

SNA 2C: Growth & Preferential Attachment Models

Lada Adamic



Online Question & Answer Forums



Uneven participation



Real-world degree distributions

Sexual networks Great variation in contact numbers 10^{-1} 10^{-1}

k [sex partners]

Power-law distribution



high skew (asymmetry)

straight line on a log-log plot

Poisson distribution



- little skew (asymmetry)
- curved on a log-log plot

Power law distribution

Straight line on a log-log plot $\ln(p(k)) = c - \alpha \ln(k)$

Exponentiate both sides to get that p(k), the probability of observing an node of degree 'k' is given by

$$p(k) = Ck^{-\alpha}$$

normalization constant (probabilities over all *k* must sum to 1)

power law exponent α

Quiz Q:

- As the exponent α increases, the downward slope of the line on a log-log plot
 - stays the same
 - becomes milder
 - becomes steeper

2 ingredients in generating power-law networks

nodes appear over time (growth)



2 ingredients in generating power-law networks

nodes prefer to attach to nodes with many connections (preferential attachment, cumulative advantage)





Ingredient # 1: growth over time

nodes appear one by one, each selecting m other nodes at random to connect to



m = 2

random network growth

- one node is born at each time tick
- at time t there are t nodes
- □ change in degree k_i of node *i* (born at time i, with 0 < i < t)



a node in a randomly grown network

how many new edges does a node accumulate since it's birth at time *i* until time *t*?

□ integrate from *i* to *t*

$$\frac{dk_i(t)}{dt} = \frac{m}{t}$$

to get

$$k_i(t) = m + m \log(\frac{t}{i})$$

born with **m** edges

age and degree

on average $k_i(t) > k_j(t)$ if i < j

i.e. older nodes on average have more edges



Quiz Q:

How could one make the growth model more realistic for social networks?
 old nodes die
 some nodes are more sociable
 friendships vane over time
 all of the above

growing random networks

Let $\tau(100)$ be the time at which node with degree e.g. 100 is born. The the fraction of nodes that have degree <= 100 is $(t - \tau)/t$

$$k_{\tau}(t) = m + m \log(\frac{t}{\tau})$$

random growth: degree distribution



The probability that a node has degree k or less is $1-\tau/t$

$$P(k < k') = 1 - e^{-\frac{k' - m}{m}}$$

Quiz Q:

The degree distribution for a growth model where new nodes attach to old nodes at random will be
 a curved line on a log-log plot
 a straight line on a log-log plot

2nd ingredient: preferential attachment

Preferential attachment:

new nodes prefer to attach to well-connected nodes over less-well connected nodes

Process also known as

- cumulative advantage
- rich-get-richer
- Matthew effect

Price's preferential attachment model for citation networks

[Price 65]

- each new paper is generated with m citations (mean)
- new papers cite previous papers with probability proportional to their indegree (citations)
- what about papers without any citations?
 - each paper is considered to have a "default" citation
 - probability of citing a paper with degree k, proportional to k+1

D Power law with exponent $\alpha = 2+1/m$

Preferential attachment





Cumulative advantage: how?



Def author at generals seen scients due to has not preparations forware study segand due to has any solid general construct to an of a local sector of the science of the science of the local sector of the science publicity sizes its polarities is had not seen to be approximately size of the second Address is created for accesses of the point. If these understands is the process and the point and phones and is not created in accelerative. The particul described is the Propose controlly: The tabled deached is the paper ine best implemental in the Folds. Hypogeneo-bathen in the Westmer Likelsheed anguest pathag-vised is modelin by persystem by a Sprincosk-encets in the descent collisionation. Martinessen, R. 1998, Inc. J. Martiness, M. Baragara, Martinesse, M. 19, Annual Science, and Annual Science, Science and Annual Science and Ann ----Regi II (102) Based aware from of former relation former strategies and the second strategies of the energy strategies and the second strategies and and the second strategies for the second strategies and the second strategies for the second strategies and the second strategies and strategies for the second strategies and the second strategies and strategies for the second strategies and strategies for strategies and strategies for the second strategies and the second strategies and strategies for strategies for the second strategies and strategies for strategies and strategies for the second strategies and strategies and strategies and strategies for the second strategies and strategies and strategies and strategies for the second strategies and strategies and strategies and strategies for the second strategies and strategies and strategies and strategies for the second strategies and strategies and strategies and strategies and strategies for the second strategies and str

the lattice of collect size of the Concern Web Concernent the first offer appendix concernent had assigned in the first offer in the con-cernent of the first offer in the con-sensing finitume (COR) (III) Reduce of COR) is done considing using off do advergence to complex and built have their the complexity of the constraints are built as the first

Kons of species Webs of Artigers Kons of Webs of Webs of Webs Konserver webs of Webs Konserver webs of Webs Konserver webs

Tantan A. (10). Mar.) also is functions second to constant life, parsing a function of the first of distances of a collection participant of designed on the manufacture for the participant of designed on the Recented & 1965, To release if we applied to the second se Read (a) (b) the holes proves of spatio good, 18

1 No. 54 Street Million for summarial collinear access Marchine State in a large for the state of a 1997 Million collinear party Collection States

The Alline Residence property of the sections fielded residence interpretation of the sections fielded material and the first of the Alline

IN URBANISH A Robert Pargent environment operation

<page-header><text><text><text><text><section-header><section-header><section-header><section-header><section-header><section-header><section-header><section-header><section-header><text><text><text><text>

copying mechanism



Barabasi-Albert model

- First used to describe skewed degree distribution of the World Wide Web
- Each node connects to other nodes with probability proportional to their degree
 - the process starts with some initial subgraph
 - each new node comes in with *m* edges
 - probability of connecting to node i

$$\Pi(i) = m \frac{k_i}{\sum_j k_j}$$

Results in power-law with exponent α = 3

Basic BA-model

- Very simple algorithm to implement
 - **\Box** start with an initial set of m_0 fully connected nodes

• e.g.
$$m_0 = 3$$

- now add new vertices one by one, each one with exactly m edges
- each new edge connects to an existing vertex in proportion to the number of edges that vertex already has → preferential attachment
- easiest if you keep track of edge endpoints in one large array and select an element from this array at random
 - the probability of selecting any one vertex will be proportional to the number of times it appears in the array – which corresponds to its degree

1 1 2 2 2 3 3 4 5 6 6 7 8

generating BA graphs - cont'd

112233

1 1 2 2 2 3 3 3 4 4

3

3

3

2

2

5

4

- To start, each vertex has an equal number of edges (2)
 - the probability of choosing any vertex is 1/3
- We add a new vertex, and it will have m edges, here take m=2
 - draw 2 random elements from the array suppose they are 2 and 3
- Now the probabilities of selecting 1,2,3,or 4 are 1/5, 3/10, 3/10, 1/5
- Add a new vertex, draw a vertex for it to connect from the array

1 1 2 2 2 3 3 3 3 4 4 4 5 5

etc.

after a while...



contrasting with random (non-preferential) growth



m = 2



random

preferential

mean field approximation

probability that node *i* acquires a new link at time *t*

$$\frac{dk_i(t)}{dt} = m\frac{k_i}{2tm} = \frac{k_i}{2t} \quad \text{with} \quad k_i(i) = m$$

$$k_i(t) = m(\frac{t}{i})^{1/2}$$

BA model degree distribution

 \Box time of birth of node of degree k': τ

$$\frac{\tau}{t} = \left(\frac{m}{k'}\right)^2$$

$$P(k < k') = 1 - \frac{m^2}{k'^2}$$

$$p(k) = \frac{2m^2}{k^3}$$

Properties of the BA graph

- The distribution is scale free with exponent α = 3 P(k) = 2 m²/k³
- The graph is connected
 - Every new vertex is born with a link or several links (depending on whether m = 1 or m > 1)
 - It then connects to an 'older' vertex, which itself connected to another vertex when it was introduced
 - And we started from a connected core
- The older are richer
 - Nodes accumulate links as time goes on, which gives older nodes an advantage since newer nodes are going to attach preferentially – and older nodes have a higher degree to tempt them with than some new kid on the block



try it yourself



http://www.ladamic.com/netlearn/NetLogo501/RAndPrefAttachment.html

Quiz Q:

- Relative to the random growth model, the degree distribution in the preferential attachment model
 - resembles a power-law distribution less
 - resembles a power-law distribution more

Summary: growth models

- Most networks aren't 'born', they are made.
- Nodes being added over time means that older nodes can have more time to accumulate edges
- Preference for attaching to 'popular' nodes further skews the degree distribution toward a power-law

Assignment: implications for diffusion

How does the size of the giant component influence diffusion?





Assignment: implications for diffusion

How do growth and preferential attachment influence diffusion?

