# CS5228 – Tutorial 9

## Graph Mining

1. **Centrality Measures.** The centrality of a node/vertex in a graph $G$ measures its relative importance among all other nodes w.r.t. the graph structure. Figure 1 shows a direct graph $G$ with 12 nodes and 13 directed edges.

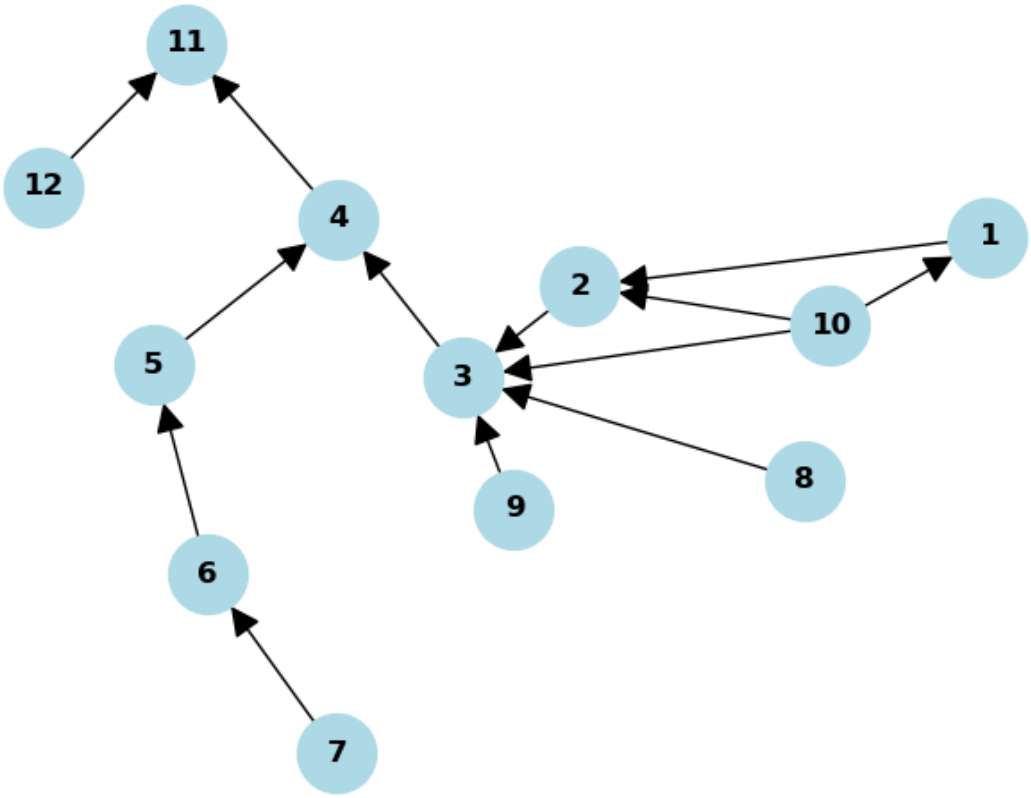


Figure 1: Example of a directed graph G

(a) Simply by eyeballing graph $G$ in Figure 1 try to identify the nodes with the highest score according to the following 5 centrality measures

- OutDegree:
- InDegree:
- PageRank:
- Closeness:
- Betweenness:

(b) Let's assume the nodes are simple websites with just a single page each. You're the owner of Site 3 and want to boost your PageRank score. Without deleting existing links and without creating additional sites (i.e., nodes), how can you boost your PageRank score to have the highest rank among all sites?

(c) PageRank has become famous for ranking websites w.r.t their relative importance compared to other sites. The underlying intuition is that a website is important (or trusted or authoritative) if many other important websites link to it. However, the idea of ranking nodes is of course not limited to websites, and there are also many other centrality measures to quantify a node's importance considering different aspects of the graph structure. For the following 5 centrality measures, for which application or data mining task would a specific measure arguably be the best choice?

- OutDegree
- InDegree
- PageRank
- Closeness
- Betweenness

(d) In the lecture, we saw the definition of PageRank being

$$c_{pr} = \alpha M c_{pr} + (1 - \alpha)E$$

where $c_{pr}$ is the vector of PageRank scores for all nodes, and $E = (1/n, 1/n, ...)^T$ with $n = |V|$. What is the intuition behind the term $(1 - \alpha)E$ and why do we need it?

(e) Which of the 5 centrality measures above can arguably also be used for finding communities in a graph?