Econ 524 Project 1

For our first several projects, we will work with data on historical outcomes of college football games and implement the Colley ranking method to rank teams in a given year. Because college football teams play only a small fraction of the total teams, a ranking is necessary to determine who plays in the championship game or who advances to the playoffs. A simple ranking by winning percentage would be misleading because there is wide variation in the strength of schedule. That is, some teams play a much more difficult schedule than others because of the conference they play in. Furthermore, since teams have some leeway to choose their out-of-conference games, teams would have incentives to schedule easy games for the out-of-conference schedule if the ranking does not control for strength of schedule. With that in mind, Colley developed a ranking method that does account for strength-of-schedule when ranking teams. The PDF file on PolyLearn provides the full details of the Colley method.

For project 1, you will implement the Colley matrix method to rank teams for the 2010 season. The data is available in the file ncaaf_2010.rdata. You can load this data frame using the command:

load("ncaaf_2010.rdata")

The data frame has 4 columns. The first is of type factor and has the 120 team names. The second is integer type and has the number of wins for that team. The third is integer type and has the number of losses for that team. The fourth is a list where each element of the list (i.e. each row of the column) is a vector of indices to the other teams played. For example, if the 1st row of this column was the vector c(4,8,2), it would mean the team in row 1 played 3 games, one against the team in row 4, one against the team in row 8, and one against the team in row 2.

You need to construct the Colley matrix (see section 6.1 of the PDF file). The Colley matrix is a square matrix with dimension equal to the number of teams. The diagonal entries (i,i) in the matrix are (2+num games played by team i). The off diagonal entries in row i (i,j) are set to -1*(# games played by team i against team j). So, each row should sum to 2.

You also need to construct a vector where the i-th element is:

1 + (team i num wins – team i num losses)/2

Then, solve the linear system Ax=b where A is the matrix and b is the vector. The solution is the score for each team where a higher score is better.

Section 6.3 shows an example with 2 teams and one game where the first team beat the second team in that game. There is also a 5 team example shown at the end of Section 6.3.

Create a new data frame with the name "solution" which has two columns. The first column should be the team name and the second column should be the score. Sort this data frame by score so that the best team (highest score) is in row 1. After I run your code, I should be able to view the solution data frame and see the ranking. Save all of your code in a file with extension ".r" and include your name in the file name. I should be able to source your code and it should work (i.e. it creates the solution data frame and does not generate any errors).