Centrality and Community

CS 579 Online Social Network Analysis

Dr. Cindy Hood
9/25/25

Homework Assignments

- HW #3 – Network Metrics
  - Assigned yesterday – due by midnight 10/3, no submissions accepted after 5pm 10/6
  - Good prep for Exam 1
- HW #4 – Chicago Community Areas + Census Data
  - You may work in groups up to 4 students (no exceptions) on this hw
  - Assigned later this week
- Please contact TAs with questions

Exams and Final Project Poster Presentation

- Exam 1 – Oct 9 in class
- Exam 2 – Dec 2 in class

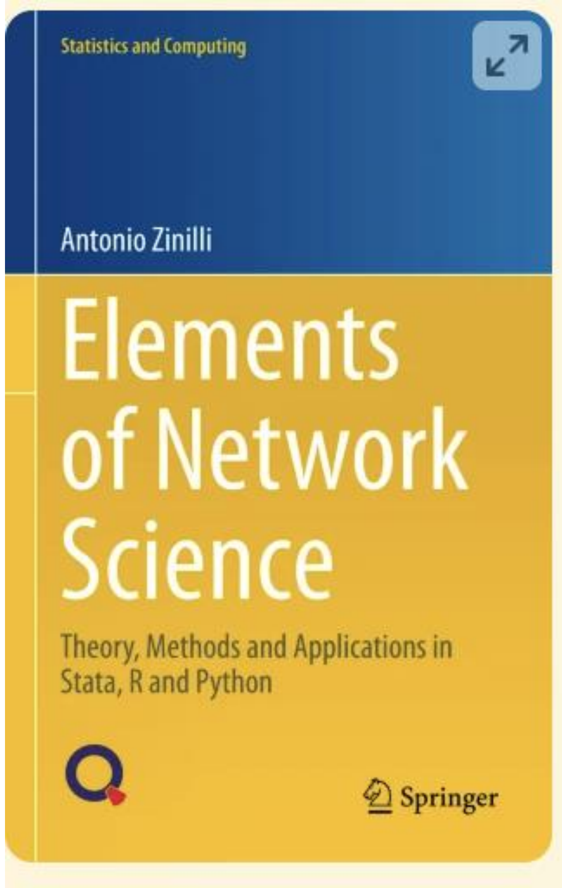
- Final Project Poster Session – Dec 4 in class

- Online students (sections 2 and 3) will have remote options

Teaching Assistants

- Siva Krishna Golla
  - sgolla2@hawk.illinoistech.edu
  - Mondays 2-3pm on zoom
- Khush Dhiren Patel
  - kpatel210@hawk.illinoistech.edu
  - Wednesdays 11-12 online
- Aswith Sama
  - asama@hawk.illinoistech.edu
  - Thursdays 3-4pm on zoom

Reference



https://link.springer.com/book/10.1007/978-3-031-84712-7

Recall Node's Neighborhood

- A nodes neighborhood includes the group of nodes that it is connected to

$$N_i(G) = \{ j : g_{ij} = 1 \}$$

- Studying the neighborhood aids in identifying nodes or communities of nodes that share specific characteristics
- When individuals are motivated to emulate the behavior of their neighbors, cascading effects can occur
  - Can be observed when a new behavior initiates with a small group of early adopters and then spreads radially outward through the network

Neighborhood Degree Sequence

The neighborhood degree sequence for the node i, s_i, is derived from:

$$s_i = \left\{k_1^i, k_2^i k_3^i \ldots, k_n^i\right\} \tag{3.19}$$

with k_i equal to the degree of the nodes to which i is connected. Being a sequence of degrees:

$$k_1^i \leq k_2^i \leq k_3^i \ldots \leq k_n^i \tag{3.20}$$

Neighborhood Degree Sequence

- Analysis
  - May help in examining variance of neighborhood degree sequences
    - Context of hierarchical complexity
  - May analyze neighbors of same degree to determine how similar they are

Jaccard Coefficient (JC) aka Jaccard Index

- Used to calculate how similar the neighbors of two vertices are

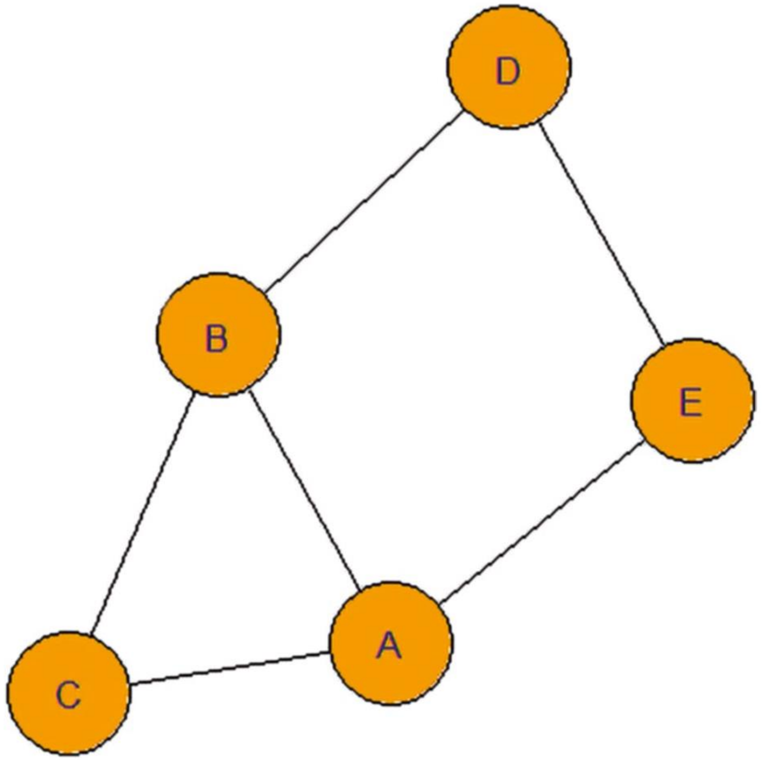
$$JC_{ij} = \frac{|N_i \cap N_j|}{|N_i \cup N_j|}$$

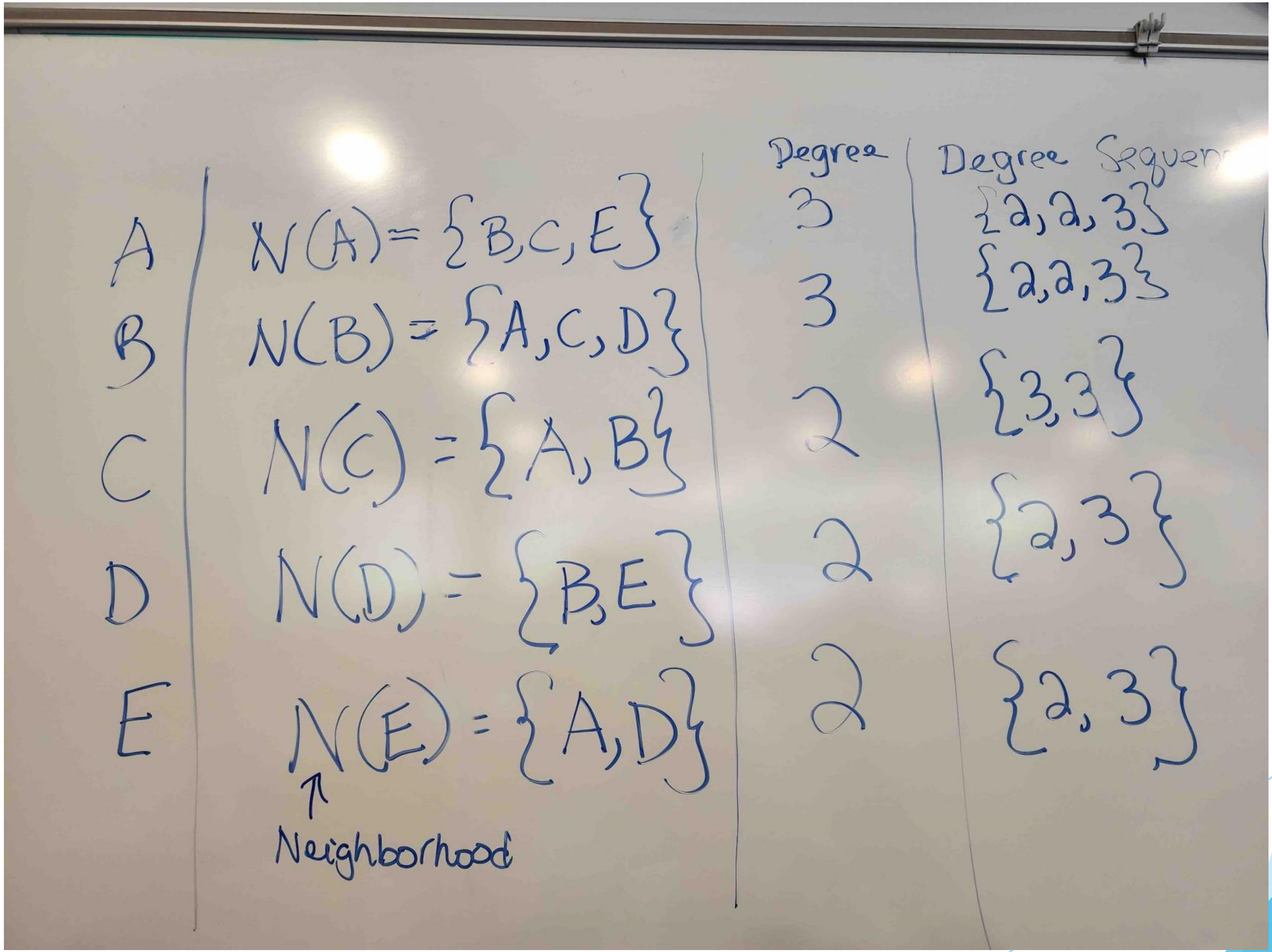
N_i = Neighborhood of i

- 0 <= JC <= 1
  - 0 is maximum dissimilarity
  - 1 is maximum similarity

Example

Fig. 3.4 Graph example




Degree
Degree Sequen
A N(A) = {B, C, E} 3 {2, 2, 3}
B N(B) = {A, C, D} 3 {2, 2, 3}
C N(C) = {A, B} 2 {3, 3}
D N(D) = {B, E} 2 {2, 3}
E N(E) = {A, D} 2 {2, 3}
Neighborhood

equence

$$JC_{AB} = \frac{|N(A) \cap N(B)|}{|N(A) \cup N(B)|} = \frac{C}{ABCDE} = \frac{1}{5}$$

$$JC_{AC} = \frac{B}{ABCE} = \frac{1}{4}$$

$$JC_{AD} = \frac{B, E}{B, C, E} = \frac{2}{3}$$

$$JC_{AE} = 0$$

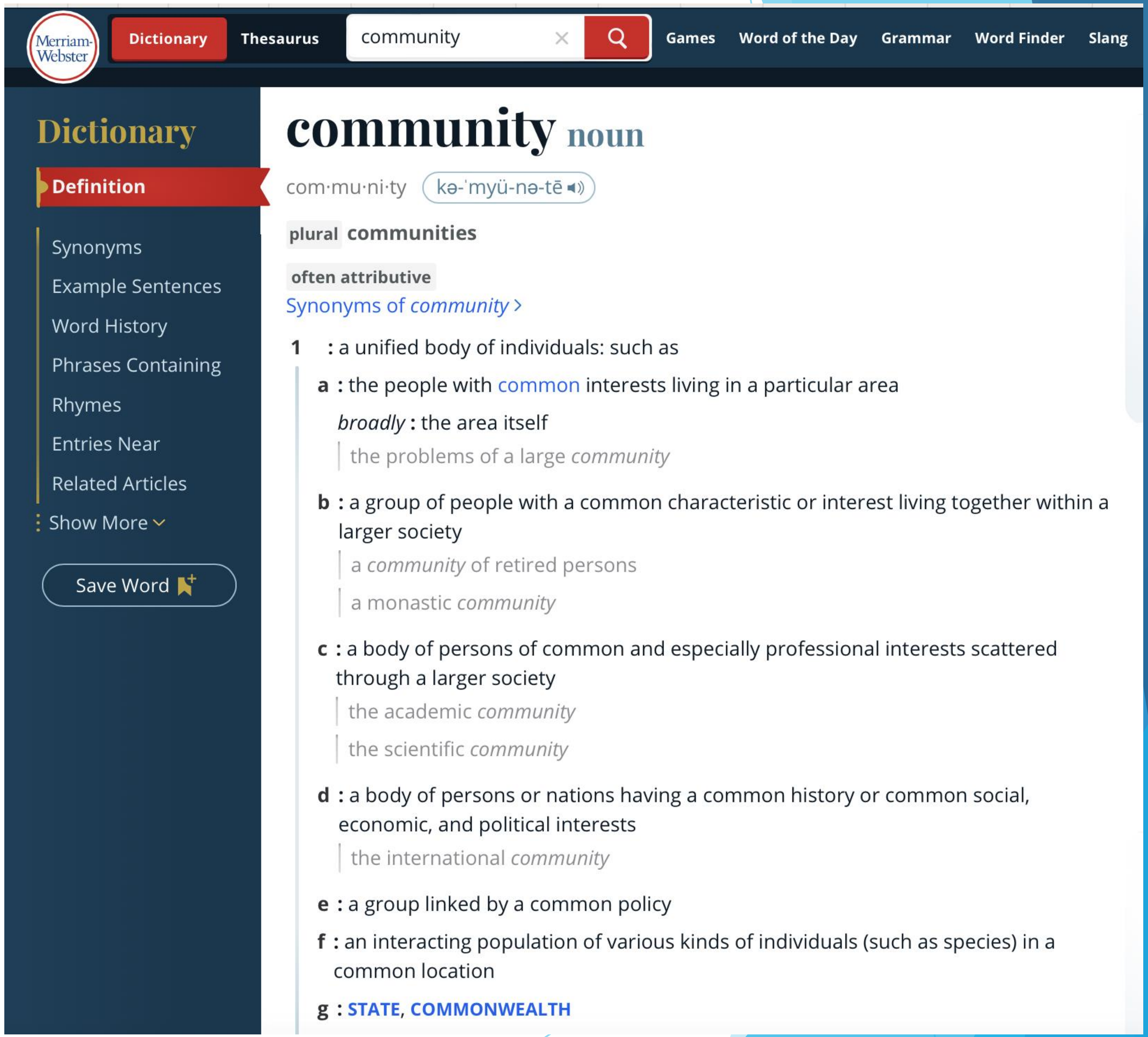
$$JC_{BC}:$$

Jaccard Distance (JD)

- Measure of dissimilarity
- Complement of Jaccard Coeffiecient
- JD = 1 - JC

Community Detection

- Involves finding groups of nodes that are, in some way, more similar to each other than to other nodes



https://www.merriam-webster.com/dictionary/community

Clustering Coefficient aka Transitivity

- Feature of real world networks
- Reflects the degree to which a node’s neighbors are related to one another
- Frequently used to understand how nodes in a network tend to cluster together
- Also used to understand whether a graph has a small world, random or scale-free property

Local Clustering Coefficient

- Assume there is a node *i* of degree *k*
  - Let b_i be the number of edges that exist between the *k* neighbors of *I*
  - The local clustering coefficient $C_{local}(i)$ is the proportion between the actual number of edges amongst the neighbors, b_i and the maximum possible value of edges among the the neighbors.
    - Likelihood that two neighbors of node *i* are also connected
  - The maximum possible value can be calculated using classical binomial coefficient $\binom{k}{2}$
    - $\binom{k}{2} = \frac{k!}{2!(k-2)!} = \frac{k!}{2(k-2)!}$
  - $C_{local}(i) = \frac{b_i}{\binom{k}{2}}$ with $0 <= C_{local}(i) <= 1$

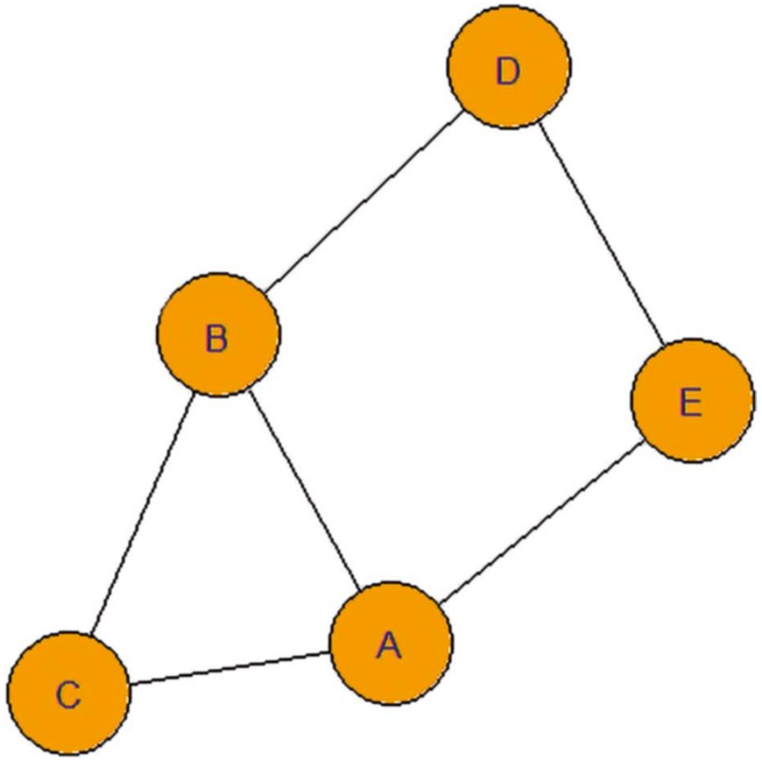
Global Clustering Coefficient

- Average value of the local clustering coefficient $C_{local}(i)$ over all nodes in the network

- $C_{\text{global}} = \frac{1}{n} \sum_{i=1}^{n} C_{local}(i)$

Example

Fig. 3.4 Graph example



$$C_{local}(i) = \frac{b_i}{\binom{k_i}{2}} = \frac{b_i}{\frac{k_i!}{2!(k_i-2)!}}$$

$$b_A = 1 \quad k_A = 3 \quad C_{local}(A) = \frac{1}{\frac{3!}{2!(3-2)!}} = \frac{1}{\frac{6}{2}} = 1/3$$

$$b_B = 1 \quad k_B = 3 \quad C_{local}(B) = 1/3$$

$$C_{local}(C) = 1$$

$$C_{local}(D) = 0$$

$$C_{local}(E) = 0$$

$= 1/3$

$n=5$

$$C_{Global} = \frac{1}{5}\left(1/3 + 1/3 + 1 + 0 + 0\right)$$

$$= 1/\cancel{5}\left(\cancel{5}/3\right) = 1/3$$