INCOMPLETE PAIRWISE COMPARISON

R.S.MACKAY

The English Premier League in football was interrupted by the coronavirus on 10 March. By the time this article is published it might well have restarted, but a valid question is how to extrapolate to an end result from what was already played.

Each team is supposed to play each other team twice: once at home and once away. Officially, the outcome of the season is decided on the basis of the sum of points, with 3 for a win, 1 for draw and 0 for a lose. If there is a tie it is decided on goal difference. If there is still a tie it is decided by goals for. If there is still a tie and it matters for championship or relegation then there are some subsidiary rules ending with a possible playoff.

I shall take a different criterion, namely to infer team strengths from goal differences. If each team played each other team twice as intended, then one could just add up the goal differences for a team's matches (given team's score minus opponent's score) and that would give a measure of its strength. But with the season incomplete, this would not be a fair reflection: some played different numbers of matches from others, some had potentially easier fixtures left, others harder ones.

Instead, we can infer strengths S_m for each team m by postulating a model of the form

$$G_{mn} = S_m - S_n + \varepsilon_{mn} \tag{1}$$

for the goal difference of a match where m plays n. Here S_m is a notional strength for m and ε_{mn} reflects a random deviation from the difference in strengths (and also rounding to produce an integer). To distinguish between the two matches where m plays home or away, we'll write the home team first. Let's introduce a matrix W_{mn} to indicate which matches were played, with $W_{mn} = 1$ if m played at home against n, 0 otherwise. Then we can determine the strengths by minimising the sum of squares

$$\sum_{mn} W_{mn}(G_{mn} - S_m + S_n)^2$$

over S (without loss of generality we put $G_{mn} = 0$ if the match was not played).

Least squares minimisation leads to the linear system:

$$\Lambda S = U, \tag{2}$$

with the "graph Laplacian"

$$\Lambda_{mn} = D_{mn} - W_{mn} - W_{nm},$$

Date: May 28, 2020.

where D is diagonal with elements $D_{mm} = 2\sum_n W_{mn}$, and the imbalance vector

$$U_m = \sum_n G_{mn} - G_{nm}.$$

There is an obvious degeneracy for (2): if S is one solution and c is a constant then $S + c\mathbf{1}$ is also a solution, where $\mathbf{1}$ is the vector of all 1s. But we could apply a shift to enforce the average of S_m to be zero. Let $\mathcal{G}$ be the undirected graph with an edge from m to n iff they played a match. A nice theorem is that if $\mathcal{G}$ is connected then there is a solution S of (2) and it is unique up to a shift. To compute it, one can replace any of the rows of the system (2) by $\sum_m S_m = 0$. Alternatively, to keep symmetry, one can augment Λ by a row of 1s and a column of 1s apart from a new diagonal element of 0, let's say we put them in row and column 0, and augment U by a new element of 0, and solve

$$\begin{bmatrix} 0 & \mathbf{1}^{\mathbf{T}} \\ \mathbf{1} & \Lambda \end{bmatrix} \begin{bmatrix} S_0 \\ S \end{bmatrix} = \begin{bmatrix} 0 \\ U \end{bmatrix}$$

for the augmented vector $\begin{bmatrix} S_0 \\ S \end{bmatrix}$. Its new element S_0 comes out to be 0 because $\sum_m U_m = 0$.

Note that if all matches had been played then $S = U/(4N-2)$ is a solution, where N is the number of teams. This corresponds to dividing the total goal difference for each team by the number of matches played. But the advantage of the method is that we can apply it to an incomplete season, as long as enough matches have been played to make the graph $\mathcal{G}$ connected.

Let us apply this to the Premier League as it stood on 9 March. The table is given in Figure 1. The results of our method are shown in the first column after the table.

Home \ Away	ARS	AVL	BOU	BHA	BUR	CHE	CRY	EVE	LEI	LIV	MCI	MUN	NEW	NOR	SHU	SOU	TOT	WAT	WHU	WOL
Arsenal	–	3–2	1–0	1–2	2–1	1–2	2–2	3–2			0–3	2–0	4–0		1–1	2–2	2–2		1–0	1–1
Aston Villa		–	1–2	2–1	2–2			2–0	1–4	1–2	1–6		2–0	1–0		1–3	2–3	2–1	0–0	
Bournemouth	1–1	2–1	–	3–1	0–1	2–2		3–1		0–3	1–3	1–0		0–0	1–1			0–3	2–2	1–2
Brighton & Hove Albion		1–1	2–0	–	1–1	1–1	0–1	3–2	0–2					2–0	0–1	0–2	3–0	1–1	1–1	2–2
Burnley	0–0	1–2	3–0		–	2–4	0–2	1–0	2–1	0–3	1–4	0–2	1–0	2–0		3–0	1–1		3–0	
Chelsea	2–2	2–1	0–1	2–0	3–0	–	2–0	4–0	1–1	1–2		0–2	1–0		2–2	0–2	2–1		0–1	
Crystal Palace	1–1	1–0	1–0	1–1			–	0–0	0–2	1–2	0–2		1–0	2–0	0–1	0–2		1–0	2–1	1–1
Everton	0–0			1–0	1–0	3–1	3–1	–		a	1–3	1–1	2–2	0–2	0–2		1–1	1–0	2–0	3–2
Leicester City	2–0	4–0	3–1		2–1	2–2		2–1	–	0–4	0–1		5–0	1–1		1–2	2–1	2–0	4–1	0–0
Liverpool	3–1		2–1	2–1				5–2	2–1	–	3–1	2–0	3–1	4–1	2–0	4–0	2–1	2–0	3–2	1–0
Manchester City		3–0		4–0		2–1	2–2	2–1	3–1		–	1–2			2–0	2–1	2–2	8–0	2–0	0–2
Manchester United	1–1	2–2		3–1	0–2	4–0	1–2	1–1	1–0	1–1	2–0	–	4–1	4–0			2–1	3–0		0–0
Newcastle United	0–1		2–1	0–0	0–0	1–0	1–0	1–2	0–3		2–2	1–0	–	0–0		2–1		1–1		1–1
Norwich City	2–2	1–5	1–0			2–3	1–1		1–0	0–1	3–2	1–3	3–1	–	1–2		2–2	0–2		1–2
Sheffield United	1–0	2–0	2–1	1–1	3–0		1–0		1–2	0–1	0–1	3–3	0–2	1–0	–	0–1		1–1	1–0	
Southampton		2–0	1–3		1–2	1–4	1–1	1–2	0–9	1–2		1–1	0–1	2–1		–	1–0	2–1	0–1	2–3
Tottenham Hotspur	a	3–1	3–2	2–1	5–0	0–2	4–0			0–1	2–0		0–1	2–1	1–1	2–1	–	1–1		2–3
Watford	2–2	3–0	0–0	0–3	0–3	1–2	0–0	2–3		3–0		2–0			0–0		0–0	–	1–3	2–1
West Ham United	1–3		4–0	3–3			1–2	1–1	1–2	0–2	0–5	2–0	2–3	2–0	1–1	3–1	2–3		–	
Wolverhampton Wanderers		2–1		0–0	1–1	2–5			0–0	1–2	3–2	1–1	1–1	3–0	1–1	1–1	1–2	2–0	2–0	–

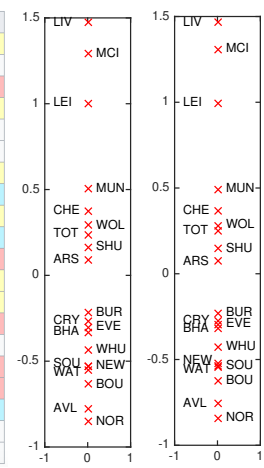

FIGURE 1. Premier league results 2019-20 from Wikipedia, inferred heights, without and with inferred home advantage 0.274.

Every football fan knows, however, that there is a home advantage, meaning that we should modify the model from (1) to

$$G_{mn} = S_m - S_n + b + \varepsilon_{mn}$$

with the home advantage b to be determined as well as the strength vector S. Minimising the sum of squares again yields a slight modification of (2) which determines b and S up to a shift. The modified system is

$$\begin{bmatrix} \Lambda & h \\ h^T & w \end{bmatrix} \begin{bmatrix} S \\ b \end{bmatrix} = \begin{bmatrix} U \\ g \end{bmatrix} \tag{3}$$

where h is the vector showing the number of home matches played by each team minus the number of away matches, w is the total number of matches played, and g is the total goal difference (home minus away goals). Again, one can take care of the shift by enforcing $\sum_m S_m = 0$ in either of the two ways above.

Carrying out the inference of both strengths and home advantage for this year's Premier League produced the last column in Figure 1. The inferred home advantage was 0.274, but it can be seen that it makes hardly any difference to the inferred strengths. Presumably this is because each team had played roughly the same numbers of home and away matches.

We can do the same for the Football Association Women's Super League. It was decided on 25 May not to restart the season, though the mechanism for the outcome was not announced. So here is what my method proposes. Figure 2 shows the table as it stood when the season was suspended, the inferred strengths of the teams without including a home advantage, and the strengths with an inferred home advantage, which this time comes out to be 0.380.

Home \ Away	ARS	BIR	BHA	BRI	CHE	EVE	LIV	MCI	MNU	REA	TOT	WHU
Arsenal	—	2–0	4–0	11–1	1–4	16 May	1–0	1–0	P–P	P–P	13 May	2–1
Birmingham City	P–P	—	P–P	0–1	0–6	0–1	2–0	0–2	P–P	6 May	1–1	P–P
Brighton & Hove Albion	0–4	3–0	—	P–P	1–1	1–0	1–0	16 May	1–1	2–2	0–1	1–3
Bristol City	P–P	0–2	0–0	—	0–4	P–P	0–1	0–5	P–P	P–P	1–2	P–P
Chelsea	2–1	2–0	P–P	6–1	—	P–P	16 May	2–1	1–0	3–1	1–0	8–0
Everton	1–3	P–P	2–0	2–0	3 May	—	P–P	0–1	2–3	3–1	3–1	P–P
Liverpool	2–3	P–P	P–P	1–1	1–1	0–1	—	P–P	P–P	0–1	P–P	1–1
Manchester City	2–1	3–0	5–0	1–0	3–3	3–1	1–0	—	1–0	P–P	P–P	5–0
Manchester United	0–1	16 May	4–0	0–1	12 May	3–1	2–0	P–P	—	2–0	3–0	P–P
Reading	0–3	1–0	P–P	3–3	P–P	3–2	P–P	0–2	1–1	—	3–1	2–0
Tottenham Hotspur	0–2	P–P	1–0	P–P	P–P	2–2	1–0	1–4	0–3	16 May	—	2–1
West Ham United	P–P	1–0	2–1	16 May	1–3	P–P	4–2	P–P	3–2	2–3	0–2	—

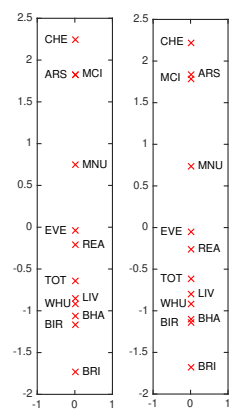

FIGURE 2. The Womens' Super League table as it stood, and inferred strengths without and with a home advantage of 0.380.

One might ask what justification the method has. If we imagine that the goal difference could be any real number, rather than integer, and that the deviations from strength difference are independent Gaussians with mean zero and common variance, then least

squares fitting corresponds precisely to Bayesian inference of the strengths with flat prior on the strengths. This result extends to the case with home advantage if we take a flat prior for it too. One could take a Gaussian prior for the home advantage and deduce a slightly modified linear system for the strengths and home advantage, but it did not seem particularly worthwhile to me. In reality the goal differences are integers and a more sophisticated probabilistic model for how the strengths produce goal differences (or even the individual goals for and against) is required, and then Bayesian inference would produce a different answer. But our method has the advantage of simplicity and relative ease of understanding.

I invite the reader to compare the final outcomes of the Premier League and the Women's Super League with the above inferred strengths.

What about the Scottish Premiership? There the rules are different. In round 1, each team plays each other team twice: once home and once away. In round 2, the teams are divided into two equal sized groups and each team plays each other in the same group twice: once home and once away. The Scottish Premiership was interrupted when Round 1 was not quite complete and Round 2 had already started.

An easy modification to the method takes care of this. We let W_{mn} be the number of matches played by m at home against n, and U_m be the sum of the goal differences over all matches played by m. Then (2) gives inferred strengths for the teams. Again, we can allow for a home advantage. It comes out to be 0.386, but again makes little difference to the inferred strengths. The results are shown in Figure 3.

Home \ Away	ABE	CEL	HAM	HOM	HIB	KIL	LIV	MOT	RAN	ROS	STJ	STM
Aberdeen	–	0–4	1–0	3–2	1–1	3–0	2–1	0–1	2–2	3–0	1–1	2–1
Celtic	2–1	–	2–1	3–1	2–0	3–1	4–0	2–0	1–2	6–0	7–0	2–0
Hamilton Academical	0–1	0–1	–	2–1	1–1	2–0	2–1	1–3	1–3	2–2	0–1	0–1
Heart of Midlothian	1–1	0–2	2–2	–	0–2	0–1	1–1	2–3	1–1	0–0	0–1	5–2
Hibernian	3–0	1–1	2–1	1–2	–	2–2	2–2	3–1	0–3	2–2	2–2	1–0
Kilmarnock	0–0	1–3	2–2	3–0	2–0	–	2–1	0–1	1–2	0–0	0–0	1–0
Livingston	0–2	2–0	0–0	0–0	2–0	3–0	–	0–0	0–2	4–0	1–0	2–1
Motherwell	0–3	2–5	1–2	1–0	3–0	2–1	2–1	–	0–2	1–2	4–0	2–0
Rangers	5–0	0–2	5–0	5–0	6–1	1–0	3–1	2–1	–	2–0		1–0
Ross County	1–3	1–4	3–0	0–0	2–1	1–0	1–4	1–2	0–4	–	2–2	2–1
St Johnstone	1–1	0–3	3–2	1–0	1–4	0–1	2–2	0–1	0–4	1–1	–	0–0
St Mirren	1–0	1–2	0–0	0–0	1–2	1–0	3–3	0–3	0–1	2–1	2–0	–

Home \ Away	ABE	CEL	HAM	HOM	HIB	KIL	LIV	MOT	RAN	ROS	STJ	STM
Aberdeen	–	1–2	–		3–1	–	–	–	–	1–2	0–1	–
Celtic	–	–	–	5–0	–	3–1	–	–	–	3–0		5–0
Hamilton Academical	1–3	1–4	–	–		1–0	2–4	0–0	–	–	–	–
Heart of Midlothian	–	–	2–2	–	–	2–3	–	1–1	2–1		–	–
Hibernian	–		–	1–3	–	–	1–1	–	–	3–0		2–2
Kilmarnock	2–2	–	–	–	1–2	–		–	2–1	3–1	–	
Livingston		2–2	–		–	–	–	1–0	–	–	–	2–1
Motherwell		0–4	–	–	0–0		–	–	–	4–1	–	1–2
Rangers	0–0	a	0–1	–	2–1	–	1–0		–	–	–	–
Ross County	–	–		–	–	–	2–0	–	0–1	–	1–1	
St Johnstone	–	–		3–3	–	2–1	1–0	2–1	2–2	–	–	–
St Mirren	0–0	–	1–1	1–0	–	–	–	–		–	0–0	–

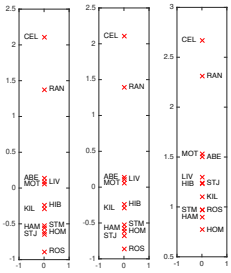

FIGURE 3. Scottish Premiership results 2019-20 from Wikipedia, inferred heights, without and with inferred home advantage 0.386, and points per game.

It was decided on 18 May to abandon the Scottish Premiership season, and settle the outcome on the basis of average points per game. The points per game are represented in the final column. One sees that there are some substantial differences: HOM is at the bottom instead of ROS, and STJ comes much higher up. This shows a significant difference between inferring strengths from goal differences versus points per game.

The big defect of the method, as we have seen with the Scottish Premiership, is that the real football leagues are decided on points, not goal differences. That could be addressed by formulating a probabilistic model of points as a function of strengths and doing Bayesian inference. Indeed there is a big literature on methods to infer rankings from pairwise

comparisons with just win/lose outcomes. Many of the methods require a complete set of pairwise comparisons, but some do not, notably the classic [BT]. A more recent reference is [SW], which promotes the Borda count, being the number of matches won by a given team, as a fair measure of strength, regardless of whether there is a complete set of comparisons. To incorporate the possibility of draws in the Borda count, one could consider a draw to be equivalent to a non-match. But neither of these address the issue that a draw in the football leagues gives fewer points than the average of a win and a loss.

The issue of inferring strengths from a set of pairwise comparisons comes up in a wide range of other contexts besides sports. For example, a market survey might produce an incomplete set of pairwise comparisons of products by a selection of consumers. A ballot might produce an incomplete set of rankings of candidates by constituents, which can be considered as providing pairwise comparisons, either by a win/lose for which candidate was ranked higher by a constituent, or by counting the difference of the ranking and using it like goal difference. Indeed, the Borda count is named after Jean-Charles de Borda who developed it in 1770 to decide on elections to the French Academy of Sciences, historians trace its use back to the Roman Senate, and variants are used for various elections today. But I am particularly keen to use available quantitative information rather than just win/lose, which is why I promote my method.

The method is an outgrowth of one for inferring heights of nodes in a directed graph, that we call trophic analysis [MJS], in a project supported by the Economic and Social Research Council under grant ES/R00787X/1. If one has more information than just pairwise comparisons, say scores by panel members for incomplete selections of REF outputs and their confidences in their scores, then another step of analysis is required. In [MKLP], we provided a method for that, which has been mentioned previously in this magazine.

I am grateful to Alex MacKay and Bazil Sansom for help getting the data into MatLab.

Robert MacKay is a past President of the IMA and became an Arsenal supporter during 1982/3 as a postdoc living in Highbury. At least Arsenal's women did well this season!

References

[BT] Bradley RA, Terry ME, Rank analysis of incomplete block designs I. The method of paired comparisons, Biometrika 39 (1952) 324–45.

[MJS] MacKay RS, Johnson S, Sansom B, How directed is a directed network? arXiv:2001.05173

[MKLP] MacKay RS, Kenna R, Low RJ, Parker S, Calibration with confidence: a principled method for panel assessment, Roy Soc Open Sci 4 (2017) 160760.

[SW] Shah NB, Wainwright MJ, Simple robust and optimal ranking from pairwise comparisons, J Mach Learning Res 18 (2018) 1–38.

Mathematics Institute, University of Warwick, Coventry CV4 7AL, U.K.