ASSORTATIVITY - HOMOPHILY
"birds of a feather flock together"

- Property of (social) networks that nodes of the same attitude tends to be connected with a higher probability than expected
- It appears as correlation between vertex properties of $x(i)$ and $x(j)$ if $(i,j) \in E$

**Vertex properties**

- age
- gender
- nationality
- political beliefs
- socioeconomic status
- habitual place
- obesity
- …
Note on interpreting homophily

Homophily can be a link creation mechanism (nodes have a preference to connect with similar ones, so the network end up to be assortative), or a consequence of influence phenomenons (because nodes are connected, they tend to influence each other and thus become more similar).

Without access to the dynamic of the network and its properties, it is not possible to differentiate those effects.
Homophily - Assortative mixing

"Opposites attract"

Disassortativity - Heterophily

• Opposite of homophily: dissimilar nodes tend to be connected

Examples

• Sexual/Sentimental networks
• Predator - prey ecological networks
To quantify homophily

- We can take into account…
  - **Categorical (Enumerative) attributes**: vertex features which are comparable but not quantifiable (e.g., gender, colour, shape, ethnicity, etc.)
  - **Scalar attributes**: vertex features which are comparable and sortable (age, weight, income, degree, …)
Homophily - Assortative mixing

Categorical attributes

\[ r = \frac{\sum_i e_{ii} - \sum_i a_i^2}{1 - \sum_i a_i^2} \]

- \( e_{ii} \): fraction of edges between nodes with same attributes
- \( a_i \): fraction of all edges having at least an end with property i.
  => Sum of degrees of nodes with property i divided by L

No assortative mixing: \( r = 0 \) (\( e_{ij} = a_i^2 \))

Perfectly assortative: \( r = 1 \)
Assortative: \( r > 0 \)
Homophily - Assortative mixing

Assortativity index - Example

Let's see a fictional example of how to compute the assortativity index. Nodes are individuals, edges represent for instance some social interaction. Columns/Rows correspond to blood types, and numbers are expressed in fraction of the total number of edges.

<table>
<thead>
<tr>
<th>Blood Types</th>
<th>A</th>
<th>AB</th>
<th>B</th>
<th>O</th>
<th>a_i</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>0.30</td>
<td>0.05</td>
<td>0.1</td>
<td>0.05</td>
<td>0.5</td>
</tr>
<tr>
<td>AB</td>
<td>0.05</td>
<td>0.05</td>
<td>0</td>
<td>0</td>
<td>0.1</td>
</tr>
<tr>
<td>B</td>
<td>0.1</td>
<td>0</td>
<td>0.2</td>
<td>0</td>
<td>0.3</td>
</tr>
<tr>
<td>O</td>
<td>0.05</td>
<td>0</td>
<td>0</td>
<td>0.05</td>
<td>0.1</td>
</tr>
</tbody>
</table>

\[
r = \frac{(0.3+0.05+0.2+0.05) - (0.5^2 + 0.1^2 + 0.3^2 + 0.1^2)}{1 - (0.5^2 + 0.1^2 + 0.3^2 + 0.1^2)} = \frac{0.6 + 0.36}{1 - 0.36} =
\]

Note on interpreting homophily.

Indeed, a preference for nodes having the same property, for instance, based on their attributes can be generalized to the concept of the notion of nodes connecting to each other with preferences reciprocallly).

Mixing Patterns

Typical examples would be in particular social networks networks such as Twitter, i.e., a greater number of connections with nodes that are different.

Homophily is considered a common feature of many networks, in political beliefs, etc.

Similarity in this case must be understood in term of nodes properties. Some typical examples can be age, gender, language, properties. Some typical examples can be age, gender, ethnicity.

Assortativity - Homophily - Mixing Patterns

As an example of mixing patterns of age in a network of interaction between nodes of the same category).

When the property for which we study homophily is categorical or numerical, it is not possible to differentiate those effects.

Without access to the dynamic of the network and its properties, thus become more similar).

Indeed, cause nodes are connected, they tend to influence phemenons (be- cause nodes are connected, we are interested in.

The way to compute homophily di- remains the same, the way to compute homophily di- remains the same, the way to compute homophily di-

Assortativity and Modularity

Let's see a fictional example of how to compute the assortativity index. Nodes are individuals, edges represent for instance some social interaction. Columns/Rows correspond to blood types, and numbers are expressed in fraction of the total number of edges.

\[
r = \frac{(0.3+0.05+0.2+0.05) - (0.5^2 + 0.1^2 + 0.3^2 + 0.1^2)}{1 - (0.5^2 + 0.1^2 + 0.3^2 + 0.1^2)} = \frac{0.6 + 0.36}{1 - 0.36} =
\]

Example of mixing patterns of age in a network of interaction between nodes of the same category).

When the property for which we study homophily is categorical or numerical, it is not possible to differentiate those effects.

Without access to the dynamic of the network and its properties, thus become more similar).

Indeed, cause nodes are connected, they tend to influence phemenons (be- cause nodes are connected, we are interested in.

The way to compute homophily di- remains the same, the way to compute homophily di- remains the same, the way to compute homophily di-

Assortativity and Modularity

Let's see a fictional example of how to compute the assortativity index. Nodes are individuals, edges represent for instance some social interaction. Columns/Rows correspond to blood types, and numbers are expressed in fraction of the total number of edges.

\[
r = \frac{(0.3+0.05+0.2+0.05) - (0.5^2 + 0.1^2 + 0.3^2 + 0.1^2)}{1 - (0.5^2 + 0.1^2 + 0.3^2 + 0.1^2)} = \frac{0.6 + 0.36}{1 - 0.36} =
\]
Homophily - Assortative mixing

\[ r = \frac{\sum_i e_{ii} - \sum_i a_i^2}{1 - \sum_i a_i^2} \]

Assortativity and Modularity

Assortativity is related to the Modularity, a measure of the quality of communities, by the following relation:

\[ r = \frac{Q}{Q_{max}} \]

Indeed, \( \sum_i e_{ii} - \sum_i a_i^2 \) corresponds to the definition of the Modularity, while \( 1 - \sum_i a_i^2 \) corresponds to the maximal value that the Modularity could reach if all nodes were in the same communities.
Homophily - Assortative mixing

Numeric attributes

**Pearson correlation coefficient** of properties at both extremities of edges

\[
\text{e}_{xy} = \frac{\sum xy (e_{xy} - a_x b_y)}{\sigma_a \sigma_b},
\]

with \( \sigma_a \) standard deviation of \( a_x \)

\( \sum e_{xy} = 1, \quad \sum_y e_{xy} = a_x, \quad \sum_x e_{xy} = b_y \)
Limit of assortativity coefficient

### Limits of Assortativity

A limit of assortativity coefficients as we have defined them is that they summarize the whole network as a single value. However, different parts of the network might have different types of assortativity.

Illustration of different local assortativity behaviors leading to the same global assortativity value (bottom: distribution of local assortativity). Figure from\(^a\), in which the authors propose a measure of **multiscale assortativity**.

\(^a\)Peel, Delvenne, and Lambiotte 2018.
Mixing patterns

Beyond assortative and disassortative, we can study more generally **Mixing patterns**, => preference of nodes with attribute a to connect with nodes with attribute b (where a,b can be identical or different)

**Mixing Patterns - example**

Example of mixing patterns of age in a network of interaction between individuals, reproduced from\(^9\).

We can see that there is some level of assortativity (high values on the diagonal), but that there are also some more complex mixing patterns, for instance between age 10 and 40, approximately, here interpreted as child-parents relationships.

\(^9\)Del Valle et al. 2007.
Mixing patterns

• [The Anatomy of the Facebook Social Graph, Ugander et al. 2011]
Degree-degree correlation

- Assortativity often used for degree assortativity
- An application of assortativity to the case of degrees used as node properties:
  - Are important nodes connected to other important nodes with a higher probability than expected?
- The degree can be used as any other scalar property

<table>
<thead>
<tr>
<th>network</th>
<th>type</th>
<th>size n</th>
<th>assortativity r</th>
<th>error $\sigma_r$</th>
</tr>
</thead>
<tbody>
<tr>
<td>physics coauthorship</td>
<td>undirected</td>
<td>52909</td>
<td>0.363</td>
<td>0.002</td>
</tr>
<tr>
<td>biology coauthorship</td>
<td>undirected</td>
<td>1520251</td>
<td>0.127</td>
<td>0.0004</td>
</tr>
<tr>
<td>mathematics coauthorship</td>
<td>undirected</td>
<td>253339</td>
<td>0.120</td>
<td>0.002</td>
</tr>
<tr>
<td>film actor collaborations</td>
<td>undirected</td>
<td>449913</td>
<td>0.208</td>
<td>0.0002</td>
</tr>
<tr>
<td>company directors</td>
<td>undirected</td>
<td>7673</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>16881</td>
<td>0.092</td>
<td>0.004</td>
</tr>
<tr>
<td>power grid</td>
<td>undirected</td>
<td>4941</td>
<td>-0.003</td>
<td>0.013</td>
</tr>
<tr>
<td>Internet</td>
<td>undirected</td>
<td>10697</td>
<td>-0.189</td>
<td>0.002</td>
</tr>
<tr>
<td>World-Wide Web</td>
<td>directed</td>
<td>269504</td>
<td>-0.067</td>
<td>0.0002</td>
</tr>
<tr>
<td>software dependencies</td>
<td>directed</td>
<td>3162</td>
<td>-0.016</td>
<td>0.020</td>
</tr>
<tr>
<td>protein interactions</td>
<td>undirected</td>
<td>2115</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>
</tbody>
</table>

Average nearest-neighbour degree

- More detailed characterization of degree-degree correlations

- $k_{\text{ann}}$: average nearest neighbours degree

- $k_{\text{ann}}$ can be written as:

$$k_{\text{ann}}(k) = \sum_{k'} k' P(k'|k) = \frac{\sum_{k'} k' e_{kk'}}{\sum_{k'} e_{kk'}}$$

- where $P(k'|k)$ is the conditional probability that an edge of a node with degree $k$ points to a node with degree $k'$

- If there are no degree correlations:

$$k_{\text{ann}}(k) = \ldots = \frac{\langle k^2 \rangle}{\langle k \rangle}$$

- $k_{\text{ann}}$ is independent of $k$ (nodes of any degrees should have the same nearest neighbours degree)

- If the network is assortative $k_{nn}(k)$ is a positive function

- If the network is disassortative $k_{nn}(k)$ is a negative function

Nearest neighbour degree

Astrophysics co-authorship network

Yeast PPI

Assortative

Disassortative

\[ \langle k' \rangle \text{ exponent: } 0.37 \pm 0.11 \]

\[ \langle k'' \rangle \text{ exponent: } -0.27 \pm 0.03 \]
Nearest neighbour degree

On Facebook

- [The Anatomy of the Facebook Social Graph, Ugander et al. 2011]