Homework 9 due April 3

Select five problems from below or from Homework 8.

1. For an undirected graph if vertices u and v are connected by an edge, then we proved that $h_{uv} + h_{vu} \le 2m$. When u and v are not connected by an edge $h_{uv} + h_{vu}$ might equal n^3. Can you explain what is going on? Write a short paragraph giving your intuitive explanation.

2. Construct a small directed graph with ten to twelve vertices. Pretend that the graph represents the World Wide Web and that nodes correspond to web pages. Pick a node for your web page and see how much you can increase you page rank. Assume a restart value of 0.15. You can only change edges from your web page. What if you are allowed to add a new web page over which you have control of the edges out? How much did you actually change your ranking, i.e. how many pages did you move above?

3. Select a web graph obtained by a crawl such as one of those on the site http://law.dsi.unimi.it/index.php?option=com_include&Itemid=65 with approximately a million nodes. Do an experiment to see how much you can improve the page rank of some page. How many pages can you move your selected page up over?

4. Assume that a similarity matrix for term-document vectors is block diagonal where all entries outside the diagonal blocks are zero. Further assume that the entries in each diagonal block are all the same within the block. What is the criterion for the best fit line to select a given block? The criterion will depend on the size of the block and the relative size of entries in the block to that of other blocks.

5. In many experiments one collects the value of a parameter at various instances of time. Let y_i be the value of the parameter y at time x_i. Suppose we wish to construct the best linear approximation to the data in the sense that we wish to minimize the mean square error. Here error is measured vertically rather then perpendicular to the line. Develop formulas for m and b to minimize the mean square error of the points $\{(x_i, y_i) \mid 1 \le i \le n\}$ to the line $y = mx + b$.

6. Let A be a real valued matrix. Prove that $B = AA^T$ is positive semi definite.

7. Prove that the eigenvalues of a symmetric real valued matrix are real.

8. We often fit a straight line to data by minimizing the sum of squares of distances. We could minimize the absolute value of the distance. What are the pros and cons of each method?

9. Form a block matrix of probabilities where the probabilities in each block are identical but values differ from block to block. Convert the matrix to a 0-1 matrix by flipping a coin for each entry in the matrix with the appropriate probability of heads for that

element. Randomly permute the rows and use the same permutation for the columns. Remember the permutation. Next do a singular value decomposition of the matrix and see if you can recover the row for each block.

10. Find the best fit line to each of the following point sets.

(a) $\{(0,1),(2,0)\}$

(b) $\{(0,1),(1,0)\}$

(c) some other interesting point set.