Assortativity and Mixing
Complex Networks
CSYS/MATH 303, Spring, 2011
Prof. Peter Dodds
Department of Mathematics & Statistics
Vermont Advanced Computing Center
University of Vermont

Licensed under the Creative Commons Attribution-NonCommercial-ShareAlike 3.0 License.

Outline
Definition
General mixing
Assortativity by degree
Contagion
References

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\textsuperscript{[4, 5]}, and Boguñá and Serano.\textsuperscript{[1]}

General mixing between node categories
- Assume types of nodes are countable, and are assigned numbers 1, 2, 3, ….
- Consider networks with directed edges.

\[ e_{\mu\nu} = \Pr(\text{an edge connects a node of type } \mu \text{ to a node of type } \nu) \]
\[ a_{\mu} = \Pr(\text{an edge comes from a node of type } \mu) \]
\[ b_{\nu} = \Pr(\text{an edge leads to a node of type } \nu) \]
- Write \( E = [e_{\mu\nu}], \tilde{a} = [a_{\mu}], \text{ and } \tilde{b} = [b_{\nu}] \).
- Requirements:
  \[ \sum_{\mu} e_{\mu\nu} = 1, \sum_{\nu} e_{\mu\nu} = a_{\mu}, \text{ and } \sum_{\mu} e_{\mu\nu} = b_{\nu}. \]

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 \( \sum_{\nu} e_{\mu\nu} = 1 \).
  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 are 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\textsuperscript{[10]}:

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

where \( \| - \|_1 \) 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 \( \text{Tr } E = 1 \), we have \( r = 1. \)
- When \( e_{\mu\nu} = a_{\mu} b_{\nu} \), we have \( r = 0. \)
Correlation coefficient:

Notes:
- \( r = -1 \) is inaccessible if three or more types are present.
- Disassortative networks simply have nodes connected to unlike nodes—no measure of how unlike nodes are.
- Minimum value of \( r \) occurs when all links between non-like nodes: \( \text{Tr} c_{ij} = 0 \).
- 
  \[
  r_{\text{min}} = \frac{-||E||_1}{1 - ||E||_1},
  \]

  where \(-1 \leq r_{\text{min}} < 0\).

Scalar quantities
- Now consider nodes defined by a scalar integer quantity.
- Examples: age in years, height in inches, number of friends...
- \( e_{jk} = Pr \) (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 (iii):

  \[
  r = \frac{\sum_j \sum_k i_k j_k (e_{jk} - a_j b_k)}{\sigma_a \sigma_b} = \frac{\langle jk \rangle - \langle j \rangle \langle k \rangle}{\sqrt{\langle j^2 \rangle_a - \langle j \rangle_a^2} \sqrt{\langle k^2 \rangle_b - \langle k \rangle_b^2}}
  \]

  This is the observed normalized deviation from randomness in the product \( jk \).

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

  \[
  e_{jk} = \begin{cases} 
  1 & \text{an edge connects a degree } j+1 \text{ node to a degree } k+1 \text{ node} \\
  0 & \text{an edge runs between a node of in-degree } j \\
  & \text{and a node of out-degree } k
  \end{cases}
  \]

  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
- Notation reconciliation for undirected networks:

  \[
  r = \frac{\sum_j \sum_k i_k j_k (e_{jk} - R_j R_k)}{\sigma_R^2}
  \]

  where, as before, \( R_k \) is the probability that a randomly chosen edge leads to a node of degree \( k + 1 \), and

  \[
  \sigma_R^2 = \sum_j \langle R_j \rangle - \left( \sum_j \langle j \rangle \right)^2.
  \]

  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</th>
<th>Assortativity ( r )</th>
<th>Error ( \sigma_r )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Social</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>a</td>
<td>Physics coauthorship</td>
<td>undirected</td>
<td>52 909</td>
<td>0.363</td>
<td>0.0002</td>
</tr>
<tr>
<td>b</td>
<td>Biology coauthorship</td>
<td>undirected</td>
<td>53 231</td>
<td>0.327</td>
<td>0.0004</td>
</tr>
<tr>
<td>c</td>
<td>Mathematics coauthorship</td>
<td>undirected</td>
<td>253 139</td>
<td>0.120</td>
<td>0.0002</td>
</tr>
<tr>
<td>d</td>
<td>Film actor collaborations</td>
<td>undirected</td>
<td>449 013</td>
<td>0.208</td>
<td>0.0002</td>
</tr>
<tr>
<td>e</td>
<td>Company directors</td>
<td>undirected</td>
<td>7 673</td>
<td>0.276</td>
<td>0.004</td>
</tr>
<tr>
<td>f</td>
<td>Student relationships</td>
<td>undirected</td>
<td>573</td>
<td>0.029</td>
<td>0.037</td>
</tr>
<tr>
<td>g</td>
<td>Email address books</td>
<td>directed</td>
<td>18 583</td>
<td>0.002</td>
<td>0.004</td>
</tr>
<tr>
<td>h</td>
<td>Social network</td>
<td>directed</td>
<td>3 362</td>
<td>0.006</td>
<td>0.020</td>
</tr>
<tr>
<td>i</td>
<td>World Wide Web</td>
<td>directed</td>
<td>269 004</td>
<td>0.067</td>
<td>0.0002</td>
</tr>
<tr>
<td>j</td>
<td>Software dependencies</td>
<td>directed</td>
<td>2 441</td>
<td>0.005</td>
<td>0.014</td>
</tr>
<tr>
<td>k</td>
<td>Protein interactions</td>
<td>undirected</td>
<td>2 115</td>
<td>0.156</td>
<td>0.010</td>
</tr>
<tr>
<td>l</td>
<td>Molecule network</td>
<td>directed</td>
<td>765</td>
<td>0.240</td>
<td>0.027</td>
</tr>
<tr>
<td>m</td>
<td>Neural network</td>
<td>directed</td>
<td>307</td>
<td>0.226</td>
<td>0.016</td>
</tr>
<tr>
<td>n</td>
<td>Marine food web</td>
<td>directed</td>
<td>134</td>
<td>0.263</td>
<td>0.037</td>
</tr>
<tr>
<td>o</td>
<td>Protein food web</td>
<td>directed</td>
<td>92</td>
<td>0.226</td>
<td>0.013</td>
</tr>
</tbody>
</table>

- Social networks tend to be assortative (homophily)
- Biological and technological networks tend to be disassortative
Spreading on degree-correlated networks

- 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 $B_{k1}$, 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.

Spreading on degree-correlated networks

- **Goal:** Find $f_{nj} = \Pr$ 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 $B_{k1}$.
- Define $\hat{B}_1 = [B_{k1}]$.
- **Plan:** Find the generating function $F_j(x; \hat{B}_1) = \sum_{n=0}^{\infty} f_{nj}x^n$.

Spreading on degree-correlated networks

- **Recursive relationship:**
  
  \[ F_j(x; \hat{B}_1) = x^0 \sum_{k=0}^{\infty} \frac{\theta_{jk}}{R_j} (1 - B_{k+1,1}) \]
  
  \[ + x \sum_{k=0}^{\infty} \frac{\theta_{jk}}{R_j} B_{k+1,1} \left( F_j(x; \hat{B}_1) \right)^k. \]

- **First term** is $\Pr$ 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 $j + 1$ node $= F_j(1; \hat{B}_1)$.

Spreading on degree-correlated networks

- Differentiate $F_j(x; \hat{B}_1)$, set $x = 1$, and rearrange.
- We use $F_k(1; \hat{B}_1) = 1$ which is true when no giant component exists. We find:
  
  \[ R_j F_j(1; \hat{B}_1) = \sum_{k=0}^{\infty} \theta_{jk} B_{k+1,1} + \sum_{k=0}^{\infty} k \theta_{jk} B_{k+1,1} F_k(1; \hat{B}_1). \]

- Rearranging and introducing a sneaky $\delta_{jk}$:
  
  \[ \sum_{k=0}^{\infty} (\delta_{jk} R_k - kB_{k+1,1} \theta_{jk}) F_k(1; \hat{B}_1) = \sum_{k=0}^{\infty} \theta_{jk} B_{k+1,1}. \]

- In matrix form, we have
  
  \[ A_{E, \hat{B}_1} \hat{F}(1; \hat{B}_1) = E \hat{B}_1 \]

  where
  
  \[ A_{E, \hat{B}_1} \left[ F_k(1; \hat{B}_1) \right]_{j+k=1} = \delta_{jk} R_k - kB_{k+1,1} \theta_{jk}, \]

  \[ E_{j+k=1} = \theta_{jk}, \] and $\hat{B}_1 = B_{k+1,1}$.

So, in principle at least:

\[ \hat{F}(1; \hat{B}_1) = A_{E, \hat{B}_1}^{-1} E \hat{B}_1. \]

- Now: as $\hat{F}(1; \hat{B}_1)$, 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:

\[ \det A_{E, \hat{B}_1} = 0. \]
Generating function: $S$

General condition details:
- The above collapses to our standard contagion condition when $\phi_k = R_k R_k$.
- When $\tilde{B} = B^T$, we have the condition for a simple disease model's successful spread
  \[ \det [\delta_{jk} R_{k-1} - B(k - 1) e_{j-1,k-1}] = 0. \]
- $\tilde{B}_1 = B^T$, we have the condition for the existence of a giant component:
  \[ \det [\delta_{jk} R_{k-1} - (k - 1) e_{j-1,k-1}] = 0. \]
- Bonusville: We'll find a much better version of this set of conditions later...

Spreading on degree-correlated networks

We'll next find two more pieces:
1. $P_{\text{rig}}$, the probability of starting a cascade
2. $S$, the expected extent of activation given a small seed.

Triggering probability:
- Generating function:
  \[ H(x; \tilde{B}_1) = x \sum_{k=0}^{\infty} P_k \left[ F_{k-1}(x; \tilde{B}_1) \right]^k. \]
- Generating function for vulnerable component size is more complicated.

Spreading on degree-correlated networks

Want probability of not reaching a finite component.

\[ P_{\text{rig}} = S_{\text{rig}} = 1 - H(1; \tilde{B}_1) = 1 - \sum_{k=0}^{\infty} P_k \left[ F_{k-1}(1; \tilde{B}_1) \right]^k. \]

- Last piece: we have to compute $F_{k-1}(1; \tilde{B}_1)$.
- Nastier (nonlinear)—we have to solve the recursive expression we started with when $x = 1$:
  \[ F_j(1; \tilde{B}_1) = \sum_{k=0}^{\infty} \frac{\delta_{jk}}{R_k} (1 - B_{k+1,1})^k + \sum_{k=0}^{\infty} \frac{\delta_{jk}}{R_k} B_{k+1,1} \left[ F_k(1; \tilde{B}_1) \right]^k. \]
- Iterative methods should work here.

Spreading on degree-correlated networks

- Truly final piece: Find final size using approach of Gleeson [3], a generalization of that used for uncorrelated random networks.
- Need to compute $\theta_{j,t}$, the probability that an edge leading to a degree $j$ node is infected at time $t$.
- Evolution of edge activity probability:
  \[ \theta_{j,t+1} = G_j(\tilde{\theta}_t) = \phi_0 + (1 - \phi_0) \sum_{k=0}^{\infty} P_k \sum_{i=0}^{k-1} \binom{k-1}{i} \theta_{k,i}^j (1 - \theta_{k,i})^{k-1-i} B_{ki}. \]
- Overall active fraction's evolution:
  \[ \phi_{j,t+1} = \phi_0 + (1 - \phi_0) \sum_{k=0}^{\infty} P_k \sum_{i=0}^{k-1} \binom{k-1}{i} \theta_{k,i}^j (1 - \theta_{k,i})^{k-1-i} B_{ki}. \]

Spreading on degree-correlated networks

- As before, these equations give the actual evolution of $\phi_j$ for synchronous updates.
- Contagion condition follows from $\tilde{\theta}_{j,t+1} = G_j(\tilde{\theta}_t)$.
- Expand $G$ around $\tilde{\theta}_t = 0$.
  \[ \tilde{\theta}_{j,t+1} = G_j(0) + \sum_{k=1}^{\infty} \frac{\partial G_j(0)}{\partial \theta_{k,t}} \theta_{k,t} + \frac{1}{2!} \sum_{k=1}^{\infty} \frac{\partial^2 G_j(0)}{\partial \theta_{k,t}^2} \theta_{k,t}^2 + \ldots \]
  \[ + \text{If } G_j(0) \neq 0 \text{ for at least one } j, \text{ always have some infection.} \]
  \[ + \text{If } G_j(0) = 0 \forall j, \text{ want largest eigenvalue } \frac{\partial G_j(0)}{\partial \theta_{k,t}} \geq 1. \]
- Condition for spreading is therefore dependent on eigenvalues of this matrix:
  \[ \frac{\partial G_j(0)}{\partial \theta_{k,t}} = \frac{\delta_{j,k-1}}{R_k} (k - 1) B_{k1} \]

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 I

   Generalized percolation in random directed networks.

   The jackknife estimate of variance.

   Cascades on correlated and modular random networks.

   Assortative mixing in networks.

References II

   Mixing patterns in networks.