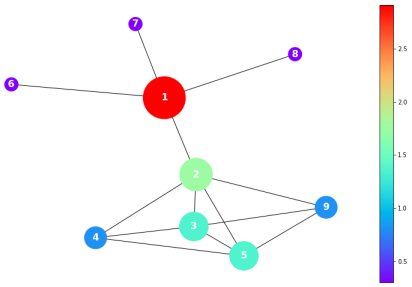
Network Analysis - Centrality Measures

What does ‘Centrality’ mean?

Centrality

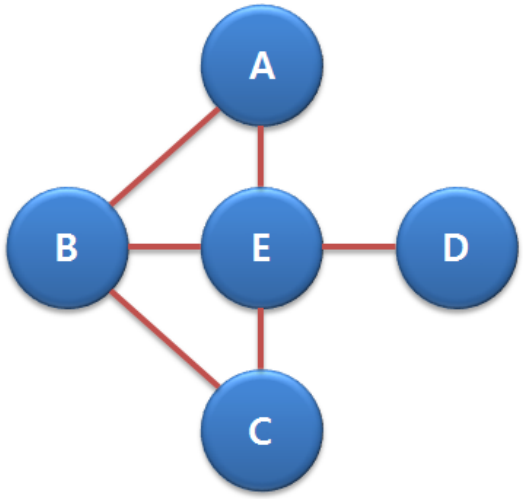
> Answer this network problem: What node is important in a network?



Node Centrality

- Centrality of node - gives us an indication of position of a node within the overall network
- common measure in **social network analysis**
    - “who is the most **important** person in this network?”
- the word **important** has multiple meanings → there are many different definitions of centrality

Centrality Measures



Adjacent Matrix

$$A = \begin{pmatrix} 0 & 1 & 0 & 0 & 1 \\ 1 & 0 & 1 & 0 & 1 \\ 0 & 1 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 1 \\ 1 & 1 & 1 & 1 & 0 \end{pmatrix}$$

Neighbor based:

Degree Centrality

> the basic one - how many direct connections(neighbors) each node has in the graph

Mechanism:

summing all the links (counts an entity's number of links)

$$C_d = A \cdot 1 = \begin{pmatrix} 2 \\ 3 \\ 2 \\ 1 \\ 4 \end{pmatrix}$$

Eigenvector Centrality (EigenCentrality)

> how important the nodes in contact with a node are - which node is better connected to important nodes in the network

Mechanism:

measures the quality of the connections (not quantity) = **few but good**

calculate eigenvector corresponding to eigenvalue (the greatest one)

$$AC_e = \lambda C_e$$

- A: adjacent matrix (N x N matrix)
- C_e: eigenvector (1 x N matrix)
- λ : eigenvalue

$$C_e = \begin{pmatrix} 0.882 \\ 1.121 \\ 0.882 \\ 0.464 \\ 1.247 \end{pmatrix}$$

Path based:

Betweenness Centrality

> detects nodes that have highest capacity of linking other nodes in the chains of contacts in a network = can have control over the flow

Mechanism:

summing propotion of the geodesics(shortest paths) including the specific node out of all the geodesics between two nodes

$$C_b(i) = \sum_{m}^{n} b_{ijm},\ (b_{ijm} = \frac{g_{ijm}}{g_{jm}})$$

- g_{ijm}: the # of geodesics between j & m containing i
  - $i \neq m, i \neq j$
- g_{jm}: the # of geodesics connecting j to m

$$C_b = \begin{pmatrix} 0 \\ 0.5 \\ 0 \\ 0 \\ 3.5 \end{pmatrix}$$

e.g. $C_b(A) = 0$ ← B-C, B-D, C-D, C-E, D-E (none of the shortest paths includes A)

Closeness Centrality

> measures the ‘average’ distance to the other nodes in the network
> = find nodes in a network that can quickly interact with other nodes

Mechanism:

average the lengths of the shortest paths from one node to all the other nodes, and take the reciprocal (shorter the length, the higher the value)

$$C_c(i) = \frac{1}{\frac{1}{N}\sum_i^n d_{ij}} = \frac{N}{\sum_i^n d_{ij}}$$

- d_{ij}: shortest path length between i & j
- N: # of paths

$$C_c = \begin{pmatrix} 0.667 \\ 0.800 \\ 0.667 \\ 0.571 \\ 1.000 \end{pmatrix}$$

e.g. $C_c(A) = 0.667$ ← A-B = 1, A-C = 2, A-D = 2, A-E = 1 → average 1.5

References

- https://symbio6.nl/en/blog/analysis/network-centrality
- https://bookdown.org/markhoff/social_network_analysis/centrality.html
- https://cambridge-intelligence.com/keylines-faqs-social-network-analysis/

- https://ddu.ext.unb.ca/6634/Lecture_notes/Lecture_4_centrality_measure.pdf
- https://www.joshobrouwers.com/series/ariannas-thread/
- https://bab2min.tistory.com/554