Homophily (or Assortativity)

Prof. Ralucca Gera,
Applied Mathematics Dept.
Naval Postgraduate School
Monterey, California
rgera@nps.edu
Learning Outcomes

✓ Understand how to measure that nodes with similar characteristics tend to cluster,
  • Based on enumerative characteristics (nationality)
  • Based on scalar characteristics (age, grade)
  • Based on degree.

✓ Analyze network using homophily by identifying the assortativity values based on various characteristics.

✓ Evaluate: consider why behind the what is the assortativity values of your network.
Are hubs adjacent to hubs?

• Real networks usually show a non-zero degree correlation (defined later compared to random).
  – If it has a positive degree correlation, the network has assortatively mixed degrees ( assort. based other attributes can also be considered).
  – If it is negative, it is disassortative.

• According to Newman, social networks tend to be assortatively mixed, while other kinds of networks are generally disassortatively mixed.
Homophily or assortativity

Sociologists have observed network partitioning based on the following characteristics:

– Friendship, acquaintances, business relationships
– Relationships based on certain characteristics:
  • Age
  • Nationality
  • Language
  • Education
  • Income level

**Homophily** is the tendency of individuals to choose friends with similar characteristics. “Like links with like.”
Homophily or assortativity is a common property of social networks (but not necessary):

- Papers in citation networks tend to cite papers in the same field
- Websites tend to point to websites in the same language
- Political views
- Race
- Obesity
Example of homophily

Assortativity by race

The Social Structure of “Countryside” School District

Points Colored by Race

- White
- Black
- Mixed/Other
Titter data: political retweet network
Red = Republicans
Blue = Democrats

Note that they mostly tweet and re-tweet to each other
• Disassortative mixing: “like links with dislike”.

• Dissasortative networks are the ones in which adjacent nodes tend to be dissimilar:
  – Dating network (females/males)
  – Food web (predator/prey)
  – Economic networks (producers/consumers)
Why?
Why?

• Identifying people of interest could be easy if the network presents homophily

• When assortivity (homophily) is low lie Pokec, RedLearnRS (machine learning algorithm that depends on the count of POIs neighbors of nodes) outperforms all other strategies.
• When attributes show high homophily, RedLearnRS performs quite similar to the other algorithms.
How?
The Radial Axis Layout groups nodes and draws the groups in axes:
• Group nodes by degree, in degree, out degree, etc.
• Group nodes by attribute sort (based on data type of attribute).
• Draw axes/spars in ascending or descending order.
• Allows top, middle or bottom "knockdown" of axes/spars, along with ability to specify number of spars resulting after knockdown.
Homophily in Gephi

Run Radial Axis Layout [here](https://gephi.org/users/tutorial-layouts/)

Run the layout by applying the following settings step by step:

- Group nodes by = “Degree”
- Group nodes by = “Modularity Class”
- Order nodes by = “Degree”

- Draw spar/axis as spiral = checked
- Draw spar/axis as spiral = unchecked
- Ascending order = checked

[Run] Homophily by degree?

Distribution of nodes by degree inside each community.

Better show links inside communities

Better show links between communities
An example: ordered by communities
• To check an attribute’s assortativity:
  
  ```python
  assortivity_val = nx.attribute_assortativity_coefficient(G, "color")
  ```

  The attribute “color” can be replaced by other attributes that your data was tagged with.

• If the attribute is “degree” then we obtain degree assortativity:
  
  ```python
  r = nx.degree_assortativity_coefficient(G)
  ```

• If the attribute is “communities” then we obtain modularity:
  
What?
Assortative mixing (homophily)

We will study two types of assortative mixing:

1. Based on enumerative characteristics (the characteristics don’t fall in any particular order):
   1. Nationality
   2. Race
   3. Gender
   4. Communities

2. Based on scalar characteristics, such as:
   1. Age
   2. Income
   3. By degree: high degree connect to high degree
Based on enumerative characteristics (characteristics that don’t fall in any particular order), such as:

- Nationality
- Race
- Gender
- Or just communities
A network is **assortative** if there is a significant fraction of edges between same-type vertices

- How to quantify the assortativity, \( r \), of a network?

**Method 1:**
- Define \( c_i \) to be the class of vertex \( I \), and tag the nodes to belong to each class \( c_i \) →

\[
r_i = \frac{\text{# edges within } C_i}{\text{all possible edges}} \quad \Rightarrow \quad r = \sum_i r_i
\]

Then: What is the assortativity if we consider \( c_i = V(G) \) as the only class? Does it make sense?
Method 2 (used): compare the assortativity of the current network to the one of a random graph:

• Compute the fraction of edges in $c_i$ in the given network,
• Compute the fraction of edges in $c_i$ in a random graph,
• $r$ is their difference.

This is the same process used for similarity, rather counting edges between nodes instead of neighbors of pairs of nodes.
Let $c_i$ be the class of vertex $i$.
Let $n_c$ be the total number of classes.

Let $\delta(c_i, c_j) = \begin{cases} 
1, & \text{if } i = j \\
0, & \text{if } i \neq j
\end{cases}$ be the Kronecker $\delta$ that accounts for vertices in the same class.

Then the number of edges of the same type is:

$$r = \sum_{ij \in E(G)} \delta(c_i, c_j) = \frac{1}{2} \sum_{ij} a_{ij} \delta(c_i, c_j)$$

Checks if vertices are in the same class
Checks for adjacent nodes
Compute the fraction of edges in $c_i$ in the random network

- Construct a random graph with the same degree distrib.
- Let $c_i$ be the class of vertex $i$
- Let $n_c$ be the total number of classes
- Let $\delta(c_i, c_j) = \begin{cases} 1, & \text{if } i = j \\ 0, & \text{if } i \neq j \end{cases}$ be the Kronecker $\delta$
- Pick an arbitrary edge in the random graph:
  - pick a vertex $i$ → there are $\deg i$ edges incident with it, so $\deg i$ choices for $i$ to be the 1st end vertex of our arbitrary edge
  - and then there are $\deg j$ choices for $j$ to be the other end-vertex.
- If $m = |E(G)|$ edges are placed at random, the expected number of edges between $i$ and $j$ is $\frac{\deg i \cdot \deg j}{2m}$

Checks if vertices are in the same class
Compute the fraction of edges in \( c_i \) in the random network

- Construct a random graph with the same degree distrib.
- Let \( c_i \) be the class of vertex \( i \)
- Let \( n_c \) be the total number of classes
- Let \( \delta(c_i, c_j) = \begin{cases} 1, & \text{if } i = j \\ 0, & \text{if } i \neq j \end{cases} \) be the Kronecker \( \delta \)
- If \( |E(G)| \) edges are placed at random, the expected number of edges between \( i \) and \( j \) is \( \frac{\deg i \cdot \deg j}{2|E(G)|} \)
- Then the number of edges between same class nodes is:
  \[
  r = \frac{1}{2} \sum_{ij \in E(G)} \frac{\deg i \cdot \deg j}{2|E(G)|} \delta(c_i, c_j)
  \]
  Check if vertices are in the same class
  dicitations: as you choose vertex \( j \) above, the edge \( ji \) will be counted after edge \( ij \) was counted.
Consider their difference

\[ r = \frac{1}{2} \sum_{ij} A_{ij} \delta(c_i, c_j) - \frac{1}{2} \sum_{ij \in E(G)} \frac{\deg i \deg j}{2|E(G)|} \delta(c_i, c_j) \]

\[ r = \frac{1}{2} \sum_{ij} [A_{ij} - \frac{\deg i \deg j}{2|E(G)|}] \delta(c_i, c_j) \]

And now normalize by \( m = |E(G)| \):

\[ \text{Modularity} = \frac{1}{2m} \sum_{ij} [A_{ij} - \frac{\deg i \deg j}{2m}] \cdot \delta(c_i, c_j) \]

The same defn as the modularity in community detection, since it measures assortativity based on predefined communities.
Modularity

• $Q = \frac{1}{2|E(G)|} \sum_{ij}[A_{ij} - \frac{\text{deg } i \text{ deg } j}{2|E(G)|}] \cdot \delta(c_i, c_j)$

Checks if vertices are in the same class

• Measure used to quantify the like vertices being connected to like vertices

• $-1 < Q < 0$ means there are fewer edges between like vertices in a class compared to a random network i.e. disassortative network

• $0 < Q < 1$ means there are more edges between like vertices in a class compared to a random network i.e. assortative network

• $Q = 0$ means it behaves like a random network.
Enumerative characteristics

- Normalizing the modularity value $Q$, by the maximum value that it can get is realistic
  - Perfect mixing is when all edges fall between vertices of the same type

$$\sum A_{ij}$$

$$Q_{\text{max}} = \frac{1}{2m} (2m - \sum \frac{k_i k_j}{2m} \delta(c_i, c_j))$$

- Then, the assortativity coefficient, $r = \frac{Q}{Q_{\text{max}}}$, is:

$$-1 \leq \frac{Q}{Q_{\text{max}}} = \frac{\sum (A_{ij} - \frac{k_i k_j}{2m}) \delta(c_i, c_j)}{2m - \sum \frac{k_i k_j}{2m} \delta(c_i, c_j)} \leq 1$$
Based on scalar characteristics, such as:
• Age
• Income
Scalar characteristics

- Scalar characteristics: enumerative characteristics taking numerical values, such as age, income
  - For example using age: two people are similar if:
    - they are born the same day or
    - within a year or within \( x \) years,
    - They are in the same class
    - Same generation
    - different granularity based on the data and questions asked.

- If people are friends with others of the same age, we consider the network assortatively mixed by age (or stratified by age)
The Social Structure of “Countryside” School District

Points Colored by Grade

James Moody
Scalar characteristics

- When we consider scalar characteristics we basically have an approximate notion of similarity between adjacent vertices (i.e. how far/close the values are)
  - There is no approximate similarity that can be measured this way when we talk about enumerative characteristics; rather present/absent
Assortativity matrix based on Scalar characteristics

Friendships at the same US high school: each dot represents a friendship (an edge from the network)

Denser along the $y = x$ line (because of the way data is displayed)

Sparser as the difference in grades increases
Data: 1995 US National Survey of Family Growth

- Top figure: A scatter plot of 1141 married couples
- Bottom figure: The same data showing a histogram of the age difference

Scalar characteristics

• How do we measure scalar assortative mixing?
• Would the idea we use for the enumerative assortative mixing work?
• That is to place vertices in bins based on scalar values:
  – Treat vertices that fall in the same bin (such as age) as “like vertices” or “identical”
  – Apply modularity metric for enumerative characteristics
Scalar characteristics

Then the assortativity coefficient $r = \frac{Q}{Q_{max}}$ is defined again as:

$$r = \frac{\sum_{ij} (A_{ij} - \frac{k_i k_j}{2m})x_i x_j}{\sum_{ij} (k_i \delta_{ij} - \frac{k_i k_j}{2m})x_i x_j}$$

$r = 1 \rightarrow$ Perfectly assortative network

$r = -1 \rightarrow$ Perfectly disassortative network

$r = 0 \rightarrow$ no correlation

Same as Modularity or Pearson correlation coeff.
88 Computer Science faculty:
- vertices are PhD granting institutions in North America
- Edge (i,j) means that PhD student at i, now faculty at j
labels are US census regions + Canada

<table>
<thead>
<tr>
<th></th>
<th>Northeast</th>
<th>Midwest</th>
<th>South</th>
<th>West</th>
<th>Canada</th>
<th>$a_i$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Northeast</td>
<td>0.119</td>
<td>0.053</td>
<td>0.074</td>
<td>0.055</td>
<td>0.022</td>
<td>0.322</td>
</tr>
<tr>
<td>Midwest</td>
<td>0.031</td>
<td><strong>0.067</strong></td>
<td>0.061</td>
<td>0.026</td>
<td>0.011</td>
<td>0.196</td>
</tr>
<tr>
<td>South</td>
<td>0.025</td>
<td>0.027</td>
<td><strong>0.083</strong></td>
<td>0.024</td>
<td>0.006</td>
<td>0.166</td>
</tr>
<tr>
<td>West</td>
<td>0.049</td>
<td>0.033</td>
<td>0.043</td>
<td><strong>0.073</strong></td>
<td>0.011</td>
<td>0.209</td>
</tr>
<tr>
<td>Canada</td>
<td>0.006</td>
<td>0.005</td>
<td>0.005</td>
<td>0.005</td>
<td><strong>0.085</strong></td>
<td>0.107</td>
</tr>
<tr>
<td>$b_i$</td>
<td>0.229</td>
<td>0.185</td>
<td>0.267</td>
<td>0.184</td>
<td>0.135</td>
<td></td>
</tr>
</tbody>
</table>

$r = 0.264$

moderately assortative
By degree:
high degree nodes connect to high degree nodes
A special case is when the characteristic of interest is the degree of the node

- Commonly used in social networks (the most used one of the scalar characteristics)
- More interesting since degree is a topological property of the network (not just a value like age or grade)
- This now reduces to Pearson Correlation Coefficient
Assortative mixing by degree

- Assortative network by degree $\rightarrow$ core of high degrees and a periphery of low degrees (Figure (a) below)
- Disassortative network by degree $\rightarrow$ uniform: low degree adjacent to high degree (Figure (b) and (c) below)

\[ r = \text{assortativity coefficient} \]

<table>
<thead>
<tr>
<th>Network</th>
<th>Type</th>
<th>( n )</th>
<th>( m )</th>
<th>( c )</th>
<th>( S )</th>
<th>( \ell )</th>
<th>( \alpha )</th>
<th>( C )</th>
<th>( C_{WS} )</th>
<th>( r )</th>
<th>Ref(s.)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Film actors</td>
<td>Undirected</td>
<td>449,913</td>
<td>25,516,482</td>
<td>113.43</td>
<td>0.980</td>
<td>3.48</td>
<td>2.3</td>
<td>0.20</td>
<td>0.78</td>
<td>0.208</td>
<td>16,323</td>
</tr>
<tr>
<td>Company directors</td>
<td>Undirected</td>
<td>7,673</td>
<td>55,392</td>
<td>14.44</td>
<td>0.876</td>
<td>4.60</td>
<td>-0.59</td>
<td>0.88</td>
<td>0.276</td>
<td>88,253</td>
<td></td>
</tr>
<tr>
<td>Math coauthorship</td>
<td>Undirected</td>
<td>253,339</td>
<td>496,489</td>
<td>3.92</td>
<td>0.822</td>
<td>7.57</td>
<td>-0.15</td>
<td>0.34</td>
<td>0.120</td>
<td>89,146</td>
<td></td>
</tr>
<tr>
<td>Physics coauthorship</td>
<td>Undirected</td>
<td>52,909</td>
<td>245,300</td>
<td>9.27</td>
<td>0.838</td>
<td>6.19</td>
<td>-0.45</td>
<td>0.56</td>
<td>0.363</td>
<td>234,236</td>
<td></td>
</tr>
<tr>
<td>Biology coauthorship</td>
<td>Undirected</td>
<td>1,520,251</td>
<td>11,803,064</td>
<td>15.53</td>
<td>0.918</td>
<td>4.92</td>
<td>-0.088</td>
<td>0.60</td>
<td>0.127</td>
<td>234,236</td>
<td></td>
</tr>
<tr>
<td>Telephone call graph</td>
<td>Undirected</td>
<td>47,000,000</td>
<td>80,000,000</td>
<td>3.16</td>
<td>2.1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>9,10</td>
<td></td>
</tr>
<tr>
<td>Email messages</td>
<td>Directed</td>
<td>59,812</td>
<td>86,300</td>
<td>1.44</td>
<td>0.952</td>
<td>4.95</td>
<td>1.5/2.0</td>
<td>0.16</td>
<td></td>
<td>103</td>
<td></td>
</tr>
<tr>
<td>Email address books</td>
<td>Directed</td>
<td>16,881</td>
<td>57,029</td>
<td>3.38</td>
<td>0.590</td>
<td>5.22</td>
<td>-0.17</td>
<td>0.13</td>
<td>0.092</td>
<td>248</td>
<td></td>
</tr>
<tr>
<td>Student dating</td>
<td>Undirected</td>
<td>573</td>
<td>477</td>
<td>1.66</td>
<td>0.503</td>
<td>16.01</td>
<td>-0.005</td>
<td>0.001</td>
<td>-0.029</td>
<td>34</td>
<td></td>
</tr>
<tr>
<td>Sexual contacts</td>
<td>Undirected</td>
<td>2,810</td>
<td></td>
<td></td>
<td></td>
<td>3.2</td>
<td></td>
<td></td>
<td></td>
<td>197,198</td>
<td></td>
</tr>
<tr>
<td>WWW nd. edu</td>
<td>Directed</td>
<td>269,504</td>
<td>1,497,135</td>
<td>5.55</td>
<td>1.000</td>
<td>11.27</td>
<td>2.1/2.4</td>
<td>0.11</td>
<td>0.29</td>
<td>-0.067</td>
<td>13,28</td>
</tr>
<tr>
<td>WWW AltaVista</td>
<td>Directed</td>
<td>203,549,046</td>
<td>1,466,000,000</td>
<td>7.20</td>
<td>0.914</td>
<td>16.18</td>
<td>2.1/2.7</td>
<td>0.3/0/</td>
<td></td>
<td>56</td>
<td></td>
</tr>
<tr>
<td>Citation network</td>
<td>Directed</td>
<td>783,339</td>
<td>6,716,198</td>
<td>8.57</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>280</td>
<td></td>
</tr>
<tr>
<td>Roget’s Thesaurus</td>
<td>Directed</td>
<td>1,022</td>
<td>5103</td>
<td>4.99</td>
<td>0.977</td>
<td>4.87</td>
<td>-0.13</td>
<td>0.15</td>
<td>0.157</td>
<td>184</td>
<td></td>
</tr>
<tr>
<td>Word co-occurrence</td>
<td>Undirected</td>
<td>460,902</td>
<td>16,100,000</td>
<td>66.96</td>
<td>1.000</td>
<td>2.7</td>
<td>2.7</td>
<td>0.44</td>
<td></td>
<td>97,116</td>
<td></td>
</tr>
<tr>
<td>Internet</td>
<td>Undirected</td>
<td>10,697</td>
<td>31,992</td>
<td>5.98</td>
<td>1.000</td>
<td>3.31</td>
<td>2.5</td>
<td>0.035</td>
<td>0.39</td>
<td>-0.189</td>
<td>66,111</td>
</tr>
<tr>
<td>Power grid</td>
<td>Undirected</td>
<td>4,941</td>
<td>6,594</td>
<td>2.67</td>
<td>1.000</td>
<td>18.99</td>
<td>-0.10</td>
<td>0.080</td>
<td>-0.003</td>
<td>323</td>
<td></td>
</tr>
<tr>
<td>Train routes</td>
<td>Undirected</td>
<td>587</td>
<td>19,603</td>
<td>66.79</td>
<td>1.000</td>
<td>2.16</td>
<td>-0.69</td>
<td>-0.033</td>
<td>294</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Software packages</td>
<td>Directed</td>
<td>1,439</td>
<td>1,723</td>
<td>1.20</td>
<td>0.998</td>
<td>2.42</td>
<td>1.6/1.4</td>
<td>0.070</td>
<td>0.082</td>
<td>-0.016</td>
<td>239</td>
</tr>
<tr>
<td>Software classes</td>
<td>Directed</td>
<td>1,376</td>
<td>2,213</td>
<td>1.61</td>
<td>1.000</td>
<td>5.40</td>
<td>-0.033</td>
<td>0.012</td>
<td>-0.119</td>
<td>315</td>
<td></td>
</tr>
<tr>
<td>Electronic circuits</td>
<td>Undirected</td>
<td>24,097</td>
<td>53,248</td>
<td>4.34</td>
<td>1.000</td>
<td>11.05</td>
<td>3.0</td>
<td>0.010</td>
<td>0.030</td>
<td>-0.154</td>
<td>115</td>
</tr>
<tr>
<td>Peer-to-peer network</td>
<td>Undirected</td>
<td>880</td>
<td>1,296</td>
<td>1.47</td>
<td>0.805</td>
<td>4.28</td>
<td>2.1</td>
<td>0.012</td>
<td>0.011</td>
<td>-0.366</td>
<td>6,282</td>
</tr>
<tr>
<td>Metabolic network</td>
<td>Undirected</td>
<td>765</td>
<td>3,686</td>
<td>9.64</td>
<td>0.996</td>
<td>2.56</td>
<td>2.2</td>
<td>0.090</td>
<td>0.67</td>
<td>-0.240</td>
<td>166</td>
</tr>
<tr>
<td>Protein interactions</td>
<td>Undirected</td>
<td>2,115</td>
<td>2,240</td>
<td>2.12</td>
<td>0.689</td>
<td>6.80</td>
<td>2.4</td>
<td>0.072</td>
<td>0.071</td>
<td>-0.156</td>
<td>164</td>
</tr>
<tr>
<td>Marine food web</td>
<td>Directed</td>
<td>134</td>
<td>598</td>
<td>4.46</td>
<td>1.000</td>
<td>2.05</td>
<td>-0.16</td>
<td>0.23</td>
<td>-0.263</td>
<td>160</td>
<td></td>
</tr>
<tr>
<td>Freshwater food web</td>
<td>Directed</td>
<td>92</td>
<td>997</td>
<td>10.84</td>
<td>1.000</td>
<td>1.90</td>
<td>-0.20</td>
<td>0.087</td>
<td>-0.326</td>
<td>209</td>
<td></td>
</tr>
<tr>
<td>Neural network</td>
<td>Directed</td>
<td>307</td>
<td>2,359</td>
<td>7.68</td>
<td>0.967</td>
<td>3.97</td>
<td>-0.18</td>
<td>0.28</td>
<td>-0.226</td>
<td>323,328</td>
<td></td>
</tr>
</tbody>
</table>

Examples (published in 2003)

Same formula:

\[
\text{degcorr\_coeff} = \frac{\sum_{ij} (A_{ij} - k_i k_j / 2m) k_i k_j}{\sum_{ij} (k_i \delta_{ij} - k_i k_j / 2m) k_i k_j}
\]

<table>
<thead>
<tr>
<th>network</th>
<th>type</th>
<th>size $n$</th>
<th>degree assortativity $\tau$</th>
<th>error $\sigma_{\tau}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>physics coauthorship</td>
<td>undirected</td>
<td>52,909</td>
<td>0.363</td>
<td>0.002</td>
</tr>
<tr>
<td>biology coauthorship</td>
<td>undirected</td>
<td>1,520,251</td>
<td>0.127</td>
<td>0.0004</td>
</tr>
<tr>
<td>mathematics coauthorship</td>
<td>undirected</td>
<td>253,339</td>
<td>0.120</td>
<td>0.002</td>
</tr>
<tr>
<td>film actor collaborations</td>
<td>undirected</td>
<td>449,913</td>
<td>0.208</td>
<td>0.0002</td>
</tr>
<tr>
<td>company directors</td>
<td>undirected</td>
<td>7,673</td>
<td>0.276</td>
<td>0.004</td>
</tr>
<tr>
<td>student relationships</td>
<td>undirected</td>
<td>573</td>
<td>-0.029</td>
<td>0.037</td>
</tr>
<tr>
<td>email address books</td>
<td>directed</td>
<td>16,881</td>
<td>0.092</td>
<td>0.004</td>
</tr>
<tr>
<td><strong>social</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>power grid</td>
<td>undirected</td>
<td>4,941</td>
<td>-0.003</td>
<td>0.013</td>
</tr>
<tr>
<td>Internet</td>
<td>undirected</td>
<td>10,697</td>
<td>-0.189</td>
<td>0.002</td>
</tr>
<tr>
<td>World-Wide Web</td>
<td>directed</td>
<td>269,504</td>
<td>-0.067</td>
<td>0.0002</td>
</tr>
<tr>
<td>software dependencies</td>
<td>directed</td>
<td>3,162</td>
<td>-0.016</td>
<td>0.020</td>
</tr>
<tr>
<td><strong>technological</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>protein interactions</td>
<td>undirected</td>
<td>2,115</td>
<td>-0.156</td>
<td>0.010</td>
</tr>
<tr>
<td>metabolic network</td>
<td>undirected</td>
<td>765</td>
<td>-0.240</td>
<td>0.007</td>
</tr>
<tr>
<td>neural network</td>
<td>directed</td>
<td>307</td>
<td>-0.226</td>
<td>0.016</td>
</tr>
<tr>
<td>marine food web</td>
<td>directed</td>
<td>134</td>
<td>-0.263</td>
<td>0.037</td>
</tr>
<tr>
<td>freshwater food web</td>
<td>directed</td>
<td>92</td>
<td>-0.326</td>
<td>0.031</td>
</tr>
<tr>
<td><strong>biological</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### Range of the value $r$ for real networks

Some statistics about real networks published in 2011

<table>
<thead>
<tr>
<th>Network</th>
<th>$N$</th>
<th>$z$</th>
<th>$\ell$</th>
<th>$\ell_1$</th>
<th>$\ell_1^F$</th>
<th>$C$</th>
<th>$\tilde{C}$</th>
<th>$r$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Power grid</td>
<td>4941</td>
<td>2.67</td>
<td>18.99</td>
<td>8.61</td>
<td>7.85</td>
<td>0.08</td>
<td>0.10</td>
<td>0.0035</td>
</tr>
<tr>
<td>PGP network</td>
<td>10680</td>
<td>4.55</td>
<td>7.49</td>
<td>5.40</td>
<td>2.66</td>
<td>0.27</td>
<td>0.38</td>
<td>0.23</td>
</tr>
<tr>
<td>AS Internet</td>
<td>28311</td>
<td>4.00</td>
<td>3.88</td>
<td>3.67</td>
<td>2.56</td>
<td>0.21</td>
<td>0.0071</td>
<td>−0.20</td>
</tr>
<tr>
<td>RL Internet</td>
<td>190914</td>
<td>6.34</td>
<td>6.98</td>
<td>5.25</td>
<td>3.17</td>
<td>0.16</td>
<td>0.061</td>
<td>0.025</td>
</tr>
<tr>
<td>Coauthorships</td>
<td>39577</td>
<td>8.88</td>
<td>5.50</td>
<td>4.45</td>
<td>2.93</td>
<td>0.65</td>
<td>0.25</td>
<td>0.19</td>
</tr>
<tr>
<td>Airports 500</td>
<td>500</td>
<td>11.92</td>
<td>2.99</td>
<td>2.76</td>
<td>1.62</td>
<td>0.62</td>
<td>0.35</td>
<td>−0.278</td>
</tr>
<tr>
<td>Interacting proteins</td>
<td>4713</td>
<td>6.30</td>
<td>4.22</td>
<td>4.05</td>
<td>2.96</td>
<td>0.09</td>
<td>0.062</td>
<td>−0.136</td>
</tr>
<tr>
<td>C. Elegans metabolic</td>
<td>453</td>
<td>8.94</td>
<td>2.66</td>
<td>2.55</td>
<td>1.93</td>
<td>0.65</td>
<td>0.12</td>
<td>−0.226</td>
</tr>
<tr>
<td>C. Elegans neural</td>
<td>297</td>
<td>14.46</td>
<td>2.46</td>
<td>2.33</td>
<td>1.84</td>
<td>0.29</td>
<td>0.18</td>
<td>−0.163</td>
</tr>
<tr>
<td>Facebook Caltech</td>
<td>762</td>
<td>43.70</td>
<td>2.34</td>
<td>2.26</td>
<td>1.55</td>
<td>0.41</td>
<td>0.29</td>
<td>−0.066</td>
</tr>
<tr>
<td>Facebook Georgetown</td>
<td>9388</td>
<td>90.67</td>
<td>2.76</td>
<td>2.55</td>
<td>1.79</td>
<td>0.22</td>
<td>0.15</td>
<td>0.075</td>
</tr>
<tr>
<td>Facebook Oklahoma</td>
<td>17420</td>
<td>102.47</td>
<td>2.77</td>
<td>2.66</td>
<td>1.79</td>
<td>0.23</td>
<td>0.16</td>
<td>0.074</td>
</tr>
<tr>
<td>Facebook UNC</td>
<td>18158</td>
<td>84.46</td>
<td>2.80</td>
<td>2.68</td>
<td>1.87</td>
<td>0.20</td>
<td>0.12</td>
<td>$7 \times 10^{-5}$</td>
</tr>
</tbody>
</table>

https://www.semanticscholar.org/paper/The-unreasonable-effectiveness-of-tree-based-theor-Melnik-Hackett/0ef76143b83257592c4155a648286ba7a80cf8474
References

Extra slides

\[ r = \frac{\sum_i e_{ii} - \sum_i a_i b_i}{1 - \sum_i a_i b_i} = \frac{\text{Tr } \mathbf{e} - \| \mathbf{e}^2 \|}{1 - \| \mathbf{e}^2 \|} = .621 \text{ for the network below (strongly assort.)} \]

\( e_{ij} \) is the fraction of edges in a network that connect a vertex of type \( i \) to one of type \( j \):

\[ \sum e_{ij} = 1, \quad \sum e_{ij} = a_i, \quad \sum e_{ij} = b_j, \]

<table>
<thead>
<tr>
<th></th>
<th>women</th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>black</td>
<td>hispanic</td>
<td>white</td>
<td>other</td>
<td></td>
</tr>
<tr>
<td>men</td>
<td>black</td>
<td>0.258</td>
<td>0.016</td>
<td>0.035</td>
<td>0.013</td>
</tr>
<tr>
<td></td>
<td>hispanic</td>
<td>0.012</td>
<td>0.157</td>
<td>0.058</td>
<td>0.019</td>
</tr>
<tr>
<td></td>
<td>white</td>
<td>0.013</td>
<td>0.023</td>
<td>0.306</td>
<td>0.035</td>
</tr>
<tr>
<td></td>
<td>other</td>
<td>0.005</td>
<td>0.007</td>
<td>0.024</td>
<td>0.016</td>
</tr>
<tr>
<td></td>
<td>( b_i )</td>
<td>0.289</td>
<td>0.204</td>
<td>0.423</td>
<td>0.084</td>
</tr>
</tbody>
</table>

TABLE I: The mixing matrix \( e_{ij} \) and the values of \( a_i \) and \( b_i \) for sexual partnerships in the study of Catania et al. [23]. After Morris [24].