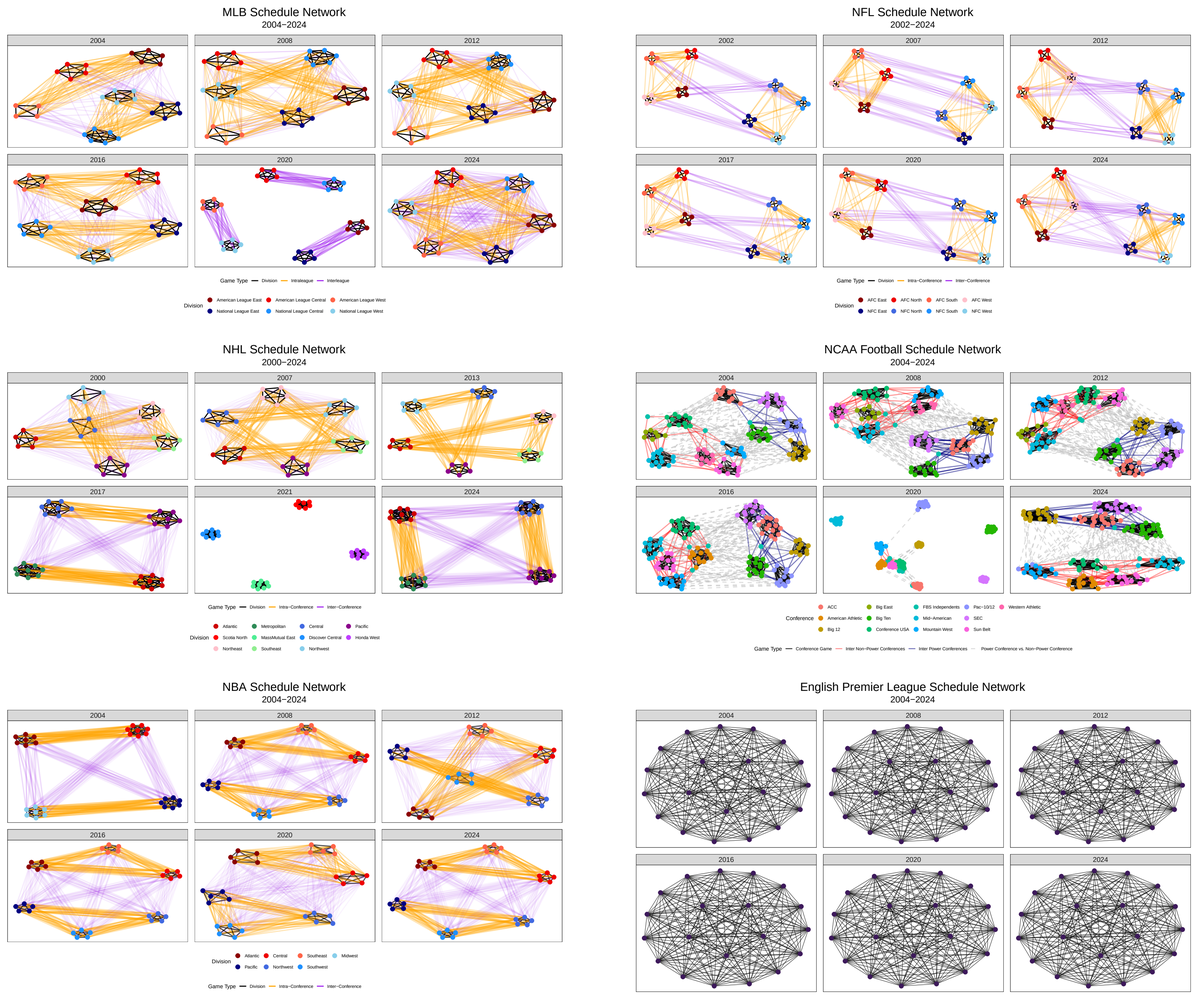
Schedule Path Ratios: A Graph-Based Approach to Quantify Schedule Balance

Luke Benz

[1]Department of Biostatistics, Harvard T.H. Chan School of Public Health, Boston, MA

Introduction: Schedule Heterogeneity Across Sport & Time



Schedules across college and professional sports display substantial heterogeneity. Moreover, scheduling patterns have changed over time, as leagues have undergone rule changes, external factors (e.g. lockouts, COVID-19) have forced structural changes, and incentives have changed (NIL, Collge Football Playoff).

In this work, we introduce a novel metric to quantify schedule balance in a manner that is agnostic to team strength. This metric, which we title schedule path ratios, compares graph-based distances in the observed schedule network to those in a network which represents an optimally balanced schedule.

We are particularly interested in how quantification of the structural balance of schedules influences identification of the best team, a task particularly salient in selection of the expanded College Football Playoff (CFP) field.

Schedule Path Ratios

Goals for a Balanced Schedule

1. Same number of games between each pair of teams
2. Consider direct comparison between teams more informative than an indirect comparison

$S = (S_{\text{teams}}, S_{\text{games}})$ denotes a schedule of N games between n teams. Schedule S induces a a graph $G = (V, E)$ where $V = S_{\text{teams}}$, $E = \{(i,j) \mid (i,j) \in S_{\text{games}}\}$ with weights $w_{ij} = \frac{1/n_{ij}}{\sum_{i<j} 1/n_{ij}} N$, where n_{ij} denotes the number of games played between teams i and j.

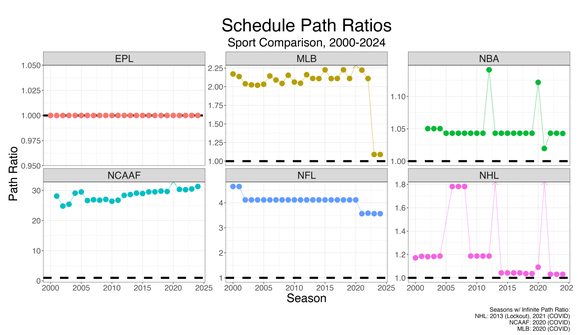
Given the goals for a balanced schedule above, we define an optimally balanced graph, $G_{\text{opt}} = (V_{\text{opt}}, E_{\text{opt}})$ with $V_{\text{opt}} = S_{\text{teams}}$ and $E_{\text{opt}} = \{(i,j) \mid i,j \in S_{\text{teams}}, i < j\}$ with weights $w^*_{ij} = \frac{2N}{n(n-1)}$. Note this is just the graph which has edges connecting each team in the league, and spreads the weight of N games played across all possible edges, of which there are $\frac{n(n-1)}{2}$.

Finally, define the distance between two teams i and j

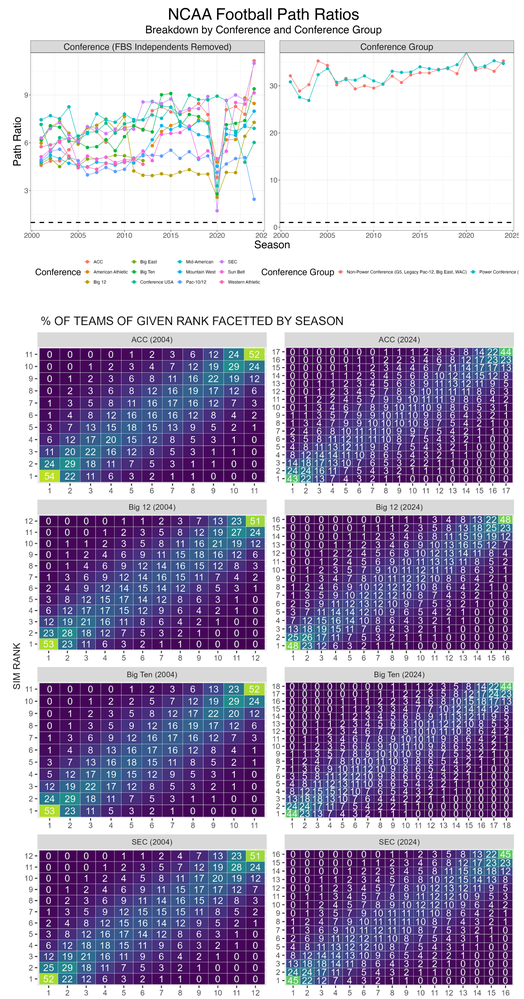
$$d_{ij}(G) = \begin{cases} w_{ij} & (i,j) \in E \\ P_{ij} & (i,j) \notin E \end{cases}$$

where P_{ij} is the length of the shortest path between i and j. With all this set-up, we finally define our metric of interest, the schedule path ratio $R(G, G_{\text{opt}})$.

$$R(G, G_{\text{opt}}) = \sqrt{\frac{\sum_{i<j \in S_{\text{teams}}} d_{ij}(G)^2}{\sum_{i<j \in S_{\text{teams}}} d_{ij}(G_{\text{opt}})^2}}$$



Case Study: College Football



Simulating 2004 and 2024 Power 4/5 conference schedules using team strengths drawn from distribution of model-based estimates for P4/5 teams shows that better teams are more likely to win their conference under 2004 schedule structure than 2024.

Acknowledgments

I would like to thank Thompson Bliss and Michael Lopez for their helpful feedback and discussion.