COMP631 Introduction to IR
Homework 2

March 11, 2022

1) Submit s single PDF file containing your answers to each question. **You are highly encouraged to use LaTex or Microsoft Word.**
2) For calculation questions, please provide the analysis and calculation steps. An answer with only a resultant value will not receive any score.
– Due Date: March 21, 2022 11:59pm

1 HITS and PageRank (35 pts)

Given the figure including a graph as below:

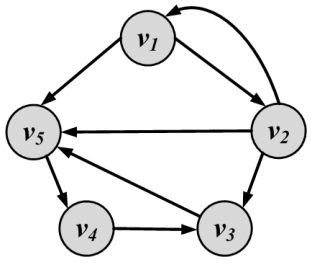

Figure 1: A toy example of a graph

1) (8 pts) For each graph node in Figure 1, calculate the degree centrality *normalized by the degree sum*, as well as the rank with respect to the centrality.

2) (8 pts) In the context of world wide web or social media, given a concrete real-world example for entities that can be recognized as "authorities".

3) (12 pts) Calculate the PageRank values for the graph in Figure 1 by applying the iterative algorithm. (Please show 2 iteration steps.)

4) (7 pts) Explain the limitations of Katz Centrality, and how PageRank overcome these limitations.

2 Graph Analysis (35 pts)

Given the graph as below:

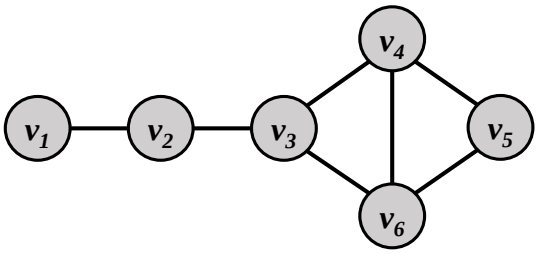

Figure 2: Another toy example of a graph

1) (7 pts) Calculate the Closeness centrality of node v_5.
2) (7 pts) Calculate the Betweenness centrality of node v_4.
3) (7 pts) Calculate the Jaccard similarity between v_3 and v_5.
4) (7 pts) Calculate the local clustering coefficient of v_4.
5) (7 pts) Explain how we can use similarity measure or clustering coefficient for friend recommendation in social network.

3 Results Assembly (30 pts)

1) (12 pts) For a multi-term query q, explain why we can only consider high-idf query terms for index elimination.

2) (18 pts) Consider a more general form of net score,

$$netscore(q, d) = \alpha \cdot g(d) + \beta \cdot cosine(q, d), \tag{1}$$

where α and β are weights, $0 \geq \alpha \geq 1$, $0 \geq \beta \geq 1$, $\alpha + \beta = 1$, explain how the IR system would behave with different relative weighting.