Week 7-3: Paper Summaries

CE-510 Seminar: Social Media Mining

Student Name: <u>Jiaqi Guo</u>   NetID: <u>JGR9647</u>

■ PageRank: Standing on the shoulders of giants

In this paper, they review the PageRank method and link it to some renowned previous techniques that we have found in the fields of **Web information retrieval bibliometrics, sociometry, and econometrics**.

The PageRank algorithm calculates the PageRank value of each page, and then ranks the importance of the page according to the size of this value. The idea is to simulate a carefree Internet user, Internet users first randomly select a web page open, and then in this web site for a few minutes, jump to the page the link is pointing to, such doing nothing, aimlessly jumped on a web page, PageRank is to estimate the distribution of Internet users leisurely on the probability of each web page.

Possible Improvement Directions:

1. PageRank algorithm only uses the link structure of the network, cannot judge the similarity of web content; In addition, the algorithm distributes the weight of the outbound links in an average way so that the webpage with no related theme gets the same attention as the webpage with related theme, and the theme drift appears. So, we should probably incorporate more content related information into this algorithm.

2. A new page has relatively few links, and even though its content is of high quality, it still takes a long time to become a high PR page. A new page has relatively few links, and even though its content is of high quality, it still takes a long time to become a high PR page.

■ Centrality in Social Networks Conceptual Clarification

This paper reviews the intuitive background for measures of structural centrality in social networks and evaluated the existing measures in terms of their consistency with intuitions and their interpretability. Besides, it proposed the concept of degree centrality.

In Graph Theory and Network Analysis, Centrality is an indicator to judge the importance of nodes in a Network, and it is the quantification of the importance of nodes. Degree Centrality is the most direct measure to describe node Centrality in network analysis. The greater the degree of a node is, the higher the degree centrality of the node is, and the more important the node is in the network.

In an Undirected Graph, degree centrality measures the degree to which a node in a network is related to all

other nodes. For an undirected graph with G nodes, the degree centrality of node I is the total number of direct

connections between I and other G-1 nodes, which can be expressed by matrix as follows:

$$C_D(N_i) = \sum_{J=1}^{g} x_{ij}(i \neq j)$$

The node degree centrality measured thus not only reflects the correlation between each node and other nodes,

but also depends on the network size (G). In other words, the larger the network size, the higher the maximum

possible value of degree centrality.

Possible Improvement Directions:

3. Degree centrality is not a general metric, and corresponding metric should be selected according to different problems, such as Closeness or Betweenness centrality

4. Degree centralities fail to consider the scale of the graph, the larger the network size, the higher the maximum possible value of degree centrality. We need to add a standardized metric to the network to eliminate the impact of network size changes on centrality