Social Network Analysis

Measures of centrality and prestige

Procheta Sen

UNIVERSITY OF LIVERPOOL

Terminology

- **Actors**: objects of interest (e.g. people)
- Actors have **interactions** or **relationships**
- This can be represented by a graph $G = (V, E)$ (nodes are actors, edges are relationships)
- Examples: Facebook, LinkedIn, Co-authorship network,…

What we can do with such a network?

- Study structural properties
- Identify “central” or “influential” nodes
- Identify critical nodes
- Detect “communities” formed by a group of actors
- Detect abnormal substructures (e.g. link farms in the web graph)
- New link prediction
- Social influence analysis
- etc.

Measures of vertex “importance”

Which nodes are more “important” and which are less “important”?

- Measures of Centrality (for undirected graphs)
    - Degree centrality
    - Closeness centrality
    - Betweenness Centrality
- Measures of Prestige (for directed graphs)
    - Degree prestige
    - Proximity prestige

Measures of centrality: **Degree Centrality**

The **degree centrality** $C_D(i)$ of a node i of an undirected network is equal to the degree of the node, divided by the maximum possible degree of the nodes.

$$C_D(i) = \frac{\deg(i)}{n-1}$$

Motivation: nodes with higher degree are often hub nodes, they tend to be more central to the network and bring distant parts of the network closer together.

Measures of centrality: **Degree Centrality**

Motivation: nodes with higher degree are often hub nodes, they tend to be more central to the network and bring distant parts of the network closer together.

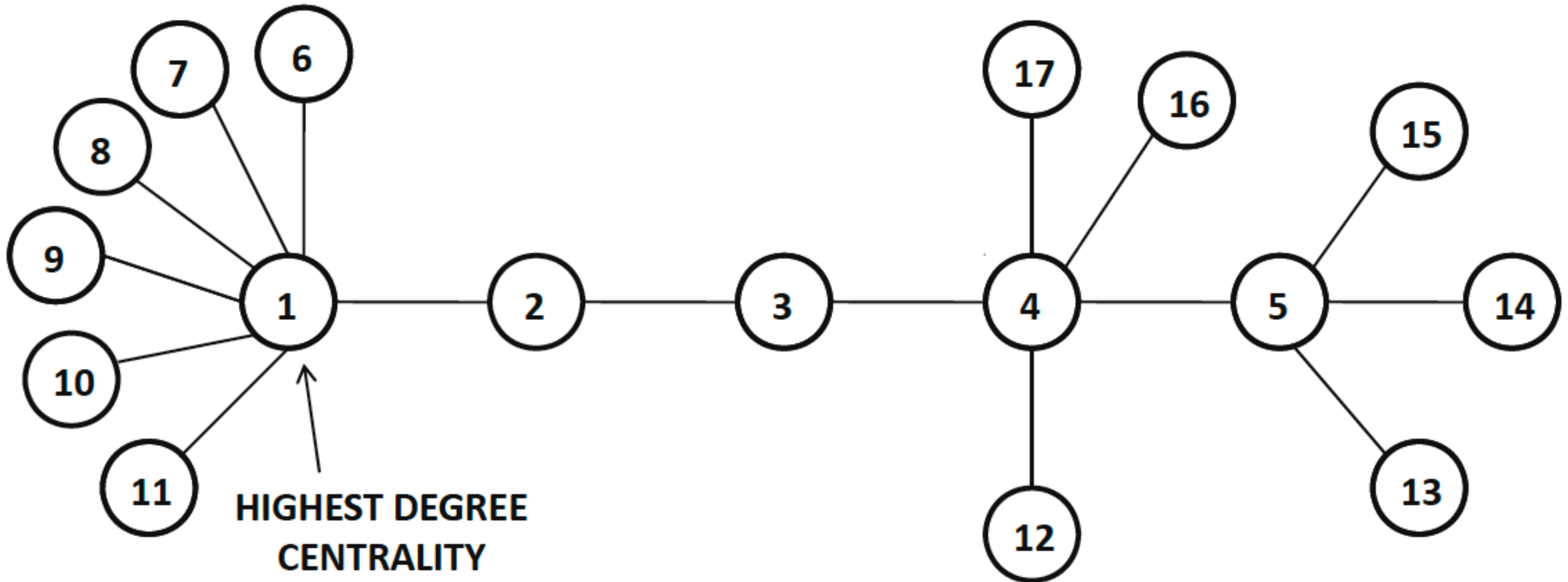
Zhang, X., Xu, G., & Xia, Y. (2018). Optimal defense resource allocation in scale-free networks. *Physica A: Statistical Mechanics and its Applications*, *492*, 2198-2204.

Measures of centrality: **Degree Centrality**

- The major problem with degree centrality is that it uses local information only: it does not consider nodes beyond the immediate neighborhood of a given node i.

- The overall structure of the network is ignored to some extent.

- **Example**: node 1 has the highest degree centrality, but it cannot be viewed as central to the network itself.



Measures of centrality: **Closeness Centrality**

The **closeness centrality** is defined for undirected and connected graphs.

$\mathrm{AvDist}(i)$: the average shortest path distance, starting from node i

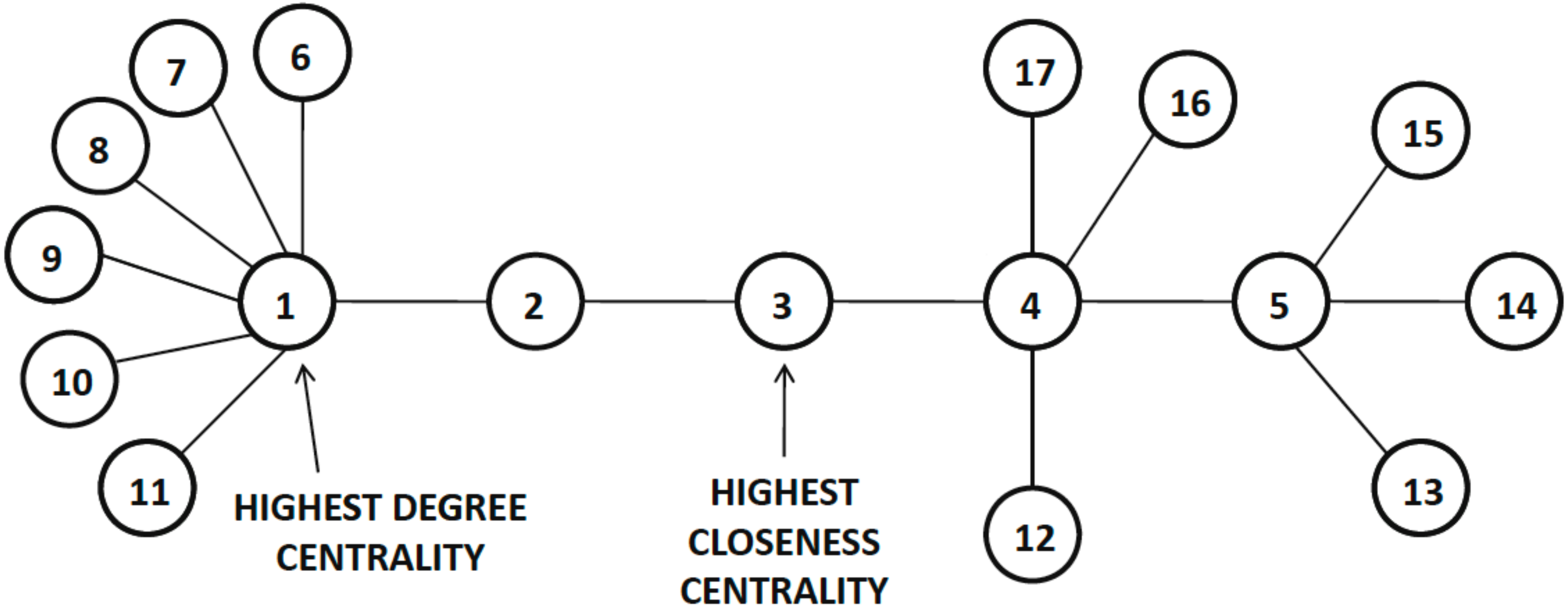
$$\mathrm{AvDist}(i) = \frac{\sum_{j=1}^{n} \mathrm{dist}(i,j)}{n-1}$$

The **closeness centrality** $C_C(i)$ of i is the inverse of the average distance $\mathrm{AvDist}(i)$

$$C_C(i) = \frac{1}{\mathrm{AvDist}(i)}$$

Because the value of $\mathrm{AvDist}(i)$ is at least 1, the closeness centrality $C_C(i)$ ranges between 0 and 1.

Measures of centrality: **Closeness Centrality**



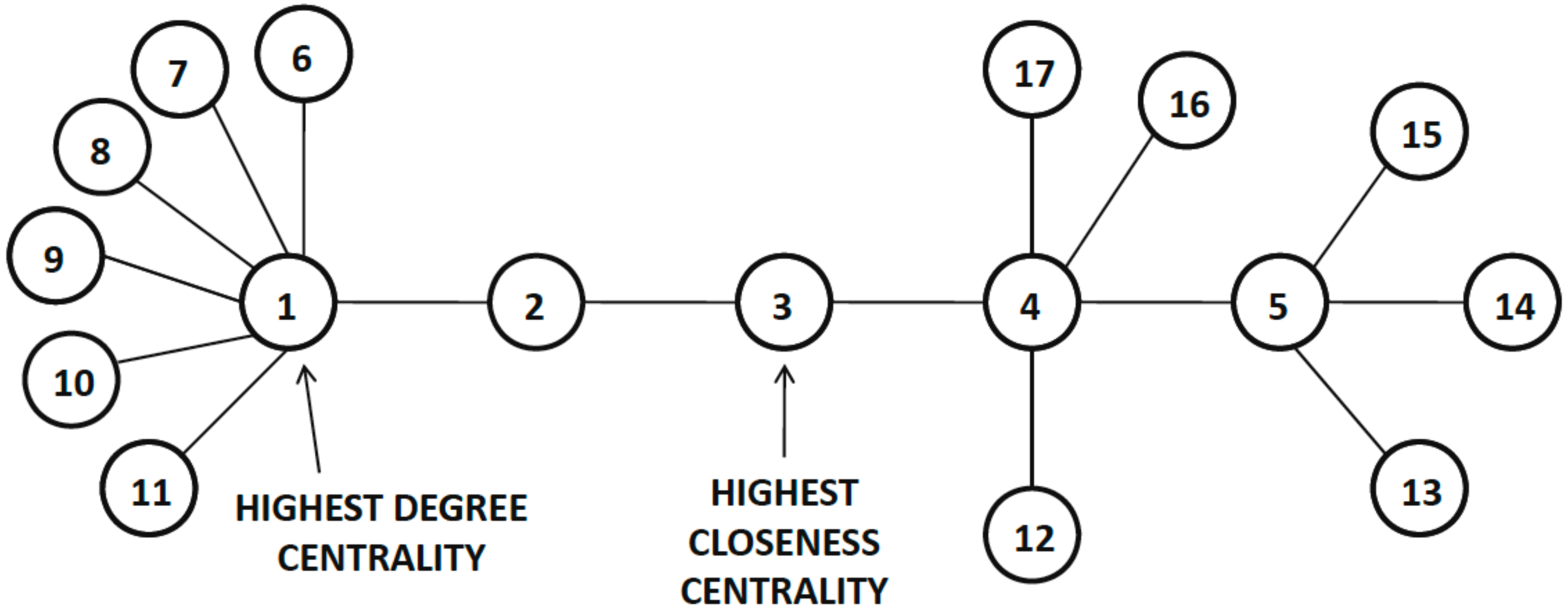
node 3 has the highest closeness centrality because it has the lowest average distance to other nodes

Measures of centrality: **Betweenness Centrality**

The **closeness centrality**

- is based on the notion of distance
- does not take into account the **criticality** of the node in terms of the number of shortest paths that pass through it
- such notions of criticality are crucial in determining actors that have the greatest control of the flow of information between other actors in a social network

Measures of centrality: **Betweenness Centrality**



- Node 3 has the highest closeness centrality, but
- Node 4 is more critical than Node 3 with respect to shortest path between different pairs of nodes:
  - Node 4 participates in shortest paths between the pairs of nodes directly incident with it, whereas Node 3 does not participate in these paths

Measures of centrality: **Betweenness Centrality**

q_{jk} : the number of shortest paths between nodes j and k

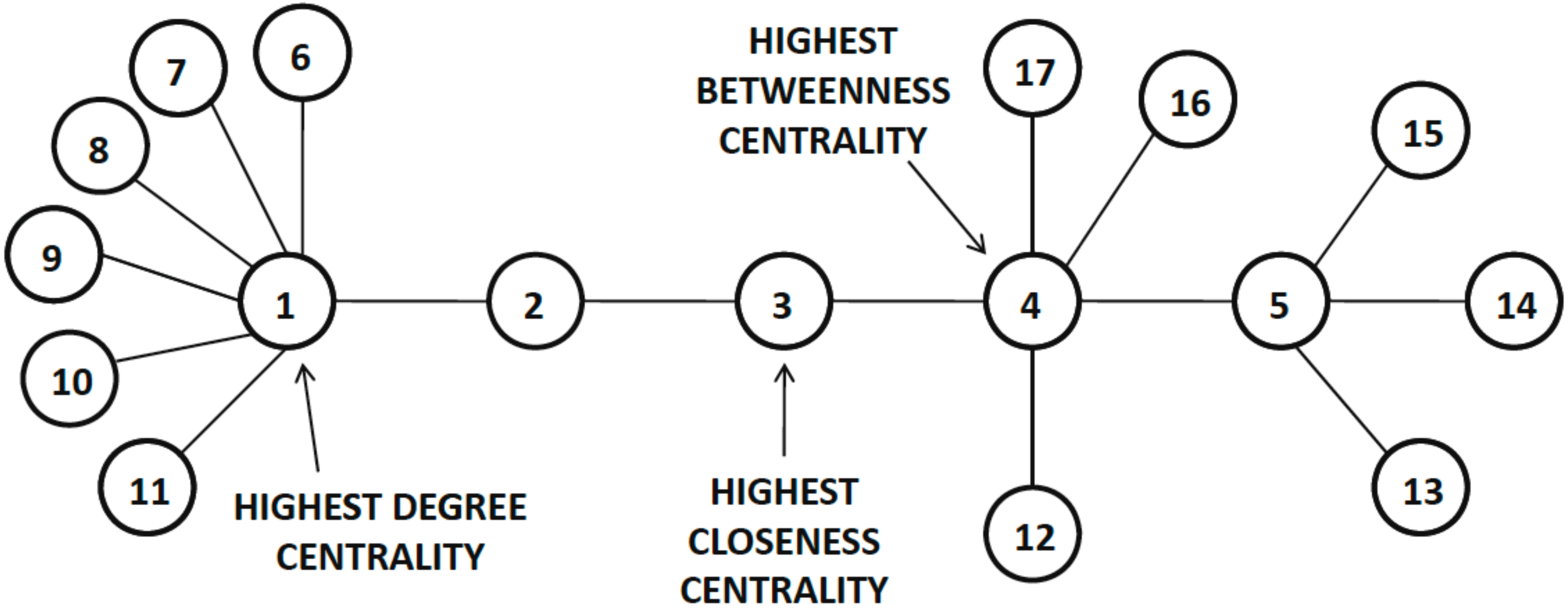
$q_{jk}(i)$: the number of shortest paths between nodes j and k that pass through node i

$f_{jk}(i) = \dfrac{q_{jk}(i)}{q_{jk}}$: the fraction of paths that pass through node i. Intuitively, $f_{jk}(i)$ is a fraction that indicates the level of control that node i has over nodes j and k in terms of regulating the flow of information between them.

The **betweenness centrality** $C_B(i)$ is the average value of $f_{jk}(i)$ over all $\binom{n}{2} = \dfrac{n(n-1)}{2}$ pairs of nodes j,k

$$C_B(i) = \frac{\sum_{j<k} f_{jk}(i)}{\binom{n}{2}}$$

Measures of centrality: **Betweenness Centrality**



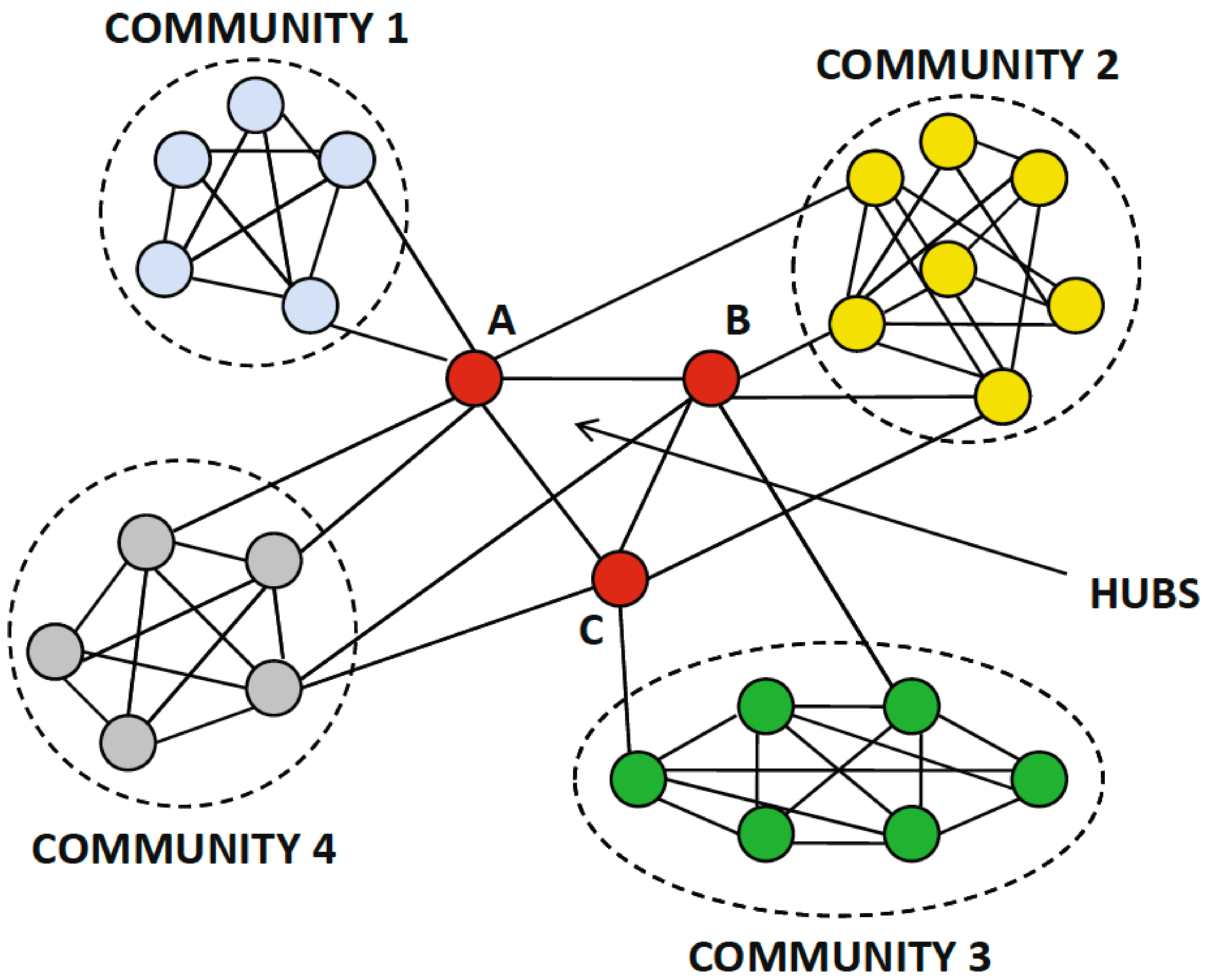
node 4 has the highest betweenness centrality

Measures of centrality: **Betweenness Centrality**

- The betweenness centrality is between 0 and 1
- Higher values correspond to better betweenness
- Betweenness centrality can be defined for disconnected networks

Measures of centrality: **Betweenness Centrality**

- Can be generalised to edges by using the number of shortest paths passing through an edge (rather than a node).
- Edges that have high betweenness tend to connect nodes from different clusters in the graph
- Node/Edge betweenness concepts are used in many community detection algorithms, such as the Girvan–Newman algorithm



the edges connected to the hub nodes have high betweenness.

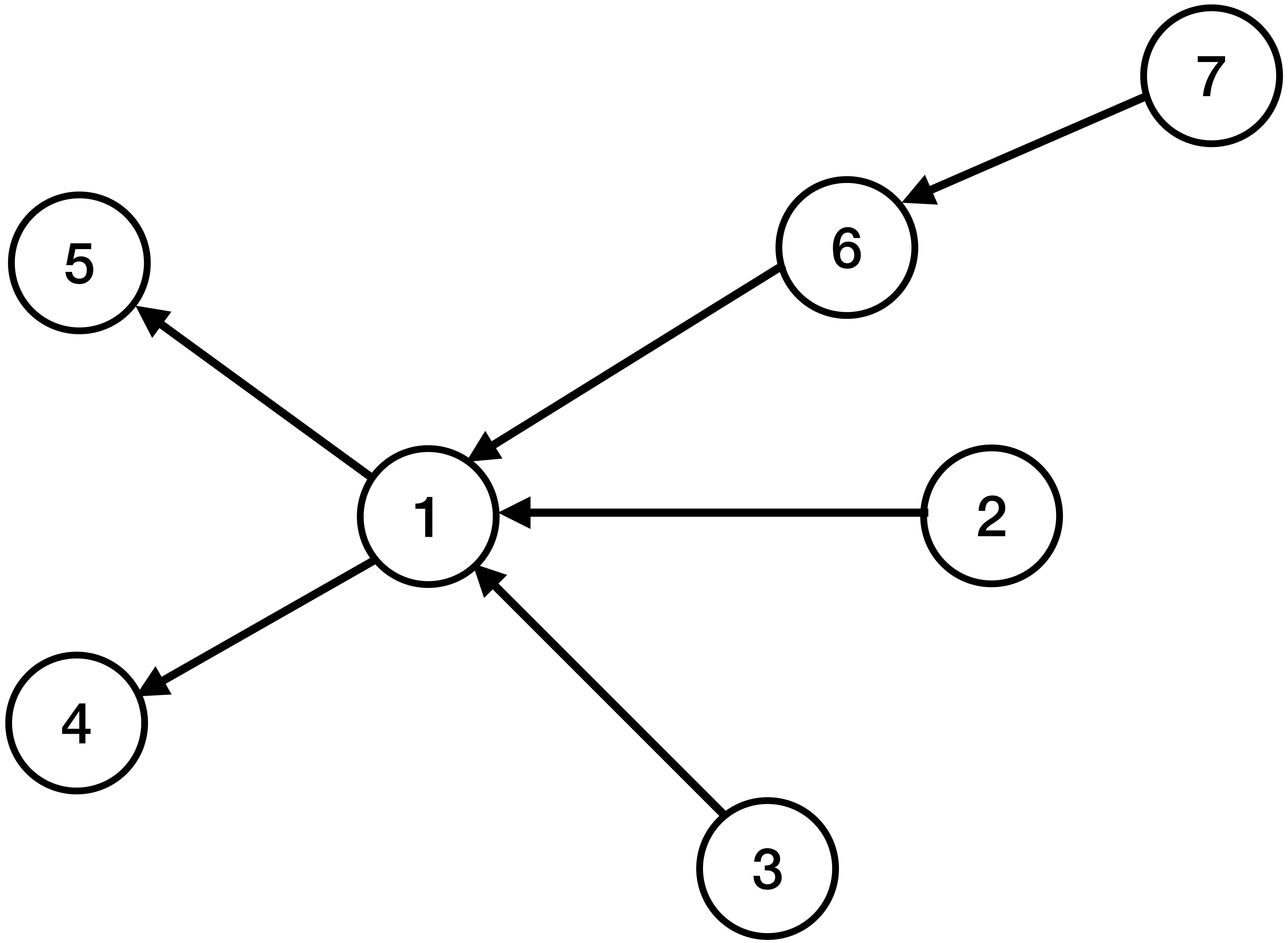
Measures of prestige: **Degree Prestige**

The **degree prestige** is defined for directed networks only, and uses the in-degree of the node, rather than its degree.

$$P_D(i) = \frac{\deg_+(i)}{n-1}$$

Motivation: only a high in-degree contributes to the prestige because the in-degree of a node can be viewed as a vote for the popularity of the node.

Measures of prestige: **Degree Prestige**

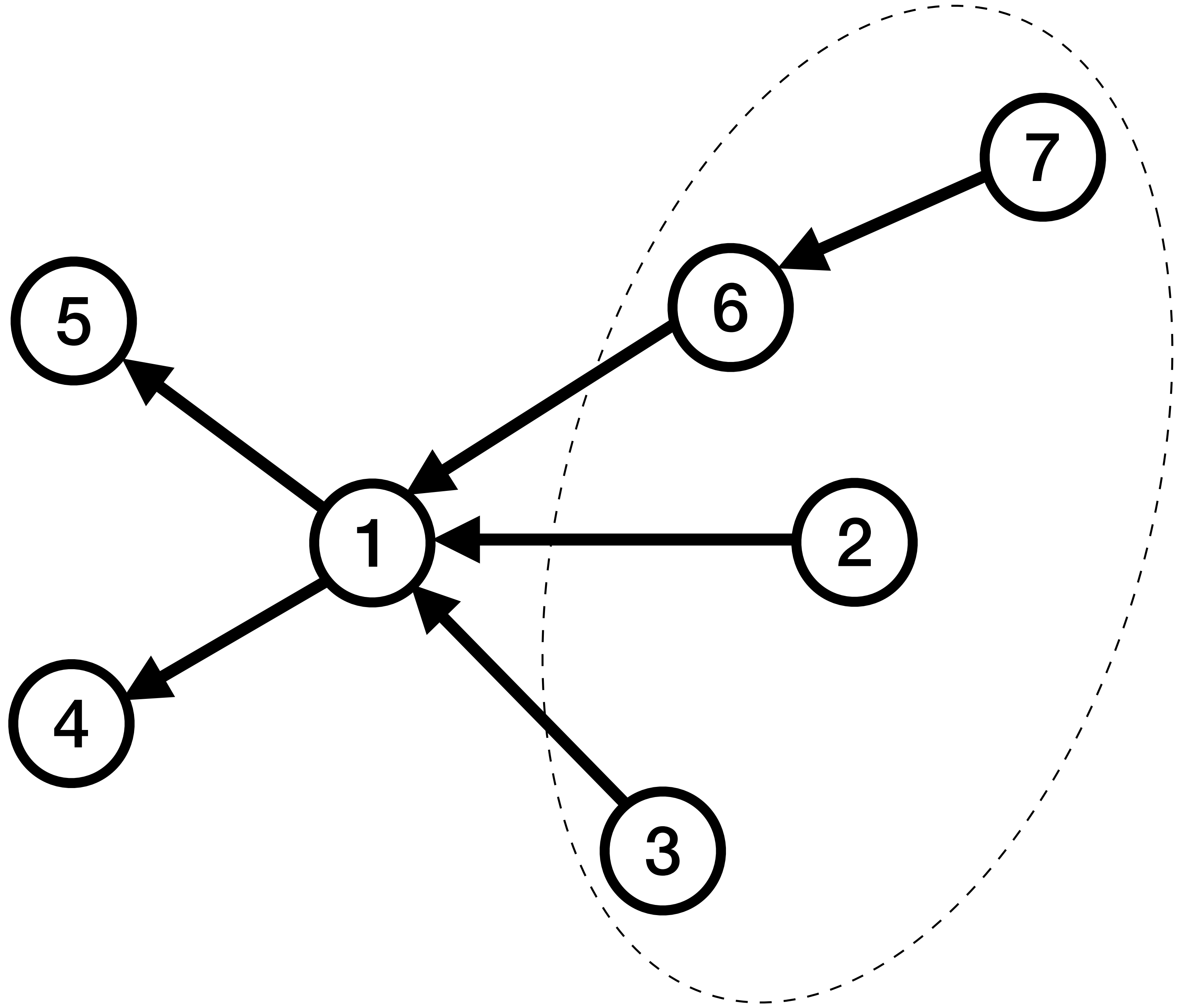


node 1 has the highest degree prestige

Measures of prestige: **Proximity Prestige**

The **proximity prestige** is defined for directed graphs.

Influence(i) : the set of nodes that can reach node i with a direct path.



Influence(1): influence set of node 1

Measures of prestige: **Proximity Prestige**

The **proximity prestige** is defined for directed graphs.

$\text{Influence}(i)$: the set of nodes that can reach node i with a direct path.

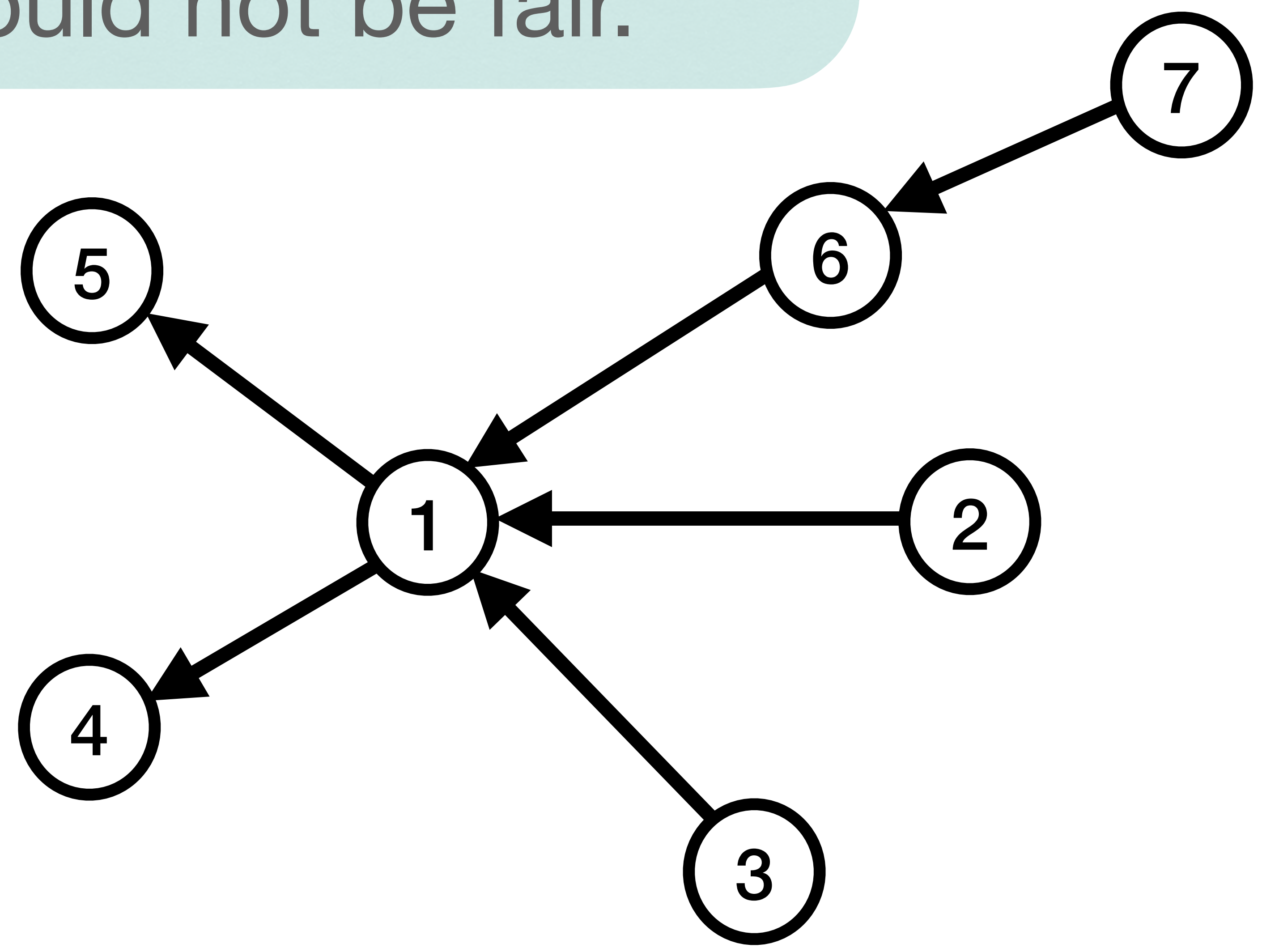
$\text{AvDist}(i)$: the average shortest path distance **to node** i

$$\text{AvDist}(i) = \frac{\sum_{j \in \text{Influence}(i)} \text{dist}(j, i)}{|\,\text{Influence}(i)\,|}$$

Measures of prestige: **Proximity Prestige**

Use of the inverse of the average distance, as in Closeness Centrality, would not be fair.

For example,



- for node 6, $\text{AvDist}(6) = 1$, but it has only one node in its influence set;
- for node 1, $\text{AvDist}(1) = 5/4$, but it has four nodes in its influence set.

While $\dfrac{1}{\text{AvDist}(6)} > \dfrac{1}{\text{AvDist}(1)}$, it is natural to say that node 1 has higher prestige than node 6.

To fix the problem, we use multiplicative penalty factor that measures the fractional size of the influence set of the node

Measures of prestige: **Proximity Prestige**

The **proximity prestige** is defined for directed graphs.

$\text{Influence}(i)$: the set of nodes that can reach node i with a direct path.

$\text{AvDist}(i)$: the average shortest path distance **to node** i

$$\text{AvDist}(i) = \frac{\sum_{j \in \text{Influence}(i)} \text{dist}(j, i)}{|\text{Influence}(i)|}$$

$$\text{InfluenceFraction}(i) = \frac{|\text{Influence}(i)|}{n-1}$$

The **proximity prestige** $P_P(i) = \dfrac{\text{InfluenceFraction}(i)}{\text{AvDist}(i)}$

Measures of prestige: **Proximity Prestige**

The **proximity prestige** is defined for directed graphs.

$$P_P(i) = \frac{\text{InfluenceFraction}(i)}{\text{AvDist}(i)}$$

- The proximity prestige lies between 0 and 1
- Higher values indicate greater prestige