# Decision Making and Social Networks

Lecture 4: Models of Network Growth

Umberto Grandi

Summer 2013

# Overview

In the previous lecture:

- We got acquainted with graphs and networks
- We saw lots of definitions: degree distributions, diameter, centrality measures, clustering...
- We studied a couple of examples: the Medici power structure, six degrees of separation, labour markets

Today we start from the observation that networks are not given nor are they static, they form and grow over time!

Topics of the day:

- Erdös-Rényi random graphs
- Growing random networks: preferential attachment
- Strategic network formation

We are following Chapters 4,5,6 of Jackson's book.

# Erdös-Rényi Random graphs

For each pair of nodes $i$ and $j$ in $N$, create the edge $ij \in g$ with probability $p$.

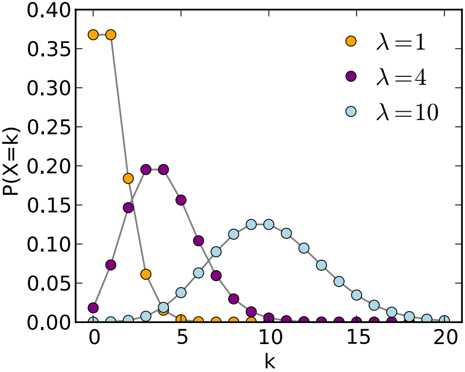
- Number of nodes $n = |N|$
- Probability $p$ of linking two nodes

Very rough model: no influence, no dynamic, but tractable:

- probability of a network with $m$ links is $p^m(1-p)^{\frac{n(n-1)}{2}-m}$
- probability of having $d$ links is $P(d) = \binom{n-1}{d} p^d (1-p)^{n-d-1}$
- Poisson distribution for small $p$ and large $n$:

$$P(d) = \frac{e^{-(n-1)p}((n-1)p)^d}{d!}$$

# Poisson Distribution




$P(d) = \frac{e^{-\lambda}(\lambda)^d}{d!}$ with $\lambda = (n-1)p$ (average degree)

# Properties of Random Graphs I

Difficult to assess: any graph can form!

Call $p(n)$ the probability of a link to form: we can study what happen when $n$ tends to infinity.

But what is a property? We assume is just a subset of graphs.

Examples:

- not having an isolated node: $I(N) = \{g \mid N_i(g) \neq \emptyset\}$
- connectedness: $C(N) = \{g \mid \ell(i,j) \text{ is finite } \forall i,j \in N\}$
- diameter less than $\log(n)$: $D(N) = \{g \mid \ell(i,j) < \log(n) \quad \forall i,j \in N\}$

# Properties of Random Graphs II

**Definition**

*A property $A$ is monotone if when $g \in (A)$ then also $g' \in A(N)$ for all supernetworks $g' \supseteq g$.*

*Observation*. Properties $I$, $C$, $D$ previously defined are monotone.

Recall from probability theory:

- Property $A(N)$ holds almost surely if its probability tends to 1 when $n$ tends to infinity
- Property $A(N)$ holds almost never if its complement holds almost surely
- Threshold functions ...

# Threshold functions (Erdös-Renyi graphs)

We study properties $A(N)$ depending on the shape of $p(n)$

A threshold function $t(n)$ indicates when does a certain phase transition occur with respect to a monotone property $A$:

$$\mathsf{Pr}_{p(n)}[A(N)] \to 1 \text{ when } \frac{p(n)}{t(n)} \to \infty$$

$$\mathsf{Pr}_{p(n)}[A(N)] \to 0 \text{ when } \frac{p(n)}{t(n)} \to 0$$

Property $A$ holds almost surely if $p(n)$ grows more than $t(n)$, and holds almost never if $p(n)$ grows less than $t(n)$.

# Notable Phase Transitions

Notable examples:

- threshold of $\frac{1}{n^2}$ for the first link to emerge (if $p$ grows less in the limit the network is empty)

*Exercise*: show it.

- threshold of $\frac{1}{n^{3/2}}$ to have at least one connected component of size 3
- threshold of $\frac{1}{n}$ to have cycles
- ...

And a theorem by Erdós and Renyi:

**Theorem**

*The threshold function for the network to be connected is* $\frac{log(n)}{n}$.

P. Erdös and A. Rényi. On the Strength of Connectedness of a Random Graph. *Acta Mathematica Hungarica*, 1961.

# An Interesting "Viral" Example

We have:

- $n$ agents linked by a random network $g$
- One of them is infected by a disease/computer virus/revolutionary seed/marketing idea
- The probability of an individual to be immune to the disease is $\pi$

Will the disease spread??

The problem is equivalent to:

- Generate an Erdös-Rényi random graph of size $n$
- Delete $\pi n$ nodes with uniform probability
- identify the component where the infected individual is present!

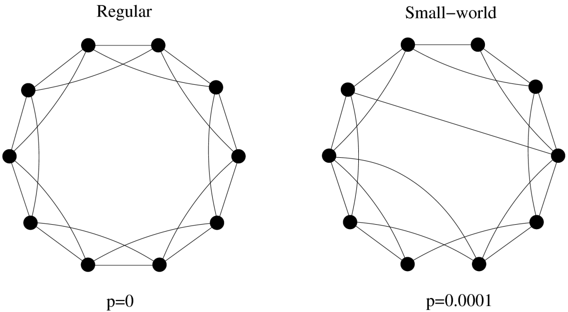
The threshold for having a large connected component is at $p(n)(1-\pi)n = 1$. So if $p(n) < \frac{1}{(1-\pi)n}$ then the disease will die out but if $p(n) > \frac{1}{(1-\pi)n}$ then we have an epidemic!!

Are random graphs realistic?

- Diameter is small: OK
- Clustering? Too small... (average degree grows more slowly than $n$)

# Small World Networks

Small world networks are obtained by rewiring a circle of small worlds connected to each other:




Regular: clustering coefficient is $\frac{1}{2}$ but diameter is in the order of $\frac{n}{4}$.
Small-world: still high clustering, but significantly lower diameter!

D.J Watts and S. Strogatz, Collective Dynamics of Small-world Networks, Nature, 1998.

# Growing Random Networks

Static networks do not take into account the time variable:

- Time increases heterogeneity among nodes of the same age
- "Rich get richer" over time: preferential attachment
- Growth can justify power law distributions
- How do hubs form?


A. Barabási and R. Albert. Emergence of Scaling in Random Networks. *Science*, 1999.

A. Barabási. *Linked: The New Science of Networks*, Perseus, 2002.

# Growing Random Networks

Growth model 1:

Start with $m$ fully connected nodes. When a new node joins it establishes $m$ links uniformly at random

Result: exponential distribution similar to Erdös-Rényi random graphs
Shortcomings: old nodes do not form links

# Scale-free distributions

Scale-free of power law degree distribution

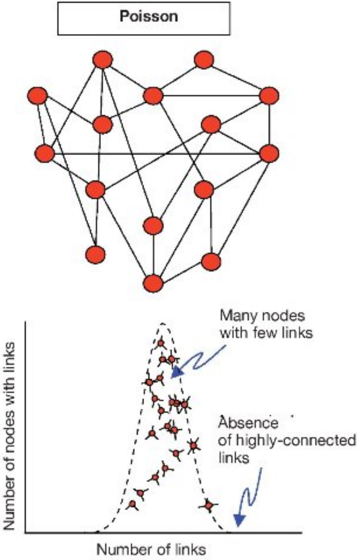
$$P(d) = cd^{-\gamma}$$

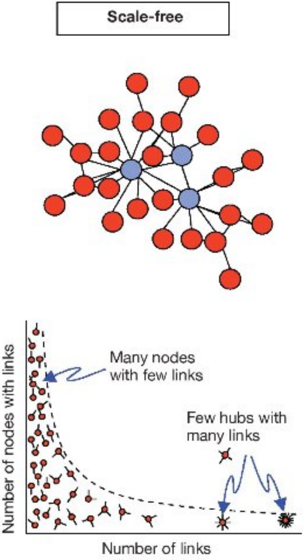
$c$ is there to normalise (sum to 1) the distribution and is positive. Depending on $\gamma$ the distribution can vary significantly $(2 < \gamma < 3)$.

Some properties of scale-free networks:

- used to study wealth distribution (Pareto, 1890).
- scale-free because $\frac{p(kd)}{p(d)} = \frac{p(kc)}{p(c)}$.
- a straight line if plotted on a log-log scale
- fat tails: existence of hubs with many many links
- how do they form? Probability of linking to a node depends on its degree!

# Erdös-Rényi vs. Scale-free



# Preferential Attachment

Growth model 2:

Start with $m$ connected nodes. When a new node joins it establishes $m$ links with a probability distribution that is proportional to the degree of the existing nodes in the network.

Result: scale-free networks.

- at time $t$ the probability that node $i$ gets a new connection is

$$p_i = m \frac{d_i(t)}{\sum_j d_j(t)}$$

- there are $tm$ edges at time $t$, the sum of degrees is $2tm$
- the probability of getting a new link is $\frac{d_i(t)}{2t}$
- from here we build a differential equation/dynamical system and find that the solution is a scale-free network

# Properties of Preferential Attachment Model I

What about the diameter:

**Theorem**

*A network growing with preferential attachment in which each newborn node forms $m \geqslant 2$ links consists of a single component with diameter that is proportional to $\frac{\log(n)}{\log\log(n)}$ (almost surely).*

The diameter is lower than that of a random network: because of hubs!

Bollobàs and Riordan. The Diameter of a Scale-Free Random Graph. *Combinatorica*, 2002.

# Properties of the Preferential Attachment Model II

What about clustering?

Not very good results. In a random network the chance of closing a triangle is:

- the number of pairs to choose from by a newborn node $\frac{t(t-1)}{2}$
- $tm$ is the number of existing links
- probability of closing the circle is $\frac{2m}{t-1}$ that tends to $0$ when $t$ goes to infinity...

Same result for preferential attachment.

# Strategic Network Formation

Basic facts:

- Players (nodes) follow incentives (does not have to be conscious)
- Individual or collective incentives: tension between stability and efficiency
- Provide an answer to why do networks take a certain form rather than how

# Using Game Theory?

Nodes can be viewed as players, each on having utility functions

$$u_i : G(N) \to \mathbb{R}$$

so $u_i(g)$ is the utility of node $i$ in a given network $g$

Why not using Nash equilibria?

- Modelling communication
- Coordination and mutual consent to form a given link! To dismantle a link one player is instead sufficient.

# Pairwise Stability

To formalise the asymmetry of the network formation process:

**Definition**

*A network $g$ is pairwise stable if the following two conditions hold:*

- $\forall ij \in g \;\; u_i(g) \geqslant u_i(g - ij)$ *and* $u_j(g) \geqslant u_j(g - ij)$
- $\forall ij \notin g \quad$ *if* $u_i(g + ij) > u_i(g)$ *then* $u_j(g + ij) < u_j(g)$

Weaknesses of the model:

- one player at a time can move (no coordinated action possible)
- only one edge at a time is considered: it is possible that by deleting all or a subset of its neighbouring links player $i$ can get to a better network

# Efficiency

Given a vector of utilities we can adapt notions of social welfare to define efficient networks.

**Definition**

*A network $g$ is efficient if $\sum_{i \in N} u_i(g) \geqslant \sum_{i \in N} u_i(g')$ for all $g' \in G(N)$*

**Definition**

*A network $g$ is Pareto-efficient if there is no $g' \in G(N)$ such that $u_i(g') \geqslant u_i(g)$ and there exists one $i^*$ such that $u_{i^*}(g') > u_{i^*}(g)$.*

Other notions form the literature may be adapted (and it seems that this problem is yet to be explored...)

# First Result: Characterisation of Efficient Networks

Special case of networks: symmetric connection model - simple version

Ingredients:

- Decreasing benefit function: $\delta^K$ where $K$ is the distance between $i$ and $j$
- Cost $c$ of maintaining a link
- Utility is based on distance:

$$u_i(g) = \sum_{\{j \neq i | j \in N(i)\}} \delta^{\ell_{ij}(g)} - c d_i(g)$$

### Theorem

*The unique efficient network under the simple symmetric connection model is:*

- *the complete network if* $c < \delta - \delta^2$
- *a star involving all nodes if* $\delta - \delta^2 < c < \delta + \frac{n-2}{2}\delta^2$
- *the empty network if* $c > \delta + \frac{n-2}{2}\delta^2$

Bloch and Jackson. The Formation of Networks with Transfers among Players. *Journal of Economic Theory*, 2007.

# Second Result: Stability and Efficiency are Incompatible I

If we refine the model we can prove that stability and efficiency cannot coexist in an optimal network:

**Definition**

*A transfer function is a* $t : G(N) \to \mathbb{R}^N$ *such that* $\sum_{i \in g} t_i(g) = 0$ *for all* $g$*.*

The second condition is the balance: no externalitites, no loss of utility.

**Definition**

*A transfer function is component balanced if* $\sum_{i \in S} t_i(g) = 0$ *for all* $g$ *and connected components* $S$ *of* $g$*.*

A transfer function satisfies equal treatment of equals if $t_i(g) = t_j(g)$ when $i$ and $j$ are "equals" (same connections, same power and utility with respect to the network)

# Second Result: Stability and Efficiency are Incompatible II

## Definition

*A profile of utility functions is component decomposable if* $u_i(g) = u_i(g_{|N_i^n(g)})$ *for all* $g$ *and players* $i$.

## Theorem

*There exist component-decomposable utility functions such that every pairwise-stable network relative to any component-balanced transfer rule satisfying equal treatment of equals is inefficient*

M. O. Jackson and J. Wolinsky. A Strategic Model of Social and Economic Networks. *Journal of Economic theory*, 1996.

# Last Slide

What have we done today:

- Erdös-Rényi random networks: base-line model
- Preferential attachment: hubs and connectors, power-law distributions
- Strategic networks: pairwise stability and efficiency cannot coexist

This is just a very brief introduction. More material:

- M. O. Jackson. Social and Economic Networks. (see also full course on Coursera)
- D. Easley and J. Kleinberg. Networks, Crowds and Markets.
- A. L. Barabási. Linked. Nice popular science book.

More from you:

- Tuesday 25 (11-13): Cristina presents *Influence and aggregation of preferences over combinatorial domains* by N. Maudet, M. S. Pini, F. Rossi and K. B. Venable.
- Tuesday 25 (11-13): Michele presents *Voting in Social Networks* by P. Boldi, F. Bonchi, C. Castillo.
- Thursday 27 (11-13): Andrea presents *Social Network Games* by S. Simon and K.R. Apt.
- Thursday 27 (11-13): (maybe) Pietro presents something about recommendations over social networks

Discussion on final projects on 25 July?