Complex-Network Modelling and Inference

Lecture 13: Random Graphs: preferential-attachment models

Matthew Roughan
<matthew.roughan@adelaide.edu.au>
https://roughan.info/notes/Network_Modelling/

School of Mathematical Sciences,
University of Adelaide

August 18, 2021

Section 1

Preferential attachment graphs

Problem with "small-world" graph

- Small-world graph replicates desired
  - short path length
  - high clustering
- Node degree is (almost) always k
  - but observed node-degree distributions are more variable

'Scale-free' Networks

- Barabási and Albert [BA99]
- Draw on idea that the "rich get richer"
- Preferential attachment model
    1. start with $N = \{1, 2\}$, and $E = \{(1, 2)\}$.
    2. for i=3:n
        a. add vertex i to N
        b. add link (i, j) to E, where $j \in \{1, 2, \ldots, i-1\}$ is chosen with probability

$$p_j = \frac{k_j}{\sum_{k=1}^{i-1} k_k},$$

where k_j is the degree of node j.

'Scale-free' Networks, mark II

- Barabási and Albert [BA99]
- Draw on idea that the “rich get richer”
- Preferential attachment model
  1. `start with` $N = \{1, 2, \ldots, m\}$, `and` $E = \{(i, j) \mid \forall i = 1, 2, \ldots, m, j = i + 1, \ldots, m\}$.
  2. `for i=3:n`
     a. `add vertex` i `to` N
     b. `add` m `links` (i, j) `to` E, `where` $j \in \{1, 2, \ldots, i-1\}$ `is chosen with probability`

$$p_j = \frac{k_j}{\sum_{k=1}^{i-1} k_k},$$

`where` k_j `is the degree of node` j.

- Note that the result will be a multi-graph unless care is take to sample from the above distribution without replacement.

Properties of preferential attachment

- connected (by construction)
- degree distribution takes power-law form

$$p_k \simeq k^{-\alpha}.$$

Degree distribution approximation

- Take degree k_i of ith node to be a continuous variable
- Take time (number of nodes added) to be continuous
- Rate of increase of degree is proportional to degree

$$\frac{dk_i}{dt} = m \frac{k_i}{\sum_{j=1}^{n} k_j}$$

- note that the total number of links in the network is

$$|E| = mt = \sum_{j=1}^{n} k_j / 2$$

Degree distribution approximation

- Substitute $2mt$ in first equations

$$\frac{dk_i}{dt} = \frac{k_i}{2t}.$$

- Solve the DE, and we get

$$k_i(t) = ct^{1/2}$$

- Use initial condition $k_i(t_i) = m$

$$k_i(t) = m(t/t_i)^{1/2}$$

Degree distribution approximation

- So

$$k_i(t) = m(t/t_i)^{1/2}$$

- Calculating the CDF we get

$$\begin{aligned}
\mathrm{Prob}\{k_i(t) < k\} &= \mathrm{Prob}\{m(t/t_i)^{1/2} < k\} \\
&= \mathrm{Prob}\{(t/t_i) < (k/m)^2\} \\
&= \mathrm{Prob}\{(t_i/t) > (m/k)^2\} \\
&= 1 - \mathrm{Prob}\{(t_i/t) \leq (m/k)^2\}
\end{aligned}$$

- Adding nodes at uniform time intervals means $t_i = i$, so in the limit as $t \to \infty$, the t_i/t are uniformly distributed on $[0, 1]$, and we get the form

$$\mathrm{Prob}\{k_i(t) < k\} \simeq 1 - (m/k)^2$$

for $(m/k)^2 \leq 1$

Degree distribution approximation

- For large k, $(m/k)^2 \leq 1$, and the density function p_k can be approximated by the derivative

$$
\begin{aligned}
p_k &\simeq \frac{d}{dk}\mathrm{Prob}\{k_i(t) < k\} \\
&\simeq -\frac{d}{dk}(m/k)^2 \\
&\simeq 2m^2k^{-3}
\end{aligned}
$$

- we usually care about the limit (for this type of distribution) so we write

$$p_k \sim k^{-3}$$

- This is a power law with exponent 3

Generalisation

Evolve the network over time

- add m edges with probability p
  - one end uniformly chosen over all nodes
  - other end chosen proportional to degree
- rewire m edges with probability q
  - choose node i at random
  - rewire one of its edges using proportional attachment
- with probability $1 - p - q$ a new node is added
  - m new edges with proportional attachment

Can generate degree distribution with power-law between 2 and ∞.

Why are they called "Scale Free"?

- degree distribution doesn't depend on the size of the network (as long as it's a limit)
- form of degree distribution doesn't depend on number of links (per node)
- power-laws exhibit a type of scale invariance

$$\begin{aligned} p(x) &= ax^{-\alpha} \\ p(bx) &= a(bx)^{-\alpha} \\ &= Ax^{-\alpha} \propto p(x) \end{aligned}$$

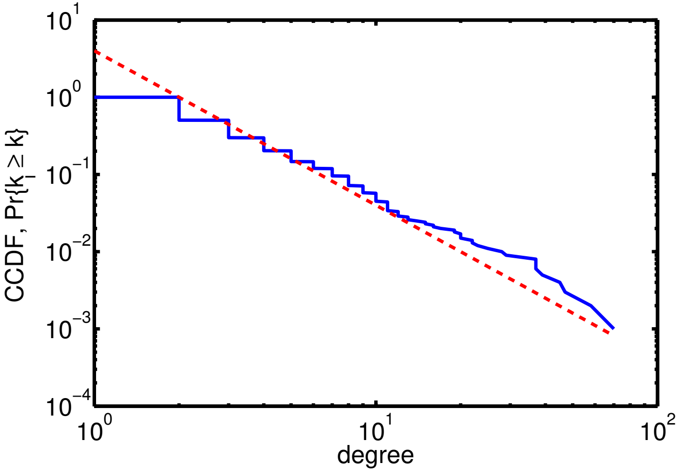
- another form of scale invariance

$$p(2x) = 2^{\alpha} p(x)$$

regardless of x

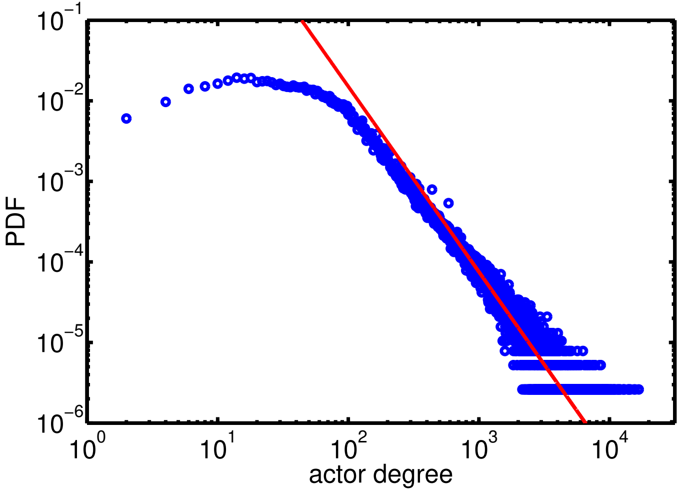
Power-laws

Power-laws look like straight lines on a log-log graph



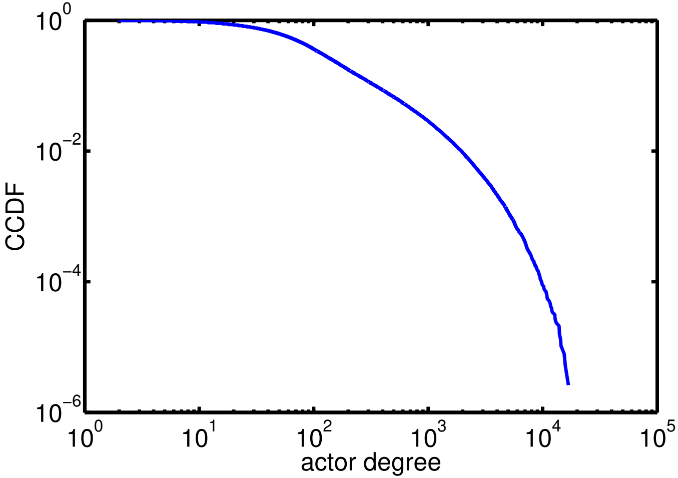
Do they match real data?

Actor collaboration graph appears to have power-law [BA99]



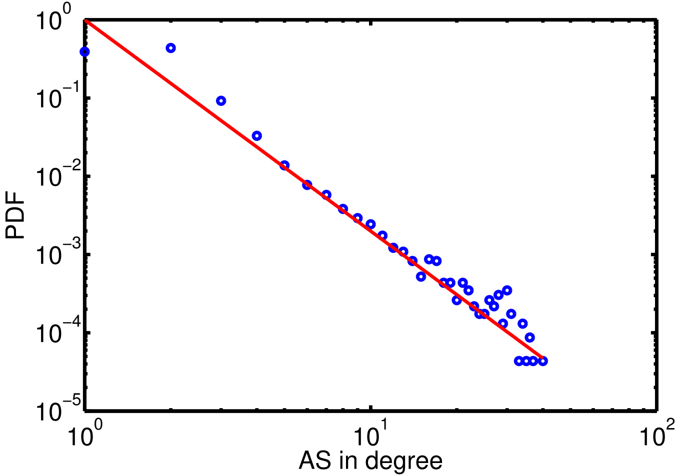
Do they match real data?

Care must be taken though



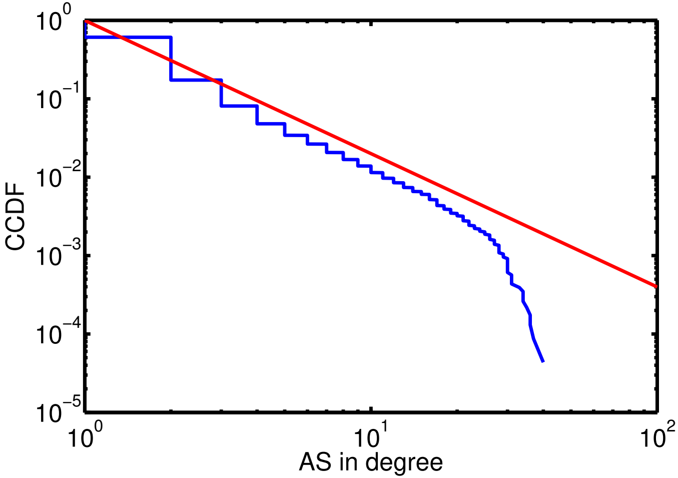
Do they match real data?

AS-graph appears to have power-law [FFF99]



Do they match real data?

Care must be taken though

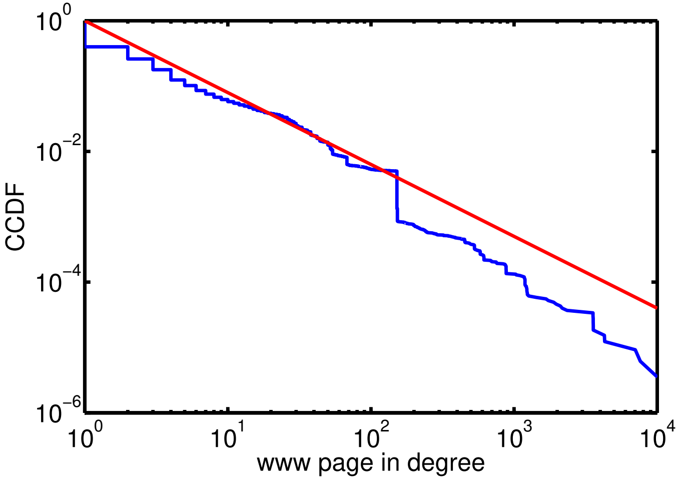


PDF vs CCDF

- for continuous distributions
  - PDF = derivative of CDF = - derivative of CCDF
  - if one has a power-law, both should
- PDF: $\mathrm{Prob}\{k_i == k\}$
  - hard to accurately estimate
  - require arbitrary choice of "binning"
  - lots of "zeros" in the tail
  - zeros don't show up on log-log graph
- CCDF: $\mathrm{Prob}\{k_i > k\}$
  - easy to estimate/plot
    ```
    loglog(sort(degree), 1 - (0:n-1)/n)
    ```
  - much more robust in the tail

Do they match real data?

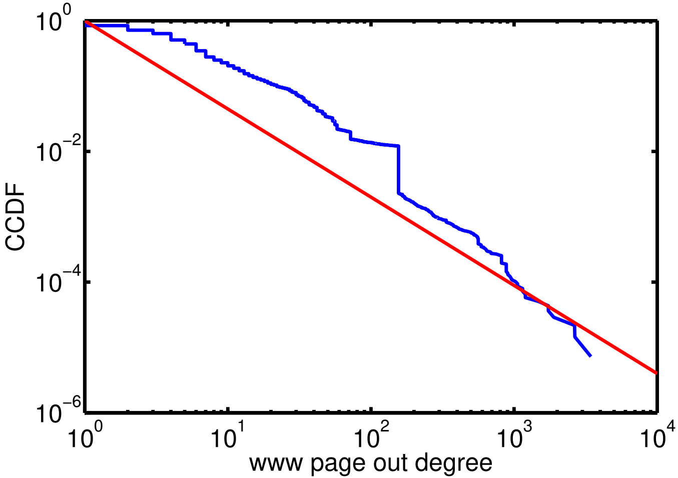
WWW page graph really appears to have power-law [AJB99]



http://www.nd.edu/~networks/resources.htm

Do they match real data?

WWW page graph really appears to have power-law [AJB99]



http://www.nd.edu/~networks/resources.htm

Power-law degree

- Appeal of the model
  - simple/parsimonious
  - real networks **sometimes** have power-law degree
    - makes more sense for virtual networks
  - power-law is an emergent phenomena
  - seems logical, but wait
- Even if they match data, does the model explain the "real" process behind network construction
  - is this the only way to generate power-laws?
  - if not, does the model tell us anything?
  - do other features match real networks?
- And they don't match as many data sets as the hype:
  "Scale-free networks are rare", Broido and Clauset, Nature Communications 2019.
- We'll come back to these topics after we consider measurements in more detail.

Preferential attachment generalisations

- Price's model: We can make the number of edges brought by a new node random
- We can allow some re-wiring
  - allows varying power-law exponent
- Can allow node birth and death of nodes

Estimation

- In BA model, it comes down to estimating average degree
- In general, need to estimate exponent of a power-law
  - more on this later

Further reading I

R'eka Albert, Hawoong Jeong, and Albert-László Barabási, *Diameter of the world wide web*, Nature **401** (1999), no. 130, 130–131.

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

Michalis Faloutsos, Petros Faloutsos, and Christos Faloutsos, *On power-law relationships of the Internet topology*, ACM SIGCOMM'99, 1999.