Network Science: Random and preferential attachment models for network growth

Ozalp Babaoglu
Dipartimento di Informatica — Scienza e Ingegneria
Università di Bologna
www.cs.unibo.it/babaoglu/

Models for network growth

- So far, we have examined models that define how a network of n nodes forms by adding edges
- These models are static with respect to nodes
- Consider starting with a small network of few nodes and define how to grow it by adding new nodes
- Why do networks grow?
  - citation networks (new papers written)
  - web graph (new pages created)
  - collaboration networks (new authors are born)
  - social networks (new individuals enter)

Models for network growth

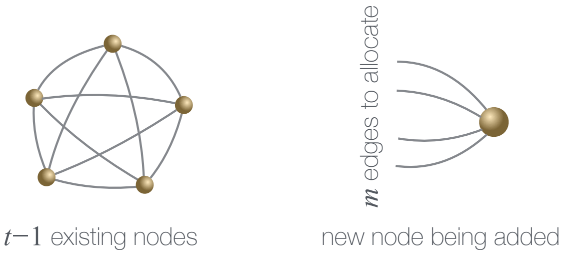
- Growth models add more realism
- Include a form of dynamism
- Add heterogeneity to the network through age (older nodes likely to have more edges)
- Natural way of obtaining varied degree distributions
- Potentially address the poor degree distribution prediction capability of previous models in search of heavy-tailed degree distributions

Models for network growth
Uniform random model

- First idea: extend the Erdős-Rényi model to dynamic nodes
  - a new node is born at each time step
  - the new node has m edges to allocate to existing nodes
  - each node to link to is selected at random uniformly
- Assume first m nodes are fully connected
- At time t there will be t nodes all together ($t-1$ existing nodes plus the new one) and an existing node will get a new link with probability m/t

Models for network growth
Uniform random model

At time t



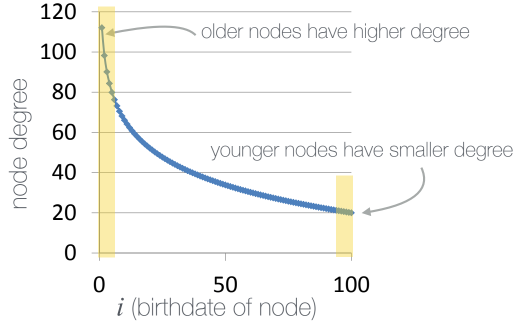
Models for network growth
Uniform random model

- At time t, how many edges will a node that was born at time i $(m<i<t)$ have on the average?

  $m+m/(i+1)+m/(i+2)+ \dots +m/t$
- These are the harmonic numbers and for large t can be approximated by $m(1+\log(t/i))$
- What distribution of degrees does this generate?
- Nodes that have average degree less than k are those such that $m(1+\log(t/i)) < k$

Models for network growth
Uniform random model

- What does $m(1+\log(t/i))$ look like as function of i?
- Look at node degrees at time $t=100$ for $m=20$



Models for network growth
Uniform random model

- Solving $m(1+\log(t/i)) < k$ for i we get the time that a node has to be born to have degree less than k

  $i > t\ e^{-(k-m)/m}$
- Nodes born after i will have expected degree less than k
- Fraction of nodes that have expected degree less than k at time t is given by (at time t, there are t nodes total)

  $F_t(k) = (t - t\ e^{-(k-m)/m})/t$

  $= 1 - e^{-(k-m)/m}$
- In other words, distribution of *expected degrees* is *exponential* with mean m

Models for network growth
Preferential attachment

- Growing networks by linking new nodes to existing nodes uniformly and randomly does not produce heavy-tailed degree distributions
- Idea: do not link new nodes uniformly and randomly but according to a “rich-get-richer” scheme which is known to produce heavy-tailed distributions
- New nodes prefer to attach to well-connected nodes rather than less-connected nodes
- This process amplifies inequality among node degrees

Models for network growth
Preferential attachment

- “Preferential attachment” is one such model where the likelihood of linking to a node is proportional to its current degree — the greater the degree, the more edges it will get making its degree even greater
- 1896 Pareto (wealth)
- 1925 Yule (relation to power-law)
- 1955 Simon (sizes of cities, etc.)
- 1976 Price (citation networks)
- 1999 Barabási–Albert (web graph)

Preferential attachment
Price model

- Each new paper generates (on the average) m citations
- New papers cite previous papers with probability proportional to their citation count
- Each paper has a default citation
- Preferential attachment rule: probability of citing a paper with k citations is proportional to $k+1$
- Generates a power-law with exponent $\alpha=2+1/m$

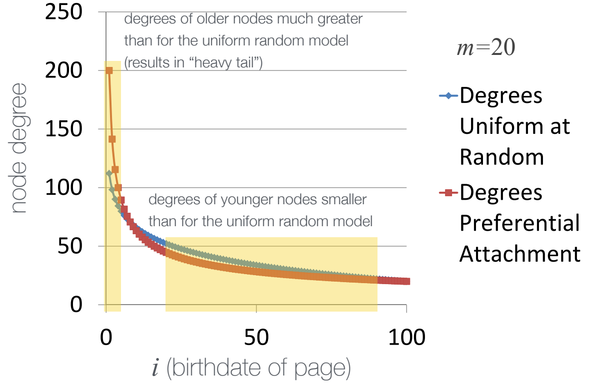
Preferential attachment
Barabási–Albert model

- A new page is born at every time step
- New pages link to m existing pages
- At time t, there will be a total of t pages and mt edges
- Sum of all page degrees is $2mt$
- Degree of page i at time t is denoted by $k_i(t)$
- Preferential attachment rule: probability of a new page linking to page i is given by $k_i(t)/2mt$
- In other words, a new page will *prefer* to link to a page that already has high degree — rich-get-richer

Preferential attachment
Barabási–Albert model

- Continuous time (mean-field) approximation:
- Pages get m edges at birth: $k_i(i)=m$
- Rate of change of degree (how many of the m new edges does page i get?)
  $dk_i(t)/dt=m(k_i(t)/2mt)=k_i(t)/2t$
- Solving the differential equation we get
  $k_i(t)=m(t/i)^{½}$

Preferential attachment
Barabási–Albert model



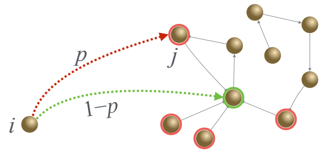
Preferential attachment
Barabási–Albert model

- Pages that have expected degree less than k at some time t are those such that
  $m(t/i)^{½} < k$
  $i > tm^2/k^2$
- The fraction of nodes with degree less than k at time t becomes (recall that at time t, there are t nodes total)
  $F_t(k) = (t - tm^2/k^2)/t = 1 - m^2/k^2$
- Taking the derivative with respect to k we get the density function
  $f_t(k) = 2m^2/k^3 = c\,k^{-3}$ (power-law with exponent α=3)
- NetLogo Library/Networks/Preferential Attachment

Preferential attachment
Alternative interpretation

- A new page is born at every time step
- New pages link to a single existing page
- The link of the i^{th} page is attached to an earlier page according to the following rule:
  - Page i chooses a page j uniformly at random from among all earlier pages, and with probability p adds an edge to page j
  - With probability $1-p$, page i instead adds an edge to the page that j *points to*

Preferential attachment
Alternative interpretation



- A new page i is born and has to be added
- Pick a page j uniformly and randomly among existing pages
- With probability p link to page j
- With probability $1-p$ link to the page that j points to
- But there are 4 different initial choices that would have resulted in i linking to this same page