Universiteit Utrecht

Faculteit Bètawetenschappen

Ranking Methods for the Dutch Football Competition

Bachelor Thesis

Thomas Breeman

Bachelor Wiskunde

Supervisor:

Dr. S. (Sjoerd) Dirksen
Mathematisch instituut

November 8, 2020

Abstract

In the season 2019-2020, the Dutch Football competition was abrupted prematurely due to COVID-19. This season was not fully completed and therefore has some interesting properties for ranking methods, due to the fact that this is an uneven paired competition. I have used three different kind of ranking methods, based on the Perron-Frobenius theorem. The results show that the methods based on Keener's method produced rankings similar to the ranking assigned by the Dutch Football Association (KNVB). The method based on PageRank was very different than the original method.

Contents

1 Introduction

Statistics and football have always been of great interest to me. I have been a football fan my entire life, and I always followed the rankings of football teams all over the world. Most of these ranking methods are similar to one another: you get 3 points for winning a match, 1 point for drawing, and 0 for losing. At the end of the season the team with the highest amount of points wins the competition, the team with the second highest amount of points is placed second and so on. When teams have equal amounts of points, the tie breaker may be determined differently in different competitions. For example in the Dutch competition, in case of an equal amount of points, a team is rated higher if the goal difference (the amount of goals scored minus the amount of goals conceded) is higher. In the Spanish league the tie breaker is decided using the results the teams got against each other. So if a team won twice against the team it is tied with in the league, it gets rated higher [1].

In the season 2019-2020 something interesting happened. Due to the COVID-19 pandemic the Dutch football competition was abrupted. The organisation of the Dutch football competition, KNVB, stood for a challenge. Normally the ranking at the end of the season determines whether a team may play in Europe or relegate to a lower division the next season. However, this particular season was not finished. This means that not every team had played each other team twice, so some teams would have had an 'easier' schedule than others. The KNVB decided that there would be no champion, as well as no relegation candidates. The ranking at the time the competition was stopped would function as the end-of-season ranking, meaning the teams playing in Europe next year would be appointed accordingly [2]

AZ Alkmaar, the second ranked team in that season, thought this was unfair. They had the same amount of points, but were second due to having a lower goal difference, which, in the Dutch Football competition, is the tie breaker. One of the arguments they made was that the matches against each other would represent a better ranking method for the tie breaker, due to the fact the competition was not finished [3].

This raises the question whether another ranking system would be more suitable for this particular season. That is why this study will investigate what other ranking algorithms have to say on this matter. The ranking methods used have been developed to rank teams in an uneven paired competition. This paper will contain a comparison between two important ranking algorithms and some manipulation of those. One is developed by W.N. Colley and J.P. Keener[4], the other is developed by Anjela Y. Govan, Carl D. Meyer and Russel Albright [5]. Both these methods rely in some fundamental way on the Perron-Frobenius Theorem. The first method uses this theorem directly and the other one uses it in an indirect way.

The remainder of this thesis is organized as follows. We first will describe the methods in a detailed way. We will describe how they work and we will finish this part with a comparison of these methods. In the next part we will see the results of these ranking methods applied to the Dutch Football Competition in the seasons of 2018-2019 and 2019-2020. We will make a conclusion of these results and we will end this thesis with a discussion of our results.

2 Method

2.1 Keener's Method

Keener's Method was originally used to rank American Football teams [4]. The competition these teams play in is an uneven paired competition. An uneven paired competition is a competition where each team is not paired with every other team an equal amount of times. Because of this, the strength of a team's schedule matters. A team could play mostly against the weaker teams of the competition, obtaining a high percentage of wins, but may be weaker than other teams with less wins. Keener's Method will compensate for this and does so by using the Perron-Frobenius Theorem.
The idea is as follows. We want to assign a rank to each team based on the interactions with the other teams. We do this by using a preference matrix A. The entries of this matrix, $\mathrm{a_{ij}}$, will be a score based on the interaction between team i and j. For example, for football we could pick $\mathrm{a_{ij}}$ to be 1 if team i won the game, 1/2 if the game ended in a draw and zero otherwise. Now, we say that the ranking of a team should be proportional to its score, e.g.:

$$A\mathbf{r} = \lambda\mathbf{r} \tag{2.1}$$

with r the ranking vector, λ the eigenvalue and A the preference matrix.
The Perron-Frobenius theorem tells us that such a system has a solution, with a positive eigenvector, under certain assumptions. One of the assumptions is that the matrix A is irreducible. The following definitions are equivalent to one another, as proven in Appendix A:

1. A is irreducible if for any two numbers i and j there is an integer p > 0 and a sequence of integers $\mathrm{k_1,.....,k_p}$, so that the product $\mathrm{a_{ik_1}a_{k_1k_2}...a_{k_pj}} \neq 0$.

2. A is irreducible if there is no permutation that transforms the matrix A into a block matrix of the form:

$$\begin{pmatrix} A_{11} & A_{12} \\ 0 & A_{22} \end{pmatrix}$$

   With $\mathrm{A_{11}}$ and $\mathrm{A_{22}}$ square matrices.

Theorem 2.1.1. *If the matrix A is irreducible and has nonnegative entries, then there exists an eigenvector r with strictly positive entries, is unique and the corresponding eigenvalue is the largest eigenvalue of A in absolute value.*

In order to use theorem 2.1.1 to guarantee a solution, and thus a ranking, we must pick our entries of the matrix A sufficiently. Therefore, A must be nonnegative and irreducible. The condition that A must be nonnegative is easy to fulfill. This is easily justifiable, because a team should not lose points simply for showing up and participating in a game. The second condition, that A must be irreducible, is a bit more complex to fulfill.
In order to make sure our preference matrix A is irreducible, there can be no partition into two sets S en T such that for all i $\in$ S and for all j $\in$ T, $\mathrm{a_{ij}} = 0$. To understand this we could use the example we used earlier. We say $\mathrm{a_{ij}}$= 1 if team i won against team j, $\mathrm{a_{ij}}$=1/2 if it ended in a draw and is zero otherwise. Matrix A is irreducible if there is no partition of the teams into two sets S and T such that no team in S plays any team in T or every game between a team in S against a team in T results in a victory for the team in S.
If we succeed in accomplishing these assumptions for our preference matrix A, the Perron-Frobenius theorem tells us we get a solution. The challenge is to find a way to get our preference matrix as fair as possible. By this we mean to assign a score that reflects the rank of a team as best, according to the results that team played. Here the subjectivity of the ranking system comes in. Our earlier choice of the preference matrix is adequate if teams play each other often enough. The higher the number of games between teams, the better $\mathrm{a_{ij}}$ describes the relative strength between teams. In football, especially in the season we are interested in, one can argue that this is not the optimal choice for the entries of the matrix, because teams play each other once or twice in the season. This suggests that information will be lost when credit is given only for the win. For example, whether a score is nearly even or quite lopsided, all of the credit for the win goes to the winner.

On the other side, this is how rankings are calculated in a full season. A team will get a higher reward for a win (3 points), than for a draw (which is 1 point). Thus, a team will get more points if it wins and loses one time, than if it draws two times. Since we are evaluating the Dutch Football competition, which adopts this sort of scoring system, we will use this configuration for a_{ij}, for one of our ranking methods.
According to Keener, a better way to distribute the points is to do it in a continuous, rather than a discrete, way [4]. He does this by looking at the game score. If team i scored S_{ij} goals and team j scored S_{ji} goals in their encounter, he awards a score:

$$a_{ij} = h(\frac{S_{ij}+1}{S_{ij}+S_{ji}+2}) \tag{2.2}$$

$$h(x) = \frac{1}{2} + \frac{1}{2} sgn(x - \frac{1}{2})\sqrt{|2x-1|} \tag{2.3}$$

This will make sure that a team cannot run up a score to significantly improve their ranking. In order to see this we can look at the graph of h(x) in Figure 1.

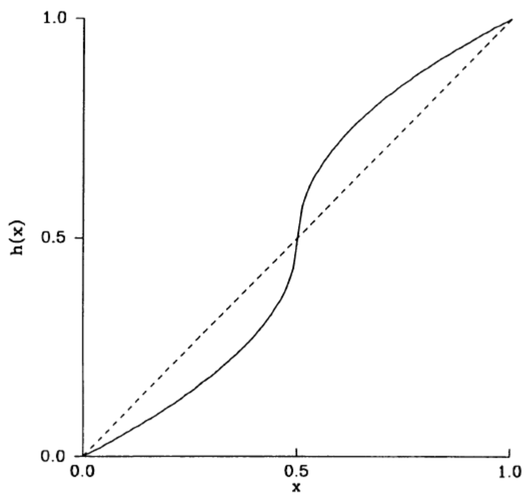

Figure 1: Plot of h(x) as a function of x (solid curve) and the line y(x)=x (dashed curve).

We can see that around x=0.5 we can quickly gain a higher score by scoring, but when we get closer to one, obtaining a higher score is more difficult. This means that if we already have a high score difference, one more goal would not get us a much higher score. If we are tied (meaning x=0.5), one more goal would get us a significant higher score.
To summarise, we can make up our method as follows:

1. Form the preference matrix A
2. Solve the following equation, with regards to r:

$$Ar = \lambda r \tag{2.4}$$

   with λ the highest eigenvalue of A.
3. Use rating scores in r to rank the teams

For the construction of the preference matrix A we will use the following methods, referred to as Keener I and Keener II in the remainder of this thesis:
Keener I: Here we will use the matrix A according to:

$$A_{ij} = \begin{cases} h(\frac{S_{ij}+1}{S_{ij}+S_{ji}+2}) & \text{if team i played team j} \\ 0 & \text{else} \end{cases}$$

With $h(x) = \frac{1}{2} + \frac{1}{2}sgn(x - \frac{1}{2})\sqrt{|2x - 1|}$
If teams played each other twice (or more) in the season, we will sum up the goals made in these encounters. For example, if team 1 and team 2 played each other twice in the season with the following results: 2-0, 1-2, than S_{12}=3 and S_{21}=2.
Keener II: This is our modification of the original Keener method. Here we will use the matrix A according to:

$$A_{\text{ij}} = \begin{cases} a_{\text{ij}} = a_{\text{ji}} = 1 & \text{if team i drew against j} \\ a_{\text{ij}} = 3 & \text{if team i won against team j} \\ a_{\text{ij}} = 0 & \text{else} \end{cases}$$

If teams played each other twice (or more) in the season, we will sum up the results accordingly. For example, if team 1 won once against team 2 and drew once, then a_{12}= 4 and a_{21} = 1.

2.2 PageRank

In this section, we will explain the second ranking method we used, the PageRank method. To understand and interpret the results of this ranking method, we will first go over some necessary theory of discrete Markov chains. Secondly, we will briefly go over the outline of the PageRank algorithm. Finally, we will explain how we used this method in the case of ranking the football teams in the Dutch Football competition.

2.2.1 Discrete Markov chains

This section will explain some necessary theory to understand the results from the PageRank algorithm. All the claims made here are proven in the course 'Stochastische processen'from Utrecht University [6]. The PageRank method makes use of the theory behind Markov chains. This section's purpose is to explain the convergence theorem of Markov chains. This theorem will help us understand why the PageRank method works. We will first explain some theory about Markov chains that is necessary to understand the convergence theorem. We will use the following definition for a Markov chain.

Definition 2.2.1. A (discrete) Markov Chain is, is a stochastic process $(X_n)_{n\geq 0}$ with the following two properties:

1. It is a chain, i.e., the X_n all take values in a countable set I. We call I the state space of $(X_n)_{n\geq 0}$ and each i ∈ I a state.

2. it statisfies the Markov Property, i.e., $P(X_{n+1} = i_{n+1}|X_0 = i_0, X_1 = i_1, ..., X_n = i_n) = P(X_{n+1} = i_{n+1}|X_n = i_n)$.

A Markov Chain is called homogeneous if for any i,j ∈ I the conditional probability does not depend on n. Under this standing assumption, a Markov chain is completely described by

1. the initial distrubion $\lambda = (\lambda_i)_{i \in I}$, given by $\lambda_i = P(X_0 = i)$

2. the transition matrix $P = (p_{ij})_{i,j \in I}$ given by $p_{ij} = P(X_1 = j|X_0 = i)$

One could imagine that it is useful to know what the probability is of moving from one state to another in n time steps. Since this is an important measure we will use the following definition for convenience.

Definition 2.2.2. We denote the n-step probability as

$$p_{ij}(n) := P(X_n = j|X_0 = i)$$

As we can see, $p_{ij}(n)$ gives us the probability of moving to state j in n time steps, if we start in state i. One could construct Markov chains in which if you leave a state i, the chain will never return to state i. To make this more precise and to investigate these kind of Markov chains, or in our case to make sure this is not the case in our Markov chain, we can make use of the following definition.

Definition 2.2.3. The first-passage time to state j is defined as:

$$T_j = min\{n \geq 1 | X_n = j\}$$

We call a state i recurrent if $\mathbb{P}_i(T_i < \infty) = 1$

With this definition, we can see that if a state i is recurrent and the Markov chain starts in this state, it is certain to come back to that state if n is large enough. It can happen that the Markov chain will settle into an equilibrium. For this to happen the equilibrium needs to have at least the following properties.

Definition 2.2.4. Let $(\mathrm{X_n})_{\mathrm{n}\geq 0}$ be a Markov chain with tranisition matrix P. A row vector $\pi = (\pi_\mathrm{i})_{\mathrm{i}\in \mathrm{I}}$ is called an invariant distribution if:

1. $\pi_\mathrm{i} \geq 0$ for all i in I and $\sum_{i\in I} \pi_\mathrm{i} = 1$
2. $\pi P = \pi$

If a Markov chain does have a unique invariant distribution, it is not necessarily that it converges to this distribution. Luckily the convergence theorem will tell us when it does, but for this we need a few more definitions. It is not enough for a Markov chain to be recurrent, but it needs to be positive recurrent. By this we mean the following.

Definition 2.2.5. Suppose i is a recurrent state of a Markov chain. Then we call i positive recurrent if its expected return time:

$$m_i = \mathbb{E}(T_i)$$

is finite.

One thing to note however, is that for an irreducible Markov chain with finite state space, every state is positive recurrent. The last definition we need to understand the convergence theorem is the definition of periodicity, which is stated as follows.

Definition 2.2.6. The period of a state i is defined as:

$$d_\mathrm{i} = gcd\{n \geq 1 | p_{ii}(n) > 0\}$$

where 'gcd 'means greatest common divisor. We call a state i aperiodic if $\mathrm{d_i}$=1

The convergence theorem now states that if we have a Markov chain that is irreducible, aperiodic and positive recurrent, then it will settle into a unique equilibrium. To be precise:

Theorem 2.2.7. *Let $(X_n)_{n\ \geq\ 0}$ be Markov(λ,P) and suppose $(X_n)_{n\geq}$ is irreducible, aperiodic and positive recurrent. Then for any i,j ∈ I:*

$$\lim_{n\to\infty} p_{ij}(n) = \pi_j$$

and as a consequence

$$\lim_{n\to\infty} \mathbb{P}(X_n = j) = \pi_j$$

where π denotes the unique invariant distribution of the chain.

In the PageRank method we will use this theorem to find an equilibrium. This theorem states the chance that a Markov chain is in a certain state if we go on forever, is the same as the corresponding entry of the unique invariant distribution. We can make sure that a unique invariant distribution exists and converges for our problem, which we will show later. We will use this unique invariant distribution as our ranking vector. This will be explained in the next sections.

2.2.2 Generalized PageRank

Google used this algorithm on the world wide web [7]. The idea of this algorithm comes from looking at the behaviour of a random surfer on the internet. With this behaviour, we will get an objective measure for the importance of webpages. We model the behaviour using a directed graph G=(N,E), with the nodes in N that represents the websites, which are connected by edges. A website u $\in$ N has an edge to a website v $\in$ N if u has a link to v. We say that at any given website, the random surfer picks one of the links on the website uniformly at random and clicks it. For our ranking we want to have some sort of objective measure, we do this by looking at the number of times the surfer visits the website if he/she keeps on surfing forever. By objective measure, we mean a measure which is independent of the relevance of the content. We want this, because we want to create a model of whether a website is popular or reliable.
To create this measure, we are going to let the random surfer perform a random walk on the graph G. We can model this by using Markov Chains. Using the notation in section 2.2.1, we can quickly see that the random walk on the graph G can be viewed as a Markov chain $(\mathrm{X_n})_{\mathrm{n}\geq 0}$ with transition matrix P given by

$$p_{\mathrm{ij}} = \begin{cases} \dfrac{1}{O(i)} & \text{if (i,j)} \in \text{E} \\ 0 & \text{else} \end{cases} \quad \text{With } O(i) = \|j \in N : (i,j) \in E\|$$

We can view $O(i)$ as the number of outgoing links on website i. However, in order for this to work there cannot be a website i with zero outgoing links. In that case, $\mathrm{p_{ij}}$=0 for all j $\in$ N and hence P does not define a stochastic matrix. A way to solve this problem is to set $p_{\mathrm{ij}} = \frac{1}{|N|}$ for all i $\in$ N.
First, we are going to assume that the web is G is connected in such a way that for any pair of nodes (i,j), there is a path connecting i to j. If this is the case, then :

1. $(\mathrm{X_n})_{\mathrm{n}\geq 0}$ is irreducible
2. Since we have a finite number of websites ($|N| < \infty$), there exists a unique invariant distribution π

We will not proof the first claim, since this can be directly done from using the definition. The proof for the second claim can be found in appendix C. The idea is to use this unique invariant distribution to rank the web: the higher π_{i}, the more important the website i should be considered. We saw by using the convergence theorem, the unique distribution can be seen as a distribution in which the Markov chain will eventually settle.
In reality, there were a few challenges to overcome for this to work properly for the web. As it turned out, the web was not really connected. There were two main problems with the state of the web in which this ranking method could not be applied. Firstly, there could be two unconnected subwebs. This means that the transition matrix is not irreducible. However, the method can produce a ranking for each of these subwebs, but there could be no unique ranking for the entire web. The second problem is that there could be cycles in the web, leading to periodicity in the Markov chain. Google has found a solution to these problems. Their idea was to alter the behaviour of the random surfer by inventing a parameter α and do the following at each step:
- With probability α the surfer will act as before.
- With probability 1-α the surfer will 'teleport' to a uniformly randomly chosen website on the internet.
If we let T be the matrix with $\mathrm{T_{ij}}=\frac{1}{|N|}$ for all i,j $\in$ N. It is clear that we have a new transition matrix:

$$\widetilde{P} = \alpha P + (1-\alpha)T. \tag{2.5}$$

Since we have a positive probability to 'jump'to any other website from every website, it is clear that this web is connected. We can see that we have a unique invariant distribution $\widetilde{\pi}$, which is called the PageRank vector. Decreasing the parameter α can make convergence faster, but also decreases the quality of the ranking. It is reported that Google set α to be 0.85 [7].

2.2.3 PageRank for Football

This idea of the PageRank method can be used to calculate a ranking for football teams. We have done that by using the idea of Govan, A. Y., Meyer, C. D., and Albright, R. [5]. They used the PageRank algorithm to come up with a ranking for American Football teams. We will use their modification of the PageRank algorithm and use it on the Dutch Football competition. To illustrate how this works, we can think of an analogous reasoning as in 2.2.2. This time instead of the random surfer on the internet, we will look at the behaviour of a disloyal fan. We model the behaviour using a weighted graph with n nodes, in our case n = 18. Each team represents a node and each game represent an edge from loser to winner with weight equal to the positive score difference. Note that if a game ended in a draw, no score is given and thus no edge is made. We say that for any given team the disloyal fan supports, the fan will support a new team, chosen from the links that are made. If a team has lost against two other teams, one with a score difference of 2 and the other with a score difference of 3, the disloyal fan has a 2/5 chance to support the first team and a 3/5 chance to support the second team. From this web, we make our transition matrix P and calculate the invariant distribution π. We can interpret this invariant distribution as the percentage of time our disloyal fan supports a team. The higher percentage of time our disloyal fan supports a team, the higher this team is ranked.

If we have enough information of our data, we can determine if such an invariant distribution exists. If our matrix P satisfies the assumptions in theorem 2.2.7, we know our Markov chain will converge to the unique invariant distribution. In some cases, for example, when there are undefeated teams in the competition, this is not the case. It can also happen that you simply do not know if the assumptions are fulfilled. If this happens, we simply use the idea google used. We set, for the undefeated team i, $p_{\text{ij}} = \frac{1}{|N|}$ for all i $\in$ N. Next, we make our matrix T accordingly to the way we described and calculate $\widetilde{\pi}$ to rank the teams. To sum it up, the algorithm contains the following steps:

1. Make the directed graph and form the corresponding matrix A.

$$A_{\text{ij}} = \begin{cases} w_{\text{ij}} & \text{if team i lost to team j} \\ 0 & \text{else} \end{cases}$$

   Where w_{ij} is the positive score difference between the match of team i and j. If team i lost to team j more than once during the season, w_{ij} is the sum of the positive score difference of the games between i and j lost by i.

2. (a) Form the transition matrix P:

$$P_{\text{ij}} = \begin{cases} \dfrac{w_{ij}}{\sum_{k=1}^{n} A_{ik}} & \text{if team i lost to team j} \\ 0 & \text{else} \end{cases}$$

   (b) Form the transition matrix $\widetilde{P}$:
   $\widetilde{P} = \alpha(\text{P} + \text{U}) + (1-\alpha) * T$

   With P as describe in 2.a, $T_{\text{ij}} = \frac{1}{|N|}$ for all i,j $\in$ N, I the identity matrix and U described as:

$$U_{\text{ij}} = \begin{cases} \dfrac{1}{|N|} & \text{for all j, if } P_{\text{ij}} = 0 \text{ for all j} \\ 0 & \text{else} \end{cases}$$

3. (a) Find the unique invariant distribution using:

$$\pi^T = \pi^T P$$

   Use rating scores in π to rank teams

   (b) Find the unique invariant distribution using:

$$\widetilde{\pi}^T = \widetilde{\pi}^T \widetilde{P}$$

   Use rating scores in $\widetilde{\pi}$ to rank teams

Since our matrix P will be irreducible for our football competitions, we will use the 'a-variant'in the algorithm.

2.3 Comparison between PageRank and the Keener method

There are some similarities between the PageRank method and the Keener method. This can be seen when we set $\alpha = 1$ in the PageRank method. If we do this, we are interested in the following vector to determine our ranking :

$$\pi^T = \pi^T P \tag{2.6}$$

If we look at the Keener method, we can deduce that:

$$(Ar)^T = r^T A^T = \lambda r^T \tag{2.7}$$

P is a stochastic matrix, meaning that if we normalize the rows of $\mathrm{A^T}$, we see that $\mathrm{A^T}$=P will yield the same ranking vector. In this case we see the direct link between the weights we have set for each match of every team and the score we assigned to the teams for each match.
The PageRank method has one big advantage. If we set $\alpha < 1$, this method will guarantee us a ranking. For the Keener method this is not the case. Here we need to choose our entries in such a way that we have an irreducible matrix.

3 Results

The methods described above are applied to the Dutch football competition in the season of 2018-2019 and 2019-2020. As mentioned, the season of 2019-2020 was not finished and because of this we can view this as an uneven paired competition. We collected the data for the season of 2018-2019 and 2019-2020 from FC Update [8,9]. We used three different ranking methods on these data sets. The methods represented in the data are as follows:

1. The original ranking, the one that the KNVB used.
2. The method as defined as Keener I in section 2.2.2.
3. The method as defined as Keener II in section 2.2.2.
4. The PageRank method for football as described in the algorithm 2.a in section 2.2.3.

In table 1 the results of the Dutch Football competition of the season 2019-2020 are shown. We can see all the rankings of the different methods and the difference between that ranking and the ranking that was used by the KNVB for this season (the first ranking method). One thing that immediately becomes clear is that AZ is ranked first in all three different ranking methods. To explain this, we can look at some of the results from AZ. AZ defeated Ajax, the second highest rated team, twice. All the ranking methods take the strength of schedule into account. Winning twice from one of the best teams, gives you very high 'points', thus likely being rated higher than other teams. Hence, it is more efficient to win against one of the best teams than to win against a lower rated team, because it will give you a higher score. If you compare this to Ajax, although Ajax has won the same amount of games, they did it against lesser oppositions. This is one of the reasons AZ comes up on top of Ajax.
Another thing that stands out is that the ranking from the PageRank methods is very different in comparison with the other rankings. One way to explain this is that, with the PageRank method, a win is rated very highly, especially against high rated oppositions. Since the only way a link is made between two teams is if one team wins, one can see that this is crucial to obtain a high rating. This can also be seen in table 2, where Ajax and AZ scored a very high rating.

Dutch Footbal Competition season 2019-2020							
Team	**#1**	**#2**	up/down	**#3**	up/down	**#4**	up/down
Ajax	1	2	-1	2	-1	2	-1
AZ	2	1	+1	1	+1	1	+1
Feyenoord	3	3	-	3	-	7	-4
PSV	4	4	-	4	-	4	-
Willem II	5	7	-2	5	-	3	+2
Utrecht	6	6	-	7	-1	6	-
Vitesse	7	5	+2	6	+1	9	-2
Heracles	8	8	-	9	-1	11	-3
Groningen	9	10	-1	8	+1	12	-3
Heerenveen	10	9	+1	10	-	10	-
Sparta	11	11	-	12	-1	5	+6
Emmen	12	13	-1	11	+1	13	-1
VVV	13	17	-4	14	-1	16	-3
Twente	14	12	+2	13	+1	8	+6
PEC	15	15	-	16	-1	17	-2
Fortuna	16	14	+2	15	+1	14	+2
ADO	17	16	+1	17	-	18	-1
RKC	18	18	-	18	-	15	+3

Table 1: Results of the three different ranking methods compared with the original ranking for the season 2019-2020. The up/down column indicates the difference with the original ranking for each team.

Dutch Footbal Competition season 2019-2020					
Team	**#1**	%	**#2** %	**#3** %	**#4** %
Ajax	56	8.66 %	7.97 %	8.30 %	11.55 %
AZ	56	8.66 %	8.28 %	8.71 %	14.72 %
Feyenoord	50	7.73 %	6.73 %	7.88 %	6.01 %
PSV	49	7.57 %	6.69 %	7.38 %	6.83 %
Willem II	44	6.80 %	6.24 %	7.27 %	9.75 %
Utrecht	41	6.34 %	6.26 %	6.28 %	6.28 %
Vitesse	41	6.34 %	6.39 %	6.42 %	5.48 %
Heracles	36	5.56 %	5.68 %	5.45 %	4.97 %
Groningen	35	5.41 %	5.55 %	5.94 %	4.91 %
Heerenveen	33	5.10%	5.66 %	5.11 %	5.14 %
Sparta	33	5.10 %	5.25 %	4.68 %	6.37 %
Emmen	32	4.95 %	4.78 %	5.01 %	3.72 %
VVV	28	4.33 %	3.64 %	4.27 %	1.99 %
Twente	27	4.17 %	4.90 %	4.53 %	5.49 %
PEC	26	4.02 %	4.14 %	3.64 %	1.40 %
Fortuna	26	4.02 %	4.32 %	3.90 %	2.25 %
ADO	19	2.94 %	4.02 %	2.71 %	1.04 %
RKC	15	2.32 %	3.50 %	2.52 %	2.08 %

Table 2: Results of the three different ranking methods compared with the original ranking for the season 2019-2020. The percentages indicate how much each team has scored compared to one another.

In table 2 the total amount of points relative to one another is shown. We see that the points with the PageRank method are much more separated than the points from the other methods. With the PageRank method, AZ got a total of 14,72% of the points. The lowest scored team got a total of 1.04%. We can explain this with the fact that draws do not count towards the points, and because AZ only got two draws and ADO

7, the difference is that high. With the other methods we see that the range is somewhat similar than the original football ranking. The original ranking differs from 2.32% to 8.66% and the keener methods are in the range between 3.50%-8.28% and 2.52%-8.71%.

Dutch Footbal Competition season 2018-2019							
Team	#**1**	#**2**	up/down	#**3**	up/down	#**4**	up/down
Ajax	1	2	-1	1	-	1	-
PSV	2	1	+1	2	-	3	-1
Feyenoord	3	3	-	3	-	2	+1
AZ	4	5	-1	4	-	4	-
Vitesse	5	4	+1	5	-	13	-8
Utrecht	6	7	-1	6	-	8	-2
Heracles	7	8	-1	7	-	6	+1
Groningen	8	9	-1	10	-2	14	-6
ADO	9	6	+3	8	+1	5	+4
Willem II	10	10	-	9	+1	7	+3
Heerenveen	11	11	-	11	-	15	-4
VVV	12	13	-1	14	-2	16	-4
PEC	13	12	+1	12	+1	9	+4
Emmen	14	14	-	13	+1	17	-3
Fortuna	15	17	-2	16	-1	10	+5
Excelsior	16	15	+1	15	+1	12	+4
De Graafschap	17	16	+1	17	-	11	+6
NAC	18	18	-	18	-	18	-

Table 3: Results of the three different ranking methods compared with the original ranking for the season 2018-2019. The up/down column indicates the difference with the original ranking for each team.

In table 3 we can see the methods being used on a full football competition. This means that every team has played each other twice, one time at home and the other away. We can see that with the PageRank method, there are a lot of changes. For example, De Graafschap is rated higher than Groningen, even though in the original ranking their score difference is 16. Since we want our ranking method to be similar to the original ranking for a full season, this shows that the PageRank method used in this way, is not a good ranking method for football.

With the Keener I method we see some shuffle of the ranking as well. The most notable is ADO. In the original ranking they scored 45 points, but in the Keener I method they are ranked higher than Utrecht, which has 53 points. The fact that they won 5-0 against Utrecht and had some good results against other top teams may have contributed to this fact. This method looks at score difference between matches, and it seems that, although the ranking is similar to the original ranking, some teams are counter intuitively ranked higher than others.

The Keener II method, seems to fit the best with the original ranking. We can see that ten teams have the same ranking as in the original ranking. The standard deviation of the difference between the position in the ranking of this method and the position in the original ranking is $\sigma_3 = \sqrt{\frac{7}{9}}$. This is the lowest of all the rankings, for Keener I it is $\sigma_2 = \sqrt{\frac{4}{3}}$ and for PageRank it is $\sigma_4 = \sqrt{\frac{133}{9}}$. We can see that there are some minor shuffles in this ranking compared to the original ranking. The biggest difference in points is three. Emmen is rated higher than VVV and they have 41 and 38 points respectively. The other shuffels are 1 point apart, or the teams have the same amount of points.

4 Conclusion

We have used three different ranking methods on the Dutch football competition for the seasons 2018-2019 and 2019-2020. For the fully completed season 2018-2019, we saw that the results of the PageRank method were very different from the original method from the KNVB ($\sigma_4 = \sqrt{\frac{133}{9}}$). Because of this fact, we do not consider this ranking method as very suitable for the ranking of football teams. The Keener II method was most similar to the original ranking method in the season 2018-2019, with $\sigma_3 = \sqrt{\frac{7}{9}}$.
For the incomplete season 2019-2020 we see that the results of the Keener II method are comparable to the KNVB ranking. The maximum difference in ranking between the teams in both ranking methods is one position. The most important difference, however, is that AZ is placed first. Since this is the case in all the ranking methods we used, we conclude that AZ has performed better in the season of 2019-2020 than Ajax.

5 Discussion

We looked if the ranking methods could produce a ranking similar to the KNVB ranking for a full competition. Next, we used these ranking methods to determine the ranking for the unfinished season. For future work one may use these methods as a ranking mechanism for a full competition. The advantage of these methods is that we can use all kinds of statistics to result in a desired ranking. In this way we can manipulate these ranking methods in such a way that we can "force" teams to play a certain type of football. For example, if we want teams to play a more attacking style of football, we could give a higher amount of points for goals scored than goals conceded. Then it is more beneficial to score more goals, hence the teams will most likely play a more attacking style of football.
If one uses one of these ranking methods to deliver the ranking during a season, one needs to be aware of the fact that the rankings can differ a lot from match to match. In the original way, a team gets 3 points if it wins, no matter how highly the opposition is rated. In these ranking methods, the rating of the teams matter for the increase in rating a team may get. This means that if a team wins against the current number one, it gets a higher ranking than if that team would win against a (much) lower ranked team.

A Equivalent definitions of an irreducible Matrix

Here we will proof our claim that our definitions of an irreducible matrix are equivalent.

Proposition A.0.1. *The following statements are equivalent:*

1. *A is irreducible if for any two numbers i and j there is an integer $p \geq 0$ and a sequence of integers $k_1,....,k_p$, so that the product $a_{ik_1}a_{k_1k_2}...a_{k_pj} \neq 0$.*
2. *A is irreducible if there is no permutation that transforms the matrix A into a block matrix of the form:*

$$\begin{pmatrix} A_{11} & A_{12} \\ 0 & A_{22} \end{pmatrix}$$

With A_{11} and A_{22} square matrices

Proof. (1) ⇒ (2):
Suppose there is a permutation that transforms the matrix A into a block matrix of the form:

$$\begin{pmatrix} A_{11} & A_{12} \\ 0 & A_{22} \end{pmatrix}$$

With A_{11} and A_{22} square matrices.
We will show that this leads to a contradiction. If we can write the matrix in this form using permutations, we see that we can make a union of two disjoint sets $V=V_1 \cup V_2$ for which $a_{ij}=0$ for all $i \in V_1$, $j \in V_2$. From definition (1), we know that for any numbers i and j, there is an integer $p \geq 0$ and a sequence of integers $k_1,....,k_p$, such that the product $a_{ik_1}a_{k_1k_2}....a_{k_pj} \neq 0$. In particular there are numbers $w \in V_1$ and $z \in V_2$ for which there is an integer $p \geq 0$ such that, $a_{wk_1}a_{k_1k_2}....a_{k_pz} \neq 0$. Since V_1 and V_2 are disjoint sets, there must be an i such that $a_{k_ik_{i+1}}$ is in the product, with $k_i \in V_1$ and $k_{i+1} \in V_2$. Since $a_{k_ik_{i+1}}=0$, the whole product must be equal to zero and this leads to our contradiction and we are done.

(2) ⇒ (1)
We will proof this by proving ¬ (1) ⇒ ¬ (2). Assume there exists two numbers i and j such that for all $p \geq 0$ and a sequence of integers $k_1,...,k_p$, the product $a_{ik_1}a_{k_1k_2}...a_{k_pj}=0$. We now define two subsets $V_1 := \{b \in I | a_{bk_1}a_{k_1k_2}...a_{k_pj} = 0\}$ and $V_2 := \{d \in I | a_{dk_1}a_{k_1k_2}...a_{k_pj} \neq 0\}$ for all $p \geq 0$ and a sequence of integers $k_1,...,k_p$. Here $I=\{1,...,n\}$ with $A \in R^{nxn}$. Clearly these subsets are disjoint. We first note that, by our assumption, $i \in V_1$. Now we have two cases:

1. $j \in V_2$:
   We note that if $b \in V_1$ and $d \in V_2$, that for all $p \geq 0$ and integers $k_1,..,k_p$, the product $a_{bk_1}a_{k_1k_2}....a_{k_pd} = 0$. If this was not the case, we could simply take $a_{bk_1}...a_{k_pd}a_{dk_{p+1}}....a_{k_{\tilde{p}}j} \neq 0$, contradicting that $b \in V_1$. Now we can make a disjoint union of sets $V=V_1 \cup V_2$ for which $a_{ij}=0$ for all $i \in V_1$ and $j \in V_2$.
2. $j \notin V_2$:
   We define $W_1 := V_1 \setminus \{j\}$ and $W_2 := V_2 \cup \{j\}$. Using the same argument as the previous case, but with replacing v_1 with W_1 and V_2 with W_2, we see that we can make a union of sets $V=W_1 \cup W_2$ for which $a_{ij}=0$ for all $i \in W_1$ and $j \in W_2$.

In both cases we can make a permutation that transforms the matrix A into a block matrix of the form:

$$\begin{pmatrix} A_{11} & A_{12} \\ 0 & A_{22} \end{pmatrix}$$

With A_{11} and A_{22} square matrices. This completes the proof.

□

B Perron-Frobenius theorem

Lemma B.0.1. *The set of all nonnegative vectors with Euclidean norm 1 is closed and bounded.*

Proof. Let A=$\{(x_1, ..., x_n) \in \mathbb{R}^n | x_i \geq 0 \ \forall i, x_i^2 + ... + x_n^2 = 1\}$
We can see that A $= S^{n-1} \cap X$, with S the n-sphere and $X = \{(x_i, ..., x_n) \in \mathbb{R}^n | x_i \geq 0 \ \forall i\}$.
We know that S^{n-1} is closed and bounded, and X is closed. Therefore A is closed and bounded. This concludes the proof.

□

Lemma B.0.2. *Let $s \in A = \{(x_1, ..., x_n) \in \mathbb{R}^n | x_i \geq 0 \ \forall i, x_i^2 + ... + x_n^2 = 1\}$. For each vector s in the set A let σ_s^* be the positive number for which $As \leq \sigma s$ whenever $\sigma \geq \sigma_s^*$. Then the smallest value of σ_s^* is attained for some s^* in A.*

Proof. We define $\sigma_s^* := inf\{\sigma \in \mathbb{R}_{>0} | As \leq \sigma s\}$. The map

$$\begin{aligned} \phi : \mathbb{R}^n &\to \mathbb{R}_{>0} \\ s &\to \sigma_s^* \end{aligned}$$

is continuous. We know that $\mathbb{R}^n$ is a metric space and according to lemma B.0.1., set A is closed and bounded, hence compact. Since ϕ is continuous, it follows that $\phi(A)$ is compact. It follows that

$$\sigma^* := inf\{\sigma_s^* \in \mathbb{R}_{>0} | s \in A\} \in \overline{\phi(A)} = \phi(A).$$

We can conclude that there exists a $s \in A$ such that $\sigma_s^* = \sigma^*$.

□

Lemma B.0.3. *If A is irreducible, then every nonnegative eigenvector is strictly positive.*

Proof. We will proof this by proving if there is a nonnegative eigenvector which is not strictly positive, then A is not irreducible.
Assume there is a nonnegative eigenvector which is not strictly positive (e.g. Ax= λx with x $\geq$ 0, x is not strictly positive.). Then there is a permutation P and k $\in \mathbb{N}$ such that

$$P^{\mathrm{T}} x = \begin{pmatrix} x_1 \\ 0 \end{pmatrix} \text{ where } 0 < x_1 \in \mathbb{R}^k$$

We note that $P^T A P P^T x = P^T A x = P^T \lambda x = \lambda P^T x$. Now take

$$P^{\mathrm{T}} A P = \begin{pmatrix} B_{11} & B_{12} \\ B_{21} & B_{22} \end{pmatrix}$$

It follows that:

$$\begin{pmatrix} B_{11} & B_{12} \\ B_{21} & B_{22} \end{pmatrix} \begin{pmatrix} x_1 \\ 0 \end{pmatrix} = \lambda \begin{pmatrix} x_1 \\ 0 \end{pmatrix}$$

thus B_{21}*x_1 = 0. Since x_1>0, it must be that B_{21}=0. We conclude that A is reducible and this completes the proof.

□

Theorem B.0.4. *If the matrix A is irreducible and has nonnegative entries, then there exists an eigenvector r with strictly positive entries, is unique and the corresponding eigenvalue is the largest eigenvalue of A in absolute value.*

Proof. We will first proof that the eigenvector has strictly positive entries. Let S be the set of all non-negative vectors with Euclidean norm one. For each vector s in the set S let σ_s^* be the positive number for which $As \leq \sigma s$ whenever $\sigma \geq \sigma_s^*$. According to lemma B.0.1 and B.0.2, the smallest value of σ_s^*, is attained for some s^* in S. We call this value σ^*. We claim that s^* is a positive eigenvector of A.
Suppose that $As^* \leq \sigma^* s^*$ and suppose s^* is not an eigenvector of A. If this is the case, then some, but not all, of the relations in the statement $As^* \leq \sigma^* s^*$ are equalities (if this is not the case, the number σ^* is incorrectly

chosen). If we separate these inequalities via permutation, then we can write the relations $As^* \leq \sigma^* s^*$ in the form

$$A_{11}s_1^* + A_{12}s_2^* < \sigma^* s_1^* \tag{B.1}$$

$$A_{21}s_1^* + A_{22}s_2^* = \sigma^* s_2^* \tag{B.2}$$

We know that A is irreducible, so A_{21} is not identically zero. Because of this, we can reduce a component of the vector s_1, in such a way that at least one of the equalities changes to a strict inequality, without changing any of the original strict inequalities. After this change, we can rescale our newly made s^* to have norm one. If we proceed this process inductively, we can continue until all of the relations in $As^* \leq \sigma^* s^*$ are strict inequalities. This is of course a contradiction for our definition of σ^*. We conclude that s^* is a nonnegative eigenvector and because of lemma B.0.3, this eigenvector is strictly positive.

Now we are going to prove the uniqueness of the eigenvector. We first note that a nonnegative eigenvector r must have all positive entries, according to lemma B.0.3. Assume we have two linearly independent eigenvectors of A, statisfying $Ar_1 = \lambda_1 r_1$ and $Ar_2 = \lambda_2 r_2$. We take r_1 as strictly positive eigenvector. If the entries of r_2 are of one sign, then without loss of generality they can be taken as positive. For a nonnegative irreducible matrix A and non zero vector $x \geq 0$, we have $Ax > 0$ (see theorem A.2 (f) and (h) in Abramov (2014) [10]). We are going to look at the vector $r(t) = r_1 - tr_2$. We define $t_0 := max\{\frac{r_1^i}{r_2^i} | r_2^i > 0\}$ with r_1^i the ith-component of vector r_1. We note that r_2 always has at least one component larger than zero and r_1 is strictly positive, hence $t_0 > 0$. We can see that for $0 \leq t \leq t_0$, we have $r(t) \geq 0$. We see that $r(t_0)$ has some zero entries, but is not identically zero. For $t > t_0$ it follows that r(t) has some negative entries. Then $Ar(t_0) = \lambda_1(r_1 - t_0 \frac{\lambda_2}{\lambda_1} r_2)$ is positive. By the maximality of t_0 we conclude that $|\lambda_2| < |\lambda_1|$. If r_2 would be strictly positive as well, then we could interchange r_1 and r_2 in the above argument and conclude that $|\lambda_1| < |\lambda_2|$. This is of course a contradiction. We conclude that the positive eigenvector is unique and all other eigenvectors have eigenvalues smaller that are smaller in absolute value.

□

C Existence of the unique distribution π

We want to proof the following statement:
If $(X_n)_{n \geq 0}$ is an irreducible Markov chain with finite state space, then there exists an invariant distribution π .
Proof:
Assume we have an irreducible Markov chain with finite state space I and transition matrix P. Since P is a stochastic matrix, each of it rows sums to 1. This means that the all ones vector $\vec{1} = (1, ..., 1) \in R^n$ is an eigenvector corresponding to the eigenvalue 1. P and P^T have the same eigenvalues, thus 1 is also an eigenvalue of P^T. We have:

$$P^T \lambda = \lambda \tag{C.1}$$

for some $\lambda \in R^n$.
What is left to prove is to show that $\lambda_i \geq 0$ (or $\lambda_i \leq 0$, but by considering - λ instead of λ if necessary, this is the same) for all $i \in I$. There are numerous ways to do this. One way is to proof that 1 is the largest eigenvalue and using the Perron-Frobenius theorem. I will do this via another method, based on the proof from S. Dirksen in his lecture notes [6].
Let $Q = P^T$ and $I = \{1, ..., n\}$. Since P is nonnegative and has rows summing to 1, Q is nonnegative and has columns summing to 1. Assume that $Q\lambda = \lambda$ and that there are some k, k' such that $x_k < 0$ and $x_{k'} > 0$. Since we have an irreducible Markov chain, there exists some $n > 0$ such that $p_{kk'}(n) = (P^n)_{kk'} = (Q^n)_{kk'} > 0$. If we set $M = Q^n$, then $Mx = x$. In particular,

$$x_{k'} = \sum_{j=1}^{n} M_{k'j} x_j = M_{k'k} x_k + \sum_{j \neq k} M_{k'j} x_j. \tag{C.2}$$

Since $x_{k'}$ is positive and $M_{k'j}x_k$ <0, we find

$$\sum_{j \neq k} M_{k'j} x_j > 0. \tag{C.3}$$

Using this we can see that:

$$x_{k'} = |x_{k'}| = |\sum_{j=1}^{n} M_{k'j} x_j| < \sum_{j=1}^{n} M_{k'j} |x_j|. \tag{C.4}$$

This leads to the following contradiction:

$$\sum_{i=1}^{n} |x_i| < \sum_{i=1}^{n} \sum_{j=1}^{n} M_{ij} |x_j| = \sum_{j=1}^{n} \sum_{i=1}^{n} M_{ij} |x_j| = \sum_{j=1}^{n} |x_j|. \tag{C.5}$$

We can conclude that the sign of all the entries of λ must the same.
We can see that the following vector:

$$\pi^T = \frac{\lambda}{\sum_{j=1}^{n} \lambda_j} \tag{C.6}$$

defines an invariant distribution.

References

[1] Goal (2020, October 1) Goal difference or head to head? How every major football competition ranks teams level on points https://www.goal.com/en/news/goal-difference-or-head-to-head-how-every-major-football/1jax9vfriz1xs13jkdpf9qzhjo

[2] KNVB (2020, October 1) AFWIKKELING VOETBALSEIZOEN 2019/’20 EEN FEIT https://www.knvb.nl/nieuws/betaald-voetbal/competitiezaken/59905/afwikkeling-voetbalseizoen-2019%E2%80%9920-een-feit

[3] AjaxShowTime (2020, October 1) Perez over AZ-beroep: ’Ajax had precies hetzelfde gedaan’ https://www.ajaxshowtime.com/article/opinie/133167/perez-over-az-beroep-ajax-had-precies-hetzelfde-gedaan

[4] Keener, J. P. (1993). The Perron-Frobeius theorem and the ranking of football teams. SIAM Review, 35(1), 80-93. doi: 10.1137/1035004

[5] Govan, A. Y., Meyer, C. D., & Albright, R. (2008). Generalizing Google’s PageRank to rank national football league teams. Proceedings of the SAS Global Forum. 1-11

[6] Dirksen, S. (2020, October 4) Stochastic Processes (WISB 362) - Lecture Notes 2019/2020 , Utrecht University

[7] Sergey Brin, Lawrence Page, The anatomy of a large-scale hypertextual Web search engine, Computer Networks and ISDN Systems, 30: 107-17, 1998.

[8] FC Update (2020, October 5) Eredivisie uitslagen 2018-2019. https://www.fcupdate.nl/uitslagen/s1398/nederland-eredivisie-2018-2019/

[9] FC Update (2020, October 5) Eredivisie uitslagen 2019-2020. https://www.fcupdate.nl/uitslagen/s1398/nederland-eredivisie-2019-2020/

[10] Abramov, A. P. (2014). Balanced and Cyclical Growth in Models of Decentralized Economy. Springer International Publishing.