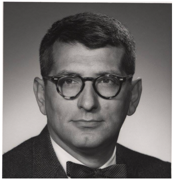
CSI 445/660 – Part 6
(Centrality Measures for Networks)

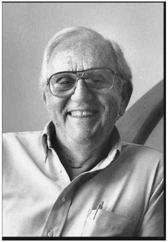
References

1. L. Freeman, "Centrality in Social Networks: Conceptual Clarification", *Social Networks*, Vol. 1, 1978/1979, pp. 215–239.
2. S. Wasserman and K. Faust, *Social Network Analysis: Methods and Applications*, Cambridge University Press, New York, NY, 1994.
3. M. E. J. Newman, *Networks: An Introduction*, Oxford University Press, New York, NY, 2010.
4. Wikipedia entry on Centrality Measures:
   `https://en.wikipedia.org/wiki/Centrality`

Some Pioneers on the Topic



- Alex Bavelas (1913–1993) **(??)**
- Received Ph.D. from MIT (1948) in Psychology.
- Dorwin Cartwright was a member of his Ph.D. thesis committee.
- Taught at MIT, Stanford and the University of Victoria (Canada).



- Harold Leavitt (1922–2007)
- Received Ph.D. from MIT.
- Authored an influential text ("Managerial Psychology") in 1958.
- Taught at Carnegie Mellon and Stanford.

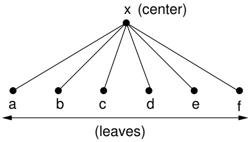
Centrality Measures for Networks

Centrality:

- Represents a “measure of importance”.
  - Usually for nodes.
  - Some measures can also be defined for edges (or subgraphs, in general).
- Idea proposed by Alex Bavelas during the late 1940’s.
- Further work by Harold Leavitt (Stanford) and Sidney Smith (MIT) led to **qualitative** measures.
- **Quantitative** measures came years later. (Many such measures have been proposed.)

Point Centrality – A Qualitative Measure

Example:



- The **center** node is "structurally more important" than the other nodes.

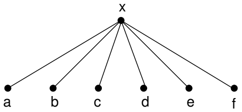
Reasons for the importance of the center node:

- The center node has the maximum possible degree.
- It lies on the shortest path ("geodesic") between any pair of other nodes (leaves).
- It is the closest node to each leaf.
- It is in the "thick of things" with respect to any communication in the network.

Degree Centrality – A Quantitative Measure

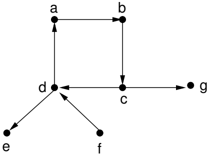
- For an **undirected** graph, the **degree** of a node is the number of edges incident on that node.
- For a **directed** graph, both **indegree** (i.e., the number of incoming edges) and **outdegree** (i.e., the number of outgoing edges) must be considered.

Example 1:



- Degree of $x = 6$.
- For all other nodes, degree $= 1$.

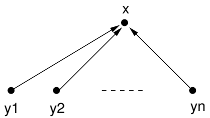
Example 2:



- Indegree of $b = 1$.
- Outdegree of $d = 2$.

Degree Centrality (continued)

When does a large indegree imply higher importance?

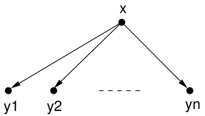


- Consider the Twitter network.
- Think of x as a **celebrity** and the other nodes as followers of x.

- For a different context, think of each node in the directed graph as a web page.
- Each of the nodes y_1, y_2, ..., y_n has a link to x.
- The larger the value of n, the higher is the "importance" of x (a crude definition of **page rank**).

Degree Centrality (continued)

When does a large outdegree imply higher importance?



- Consider the hierarchy in an organization.
- Think of x as the manager of y_1, y_2, ..., y_n.
- Large outdegree may mean more "power".

Undirected graphs:

- High degree nodes are called **hubs** (e.g. airlines).
- High degree may also also represent higher **risk**.

  Example: In disease propagation, a high degree node is more likely to get infected compared to a low degree node.

Normalized Degree

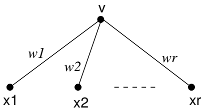
Definition: The **normalized degree** of a node x is given by

$$\text{Normalized Degree of } x = \frac{\text{Degree of } x}{\text{Maximum possible degree}}$$

- Useful in comparing degree centralities of nodes between two networks.

Example: A node with a degree of 5 in a network with 10 nodes may be relatively more important than a node with a degree of 5 in a network with a million nodes.

Weighted Degree Centrality (Strength):



- Weighted degree (or strength) of $v = w_1 + w_2 + \ldots + w_r$.

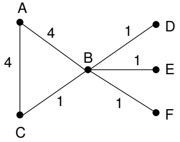
Degree Centrality (continued)

Assuming an **adjacency list** representation

- for an undirected graph $G(V, E)$, the degree (or weighted degree) of all nodes can be computed in **linear** time (i.e., in time $O(|V| + |E|)$) and
- for a directed graph $G(V, E)$, the indegree or outdegree (or their weighted versions) of all nodes can be computed in **linear** time.

Combining degree and strength: (**[Opsahl et al. 2009]**)

Motivating Example:



- A and B have the same strength.
- However, B seems more central than A.

Combining Degree and Strength (continued)

Proposed Measure by Opsahl et al.:

- Let d and s be the degree and strength of a node v respectively.
- Let α be a parameter satisfying the condition $0 \leq \alpha \leq 1$.
- The combined measure for node $v = d^{\alpha} \times s^{1-\alpha}$.
- When $\alpha = 1$, the combined measure is the **degree**.
- When $\alpha = 0$, the combined measure is the **strength**.
- A suitable value of α must be chosen for each context.

Farness and Closeness Centralities

Assumptions:

- Undirected graphs. (Extension to directed graphs is straightforward.)
- Connected graphs.
- No edge weights. (Extension to weighted graphs is also straightforward.)

Notation:

- Nodes of the graph are denoted by v_1, v_2, ..., v_n.
- For any pair of nodes v_i and v_j, d_{ij} denotes the number of edges in a shortest path between v_i and v_j.

The **farness centrality** f_i of node v_i is given by

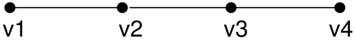
$$\begin{aligned} f_i &= \text{Sum of the distances between } v_i \text{ and the other nodes} \\ &= \sum_{v_j \in V - \{v_i\}} d_{ij} \end{aligned}$$

Farness and Closeness Centralities (continued)

The **closeness centrality** (or **nearness centrality**) η_i of node v_i is given by $\eta_i = 1/f_i$.

Note: If a node x has a larger closeness centrality value compared to a node y, then x is more central than y.

Example 1:

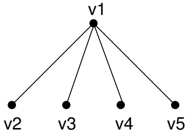


- $f_1 = 1 + 2 + 3 = 6$. So, $\eta_1 = 1/6$.
- $f_2 = 1 + 1 + 2 = 4$. So, $\eta_2 = 1/4$.
- $f_3 = 2 + 1 + 2 = 4$. So, $\eta_3 = 1/4$.
- $f_4 = 3 + 2 + 1 = 6$. So, $\eta_4 = 1/6$.

So, in the above example, nodes v_2 and v_3 are more central than nodes v_1 and v_4.

Farness and Closeness Centralities (continued)

Example 2:



- $f_1 = 4$. So, $\eta_1 = 1/4$.
- For every other node, the farness centrality value $= 7$; so the closeness centrality value $= 1/7$.
- Thus, v_1 is more central than the other nodes.

Additional Remarks:

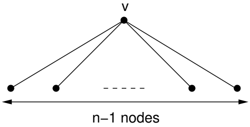
- For any graph with n nodes, the **farness centrality** of each node is **at least** $n-1$.

  Reason: Each of the other $n-1$ nodes must be at a distance of at least 1.

Farness and Closeness Centralities (continued)

Additional Remarks (continued):

- Since the farness centrality of each node is at least $n-1$, the **closeness centrality** of any node must be **at most** $1/(n-1)$.



- For the star graph on the left, the closeness centrality of the center node v is exactly $1/(n-1)$.

- If G is an n-clique, then the closeness centrality of each node of G is $1/(n-1)$.

Eccentricity Measure

- Recall that **farness centrality** of a node v_i is given by

$$f_i \;=\; \sum_{v_j \in V - \{v_i\}} d_{ij}$$

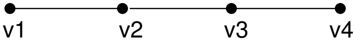
- The **eccentricity** μ_i of node v_i is defined by replacing the **summation** operator $\left(\sum\right)$ by the **maximization** operator; that is,

$$\mu_i \;=\; \max_{v_j \in V - \{v_i\}} \{d_{ij}\}$$

- This measure was studied by two graph theorists (Gert Sabidussi and Seifollah L. Hakimi).
- **Interpretation:** If μ_i denotes the eccentricity of node v_i, then every other node is within a distance of **at most** μ_i from v_i.
- If the eccentricity of node x is less than that of y, then x is more central than y.

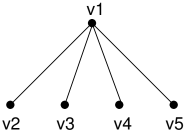
Examples for Eccentricity Measure

Example 1:



- $\mu_1 = \max\{1, 2, 3\} = 3.$
- $\mu_2 = \max\{1, 1, 2\} = 2.$
- $\mu_3 = \max\{2, 1, 1\} = 2.$
- $\mu_4 = \max\{3, 2, 1\} = 3.$

Example 2:

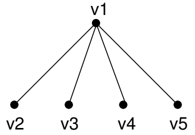


- $\mu_1 = 1.$
- For every other node, eccentricity $= 2$.

Eccentricity – Additional Definitions

Definition: A node v of a graph which has the smallest eccentricity among all the nodes is called a **center** of the graph.

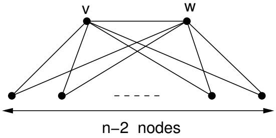
Example:



- The center of this graph is v_1. (The eccentricity of $v_1 = 1$.)

Note: A graph may have two or more centers.

Example:



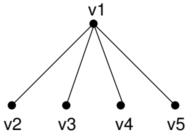
- Both v and w are centers of this graph. (Their eccentricities are $= 1$.)
- If G is clique on n nodes, then every node of G is a center.

Eccentricity – Additional Definitions (continued)

Definition: The smallest eccentricity value among all the nodes is called the **radius** of the graph.

Note: The value of the radius is the eccentricity of a center.

Example:



- The radius of this graph is 1 (since v_1 is the center of this graph and the eccentricity of $v_1 = 1$.)

Facts:

- The **largest eccentricity value** is the **diameter** of the graph.
- For any graph, the diameter is at most twice the radius. **(Students should try to prove this result.)**

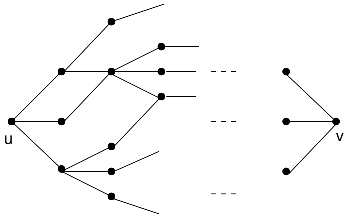
Random Walk Based Centrality (Brief Discussion)

Ref: **[Noh & Rieger 2004]**

Motivation:

- Definitions of centrality measures (such as **closeness** centrality) assume that "information" propagates along shortest paths.
- This may not be appropriate for certain other types of propagation. For example, propagation of diseases is a **probabilistic** phenomenon.

Idea of Random Walk Distance in a Graph:



Note: Assume that we want to do a random from u to v.

Random Walk ... (Brief Discussion)

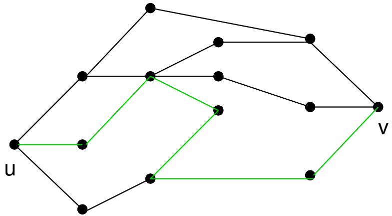
Random Walk Algorithm – Outline:

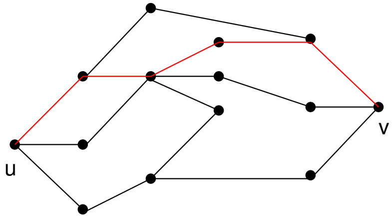
- Suppose we want to find the random walk distance from u to v.
- **Initialize:** Current Node $= u$. No. of steps $= 0$.
- **Repeat**
  1. Randomly choose a neighbor x of the Current Node.
  2. No. of steps $=$ No. of steps $+ 1$.
  3. Set Current Node $= x$.

  Until Current Node $= v$.

Note: In Step 1 of the loop, if the Current Node has degree d, probability of choosing any neighbor is $1/d$.

Examples of Random Walks:





Random Walk ... (Brief Discussion)

Definition: The **random walk distance** (or **hitting time**) from u to v is the expected number of steps used in a random walk that starts at u and ends at v.

- One can define farness/closeness centrality measures based on random walk distances.
- **Weakness:** Even for undirected graphs, the random walk distances are **not symmetric**; that is, the random walk distance from u to v may **not** be the same as the random walk distance from v to u.