

May 2015

Discovering Key Players and Key Groups in a Soccer Team Using Centrality Measures

Ying Zhu

University of Wisconsin-Milwaukee

Follow this and additional works at: <https://dc.uwm.edu/etd>



Part of the [Electrical and Electronics Commons](#)

Recommended Citation

Zhu, Ying, "Discovering Key Players and Key Groups in a Soccer Team Using Centrality Measures" (2015). *Theses and Dissertations*. 941.

<https://dc.uwm.edu/etd/941>

This Thesis is brought to you for free and open access by UWM Digital Commons. It has been accepted for inclusion in Theses and Dissertations by an authorized administrator of UWM Digital Commons. For more information, please contact open-access@uwm.edu.

DISCOVERING KEY PLAYERS AND KEY GROUPS
IN A SOCCER TEAM USING CENTRALITY
MEASURES

by

Ying Zhu

A Thesis Submitted in

Partial Fulfillment of the

Requirements for the Degree of

Master of Science

in Engineering

at

The University of Wisconsin-Milwaukee

May 2015

ABSTRACT

DISCOVERING KEY PLAYERS AND KEY GROUPS
IN A SOCCER TEAM USING CENTRALITY
MEASURES

by

Ying Zhu

The University of Wisconsin-Milwaukee, 2015
Under the Supervision of Professor Jun Zhang

In this thesis, I introduce that passing performance is crucial skill in the soccer game. I provide network centrality approaches to discover key players and key groups in a soccer team. The Utility Model of game theory evaluates each soccer player's contribution to his team outcome. The approach of finding key players is to implement soccer passing network data with the combination of Nash Equilibrium with Bonacich Centrality Measure. We identify the key player by finding the top individual Inter-Centrality Measure, and also identify the key group of players that match better together in the game. The results verification will use

2013 market values, media attention, and team unbeaten probability by his appearance/absence.

© Copyright by Ying Zhu, 2015
All Rights Reserved

TABLE OF CONTENTS

1	INTRODUCTION	1
2	APPROACH.....	8
2.1	<i>An example of “best friends selector network”</i>	8
2.2	<i>Mathematical modeling</i>	9
2.3	<i>Centrality measures.....</i>	12
2.4	<i>Linkage between Bonacich alpha centrality and Nash Equilibrium</i>	22
2.5	<i>Key player Inter-centrality measure</i>	24
2.6	<i>Key group of inter-centrality measure.....</i>	27
2.7	<i>Approach Extension.....</i>	30
3	IMPLEMENTATION.....	34
3.1	<i>Passing data extraction</i>	34
3.2	<i>Introduce our passing network data</i>	36
3.3	<i>Individual ability weights.....</i>	37
3.4	<i>Find key players and key groups by inter-Centrality (ICM).....</i>	40
3.5	<i>Results verification of the key group</i>	41
4	EXPERIMENT RESULTS.....	42
4.1	<i>Local centrality of each player.....</i>	42
4.2	<i>Key players of Manchester City and Bolton Wanderers</i>	43
4.3	<i>Key groups of MC and BW in size 2, 3, and 4</i>	45
4.4	<i>Result verification</i>	47
4.5	<i>Closeness Centrality results in MC and BW</i>	48
4.6	<i>Betweenness Centrality results</i>	48
5	OBSERVATIONS	49
5.1	<i>Key players and key groups analysis.....</i>	49
5.2	<i>Extensions.....</i>	51
6	CONCLUSION	56
	REFERENCES.....	57

LIST OF FIGURES

Figure 1: Passes completion.....	2
Figure 2: "Best friends selector " network graph.....	9
Figure 3: (b) The top Betweenness removal.....	33
Figure 4: Passing network graph.....	37
Figure 5: Passing Matrix.....	37
Figure 6: Goal probability.....	38
Figure 7: Final Third passes completion.....	39
Figure 8: Successful/ unsuccessful passes ratio.....	40
Figure 9: Local Centrality.....	43
Figure 10: ICM rankings	44
Figure 11: MC network graph for the key player	45
Figure 12: BW network graphs for the key player	45
Figure 13: MC network graphs for the key group	46
Figure 14: BW network graph for the key group	47
Figure 15: Closeness Centrality rankings	48
Figure 16: Betweenness Centrality rankings.....	48
Figure 17: The top Betweenness and the key player graph.....	49
Figure 18: Player integrated centralities rankings.....	53

LIST OF TABLES

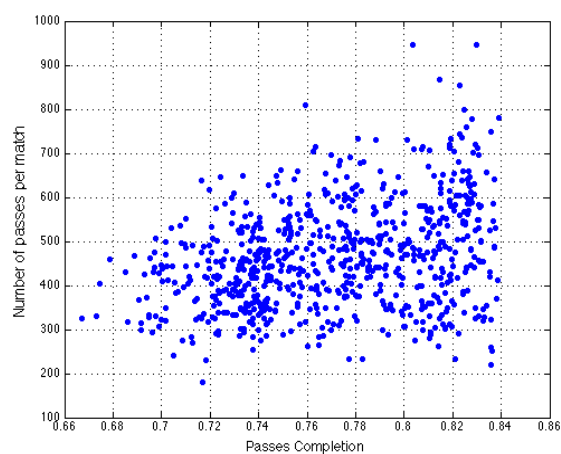
Table 1: ICM standings: (a) MC (b) BW.....	44
Table 2 : Key groups.....	46
Table 3: Result verification	47
Table 4: Team comparison.....	55

1 Introduction

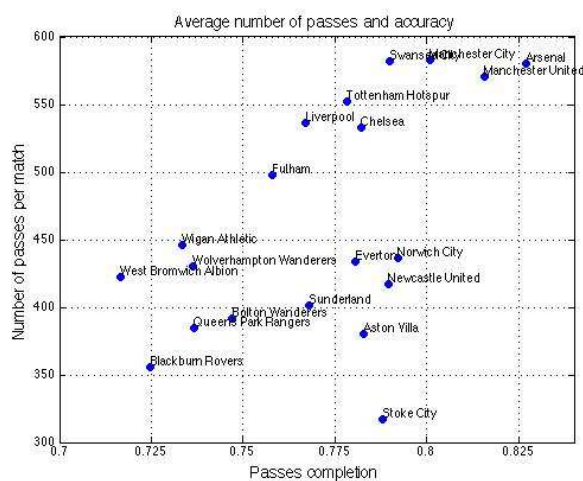
Soccer is the most popular sport throughout the world. It is a hot topic to evaluate soccer players' performance, like who is the most influential, who is the most important player (key player) in a soccer team. A general naive answer is that the player with the most goals is the key player, but, in modern football, a more thoughtful answer is the player at the center of the passes may be the key player. Actually, a goal often requires a sequence of good passes. Passes play a significant role in soccer game. The book named the *Numbers Game: why everything you know about football is wrong* said: "the single most common action of soccer players are passes in all shapes and sizes: passes account for well over 80 percent of events on the pitch. This means that possession requires a collective, rather individual effort. It is a measure of team competence ability, not a specific player's brilliance."

In this thesis, the soccer data analysis is based on data collection of England Premier League 2011-2012 season, which supplied by Manchester City Club.

Through the 20 teams overall passes performance can tell us some stories. From figure 1, we can see each spots represent a team passing performance. As the increasing of the percentage of pass completion, the number of passes accomplishes go up in per game.



(a)



(b)

Figure 1: Passes completion (a) Team's each game (b) Average

Averaged over the entire England Premier League 2011-2012 season, the picture of possession looks straightforward shown on figure 1 (b). Three teams, Manchester City, Arsenal or Manchester United played more than 550 passes in the typical match, and completed 80% or better passes, their standings are excellent in the end of the season, which are the third, second, and first, respectively. Otherwise, A team named Blackburn Rovers complete around 72% passes, and its standing position is the second from the end. Even though the team Stoke City complete passes around 78%, his average passes just over 300 results in not better rank position (No.14th, total 20 teams). It follows that teams better at passing the ball should concede fewer turnovers. Those teams that complete passes at a higher rate are less prone to giving the ball back to the opposition. After all, for a team, possession quality is truly dependent on the rate and the accuracy of passes. For an individual player's completion percentage, at the elite level, it affects the particular situation not his foot skills. Further, a good

team manages to create and find space for both the passer of the ball and his intended target, making the passing situation easier.

Currently, we understand that passing the ball plays the most important role in the modern soccer game. In this thesis, based on players' interactive passing performance, I use mathematical models to discover the most important player, represented by the key player, and to discover the most important group of that team, represented by the key group.

In the real world, we are living in a big social network, and everyone is one simple node in his/her social network. In general, we always think of one question: who is the central person, who is most popular in our social circle. A soccer game is an interactive game. Also, our objective is to discover the most valuable player on a soccer team. First, I use the Utility Model of game theory, which involves individual impact parameters and interactive performance parameters between players. Each player has his own individual impact on the team game outcome, and players' interactions with one another have an impact on the game outcome.

Player's interaction means that each player's passing performance contributes to group actions. Second, a network method, Inter-Centrality Measure, can help us identify the most important soccer player from his passing network. This measure reflects a player's contribution to their teammates in the Utility Model. This measure was developed by Katz Centrality (1953), PageRank Centrality (1998), and Bonacich Centrality (2006). Nash equilibrium theory is involved through the explanation process. Nash proved that every finite game has a unique interior Nash Equilibrium. When each player gets his pure strategy in team outcome, his Inter-Centrality measure will be the proportioned to his impact on team outcome. Based on the linkage between Inter-Centrality Measure and Nash Equilibrium, we use Inter-Centrality Measure to rank players' team contribution. The key player has the highest inter-centrality. That is to say, the total team performance suffers the most, if the key player is removed for any other player. The key player removal leads to the greatest overall reduction in team outcome performance. Moreover, we add individual ability weights to represent each player's personal skill, which

involves shot ratio, ratio of successful and unsuccessful passes and final third pass completion. It should be more persuasive to identify the key player by a combination of personal ability weights and group effort. In addition, we can use the same idea to identify the key group in a soccer team. The key group's removal will make total performance suffer most compared to other groups of players on that team. Knowledge of the key group is crucial for coaches to determine who performs better together.

Furthermore, using other centrality methods such as Local Centrality, Betweenness Centrality, and Closeness Centrality can reflect each player's performance characteristics. And taking comparison between two teams by Density Centrality and Entropy methods can help us know which one performs more cohesively, like a "total soccer team".

Finally, I will use four parameters to verify our key player and key group results. They are soccer player market value in 2012 and 2013, team unbeaten probability by his appearance/absence, media attention by number of articles in *Guardian*,

and Integrated performance of centralities.

My research's technique can be utilized on many potential applications.

I mentioned player evaluation is a hot topic in professional sports. This technique can be used for player performance evaluation and used to determine a player's value in the trade market, as well as determine a player's monetary value. These techniques also can be used in other applications. For example, in crime analysis, we select a number of suspects in a criminal network to discover the key suspect, or key suspect groups, and then to neutralize them in order to maximally disrupt their crime network's ability. In business, this technique can help human resources discover the best employee. In a military justice context, these techniques can help select an efficient set of actors to perform at the highest level of discretion. In public health, it is helpful to select the key subset of the population to immunize or quarantine in order to optimally slow epidemics [5].

This thesis is organized as follows: in section 2, the approaches are described in this section: payoff model, inter-centrality measure, centrality

measures extension. In section 3, I will show the details of implementation:

passing data extraction of MC vs. BW game, explanation of our data information

and three players' personal skill weights. In section 4, it is experiment results.

Finally, this paper has observations in section 5, and a conclusion in section 6.

2 Approach

In this section, first, we will give an example of network. Second, I will introduce a network matrix demo. And then, I will define the utility function for the game. Third, I will introduce several centrality measures. Forth, I will explain the linkage between centrality measure and Nash Equilibrium of the game. Moreover, I will introduce other two centralities to reflect player's different character on the pitch. Finally, I will provide Density Centrality and Entropy concept for two teams performance comparison.

2.1 An example of “best friends selector network”

Firstly, I will introduce the idea from Katz [12], an example named “best friend

chooser” for seven people, who are acquainted with each other, shows how Katz comes up with. A chooses F as her best friend. The option of B is C and F. And F chooses A and D. etc. (Table one. shows). It is a 7×7 adjacency matrix with diagonal zeros. After we saw the datasets below, a question comes up with: what information is made from such network? And who is be central or popular person between them?

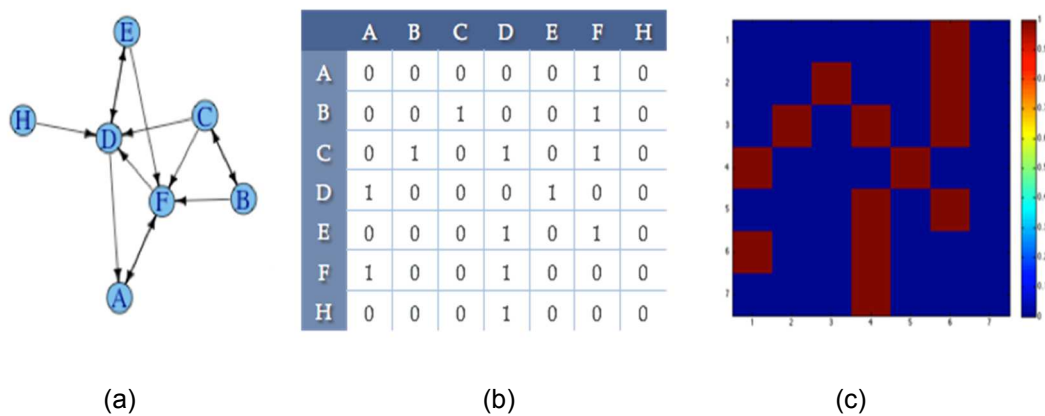


Figure 2: "Best friends selector " network graph

2.2 Mathematical modeling

2.2.1 Payoff model of game theory

This section illustrates a model for identifying key players and key groups in

a game. First, the model is a utility function of players. Second, the illustration of Inter-centrality measure (ICM) will provide the results of key players and key groups.

A team outcome is mentioned previously. That will use an individual payoff function (utility function) for each player using the notation of Ballester et al [4,2006]. A player's performance is dependent on those of other players, through a quadratic utility function as

$$u_i(x_1, x_2, \dots, x_n) = a_i x_i + \frac{1}{2} \theta_{ii} x_i^2 + \sum_{i \neq j} \theta_{ij} x_i x_j + \delta_i x_i \quad (18)$$

where the parameter $a_i > 0$ stands for the coefficient of individual efforts. We set all players' individual efforts a_i the same as a . The second term, θ_{ii} , will captures the concavity of individual actions for all $i=1,2,3\dots n$. θ_{ij} stands for the complementary action of player i on player j , as Σ denoted the cross-effects. Each player has his own performance efforts or impact x_i . In the Ballester [4], they decomposed Σ into three components following expression defined as

$$\Sigma = [\theta_{ij}] = -\omega I - \xi O + \lambda G \quad (19)$$

, coming from

$$\Sigma = \begin{bmatrix} \theta_{ii} & \dots & \theta_{ij} \\ \theta_{ij} & \ddots & \vdots \\ \theta_{ij} & \dots & \theta_{ii} \end{bmatrix} = -\omega \begin{bmatrix} 1 & 0 & \dots & 0 \\ \vdots & 1 & \dots & 0 \\ \vdots & 0 & \ddots & 0 \\ 0 & 0 & \dots & 1 \end{bmatrix} - \xi \begin{bmatrix} 1 & 1 & \dots & 1 \\ 1 & 1 & \dots & \vdots \\ \vdots & \vdots & \ddots & \vdots \\ 1 & 1 & \dots & 1 \end{bmatrix} + \lambda \begin{bmatrix} 0 & & & \frac{\theta_{ij} + \xi}{\lambda} \\ & \ddots & & \\ & & \ddots & \\ \frac{\theta_{ij} + \xi}{\lambda} & & & 0 \end{bmatrix},$$

where I is a $n \times n$ square identity matrix, and 0 denotes $n \times n$ squares matrix of

ones. How can I get those three components? The payoff function of game theory

is indeed a quadratic function. If $a_i > 0$, $\frac{\partial^2 u_i}{\partial x_i^2} = \theta_{ii} < 0$ defined the concave in

own effort. Bilateral influence between players are achieved by Jacobian of

x_i : $\frac{\partial^2 u_i}{\partial x_i \partial x_j} = \theta_{ij}, i \neq j$. We should understand player i 's response will be upward

with the increase of player j 's effort, when $\theta_{ij} > 0$. Let $\bar{\theta} = \max\{\theta_{ij} | i \neq j\}$, and

$\underline{\theta} = \min\{\theta_{ij} | i \neq j\}$; $\xi = -\min\{\theta_{ij}\}$. By construction, $g_{ij} \in [0,1]$, let $\lambda = \bar{\theta} + \xi \geq$

0 , and $\theta = -\omega - \xi$, So $g_{ij} = \frac{\theta_{ij} + \xi}{\lambda}$, when $i \neq j$ and $g_{ii} = 0$. Now, $G = [g_{ii}]$ is a

non-negative square and weighted adjacency matrix. Its zero diagonal property

means no cross effects to themselves. Finally, the new form of payoff function

constructed as

$$u_i(x_1, x_2, \dots, x_n) = a_i x_i - \frac{1}{2}(\omega - \xi)x_i^2 + \xi \sum_{j=1}^n x_i x_j + \lambda \sum_{j=1}^n g_{ij} x_i x_j + \delta_i x_i \quad (20)$$

where $\omega > 0$ for all $i = 1, 2, \dots, n$. The G is a zero diagonal nonnegative square matrix, interpreted as an adjacency matrix of g . G is a passing matrix which recorded the number of successful passes from player i to player j .

2.3 Centrality measures

2.3.1 Local centrality

Local centrality is a method to measure degree of each node in the network.

There are three degrees: in-degree and out-degree. In a directed graph,

In-degree represents the number of incoming edges from other nodes. On the

other side, out-degree of node is the number of edges stemmed from that node to

the others. For indirect graphs, its in-degree is equal to out-degree. By the way,

total degree equals to the sum of in-degree and out-degree. For our example,

player A's in-degree is 2, which means A has been passed twice. And looking at

player C's out-degree is 3, which hints C passes ball to other three players.

According to local centrality, we can directly know who get the most “ballots”, and also who contributes a lot to others.

2.3.2 PageRank Centrality

This measure is a reasonable method to find out the relative “important” person within the interactive network, so as to recognize more valuable sources in a communication network, and to compute the best performance status of soccer player in a game. A method named PageRank [14] is proposed for computing a relative rank for every web page based on the graph of the web. This paper states a point: highly linked pages play more important role on pages with few links.

What is meaning of PageRank? The intuitive interpretation of PageRank: a page having high rank is the sum of the ranks of its backlinks is high. The author of PageRank algorithm takes a smart action: to make web pages being link structure. For example, web pages are seen as nodes, and web pages links each other are regarded as edges. Every web page has the number of forward links

(out-degrees) and backlinks (in-degrees). Generally, such measure idea built by a general formula, it is

$$\lambda v_j = g_{1j}v_1 + g_{2j}v_2 + \dots + g_{nj}v_n \quad (4),$$

where g_{ij} represents node i contribute to node j , v is a vector that represents each node's PageRank centrality value, and λ is a constant that makes the equation with non-zero. A network graph can be transformed to an adjacency matrix G .

Equation (4) is a function of the centrality or status of an actor, which tells us the status of those who choose that node. Equation (5) is another representation for equation (4) as

$$\lambda v = G^T v \quad (5)$$

Obviously, such measure looks like the general eigenvector equation. Equation (5) represents the status of each individual is a linear proportional relationship of the actors to whom it is interacted with. Actually, equation (5) has n values of λ .

Another form of equation (5) is $G^T V = V \Lambda$, which Λ stands for a diagonal matrix of eigenvalues λ , and V describes n by n matrix, whose column is eigenvectors of G .

As it is, an eigenvector corresponds one of standard centrality measures, named eigenvector centrality. For our example, it is easy to be interpreted the status based on the PageRank centrality description. In example network, person A's popularity becomes the greatest, cause A's receiving options from player D and player F, who have already got high ranks of being popular. It is zero status on player H, who is not supposed to be selected from any others. In this case, H contributes nothing to others' centrality. That is to say, like a "balloting" case, a person, who has no power, or makes no sense for any others status.

2.3.3 Bonacich Centrality

In fact, Bonacich centrality [4] is the extension of PageRank centrality. Why should we improve PageRank? It exists a small problem in PageRank function.

Taking into account a situation that two nodes points to each other but nothing else, then, the nodes never distribute any power, because there are no out-edges.

Such case is in trouble and is called a rank sink. In my example, B and C select

each other, but nothing else in the network chooses them. So the PageRank centrality of B and C are all zeros. How can we improve this rank sink problem? A solution to this problem should allow every node having its own source status or characteristics before connection to others. Let c be the vector of each individual characteristics status out of network. Further, the alpha centrality is formulated as

$$v = \alpha G^T v + c \quad (6)$$

$$v = (I - \alpha G^T)^{-1} \cdot c \quad (7)$$

where the determination of centrality (6) is that the parameter α reflects the relative strong affecting of internal status versus external factors. I is identity matrix with one diagonal.

Even though c could be individual external status, at present, I focus on the influences of graph structure. I assume every node's own external status to be identical as one. The constant α would be less or equal to the inverse of greatest eigenvalue of G . I will explain for its range limit later. From our example structure, for $\alpha = 0.58$ (the largest eigenvalue is 1.6838), alpha centrality ranking order

should be $A > F > D > E > B = C > H$. The results verified the assumption of alpha centrality. F and D are apparently chosen a lot by others. Though A is chosen by just two players, but is chosen by F and D other than “small fry” in the group. In addition, B, C, and E are all chosen by one of them. But E is linked by D, so is in somewhat different position than B and C, both selected by each other.

2.3.4 Linkage between Katz centrality and Bonacich centrality

Bonacich centrality is approximately identical to Katz centrality. The purpose of Katz [4] is also to suggest a method of computing node status in the socio-group. Katz pointed out the idea: status is the total number of successful direct and indirect influences produced by each individual. At the first glance the example, we have already know each individual choices, and 7×7 matrix is referred to be matrix G, whose “1” represents an individual passes one ball to the other one. Katz said: “the powers of G as the elements is the numbers of chains of

corresponding lengths going from i through intermediaries to j ". That is to

say, $G^2 = g_{ij}^2 = \sum_k g_{ik}g_{kj}$, if and only if $g_{ik}g_{kj} = 1$, i chooses k , and k chooses j .

Consider not only the numbers of direct passing, but also the status of the

individual who passes ball first, the action of each player who passes these in turn,

etc. For our example, player B selects C, and C selects D, so $g_{BC}g_{CD} =$

1. Actually, the column sum of G^2 is the two-step choices from group to each

other. Higher power G has the similar explanation with power two of G . How is

able to compute each individual's influence within social network? Katz answered

by constructed a weighted sum of all the powers of the network matrix G as

$$V = \alpha G + \alpha^2 G^2 + \dots + \alpha^k G^k \quad (8)$$

$$\sum_{i=1}^{\infty} \alpha^i G^i = \sum_{i=0}^{\infty} \alpha^i G^i - I = (I - \alpha G^T)^{-1} - I \quad (9)$$

where α is an "attenuation" factor. Equation (8) works out as long as $|\alpha| < \frac{1}{\lambda_1}$

satisfied, where λ_1 is maximum value of an eigenvalue of G , because the infinite

sum of $|\alpha| < \frac{1}{\lambda_1}$ converges to equation (8). Now, another question come up with

my mind: does Katz centrality and Bonacich alpha centrality have a bridge? See

proof below:

$$\begin{aligned} \sum_{i=1}^{\infty} \alpha^k (G^T)^k &= \left[\sum_{i=0}^{\infty} \alpha^k (G^T)^k - I \right] \cdot c = [(I - \alpha G^T)^{-1} - I] \cdot c = [(I - \alpha G^T)^{-1}] \\ &= v - c \end{aligned} \quad (10)$$

where it has just a constant, c , difference between equation (7) and equation (8).

As a result, Katz's algorithm is the same as Bonacich alpha-centrality. Both have already helped each other verify their own assumption.

2.3.5 The relation between an asymmetric matrix and alpha-centrality

Based on equation $G^T V = V \Lambda$, it is for sure that $G^T = V \Lambda V^{-1}$. The eigenvectors of asymmetries are not orthogonal, so the transpose of V is not equal to the inverse of V . Let y_i be the i -th row of V^{-1} . Thus, we can attain equation as:

$$v = (I - \alpha G^T) \cdot \mathbf{1} = \sum_{i=0}^n \frac{y_i \cdot \mathbf{1}}{(1 - \alpha \lambda_i)} v_i \quad (11)$$

where λ_1 is the largest eigenvalue of G^T . If α is very approach λ_1 , the equation

(11)'s final sum will become more and more dominant and also converge. I mean

$$\lim_{\alpha \rightarrow (\frac{1}{\lambda_i})} x = \left(\frac{y_i \cdot 1}{1 - \alpha \lambda_i} \right) x_i \quad (12)$$

The process of proof equation (11) is $v = (I - \alpha G^T) \cdot 1 = (\sum_{k=0}^{\infty} \alpha^k G^{kT}) \cdot 1 =$

$$\left(\sum_{k=0}^{\infty} \alpha^k \sum_{i=1}^{\infty} \lambda_i^k v_i y_i \right) \cdot 1 = \left(\sum_{i=1}^{\infty} \left(\sum_{k=0}^{\infty} \alpha^k \lambda_i^k \right) v_i y_i \right) \cdot 1 = \sum_{i=0}^n \frac{y_i \cdot 1}{(1 - \alpha \lambda_i)} v_i, \text{ which proof}$$

the relation between asymmetric matrix and alpha centrality.

2.3.6 Key node of centrality measure

Bonacich has proposed the purpose of measure of node status. A player is being more popular because he/she is opted by a popular individual. I mean someone having more power transmits information to the other one, who will become more powerful. The below matrix keeps track of the numbers of indirect links that start from node i and end at node j with a parameter α and an adjacency matrix G . Based on the range proof of α , at present, α should be chosen a little less than the inverse of the norm of highest eigenvalues of G matrix transpose, in

which case $g_{ij} \in [0,1]$, measures the weight related to direct connections. Again,

G^k keeps track of the indirect links in the network matrix, as I explained previously.

Currently, according to equation (9), it defines another form of equation as

$$\sum_{k=0}^{+\infty} \alpha^k G^k = [I - \alpha G]^{-1} = M(\alpha, g), \quad 0 < \alpha < \lambda_1 \quad (13)$$

$$m_{ij}(\alpha, g) = \sum_{k=0}^{+\infty} \alpha^k g_{ij}^k \quad (14)$$

If M is nonnegative network matrix, its coefficient represents the number of links in

g that start at node i and end at j , where paths of length k are weighted by α^k .

Previously, we found out linkage between Katz centrality and Bonacich alpha

centrality. The determination of centrality equation (7) is that the centrality impact

associated to the external status with external factors c . At present, we primary

pay attention to the effects of network native structure. Each node is regard as the

same “characteristic” so that parameter c is a vector of ones. The vector of

Bonacich alpha centrality with alpha factor in g is equation (15) as

$$b(\alpha, g) = [I - \alpha G]^{-1} \cdot \mathbf{1} \quad (15).$$

The equation (16) counts the total number of passes that node i contributes to

node j in matrix G .

$$b_i(\alpha, g) = \sum_{j=1}^n m_{ij}(g, \alpha) \quad (16)$$

All sums is from i to itself, and sum of all the out-coming edges from player i to

player j by equation (17) as

$$b_i(g, \alpha) = m_{ii}(g, \alpha) + \sum_{i \neq j} m_{ij}(g, \alpha) \quad (17)$$

2.4 Linkage between Bonacich alpha centrality and Nash

Equilibrium

In this section, the team game has a unique and interior Nash Equilibrium.

Theorem 1(Ballester, 2006)[3]: the matrix $\left(\omega \left[I - \frac{\lambda}{\omega} G\right]^{-1}\right)$ is nonnegative, if and

only if $\frac{\lambda}{\omega} < \frac{1}{\lambda_1}$. The game Σ exists a unique and interior Nash equilibrium x^* of the

team game, which is defined as

$$x_i^* = \frac{\alpha \cdot b_i(g, \lambda^*) + b_{\delta_i}(g, \lambda^*)}{\omega + \xi \sum_{j=1}^n b_j(g, \lambda^*)} \quad (21),$$

where equation (21) implies that each individual's performance is proportion to his

own Bonacich centrality, and also contributes to the group actions.

The proof of Theorem 1: the game of payoff function is a quadratic equation, when an interior Nash Equilibrium in pure strategy exists, payoff function will satisfy: $\frac{\partial u_i}{\partial x_i}(\mathbf{x}^*) = 0$, and $x_i^* > 0$, for $i = 1, 2, \dots, n$. Thus, for maximizing u_i : $\frac{\partial u_i}{\partial x_i} =$

$a_i + \theta_{ii}x_i + \sum_{i \neq j} \theta_{ij}x_j + \delta_i = 0$, we can get equation: $-\sum x = [\omega I + \xi O - \lambda G] = a \cdot$

$\mathbf{1} + \delta \cdot \mathbf{1}$, and $\lambda^* = \frac{\lambda}{\omega}$. $\omega(I - \lambda^*G) \cdot \mathbf{x}^* = a \cdot \mathbf{1} + \delta \cdot \mathbf{1} - \xi O \mathbf{x}^*$, with $O \mathbf{x}^* = \mathbf{x}^* \cdot$

$\mathbf{1}$, and $\mathbf{x}^* = \sum_{i=1}^n x_i^*$, and then, $\omega \mathbf{x}^* = (I - \lambda^*G)^{-1}(a - \xi \mathbf{x}^*) \cdot \mathbf{1} + (I - \lambda^*G)^{-1} \cdot \delta \cdot$

1. Based on equation (13), it should be $\omega \mathbf{x}^* = (a - \xi \mathbf{x}^*) \cdot \mathbf{b}(\mathbf{g}, \lambda^*) + \mathbf{b}_\delta(\mathbf{g}, \lambda^*)$,

which is also equivalent to $\mathbf{x}^*(\omega + \xi \cdot \mathbf{1} \cdot \mathbf{b}(\mathbf{g}, \lambda^*)) = a \cdot \mathbf{b}(\mathbf{g}, \lambda^*) + \mathbf{b}_\delta(\mathbf{g}, \lambda^*)$.

Finally, I get the equation (21). Given $a + \mathbf{b}_\delta(\mathbf{g}, \lambda^*) > 0$, and $\mathbf{b}(\mathbf{g}, \lambda^*) +$

$\mathbf{b}_\delta(\mathbf{g}, \lambda^*) \geq 1$, \mathbf{x}^* is the unique interior equilibrium, for all $i = 1, 2, \dots, n$. There is

only one maximum point, where $\frac{\partial^2 u_i}{\partial^2 x_i^2} = \theta_{ii} < 0$ is concave, definitely, x_i^* is an

player i 's unique solution in pure strategies. The other form of equation (21) is

similar as

$$x_i^* = \frac{b_i(\mathbf{g}, \lambda^*)}{b(\mathbf{g}, \lambda^*)} x^* + b_{\delta_i}(\mathbf{g}, \lambda^*) \quad (22)$$

where x^* is total sum of each player's Bonacich centrality, which is the aggregate equilibrium level. As a result, on one hand, the Nash equilibrium of the team game identifies that the best response of each player based on the peer effect in the network. Peer effect means passing interaction between players in the same team. On the other hand, with the increase of player's Bonacich centrality, he will motivate the perform action.

2.5 Key player Inter-centrality measure

Our objective is to discover a method that can find out key players in a team.

This method is defined as inter-centrality measure, named as ICM, which idea based on Katz-Bonacich centrality policy. I expect player i^* is the key player in the team. I mean, when we remove the player i^* from the network, this behavior makes the team performance suffers the most. The planner expects that reduce x^* optimally by picking the appropriate player from the population. Formally, from the Ballester (2006), $\max\{x^* - x^{*-i} | i = 1, 2, \dots, n\}$ is equal as

$$\min\{x^{*-i} | i = 1, 2, \dots, n\} \quad (23)$$

where it has at least one solution. We define that G^{-i} is the new adjacency matrix, which means the i -th row and column of G will be set to zeros. Our G is an asymmetric adjacency matrix.

Theorem 2: Given $\frac{\lambda}{\omega} < \frac{1}{\lambda_1}$, the key player i^* of team game solves equation (23),

and has highest inter-centrality of parameter λ^* in g , that is $c_{i^*}(g, \lambda^*) >$

$c_i(g, \lambda^*)$ for all $i = 1, 2, \dots, n$.

The proof of Theorem 2: the dependence of aggregate Nash equilibrium in the team game is the Bonacich centrality and the Bonacich centrality weighted by player's ability parameters, which are score ability, passes final third completion and total passes ability. I have already got well defined and nonnegative matrix M in equation (13), so $M^{\{-i\}} = [I - \alpha G^{\{-i\}}]^{-1}$, for all $i = 1, 2, \dots, n$, where $\lambda_1(G) > \lambda_1(G^{-i})$. As we all know, x^{*-i} also increases in $b(g^{-i}, \lambda^*)$, when $\alpha > 0$. At this time, $\min\{x^{*-i} | i = 1, 2, \dots, n\} = \min\{b(g^{-i}, \lambda^*) | i = 1, 2, \dots, n\}$, $b(g, \lambda^*)$ sum of $b_i(g, \lambda^*)$, and $b(g^{-i}, \lambda^*)$ is total centrality measure of player when player i removed. Using Lemma 1(Ballester, 2006): let matrix $M = [I - \alpha G]^{-1}$ be we

defined and nonnegative. Then, $m_{ji}(g, \alpha)m_{ik}(g, \alpha) = m_{ii}(g, \alpha)[m_{jk}(g, \alpha) - m_{jk}(g^{-i}, \alpha)]$, for all $k \neq i \neq j$. Define $b_{ji}(g, \lambda^*)$ as the contribution of Bonacich centrality of player i to player j in network G for all $i \neq j$, $l_i(g, \lambda^*) = b(g, \lambda^*) - b(g, \lambda^*) - r_\delta^i(g, \lambda^*)$; $b_{ji}(g, \lambda^*) = b_j(g, \lambda^*) - b_j(g^{-i}, \lambda^*)$; So $l_i(g, \lambda^*) = b_i(g, \lambda^*) + \sum_{i \neq j} b_{ji}(g, \lambda^*) + r_\delta^i(g, \lambda^*)$. Define the player i^* is key player, whose removal will result the largest value of $l_i(g, \lambda^*)$, $l_{i^*}(g, \lambda^*) \geq l_i(g, \lambda^*)$, for $i^* \neq i, i = 1, 2, \dots, n$.

$l_i(g, \lambda^*) = b_i(g, \lambda^*) + \sum_{i \neq j} [b_i(g, \lambda^*) - b_i(g^{-i}, \lambda^*)] + r_\delta^i(g, \lambda^*)$, using equation 17, we get $b_i(g, \lambda^*) = m_{ii}(g, \lambda^*) + \sum_{i \neq j} m_{ij}(g, \lambda^*)$. And then,

$$\begin{aligned} l_i(g, \lambda^*) &= b_i(g, \lambda^*) + \sum_{i \neq j} \sum_{k=1}^n [m_{jk}(g, \lambda^*) - m_{jk}(g^{-i}, \lambda^*)] + r_\delta^i(g, \lambda^*) \\ &= b_i(g, \lambda^*) + \sum_{i \neq j} \sum_{k=1}^n \left[\frac{m_{ji}(g, \lambda^*)m_{ik}(g, \lambda^*)}{m_{ii}(g, \lambda^*)} \right] + r_\delta^i(g, \lambda^*) \\ &= b_i(g, \lambda^*) \cdot \sum_{j=1}^n \left[\frac{m_{ji}(g, \lambda^*)}{m_{ii}(g, \lambda^*)} \right] + r_\delta^i(g, \lambda^*) \end{aligned}$$

Definition 1: For an asymmetric network G matrix, the highest inter-centrality measure determines the key players in the team. The inter-centrality measure(ICM):

$$c_i(g, \lambda^*) = b_i(g, \lambda^*) \cdot \left\{ \sum_{j=1}^n \left[\frac{m_{ji}(g, \lambda^*)}{m_{ii}(g, \lambda^*)} \right] \right\} + \sum_{j=1}^n m_{ji}(g, \lambda^*) \cdot \delta_i \quad (24),$$

where the last part of equation(25) is Receiving centrality of player weighted by goal probability and successful passing probability as $r_{\delta}(g, \alpha) = \sum_{j=1}^n m_{ji} \cdot \delta_i$ that measures player's contribution with his own ability parameter. Bonacich centrality of each player counts the number of paths in network G that start from i . But, inter-centrality measure takes into account the sum of, not only player i 's Bonacich centrality, but also the player i contributes to every other player's Bonacich centrality.

2.6 Key group of inter-centrality measure

In this part, I want to introduce how to discover key group in the team game, the idea from key player seeking. From the theorem 2, we know each individual equilibrium outcome is proportion to the Katz-Bonacich centrality. The key player is that his removal in the team will result the largest reduction the Nash equilibrium outcome. Previously, equation (19) exists at least one solution. Thus, I expect to pick up k players from the team, for all $k=1,2,\dots,n$. And we define k players as a group. If I find out one of a group, whose loss makes the team suffer

most, it will be the key group. Formally, it solves as $\max\{x^* -$

$x^{*-\{i_1, i_2, \dots, i_k\}} | i_1, i_2, \dots, i_k = 1, 2, \dots, k; i_r \neq i_s\}$, where it equals to

$$\max\{x^{*-\{i_1, i_2, \dots, i_k\}} | i_1, i_2, \dots, i_k = 1, 2, \dots, k; i_r \neq i_4\} \quad (25).$$

Now, the network adjacency matrix G will be $G^{-\{i_1, i_2, \dots, i_k\}}$ matrix. For an

asymmetric matrix G , the highest group inter-centrality measure determines the

key group of the team. According to equation (11), currently,

$$M^{-\{i_1, i_2, \dots, i_k\}} = [I - \alpha G^{-\{i_1, i_2, \dots, i_k\}}]^{-1}, i \in [1, n] \quad (26).$$

Theorem 2: let $M^{-\{i_1, i_2, \dots, i_k\}}$ exist and nonnegative. If $\frac{\lambda}{\omega} < \frac{1}{\lambda_1}$, the key group of size

$k\{i_1^*, i_2^*, \dots, i_k^*\}$ has the highest $k - th$ order group inter-centrality. i.e,

$$c_{\{i_1^*, i_2^*, \dots, i_k^*\}}(g, \alpha) \geq c_{\{i_1, i_2, \dots, i_k\}}(g, \alpha) \text{ for all } i_1, i_2, \dots, i_k = 1, 2, \dots, n.$$

The method to find out the key group named as group inter-centrality measure

(GICM) is

$$c_g = b'E(E'ME)^{-1}E'b + (1'ME)(E'\delta) \quad (27)$$

where E is $n \times k$ identity matrix defined as $E = \{e_{i_1}, e_{i_2}, \dots, e_{i_k}\}$, e_{i_r} being

the i_r^{th} column of the identity matrix E , $k \in [1, n]$. Removing the set of players from

the team will reduce interactive action between players, as well as decreases the probability of goal score and passing numbers.

Considering the size of group, if $k=1$, the equation (23) will count a single player's inter-centrality measure. Also, if $k=n$, at the same time, the $E = I$, it will immediately count the total summation of the Katz-Bonacich centrality measure of all players in the team. Actually, key groups of the team will supply information about the joint performance of players in the group, and will decrease the size of G . So, the choice of size k will discuss.

For my network example, the key group in size 2 is $\{A,H\}$, and the size of key group 3 is $\{A,F,H\}$. Note that two players $\{D,F\}$ with his own the highest inter-centrality measures are both not in the first key group in size 2, hence are more homogenous than a pair of players from the two different clusters. The key group in size 2 consists of players A and H that are part of the two different clusters, thus being less redundant with respect to each other than the pair $\{D,F\}$. Moreover, player A and H are less similar to all other players. At the same time,

player A, D, and F are not in the first key group in size 3, even though they are top 3 in the individual inter-centrality measure. The top one key group in size 3 is {A,F, H}. Further, the coincidence of the key group problem and cluster analysis can be also shown for larger size of the key group and lower level of agglomeration, which confirm our expectation that the key groups members are rather non-redundant with respect to each other in terms of their linking patterns in a network. Finally, key group includes less structure equivalent players. The results show that players with the highest inter-centrality measures may not necessarily make up of the top key group, because players may be redundant with respect to each other.

2.7 Approach Extension

2.7.1 Density

The definition of density in a network is the total number of edges divided by the total number of possible edges. It is 0.4 density measure for our “best friends

chooser". We can use density to compare different graphs. The higher density measure is the more strongly connected and better resist link failures graph can.

2.7.2 Closeness centrality

Closeness centrality of node i is the inverse of the average shortest path between node i and all other nodes in the network [15]:

$$C_c(i) = \frac{1}{\sum_{i \neq j} d_{ij} + \sum_{j \neq i} d_{ji}} \quad (28)$$

where d_{ij} is the shortest path between node i and node j . For some network graphs, some nodes have relatively high degree but low closeness. Closeness measure reflects which particular node is easier to be reachable, wants to be in the "middle" of things, not too far from the center. A high closeness score corresponds to a small average short path, indicating a well-connected player with the team.

2.7.3 Betweenness centrality

How many pairs of individuals would have to go through a particular node in

order to reach one another in the minimum number of links? Betweenness score defines the percentage of shortest path between node j and node k that go through node i , defined as

$$C_B(i) = \sum_{i \neq k \neq j} \frac{n_{jk}^i}{g_{jk}} \quad (29)$$

where g_{jk} is the total number of shortest path, n_{jk}^i is the number of shortest paths from j to k going through i . If many pairs of other nodes have passed through a node with shortest paths, this node will be high betweenness score. Thus, the damage of the node with high betweenness in the network structure results a greater effect on the flow in the network. In our example, let us see the removing effect of highest betweenness player D as figure [3b]. We can see when node D removed, it plays great damage on the 'information flow' of 'Best friends network'.

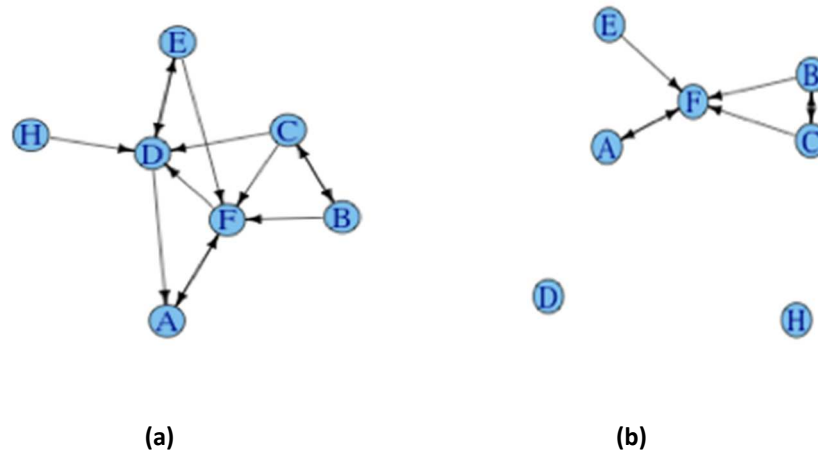


Figure 3: (b) The top Betweenness removal

2.7.4 Entropy and Randomness

In fact, the statistical definition of entropy can be regarded as randomness, because a uniform probability distribution reflects the largest randomness, a system following n status will have the greatest entropy when each status is equally likely [xx]. In such case, the probabilities become $p_i = p = \frac{1}{\Omega}$, where Ω is total number of small statuses. The entropy is $s = -\sum_{j=1}^{\Omega} \frac{1}{\Omega} \log_2\left(\frac{1}{\Omega}\right)$, which states that the larger the number of possible status, the larger the entropy. Actually, entropy in statistics field is the measure of uncertainty. In a system, distribution gets the greatest randomness and uncertainty and also the largest entropy. I

mean, I want to implement entropy to compare performance between MC and BW, in order to judge which team's distribution more uncertainly.

3 Impementation

3.1 Passing data extraction

Firstly, the passing data information is stored in a XML file, which describes the real-time game events. Manchester City soccer club supplied this data. For passing data extraction, my goal is to read that xml file into data structures by Java programing language. My strategy is to implemented native Java published API for representing XML as data structure, named JDOM. Dom stands for Document Object Model by W3C (`org.w3d.dom.Document`). From figure 4, we can see the architecture of DOM abstraction layer in Java. Java provides simple DOM implementation. Under Eclipse IDE for Java developer, I created a DocumentBuilder Factory instance, which allowed ease of switching parser

implementations. And then, I created a Document object, and used API (Application Program Interface) to manipulate, to parse as sequence of soccer events by XML file document element tag name "<Event>". I can get all players' passing information of that game, including Free Kicks, Goal Kicks, Corners, Crosses and Throw-ins. Finally, two passing network datasets have already extracted from one match real-time data of England Premier League 2011-2012 season: Manchester City (MC) vs. Bolton Wanderers (BW), happened on 08/21/2011. The result was that MC beat BW.

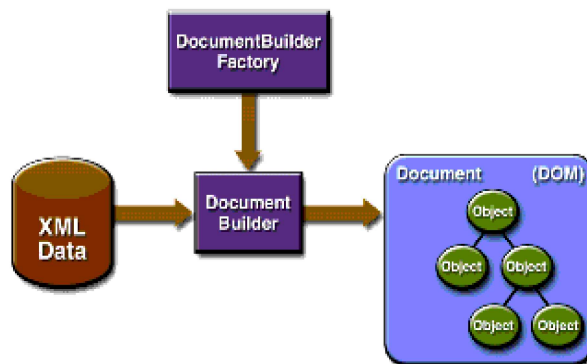


Figure 4: The architecture of DOM abstraction layer in Java

3.2 Introduce our passing network data

The Manchester City (MC) team beats Bolton Wanderers (BW) on 8/21/2011.

Passing data records successful passes performance completed by each soccer player of his own team. Firstly, there are two passing network graphs shown in figure 4 (a) and (b). These yellow, blue, green and orange nodes stand for goalkeepers, defenders, midfielders, and forwards, respectively. Also, MC 's lineup is 4-3-3 and BW's lineup is 4-4-2. Secondly, it is translated from passing graph network to passes network matrix, MC (14×14) and BW (13×13) shown in figure 5 (a) and (b). Each row of the passing network matrix is the number of successful passes made by each player, and each column represents number of successful receives by player.

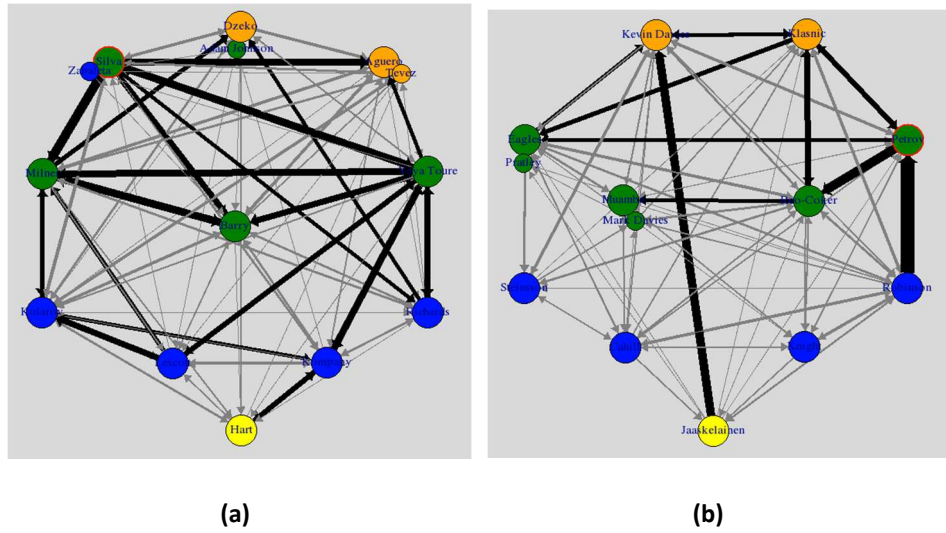


Figure 5: Passing network graph: (a) MC (b) BW

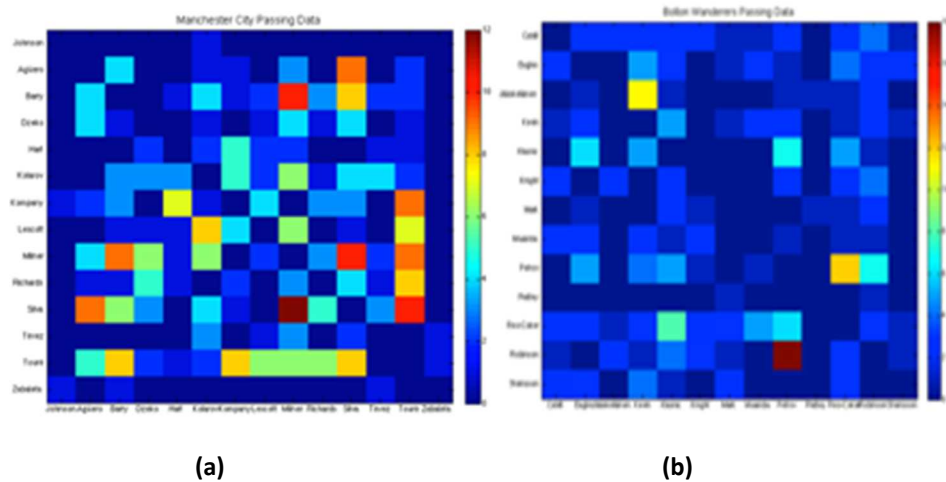


Figure 6: Passing Matrix: (a) MC (b) BW

3.3 Individual ability weights

First of all, I achieve each player's ability weights using the statistical dataset of England Premier League 2011-2012 season. Generally, for soccer game, soccer fans know goal ability plays a key role on game winning. So, the first ability

weight is score probability of each player in whole season 2011-2012. From figure 6, we see Manchester City's Forward, Aguero, shoots a lot and also goals a lot. In the contrast, Bolton Wanderers's forward shots also a lot, but he does not have good scores on goal. In addition, based on linear regression model, goals performance of MC is better than BW. Further, most soccer positions on some Midfielders and Defenders shot less and goal less.

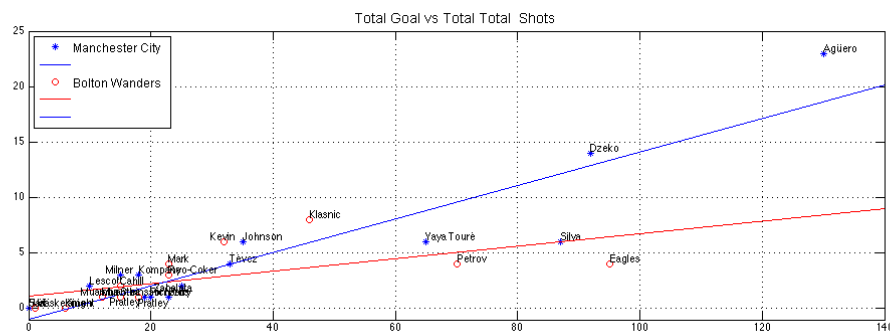


Figure 7: Goal probability

Second, considering most of soccer players' goal probability equals to zero on just one game, I add the second weight: Final Third passes completion.

Because a player who completes passes in the difficult field will make chances as assists. Looking at data analyzed by Jason Rosenfeld of *StatDNA* observes: to evaluate a player's passing skill, it depends on pass completion by the difficulty of

the pass field (the final third of the pitch) and under defensive pressure. See the figure xx below, Silva (MC) passes a lot in the final third area, and can get almost 90% completion. In contrast, Kevin (BW) also passes most in the final third field, he just succeed just around 24%. Comparing MC's performance with BW in Final Third field on figure 7, the average Final Third passes completion of MC is around 70%, which is better than BW (58%).

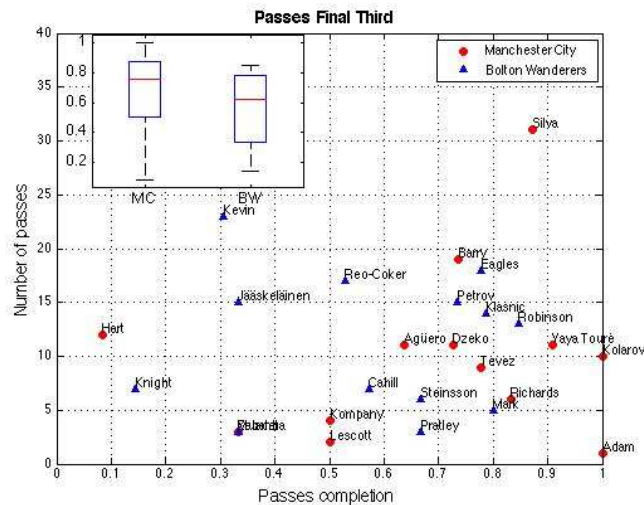


Figure 8: Final Third passes completion

The third weight is the pass ratio of successful and unsuccessful of each player in this MC vs. BW match. Figure 8 up histogram is Manchester City's each player's successful pass (blue) and unsuccessful pass (green) ratio, and the figure down histogram is Bolton Wanderers's. For each bar, the blue part plus

green part equals one. From figure 8, passing performance of Yaya Touré (MC) is the best. In the contrast, passing performance of Hart and Jaaskelainen, as goalkeeper role, play not well. For teams' comparison, the overall level in passing performance of MC's is better than BW's.

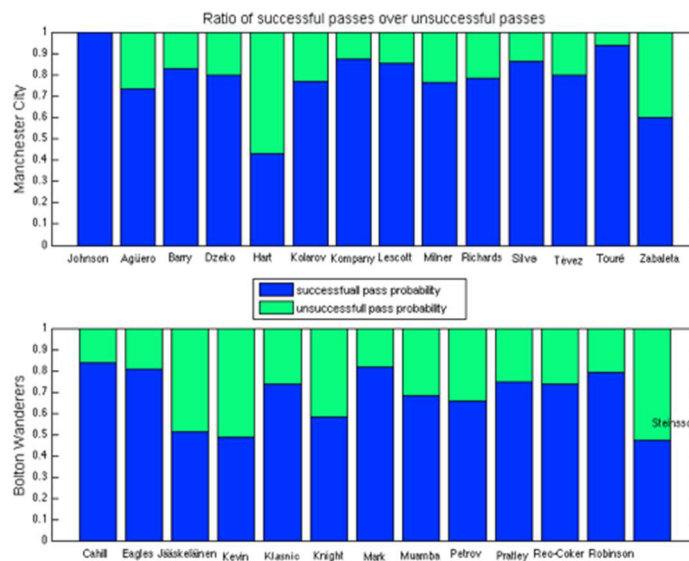


Figure 9: Successful/ unsuccessful passes ratio

2.3 Find key players and key groups by inter-Centrality (ICM)

Firstly, after we understand the Inter-Centrality measure (ICM), it will be utilized to identify the key player of MC and BW. The top ICM player is the key player of that team. And, looking the graph variation after the key player removal,

so as to see the key player's direct influence in his teams' passing network. Also, I will rank all players' ICM from high to low in his team, which help evaluate player's contribution on that team's outcome. Second, as we have known, ICM can also discover key groups. The key group sizes should be 2 players, 3 players and 4 players of each team. To rank Inter-Centrality measure reflects each groups' contribution from the highest Inter-Centrality value to the lowest Inter-Centrality value. My main option is top five groups in size 2, 3 and 4 from a great of permutation results. Through two passing network graph of key group removal in size 3 will be a way to reflect the impact of their disappearances.

3.5 Results verification of the key group

How can I see Inter-Centrality measure as reasonable? It can be verified by four approaches. First, taking into account market value 2012 and 2013, professional soccer players must have his own market value, which was published by web site (www.transfermarket.com). Market value will be a reliable source to proxy the salary of players. High salary is directly reflection of high skill.

Second, considering the effect of team unbeaten probability conditioned on the key player appearance or absence. Finally, players' media attention: the numbers of articles via *The Guardian* (a British national daily newspaper).

4 Experiment Results

4.1 Local centrality of each player

First, Local Centrality results shown on the figure 9 can directly know some players with high in-degree measure (blue square spots) stands for receiving a lot passes. And some players with high out-degree (green square spots) stands for passing a lot balls to others. Also, some players, having both high values (red square spots), can be regarded as directly good interactive passing performance with others. For team comparison, MC average passes around 28 is obviously better than BW's around 18.

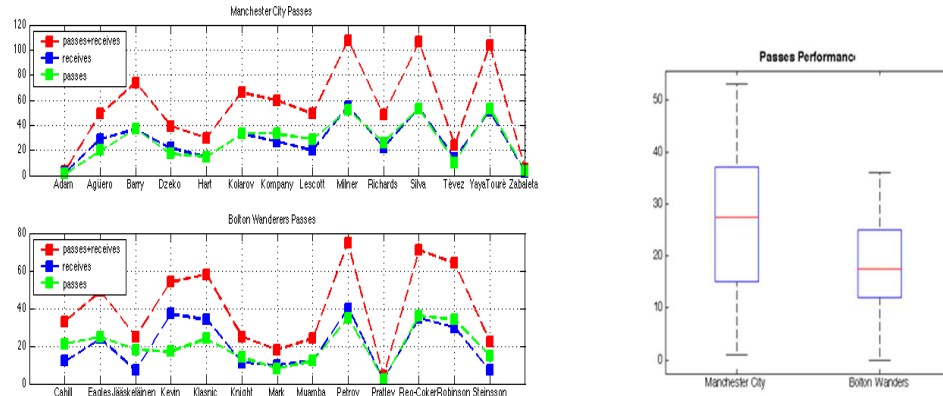


Figure 10: Local Centrality

2.4 Key players of Manchester City and Bolton Wanderers

Table 2 shows the standings of Inter-Centrality Measure of MC and BW.

That position's F, M, and D stand for Forward, Midfielder and Defender, respectively. And the figure 10 is the comparison of Inter-Centrality between MC

and BW. We can see the MC top 4 key players are midfielders. This implies

that MC's midfielders are so strong. From box plot of figure 10, we can see the

ICM average value in MC is higher than BW's. This implies top 4 key players in

BW over perform. Through figure 11 and figure 12, we can directly see that graph

variation between original passing network and the key player removal network

graph. Figure 11(c) and (f) show us the passing links graph of the MC's key and the BW's key player.

Name	Position	ICM	Name	Position	ICM
Silva (1)	M	138	Petrov(1)	M	149
Milner(2)	M	137	Reo-Coker(2)	M	140
Yaya Touré (3)	M	135	Robinson(3)	D	140
Barry	M	115	Klasnic	F	128
Kolarov	D	91	Eagles	M	108
Agüero	F	85	Kevin	F	100
Kompany	D	81	Cahill	D	69
Richards	D	80	Knight	D	59
Lescott	D	75	Muama	M	56
Dzeko	F	59	Jääskeläinen	G	41
Hart	G	31	Steinsson	D	40
Tévez	F	30	Mark	M	33
Adam	M	4	Pratley	M	5
Zabaleta	D	3			

(a)

(b)

Table 1: ICM standings: (a) MC (b) BW

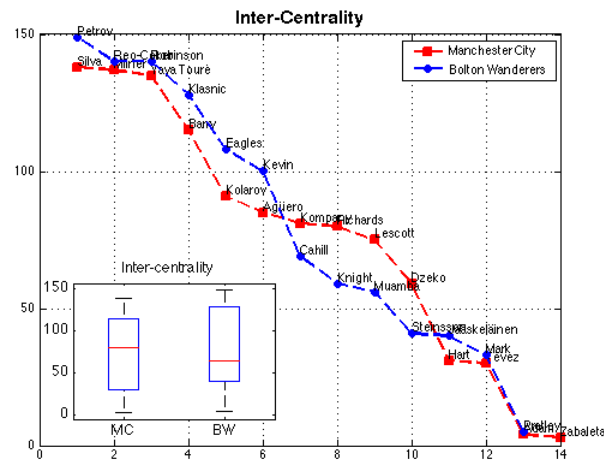


Figure 11: ICM rankings

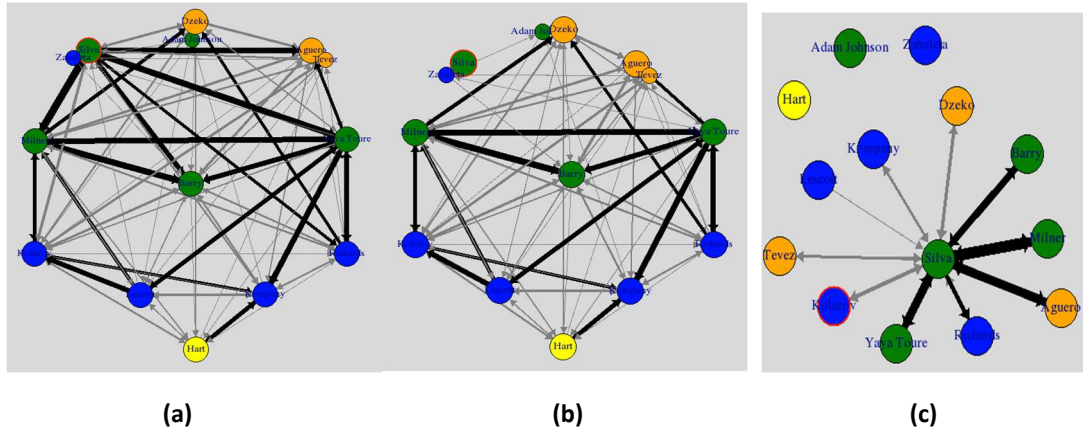


Figure 12: MC network graph: (a) Original (b) The key player removal (c) The key player links

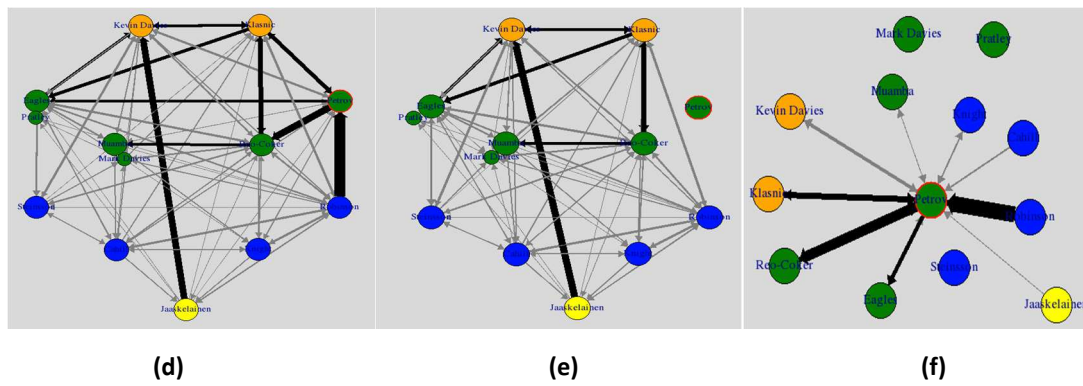


Figure 13: BW network graph: (d) original (e) The key player removal (f) The key player links

4.3 Key groups of MC and BW in size 2, 3, and 4

Rank	Manchester City	ICM	Bolton Wanderers	ICM
1	{Silva, Yaya Touré}	191	{Kevin, Klasnic }	213
2	{Milner, Yaya Touré}	190	{Klasnic, Petrov}	212
3	{Milner, Silva}	189	{Klasnic, Reo-Coker}	210
4	{Barry, Silva}	178	{Kevin, Reo-Coker}	204
5	{Barry, Yaya Touré}	177	{Petrov, Reo-Coker}	202

(a)

Rank	Manchester City	ICM	Bolton Wanderers	ICM
1	{Milner, Silva, Yaya Touré}	228	{Klasnic, Petrov, Reo-Coker}	238
2	{Barry, Silva, Yaya Touré}	219	{Kevin, Klasnic, Reo-Coker}	236
3	{Barry, Milner, Yaya Touré}	216	{Klasnic, Kevin, Petrov}	235
4	{Barry, Milner, Silva}	215	{Klasnic, Eagles, Petrov}	234
5	{Kolarov, Silva, Yaya Touré}	214	{Klasnic, Eagles, Reo-Coker}	233

(b)

Rank	Manchester City	ICM	Bolton Wanderers	ICM
1	{Barry, Milner, Silva, Yaya Touré}	252	{Klasnic, Eagles, Petrov, Reo-Coker}	259
2	{Barry, Silva, Yaya Touré, Kolarov}	249	{Klasnic, Kevin, Reo-Coker, Petrov}	258
3	{Agüero, Milner, Silva, Yaya Touré}	246	{Klasnic, Kevin, Eagles, Reo-Coker}	254
4	{Milner, Richards, Silva, Yaya Touré}	244	{Kevin, Klasnic, Eagles, Petrov}	254
5	{Kompany, Milner, Silva, Yaya Touré}	243	{Klasnic, Petrov, Reo-Coker, Robinson}	253

(c)

Table 2 : Key groups: (a) in size 2 (b) in size 3 (c) in size 4

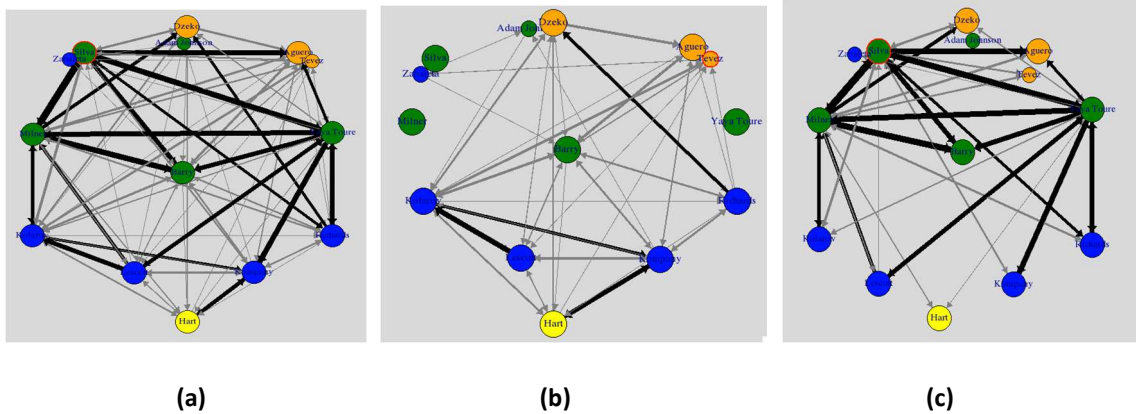


Figure 14: MC network graph: (a) Original (b) The key group Removal (c) The key group links

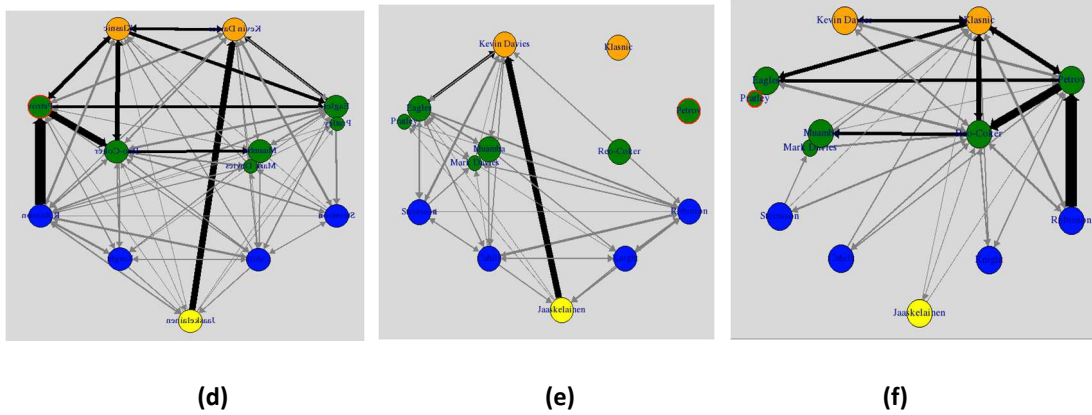


Figure 15: BW network graph: (a) Original (b) The key group removal (c) The key group links

2.5 Result verification

	Position	Key player (rank)	Market value 2011/12	Market value 2012/2013	The Guardian (Article Numbers)	unbeaten probability (Appearance)	unbeaten probability (Absence)
Silva	Attacking Midfield	1	32,00 M€	50,00 M€	199,000	91%	80%
Milner	Right Midfield	2	20,00 M€	21,00 M€	85,900	86%	88%
Yaya Touré	Center Midfield	3	25,00 M€	36,50 M€	80,600	90%	75%
Barry	Defensive Midfield	4	17,00 M€	10,00 M€	72,700	86%	100%
Agüero	Center Forward	6	36,00 M€	51,00 M€	286,700	86%	100%

(a)

	Position	Key player (rank)	Market Value 2011/2012	Market Value 2012/2013	The Guardian (Article Numbers.)	unbeaten probability (Appearance)	unbeaten probability (Absence)
Petrov	Right wing	1	5,50M€	1,00 M€	1190	50%	14%
Reo-Coker	Defensive Midfield	2	5,50 M€	5,00 M€	779	42%	0%
Robinson	Center Back	3	2,50 M€	0,5 M€	638	31%	11%
Klasnic	Center Forward	4	2,50M€	1,25 M€	348	60%	50%
Eagle	Left Wing	5	3,5 M€	5,50 M€	13,000	35%	100%

(b)

Table 3: Result verification

4.5 Closeness Centrality results in MC and BW

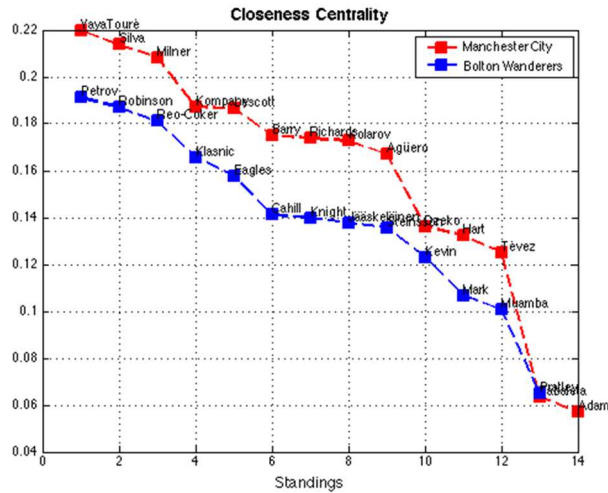


Figure 16: Closeness Centrality rankings

4.6 Betweenness Centrality results

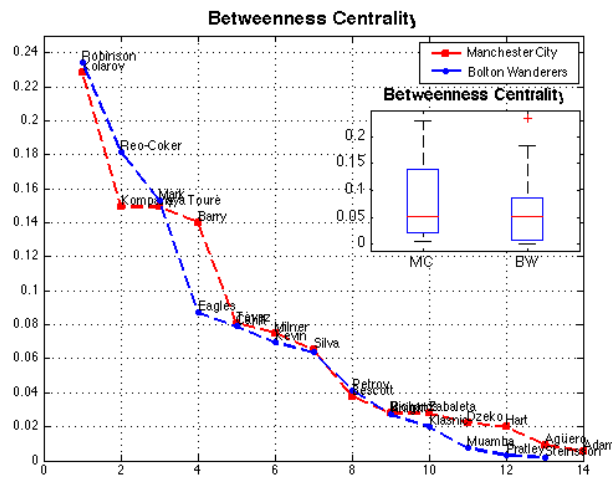


Figure 17: Betweenness Centrality rankings

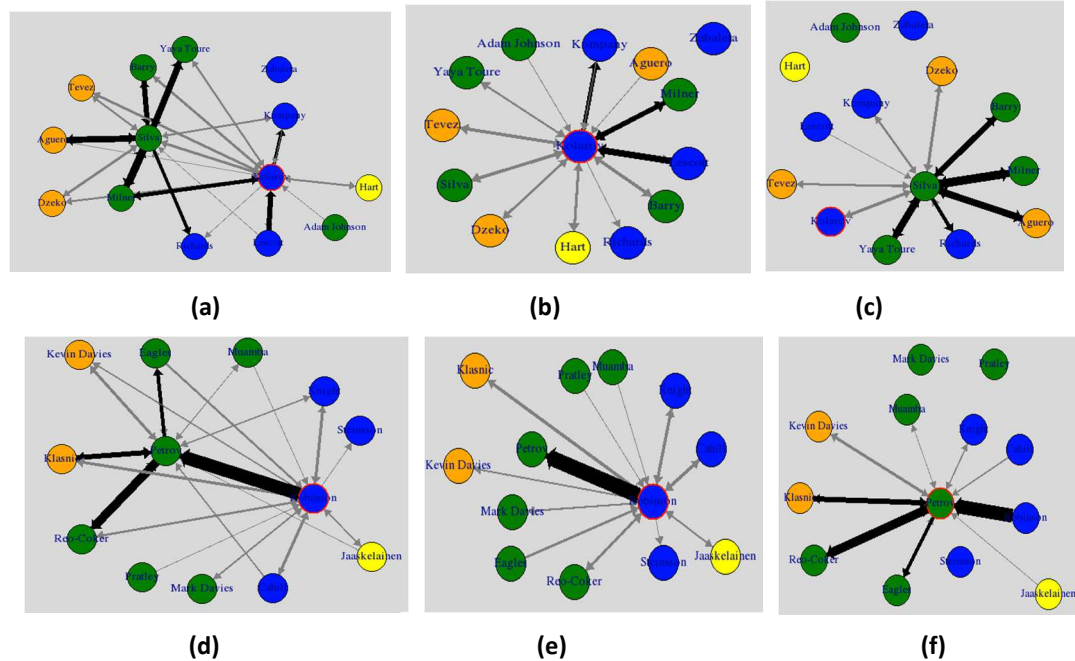


Figure 18: The top Betweenness and the key player graph: (a) MC (d) BW;

The top Betweenness player links: (b) MC (e) BW; The key player links: (c) MC (f) BW

5 Observations

5.1 Key players and key groups analysis

Silva is the MC key player "discovered" by our analysis. Soccer amateurs know he is a recognized world-class player. First, his market value went up from 32,00 Million € in 2011/2012 season to 50,00 Million € in 2012/2013 season, and is the second high price in MC team (Table 4). Secondly, when Silva appears in

the games, the MC unbeaten probability is up to 91% with him, but is 60% without him. Furthermore, the number of articles centered on Silva, 19,900, manifests his good skill get high media attention. In addition, the market value of Milner, Yaya Touré, and Agüero also went up in 2012/2013 Season. Also, appearances of these players have a great impact on MC's unbeaten probability. On the other side, Petrov is the key BW player of "discovered" by our analysis. He is a recognized as the top player in his team. Also, the team with Petrov is unbeaten 50% of time, but, only 14% of time without him. Otherwise, the absence of Reo-Coker and Robinson (the key group player) results in Unbeaten 0% and 11% of Bolton Wanderers.

For key groups aspect, looking at the table 3 (b), figure 13 (b) and figure 14 (e), top key groups removal in size 3 definitely and severely damage the passing network structures of each team. The MC key group in size 2, 3, and 4 are made up of midfielders, which means midfielders play the important role in the pitch, and is pioneered by Silva, as an attacking midfielder. The BW key group in size 3 and

4 also center on midfields. So, we can say MC and BW have something in common that midfielders' performance is crucial for the game. Further, the BW key group removal results in more damage of its passing, because the right side of lineup is damaged a lot, which illustrates the BW key group 'over-performance' in the match. In the end of the 2011-2012 season, Bolton Wanderers was relegated from England Premier League. And one of BW top key group (Cahill) transferred to a top team named Chelsea. Also, Key players "discovered" by our technique correlate well with their media attention. Comparing to goalkeepers, both teams' goalkeeper have very small ICM values, which verified that goalkeepers play a different role on the match.

5.2 Extensions

Firstly, using Closeness Centrality, we can know how easy a particular player is reachable within a team. Figure 15 shows Yaya Toure (MC) and Petrov are top one Closeness measure in his own team. Combined Local Centrality with Closeness Centrality (see figure 9 and figure 15), Yaya Toure and Petrov also

have very high number of passes and receives. But, it is different that players has relatively low degree but high closeness, such as Kompany (MC); or high degree but relatively low closeness, such as Kevin (BW).

Secondly, Betweenness measures how “ball-flow” is dependent on a particular player between others players, not how good-link is with the others. The results of Betweenness in MC vs. BW match shown by figure 17. Kolarov(MC) and Robinson(BW) get top one Betweenness centrality. Assuming Kolarov and Robinson are removed, do they damage the “ball-flow” of his own team? We can found out Kolarov is directly reachable to the others except only Zabaleta(MC), also Robinson(BW) links with all of his teammates. Both removal will obviously damage the “ball flow” of their own teams. Top one Betweenness player almost passes the ball to all his teammates, act like a “bridge” in the game. In the contrast, Silva (MC) and Petrov (BW) have highest ICM but relatively less connections with their own teammates. That’s why Silva (MC)’ and Petrov (BW)’s Betweenness centrality are smaller than Kolarov and Robinson. But, both top one

Betweenness players result in not being popular than Silva, which will be illustrated by PageRank centrality. Kompany (MC) and Mark (BW) have high Betweenness but relative low Local centrality. Furthermore, players having low closeness and low Betweenness can almost be identified as forwards, because forward (striker) mainly task is to receive passes to make chance to shots, and does not act as critical piece of “ball-flow” between other players as usual.

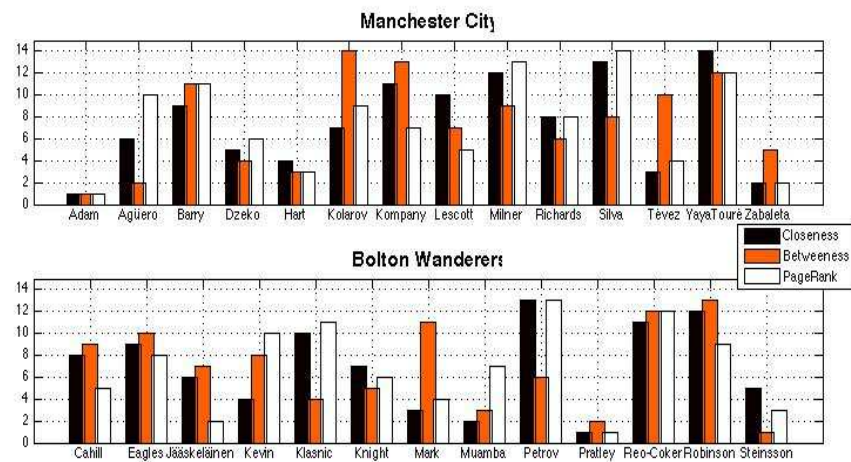


Figure 19: Player integrated centralities rankings

The figure 19 results show that forwards as Agüero, Dzeko and Tévez in MC giving low closeness centrality and Betweenness centrality, and also Klasnic(BW) giving very low Betweenness and Kevin(BW) giving low closeness. Especially,

Silva (MC) acts as Midfielder generally, but in this game, he played a forward role, giving the highest PageRank centrality, very high closeness. Those are explained the fact that he is a well-connected attacking midfielder, who rank No. 1 on integrated centrality ranking, and also shot a goal on this game.

Table [3] has shown comparison between Manchester City and Bolton Wanderers. Firstly, the scores for both teams point to a big difference in total number of passes, which means MC players' interaction are better. And, MC with higher density implies the fact that MC is more strongly connected and better to resist turnovers. In addition, MC with higher Closeness Centrality illustrates players are much closer to each other, which is a reflection of the 'total football'. In addition, the largest entropy equals to 3.585 in our passing network system. According to entropy of ICM and Betweenness, MC is closer to the largest entropy. Each player's passing performance of Manchester City is more evenly spreading out. But Bolton Wanderers's passing performance is relatively not evenly distributed. Thus, BW depends more on the a few players. I mean Petrov and

Reo-Coker play a predominant role in BW's passing strategy. Furthermore, the implementation of entropy method can help us to understand that team MC plays more like "total football". MC's Entropy in Betweenness is larger than BW's, so Manchester City players played more cohesive, balanced expectation. Overall, it is apparently that BW is an unbalanced use of pitch from figure 14, and obvious preference to the right side; however, MC shows more balanced use of pitch, especially on opposition half of pitch, which is indicative of attacking effectively and efficiently.

Team	Total Passes	Density	Closeness	Betweenness Entropy	ICM Entropy
MC	382	2.10	2.22	3.20	3.45
BW	261	1.68	1.83	3.04	3.41

Table 4: Team comparison

6 Conclusion

Firstly, the inter-centrality measure (ICM) produces good and reasonable results as verified by market value, unbeaten probabilities with/without him, and media attention. In addition, in order to increase the chance of winning, first, team should increase individual, group in ICM. Team should increase the passing sequence number. The winning team has good passing interaction than the losing team. Second, "Take-out" the key players/groups of the opposing team. There are some good strategies, for example, by closer man-marking, by increasing the number of defensive midfielders. Because it is apparently that the most top key groups are made up of Midfielders. Furthermore, Closeness Centrality and Betweenness Centrality hint coaches to know players' passing characteristics. Take good defense on top closeness and top Betweenness will better cut team's connection and prevent the "ball-flow" of that team. And Density is a way to judge team cohesion. Entropy is a factor that can judge team's balanced expectation. Finally, ICM can be used as an important factor to determine the potential market

value of a player, except for goalkeeper position, who plays different role in the soccer game.

References

[1] C. Anderson and D. Sally, *The numbers game: why everything you know about football is wrong*, Penguin UK, 2013.

[2] Anderson, C., & Sally, D. (2013). *The numbers game: why everything you know about football is wrong*: Penguin UK.

[3] Ballester, C., Calvó-Armengol, A., & Zenou, Y. (2006). Who's Who in Networks. Wanted: The Key Player. *Econometrica*, 74(5), 1403-1417.

[4] Bonacich, P., & Lloyd, P. (2001). Eigenvector-like measures of centrality for asymmetric relations. *Social Networks*, 23(3), 191-201.

[5] Borgatti, S. (2006). Identifying sets of key players in a social network. *Computational & Mathematical Organization Theory*, 12(1), 21-34.

[6] Borgatti, S. P. (2005). Centrality and network flow. *Social Networks*, 27(1),

55-71.

[7] Calvó-Armengol, A., Patacchini, E., & Zenou, Y. (2009). Peer Effects and Social Networks in Education. *The Review of Economic Studies*, 76(4), 1239-1267.

[8] Finnveden, G., Hauschild, M. Z., Ekvall, T., Guinée, J., Heijungs, R., Hellweg, S., Suh, S. (2009). Recent developments in Life Cycle Assessment. *Journal of Environmental Management*, 91(1), 1-21.

[9] Freeman, L. (1977). A Set of Measures of Centrality Based on Betweenness. *Sociometry*, 40(1), 35-41

[10] Friedkin, N. E. (1991). Theoretical Foundations for Centrality Measures. *The American Journal of Sociology*.

[11] Jackson, M. O. (2008). *Social and economic networks* (Vol. 3): Princeton University Press Princeton

[12] Katz, L. (1953). A new status index derived from sociometric analysis. *Psychometrika*, 18(1), 39-43

[13] Lago-Ballesteros, J., & Lago-Peñas, C. (2010). Performance in Team Sports: Identifying the Keys to Success in Soccer *Journal of Human Kinetics* (Vol. 25, pp. 85).

[14] Page, L., Brin, S., Motwani, R., & Winograd, T. (1999). The PageRank Citation Ranking: Bringing Order to the Web. Stanford InfoLab.

[15] Peña, J. L., & Touchette, H. (2012). *A network theory analysis of football strategies*. Paper presented at the Euromech Physics of Sports Conference.

[16] Sarangi, S., & Unlu, E. (2010). Key Players and Key Groups in Teams: A Network Approach Using Soccer Data. Retrieved 1/1, 2014

[17] Temurshoev, U. (2008). Who's Who in Networks. Wanted: the Key Group. Retrieved 12/31, 2013

[18] Wasserman, S. (1994). *Social network analysis: Methods and applications* (Vol.8): Cambridge university press.