Basic idea:

- Random networks with arbitrary degree distributions cover much territory but do not represent all networks.
- Moving away from pure random networks was a key first step.
- We can extend in many other directions and a natural one is to introduce correlations between different kinds of nodes.
- Node attributes may be anything, e.g.:
  1. degree
  2. demographics (age, gender, etc.)
  3. group affiliation
- We speak of mixing patterns, correlations, biases...
- Networks are still random at base but now have more global structure.
- Build on work by Newman [3, 4].

General mixing between node categories

- Assume types of nodes are countable, and are assigned numbers 1, 2, 3, ... .
- Consider networks with directed edges.

\[
e_{\mu\nu} = \text{Pr} \left( \text{an edge connects a node of type } \mu \text{ to a node of type } \nu \right)
\]

\[
a_{\mu} = \text{Pr}(\text{an edge comes from a node of type } \nu)
\]

\[
b_{\mu} = \text{Pr}(\text{an edge leads to a node of type } \nu)
\]

- Write \( E = [e_{\mu\nu}] \), \( \bar{a} = [a_{\mu}] \), and \( \bar{b} = [b_{\mu}] \).
- Requirements:

\[
\sum_{\nu} e_{\mu\nu} = 1, \quad \sum_{\nu} e_{\mu\nu} = a_{\mu}, \quad \text{and} \quad \sum_{\mu} e_{\mu\nu} = b_{\nu}.
\]
Notes:

- Varying $e_{\mu\nu}$ allows us to move between the following:
  
  1. **Perfectly assortative networks** where nodes only connect to like nodes, and the network breaks into subnetworks. Requires $e_{\mu\nu} = 0$ if $\mu \neq \nu$ and $\sum_{\mu} e_{\mu\mu} = 1$.
  
  2. **Uncorrelated networks** (as we have studied so far) For these we must have independence: $e_{\mu\nu} = a_{\mu} b_{\nu}$.
  
  3. **Disassortative networks** where nodes connect to nodes distinct from themselves.

- Disassortative networks can be hard to build and may require constraints on the $e_{\mu\nu}$.

- Basic story: level of assortativity reflects the degree to which nodes are connected to nodes within their group.

**Correlation coefficient:**

- Quantify the level of assortativity with the following **assortativity coefficient**[^4]:

$$ r = \frac{\sum_{\mu} e_{\mu\mu} - \sum_{\mu} a_{\mu} b_{\mu}}{1 - \sum_{\mu} a_{\mu} b_{\mu}} = \frac{\text{Tr} \ E - ||E^2||_1}{1 - ||E^2||_1} $$

where $|| \cdot ||_1$ is the 1-norm = sum of a matrix’s entries.

- $\text{Tr} \ E$ is the fraction of edges that are within groups.

- $||E^2||_1$ is the fraction of edges that would be within groups if connections were random.

- $1 - ||E^2||_1$ is a normalization factor so $r_{\text{max}} = 1$.

- When $\text{Tr} \ e_{\mu\mu} = 1$, we have $r = 1.$ ✓

- When $e_{\mu\mu} = a_{\mu} b_{\mu}$, we have $r = 0.$ ✓

**Scalar quantities**

- Now consider nodes defined by a scalar integer quantity.

- Examples: age in years, height in inches, number of friends, ...

- $e_{jk} = \text{Pr}$ a randomly chosen edge connects a node with value $j$ to a node with value $k$.

- $a_j$ and $b_k$ are defined as before.

- Can now measure correlations between nodes based on this scalar quantity using standard **Pearson correlation coefficient** ($\mathbb{E}$):

$$ r = \frac{\sum_{j,k} k(e_{jk} - a_j b_k)}{\sigma_a \sigma_b} = \frac{\langle jk \rangle - \langle j \rangle \langle k \rangle}{\sqrt{\langle j^2 \rangle \langle k \rangle} - \langle j \rangle^2 \sqrt{\langle k^2 \rangle} - \langle k \rangle^2} $$

- This is the observed normalized deviation from randomness in the product $jk$. 

[^4]: The correlation coefficient quantifies the level of assortativity in a network, reflecting the degree to which nodes are connected to nodes within their group.
Degree-degree correlations

- Natural correlation is between the degrees of connected nodes.
- Now define \( e_{jk} \) with a slight twist:

\[
e_{jk} = \Pr(\text{an edge connects a degree } j + 1 \text{ node to a degree } k + 1 \text{ node})
\]

\[
= \Pr(\text{an edge runs between a node of in-degree } j \text{ and a node of out-degree } k)
\]

- Useful for calculations (as per \( R_k \))
- Important: Must separately define \( P_0 \) as the \( \{e_{jk}\} \) contain no information about isolated nodes.
- Directed networks still fine but we will assume from here on that \( e_{jk} = e_{kj} \).

Degree-degree correlations

Error estimate for \( r \):
- Remove edge \( i \) and recompute \( r \) to obtain \( r_i \).
- Repeat for all edges and compute using the jackknife method (\[\text{Box}\])\[\text{1}\]

\[
\sigma_r^2 = \sum_i (r_i - r)^2.
\]

- Mildly sneaky as variables need to be independent for us to be truly happy and edges are correlated...

Measurements of degree-degree correlations

<table>
<thead>
<tr>
<th>Group</th>
<th>Network</th>
<th>Type</th>
<th>Size ( n )</th>
<th>Assortativity ( r )</th>
<th>Error ( \sigma_r )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Social</td>
<td>Physics coauthorship</td>
<td>undirected</td>
<td>52 909</td>
<td>0.363</td>
<td>0.002</td>
</tr>
<tr>
<td></td>
<td>Biology coauthorship</td>
<td>undirected</td>
<td>1 520 251</td>
<td>0.127</td>
<td>0.0004</td>
</tr>
<tr>
<td></td>
<td>Mathematics coauthorship</td>
<td>undirected</td>
<td>253 339</td>
<td>0.120</td>
<td>0.002</td>
</tr>
<tr>
<td></td>
<td>Film actor collaborations</td>
<td>undirected</td>
<td>449 913</td>
<td>0.208</td>
<td>0.0002</td>
</tr>
<tr>
<td></td>
<td>Company directors</td>
<td>undirected</td>
<td>7 673</td>
<td>0.276</td>
<td>0.004</td>
</tr>
<tr>
<td></td>
<td>Student relationships</td>
<td>undirected</td>
<td>573</td>
<td>0.029</td>
<td>0.037</td>
</tr>
<tr>
<td></td>
<td>Email address books</td>
<td>directed</td>
<td>16 881</td>
<td>0.092</td>
<td>0.004</td>
</tr>
<tr>
<td>Technological</td>
<td>Power grid</td>
<td>undirected</td>
<td>4 941</td>
<td>0.003</td>
<td>0.013</td>
</tr>
<tr>
<td></td>
<td>Internet</td>
<td>undirected</td>
<td>10 697</td>
<td>0.189</td>
<td>0.002</td>
</tr>
<tr>
<td></td>
<td>World Wide Web</td>
<td>directed</td>
<td>269 504</td>
<td>0.067</td>
<td>0.0002</td>
</tr>
<tr>
<td></td>
<td>Software dependencies</td>
<td>directed</td>
<td>3 162</td>
<td>0.016</td>
<td>0.020</td>
</tr>
<tr>
<td>Biological</td>
<td>Protein interactions</td>
<td>undirected</td>
<td>2 115</td>
<td>0.156</td>
<td>0.010</td>
</tr>
<tr>
<td></td>
<td>Metabolic network</td>
<td>undirected</td>
<td>765</td>
<td>0.240</td>
<td>0.007</td>
</tr>
<tr>
<td></td>
<td>Neural network</td>
<td>directed</td>
<td>307</td>
<td>0.226</td>
<td>0.016</td>
</tr>
<tr>
<td></td>
<td>Marine food web</td>
<td>directed</td>
<td>134</td>
<td>0.263</td>
<td>0.037</td>
</tr>
<tr>
<td></td>
<td>Freshwater food web</td>
<td>directed</td>
<td>92</td>
<td>0.326</td>
<td>0.031</td>
</tr>
</tbody>
</table>

- Social networks tend to be assortative (homophily)
- Technological and biological networks tend to be disassortative
Next: Generalize our work for random networks to degree-correlated networks.
As before, by allowing that a node of degree $k$ is activated by one neighbor with probability $\beta_k$, we can handle various problems:
1. find the giant component size.
2. find the probability and extent of spread for simple disease models.
3. find the probability of spreading for simple threshold models.

Goal: Find $f_{n,j} = \Pr$ an edge emanating from a degree $j + 1$ node leads to a finite active subcomponent of size $n$.

Repeat: a node of degree $k$ is in the game with probability $\beta_k$.
Define $\vec{\beta}_1 = [\beta_k]$.
Plan: Find the generating function $F_j(x; \vec{\beta}_1) = \sum_{n=0}^{\infty} f_{n,j} x^n$.

Recursive relationship:

$$F_j(x; \vec{\beta}_1) = x^0 \sum_{k=0}^{\infty} \frac{e_k}{R_j} (1 - \beta_{k+1,1}) + x \sum_{k=0}^{\infty} \frac{e_k}{R_j} \beta_{k+1,1} \left[ F_k(x; \vec{\beta}_1) \right]^k.$$

First term = $\Pr$ that the first node we reach is not in the game.
Second term involves $\Pr$ we hit an active node which has $k$ outgoing edges.
Next: find average size of active components reached by following a link from a degree $j$ node = $F_j'(1; \vec{\beta}_1)$.

Differentiate $F_j(x; \vec{\beta}_1)$, set $x = 1$, and rearrange.
We use $F_k(1; \vec{\beta}_1) = 1$ which is true when no giant component exists. We find:

$$R_j F_j'(1; \vec{\beta}_1) = \sum_{k=0}^{\infty} e_{jk} \beta_{k+1,1} + \sum_{k=0}^{\infty} k e_{jk} \beta_{k+1,1} F_k'(1; \vec{\beta}_1).$$

Rearranging and introducing a sneaky $\delta_{jk}$:

$$\sum_{k=0}^{\infty} (\delta_{jk} R_k - k \beta_{k+1,1} e_{jk}) F_k'(1; \vec{\beta}_1) = \sum_{k=0}^{\infty} e_{jk} \beta_{k+1,1}.$$
Spreading on degree-correlated networks

- In matrix form, we have

\[ \mathbf{A}_{E,\vec{\beta}_1} \vec{F}'(1; \vec{\beta}_1) = \mathbf{E}\vec{\beta}_1 \]

where

\[
\begin{align*}
[A_{E,\vec{\beta}_1}]_{j+1,k+1} &= \delta_{jk} R_k - k \beta_{k+1,1} e_{jk}, \\
[\vec{F}'(1; \vec{\beta}_1)]_{k+1} &= F_k'(1; \vec{\beta}_1), \\
[E]_{j+1,k+1} &= e_{jk}, \text{ and } [\vec{\beta}_1]_{k+1} = \beta_{k+1,1}. 
\end{align*}
\]

Spreading on degree-correlated networks

- General condition details:

\[ \det \mathbf{A}_{E,\vec{\beta}_1} = \det \left( \delta_{jk} R_k - (k - 1) \beta_{k,1} e_{j-1,k-1} \right) = 0. \]

- The above collapses to our standard contagion condition when \( e_{jk} = R_j R_k \).

- When \( \vec{\beta}_1 = \vec{\beta} \bar{T} \), we have the condition for a simple disease model's successful spread

\[ \det \left( \delta_{jk} R_k - \beta (k - 1) e_{j-1,k-1} \right) = 0. \]

- When \( \vec{\beta}_1 = \vec{1} \), we have the condition for the existence of a giant component:

\[ \det \left( \delta_{jk} R_k - (k - 1) e_{j-1,k-1} \right) = 0. \]

- Bonusville: We'll find another (possibly better) version of this set of conditions later...

Spreading on degree-correlated networks

- So, in principle at least:

\[ \vec{F}'(1; \vec{\beta}_1) = \mathbf{A}_{E,\vec{\beta}_1}^{-1} \mathbf{E}\vec{\beta}_1. \]

- Now: as \( \vec{F}'(1; \vec{\beta}_1) \), the average size of an active component reached along an edge, increases, we move towards a transition to a giant component.

- Right at the transition, the average component size explodes.

- Exploding inverses of matrices occur when their determinants are 0.

- The condition is therefore:

\[ \det \mathbf{A}_{E,\vec{\beta}_1} = 0. \]

- We’ll next find two more pieces:

  1. \( P_{\text{trig}} \), the probability of starting a cascade

  2. \( S \), the expected extent of activation given a small seed.

Triggering probability:

- Generating function:

\[ H(x; \vec{\beta}_1) = x \sum_{k=0}^{\infty} P_k \left[ F_{k-1}(x; \vec{\beta}_1) \right]^k. \]

- Generating function for vulnerable component size is more complicated.
Spreading on degree-correlated networks

- Want probability of not reaching a finite component.
  
  \[ P_{\text{orig}} = S_{\text{orig}} = 1 - H(1; \beta_1) \]
  
  \[ = 1 - \sum_{k=0}^{\infty} P_k \left[ F_{k-1}(1; \beta_1) \right]^k. \]

- Last piece: we have to compute \( F_{k-1}(1; \beta_1) \).
- Nastier (nonlinear)—we have to solve the recursive expression we started with when \( x = 1 \):
  
  \[ F_j(1; \beta_1) = \sum_{k=0}^{\infty} \frac{\phi_k}{R_j} (1 - \beta_{k+1,1}) + \sum_{k=0}^{\infty} \frac{\phi_k}{R_j} \beta_{k+1,1} \left[ F_k(1; \beta_1) \right]^k. \]

- Iterative methods should work here.

Spreading on degree-correlated networks

As before, these equations give the actual evolution of \( \phi_t \) for synchronous updates.

- Contagion condition follows from \( \bar{\theta}_{t+1} = \bar{G}(\bar{\theta}_t) \) equation.
- Need small \( \bar{\theta}_0 \) to take off so we linearize \( \bar{G} \) around \( \bar{\theta}_0 = 0 \).

\[ \bar{\theta}_{j,t+1} = G_j(0) + \sum_{k=1}^{\infty} \frac{\partial G_j(0)}{\partial \theta_{k,t}} \bar{\theta}_{k,t} + \sum_{k=1}^{\infty} \frac{\partial^2 G_j(0)}{\partial \theta_{k,t}^2} \bar{\theta}_{k,t}^2 + \ldots \]

- Largest eigenvalue of \( \frac{\partial G_j(0)}{\partial \theta_{k,t}} \) must exceed 1.
- Condition for spreading is therefore dependent on eigenvalues of this matrix:

\[ \frac{\partial G_j(0)}{\partial \theta_{k,t}} = \frac{\epsilon_{j-1,k-1}}{R_{j-1}} (k - 1) (\beta_{k1} - \beta_{k0}). \]

Insert question from assignment 5 (aaS)

How the giant component changes with assortativity

- More assortative networks percolate for lower average degrees.
- But disassortative networks end up with higher extents of spreading.

References
References I

B. Efron and C. Stein.

J. P. Gleeson.

M. Newman.