Network models

Leonid E. Zhukov

School of Applied Mathematics and Information Science
National Research University Higher School of Economics

21.12.2013

Main properties

- Power law degree dsitribution
- Small average distance (diameter)
- Large clastering coefficient
- Hierachical structure (often)

Generative models

Models:

- Random graphs (Erdos & Renyi, 1959)
- Small world networks (Watts & Strogatz, 1998)
- Preferential attachment (Barabasi & Albert, 1999)
- .....
- Strategic network formation (Jackson & Wolinksy, 1996)

Random graph models

Erdos and Renyi, 1959.
Graph $G\{E, V\}$, nodes $n = |V|$, edges $m = |E|$
Random graph models

- $G_{n,m}$, a randomly selected graph from the set of $C^m_{n(n-1)/2}$ graphs with n nodes and m edges
- $G_{n,p}$, each pair from n nodes is connected with probability p, m - random number
- A graph with m edges and n nodes

$$\langle m \rangle = p\frac{n(n-1)}{2}$$

$$\langle k \rangle = \frac{1}{n}\sum_i k_i = \frac{2m}{n} = p\ (n-1) \approx pn$$

Phase transition

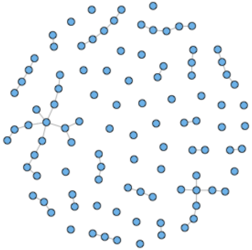
- Graph $G_{n,p}$ -function of p: $p = 0$ - empty graph, $p = 1$ - full graph
- There are exist critical p_c, structural changes from $p < p_c$ to $p > p_c$
- Critical value:

$$p_c = \frac{1}{n}, \ \langle k \rangle = p_c n = 1$$

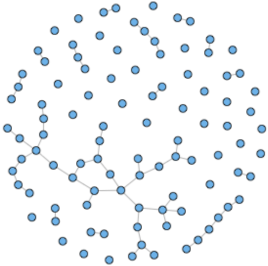
  - when $p < p_c$, $(\langle k \rangle = pn < 1)$ there is no components with more than $O(\ln n)$ nodes, all components are either trees or have maximum one cycle, largest component is a tree
  - when $p = p_c$, $(\langle k \rangle = pn = 1)$ the largest component has $O(n^{2/3})$ nodes
  - when $p > p_c$, $(\langle k \rangle = pn > 1)$ gigantic component has all $O(n)$ nodes
- Phase transition:

  separate connected islands -> gigantic connected component

Random graph model

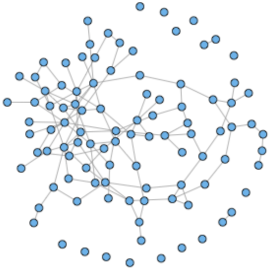


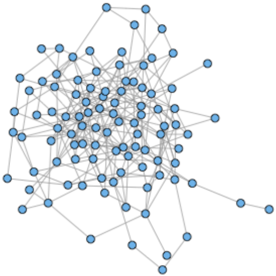
$p < p_c$



$p = p_c$

Random graph model





$p > p_c$ $p >> p_c$

R: erdos.renyi.game {igraph}

Random graph properties

- Node degree distribution - Poisson ($n \to \infty$ at fixed $\langle k \rangle = pn$):

$$P(k) = \frac{\langle k \rangle^k e^{-\langle k \rangle}}{k!}$$

- Graph diameter when $p > p_c$:

$$d = \frac{\ln n}{\ln \langle k \rangle}$$

- Clustering coefficient:

$$C = p = \frac{\langle k \rangle}{n}$$

$C \to 0 \quad \text{when} \quad n \to \infty$

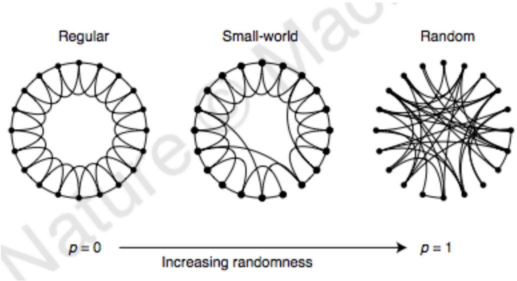
Small world model

Watts and Strogatz, 1998

One parameter model, interpolation between regular lattice and random graph

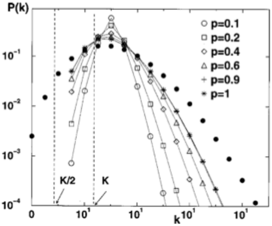
- start with regular lattice with n - nodes, k - nearest neighbors (node degree)
- randomly connect with other nodes with probability p, forms $pnk/2$ "long distance"connections
- $p = 0$ - regular lattice, $p = 1$ - random graph

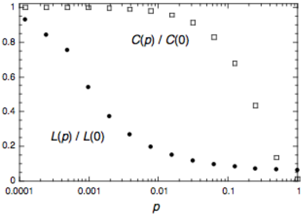
Small world model



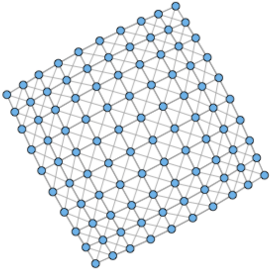
Small world model

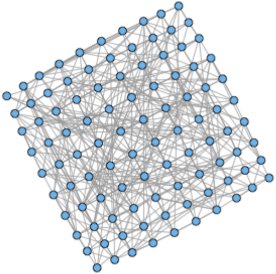
- Node degree distribution:
  Poisson like
- Ave. path length $\langle L(p) \rangle$:
  $p \to 0$, ring lattice, $\langle L(0) \rangle = 2n/k$
  $p \to 1$, random graph, $\langle L(1) \rangle = \log(n)/\log(k)$
- Clustering coefficient $C(p)$:
  $p \to 0$, ring lattice, $C(0) = 3/4 = const$
  $p \to 1$, random graph, $C(1) = k/n$





Small world model





20% rewiring:

ave. path length = 3.58 -> ave. path length = 2.32

clust. coeff = 0.49 -> clust. coeff = 0.19

R: watts.strogatz.game, graph.lattice,rewire.edges {igraph}

Preferential attachment model

Barabasi and Albert, 1999

Dynamical (growth) model

- $t = 0$, n_0 unconnected nodes
- growth: on every step add a node with m edges ($m \leq n_0$)
- Preferential attachment: probability of linking to existing node is proportional to the node degree k_i

$$k_i(t + \delta t) = k_i(t) + \frac{k_i(t)}{2t}\delta t$$

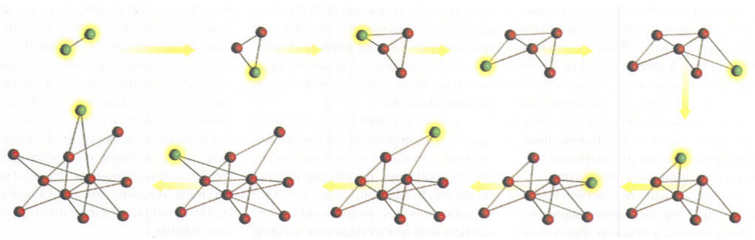
$$\frac{dk_i(t)}{dt} = \frac{k_i(t)}{2t}$$

initial conditions: $k_i(t_i) = m$

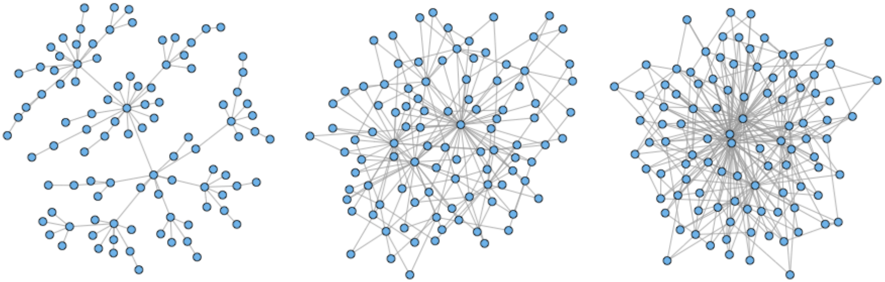
- Time evolution of a node degree

$$k_i(t) = m\left(\frac{t}{t_i}\right)^{1/2}$$

Preferential attachment model



Preferential attachment model



$m = 1$ $m = 2$ $m = 3$

R: barabasi.game {igraph}

Preferential attachment model

- Node distribution function - "power law":

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

- Average path length:

$$\langle L \rangle = \log(n) / \log(\log(n))$$

- Clustering coefficient:

$$C \sim \frac{1}{n^{0.75}}$$

Model comparison

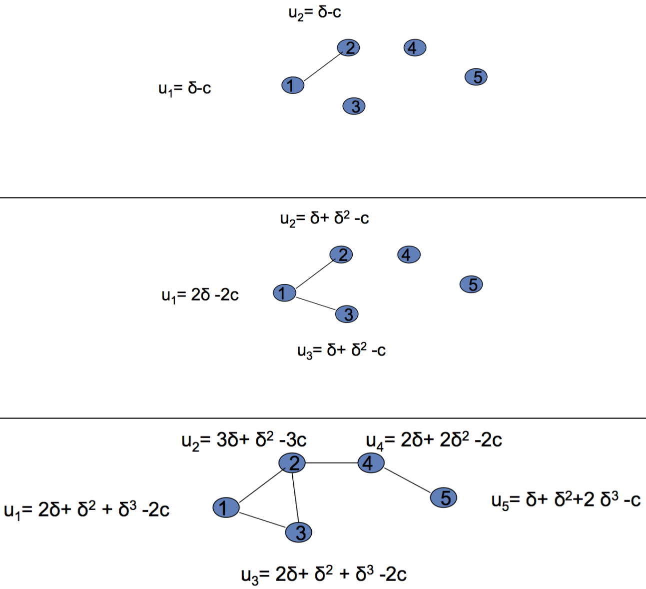
	Random	BA model	WS model	Real World
P(k)	$\frac{z^k e^{-z}}{k!}$	k^{-3}	poisson like	power law
C	$p = \frac{\langle k \rangle}{n}$	$n^{-0.75}$	const	large
$\langle L \rangle$	$\frac{log(n)}{log(\langle k \rangle)}$	$\frac{log(n)}{loglog(n)}$	$log(n)$	small world

Strategic network formation

Jackson and Wolinksy, 1996
"A Strategic Model of Social and Economic Networks"

- why networks becomes the way they are
- people (agents) making rational choices establishing connections
- maximizing individual utility (incentives)
- connections brings costs and benefits
- stability of the network
- social efficency (best for the society)
- friendship, professional, political, trade networks

Distance-based utility function



Utility functions

- Distance-based utility

$$u_i(G) = \sum_{j \in N_i^{n-1}} \delta^{l_{ij}} - d_i \cdot c$$

l_{ij} - shortest path, d_i - node degree, δ, C - parameters, $\delta < 1$

- Co-author model

$$u_i(G) = \sum_{j \in N_i} \left(\frac{1}{d_i} + \frac{1}{d_j} + \frac{1}{d_i d_j}\right), \quad d_i \neq 0$$

$u_i(G) = 1$ if $d_i = 0$, d_i - node degree

Network stability and efficiency

- Evolution: forming a link - mutual consent, removing a link - one person decision
- Network is pairwise stable if no palyer wants to remove a link and no two players want to add a link:

$$\begin{aligned} \forall i \; u_i(G) &\geq u_i(G - e_{ij}) \\ \forall i,j \text{ if } u_i(G + e_{ij}) &> u_i(G), \text{then } u_j(G + e_{ij}) < u_j(G) \end{aligned}$$

- Strong efficiency (”best network”, maximizing total utility for the society):

$$G^* = \max_G \sum_i u_i(G)$$

- Pareto efficiency (impossible to increase anyone’s utility without decresing someone’s else):

$$\begin{aligned} \nexists \; G' : \; u_i(G') &\geq u_i(G) \text{ for all } i \\ \text{and } u_i(G') &> u_i(G) \text{ for one } i \end{aligned}$$

Conclusion

- Complex networks (degree distr, diameter, clustering coeff)
- Node and link analysis (centrality, pagerank)
- Network communities (modularity, detection methods)
- Node similarity (strucutural equiv, distance, similarity matrix)
- Network structure (cliques,cores, motifs, dyads, triads)
- Network visualization
- Network models (random, small world, pref. attachment)