Assignment: 4

March 20, 2012

Problem 0.1 *A popular pastime of Computer Scientists is to compute their Erdos number. Paul Erdos is a Hungarian mathematician known for this excellent contributions to several disciplines of mathematics including probability, combinatorics, graphs, and the like. To recognize his efforts, an Erdos number is defined as follows.*

The Erdos number of Erdos himself is 0. If person X co-authors a paper along with Erdos, then the Erdos number of X is 1. The Erdos number of a person Y is one more than the smallest Erdos number of any of his current co-authors. We say that A and B are co-authors if A and B authored a paper together at some point of time, possibly along with other co-authors. It is beleived that an Erdos number greater than four is nothing to be proud of, where as an Erdos number of two is a big acheivement.

In this problem you will be given an input that has the following form: the first line of input is the number of lines in the input. An input line is of the form $a, b, c, d,$

where it means that $a, b, c,$ *and* d *are co-authors of some paper. We will assume that* a, b, c *and* d *are positive integers. We will consider that vertex numbered 1 is the vertex corresponding to Erdos. The output will be a listing of the Erdos number of the vertices in the following format:* $a_1, a_2, \ldots, a_N$.

In the above, we arrange the vertices in increasing order and print their corresponding Erdos numbers. So, the Erdos number of vertex 1 is printed first, followed by the Erdos number of vertex 2, and so on.

Problem 0.2 *A graph* G *is said to be bipartite if its vertex set can be partitioned into two sets (disjoint)* V_1 *and* V_2 *such that the set of edges* E *is a subset of* $V_1 \times V_2$. *A necessary and sufficient condition for a graph* G *to be bipartite is that it has no cycles of odd length. Given an undirected graph* G,

(a) check whether it is bipartite or not.

(b) If it is, produce the bipartition also.

The input to the program will be a $V \times V$ *adjacencey matrix. The bipartition can beproduced in two lines of output in the sorted order of the vertex numbers separated by space.*

Problem 0.3 *One measure that is gaining popularity recently for researcher effectiveness is the h-index. Read about it on wikepedia. In this problem, we compute a slightly new mesaure as follows. Conisder a graph with two kinds of nodes: author nodes corresponding to authors, and paper nodes that correspond to papers. There are directed edges from an author node A to all the papers that are authored/co-authored by A. There are also directed edges between papers such that an edge from* P_1 *to* P_2 *indicates that paper* P_2 *cites paper* P_1.

Given this graph, for each author find his h-index defined as follows. An author is said to have a h- index of k *if exactly* k *of his papers have at least* k *citations and the remaining papers have at most* k *citations each. In counting the citations of a paper* P_i *you should discount citations* P_k *such that there exists a paper* P_j *where* P_j *cites* P_i*, and* P_k *cites both* P_j *and* P_i*. In case there is such a* P_k *with 2 or more such* P_j*'s, discount* P_k *only once overall. Under the above definition, find the h-index of all authors in the input.*

The input is given as a set of author-paper tuples followed by a set of paper tuples which correspond to the first element in the tuple is cited by the second element. The number of (author, paper) tuples is at the first line of the input. A sample input is as given:

4
a_1, p_1
a_2, p_2
a_3, p_3
a_2, p_3
p_2, p_3
p_3, p_1

Problem 0.4 *In an undirected graph G, a connected component is a maximal connected subgraph of G. Let us call the number of nodes present in each subgraph as a the size of the connected component. Given a graph, report the size of the largest connected component.*
The input graph G is given as an adjacency matrix.

Problem 0.5 *A directed acyclic graph is a directed graph with no directed cycles. Given a directed acyclic graph G and two vertices of the the graph, say u and v, and a positive number k, report all, or at most k, paths (whichever is smaller), the graph G between the two nodes u and v. The paths are to be reported in increasing order of length and in lexicographic order amongst paths of the same length. So, the path $1 \to 2 \to 4 \to 7$ should precede the path $1 \to 3 \to 5 \to 7$, and similarly these two paths appear after the path $1 \to 6 \to 7$.*

The input graph G is in form of an adjacency matrix.