac{n^i_{st}}{g_{st}} \text{and note that } 0 \leq x_i \leq 1$$

where $\frac{1}{n^2}$ normalizes over a total number of nodes and $\frac{1}{g_{st}}$ normalizes over a number of possible shortest paths from s to t

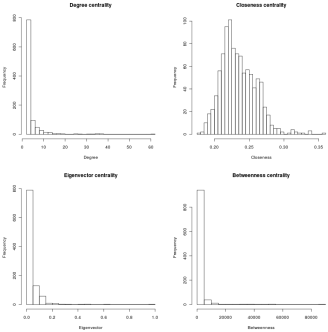
Centrality measures:

- Degree
- Eigenvector
- Closeness
- Betweeness
- More to come: flow and random-walk betweeness/closeness, hub centrality, cluster centrality, etc
- What are the distribution laws for the eigenvector, closeness and betweeness centrality measures?

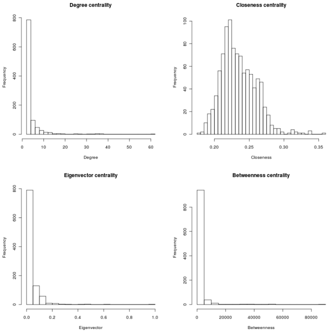
Centrality measures:

- Degree
- Eigenvector
- Closeness
- Betweeness
- More to come: flow and random-walk betweeness/closeness, hub centrality, cluster centrality, etc
- What are the distribution laws for the eigenvector, closeness and betweeness centrality measures?

- Eigenvector and betweeness centralities follow power disitrbutions (roughly)
- Closeness has a more complex shape
- Why?



- Eigenvector and betweeness centralities follow power disitrbutions (roughly)
- Closeness has a more complex shape
- Why?



Three major sources:

- http://networksciencebook.com/translations/en/resources/data.html
- http://www-personal.umich.edu/~mejn/

Some ideas (for now):

- Compute centrality measures
- Compute centrality distributions
- Visualize networks
- Compare all of the above for different networks

Books:

- Neman: Sections 6.10, 10.3-10.5

Other sources:

- Power laws, Pareto distributions and Zipfs law, M. E. J. Newman, Contemporary Physics, pages 323351, 2005.
- Power-Law Distribution in Empirical Data, A. Clauset, C.R. Shalizi, M.E.J. Newman, SIAM Review, Vol 51, No 4, pp. 661-703, 2009.
- A Brief History of Generative Models for Power Law and Lognormal Distributions, M. Mitzenmacher, Internet Mathematics Vol 1, No 2, pp 226-251.