Structural Equivalence and Assortative Mixing

Leonid E. Zhukov

School of Data Analysis and Artificial Intelligence
Department of Computer Science
National Research University Higher School of Economics

Structural Analysis and Visualization of Networks
1. Node equivalence
   - Structural equivalence
   - Regular equivalence

2. Node similarity
   - Jaccard similarity
   - Cosine similarity
   - Pearson correlation

3. Assortative mixing
   - Mixing by value
   - Degree correlation
Patterns of relations

- Global, statistical properties of the networks:
  - average node degree (degree distribution)
  - average clustering
  - average path length

- Local, per vertex properties:
  - node centrality
  - page rank

- Pairwise properties:
  - node equivalence
  - node similarity
  - correlation between pairs of vertices (node values)
Structural equivalence

Definition

Structural equivalence: two vertices are structurally equivalent if their respective sets of in-neighbors and out-neighbors are the same.

\[
\begin{array}{cccccc}
\text{u1} & \text{u2} & \text{v1} & \text{v2} & \text{w} \\
\text{u1} & 0 & 0 & 1 & 1 & 0 \\
\text{u2} & 0 & 0 & 1 & 1 & 0 \\
\text{v1} & 0 & 0 & 0 & 1 & 1 \\
\text{v2} & 0 & 0 & 1 & 0 & 1 \\
\text{w} & 0 & 0 & 0 & 0 & 0
\end{array}
\]

rows and columns of adjacency matrix of structurally equivalent nodes are identical, ”connect to the same neighbors”
In order for adjacent vertices to be structurally equivalent, then might have self loops.

Sometimes called ”strong structural equivalence”

Sometimes relax requirements for self loops for adjacent nodes
Similarity measures

- Jaccard similarity

\[ J(v_i, v_j) = \frac{|\mathcal{N}(v_i) \cap \mathcal{N}(v_j)|}{|\mathcal{N}(v_i) \cup \mathcal{N}(v_j)|} \]
Similarity measures

- Undirected graph
- Cosine similarity (vectors in $n$-dim space)

$$\sigma(v_i, v_j) = \cos(\theta_{ij}) = \frac{v_i^T v_j}{|v_i||v_j|} = \frac{\sum_k A_{ik} A_{kj}}{\sqrt{\sum A_{ik}^2} \sqrt{\sum A_{jk}^2}}$$

- Pearson correlation coefficient:

$$r_{ij} = \frac{\sum_k (A_{ik} - \langle A_i \rangle)(A_{jk} - \langle A_j \rangle)}{\sqrt{\sum_k (A_{ik} - \langle A_i \rangle)^2 \sqrt{\sum_k (A_{jk} - \langle A_j \rangle)^2}}}$$

### Graph and adjacency matrix example:

```
A  B  C  D  E
A  0  1  0  1  1
B  1  0  1  0  1
C  0  1  0  1  0
D  1  0  1  0  1
E  1  1  0  1  0
```

Leonid E. Zhukov (HSE)
Similarity measures

- Unweighted undirected graph $A_{ik} = A_{ki}$, binary matrix, only 0 and 1
- $k_i = \sum_k A_{ik} = \sum_k A_{ki}^2$ - node degree
- $n_{ij} = \sum_k A_{ik}A_{kj} = (A^2)_{ij}$ - number of shared neighbors
- $\langle A_i \rangle = \frac{1}{n} \sum_k A_{ik}$

- Cosine similarity (vectors in $n$-dim space)

$$\sigma(v_i, v_j) = \cos(\theta_{ij}) = \frac{n_{ij}}{\sqrt{k_i k_j}}$$

- Pearson correlation coefficient:

$$r_{ij} = \frac{n_{ij} - \frac{k_i k_j}{n}}{\sqrt{k_i - \frac{k_i^2}{n}} \sqrt{k_j - \frac{k_j^2}{n}}}$$
Similarity matrix

Graph Node similarity matrix
Regular equivalence: two vertices are regularly equivalent if they are equally related to equivalent others.

Equivalent definition in terms of role assignment (coloring): vertices that are colored the same, have the same colors of their neighborhoods.

White and Reitz, 1983; Everette and Borgatti, 1991
Equivalence example

- **structural equivalence**

- **regular equivalence**
Recursive definition: two vertices are regularly equivalent if they are equally related to equivalent others. Quantitative measure of regular equivalence - $\sigma_{ij}$, similarity score

$$\sigma_{ij} = \alpha \sum_{k,l} A_{ik} A_{jl} \sigma_{kl}$$

$$\sigma = \alpha A \sigma A$$

should have high $\sigma_{ii}$ - self similarity

$$\sigma_{ij} = \alpha \sum_{k,l} A_{ik} A_{jl} \sigma_{kl} + \delta_{ij}$$

$$\sigma = \alpha A \sigma A + I$$
A vertex $j$ is similar to vertex $i$ (dashed line) if $i$ has a network neighbor $v$ (solid line) that is itself similar to $j$

$$\sigma_{ij} = \alpha \sum_v A_{iv} \sigma_{vj} + \delta_{ij}$$

Closed form solution:

$$\sigma = \alpha A \sigma + I$$

$$\sigma = (I - \alpha A)^{-1}$$

Leicht, Holme, and Newman, 2006
- \( s(a, b) \) - similarity between \( a \) and \( b \)
- \( I() \) - set of in-neighbours

\[
s(a, b) = \frac{C}{|I(a)||I(b)|} \sum_{i=1}^{l(a)} \sum_{j=1}^{l(b)} s(l_i(a), l_j(b)), \quad a \neq b
\]

\[
s(a, a) = 1
\]

- Matrix notation:

\[
S_{ij} = \frac{C}{k_i k_j} \sum_{k,m} A_{ki} A_{mj} S_{km}
\]

Jeh and Widom, 2002
Mixing patterns

Network mixing patterns

- **Assortative mixing**, "like links with like", attributed of connected nodes tend to be more similar than if there were no such edge.

- **Disassortative mixing**, "like links with dislike", attributed of connected nodes tend to be less similar than if there were no such edge.

Vertices can mix on any vertex attributes (age, sex, geography in social networks), unobserved attributes, vertex degrees.

Examples:
- assortative mixing - in social networks political beliefs, obesity, race
- disassortative mixing - dating network, food web (predator/prey), economic networks (producers/consumers)
Assortative mixing

- Political polarization on Twitter: political retweet network, red color - "right-learning" users, blue color - "left learning" users

- Assortative mixing = homophily

Conover et al., 2011
**Assortative mixing**

- The Spread of Obesity in a Large Social Network over 32 Years

Node colors - person’s obesity status: yellow denotes an obese person (body-mass index > 30) and green denotes a nonobese person.

Edge colors - relationship between them: purple denotes a friendship or marital tie and orange denotes a familial tie.

Christakis and Fowler, 2007
Mixing by categorical attributes

- Vertex categorical attribute ($c_i$ - label): color, gender, ethnicity
- How much more often do attributes match across edges than expected at random?
- Modularity :

$$Q = \frac{m_c - \langle m_c \rangle}{m} = \frac{1}{2m} \sum_{ij} \left( A_{ij} - \frac{k_i k_j}{2m} \right) \delta(c_i, c_j)$$

- $m_c$ - number of edges between vertices with same attributes
- $\langle m_c \rangle$ - expected number of edges within the same class in random network
- Assortativity coefficient:

$$\frac{Q}{Q_{max}} = \frac{\sum_{ij} \left( A_{ij} - \frac{k_i k_j}{2m} \right) \delta(c_i, c_j)}{2m - \sum_{ij} \frac{k_i k_j}{2m} \delta(c_i, c_j)}$$
Mixing by scalar values

- Vertex scalar value (attribute) - \( x_i \)
- How much more similar are attributes across edges than expected at random?
- Average and covariance over edges

\[
\langle x \rangle = \frac{\sum_i k_i x_i}{\sum_i k_i} = \frac{1}{2m} \sum_i k_i x_i = \frac{1}{2m} \sum_{ij} A_{ij} x_i
\]

\[\text{var} = \frac{1}{2m} \sum_{ij} A_{ij} (x_i - \langle x \rangle)^2 = \frac{1}{2m} \sum_i k_i (x_i - \langle x \rangle)^2\]

\[\text{cov} = \frac{1}{2m} \sum_{ij} A_{ij} (x_i - \langle x \rangle)(x_j - \langle x \rangle)\]

- Assortativity coefficient

\[
r = \frac{\text{cov}}{\text{var}} = \frac{\sum_{ij} \left( A_{ij} - \frac{k_i k_j}{2m} \right) x_i x_j}{\sum_{ij} \left( k_i \delta_{ij} - \frac{k_i k_j}{2m} \right) x_i x_j}
\]
Mixing by node degree

- **Assortative mixing by node degree**, $x_i \leftarrow k_i$

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

- **Computations:**
  
  \begin{align*}
  S_1 &= \sum_i k_i = 2m \\
  S_2 &= \sum_i k_i^2 \\
  S_3 &= \sum_i k_i^3 \\
  S_e &= \sum_{ij} A_{ij} k_i k_j
  \end{align*}

- **Assortatitivity coefficient**

  $$r = \frac{S_e S_1 - S_2^2}{S_3 S_1 - S_2^2}$$
Mixing by node degree

- **Assortative network**: interconnected high degree nodes - core, low degree nodes - periphery
- **Disassortative network**: high degree nodes connected to low degree nodes, star-like structure

![Assortative network](image1)
![Disassortative network](image2)

S. Borgatti, M. Everett. The class of all regular equivalences: algebraic structure and computations. Social Networks, v 11, p65-68, 1989


M. D. Conover, J. Ratkiewicz, et al, Political Polarization on Twitter. Fifth International AAAI Conference on Weblogs and Social Media, 2011