A network growth model with intrinsic vertex fitness

G. J. Rodgers

Brunel University
School of Information Systems, Computing and Mathematics

Outline

Development of the field

- Network theory emerged from the study of graphs
- The field has changed significantly over time:
  - Foundation of **graphs**: Seven Bridges of Kőnigsberg (Euler, 1735)
  - Foundation of **random graphs**: Erdős, Renyi (1959)
  - Foundation of **complex networks**: Price (1976), Barabasi,Albert (1999).
- Today, complex networks receive a lot of attention in the academia, but also in the popular science community
  - i.e. Wolfram Research recent efforts on analysing Facebook data

Network models

- Networks can be coarsely classified into
  - Dynamic vs. static
  - Degree preferential vs. intrinsic variables
- Degree preferential models have been investigated thoroughly since the work of BA in 1999
- Static networks driven by intrinsic variables have been investigated well since the work of Caldarelli et.al (2002)

Classical Preferential Attachment

$$\Pi_{\text{new}\to i} = \frac{k_i}{\sum_j k_j} \qquad \Rightarrow \qquad p(k) \propto k^{-\alpha}$$

- Empirics show that this is **not the exact underlying mechanism** for real world networks
- A cross-temporal study has shown that not all networks that have a power-law degree distribution are built in this way:
  - Internet ✓
  - Citation network ✓
  - Science Collaboration ✗
  - Hollywood actor network ✗
- For the latter 2 networks non linear attachment kernels have been found (Jeong,Néda and Barabási, 2003)
- Theory does not predict power-laws in this case, although empirics have found $p(k) \propto k^{-\alpha}$

Hidden variables

- Having perfect information about **connectivity** is a very unrealistic assumption in many circumstances
  - It applies for some networks, i.e. the migration towards cities
  - But for example investment decisions are usually not disclosed
- A model that assumes perfect information about **attractiveness** of nodes is called the hidden variable model
  - Attractiveness can be anything: RoE, Life quality, GDP, etc...
  - Attractiveness is in the recent literature an intrinsic random variable, that follows some distribution $\rho(x)$
- Attachment is driven by mutual attractiveness for two nodes with fitness x and y: $f(x, y)$.

The model

- In this dynamic model, one of two things can happen **each time-step**:
  - With probability q **a new node is created**, endowed with a fitness x and attached to a node in the existing network with fitness y with probability proportional to $f(x, y)$.
  - With probability $1 - q$ **an edge is added** between two nodes of fitness x and y with probability prop. to $f(x, y)$.
- Note: For large t: $N \sim qt$

The model cont'd.

- Static networks with hidden variables have been studied well
- **Nobody has yet obtained a general theory for growing networks** purely induced by a linking propensity $f(x, y)$ and a fitness distribution $\rho(x)$
- We suggest a **new paradigm**.
- Model driven by attractiveness $\lambda(x, t)$:

$$\lambda(x,t) = \frac{1}{t}\left[\int_0^\infty \frac{\rho(y) f(y,x)}{\int_0^\infty f(y,x)\rho(x)dx} dy + \frac{2(1-q)}{q} \frac{\int_0^\infty f(y,x)\rho(y)dy}{\int_0^\infty \int_0^\infty f(y,x)\rho(x)\rho(y)dxdy} \right] \quad (1)$$

- Note that $\lambda(x, t)$ is factorisable in t. In the following:

$$\lambda(x,t) = \frac{1}{t}\lambda(x) \quad (2)$$

The model cont'd.

- **Attractiveness is comparative** $\Rightarrow$ ranks are sufficient
- Hence, measure everything in **ranks** $u \in [0,1]$

$$\lambda(u) \qquad \text{and} \qquad \tilde{\rho}(\lambda) = \frac{du}{d\lambda}$$

- It is closer to the original ideas of preferential attachment, that is defined purely by an attachment probability, not by two functions

A note on the range of power-law exponents

- For factorisable $f(x, y)$, from the definition of $\lambda(x, t)$, it follows directly that

$$\int_0^{\infty} \lambda(x)\rho(x)dx = \int_0^1 \lambda(u)du = 1 \tag{3}$$

- This condition is equivalent to

$$\sum_k kp(k) = \frac{2}{q} \quad \left(= \langle k \rangle = \frac{\sum_i k_i}{N(t)} = \frac{2t}{qt}\right) \tag{4}$$

A note on the range of power-law exponents cont'd.

- Imposing now that

$$p(k) \propto k^{-\alpha} \tag{5}$$

- Normalisation $\sum_k p(k) = 1$, since $p(k)$ is a density leads to

$$p(k) = \frac{1}{\zeta(\alpha)} k^{-\alpha} \tag{6}$$

- Substituting this into the equation for $\langle k \rangle$ leads to

$$\frac{2}{q} = \langle k \rangle = \frac{1}{\zeta(\alpha)} \sum_k k \cdot k^{-\alpha} = \frac{\zeta(\alpha - 1)}{\zeta(\alpha)} \tag{7}$$

- This limits the range of feasible exponents for a pure power law emerging from the described mechanism to

$$\alpha \in (2, 2.478 \ldots] \tag{8}$$

The solution for the full analytic theory

- The solution of this model can be obtained with a fairly detailed expression:
- The probability that a node has degree k at time t, given that it joined with ranking u at time τ:

$$p(k, t|u, \tau)$$

- Rate equation on this distribution function:

$$\begin{aligned} p(k, t+1|u, \tau) = p(k-1, t|u, \tau) \cdot \lambda(u, t) \\ + p(k, t|u, \tau) \cdot (1 - \lambda(u, t)) \end{aligned} \tag{9}$$

- The solution can be obtained in terms of a generating function:

$$G(s, u, \tau, t) = \sum_{k \geq 1} p(k, t|u, \tau) s^k \tag{10}$$

Solution cont'd

- That leads to a differential equation that can be solved to obtain for large t.

$$G(s, u, \tau, t) = s \cdot \left(\frac{t}{\tau}\right)^{(s-1)\lambda(u)} \tag{11}$$

- This solution uses the earlier result that $\lambda(u, t) = \lambda(u)/t$
- Averaging over ranks and entry time and expanding in k gives the stationary distribution of node degrees

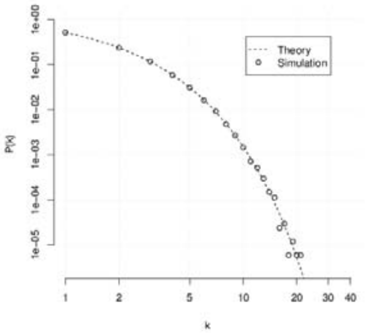
$$p(k) = \int_0^1 \frac{du}{\lambda(u) + 1} \left(\frac{\lambda(u)}{1 + \lambda(u)}\right)^{k-1} \tag{12}$$

An Example

- We have found so far the formalism that gives for a given attachment propensity $\lambda(u)$ the resulting degree distribution:

$$p(k) = \int_0^1 \frac{du}{\lambda(u)+1} \left(\frac{\lambda(u)}{1+\lambda(u)} \right)^{k-1}$$

- This result is numerically confirmed.
- E.g. $\lambda(u) = \frac{3}{2} u^{\frac{1}{2}}$



Power law degree distribution

- We have seen earlier that **power-law degree distributions are of special interest**, hence finding an attachment propensity for this class of models is of special interest.
- Solving for a power law can be done in two different ways
  - One way is to directly postulate a power law:

$$G(s) = A \cdot \sum_{k \geq 1} k^{-\alpha} s^k, \qquad A = \frac{1}{\zeta(\alpha)} \tag{13}$$

The rhs is a poly-logarithm. Using the integral representation of rhs, and substituting explicit $G(s)$ on the lhs., obtain:

$$\int_0^1 \frac{du}{\lambda(u)} \frac{s}{\frac{1+\lambda(u)}{\lambda(u)} - s} = \frac{s}{\zeta(\alpha)\Gamma(\alpha)} \int_0^\infty \frac{t^{\alpha-1}}{e^t - s} dt \tag{14}$$

Power law degree distribution cont'd.

$$\int_0^1 \frac{du}{\lambda(u)} \frac{s}{\frac{1+\lambda(u)}{\lambda(u)} - s} = \frac{s}{\zeta(\alpha)\Gamma(\alpha)} \int_0^\infty \frac{t^{\alpha-1}}{e^t - s} dt$$

- The solution to this equation can be found by changing variables on the lhs and matching:

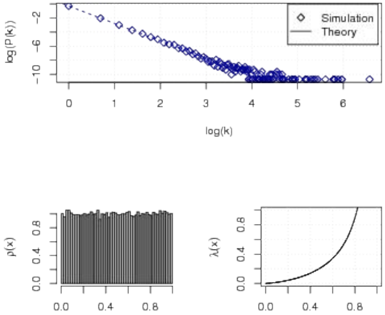
$$\frac{1+\lambda(u)}{\lambda(u)} = e^t \quad \text{and} \quad \frac{du}{\lambda(u)} = -\frac{1}{\zeta(\alpha)\Gamma(\alpha)} t^{\alpha-1} dt \qquad (15)$$

Solution to the inverse and direct problem

- The solution for the power-law degree distribution is implicitly given by

$$\zeta(\alpha)\Gamma(\alpha)\cdot u = \int_0^{\ln\left(\frac{1+\lambda(u)}{\lambda(u)}\right)} \frac{\nu^{\alpha-1}}{1-e^{\nu}} d\nu \qquad (16)$$

- From the solution $\lambda(u)$, necessary conditions on $\rho(x)$ and $f(x,y)$ respectively $\lambda(x)$ can be recovered
- Solution implies a singularity in $\lambda(u)$ at $u = 1$
  - Infinite domain of fitness is compressed on a finite interval of ranks



The general solution

- The above approach is **tailored specifically for the power law**
- In the following we present a general theory for the suggested model
- The solution is obtained in terms of the the **exponential generating function** for the node degree distribution:

$$f(s) = \sum_{k \geq 0} \frac{s^k}{k!} p(k) \tag{17}$$

- $f(s)$ has an infinite radius of convergence, that allows us to do an analytic continuation.

The general solution cont'd

- the **general solution** for $\tilde{\rho}(\lambda)$ in terms of the analytic continuation of $f(s)$, is given by

$$\boxed{\tilde{\rho}(\lambda) = \frac{1}{1+\lambda} \int_{-\infty}^{\infty} \frac{d\sigma}{2\pi} f'(i\sigma) e^{-i\sigma \frac{\lambda}{1+\lambda}}} \quad (18)$$

- Therefore obtaining the necessary form of $\lambda(u)$ to yield an arbitrary solution just needs the calculation of $f'(s)$
- Again, the solutions to the inverse and the direct problem can be obtained from $\lambda(u)$ by applying its definition.

Example

- Consider the following example for a power-law
  - The exp. generating function for a power law is given by:

$$f(s) = \sum_{k \geq 1} \frac{As^k}{k^\alpha k!} = A \int_0^\infty t^{\alpha-1} \left(e^{se^{-t}} - 1 \right) dt \qquad (19)$$

  - substituting into the solution gives:

$$\tilde{\rho}(\lambda) = \frac{1}{\Gamma(\alpha)\zeta(\alpha)} \cdot \frac{1}{1+\lambda} \ln \left(\frac{1+\lambda}{\lambda} \right)^{\alpha-1}. \qquad (20)$$

  - and that is exactly what we had before

Summary and Outlook

- **So far only** fitness induced **static** networks have been studied well
- We have studied their dynamic counterpart
- Two problems are posed in the fitness based literature
  - The direct problem (given $\rho(x)$, what is $f(x,y)$ such that $p(k) \propto k^{-\alpha}$)
  - The inverse problem (given $f(x,y)$ what is $\rho(x)$)
- We found...
  - ... the necessary linking propensity for given node degree distribution
  - ... the resulting degree distribution for given linking propensity

Thanks for your attention.

The work that has been presented is based on:

I.E. Smolyarenko, K. Hoppe, G.J. Rodgers, "A network growth model with intrinsic vertex Fitness.", submitted to Phys. Rev. E