Structural Equivalence and Assortative Mixing

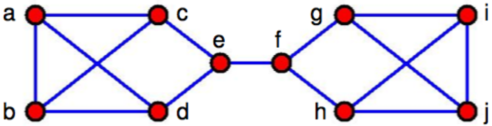
Leonid E. Zhukov

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

28.02.2013

НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ
УНИВЕРСИТЕТ

Node Equivalence

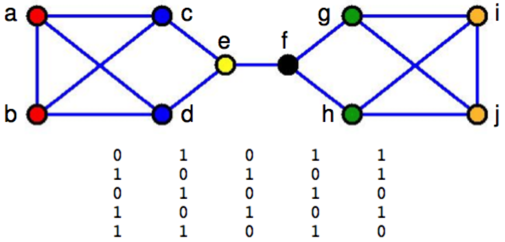


- Structural equivalence
- Regular equivalence

Structural Equivalence

Definition

Structural equivalence: two vertices are structurally equivalent if they share the same neighbours



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

- Hamming distance: number of positions where the vectors are different. (Eucledean distance between vectors in n-dim space)

$$d_{ij}^2 = \sum_k (A_{ik} - A_{jk})^2$$

- Maximal possible:

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

- Normalized Hamming distance:

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

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

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

$$\sigma_{ij} = cos(\theta_{ij}) = \frac{xy}{|x||y|} = \frac{\sum_k A_{ik}A_{kj}}{\sqrt{\sum A_{ik}A_{ki}}\sqrt{\sum A_{jk}A_{kj}}} = \frac{n_{ij}}{\sqrt{k_i k_j}}$$

- Pearson correlation coefficient:

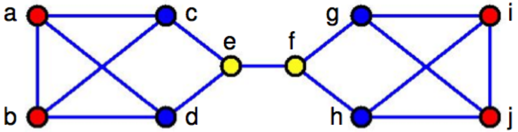
$$r_{ij} = \frac{\sum_k (A_{ik} - \langle A_i \rangle)(A_{jk} - \langle A_j \rangle)}{\sqrt{\sum_k (A_{ik} - \langle A_i \rangle)^2}\sqrt{\sum_k (A_{jk} - \langle A_j \rangle)^2}} = \frac{n_{ij} - \frac{k_i k_j}{n}}{\sqrt{k_i - \frac{k_i^2}{n}}\sqrt{k_j - \frac{k_j^2}{n}}}$$

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

Regualar Equivalence

Definition

Regular equivalence: two vertices are regular equivalent if they have neighbours that are themselves similar



- structural equivalence $\rightarrow$ regular equivalence

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

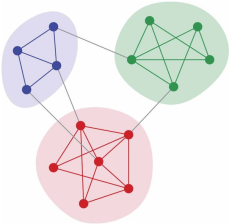
$$\sigma_{ij} = \alpha \sum_{k,l} A_{ik} A_{jl} \sigma_{kl} + \delta_{ij}$$

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

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

Assortative Mixing

- Assortative mixing (homophily) - tendency to associate and form connections with those percieved to be similar.
- Let n_c - number of classes, c_i - class label per node, $\delta(c_i, c_j)$ kronecker delta, takes values $\{0, 1\}$



- Number of edges between nodes of the same class in

$$m_c = \frac{1}{2}\sum_{ij} A_{ij}\delta(c_i, c_j)$$

Modularity

- Random network
  - consider an edge e attached to node i, degree k_i
  - probability that it is attached to node j, degree k_j is $k_j/2m$
  - expected number of edges (average) between i and j is $k_i k_j/2m$
- expected number of edges within the same class

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

- Modularity:

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

- Modularity matrix:

$$B_{ij} = A_{ij} - \frac{k_i k_j}{2m}$$

Mixing by node value

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

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

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

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

- Assortativity coefficient

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

Mixing by node degree

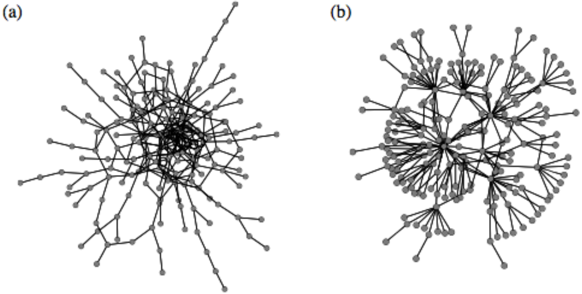
- Assortative mixing by node degree, $x_i \leftarrow k_i$

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

- 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

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

Mixing by node degree



Referenes

- M. E. J. Newman. Assortative mixing in networks. American Physical Society, 89(20):5, 2002.
- M. E. J. Newman, M. Girvan. Mixing patterns and community structure innetworks, 2002.