7 Markov Methods and PageRank

Here you create a network out of your elements. The standard is websites, but could be sports teams again. The idea is that, at first approximation, your importance is determined by how many sites point to you. But of course, the importance of the websites pointing to you is also important. In a sporting context, you could point to the teams that you lost to as sharing your vote for the best.

Then use what is called the Markov method. The idea is that one takes a random walk through the network. The rating of a site is the steady-state average time spent at that vertex.

Take an $n \times n$ matrix where entry $s_{ij}$ is 1 if team $i$ lost to team $j$ and 0 otherwise. We immediately have a problem if a team is undefeated, since that is an absorbing vertex in the random walk. So one fix is to make that row all 1’s. In any event, normalize each row so that it sums to 1; call the result $S$.

We want the stationary distribution. That is, $S^t \vec{r} = \vec{r}$. This is equivalent to the dominant eigenvector, and software can calculate that for us. In this case, there is a simple numerical idea called the power method. Start with all entries in $\vec{r}$ being $1/n$. Repeatedly calculate $S^t \vec{r}$ and normalize each time, until converges.

This is equivalent to the idea that the weight of a team’s vote should be based on how good they are. There are mathematical theorems about convergence et cetera.

For data set, get surprising results: Scaled as percentages.

FloridaSt 22.20
VirginiaTech 9.59
Duke 9.37
Clemson 9.19
GeorgiaTech 8.57
MiamiFL 7.55
BostonCollege 6.49
Pittsburgh 6.14
Maryland 5.18
Syracuse 4.91
North Carolina 4.83
Wake Forest 2.82
NC State 1.59
Virginia 1.59

Computational issues. For football, one can scale the links by the margin of victory.

PageRank and adjustments are discussed at website whydomath.org

References

Wikipedia