SC435
Introduction to Complex Networks

Similarity

Similarity in networks

- Nodes that are connected to each other in *(social)* networks tend to similar in their features.
  - friends may have similar features.
  - similar webpages may link to similar others
  - recommendation systems
  - circle of friends tell us about the person
- Assortativity: technical name of measuring similarity. We can make predictions about a person's qualities by inspecting their neighbors.
  - homophily: like attracts likes **or** due to social influence??
  - segregation and polarization of online communities on social media (Echo Chamber)
  - Degree assortativity: core-periphery structure.
- Disassortative: converse of assortative

questions

- In what way can the vertices in a network be similar?
- How can we quantify similarity?

Similarity

Constructing measures of similarity

- Structural equivalence: Two vertices of a network are structurally equivalent if they share many of the same neighbors $\Rightarrow$ in social networks two nodes are similar if they share many of the same neighbors
- Regular equivalence: Two vertices are regularly equivalent if they are equally related to equivalent others $\Rightarrow$ People with similar roles have same local neighborhood.

Similarity

Measures of Structural equivalence (extent)

- **Count of the number of common neighbors**

$$n_{ij} = \sum_k A_{ik} A_{kj} = [A^2]_{ij}$$

- **Cosine Similarity:**

$$\sigma_{ij} = \frac{\sum_k A_{ik} A_{kj}}{\sqrt{\sum_k A_{ik}^2}\sqrt{\sum_k A_{jk}^2}}$$

$$\sigma_{ij} = \frac{n_{ij}}{\sqrt{k_i k_j}} \qquad \textbf{(simple, unweighted)}$$

- **Jaccard Similarity:**

$$J_{ij} = \frac{n_{ij}}{k_i + k_j - n_{ij}}$$

Similarity

Measures of Structural equivalence (extent)

- **Pearson correlation coefficient:**

$$r_{ij} = \frac{\sum_k \left(A_{ik} - \langle A_i \rangle\right) \sum_k \left(A_{jk} - \langle A_j \rangle\right)}{\sqrt{\sum_k \left(A_{ik} - \langle A_i \rangle\right)^2} \sqrt{\sum_k \left(A_{jk} - \langle A_j \rangle\right)^2}} = \frac{\mathrm{Cov}(\sigma_\mathrm{i}, \sigma_\mathrm{j})}{\sigma_i \sigma_j}$$

- **For unweighted, undirected graph**

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

$$-1 \leq r_{ij} \leq 1$$

Similarity

Measures of Structural equivalence (extent)

- **normalize common neighbors by the expected number of common neighbors when they are picked at random**

$$g_{ij} = \frac{n_{ij}}{\frac{k_i k_j}{n}} = \frac{n \sum_k A_{ik} A_{jk}}{\sum_k A_{ik} \sum_k A_{jk}}$$

- **Euclidean (Hamming) distance:**

$$d_{ij} = 1 - \frac{2n_{ij}}{k_i + k_j}$$

Similarity

Measures of Regular equivalence (extent)

- Two nodes i and j have high similarity score if they have neighbors k and l that themselves have high similarity. For an undirected network

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

- Should have high self-similarity (σ_{ii})

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

- problem: assuming a null matrix as the initial condition after k iterations we have

$$\sigma^{(k)} = \sum_{r=0}^{k-1} \alpha^r A^{2r}$$

  This measure of similarity is a weighted sum over the number of paths of even length between pairs of vertices.

Similarity

Measures of Regular equivalence (extent)

- **Modified definition of regular equivalence:** Nodes i and j are similar if i has a neighbor of k that is itself similar to j.

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

- Iterating again starting with $\sigma = 0$, we get:

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

  Similarity of two nodes is the weighted sum of the number of paths of different length that connect them.

- Bias in favor of high degree node can be removed by dividing with node degree

$$\sigma_{ij} = \frac{1}{k_i}\left[\alpha \sum_k A_{ik}\sigma_{kj} + \delta_{ij}\right]$$

Similarity

Homophily and Assortative Mixing

- People have a strong tendency to associate with others whom they perceive as being similar to themselves. This property is called *homophily* or *assortative* mixing.
- Disassortative mixing: tendency to people to associate with others who are unlike them.
- Political polarization, mixing on the basis of race, obesity etc. (Assortative)
- Dating networks, food web (predator-prey), economic networks (producers/consumers) (Disassortative)

Similarity

Mixing by Categorical attributes

- Characterize by some numbers the value of the mixing in the network. One of the ways to do this is called (modularity)
- Every vertex has a label (c_i)
- How much more often do attributes match across edges than what is expected at random? (Pearson correlation coeff.)
- Modularity

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

- The total number of edges that run between nodes of the same class

$$\sum_{ij} A_{ij} \delta(c_i, c_j)$$

- expected number of edges between all pairs of vertices of the same type if edges are placed at random

$$\frac{k_i k_j}{2m}$$

Similarity

Mixing by Categorical attributes

- Modularity

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

$$\begin{aligned} Q &= 0 \quad \textbf{(single class or completely random)} \\ Q &> 0 \quad \textbf{(Assortative mixing)} \\ Q &< 0 \quad \textbf{(Disassortative)} \end{aligned}$$

Similarity

Mixing by Categorical attributes

- Modularity matrix B

$$[B]_{ij} = A_{ij} - \frac{k_i k_j}{2m}$$

- Q is never equal to 1

$$Q_{max} = \frac{1}{2m}\left(2m - \sum_{ij}\frac{k_i k_j}{2m}\right)$$

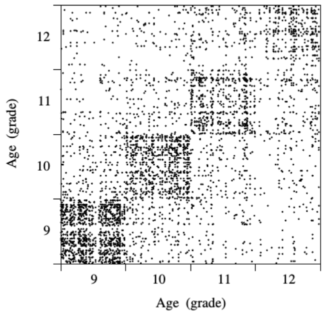
- Value of Q can be significantly smaller than 1 even for perfectly mixed networks. So we normalize it by the maximum value (Assortativity coefficient Q/Q_{max})

Similarity

Mixing by ordered characteristics

- We can also have mixing if characteristics are approximately the same.

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



Similarity

Assortative mixing by degree

- Assortative: core-periphery
- Disassortative: star-like structure

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