The Minnesota Journal of Undergraduate Mathematics

Measurement and Comparison of Passing Networks in Collegiate Soccer

Andre P. Oliveira and Helene R. Tyler

Manhattan College

The Minnesota Journal of Undergraduate Mathematics

Volume 1 Number 1 (2015)

Sponsored by
School of Mathematics
University of Minnesota
Minneapolis, MN 55455

The Minnesota Journal of Undergraduate Mathematics

Volume 1 Number 1 (2015)

Measurement and Comparison of Passing Networks in Collegiate Soccer

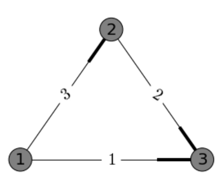
Andre P. Oliveira and Helene R. Tyler

Manhattan College

Abstract. In this work we interpret the vertices of a weighted directed graph as players on a soccer team. A weighted directed arrow between two vertices represents the number of successfully completed passes between the corresponding players. We use various centrality measures, both established and new, to study the Manhattan College women's soccer team, the Jaspers. We also utilize a Borda Count voting system to compare the play style of the Jaspers to those of several professional teams.

1. Background

The passing network of a soccer team is a weighted directed graph in which each of the eleven vertices corresponds to a player. A weighted directed arrow from vertex i to vertex j represents the number of successfully completed passes from player i to player j. This gives rise to the weighted adjacency matrix A, where A_{ij} is the number of passes from player i to player j. For illustrative purposes, we portray in Figure 1 the passing network and associated matrix for a team of three players. Arrowheads are represented by heavier rule at one end of an edge. Note, for example, that the arrow from player 1 to player 2 has a weight of 3 and that $A_{1,2} = 3$, both indicating that player 1 successfully passed to player 2 three times.



(a)

$$\begin{bmatrix} 0 & 3 & 1 \\ 0 & 0 & 2 \\ 0 & 0 & 0 \end{bmatrix}$$

(b)

Figure 1. A passing network for a team with 3 players and its associated adjacency matrix.

We wish to quantify both player and team performance. To do this we use five network invariants: closeness, betweenness, PageRank, clustering, and defensive score. The first three were used by Peña and Touchette in their analysis of the 2010 Men's World Cup Soccer Tournament [6]; they are discussed in depth in [7, chapters 7 and 8]. The final four teams in the 2010 Men's World Cup were Germany, Netherlands, Spain, and Uruguay. Three of these teams displayed notably high average clustering scores. In addition, Spain, the winner of the tournament, had uniformly low betweenness scores, which Peña and Touchette observed as evidence of a well-balanced passing strategy.

2. Centrality Measures

Two players on a team are considered "close" to one another if numerous successful passes flow between them. We define the length of an arrow in a passing network as the inverse of its weight. The geodesic distance from player i to player j, denoted d_{ij}, is given by the length of a shortest path from vertex i to vertex j, where the length of a path is obtained by adding the lengths of the arrows comprising it. If there is no path from vertex i to vertex j, then the distance from i to j is infinite. Looking again at Figure 1, we see that the arrow from player 1 to player 3 has length 1. However, the path from player 1 to player 3 that goes through player 2 has length $\frac{5}{6}$, making this the shorter path. It is worth noting here that distance between players is not necessarily symmetric. We have seen that $d_{13} = \frac{5}{6}$, but d_{31} is infinite, since there is no sequence of passes from player 3 to player 1. In the passing network of a collegiate soccer team, however, it is unlikely that a pair of players will have infinite distance between them.

Formally now, the **closeness score** of player i, denoted by clo(i), is defined as the inverse of her average geodesic distance to and from every other player. Since for each player we can consider paths to or from each of the 10 other players on the team, we have

$$\text{clo}(i) = \frac{20}{\sum_{j \neq i} d_{ij} + \sum_{j \neq i} d_{ji}}.$$

Another important attribute of an individual player is her ability to help the team move the ball from one player to another in the shortest amount of passes. The **betweenness score** of player i, denoted by bet(i), quantifies how often a player lies on a shortest path between other players. If n^i_{jk} denotes the number of geodesic paths from j to k going through i and g_{jk} gives the total number of geodesic paths from j to k, then

$$\text{bet}(i) = \frac{1}{90} \sum_{j \neq k \neq i} \frac{n^i_{jk}}{g_{jk}}.$$

The factor 1/90 normalizes the score since there are 90 possible (j,k) pairs. Returning to our simplified example in Figure 1, our normalization factor would be $\frac{1}{2}$, giving us $\text{bet}(1) = \text{bet}(3) = 0$ and $\text{bet}(2) = \frac{1}{2}$.

Note that closeness and betweenness are similar notions, but there is a subtle difference between them. Closeness measures how easy it is to get the ball to or from a particular player, while betweenness measures the extent to which a player helps her teammates move the ball between other players. Ideally, one would want every player on a team to have a high closeness score. This would mean that all players are well-connected, making it easy for the team to move the ball from player to player. To interpret a player's betweenness score, we view it holistically. While it is not bad for an individual player to have a high betweenness score, it may harm the team if she is an outlier. If she is, this may indicate that the team relies too much on that particular player. A more even distribution of betweenness scores, even if none is particularly high, suggests a more cohesive team. A more balanced team could lose a player to injury, penalty, or good defense and not have its performance suffer.

A player is said to be "important" if other "important" players pass to her. This recursive notion of importance is measured by **PageRank** centrality, which is the basis of the algorithm by which Google initially ranked webpages [3]. In the original context, a webpage is deemed important if other important pages link to it. The PageRank of a player i, denoted by $\text{Page}(i)$, satisfies

$$\text{Page}(i) = p \sum_{j \neq i} \frac{A_{ji}}{\deg^{out}(j)} \text{Page}(j) + q,$$

where $\deg^{out}(j)$ is the total number of passes made by player j, also called the **weighted out-degree** of player j. Likewise, $\deg^{in}(j)$ denotes the **weighted in-degree** of player j and counts the number of passes received by player j. The weighted in-degree will be used in our next measure. In the formula for PageRank, note that the coefficient $\frac{A_{ji}}{\deg^{out}(j)}$ is the percentage of passes player j makes to player i, so $\frac{A_{ji}}{\deg^{out}(j)}\text{Page}(j)$ computes the percentage of player j's importance that she gives to player i. The parameter p represents the probability that a player will keep the ball versus passing it to another player. By assigning a positive value to q, we ensure that each player has a positive PageRank. We have used the values declared by Peña and Touchette of $p = 0.85$ and $q = 1$ in order to compare our results to theirs. More accurate scores would result from using player dependent probabilities, but such information was not available to us. For a thorough, yet accessible explanation of the linear algebra involved in the computation of PageRank, which includes an explanation of the necessity of the parameter q, see [4].

The notion of clustering originates in undirected, unweighted graphs. In such a graph, the **clustering coefficient** of a vertex i, denoted by $\text{clust}(i)$, gives the probability that a pair of randomly selected neighbors of i are adjacent to each other. Hence,

$$\begin{aligned}\text{clust}(i) &= \frac{\text{the number of pairs of neighbors of } i \text{ that are adjacent}}{\text{the number of pairs of neighbors of } i}\\ &= \frac{2t_i}{\deg(i)(\deg(i)-1)},\end{aligned}$$

where t_i is the number of triangles involving vertex i and $\deg(i)$ is the degree of i.

Defining clustering coefficients for a weighted directed graph is more complicated. In fact, there is no consensus in the literature concerning how this should be done. A comparison of several commonly used definitions is found in [9] and [5]. We are interested here in directed triples of distinct players, (i,j,k), where j passes to i and k, and i passes to k, such as we see in Figure 1. We call such a triple a **directed triangle around** i, and introduce the following:

Definition: In the passing network of a soccer team, the **clustering coefficient** of player i, is

$$\text{clust}(i) = \frac{1}{\deg^{out}(i)\deg^{in}(i) - E_{ii}^2} \sum_{j,k} \frac{\sqrt[3]{A_{jk}A_{ji}A_{ik}}}{\max(A)}.$$

The numerator in the sum above is the geometric mean number of passes in a directed triangle around i, which is nonzero if and only if the triple (i,j,k) does indeed define a directed triangle around i. The denominator is the maximum number of passes between any two players. Hence, the sum is a weighted count of all the directed triangles around i. As explained in Table I of [5], the denominator of the term outside the sum is the maximum possible number of directed triangles around i that could be formed by pairs of the neighbors of i. Here E denotes the incidence matrix for the underlying undirected network; i.e., $E_{ij} = 1$ if $A_{ij} \neq 0$ and $E_{ij} = 0$ otherwise. Since E_{ii}^2 gives the number of paths of length 2 in the unweighted graph that begin and end at i, subtracting this term eliminates degenerate triangles.

The clustering coefficient of player i, tells us how well she helps complete directed triangles around her. Hence, it gives an indication of how well a player helps her teammates to bypass well-defended passing lines. Note that in a particular triple, the alternative path created by player i may not be the shortest one. However, if a player has a high clustering score, she is consistently providing an alternate path to move the ball between other players. It is highly desirable for a team to have a high average clustering score, as this indicates a resilient passing network, against which the opposing team may have trouble defending.

Our last measure is simple but revealing. To identify players who are good defenders, we look at the difference between the number of passes originating with a particular player and the number of passes terminating with her. Formally we have, *Definition:* The **defensive score** of player i is

$$\text{def}(i) = \sum_{i \neq j} A_{ij} - \sum_{i \neq j} A_{ji}.$$

If def(i) is positive, then we may assume that player i is winning the ball from the other team; hence, she is a good defender. The converse is not necessarily true. A player may be quite good at winning the ball from the other team, but then makes passes that are intercepted, or takes shots at goal, possibly resulting in a negative defensive score. While having passes intercepted is not desirable, one cannot conclude from a negative defensive score that a player is a poor defender.

3. The Passing Data

3.1. Collecting It. For their analysis of the 2010 World Cup, Peña and Touchette obtained the passing data directly from the website of the International Federation of Association Football (FIFA). Unfortunately, neither the National Collegiate Athletic Association (the NCAA) nor the Metro Atlantic Athletic Conference (the MAAC) provides such data to the public. Hence, we relied on three Jasper team members, Kaelyn Angelo, Alexandra Iovine, and Janie Schlauder, who watched hours of film to collect the passing data. We are grateful for their expertise, as our analysis would not have been possible without their contributions.

As stated earlier, an arrow in the passing network represents the successfully completed passes from one player to another. Since we collected our own data, we first needed to define what we consider to be a successfully completed pass. In our analysis, we consider only completed passes that are clearly directed from one player to another. That is, if a Jasper kicks the ball upfield with no clear target teammate, this pass is not counted, even if the Jaspers retain possession. Under certain circumstances when the ball goes out of bounds, a player throws the ball back onto the field to restart the play. A throw-in, when clearly directed at a particular teammate, is counted as a completed pass. The collected passing data is available in appendix A.

3.2. Processing It. Our analysis made use of iPython Notebook, a web-based interactive computational environment. To simplify data manipulation, we created a Game and Aggregated Game class. Then, to compute the network-theoretic measures and to produce the various graphics, we used version 1.9.1 of the NetworkX package.

4. Analyzing the Jaspers

4.1. Substitutions and The Positional Approach. According to official FIFA rules [1], a team is permitted to make at most three player-substitutions during a single match, and a substituted player takes no further part in the match. Hence, at most 14 players may participate in a match. Moreover, once a player is assigned to a position, he will likely play in only that position throughout his time on the field. Hence, there is nearly a one-to-one correspondence between players and positions. The Manhattan College Jaspers, however, are governed by a different organization. The NCAA allows a team to substitute up to 11 players at a time [2]. Further, a player who has been taken off the field during the first period may re-enter the match during the second period, and a player who is taken off the field during the second period may re-enter during that same period.

According to the data available on gojaspers.com, the Jaspers played 18 matches during the 2013–2014 season. On average, the Jaspers fielded 16 players per match, which is two players greater than the maximum allowed in a professional match. Moreover, this average does not account for players exiting and reentering the field during a match. In light of this, we take a positional approach in our analysis. That is, we redefine the passing network so that the vertices represent positions, rather than players. We acknowledge, of course, that each player brings her own skills and style to a particular position. However, each position plays a particular role, so we find it reasonable to assume that a player will adapt as necessary. Since in distinct matches a different collection of players may occupy a particular position, we have chosen to aggregate the passing data over several matches. The passing network was then created by weighting each arrow with the arithmetic mean number of passes per match.

4.2. The Jaspers's passing network. The formation a soccer team assumes on the field is the most basic expression of its style of play. Each formation includes a goalkeeper; the remaining 10 players are distributed among defensive, midfield, and offensive positions. Since the forwards are positioned closest to the opponent's goal, one might assume they are the team's most offensive players; however, any player on the team may make a direct attempt at scoring. Likewise, since the play on the field is dynamic, any player may find herself defensively positioned, able to stop the other team from taking a shot at goal. The Jaspers utilized several formations throughout the 2013-2014 season. Most often, they used a 4-4-2 formation, which is comprised of the goalkeeper, four defenders (left and right outside backs, left and right center backs), four midfielders (left and right wing midfielders, left and right center midfielders), and two forwards (high and underneath forwards). Our analysis focuses on eight of the matches played in the 4-4-2 formation. The opponents for these matches were: College of the Holy Cross (9/3), University of Rhode Island (9/8), Wagner College (9/14), Delaware State University (9/16), Marist College (9/21), Iona College (9/25), Niagara University (9/28), and Monmouth University (10/2). In Figure 2 we show the aggregated passing network for these eight matches. Note that we use the thickness of the arrows to represent number of passes; i.e., a thicker arrow corresponds to a higher number of passes.

We observe heavy activity along the sides of the field, with the passes between the left outside back (LOB) and the left wing midfielder (LWM), and between the right outside back (ROB) and the right wing midfielder (RWM) far outweighing any other pairings. To interpret this activity in a larger context, we turn to the centrality measures.

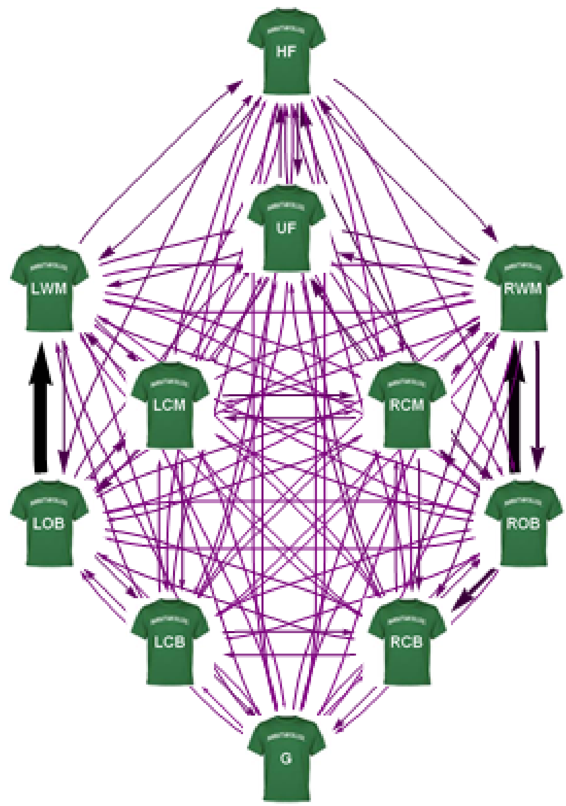


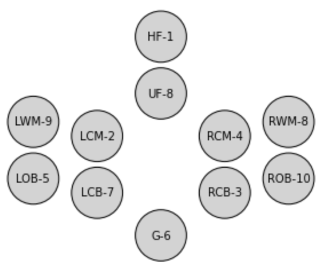
Figure 2. The Jaspers's Passing Network

4.3. Measuring the Jaspers. Table 1 summarizes the data for each position. The integer in brackets following each score indicates its ranking among the scores for that measure.

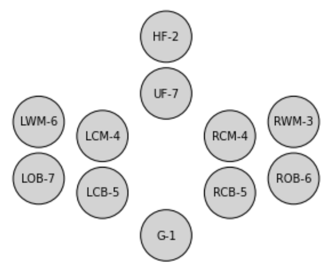
Recall that a high closeness score identifies a player as easy to reach. We observe that the high-ranked and low-ranked players are dispersed throughout the formation. That is, there are no pockets of the team that are difficult to reach. Additionally, the team displayed a well-balanced passing strategy, as indicated by a mean betweenness score of .81 with a standard deviation of .45. This inference will be confirmed in a forthcoming section, when we compare the Jaspers's performance to that of other teams. As Peña and Touchette observed, an imbalance in betweenness can be problematic for a team. If players with exceptionally high betweenness scores are identified by the opposing team, the opposition may then concentrate their defensive efforts around these players. While there were no players with exceptionally high betweenness scores, we do see that the underneath forward (UF) had a locally-high score relative to the high forward (HF) and the center midfielders. Hence, the team is clearly open to an isolated player attack. To aid us in visualizing this, we arrange the rankings from Table 1 into **Measure-Ranked Formation Diagrams (MRFDs)**, as seen in Figure 3.

Player	clo(i)	bet(i)	Page(i)	clust(i)	def(i)
HF	0.82 [1]	0.26 [2]	0.12 [8]	9.39 [10]	−6.75 [2]
UF	1.63 [8]	1.34 [7]	0.11 [7]	3.13 [4]	−4.50 [3]
RWM	1.63 [8]	0.39 [3]	0.13 [9]	7.66 [9]	−10.63 [1]
RCM	1.30 [4]	0.52 [4]	0.09 [5]	4.31 [7]	3.63 [8]
LCM	0.95 [2]	0.52 [4]	0.08 [4]	2.55 [1]	3.00 [7]
LWM	1.82 [9]	1.22 [6]	0.12 [8]	3.23 [5]	−10.63 [1]
ROB	1.90 [10]	1.22 [6]	0.10 [6]	5.07 [8]	13.88 [10]
RCB	1.03 [3]	0.98 [5]	0.07 [3]	3.70 [6]	0.00 [4]
LCB	1.60 [7]	0.98 [5]	0.05 [2]	2.59 [2]	2.38 [6]
LOB	1.43 [5]	1.34 [7]	0.10 [6]	3.70 [6]	8.75 [9]
G	1.54 [6]	0.12 [1]	0.04 [1]	2.98 [3]	0.88 [5]

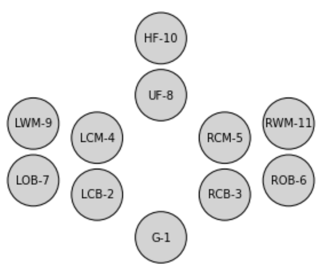
Table 1. Computed scores for individual positions. Note that clustering and betweenness are expressed as percentages.



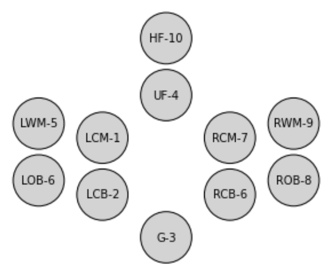
(a) Closeness



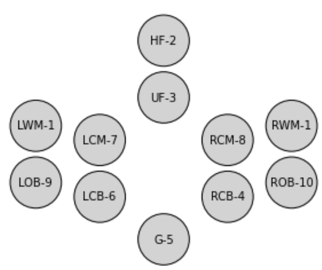
(b) Betweenness



(c) PageRank



(d) Clustering



(e) Defensive

Figure 3. The Jaspers's MRFDs

We now turn to PageRank, our notion of importance or popularity. A useful interpretation, as noted by Peña and Touchette, is that PageRank gives the probability that a particular player will have possession of the ball after a reasonable number of passes has been made. We see from the PageRank MRFD, Figure 3c, that the Jaspers clearly prefer to have the ball in the front of the field. This makes intuitive sense, of course, as possessing the ball near the opposition's goal would increase the Jaspers's chance of scoring. What is not so clear is why the Jaspers prefer to have the ball on the right side of the field. A simple Google search yields various theories regarding right-footed players and passing preference, but all evidence cited is anecdotal.

During a match, the player in possession of the ball can be in one of two states. She is either free to progress the ball forward, or she is being actively defended. As we have just observed from the PageRank scores, a Jasper in possession of the ball who is free to move will aim to put the ball in the front right portion of the field. The high betweenness ranking of the underneath forward, as seen in Figure 3b, suggests that she is instrumental in accomplishing this aim. If the player in possession of the ball is being heavily defended, she may look for help from the players on her right. This is told to us by the clustering coefficients, which identify players who help their teammates reroute the ball around defenders. We see in Figure 3d that the clustering scores are significantly higher for the players on the right side of the field.

Next we note that the outside backs had the highest defensive scores, Figure 3e. Curiously, the players with the lowest defensive scores are their neighbors, the left and right wing midfielders. While this does not say they are poor defenders, it does say they are prone to losing possession of the ball. Pairing this information with the passing network (Figure 2) paints a picture of what likely happened during those matches. The outside backs may have stolen the ball from the opposition, moved it upfield through the wing midfielders, who then lost possession. Note that this loss could have happened for a variety of reasons. For example, one of these players could have lost the ball to an opposing defender, or she could have attempted a cross, which is a medium to long-range pass aimed toward the general vicinity of the opponent's goal.

4.4. The Round of 17. Next we were interested in determining if the Jaspers's play style shares characteristics with any of the professional teams analyzed by Peña and Touchette. We began with Table 2, which was constructed by inserting the Jaspers's data into the table given in [6]. Thus, we obtain the "Round of 17". Before performing any of the computations, we naively assumed that the Jaspers's average number of passes would be far lower than those of the professional teams, given the disparity in skill and fitness level. We were surprised to see that this score for the Jaspers is exactly the median. It was also notable that for all three measures of connectivity, the Jaspers have the highest scores. Upon further reflection, though, we realized that this is likely due to our aggregation of the Jaspers's passing data over eight matches.

What is remarkable, and not so influenced by the aggregation of the passes, is the Jaspers's average betweenness score of 0.81 with a standard deviation of 0.45, as noted earlier. This is vastly different from the mean average betweenness of the professional teams, which is 3.6. At the moment it is unclear why this difference is present. In future work we plan to compare the Jaspers to the teams from the 2015 Women's World Cup and to other

Team	P	k	k_u	$\overline{\text{bet}}$	C_q
Argentina	227	4	5	2.7	8
Brazil	321	**5**	**7**	2.0	8
Chile	120	0	1	5.1	6
England	239	2	3	3.6	7
Germany	220	2	2	4.6	6
Ghana	184	3	4	3.5	8
Japan	180	1	5	3.3	8
Jaspers 4-4-2	199	**6**	**9**	**0.8**	**10**
Korea Rep.	227	3	5	2.6	8
Mexico	225	0	0	**1.9**	7
Netherlands	266	**5**	**7**	**1.9**	8
Paraguay	103	0	2	7.5	5
Portugal	175	3	4	4.1	7
Slovakia	166	3	6	3.0	7
Spain	417	3	5	**1.9**	**9**
USA	160	1	4	4.6	7
Uruguay	117	2	3	4.8	6

Table 2. Data for the teams in the 2010 Men's World Cup Round of 16, taken from Table 1 in [6], plus The Jaspers's scores. P: average number of passes; k: edge connectivity; k_u: undirected connectivity; $\overline{\text{bet}}$: average betweenness (expressed as a percentage); C_q: largest clique. The highest two values for the connectivity measures (k, k_u, C_q) and the lowest average betweenness ($\overline{\text{bet}}$) are highlighted.

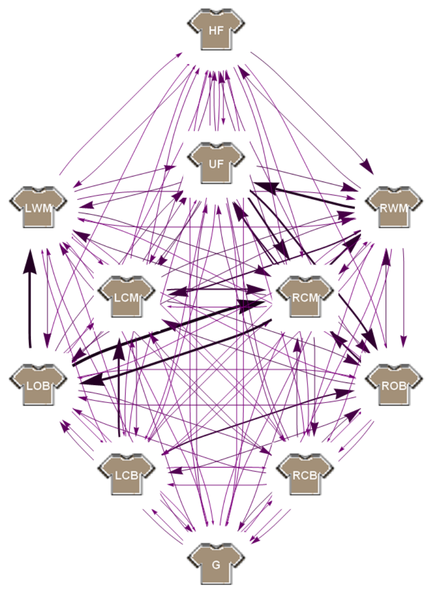
collegiate teams, in order to determine if those teams display a similarly well-balanced passing strategy.

4.5. Similarity. To obtain a more robust comparison between the Jaspers and the Top 16, we make use of a modified Borda Count voting method [8]. This is a preferential voting scheme in which voters rank candidates on the basis of preference. Traditionally, each candidate is awarded one point for each last place ranking, two points for each next-to-last ranking, and so on. The candidate with the largest point total is declared the winner. A Borda Count is viewed as a consensus-based voting system. That is, it tends to select the candidate who is most broadly acceptable to the electorate, rather than the candidate who is preferred by the majority. In our scenario, the teams that played in the same formation as the Jaspers are the candidates. We take the weighted arrows of the Jaspers's passing network to be the voters. If there is no arrow between two vertices, we consider it to have zero weight. Hence, there are 110 voters, who will choose the team whose passing network is most similar to the Jaspers's network. For each arrow in the Jaspers's network, a team in which the corresponding edge has the least difference in weight is given the highest ranking. The rankings then decrease as the magnitude of the difference in the weights increases. Note that in our system, a voter may give the same ranking to different candidates. In this way, our scheme differs from a standard Borda Count. The Borda Count for each team will be called its **similarity score**.

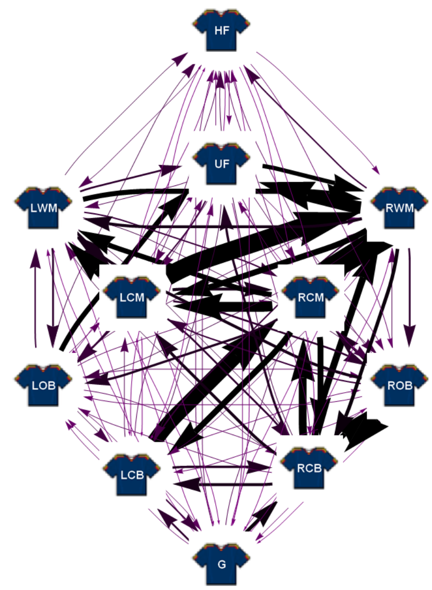
The official FIFA website for the 2010 World Cup provides the tactical line-up and actual formation for each match of the tournament. In the Round of 16, there were three teams whose formations closely resembled the Jaspers's 4-4-2 formation: Germany, Spain, USA. These teams, therefore, are the candidates in our Borda Count. Using the passing distributions from the FIFA website, we created the passing networks seen in Figure 4. Referring to the Jaspers's data in appendix A and recalling that the Jaspers played eight games in the 4-4-2 formation, we see that the arrow in the Jaspers's passing network from the left outside back to the left wing midfielder has a weight of 11. We consider this weighted arrow to be one of our 110 "voters". Now we use the FIFA data to determine that the corresponding arrow in the passing network for Germany has weight 8. In the passing network for Spain it has weight 6, and for the USA it has weight 3. Hence, this voter ranks Germany, Spain, and USA as 1st, 2nd, and 3rd, respectively. When the rankings for each arrow were tallied, Spain earned a similarity score of 56, USA was much closer with a similarity score of 107, and Germany was the winner with a similarity score of 110. While the German team is the most similar to the Jaspers among the three teams considered here, differences in their passing networks are readily apparent. In particular, the passing network for Germany shows much more activity in the center of the field. We also note differences between MRFDs for Germany, Figure 5, and the Jaspers, Figure 3. In particular, we note that Germany's betweenness scores are much higher on the right, while the Jaspers's are slightly higher on the left. Also of note are the high PageRank scores for Germany's left center back and right outside back. On the other hand, all of the Jaspers's most popular players are at the front of the field. For both teams, the high forward had the highest clustering score. The other high clustering scores for the Jaspers are heavily weighted on the right side of the field. However, the German team presents a more even distribution of its players' clustering scores.

5. Future Work

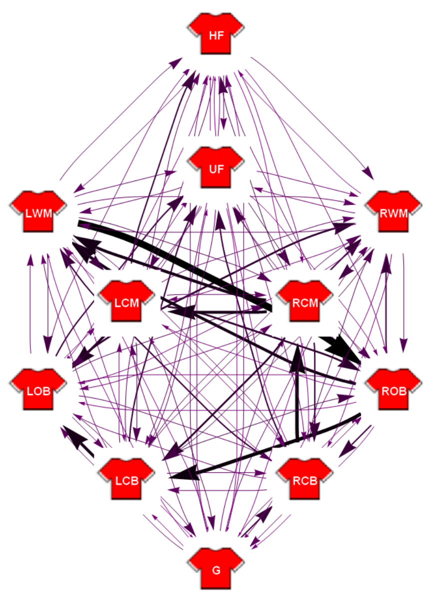
In future work, we may look more closely at the differences noted in the previous section to understand how they arise. Armed with the knowledge that the Jaspers's passing network bears resemblance to network of the very successful German team, a coach may wonder how to strengthen this resemblance. Of course, this strategy of simply emulating a successful team ignores the fact that the Jaspers aren't the only team on the field. Another project could involve simultaneously analyzing the networks and measures of both opponents in a given match. Additionally, one could apply these techniques to other collegiate sports, such as field hockey or lacrosse, that possess a similar network structure.



(A) Germany

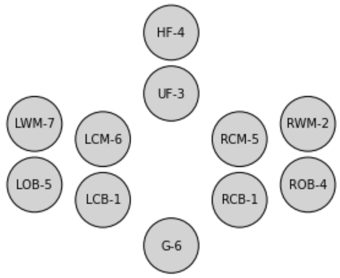


(B) Spain

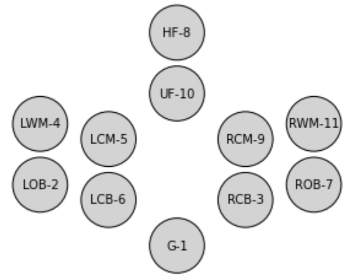


(C) USA

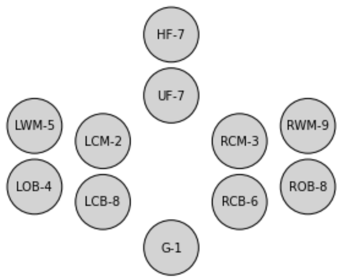
FIGURE 4. Candidates for the Borda Count Comparison



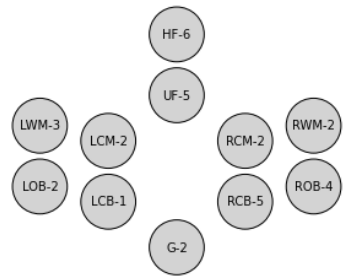
(a) Closeness



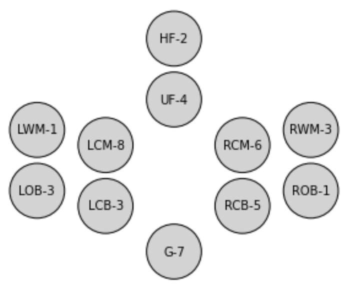
(b) Betweenness



(c) Pagerank



(d) Clustering



(e) Defensive

Figure 5. Germany's MRFDs

6. Acknowledgements

The bulk of this research was conducted in the summer of 2014 when the first named author received financial support from the Manhattan College School of Science Scholars Program. Both authors are grateful to Constantine Theodosiou, Dean of Science, and to Rani Roy, Director of Graduate School and Fellowship Advisement, for the support and encouragement they provided to him. The Department of Mathematics provided financial support for the data collection process. The invaluable contributions of Kaelyn Angelo, Alexandra Iovine, and Janie Schlauder were already mentioned, but they bear repeating. We are particularly grateful for the advice we received from Javier Lopez Peña and Hugo Touchette regarding the capabilities and usage of the NetworkX package. We thank Matthew Jura, Richard Goldstone, Janet McShane, and Kathryn Weld for many productive conversations. Lastly, we are grateful to the referees for their helpful comments and suggestions.

Appendix A. Passing Data

Players	HF	UF	RWM	RCM	LCM	LWM	ROB	RCB	LCB	LOB	G
HF	0	29	32	16	4	29	13	0	6	12	0
UF	31	0	32	16	15	22	10	5	4	6	1
RWM	29	26	0	17	5	3	47	20	0	1	0
RCM	18	31	31	0	21	12	28	22	5	7	0
LCM	10	13	13	23	0	26	5	5	14	31	0
LWM	29	13	5	5	20	0	0	1	9	39	1
ROB	40	26	82	31	5	2	0	49	3	2	8
RCB	15	10	25	10	6	4	18	0	12	11	18
LCB	2	8	5	9	7	12	2	15	0	17	18
LOB	21	19	4	10	25	88	2	7	24	0	6
G	0	5	6	5	7	9	10	4	3	9	0

Table 3. The Jaspers's aggregated passes.

References

[1] FIFA official laws and regulations. `http://www.fifa.com/aboutfifa/officialdocuments/doclists/laws.html`.
[2] NCAA soccer rules of the game. `http://www.ncaa.org/championships/playing-rules/soccer-rules-game`.
[3] Sergey Brin and Lawrence Page. The anatomy of a large-scale hypertextual web search engine. *Comput. Netw. ISDN Syst.*, 30(1-7):107–117, April 1998.
[4] Kurt Bryan and Tanya Leise. The $25, 000, 000, 000 eigenvector: The linear algebra behind google. *SIAM Rev.*, 48(3):569–581, jan 2006.
[5] Giorgio Fagiolo. Clustering in complex directed networks. *Physical Review E*, 76(2), aug 2007.
[6] Javier López Peña and Hugo Touchette. A network theory analysis of football strategies. In C. Clanet (ed.), Sports Physics: Proc. 2012 Euromech Physics of Sports Conference, p. 517-528, Éditions de l'École Polytechnique, Palaiseau, 2013. (ISBN 978-2-7302-1615-9), 2012.
[7] M. E. J. Newman. *Networks : an introduction*. Oxford University Press, Oxford New York, 2010.
[8] Donald G. Saari. *Chaotic Elections! A Mathematician Looks at Voting*. American Mathematical Society, 2001.
[9] Jari Saramäki, Mikko Kivelä, Jukka-Pekka Onnela, Kimmo Kaski, and János Kertész. Generalizations of the clustering coefficient to weighted complex networks. *Physical Review E*, 75(2), feb 2007.

Student biography

Andre Oliveira (aoliveira@wesleyan.edu) Andre graduated from Manhattan College in May 2015 with majors in Mathematics and Computer Science. He currently attends Wesleyan University, where he is pursuing a PhD in Mathematics. When he's not programming or thinking about Mathematics, Andre can be found on a rugby pitch.