Structural Equivalence and Assortative Mixing

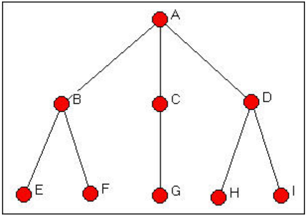
Leonid E. Zhukov

School of Applied Mathematics and Information Science
National Research University Higher School of Economics

17.02.2014

Patterns of relations

- Equivalence of positions and roles



Equivalnce classes:

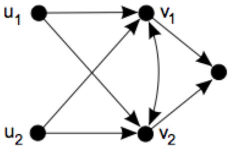
1. structural equivalence (same relationships to other nodes):
   $\{A\}$, $\{B\}$, $\{C\}$, $\{D\}$, $\{E, F\}$, $\{G\}$, $\{H, I\}$
2. automorphic equivalence (parallel structures):
   $\{A\}$, $\{C\}$, $\{G\}$, $\{B, D\}$, $\{E, F, H, I\}$
3. regular equivalence (identical patterns of ties with other classes):
   $\{A\}$, $\{E, F, G, H, I\}$,$\{B, C, D\}$

- Approximate equivalence, similarity between nodes

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 structually equivalent nodes are identical, "connect to the same neighbors"

Approximate equvivalence

- Unweighted graph - binary matrix, only 0/1
- Eucledean distance between vectors

$$d(v_i, v_j) = \sqrt{\sum_k ((A_{ik} - A_{jk})^2 + (A_{ki} - A_{kj})^2)}$$

- Hamming distance - number of positions where vectors are different (Manhattan distance for binary matrix)

$$h(v_i, v_j) = \sum_k |A_{ik} - A_{jk}|$$

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

Matrix properties

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

Hamming distance

- Maximal possible:

$$max(d_{ij}^2) = k_i + k_j$$

- Normalized Hamming distance:

$$d_{ij\,N}^2 = \frac{d_{ij}^2}{k_i + k_j} = \frac{\sum_k (A_{ik}^2 - 2A_{ik}A_{jk} + A_{jk}^2)}{k_i + k_j} = 1 - \frac{2n_{ij}}{k_i + k_j}$$

Similarity measures

- Cosine similarity (vectors in n-dim space)

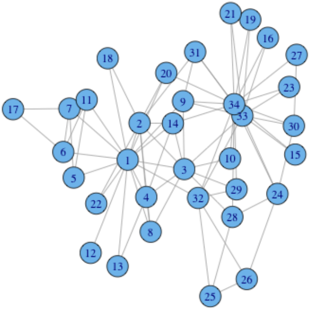
$$\sigma(v_i, v_j) = cos(\theta_{ij}) = \frac{v_i v_j}{||v_i|| \, ||v_j||} = \frac{\sum_k A_{ik} A_{kj}}{\sqrt{\sum A_{ik} A_{ki}} \sqrt{\sum A_{jk} A_{kj}}} = \frac{n_{ij}}{\sqrt{k_i k_j}}$$

- Jaccard similarity

$$J(v_i, v_j) = \frac{|\mathcal{N}(v_i) \cap \mathcal{N}(v_j)|}{|\mathcal{N}(v_i) \cup \mathcal{N}(v_j)|}$$

- Pearson correlation coefficient:

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

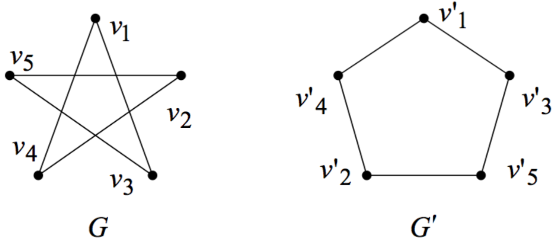


Graph

Similarity matrix

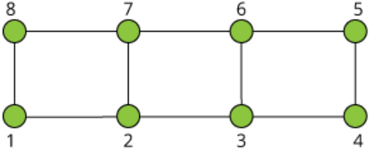
Graph isomorphism

Two graphs are isomorphic if there exists a one-to-one mapping of nodes, such that for every edge in one graph, there is a uniqe edge in another graph between the corresponding mapped vertices ("the same structure")



Graph automorphism

Automorhism is a one-to-one mapping of nodes, such that for every edge in the graph, there is unique edge between the corresponding mapped vertices. This is a form of graph symmetry, isomorphism to istelf.

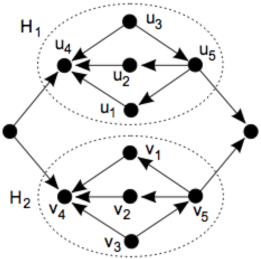


Automorphic equivalence

Definition

Two vertices are automorphically equivalent if there exist an automorphic mapping interchanging these nodes.

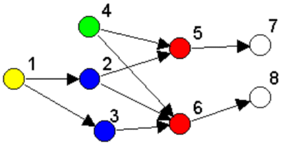
All vertcies relabeled forming isomorphic graph with two interchanged. All distance between nodes are preserved



Regular equivalence

Definition

Regular equivalence: Two vertices are regularly equivalent if they are equally related to equivalent others.



- when coloring, connected to the nodes of the same color
- σ_{ij} - similarity score

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

"connect to the same colors"

Regualar Equivalence

- σ_{ij} - similarity score

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

- should have high σ_{ii} - self similarity

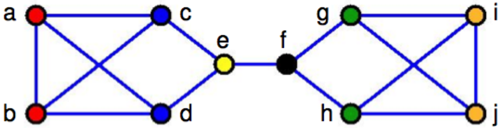
$$\sigma_{ij} = \alpha \sum_{k,l} A_{ik} A_{jl} \sigma_{kl} + \delta_{ij}$$

- variation: vertices i and j are similar if i has a neighbor k similar to j

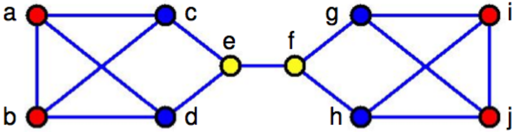
$$\sigma_{ij} = \alpha \sum_{k} A_{ik} \sigma_{kj} + \delta_{ij}$$

Equivalence

- structural equivalence



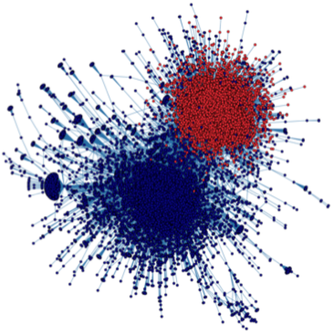
- regular equivalence



structural equivalence > automorphic equivalence > regular equivalence

Assortative Mixing

- Assortative mixing (homophily) - tendency to associate and form connections with those percieved to be similar.



Conover et al., 2011

Mixing by node value

- Let every node has a scalar value x_i associated with it
- Average and covariance over edges

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

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

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

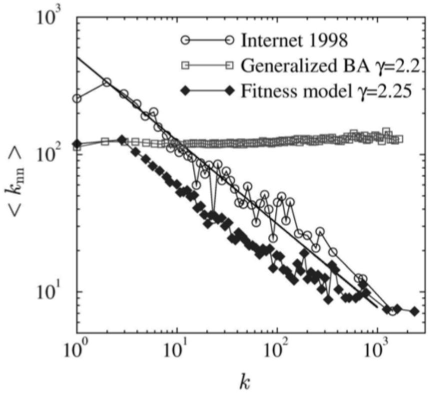
- Assortativity coefficient

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

Node degree correlation

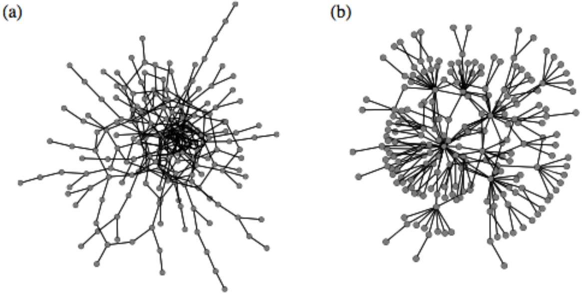
Nearest neighbours average connectivity of nodes with degree k:

$$\langle k_{nn} \rangle = \sum_{k'} k' P(k'|k)$$



PastorSatorras et al. , 2001

Degree correlation



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

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

Referenes

- M. E. J. Newman. Assortative mixing in networks. American Physical Society, 89(20):5, 2002.
- M. E. J. Newman. Mixing patterns in networks, 2003.
- R. Pastor-Satorras, A. Vazquez and A. Vespignani. Dynamical and Correlation Properties of the Internet. Physical Review Letters, V 87, N 25, 258701-2, 2001
- M. Conover, J. Ratkiewicz, et al., Fifth ICWSM 2011