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

Network Science

Lecture outline

1. Node equivalence
   - Structural equivalence
   - Regular equivalence
2. Node similarity
   - Jaccard similarity
   - Cosine similarity
   - Pearson correlation
3. Assortative mixing
   - Mixing by value
   - Degree correlation

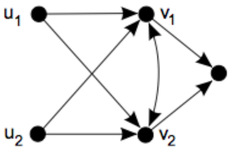
Patterns of relations

- 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

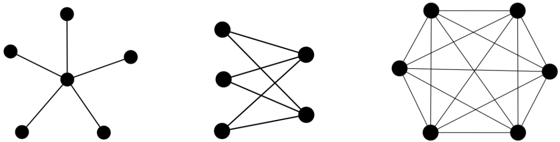


	u1	u2	v1	v2	w
u1	0	0	1	1	0
u2	0	0	1	1	0
v1	0	0	0	1	1
v2	0	0	1	0	1
w	0	0	0	0	0

rows and columns of adjacency matrix of structurally equivalent nodes are identical, "connect to the same neighbors"

Structural equivalence

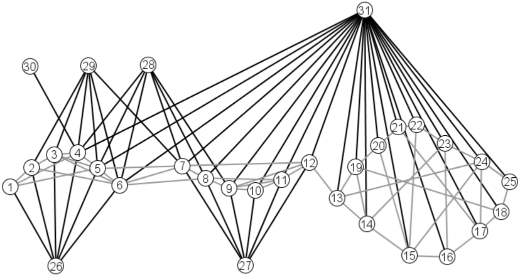
- In order for adjacent vertices to be structurally equivalent, they should have self loops.
- Sometimes called "strong structural equivalence"
- Sometimes relax requirements for self loops for adjacent nodes



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)

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

- Pearson correlation coefficient:

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

0	1	0	1	1
1	0	1	0	1
0	1	0	1	0
1	0	1	0	1
1	1	0	1	0

Similarity measures

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

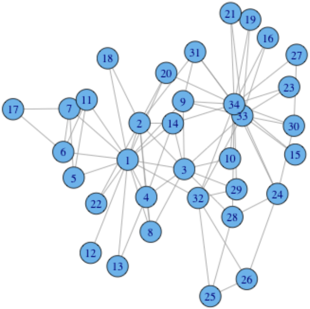
- Cosine similarity (vectors in n-dim space)

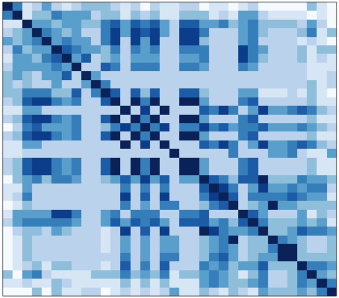
$$\sigma(v_i, v_j) = cos(\theta_{ij}) = \frac{n_{ij}}{\sqrt{k_i k_j}}$$

- 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





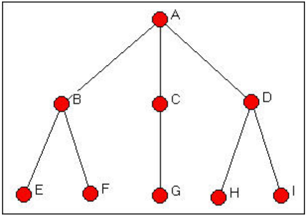
Graph

Node similarity matrix

Regular equivalence

Definition

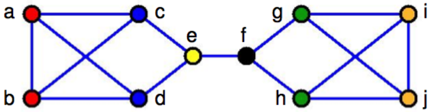
Regular equivalence: two vertices are regularly equivalent if they are equally related to equivalent others (not necessarily the same number of connections)



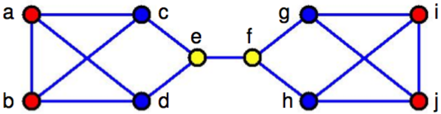
- Equivalent nodes:
  $\{A\}$, $\{E, F, G, H, I\}$,$\{B, C, D\}$

Equivalence example

- structural equivalence



- regular equivalence



Vertex similairty

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

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

$$\boldsymbol{\sigma} = \alpha \mathbf{A} \boldsymbol{\sigma} \mathbf{A}$$

- should have high σ_{ii} - self similarity

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

$$\boldsymbol{\sigma} = \alpha \mathbf{A} \boldsymbol{\sigma} \mathbf{A} + \mathbf{I}$$

Vertex similarity

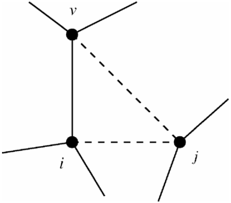
- A vertex j is similar to vertex i (dashed line) if i has a network neighbor v (solid line) that is itself similar to j

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

$$\boldsymbol{\sigma} = \alpha \mathbf{A} \boldsymbol{\sigma} + \mathbf{I}$$

- Closed form solution:

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



Leicht, Holme, and Newman, 2006

SimRank

- G - directed graph
- Two vertices are similar if they are referenced by similar vertices
- $s(a,b)$ - similarity between a and b, $I()$ - set of in-neighbours

$$s(a,b) = \frac{C}{|I(a)||I(b)|} \sum_{i=1}^{I(a)} \sum_{j=1}^{I(b)} s(I_i(a), I_j(b)), \quad a \neq b$$

$$s(a,a) = 1$$

- Matrix notation:

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

- Iterative solution starting from $s_0(i,j) = \delta_{ij}$

Jeh and Widom, 2002

Mixing patterns

Network mixing patterns

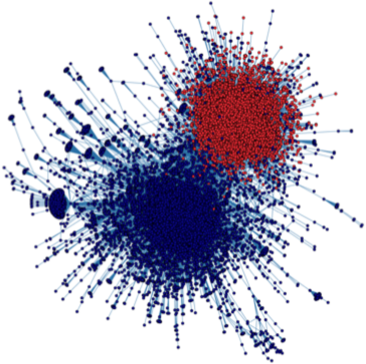
- **Assortative mixing**, "like links with like", attributed of connected nodes tend to be more similar than if there were no such edge
- **Disassortative mixing**, "like links with dislike", attributed of connected nodes tend to be less similar than if there were no such edge

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

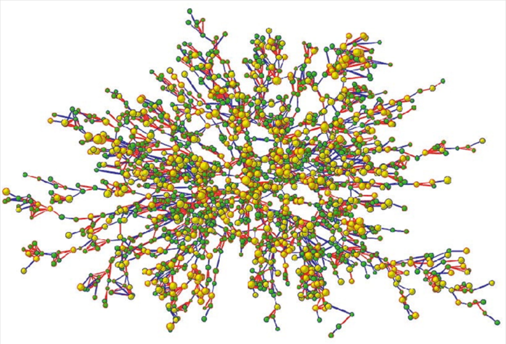


- Assortative mixing = homophily

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

$$
\begin{aligned}
\langle x \rangle &= \frac{\sum_i k_i x_i}{\sum_i k_i} = \frac{1}{2m} \sum_i k_i x_i = \frac{1}{2m} \sum_{ij} A_{ij} x_i \\
var &= \frac{1}{2m} \sum_{ij} A_{ij} (x_i - \langle x \rangle)^2 = \frac{1}{2m} \sum_i k_i (x_i - \langle x \rangle)^2 \\
cov &= \frac{1}{2m} \sum_{ij} A_{ij} (x_i - \langle x \rangle)(x_j - \langle x \rangle)
\end{aligned}
$$

- Assortativity coefficient

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

Mixing by node degree

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

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

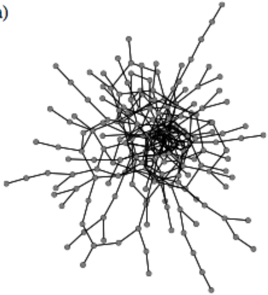
- Computations:
  $S_1 = \sum_i k_i = 2m$
  $S_2 = \sum_i k_i^2$
  $S_3 = \sum_i k_i^3$
  $S_e = \sum_{ij} A_{ij} k_i k_j$
- Assortatitivity coefficient

$$r = \frac{S_e S_1 - S_2^2}{S_3 S_1 - S_2^2}$$

Mixing by node degree

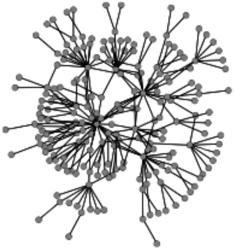
- Assortative network: interconnected high degree nodes - core, low degree nodes - periphery
- Disassortative network: high degree nodes connected to low degree nodes, star-like structure

(a)



Assortative network

(b)



Disassortative network

References

- White, D., Reitz, K.P. Measuring role distance: structural, regular and relational equivalence. Technical report, University of California, Irvine, 1985
- S. Borgatti, M. Everett. The class of all regular equivalences: algebraic structure and computations. Social Networks, v 11, p65-68, 1989
- E. A. Leicht, P.Holme, and M. E. J. Newman. Vertex similarity in networks. Phys. Rev. E 73, 026120, 2006
- G. Jeh and J. Widom. SimRank: A Measure of Structural-Context Similarity. Proceedings of the eighth ACM SIGKDD , p 538-543. ACM Press, 2002
- M. E. J. Newman. Assortative mixing in networks. Phys. Rev. Lett. 89, 208701, 2002.
- M. Newman. Mixing patterns in networks. Phys. Rev. E, Vol. 67, p 026126, 2003
- M. McPherson, L. Smith-Lovin, and J. Cook. Birds of a Feather: Homophily in Social Networks, Annu. Rev. Sociol, 27:415-44, 2001.