The Sports League Formation Problem: Case of the Washington Area Girls Soccer League

Robert Axtell[1] and Ovi Chris Rouly
Department of Computational Social Science and Center for Social Complexity
Krasnow Institute for Advanced Study
George Mason University
Fairfax, VA 22030 USA
{rax222, rouly}@gmu.edu

Abstract. Consider *n* athletic teams competing pairwise (head-to-head) in some prescribed sequence (schedule) such that the total number of games is less than $n (n - 1) m / 2$ necessary for each team to play every other *m* times. The goal of each team is to be ranked as highly as possible at the conclusion of the sequence. Because teams have heterogeneous talent and skill the best ones will presumably rise to the top of the rankings while the weaker ones will populate the bottom of the list, despite intrinsic stochasticity in game outcome. The final rank order may or may not closely resemble the 'true' underlying distribution of talent and skill, depending on the mechanism used to form the sequence of team pairings. In this paper we analyze a variety of such scheduling mechanisms and investigate how well they do both on average and in worst-case scenarios, when the actual distribution of ability is known. In essence, we compare teams' actual performance records under some mechanism, as realized on the field, with their 'true' ranking, as well as with the ranking they would expect to receive if they played each other head-to-head *m* times. A case study involving a youth soccer league serves as motivation.

Keywords: Sports league formation, football, soccer, adaptive scheduling, Washington Area Girls Soccer League (WAGS)

1 Introduction to the Problem

By the 'sports league formation problem' we mean simply this: how to efficiently arrange a fixed number of athletic teams, *n*, of unknown quality into a specified number of divisions, *d* (< *n*), such that the best teams are in higher divisions and teams of similar quality are in the same divisions. Contests between pairs of teams are used by league officials to assess the quality of teams. Specific kinds of contest sequences are called *mechanisms*, as in 'such-and-such a league uses a particular mechanism to rank its teams and form divisions.' The main questions we attempt to shed light on include 'What kinds of mechanisms yield "good" results, in the sense

[1] Participants in GMU CSS 610 contributed mightily to this paper by virtue of their class participation. For helpful conversations on WAGS we thank Gilberto Vera and Rich Shelton, along with several people at WAGS.

that the final ordering of teams closely reflects their "true" quality?'; 'What are the properties of "good" mechanisms that are primarily responsible for producing few type I and type II errors (excluding teams that should be included and including teams that should excluded, respectively)?'; 'What are the typical worst case results for specific classes of mechanisms?'; and 'What classes of mechanisms yield play that can be arranged into preliminary divisions, and do such divisions adversely affect mechanism efficiency? We shall address the problem in general but also with respect to particular mechanisms that have found widespread use.

Indeed, this problem has many real-world instantiations. For instance, imagine a youth basketball league within a community consisting of players having heterogeneous skills, and where the (amateur) coaches meet before the season to select teams according to a draft. The combination of coaches having disparate coaching skills and incomplete knowledge of player abilities, together with players of uneven talent means the resulting teams will, with high probability, have a wide range of performances on the basketball court. The job of the league organizer is to schedule league play to avoid match-ups between extremely strong and weak teams, but to do so without any *a priori* knowledge about which teams will be strong and which weak, although he or she may have reasonable priors. Specifically, the organizer wants to break the teams up into relatively equal divisions and then have 'regular season' games between teams within divisions, perhaps giving out awards at the end of the season to the teams who win their divisions.

More generally, consider the process of selecting a champion in any season of any sport. Usually there is some 'regular season' play followed by playoffs. At the beginning of the season the competitors are of unknown quality and it is hoped by league officials that the process of playing the regular season and then the playoffs properly awards those with superior skill and performance. In the language used above, the regular season/playoffs schedule together with the rules for determining who makes the playoffs constitute the *mechanism* whose properties we wish to explore. So in a certain sense what we have termed the 'sports league formation problem' is the general case of the more particular 'sports league champion determination problem'. We shall return to this general interpretation in section V below, but first we discuss a particular case in detail.

2 Instance of the Problem: Washington Area Girls Soccer League

An example of the 'sports league formation problem' occurs during the 'under 12' (U12) year of soccer in the Washington Area Girls Soccer League (WAGS), the premier soccer league for female youth in the Washington DC metro area (including Northern Virginia and Southern Maryland). WAGS runs divisions for players in the under 9 (U9) age group through under 19 (U19).[2] The U9 through U11 ages do not compete in quality-ranked divisions, but are arranged into geographically-related divisions in order to reduce the burden of travel to games. In U13 and older divisions, there are at most five divisions of roughly ten teams each—by its charter WAGS can

[2] For many years WAGS admitted teams only at the U11 age group, but for the 2006-7 season U9s and U10s were also permitted.

have at most fifty teams in an age group—arranged in order of soccer prowess: Division 1 teams are the strongest, Division 2 teams next best, and so on. Division 5 teams are the weakest and failure of a team to do well in Division 5 may lead to its relegation from the league. There are waiting lists to enter WAGS beyond U12.

It is at the U12 year—typically 6th graders—that WAGS invokes a particular process of 'league formation' of the type we study here. Some number of teams in excess of the fifty that will survive into the U13 season compete for berths in Divisions 1-5.

During the 2006-7 season some seventy-two teams competed for these fifty slots. The procedure used to accomplish this was relatively simple and works as follows:

- In the fall season, divisions were made up having relatively similar *cross-sections* of strong and weak teams. For example in 2006-7 there were 7 divisions in the fall of approximately 10 teams each, with each division consisting of 1-2 of the top teams, 2-3 additional well-ranked teams, and the remainder weaker teams. The fall season is then held with each team playing every other in its division once, and final standings tallied on the basis of a conventional formula in soccer (3 points for each win and 1 point for each tie, with no points for losses).
- Then, for the spring season, divisions are formed on the basis of team point totals. The top forty teams are placed into Division 1-4, with the 10 best teams in the first division, the next 10 best in the second division, and so on. The weakest teams on the basis of their fall performance—there were 30 such teams in 2006-7—are placed into equal-sized divisions labeled '5A', '5B', '5C' and so on, consisting of from 8-10 teams, depending on how many teams there are in total. The performance of the Division 1-4 teams in the spring season serves to reposition them for the U13 season, with some teams moving up a division and others moving down. (Teams very rarely move more than one division between seasons.) Such rearrangements can serve to 'correct' any imperfections put into the top 40 rankings in the fall. The Division 5 teams are playing with somewhat different motivation in the spring, as they are competing to stay in WAGS. The top few teams in each division 5 variant are the ones who will stay in WAGS at U13, and the remainder will be relegated. This works by ranking the spring point total of each team playing in a Division 5 against all Division 5x teams, with the top teams invited to return, their exact number adjusted from year to year to limit the total number of teams in an age group to fifty. The top division 5 team might be promoted to Division 4 for U13, especially if there is one or more weak Division 4 teams who are more appropriate for Division 5.

This procedure had been employed by WAGS for several years, with only slight variations. From an unscientific sample of coaches and managers I believe there was modest support for this system, although some complained about it. People seemed to believe that it worked pretty well although far from perfectly. Usually the strongest teams ended up in the highest divisions and the weakest teams were relegated, but in each season ostensibly relatively strong teams (from their tournament performances, for example) did not do well in the fall and/or either end up in a much lower division than expected or fall out of WAGS altogether. Alternatively, the opposite result—a relatively weak team that does very well, seemingly too well—was also observed. What is clear is that type I and II errors are to be expected, but with unknown frequency. Does this second league formation/team selection process lead to better results, either in practice or in principle? Do there exist mechanisms that outperform each of these mechanisms? Are mechanisms with good performance also robust to noise, the need to reschedule games, and other practical problems?

In this paper we try to answer some of these questions both for the WAGS league and in general. After analyzing the current and previous WAGS mechanisms, we point out some weaknesses and then go on to propose alternatives that apparently have the potential to outperform the team selection mechanism currently used by WAGS. We eventually discuss the problem in general and compare and contrast disparate mechanisms that are in use today across the sporting world and in related domains involving competition, including politics, biology, and even computer science, where biological selection is increasingly used as an efficient way to 'discover' high performance (*1*).

3 Analysis of the Performance of the WAGS U12 Mechanisms

To study the properties of the mechanisms employed by WAGS we must first develop a methodology to simulate the outcome of soccer matches. We need to have some measure of the distribution of skills across teams, and then we have to figure out how to express the chances of a team winning (or playing to a tie) as a function of its skill difference with its opponent. It turns out this can be done for WAGS data. With these relationships in hand it is then possible to build microsimulations of the WAGS mechanisms, which we use to assess the properties of these mechanisms.

3.1 Distribution of Team Skill

To establish the distribution of skills across the population of teams, we use the total points earned in the fall season. This assumes that the average level of competition is about the same across divisions. This is a not unreasonable assumption for the 2006-7 season. Figure 1 is a plot of the number of points earned in the fall during the 2006-7 season, with each division a different color.

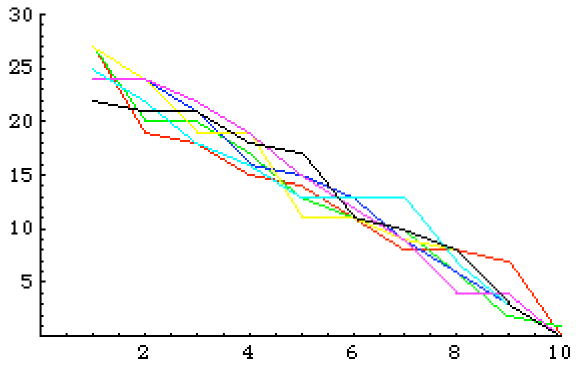

Fig. 1. Points earned during the U12 2006-7 fall season, by WAGS division

There is an approximately linear decline in the number of points accumulated by teams. However, this plot masks substantial differences between teams insofar as the best teams did not play one another and therefore cannot be separated. Indeed, the linear decline with rank is more a statement of how points are awarded than a measure of team skill. To get at skill, we plot in figure 2 the distribution of goal differentials (i.e., goals scored less goals given up) during the fall 2006-7 season.

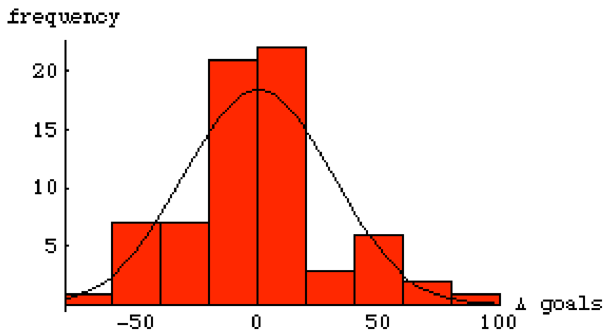

Fig. 2. Distribution of goal differential in the U12 2006-7 fall season, all divisions

While an imperfect measure of skill, goal differential does have the property that it is highly correlated with winning percentage and other measures of success. Superimposed on the empirical data is a normal distribution having the same mean and variance. This distribution does a reasonable job describing the data, so we treat team skill as normally distributed in what follows.

3.2 Representing Match Outcomes

When two teams meet, which will win? This is of course a difficult question, for if there were a completely satisfactory way to determine a winner without playing the game then there would be little point to the match. We approach the problem by first computing the overall probabilities of ties, and then disaggregating by the differences in goal differential of the teams involved. Doing so for WAGS data gives figure 3.

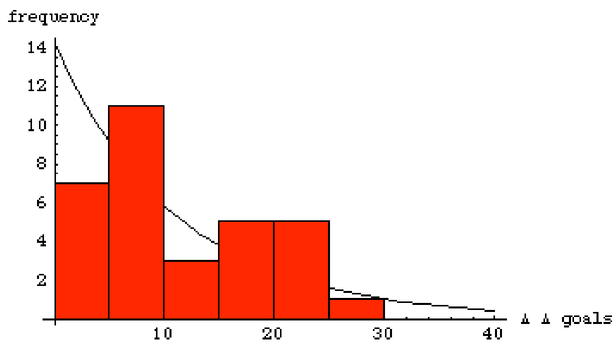

Fig. 3. Probability of a tie as a function of the difference in goal differential

For our purposes we could use the actual data in our model, but believe there are good reasons why an exponential function well-describes the underlying relationship: the highest probability of a tie should be for similar teams and unlikely between teams having disparate goal differential statistics.

Next we investigate the probability of a win by the team with the higher goal differential. These data are shown in figure 4 for the fall 2006-7 WAGS U12 season.

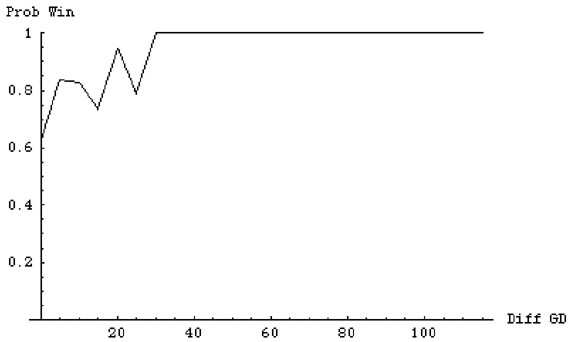

Fig. 4. Probability of winning as a function of the difference in goal differential

Clearly, the greater the difference in goal differential the larger the chance of winning by the team with the higher GD.

3.3 A U12 WAGS Season Simulated

We use these functions to simulate an entire WAGS fall season, the rearrangement into spring divisions, and then the outcome of the spring season. We do this for a specified distribution of team skills. We are the gods of our artificial soccer world and, unlike real life league organizers, know the true rank of the various teams.

For WAGS we find is that each set of fall and spring seasons produces type I and II errors: some strong teams are undeservingly relegated while weak teams who should be relegated do not get sent out. We can represent this as in figure 5, a plot of the true rank of a team (dark line) with the realized rank (in red) superimposed.

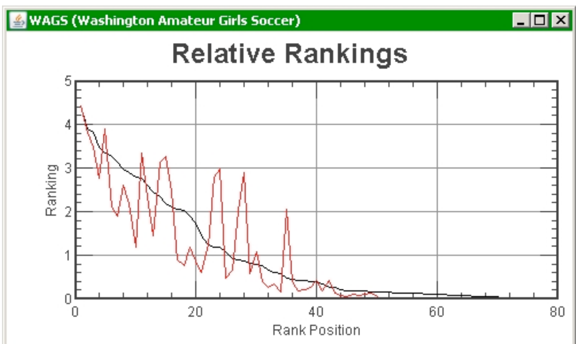

Fig. 5. Position of a team after a two seasons (red line) vs underlying rank

Clearly, a single year of U12 WAGS soccer does an imperfect job ranking the teams. A different way to look at this is to make many realizations of the model and study how many times the i^{th} best teams get relegated for all i. This is done in figure 6.

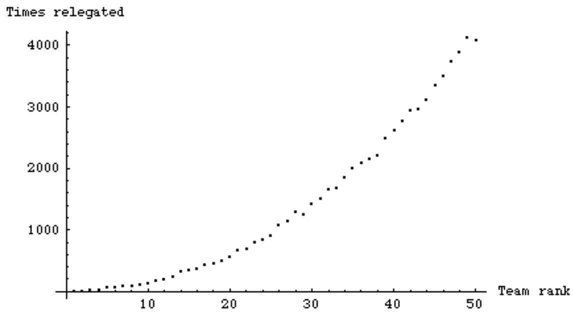

Fig. 6. Number of times a team of rank *i* is relegated over 10K realizations

We imagine that even the best mechanisms will have some error. The sports league design problem is really about finding a mechanism that will have small error.

4 Alternative Mechanisms and Some of Their Properties

We propose the following mechanism to reduce the number of teams that are relegated for no reason other than bad luck in the extant system. Instead of two nine game seasons, we propose three six game seasons, with the first two being used to sort the teams and the final one a relegation round, as at present. Figure 7 demonstrates that such a mechanism has superior performance insofar as it results in fewer good teams getting relegated and, by conservation, fewer poor teams given berths in the U13 league.

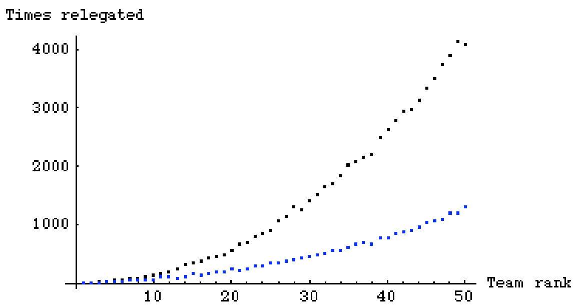

Fig. 7. Number of times a team of rank *i* is relegated over 10K realizations, previous league structure (black line) and improved performance with three seasons (blue line)

However, it turns out that it is possible to reduce these errors even further if one is willing to dispense with divisions altogether. Figure 8 gives a typical result from an ‘adaptive’ scheduler in which closely matched teams play each other repeatedly and ‘upsets’ are replayed to determine whether or not they are really aberrations.

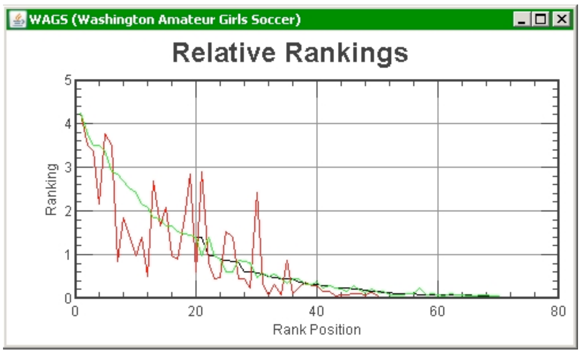

Fig. 8. Adaptive scheduling (green line) and true rank (black line)

Clearly, this mechanism goes a long way toward fixing the problem (green line).

We have not exhaustively studied all possible scheduling mechanisms, although more comprehensive work than that described above is planned. The main point of the above has been to show that it is possible to improve upon the recent WAGS practice, and that a model-based approach gives concrete, intelligible results.

5 Summary and Conclusions

We have described the ‘sports league formation problem’ in some detail, and studied a particular instantiation of it at some length. We have demonstrated imperfections of the current systems and proposed new scheduling mechanisms, both division based and divisionless, that would end up relegating fewer deserving teams.

However, much more work needs to be done. We need to have a more thorough representation of the statistical properties of winning and losing. Some further disaggregation of these data into home and away game venues would probably prove revealing. A richer specification of several model features would be desirable. We need to make larger numbers of realizations and study the robustness of alternative scheduling mechanisms to uncertainty, such as when only imperfect knowledge of underlying skill mechanisms is available, as in the real-world. A more thorough search of the mechanism space needs to be undertaken. Indeed, the present work has only scratched the surface of these problems, and hopefully demonstrated the fertility of the problem and a computational approach.

References

1. K. A. De Jong, *Evolutionary Computationl* (MIT Press, Cambridge: Mass., 2006), pp.