A MEASURE OF COMPETITIVENESS IN LEAGUES: a network approach

Alan Jessop

Durham Business School

Mill Hill Lane

Durham

DH1 3LB

tel: 0191 334 5403

e-mail: a.t.jessop@durham.ac.uk

Abstract

Many sports are played competitively in a league format. Final positions are based on the aggregations of the points won at each game. Issues of promotion, relegation and much else will depend on the position in the league. However, the results may also be seen to constitute a network of inter-team relations in which the links represent the degree to which a pair of teams have similar performance. This idea is taken as the basis for the construction of a systemic measure of competitiveness in the league. The basis for the model is the construction of a blockmodel on a network of binary relations. The method is illustrated by application to nine seasons of the English soccer Premier League.

Keywords: sports; cluster analysis; mathematical programming; networks and graphs.

A MEASURE OF COMPETITIVENESS IN LEAGUES: A network approach

Introduction

A league is a set of teams or, less frequently, players. Each plays all of the others. If teams have home grounds or stadia then each pair plays twice – home and away. Points are awarded depending on the outcome of each encounter and the team or player with the greatest number of points is the winner. This form of competition is almost universal in organised, especially professional, sport. For the spectator the enjoyment arises in part from the quality of the encounters (was it a good game?) and in part from the accumulation of points and the rewards that follow. But there is, arguably, another consideration; the extent to which the tournament is characterised by competitors of broadly similar competence resulting in some real uncertainty as to the outcome or, at the other extreme, dominance by one or a small number of competitors leading to predictable outcomes which, while the performances may be technically admirable, nonetheless detracts from the contest as spectacle. As well as competition for the top position there will usually be other outcomes of interest to fans: whether local rivals are beaten, whether performance this year is better than that of last year, and so on. The closeness of these results within the main competition add considerably to the wide appeal of the league across the whole season.

It is reasonable to seek some structural description of the results of a season of matches. The final league positions or points total provide a one-dimensional description. A two-dimensional description is provided by a network the links of which describe interactions between pairs of teams and may give both a better understanding of how the final positions were obtained and also a feeling for the characteristics of the matches themselves. It is the purpose of this paper to propose just such a model. While in principle the greater level of detail offered by this approach should provide a fuller structural description this may be mitigated by the extra demands made of the data. It may also be the case that the structure which is apparent at a high level of aggregation (the final result) is inherently less apparent when the data are disaggregated.

The purpose is therefore twofold: to describe and apply a suitable network model and to use the results obtained to investigate the degree to which a particular soccer league may be said to exhibit structure.

This paper is organised as follows: different ideas of competitiveness are considered and a network model is proposed; the model is applied and the results obtained are presented; the results and application of the model are discussed.

The data

The data used as illustration are taken from football (soccer) and are the results in the English Premiership. English professional football is organised into four leagues or divisions (the names seem to vary but since "league" has a better marketing ring it seems to be gaining the upper hand). These used to be called simply Division One down to Division Four (or, for those of us of the necessary antiquity, Division Three South and Division Three North). In the season 1995-6 (hereafter just 1995) the highest division set itself apart with a separate management in order better to exploit the commercial revenues available, primarily from the sale of television rights. This highest division was called the Premiership, now the Barclays Premiership in recognition of sponsorship. The next division is now called the Coca Cola Championship and the third and fourth divisions are Coca Cola League One and Coca Cola League Two (the marketing folk march on).

There are twenty teams in the Premiership. Each plays all of the other nineteen twice, at home and away, giving 380 matches in all. Three points are awarded for a win and one for a draw. These points are aggregated to give a final table of results. In the event that two teams have the same number of points the tie is resolved by the goal difference of each; the goal difference being the difference between the total number of goals scored and conceded by that team. Figure 1 shows the final points scores for each of the nine seasons studied in this paper.

The top four teams are eligible to play in a pan European competition, the Champions League, the following year. The next three teams play in a less prestigious European competition, the UEFA cup. As well as kudos participation in these European competitions also brings considerable additional revenues.

The bottom three clubs are relegated and replaced by the top three clubs from the Championship. Again, the financial implications are great.

The data used in this paper are the results of the nine seasons from 1995 until 2003.

Competitiveness

The award and aggregation of points provides some measure of competitiveness: the spread of points might reasonably be seen as an indication of how competitive the season of matches has been. From Figure 1 it is easy to see that the 2003 season in which Arsenal carried all before them, not losing one of their 38 matches, was not prima facie as competitive as, say, the 1997 season in which the spread of points was less. The first two lines of Table 1 give the mean number of points and the standard deviation.

However, while the total number of points awarded gives an idea of overall performance it misses an important dimension of competition. For those fans following a particular team and for neutral observers much of the excitement and interest of the season is to do with the degree to which that team has not been dominated by, nor has dominated, others. It is generally assumed that some degree of uncertainty about the outcome of the match has a beneficial effect on attendance at football matches (Peel and Thomas, 1992; Forrest and Simmons, 2002) and also at rugby games (Peel and Thomas, 1997). This uncertainty is determined by a number of factors, some of which are structural – broadly the underlying differences in ability between teams – and some of which are specific – local derbies (matches between local rivals) being an often cited example. The competitiveness inherent in a match will also vary through time; the end of the season seeing both vigorously contested matches to determine matters of qualification at the top of the table and relegation at the bottom, and also fairly meaningless middle of the table contests. (This latter is not a necessary condition of league competitions. In the NFL football league in the US the order in which teams select incoming players for a new season – the draft – is determined by league position, the bottom team having first pick.) In trying to estimate the outcomes of particular matches it is sometimes assumed that betting odds contain useful information about these various factors (Peel and Thomas, 1992, 1997; Forrest and Simmons, 2002) although a Poisson model has been used to simulate match results (Crowder *et al*, 2002; Croucher, 2004). A regression model using several variables has also been used for this purpose (Dobson and Goddard, 2004).

In this paper we focus not on forecasting individual results but rather on retrospective description of a season's play and on whether this has changed over the nine years of the Premiership. The strategy essentially is that the results of the 380 matches define the links in a twenty node network and that this permits the construction of a measure describing the degree to which nodes can be grouped into blocks of similarly performing teams. It is these pairwise competitions which both underlie the final league positions of the teams and also describe the individual tussles which give richness and variety to the league season for the fans of all twenty teams. Describing the season at this more detailed level will provide a fuller appreciation of the structure of competitiveness within the league. The two descriptions of the results – blocks and the league table of points – are both mappings of the results and should be seen as complementary. Blocks are based on pairwise competitions and so motivations may be particular to the pair of teams, as in local derbies or other contests between foes of long standing. Points, on the other hand, determine the all-important league positions and the financial consequences thereof.

Differences in performance

One of the main and consistent factors affecting match outcomes is home advantage. We avoid this bias by looking only at results aggregated over both games that a pair of teams play.

Perhaps the most obvious measure of the difference in performance between two teams, i and j, is the goal difference. Define

$$g_{ij} = \text{goals scored by } i - \text{goals scored by } j\text{, taken over both matches}$$

However, league positions are determined not by goals but by points. It is not uncommon that a team, once ahead in a match, just protects that lead by playing the rest of the game defensively. There arise occasionally situations in which teams may have the same number of points and so, at either end of the league, goal difference is important, but these are infrequent. In recognition of this primacy of points we may prefer a measure of the difference in performance based on points rather than goals:

$$p_{ij} = \text{points scored by } i - \text{points scored by } j\text{, taken over both matches}$$

A network-based measure

Any network can be made into a symmetrical binary network, X, by setting

$$\begin{aligned} x_{ij} &= 1 \text{ if } |g_{ij}| \leq \alpha \quad ; \quad \alpha>0 \\ &= 0 \text{ otherwise} \end{aligned} \qquad (1)$$

and similarly for p_{ij}. In what follows it is convenient also to set

$$x_{ii} = 1 \qquad ; \quad \forall\, i$$

X then indicates pairs of teams whose performances are practically indistinguishable according to the parameter α.

We may group pairs of teams into blocks which are maximally dense:

$$x_{ij} = 1 \text{ for all pairs } i,j \text{ in the same block}$$

For instance, if teams *a* and *b* and *c* constitute a block then *a* is similar to *b*, *b* is similar to *c* and *a* is similar to *c* and each of these three similarities have to be established explicitly; there is no commutative effect whereby the first two imply the third.

The number, B, of these maximally dense blocks which it is possible to construct gives a measure of the competitiveness in the league. The smaller the number of blocks the more competitive the league: in the extreme, if $B = 1$ then it is impossible to distinguish the performance of any of the teams and the league is maximally competitive, while if $B = 20$ then no blocking is possible and the league is minimally competitive, each team is distinguished from all others.

Following the general approach adopted by Jessop (2003) define a membership matrix Λ in which the binary variable $\lambda_{ik} = 1$ if team i is in block k and 0 if it is not. Since each team may be in only one block,

$$\sum_k \lambda_{ik} = 1 \qquad ; \quad \forall\, i \tag{2}$$

Block k comprises s_k teams where

$$s_k = \sum_i \lambda_{ik} \qquad ; \quad \forall\, k \tag{3}$$

Since each block must be a maximal density block we have

$$\sum_i \sum_j x_{ij}\lambda_{ik}\lambda_{jk} = s_k^2 = (\sum_i \lambda_{ik})^2 \qquad ; \ \forall\, k \tag{4}$$

Finding block membership involves solutions for Λ which are in some sense optimal but the formulation in (4) leads to an integer quadratic programme in which only local optima will be found. Reformulation as an integer linear programme leads to globally optimal solutions but at the expense of an increased problem size (though this increase may be mitigated by exploiting the symmetry of the problem and computing using only half matrices). Introduce the three dimensional matrix Φ such that

$$\phi_{ijk} = \lambda_{ik}\lambda_{jk} \tag{5}$$

which can be achieved via the constraints

$$\left.\begin{aligned} \phi_{ijk} &\le \lambda_{ik} \\ \phi_{ijk} &\le \lambda_{jk} \end{aligned}\right\} \tag{6}$$

$$\phi_{ijk} \geq \lambda_{ik} + \lambda_{jk} - 1$$

so that (4) becomes:

$$\sum_i \sum_j \phi_{ijk} x_{ij} = \sum_i \sum_j \phi_{ijk} \quad ; \quad \forall k \qquad (7)$$

To find the minimum number of maximum density blocks introduce the vector Γ of binary variables γ_k which have value 0 if block k is empty and 1 otherwise. Ensure this by the constraints

$$\gamma_k \leq s_k = \sum_i \lambda_{ik} \quad ; \quad \forall k \qquad (8)$$

and $$Q.\gamma_k \geq s_k = \sum_i \lambda_{ik} \quad ; \quad \forall k \qquad (9)$$

where Q is a number large in the context of the problem: Q = 100 was used.

Set the matrices Λ and Γ to have dimensions $i = 1 \ldots 20$ and $k = 1 \ldots B \leq 20$ and then solve the programme

$$\min \sum_k \gamma_k \qquad (10)$$

subject to constraints (2), (6), (7), (8) and (9). The result is the minimum number, $B_{\min}$, of non-empty blocks.

In general the solution is not unique: there may be more than one blocking pattern for any $B_{\min}$. This is unimportant if the purpose of the analysis is to use $B_{\min}$ as a measure of competitiveness but is important if we also wish to identify which teams comprise the blocks. There is another consideration. Just as we define the density of the network to be

$$D = \sum_i \sum_j x_{ij} / 20^2 \qquad (11)$$

we may also define the proportion of links which are accounted for by the block structure as

$$A = \sum_k s_k^2 / \sum_i \sum_j x_{ij} \qquad (12)$$

the numerator being just the number of links in all $B_{\min}$ blocks. It is natural to wish to maximise A and this can be done via the programme

$$\max \; S = \sum_k s_k^2 = \sum_k \sum_i \sum_j \phi_{ijk} \qquad (13)$$

subject to (2), (6) and (7). The number of blocks for the summation of k is set to $B_{\min}$ as determined by the previous model. S is a measure of concentration in that it reflects the allocation of the nodes into (maximum density) blocks, higher values of S corresponding to the presence of larger blocks. (Although argued from a slightly different standpoint this is formally the same as the measure of industrial concentration due to Herfindahl (1950) and Hirschman (1964) as applied to the problem of blockmodel construction by Jessop

(2003).) Again, the solution may not be unique, although the number of optimal solutions will be smaller, of course.

As well as the pairwise relations X there may be other measures which characterise the relative fortunes of two teams; the difference in their final league positions, for instance. Given such a measure, say q_i for team i, then $y_{ij} = |q_i - q_j|$ is a measure of distance between pairs of teams. Given two or more block structures with the optimal value $S_{\max}$ one may be preferred if the constituents of blocks exhibit greater *compactness* in that teams in each block are in aggregate closer according to Y. We may, in other words, prefer a structure which results from the programme

$$\min \; C = \sum_k \sum_i \sum_j \phi_{ijk} y_{ij} \tag{14}$$

subject to

$$\sum_k \sum_i \sum_j \phi_{ijk} = S_{\max} \tag{15}$$

and also subject to (2), (6) and (7) as before.

Three models have been described:

(a) min B provides a good first descriptor and, importantly, provides a value for $B_{\min}$ which permits much reduced computing time for (b) and (c) below;

(b) max *S* describes the aggregation into maximum density blocks and so the proportion of interactions accounted for by the blocking;

(c) min *C* provides an improved set of block memberships and so is only useful if this is of interest, since nothing is added to the systemic description provided by $B_{\min}$ and $S_{\max}$.

Application and results

System description

Three binary networks are defined; one based on goal difference and two on points. Because of the infrequency of high scoring matches (Table 1, line 3) a modest goal difference was used, teams being deemed similar if the goal difference, over both matches, was no more than 1.

The absolute value of the points difference across two matches can be either 0, 3 or 6. The first two were used. A points difference of 0 arises when either both matches are drawn or each team wins one match. A difference of at most 3 points arises in all cases except where one team wins both matches. The three networks are

$$
\begin{aligned}
x_{ij} &= 1 \text{ if } |g_{ij}| \leq 1 \quad \text{the } \textit{goals} \text{ network} && (16)\\
&= 1 \text{ if } |p_{ij}| = 0 \quad \text{the } \textit{0points} \text{ network} && (17)\\
&= 1 \text{ if } |p_{ij}| \leq 3 \quad \text{the } \textit{3points} \text{ network} && (18)\\
&= 0 \text{ otherwise}
\end{aligned}
$$

As illustration, Figure 2 shows an application to the results for the 2000 season. Figure 2(a) shows the results for each match. Figure 2(b) shows the binary *0points* network. A 1 in the table show where each of a pair of teams gained the same number of points from their two matches. Figure 2(c) shows the effect of the blockmodel which is essentially a reordering of rows and columns to give the blocks as maximum density squares along the diagonal. The densities of the off-diagonal elements measure the interaction or connectedness between the blocks. For some applications a further hierarchical clustering of blocks using these inter-block densities may be helpful. This was not judged to be the case here since it is the systemic description provided by the blocks themselves which is the main concern.

Table 1 shows the results of applying the models. The densities of each type of network do not change much: the *goals* model accounts for about half the matches played; the *0points* model for about a third and the *3points* model for about three quarters.

The minimum numbers of blocks does not alter much, although with the *0points* model there is a small increase in the number of blocks over time. As is to be expected, the number of blocks is smaller the greater the network density.

The proportion of the links accounted for by the blocks, *A*, initially varies but over time appears to converge to a similar value of about 0.4 for all three models. This is further discussed below.

Team performance

The above describe the structural properties of the competitive system. Do the block structures correlate with the results for individual teams?

In minimising *C* we choose to use the points difference between teams as a measure of compactness. Table 2 shows the results for the *0points* network. Points achieved rather than team names are shown. Blocks are arranged in descending order of mean points score per block. Within each block points are listed in descending order.

A measure of the extent to which the order as presented is different from that based on unblocked points is given by computing Spearman's rank correlation coefficient, r_s, between the orderings such as those given in Table 2 and the final league positions. The degree to which the blocking alters the ranking will depend in part on the number of blocks. Because of this it is probably only sensible to compare coefficient values for each network in turn (the number of blocks varying only a little), rather than between networks. It ought also to be borne in mind that the differences in block means are not always great. However, values of r_s do give an indication of blocking effects (Table 1). Figure 3 shows two illustrative examples to assist in the interpretation of r_s. High values of r_s indicate that similarly performing teams are in the same block whereas low values and so may be seen to indicate more unexpected (and so entertaining?) results.

Discussion

Blockmodels have been presented as a description of the results of a league competition. The results may be considered at two levels; at the level of structural description and at the level of team performance. In considering these it must be remembered that block structures are not necessarily unique: there may be more than one partitioning of teams into blocks that gives the same distribution of block sizes corresponding to a given minimum number of blocks or maximum value of S. At the level of structural description the existence of multiple optima is of no concern.

In describing structure two parameters are used in finding blocks. The first is the performance measure used and the second is the cut value α (1). Three combinations were used ((16) to (18)) resulting in corresponding criteria and binary networks. Because it is points rather than goal differences which determine so many issues important to the clubs the measures based on points are preferred. Network density (the fraction of matches with results meeting the criterion) is, of course, different depending upon the network used, ranging from about 30% to about 75%. These densities do not change much over time. Although the different networks are of different densities the degree to which the interactions are accounted for by the blockmodels, A, is roughly constant for each network. This may be due to a characteristic inherent in the game.

Wesson (2002) makes the point that soccer is, by design or otherwise, a low scoring game (for the matches studied here the mean number of goals per game is about 2.6) and that as a consequence the result of any match is not very predictable. This is a function of the rules which govern the conduct of each match, the fact that it is in the end points and not goals which matter and so the defence of a narrow advantage, and, as in any such sport, the physical opposition of two teams resulting in an Ashby-like mutual regulation ("If you want to contain the behaviour of twelve people in red shirts running madly all over a football field, you will need at least twelve men in white shirts to do it. (Beer, 1975))

The argument is effectively one of small sample size: if the number of goals per match were higher then the abilities of the superior (more skilled and better organised) team would more often lead to the defeat of lesser teams. Although the binary networks here are formed by the aggregation of home and away results this hardly represents a great increase in sample size. In the extreme, the result of each match may be determined randomly, which is to say without regard to the skills or strengths of the teams involved. Figure 4 shows that this might well be so. For each of the three network types twenty random binary networks were constructed with probabilities of 0.33, 0.5 and 0.75, which are the average network densities for the networks. Each link was assigned a random number. If this was less than the characteristic probability the link was actuated; x_{ij}=1. The results are shown together with the results for the nine seasons of the Premiership. At the extremes of density only one configuration is possible: when D=1 all possible links are realised and there is just one block, while if D=1/n (n is the number of nodes) there are just n singletons. In both cases A=1. For intermediate densities the number of possible configurations, and so the possible range of A, is higher.

Two features of Figure 4 are worth noting: first, that there is a quite wide spread of A values for any density and, second, that the actual results are indistinguishable from those of the simulated random networks. In consequence we may believe that the lower bound for A as indicated by the random networks is not sharply defined and that the results obtained from the football networks lie in that region.

The similarity of the results for the random and football networks is an initial indication that the structure in football league results, considered as pairs of matches, is at best weak.

However, the indications provided by the rank correlation measures are that there is some correspondence between block membership and seasonal performance as shown by the final points total. This total is based on the aggregation of the results of 38 matches and so should be a quite good indicator of underlying relative team performance. The correspondence between block membership and points obtained were reasonably high: r_s>0.7 for 14 of the 27 results and r_s>0.5 for 24. To the extent that the correlations are nor perfect it may be inferred that some unexpected results are to be found and this is to be welcomed as an indicator, albeit a crude indicator, of the entertainment value of the matches throughout the season.

This is an apparently paradoxical result: that meaningful groups can be found in data even though the foundation for the construction of those groups is weak. One interpretation is that this demonstrates the power of the aggregation of 38 results into overall points when compared with the aggregation over just two games to form networks. In addition, recall the relation between the max S and min C models. Because of the weak network structure there are likely to be a number of possible patterns of block membership which are consistent with the value of $S_{\max}$. The pattern found in the optimisation will depend in part on the implementation of the algorithm in the software and is to that extent arbitrary. Minimising C was introduced as a means of finding a particular solution for $S_{\max}$. It is not possible to assess the impact of the min C criterion by examining the reduction of C consequent upon its introduction because of the arbitrariness mentioned above. As an alternative consider the change in S (and so in A) of dropping from the min C model the constraint that S=$S_{\max}$. The results are shown in Table 1 as the last line in each section. The reduction is usually no more than 10%. This difference may be seen as an indication that imposing the S=$S_{\max}$ constraint gives a modest but perceptible improvement in S indicating the existence of some structure in the network.

The existence of a weak structure at this level is consistent with attempts made to forecast match outcomes. Croucher (2004) used two Poisson models the parameters of which were the mean number of goals per match scored at home and away by all teams in the league. Match results were then forecast assuming independence of the performance of the two teams (ie.using just the product of the two Poisson probabilities). It was concluded that while this gave reasonable results in some cases the assumption of independence was unsustainable (see also Dobson and Goddard, 2004) and that there was some interaction between teams. It is the interactions between team pairs which is the basis for the network models, though the characterisation of the results in the random networks is not so fine as that provided by the two Poisson model.

Conclusions

The feasibility of constructing blockmodels as descriptors of league performance has been demonstrated.

Block structures do not vary significantly over time. This may be because their characteristics are similar to those of random networks of the same density and that it is this comparatively weak structure which, getting no stronger, is at least stable.

This similarity to random networks may be inherent in the game of soccer, in particular as a consequence of the low scoring rate. While this helps to provide agreeably unexpected results it also means that at the level of team pairs little overall structure can be determined.

Meaningful identification of the teams in each block is possible. To quite a good degree block membership is related to team strength as measured by aggregate points totals. The small change in S which results from relaxing the requirement that $S=S_{max}$ reinforces the view that while some structure exists in the network it is not great.

The Barclays Premiership has enough structure to ensure that over a season the better teams succeed and the poorer teams are relegated but, at the network level of matches, a much weaker structure so that the outcomes are sufficiently uncertain that entertainment is maintained. The stronger conclusion is that the networks have characteristics indistinguishable from those of random networks and so exhibit no structure at all and so no information is useful once network density is known. That blocks may be constructed in which a non-trivial relation exists between block membership and points total indicates that the stronger conclusion is too strong and that some further investigation may be worthwhile.

These investigations into league structure have usefully been made using the network models developed.

Acknowledgement

I should like to thank Dr Les Proll of the University of Leeds for helpful discussion on the LP formulation of the blockmodel.

References

Beer S (1975). *Platform for Change*. Wiley: Chichester.

Croucher JS (2004). Using statistics to predict scores in English Premier League soccer. In: Butenko S, Gil-Lafuente J and Pardalos PM (eds). *Economics, Management and Optimization in Sport*. Springer: Berlin.

Crowder M, Dixon M, Ledford A and Robinson M (2002). Dynamic modelling of English Football League matches for betting. *The Statistician* **51** part 2: 157–168.

Dobson S and Goddard J (2004). Modelling and forecasting match results in the English Premier League and Football League. In:Butenko S, Gil-Lafuente J and Pardalos PM (eds). *Economics, Management and Optimization in Sport*. Springer: Berlin.

Forrest D and Simmons R (2002). Outcome uncertainty and attendance demand in sport: the case of English soccer. *The Statistician* **51** part 2: 229–241.

Herfindahl OC (1950). *Concentration in the U.S. Steel Industry*. Ph.D. Thesis, Columbia University: New York.

Hirschman AO (1964). The paternity of an index. *Amer Econ Rev* **54**: 761–762.

Jessop A (2003). Blockmodels with maximum concentration. *Eur J Opl Res* **148**: 56–64.

Peel D and Thomas D (1992). The demand for football: some evidence on outcome uncertainty. *Empirical Economics* **17**: 323–331.

Peel D and Thomas D (1997). Handicaps, outcome uncertainty and attendance demand. *Applied Economics Letters* **4**: 567–570.

Wesson J (2002). Golden goals. *New Scientist* 10 August:36–39.

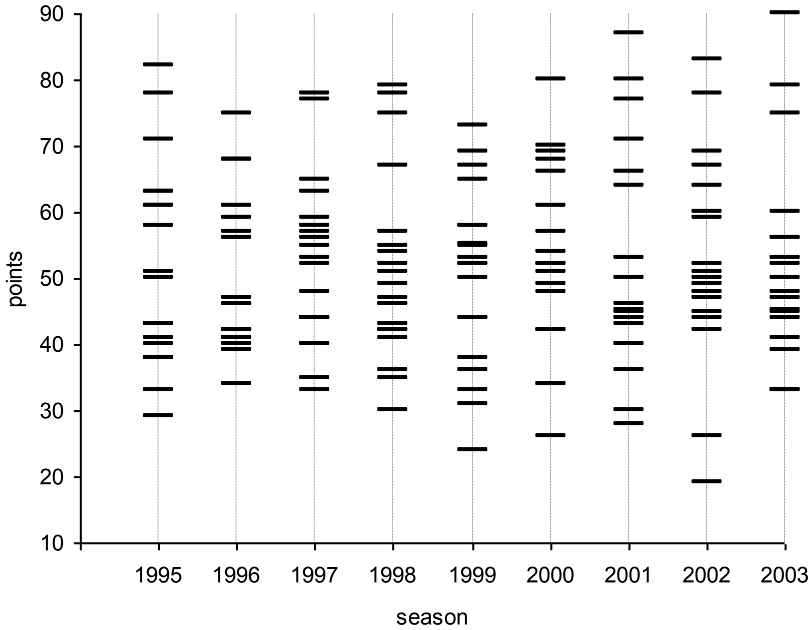


Figure 1 Final points score.

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	
1		1.0	2.0	5.3	1.1	2.1	0.0	4.1	1.0	2.1	6.1	2.0	5.0	1.0	0.3	5.0	1.0	2.2	2.0	3.0	70
2	0.0		2.0	2.1	1.1	3.2	4.1	2.1	2.1	1.2	2.1	0.3	2.2	0.1	1.1	1.1	0.0	0.0	2.0	2.2	54
3	1.1	0.3		2.0	2.0	2.1	2.0	0.1	0.2	1.1	0.0	0.2	2.2	0.3	1.1	2.2	0.1	1.4	3.3	1.2	26
4	1.0	3.3	2.0		2.0	2.2	2.1	1.0	2.1	1.2	2.0	0.4	4.0	3.3	1.0	2.0	1.1	0.1	1.0	1.1	52
5	2.2	1.0	3.0	0.1		6.1	4.1	2.1	4.1	1.1	0.2	3.0	2.1	1.1	2.1	3.1	1.0	2.4	3.0	4.2	61
6	0.1	1.1	0.0	2.2	0.0		2.0	1.3	0.1	0.0	1.0	0.2	1.1	1.2	1.3	0.2	1.1	1.0	2.1	0.3	34
7	1.2	1.0	2.0	2.2	0.4	1.0		1.0	1.1	1.1	2.0	0.4	1.1	0.3	3.3	2.0	2.2	1.0	2.1	0.0	42
8	2.0	0.1	2.1	3.0	2.1	1.2	2.2		0.3	2.2	2.1	2.3	3.1	1.3	2.2	1.1	1.1	2.2	0.0	1.1	42
9	1.1	1.2	3.1	2.0	2.2	2.0	0.1	2.0		1.2	2.0	1.1	2.1	1.1	2.1	1.0	3.1	1.0	3.0	1.1	66
10	1.0	1.2	6.1	3.1	2.0	1.0	0.0	2.0	1.2		3.1	4.3	1.2	1.1	1.1	1.3	2.0	2.0	4.3	0.1	68
11	0.0	0.0	1.2	3.1	2.1	1.3	2.1	1.1	2.1	3.1		2.0	1.2	0.3	0.3	1.1	1.0	2.0	4.2	2.1	48
12	4.0	3.1	1.0	3.0	2.2	4.1	1.1	3.1	0.1	1.2	1.0		3.2	2.0	0.0	3.0	2.1	1.1	3.1	3.0	69
13	0.4	1.3	2.0	1.4	1.2	1.2	0.0	5.0	2.3	0.4	0.1	1.1		0.1	1.1	0.1	0.1	4.2	0.1	1.0	34
14	6.1	2.0	6.0	2.1	3.3	4.2	0.1	1.0	2.0	3.0	2.0	0.1	1.1		2.1	2.0	5.0	3.0	2.0	3.1	80
15	0.1	1.1	2.2	0.0	1.0	1.1	4.0	1.2	1.2	1.2	0.3	1.0	1.1	0.2		1.3	0.1	0.0	1.1	2.1	42
16	0.0	3.0	2.1	0.1	0.0	3.1	3.2	0.1	2.1	2.1	1.0	2.1	0.1	1.1	1.2		1.1	1.2	2.0	2.1	51
17	3.2	2.0	2.0	0.0	3.2	1.2	1.0	1.0	0.3	1.0	1.0	3.3	0.2	2.1	1.3	2.0		0.1	2.0	2.3	52
18	1.0	1.1	0.0	3.2	1.0	1.0	2.1	2.0	4.1	0.2	0.0	1.1	1.0	0.1	1.0	1.1	2.2		2.3	1.1	57
19	1.1	0.0	2.1	0.0	0.3	3.0	3.1	3.2	3.1	1.2	3.0	2.1	0.0	3.1	0.0	4.2	0.0	2.1		1.0	49
20	1.2	1.1	1.1	5.0	0.2	1.1	3.1	0.2	0.1	0.2	0.1	1.1	4.1	2.2	1.0	1.0	3.0	0.2	0.0		42

(a) Results for 2000 season.
Rows are home team and columns away team.
Figures at right hand margin show points for season.
Figures in table show score, home team first (e.g. 1.2 means home team scored 1 goal and away team 2)

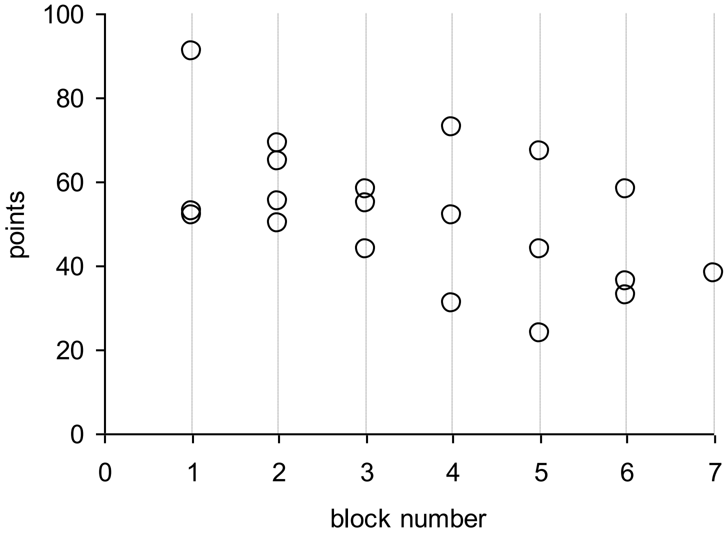
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
1	1			1	1			1		1		1		1	1		1			
2		1					1			1					1			1		1
3			1	1	1		1								1					
4	1		1	1		1		1	1		1						1			
5	1		1		1			1						1	1		1			
6				1		1	1	1										1	1	
7		1	1			1	1			1	1		1	1		1		1	1	
8	1			1	1	1		1					1							
9				1					1	1	1					1		1	1	
10	1	1					1		1	1	1		1				1			1
11				1			1		1	1	1	1	1		1		1		1	
12	1										1	1				1		1	1	
13							1	1		1	1		1		1	1	1	1		1
14	1				1		1							1			1		1	
15	1	1	1		1						1		1		1	1	1		1	1
16							1		1			1	1		1	1			1	1
17	1			1	1					1	1		1	1	1		1			
18		1				1	1		1			1	1					1		
19						1	1		1		1	1		1	1	1			1	
20		1								1			1		1	1				1

(b) original network: rows and columns ordered according to league position

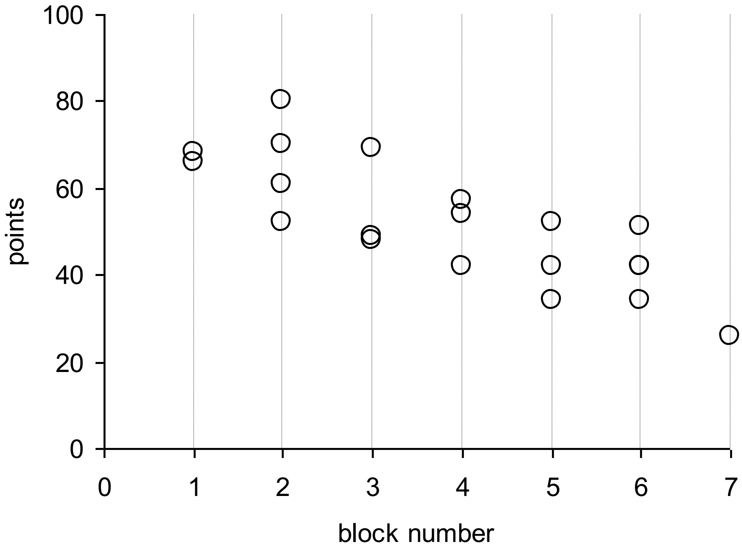
	9	10	1	5	14	17	11	12	19	2	7	18	4	6	8	13	15	16	20	3
9	1	1					1		1			1	1					1		
10	1	1	1			1	1			1	1					1			1	
1		1	1	1	1	1		1					1		1		1			
5			1	1	1	1									1		1			1
14			1	1	1	1			1		1									
17		1	1	1	1	1	1						1			1	1			
11	1	1				1	1	1	1		1		1			1	1			
12			1				1	1	1			1						1		
19	1				1		1	1	1		1			1			1	1		
2		1								1	1	1					1		1	
7		1			1		1		1	1	1	1		1		1		1		1
18	1							1		1	1	1		1		1				
4	1		1			1	1						1	1	1					1
6									1		1	1	1	1	1					
8			1	1									1	1	1	1				
13		1				1	1				1	1			1	1	1	1	1	
15			1	1		1	1		1	1						1	1	1	1	1
16	1							1	1		1					1	1	1	1	
20		1								1						1	1	1	1	
3				1							1		1				1			1

(c) blockmodel: rows and columns reordered to reveal blocks with blocks ordered by mean number of points (see also Figure 3(b) and Table 2)

Figure 2 *0points* and min *C* model for 2000



(a) 1999, r_s = 0.59



(b) 2000, r_s = 0.84

Figure 3

Relation between blocks and performance: *0points* model

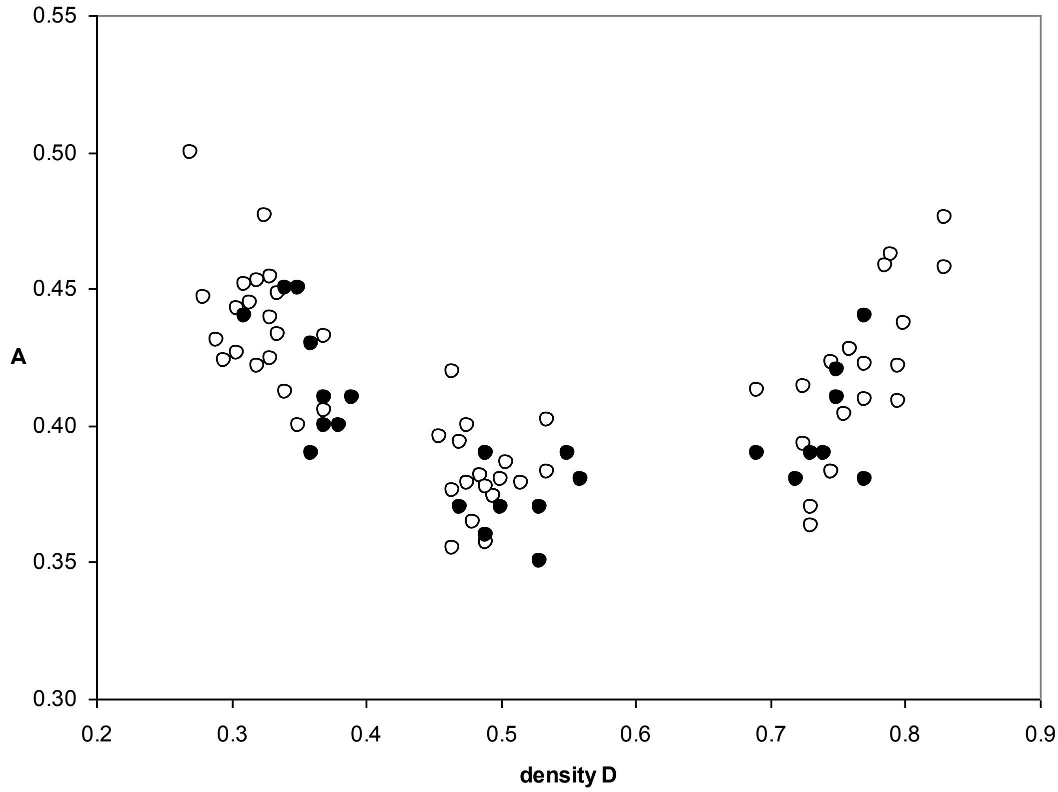


Figure 4 Relation between *A* and *D*.

The open circles show random networks and the filled circles the football league networks.

		1995	1996	1997	1998	1999	2000	2001	2002	2003
points scored	*mean*	52.1	50.9	52.3	51.3	52.4	52.0	52.0	52.5	51.6
	st. dev.	15.2	12.0	12.4	14.0	16.1	14.0	16.7	15.3	15.2
goals / match	*mean*	2.60	2.54	2.68	2.53	2.78	2.61	2.63	2.63	2.66
goals model	$B_{\min}$	6	6	6	6	6	6	6	5	5
	D	0.53	0.49	0.49	0.50	0.47	0.53	0.49	0.56	0.55
	$S_{\max}$	74	70	76	74	70	78	76	84	86
	A	0.35	0.36	0.39	0.37	0.37	0.37	0.39	0.38	0.39
	$C_{\min}$	712	580	760	900	770	712	856	866	1288
	r_s	0.74	0.65	0.78	0.56	0.55	0.81	0.81	0.51	0.44
	% change in S	5	3	11	8	0	8	3	0	5
0points model	$B_{\min}$	7	8	7	8	7	7	9	8	8
	D	0.35	0.36	0.37	0.34	0.36	0.39	0.31	0.38	0.37
	$S_{\max}$	62	56	60	60	62	64	54	60	58
	A	0.45	0.39	0.41	0.45	0.43	0.41	0.44	0.40	0.40
	$C_{\min}$	732	294	496	402	786	508	236	498	386
	r_s	0.60	0.79	0.73	0.84	0.59	0.84	0.91	0.71	0.84
	% change in S	3	7	3	13	3	3	7	13	10
3points model	$B_{\min}$	4	4	4	3	4	5	4	4	4
	D	0.74	0.75	0.75	0.77	0.69	0.77	0.72	0.72	0.73
	$S_{\max}$	114	126	122	134	108	116	114	108	114
	A	0.39	0.42	0.41	0.44	0.39	0.38	0..40	0.38	0.39
	$C_{\min}$	1798	1372	1138	1602	1520	848	1512	1120	1450
	r_s	0.42	0.69	0.85	0.61	0.65	0.92	0.74	0.54	0.40
	% change in S	11	17	15	0	6	28	12	2	12

Table 1 Summary of results

	1995	1996	1997	1998	1999	2000	2001	2002	2003
	82	68	77	79	91	68	87	83	90
	61	68	59	75	53	66	80	78	75
	78	68	48	78	52	80	77	69	79
	38	59	65	67	69	70	71	67	60
	71	75	63	42	65	61	66	48	56
	43	61	57	52	55	52	64	47	53
	61	40	56	57	50	69	53	64	56
	58	56	78	51	58	49	45	51	45
	41	46	44	46	55	48	50	59	45
	63	57	44	55	44	57	46	49	44
	50	42	58	54	73	54	40	45	52
	43	41	40	47	52	42	45	52	50
	40	47	53	42	31	52	44	49	39
	63	46	44	49	67	42	44	48	53
	61	46	55	46	44	34	43	44	47
	38	39	52	41	24	51	40	60	33
	29	42	35	36	58	42	36	50	41
	51	42	44	43	36	42	50	19	48
	38	41	40	30	33	34	30	42	33
	33	34	33	35	38	26	28	26	33
r_s	0.60	0.79	0.73	0.84	0.59	0.84	0.91	0.71	0.84

Table 2 Blocks for the *0points* model showing points. Heavy lines show block boundaries.

Figure 1 Final points score.

Figure 2 *0points* and min *C* model for 2000

(a) Results for 2000 season.
Rows are home team and columns away team.
Figures at right hand margin show points for season.
Figures in table show score, home team first (e.g. 1.2 means home team scored 1 goal and away team 2)

(b) original network: rows and columns ordered according to league position

(c) blockmodel: rows and columns reordered to reveal blocks with blocks ordered by mean number of points (see also Figure 3(b) and Table 2)

Figure 3

Relation between blocks and performance: *0points* model

(a) 1999, $r_s = 0.59$

(b) 2000, $r_s = 0.84$

Figure 4 Relation between *A* and *D*.

The open circles show random networks and the filled circles the football league networks.

Table 1 Summary of results

Table 2 Blocks for the *0points* model showing points. Heavy lines show block boundaries.