Problem 1 (PageRank). Consider the following directed graph $G(V,E)$

(a) Determine the transition probability matrix $P$ of the matrix chain induced by PageRank, assuming that the probabilities of outgoing links for each node are uniform.

(b) With teleportation probability $\epsilon = 0.15$, compute the stationary state distribution $\pi$ using the power iteration method. Begin with $\pi(0) = \langle \frac{1}{6}, \frac{1}{6}, \frac{1}{6}, \frac{1}{6}, \frac{1}{6}, \frac{1}{6} \rangle^T$ as initial state probabilities and perform three iterations. You may use a suitable software for calculation. Write down the values in each step, with float precision rounded to 3 digits after decimal point.

(c) Continue the calculation in (b) to get the final PageRank vector. Write the final PageRank vector (you don’t have to write the calculation results for intermediate steps). Which node has the highest PageRank?

(d) What does it mean when setting $\epsilon = 0$, $\epsilon = 0.15$ and $\epsilon = 1$?

Problem 2 (Special cases of HITS). Consider the HITS method for computing hub and authority scores, with scaling using $L_{\infty}$-norm.

(a) Discuss how the special case of a graph with only reciprocal edges (i.e., $(u,v) \in E \iff (v,u) \in E$) affects the computation of hub and authority vectors.

(b) Assume a graph consisting of $n$ vertices having the following cycle structure:

Compute the hub and authority vectors.

(c) Consider the graph in (b), but with the link between $n$ and 1 being removed. Compute the hub and authority vectors.

(d) Compute the hub and authority vectors for a graph with $n+2$ nodes where nodes that are connected by a simple lattice (“diamond”) where node 0 points to each of the nodes 1 through $n$ and each of the nodes 1 through $n$ points to node $n+1$. In addition, node $n+1$ points back to node 0. There is no link among the “middle” nodes 1 through $n$. The figure below depicts this kind of graph.
Problem 3 (Random Walks on Query-Click Graphs\cite{1}). Consider two queries $q_1, q_2$ and three documents $d_1, d_2, d_3$ with the number of clicks from the history given in the following table:

<table>
<thead>
<tr>
<th></th>
<th>$d_1$</th>
<th>$d_2$</th>
<th>$d_3$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$q_1$</td>
<td>2</td>
<td>2</td>
<td>4</td>
</tr>
<tr>
<td>$q_2$</td>
<td>3</td>
<td>2</td>
<td>1</td>
</tr>
</tbody>
</table>

Table 1: Click history

(a) Using random walks on query-click graphs, determine the transition matrix with self-transition rate $s = 0.1$.

(b) Apply forward random-walk with path length $t = 3$ and give the list of ranked documents for each query, along with their scores. You may use a suitable software for calculation.

(c) With the docs ranking from (b), comment on how the ranks have changed compared to the ranks we obtain if we rely only on the number of clicks. Provide a short explanation for this change.

(d) Apply backward random-walk with path length $t = 3$ and give the list of ranked documents for each query, along with their scores. You may use a suitable software for calculation.

Problem 4 (Optimizing Search Engine using Clickthrough\cite{2}). Consider a search query $q$ with 5 documents $d_1, d_2, d_3, d_4, d_5$ and their matching feature vectors: $\Phi(q, d_1) = \langle 3, 3 \rangle$; $\Phi(q, d_2) = \langle 0, 0 \rangle$; $\Phi(q, d_3) = \langle 6, 1 \rangle$; $\Phi(q, d_4) = \langle 2, 7 \rangle$; $\Phi(q, d_5) = \langle -1, 4 \rangle$.

Suppose the optimal ranking of these documents regarding query $q$ is (in decreasing order of relevance): $r^* = d_5, d_1, d_3, d_4, d_2$. However, the search engine returns the following ranking: $r = d_3, d_1, d_2, d_4, d_5$.

(a) Compute the Kendall rank correlation between the ranking of the returned documents and the ideal ranking.

(b) Now we want to optimize the search engine using the click history so that the ranking returned by the search engine is as close as possible to the ideal ranking. Suppose with the above returned ranking $r$, there is a single user clicking on two documents $d_1$ and $d_5$. Determine the weak ordering of documents by extracting preference feedback from this click data.

\footnotetext[1]{Craswell et al. SIGIR 2007}
\footnotetext[2]{Joachims, SIGKDD 2002}
(c) Consider the weak ordering extracted from (b) as training data. We perform SVM ranking to find the weight vector $w$. Which of the following values of $w$ can perfectly fit the training data: $w_1 = (2, 1)$; $w_2 = (0, 1)$; $w_3 = (-3, 1)$? Justify your answer.

(d) With the best value of $w$ from (b), find the new rank of documents $r_{new}$ returned by the search engine. Recompute the new Kendall rank correlation of $r_{new}$ with the ideal ranking $r^*$. Comment on the result by comparing it with the correlation calculated in (a).