Generational Random Graphs – a “natural” model for heterogeneous temporal networks?

Jon crowcroft 9/3/2017
Graphs aren’t static or homoeogenous

• Re-do two simple small world & clustered models
  • Preferential attachment & re-wiring (alpha & beta) models

• Add one simple idea, but in two guises:
  • Nodes are taken (in batches) from a sequence of generations
  • There's a birth (death) process of new (old) generations
  • To note: discrete generations, but continuous time...

• Two use cases
  • Social media/graphs – parents, siblings, children
  • Tech nets (internet, transport) – dialup, broadband, fiber, 3G/4, ISP/IXP or horse, car, plane, drone...
Random Graphs & Nature

Erdős and Renyi (1959)

$N$ nodes

A pair of nodes has probability $p$ of being connected.

Average degree, $k \approx pN$

What interesting things can be said for different values of $p$ or $k$?

(that are true as $N \to \infty$)

$p = 0.0 ; k = 0$

$p = 0.09 ; k = 1$

$p = 1.0 ; k \approx \frac{1}{2}N^2$
Random Graphs

Erdős and Renyi (1959)

1. Size of the largest connected cluster
2. Diameter (maximum path length between nodes) of the largest cluster
3. Average path length between nodes (if a path exists)
Random Graphs

Erdős and Renyi (1959)

\[ p = 0.0 ; k = 0 \]

\[ p = 0.045 ; k = 0.5 \]

\[ p = 0.09 ; k = 1 \]

\[ p = 1.0 ; k \approx \frac{1}{2} N^2 \]

<table>
<thead>
<tr>
<th>Size of largest component</th>
<th>1</th>
<th>5</th>
<th>11</th>
<th>12</th>
</tr>
</thead>
<tbody>
<tr>
<td>Diameter of largest component</td>
<td>0</td>
<td>4</td>
<td>7</td>
<td>1</td>
</tr>
<tr>
<td>Average path length between nodes</td>
<td>0.0</td>
<td>2.0</td>
<td>4.2</td>
<td>1.0</td>
</tr>
</tbody>
</table>
Random Graphs

Erdős and Renyi (1959)

If $k < 1$:
- small, isolated clusters
- small diameters
- short path lengths

At $k = 1$:
- a giant component appears
- diameter peaks
- path lengths are high

For $k > 1$:
- almost all nodes connected
- diameter shrinks
- path lengths shorten

phase transition
Random Graphs

What does this mean?

- If connections between people can be modeled as a random graph, then...
  - Because the average person easily knows more than one person ($k \gg 1$),
  - We live in a “small world” where within a few links, we are connected to anyone in the world.
  - Erdős and Renyi showed that average path length between connected nodes is

\[
\frac{\ln N}{\ln k}
\]

* An example researcher with Erdos #: 4
Random Graphs

What does this mean?

- If connections between people can be modeled as a random graph, then...
  - Because the average person easily knows more than one person ($k \gg 1$),
  - We live in a “small world” where within a few links, we are connected to anyone in the world.
  - Erdős and Renyi computed average path length between connected nodes to be:

$$\frac{\ln N}{\ln k}$$
The Alpha Model

Watts (1999)

The people you know aren’t randomly chosen.

People tend to get to know those who are two links away (Rapoport *, 1957).

The real world exhibits a lot of clustering.

* Same Anatol Rapoport, known for TIT FOR TAT!
The Alpha Model

Watts (1999)

\(\alpha\) model: Add edges to nodes, as in random graphs, but makes links more likely when two nodes have a common friend.

For a range of \(\alpha\) values:

Probability of linkage as a function of number of mutual friends

(\(\alpha\) is 0 in upper left, 1 in diagonal, and \(\infty\) in bottom right curves.)
The Alpha Model

Watts (1999)

α model: Add edges to nodes, as in random graphs, but makes links more likely when two nodes have a common friend.

For a range of α values:

• The world is small (average path length is short), and
• Groups tend to form (high clustering coefficient).
The Beta Model

Watts and Strogatz (1998), circular lattice, rewiring to random other link w/ probability $\beta$

$\beta = 0$
People know their neighbors.
Clustered, but not a “small world”

$\beta = 0.125$
People know their neighbors, and a few distant people.
Clustered and “small world”

$\beta = 1$
People know others at random.
Not clustered, but “small world”
The Beta Model

First five random links reduce the average path length of the network by half, regardless of $N$!

Both $\alpha$ and $\beta$ models reproduce short-path results of random graphs, but also allow for clustering.

Small-world phenomena occur at threshold between order and chaos.

Clustering coefficient ($C$) and average path length ($L$) plotted against $\beta$
Power Laws

Albert and Barabasi (1999)

What’s the degree (number of edges) distribution over a graph, for real-world graphs?

Random-graph model results in Poisson distribution.

But, many real-world networks exhibit a *power-law* distribution.

Degree distribution of a random graph,

\[ N = 10,000 \quad p = 0.0015 \quad k = 15 \]

(Curve is a Poisson curve, for comparison.)
Power Laws

Albert and Barabasi (1999)

What’s the degree (number of edges) distribution over a graph, for real-world graphs?

Random-graph model results in Poisson distribution.

But, many real-world networks exhibit a *power-law* distribution.

Typical shape of a power-law distribution.
Power Laws

Albert and Barabasi (1999)

Power-law distributions are straight lines in log-log space.

*How should random graphs be generated to create a power-law distribution of node degrees?*

Hint:

Pareto’s* Law: Wealth distribution follows a power law.

Power laws in real networks:
(a) WWW hyperlinks
(b) co-starring in movies
(c) co-authorship of physicists
(d) co-authorship of neuroscientists

* Same Velfredo Pareto, who defined Pareto optimality in game theory.*
Hippogriffically

• Spatial parameter(s) –
  • #generations – e.g. 1,3, infinity
  • Alpha’ (Beta’) now – ratio of preferential attachment (rewire) probability within and between generations –
  • e.g. between siblings, children, parents e.g. (.25, .5, .25) for 3 generations,
  • could be 1/n for n generations or could have a 1/d_i,j for distance between generations or whatever, or pick your distr...

• Temporal parameter(s) markovish...
  • #New Nodes/generation epoch
  • Removal process (perhaps)
So lots of data out there (fb, internet topo over time)

• Fit model params
• Properties now indexed by generation (for example)
  • E.g. cliques for sibling v. family, centrality for grandparents, etc
• What other nets does this describe, intuitively?
• Is it still too complicated/complex?
• Does it make some things easier (or harder)?
• Do we need generational properties to keep global properties
  • Global mean diameter, cluster science, centrality=mean of mean each generation, etc
  • Or can they deviate in weird ways?