Centrality & Clustering

Social Computing

Department of Computer Science
University of Massachusetts, Lowell

Hadi Amiri
hadi@cs.uml.edu

Lecture Topics

- Centrality
  - Degree Centrality
  - Closeness Centrality
  - Betweenness Centrality

Centrality

- What characterizes an important node in a network?
  - Most influential people in social nets,
  - Key infrastructure nodes in the Internet
  - Main spreaders of disease
  - Etc.

- Structural view:
  - Importance of a node is related to its **position in the network**.

Centrality Measures

- Different **centrality measures** capture different **structural characteristics of nodes**!
- There is often a high correlation between these measures!
- Sometimes the most important node might depend on which measure is used!

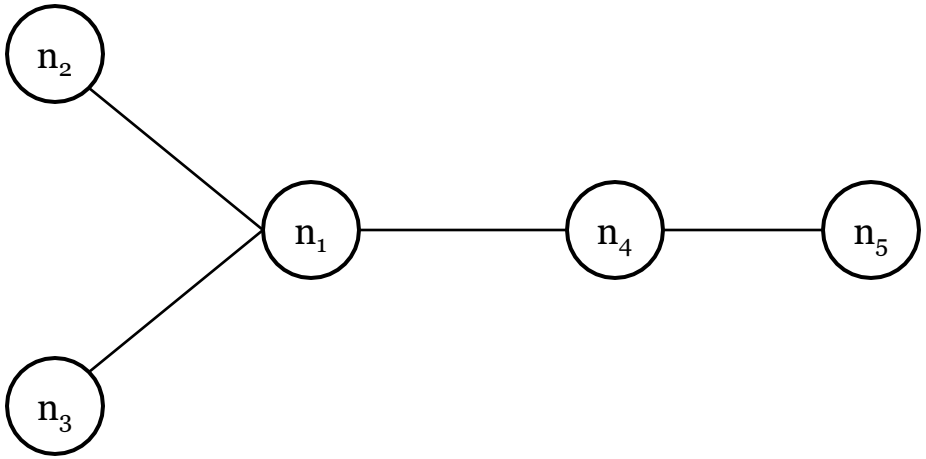
- C : Centrality
  - C (i) : Centrality for node i
  - C($\mathbf{A}$) : Centrality for a group of nodes $\mathbf{A} \in \mathbf{N}$

Centrality Measures- Cnt.

- Centrality
  - **Degree Centrality**
  - Closeness Centrality
  - Betweenness Centrality

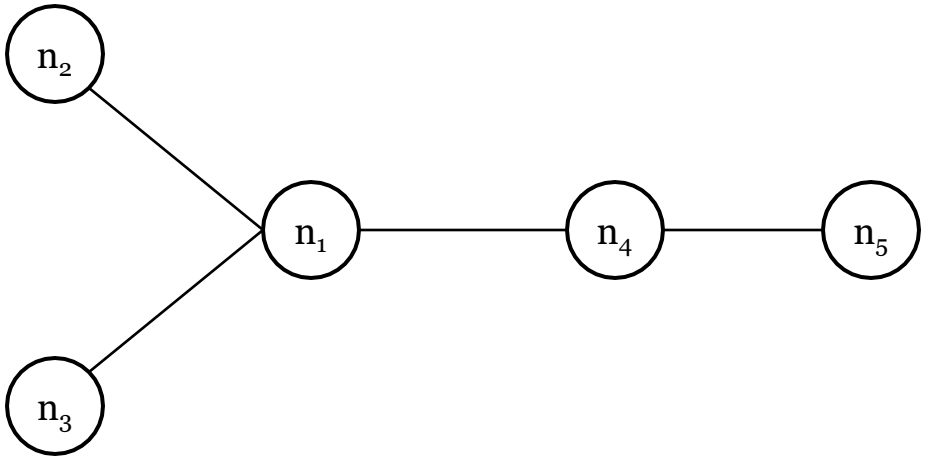
Degree Centrality

- A node is central if it has **links to many nodes**.
  - Look at the node degree



Degree Centrality- Cnt.

- A node is central if it has links to many nodes.
  - Look at the node degree

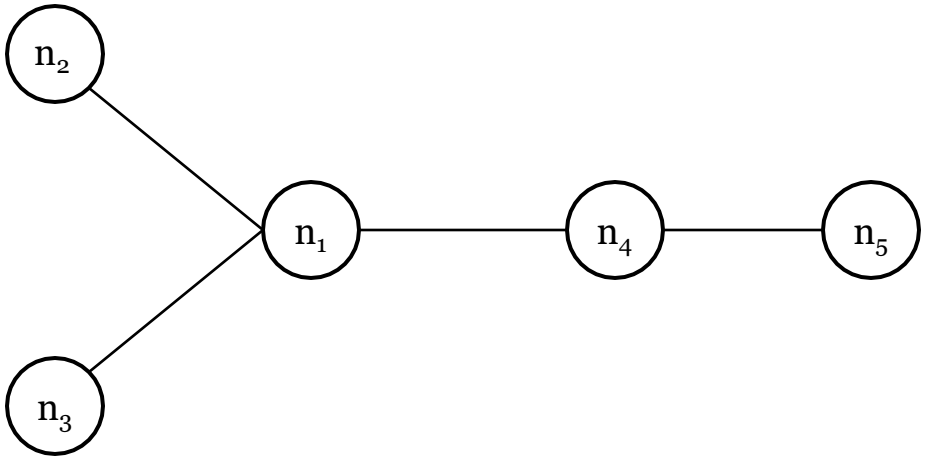


	n1	n_2	n_3	n_4	n_5	
n_1	0	1	1	1	0	**3**
n_2	1	0	0	0	0	**1**
n_3	1	0	0	0	0	**1**
n_4	1	0	0	0	1	**2**
n_5	0	0	0	1	0	**1**
	3	**1**	**1**	**2**	**1**	

Adjacency Matrix (A)

Degree Centrality- Cnt.

- Standardized Degree Centrality
  - Divide by the maximum possible degree centrality value!
    - N-1



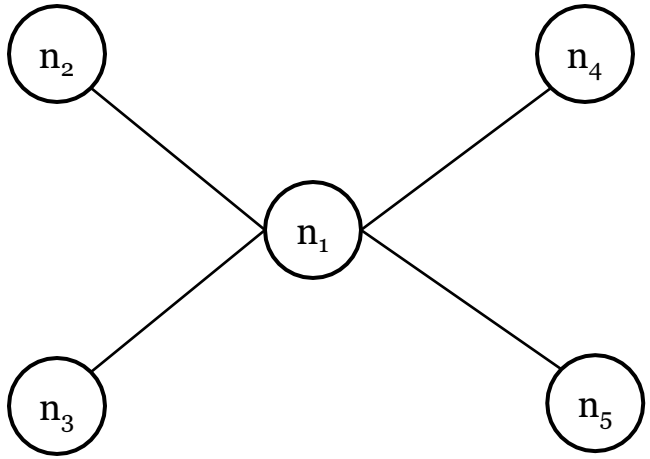
	n1	n_2	n_3	n_4	n_5	
n_1	0	1	1	1	0	3/4
n_2	1	0	0	0	0	1/4
n_3	1	0	0	0	0	1/4
n_4	1	0	0	0	1	1/2
n_5	0	0	0	1	0	1/4

Centrality Measures- Cnt.

- Centrality
  - Degree Centrality
  - **Closeness Centrality**
  - Betweenness Centrality

Closeness Centrality

- A node is central if it is **close to other nodes**.
  - Look at distance btw nodes
  - Closeness: 1 / Sum of distance to other nodes



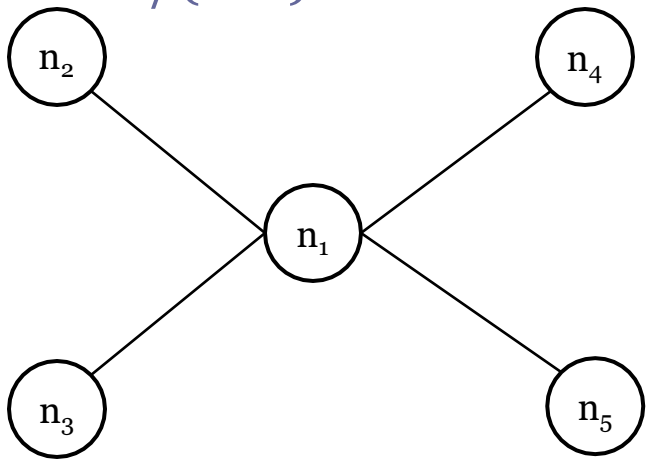
$$C(n_1) = 1/(\sum_{j=1}^{n} D_{1j}) = 1/(\sum_{i=1}^{n} D_{i1}) = 1/4$$

	n1	n_2	n_3	n_4	n_5	
n_1	0	1	1	1	1	1/4
n_2	1	0	2	2	2	1/7
n_3	1	2	0	2	2	1/7
n_4	1	2	2	0	2	1/7
n_5	1	2	2	2	0	1/7

Distance Matrix (D)

Closeness Centrality- Cnt.

- Standardized Closeness Centrality
  - Divide by the maximum possible closeness centrality value!
    - 1/(N-1)



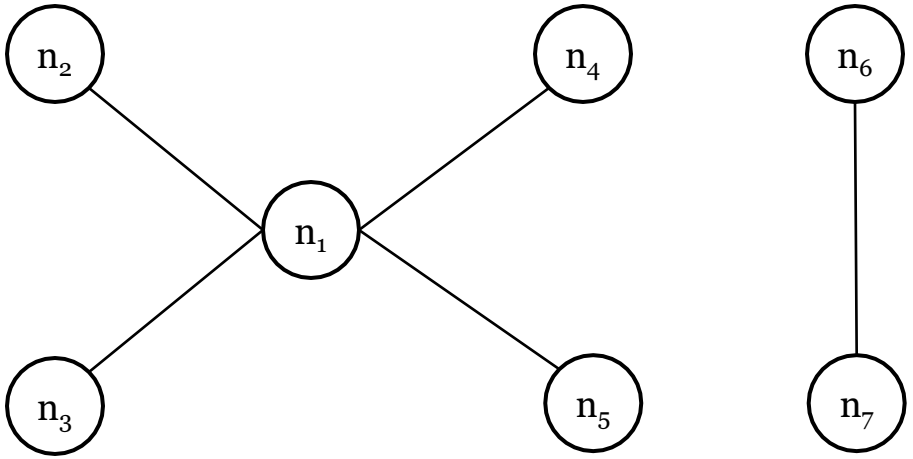
$$C(n_1) = (N-1)/(\sum_{j=1}^{n} D_{1j}) = (N-1)/(\sum_{i=1}^{n} D_{i1}) = 4/4$$

	n1	n_2	n_3	n_4	n_5	
n_1	0	1	1	1	1	4/4
n_2	1	0	2	2	2	4/7
n_3	1	2	0	2	2	4/7
n_4	1	2	2	0	2	4/7
n_5	1	2	2	2	0	4/7

Distance Matrix (D)

Closeness Centrality- Cnt.

- How to compute Closeness Centrality in networks with disconnected components?



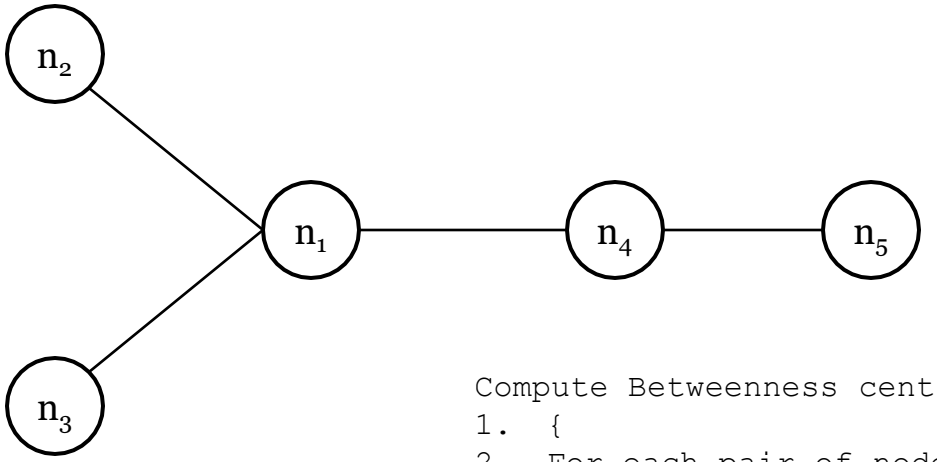
- Only consider the giant component or do graph sampling?
- Only consider nodes that are reachable in paths of length 1, 2, …
  This is called k-Step Reach!

Centrality Measures- Cnt.

- Centrality
  - Degree Centrality
  - Closeness Centrality
  - **Betweenness Centrality**

Betweenness Centrality

- A node is central if **other nodes have to go through it** to reach each other.
  - Look at shortest paths between nodes




```
Compute Betweenness centrality for a target node:
1.  {
2.  For each pair of nodes, compute the shortest paths
    between them.
3.  For each pair of nodes, determine the fraction of
    shortest paths that pass through the target node.
4.Sum the fractions over all pairs of nodes
5.  }
```

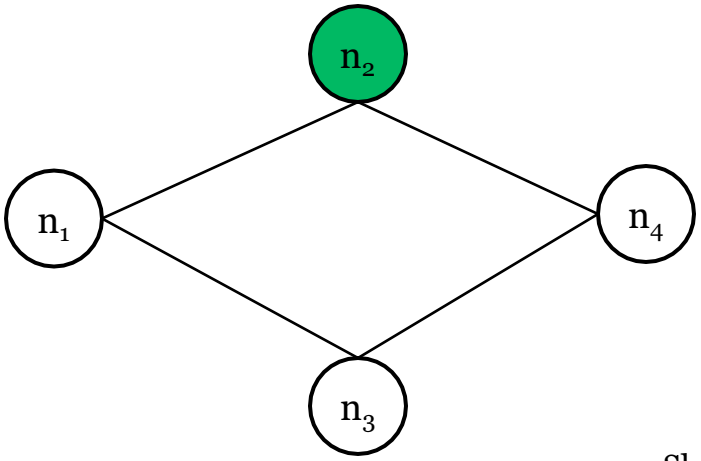
Betweenness Centrality- Cnt.

s_{jk} Number of shortest paths btw nodes n_j and n_k

$s_{jk}(n_i)$ Number of shortest paths btw nodes n_j and n_k that include node n_i

$\frac{s_{jk}(n_i)}{s_{jk}}$ Proportion of shortest paths btw nodes n_j and n_k that include node n_i

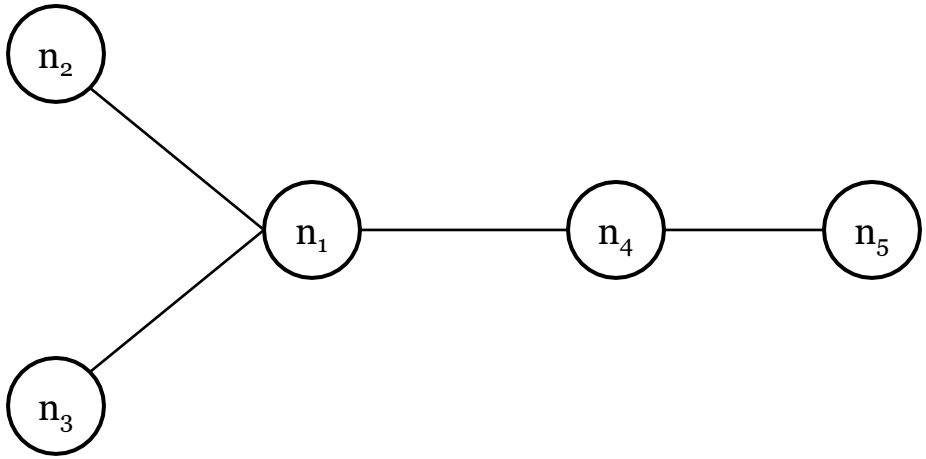
$\text{Sum}(_{j,k!=i} \frac{s_{jk}(n_i)}{s_{jk}})$ Proportion of shortest paths btw all nodes that include node n_i



Shortest paths n_1-n_4	n_1-n_2-n_4, n_1-n_3-n_4
s_{14}	2
$s_{14}(n_2)$	1
$s_{14}(n_2)$ /s_{14}	½
$C(n_2)$	½

Shortest paths btw n_1-n_3 and n_3-n_4 don't include n_2! Their corresponding proportions are 0.

Betweenness Centrality- Cnt.

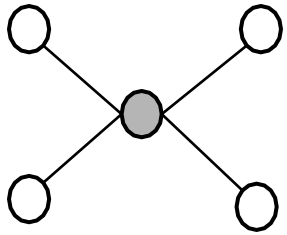


Pair	Shortest path
n1 n2	n1-n2
n1 n3	n1-n3
n1 n4	n1-n4
n1 n5	n1-n4-n5
n2 n3	n2-n1-n3
n2 n4	n2-n1-n4
n2 n5	n2-n1-n4-n5
n3 n4	n3-n1-n4
n3 n5	n3-n1-n4-n5

Betweenness	
n1	5
n2	0
n3	0
n4	3
n5	0

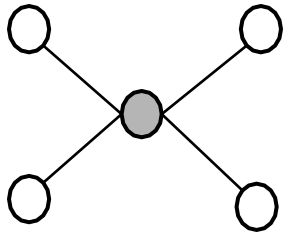
Betweenness Centrality- Cnt.

- Standardized Betweenness Centrality
  - Divide by the maximum possible betweenness centrality value!
    - ?



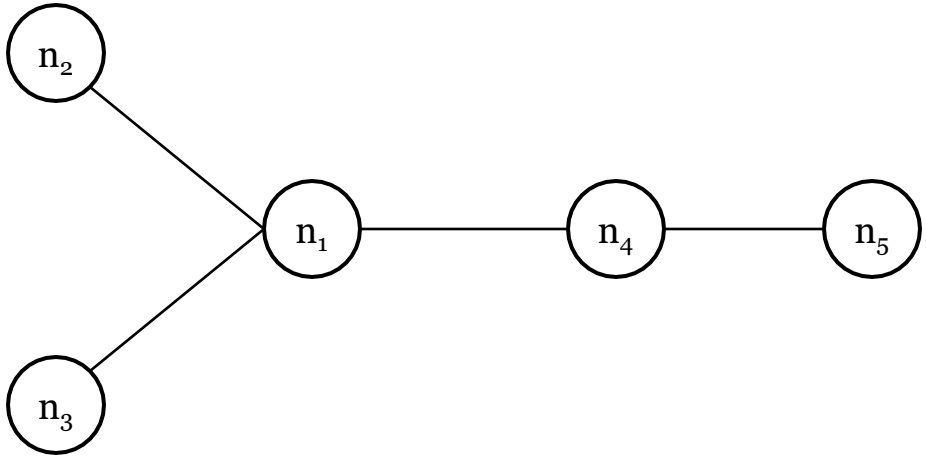
Betweenness Centrality- Cnt.

- Standardized Betweenness Centrality
  - Divide by the maximum possible betweenness centrality value!
    - (N-1)(N-2)/2 : the number of other pairs of nodes (exclude the node itself)



Betweenness Centrality- Cnt.

- Standardized Betweenness Centrality
  - Divide by the maximum possible betweenness centrality value!
    - (N-1)(N-2)/2 : the number of other pairs of nodes (exclude the node itself)



	Betweenness	Stnd. Betweenness
n1	5	5/6 = 0.83
n2	0	0/6 = 0.00
n3	0	0/6 = 0.00
n4	3	3/6 = 0.50
n5	0	0/6 = 0.00

Reading

- Ch.03 Strong and Weak Ties [NCM]