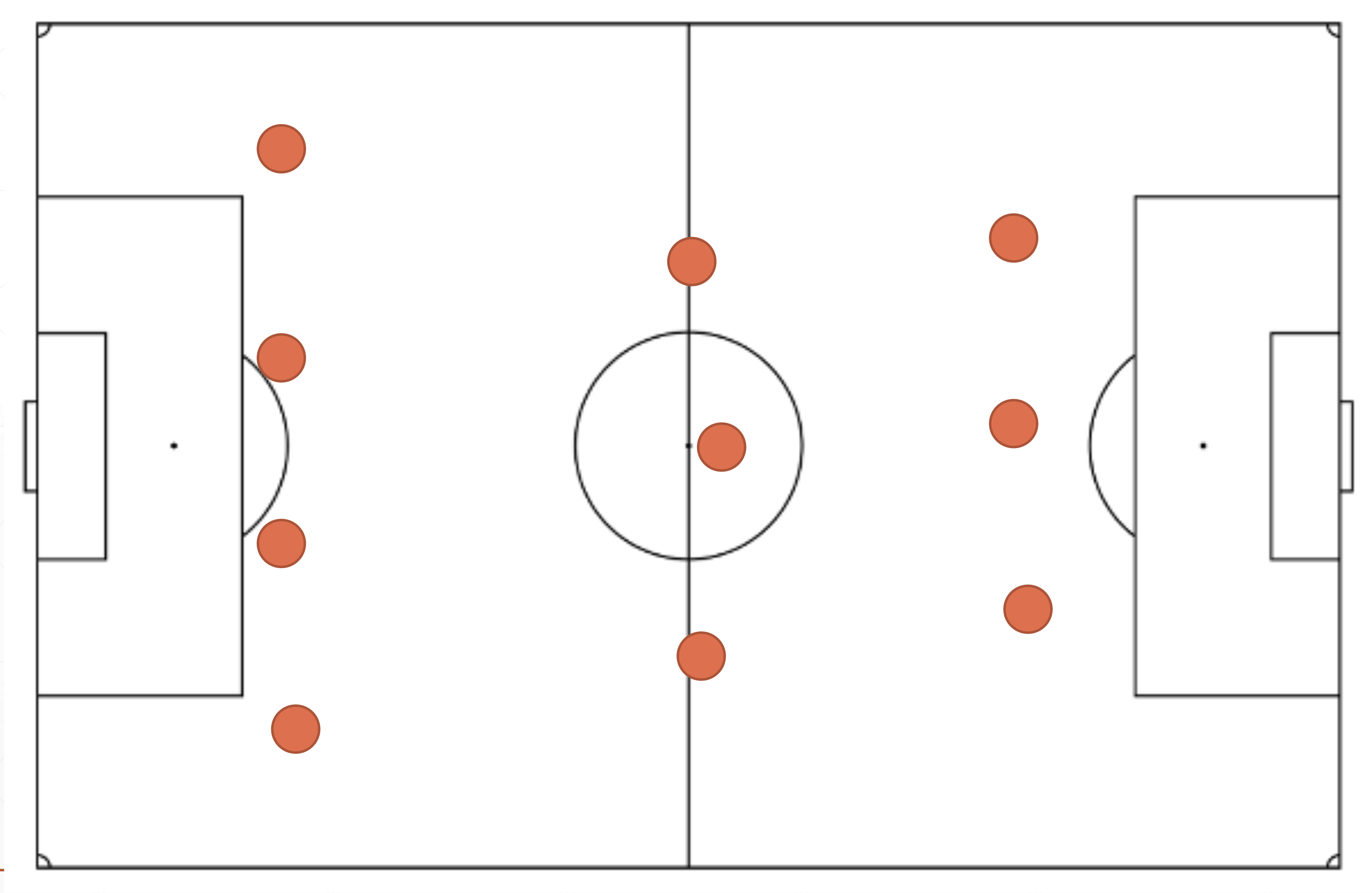
Analysis of Soccer Team Players' Connection: A Matrix Approach to Network Study

Homer (Nutthakorn Intharacha)

Background of Soccer Game

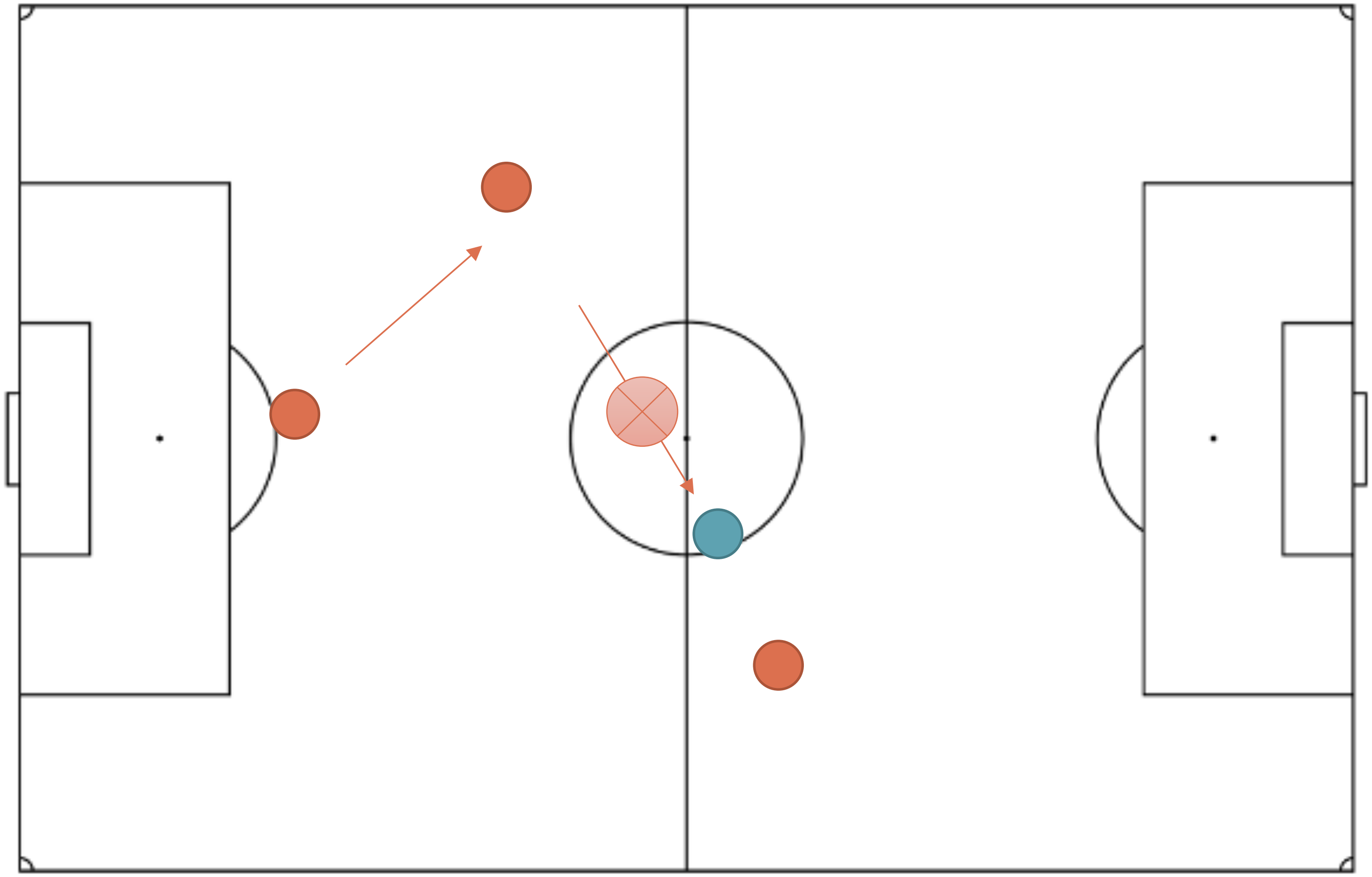
- Classical formations in modern soccer are 1-3-4-2-1, 1-4-2-3-1, 1-4-3-3, etc..
- But on average, there are 4 players on the defensive zone, 3 on the mid field , and 3 on the offensive zone.
- Each zone position can be broken down into more specific roles
- Defender (4): 1 Right / 1 Left / 2 central
- Midfielder (3) : 1 Right / 1 Left / 1 central
- Forward (3) : 1 Rightwinger / 1 Leftwinger / 1 central

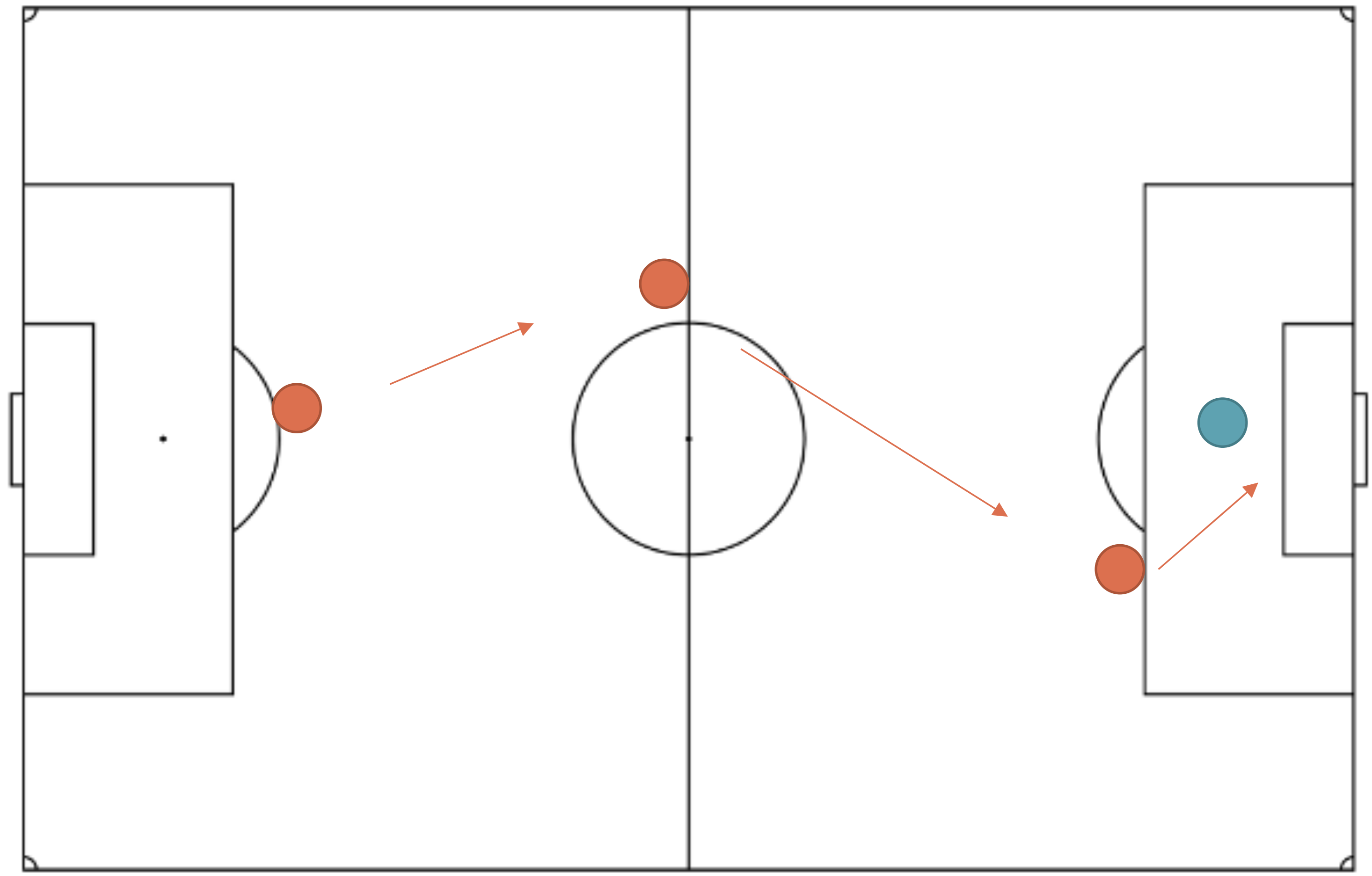


Introduction

- Soccer games can be viewed as a network/ graph with fixed number of nodes (11 players) and variation of edges for each game (successful passes between players)
- At the same time, I think we can apply Matrix approach to study players' connections
- Because of these metrics, we can identify centroid player, the team's connectivity, or even the clusters inside the team
- I wish to specifically analyze how each player contribute to the **offensive play** i.e. the process of building the attack which results in shots.
- In fact this *attacking process* is defined by Bourbousson *et al*. (2010) and Passos *et al*. (2011) as they came up with the term ' **Unit of Attack**'

Example of attacking process





Matrix Methodology

- Adjacency matrix : A =[a_{ij}] ϵ R^{nxn}
- Define: a_{ij} = 1 if there exist connection between node i and j ; a_{ij} = 0 otherwise
- Here this A matrix represents a successful pass, and the diagonal elements are set equal to 1 as it identify if player i participate in the attacking process.
- For example, if five players (4 defenders and a goalkeeper) did not contribute to the offensive play, the A matrix would be

	1	2	3	4	5	6	7	8	9	10
1	1	1	1	1	1	1	0	0	0	0
2	1	1	1	1	1	1	0	0	0	0
3	1	1	1	1	1	1	0	0	0	0
4	1	1	1	1	1	1	0	0	0	0
5	1	1	1	1	1	1	0	0	0	0
6	1	1	1	1	1	1	0	0	0	0
7	0	0	0	0	0	0	0	0	0	0
8	0	0	0	0	0	0	0	0	0	0
9	0	0	0	0	0	0	0	0	0	0
10	0	0	0	0	0	0	0	0	0	0
11	0	0	0	0	0	0	0	0	0	0

Matrix Methodology

- I think we need to take an account of how different edges and vertices affect this network
- Employing **The Edge-Weighted Edge-Adjacency Matrix (**^{w}A), we can formulate a better players cooperation model. (Using Matlab *wgPlot* package)
- The ^{w}A can be defined by the sum of all adjacency graphs each one generated by a single offensive play
- Note: w_{ij} represent a weighted edge between players i and j. In other words, it shows how strong of the cooperation between players i and j. and is proportional to the number of offensive plays.
- For simplicity, I will denote w_{ij} a total number of successful passes from player i to player j in the attacking plays

Example ^{w}A

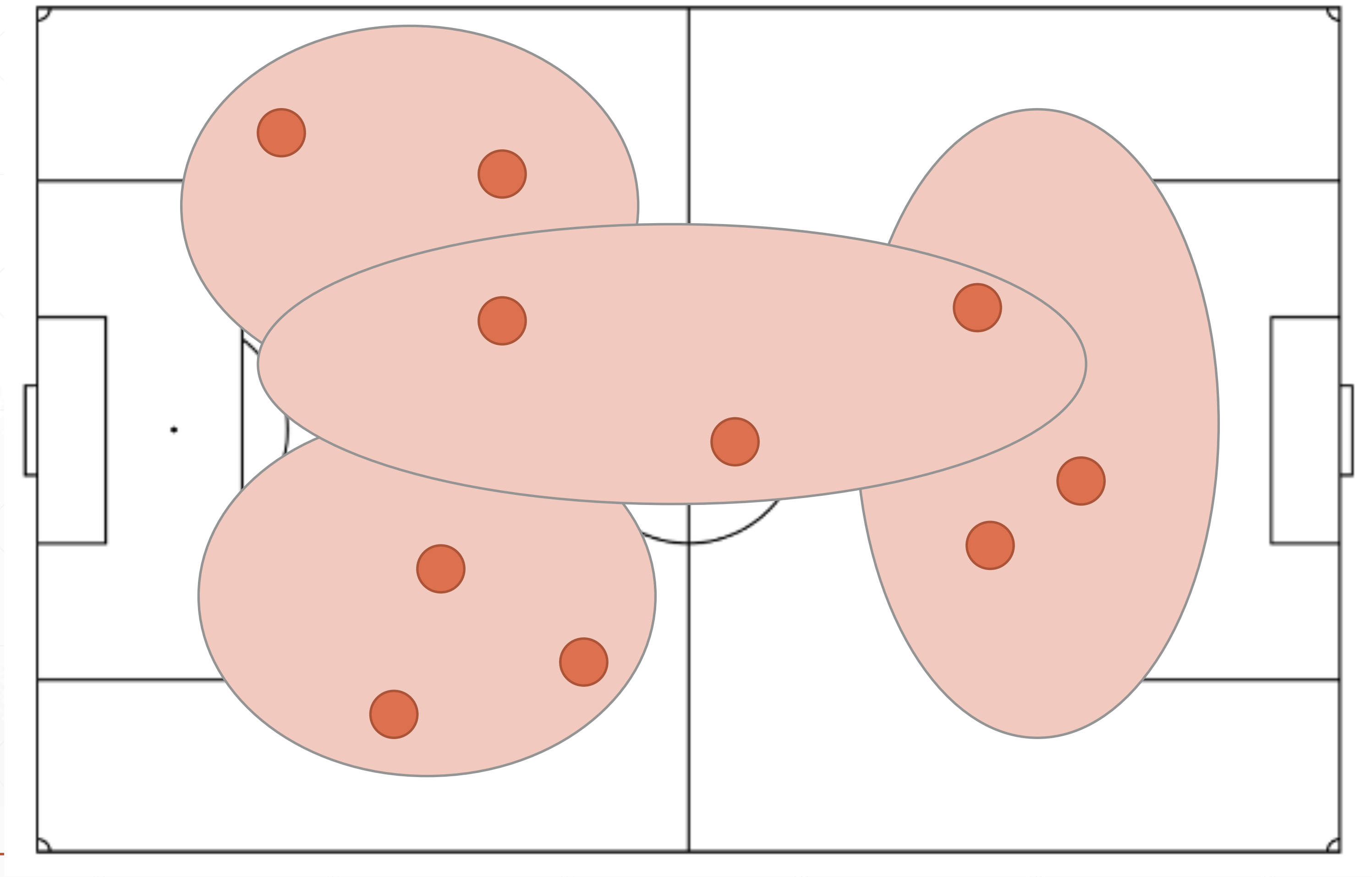
	1 GK	2 RD	3 CD	4 CD	5 LD	6 M	7 LM	8 RM	9 RF	10 LF	11 S
1 GK	1	2	2	0	0	2	0	1	0	0	1
2 RD	1	1	0	1	2	3	2	1	2	1	0
3 CD	2	1	1	0	2	4	1	2	1	0	0
4 CD	0	0	2	1	2	1	2	3	2	1	0
5 LD	0	0	1	0	1	3	2	1	1	5	2
6 M	0	1	0	1	1	1	5	4	3	4	4
7 LM	1	0	1	2	2	3	1	1	1	5	3
8 RM	0	0	0	3	0	2	2	1	7	1	4
9 RF	0	2	1	0	1	3	0	3	1	2	6
10 LF	2	1	2	1	2	4	3	2	2	1	3
11 S	0	0	0	1	0	2	1	1	3	3	1

Clustering within a team

- Now after setup ${}^{w}A$ matrix, I want to find communities i.e subgroups with in a team
- Graph Theory provides a way to constitute partitions and I can use it to generate communities
- Formally graph partition is defined by $G = (V,E)$. I can then partition G into smaller components i.e collection $P = \{V_1, .., V_k\}$ where $k < 11$ in our case

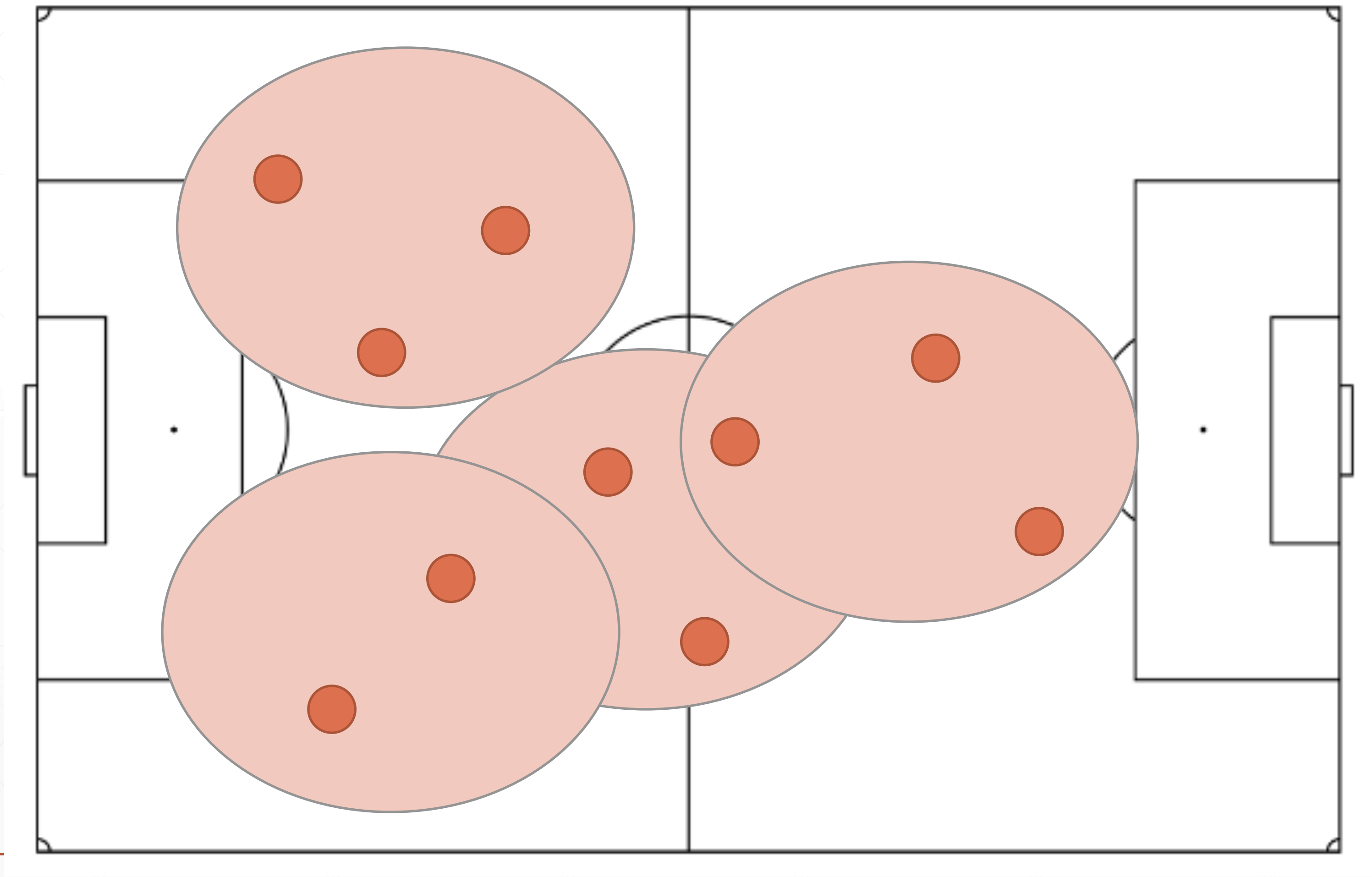
Clustering within a team

▪



Clustering within a team

▪



Clustering within a team

- To allow the use of Network model, I construct a new relative **weighted adjacency matrix** $A_r = [r_{ij}] \in R^{nxn}$
- $r_{ij} = wij/\boldsymbol{max}\, wA$ if $i \neq j$ and $r_{ij} = w_i$ if $i=j$
- Note $0 \leq r_{ij} \leq 1$
- **max** wA ($i \neq j$) represent the players that participate most in the offensive plays
- At this point, we have came up with a very powerful matrix model ready to be analyzed on both the macro (as a whole team) and micro (as individual) levels.

Example A_r

	1 GK	2 RD	3 CD	4 CD	5 LD	6 M	7 LM	8 RM	9 RF	10 LF	11 S
1 GK	0.14	0.29	0.29	0.00	0.00	0.29	0.00	0.14	0.00	0.00	0.14
2 RD	0.14	0.14	0.00	0.14	0.29	0.43	0.29	0.14	0.29	0.14	0.00
3 CD	0.29	0.14	0.14	0.00	0.29	0.57	0.14	0.29	0.14	0.00	0.00
4 CD	0.00	0.00	0.29	0.14	0.29	0.14	0.29	0.43	0.29	0.14	0.00
5 LD	0.00	0.00	0.14	0.00	0.14	0.43	0.29	0.14	0.14	0.71	0.29
6 M	0.00	0.14	0.00	0.14	0.14	0.14	0.71	0.57	0.43	0.57	0.57
7 LM	0.14	0.00	0.14	0.29	0.29	0.43	0.14	0.14	0.14	0.71	0.43
8 RM	0.00	0.00	0.00	0.43	0.00	0.29	0.29	0.14	1.00	0.14	0.57
9 RF	0.00	0.29	0.14	0.00	0.14	0.43	0.00	0.43	0.14	0.29	0.86
10 LF	0.29	0.14	0.29	0.14	0.29	0.57	0.43	0.29	0.29	0.14	0.43
11 S	0.00	0.00	0.00	0.14	0.00	0.29	0.14	0.14	0.43	0.43	0.14

Macro analysis

- The first approach which is to analyze **Connectivity** is widely used in the literature.
- This analysis will distinguish a vertex of a network
- Define players' connectivity :
  - k_i = *sum of* ***connection weights*** *between player i and other players*
  - k_i = # ball passes + # ball received
  - *The most cooperative player :* $k_{max} = \max k_i$
- Therefore, we define the **Scaled Connectivity** as $S_i = ki/kmax$

	1 GK	2 RD	3 CD	4 CD	5 LD	6 M	7 LM	8 RM	9 RF	10 LF	11 S	K_i^c
1 GK	0.14	0.29	0.29	0.00	0.00	0.29	0.00	0.14	0.00	0.00	0.14	1.29
2 RD	0.14	0.14	0.00	0.14	0.29	0.43	0.29	0.14	0.29	0.14	0.00	2.00
3 CD	0.29	0.14	0.14	0.00	0.29	0.57	0.14	0.29	0.14	0.00	0.00	2.00
4 CD	0.00	0.00	0.29	0.14	0.29	0.14	0.29	0.43	0.29	0.14	0.00	2.00
5 LD	0.00	0.00	0.14	0.00	0.14	0.43	0.29	0.14	0.14	0.71	0.29	2.29
6 M	0.00	0.14	0.00	0.14	0.14	0.14	0.71	0.57	0.43	0.57	0.57	3.43
7 LM	0.14	0.00	0.14	0.29	0.29	0.43	0.14	0.14	0.14	0.71	0.43	2.86
8 RM	0.00	0.00	0.00	0.43	0.00	0.29	0.29	0.14	1.00	0.14	0.57	2.86
9 RF	0.00	0.29	0.14	0.00	0.14	0.43	0.00	0.43	0.14	0.29	0.86	2.71
10 LF	0.29	0.14	0.29	0.14	0.29	0.57	0.43	0.29	0.29	0.14	0.43	3.29
11 S	0.00	0.00	0.00	0.14	0.00	0.29	0.14	0.14	0.43	0.43	0.14	1.71
K_i^r	1.00	1.14	1.43	1.43	1.86	4.00	2.71	2.86	3.29	3.29	3.43	

	1 GK	2 RD	3 CD	4 CD	5 LD	6 M	7 LM	8 RM	9 RF	10 LF	11 S
K_i	2.285714	3.142857	3.428571	3.428571	4.142857	7.428571	5.5714286	5.7142857	6	6.571429	5.142857

	1 GK	2 RD	3 CD	4 CD	5 LD	6 M	7 LM	8 RM	9 RF	10 LF	11 S
S_i	0.307692	0.423077	0.461539	0.461539	0.557692	1	0.7500001	0.7692309	0.8076925	0.884616	0.692308

Macro analysis

- Another approach to this analysis is to measure the **degree of interconnectivity** in the neighborhood of each player
- Recall the degree k_i of a node i is defined as the number of its neighborhood
- $k_i = \sum_j a_{ij}$
- This tendency of the neighbors of any node i to connect to each other, is called clustering and is quantified by the **clustering coefficient** C_i
- C_i can be interpreted as the fraction of triangles in which node i participates
- By convention,

$$C_i = \frac{n_i}{k_i(k_i-1)} = \frac{\sum_{j,k} a_{ij}a_{jk}a_{ki}}{k_i(k_i-1)}, \; k_i \neq 0,1$$

Macro analysis

- Using ***Weighted Clustering Coefficient*** proposed by Zhang et. al. (2005)

$$ClusterCoef_i = \frac{\sum_{j \neq i} \sum_{j \neq i} r_{ij} r_{ji} r_{ki}}{\left(\sum_{j \neq i} r_{ij}\right)^2 - \sum_{j \neq i} \left(r_{ij}\right)^2}$$

- Recall that r_{ij} = $wij/\boldsymbol{max}\, wA$
- The higher the coefficient of a player, the higher is the cooperation among his teammates

Proof

$$C^{Z}_{w,1} = \frac{\sum_{j\neq 1}^{4}\sum_{k\neq 1}^{4} w_{1j}w_{jk}w_{k1}}{\left(\sum_{j\neq 1}^{4} w_{1j}\right)^2 - \sum_{j\neq 1}^{4} w_{1j}^2} = \frac{\sum_{j\neq 1}^{4} w_{1j}\left(w_{j2}w_{21} + w_{j3}w_{31} + w_{j4}w_{41}\right)}{\left(w_{12} + w_{13} + w_{14}\right)^2 - w_{12}^2 - w_{13}^2 - w_{14}^2} =$$

$$= \frac{w_{12}w_{23}w_{31} + w_{12}w_{24}w_{41} + w_{13}w_{32}w_{21} + w_{13}w_{34}w_{41} + w_{14}w_{42}w_{21} + w_{14}w_{43}w_{31}}{w_{12}w_{13} + w_{13}w_{14} + w_{12}w_{14}} =$$

$$= \frac{w_{12}w_{23}w_{31} + w_{13}w_{34}w_{41} + w_{12}w_{24}w_{41}}{w_{12}w_{13} + w_{13}w_{14} + w_{12}w_{14}}$$

$C^{Z}_{w,1} = 1$ when $w_{23} = w_{34} = w_{24} = 1$

Micro analysis

- Now I want to take a look more specifically at individual's contribution to the game
- This is called the **Network centroid** which can define the centrally located node.
- Since the **weighted adjacent matrix** (A_r =[r_{ij}]) will tell us the most connected node, we can easily formulate the ***centroid coefficient*** which express players' connectivity strength to all other teammates
  - i.e a player with $k_{max} = max\ k_i$
- This *centroid coefficient* could be interpreted as the cooperation level of the ith player with the centroid player
- $CC_{i,\ centeroid} = r_{i,centroid}$ if $i \neq j$ and 1 if i =j

Implementation

- Matlab functions *wgPlot* and *grPatition*

Result (Bayern 16/17)

- Scaled Connectivity

	1st	2nd	3rd	Overall
1 GK	0.307692	0.306	0.385	0.3328973
2 RD	0.423077	0.791	0.852	0.6886923
3 CD	0.461539	0.851	0.800	0.7041797
4 CD	0.461539	0.888	1	0.7831797
5 LD	0.557692	1	0.381	0.6462307
6 M (Tolliso)	1	0.970	0.649	0.873***
7 LM	0.75	0.784	0.528	0.6873333
8 RM	0.769231	0.561	0.718	0.6827437
9 RF	0.807692	0.285	0.712	0.601564
10 LF (Ribery)	0.884616	0.781	0.823	0.8295387***
11 S	0.692308	0.12	0.55	0.4541027

Result

- Clustering Coefficient

	1st	2nd	3rd	Overall
1 GK	0.325	0.544	0.447	0.438667
2 RD	0.509	0.532	0.434	0.491667
3 CD	0.478	0.506	0.455	0.479667
4 CD	0.471	0.510	0.441	0.474
5 LD	0.541	0.478	0.430	0.483
6 M (Tolisso)	0.524	0.529	0.624**	0.559
7 LM	0.456	0.601	0.452	0.503
8 RM	0.598	0.502	0.412	0.504
9 RF	0.535	0.571	0.397	0.501
10 LF	0.477	0.533	0.597	0.535667
11 S ()	0.605**	0.640**	0.540	0.595**

Result

- Clustering Coefficient

	1st	2nd	3rd	Overall
1 GK	0.256	0.200	0.340	0.265333
2 RD	0.846	0.933	0.115	0.631333
3 CD	0.769	0.196	0.235	0.4
4 CD	0.333	0.591	0.867	0.597
5 LD	0.691	0.422	1 ***	0.704333
6 M (Tolisso)	1 ***	1 ***	0.741	0.913667***
7 LM	0.539	0.923	0.478	0.646667
8 RM	0.615	0.488	0.435	0.512667
9 RF	0.912	0.371	0.634	0.639
10 LF	0.741	0.821	0.502	0.688
11 S ()	0.606	0.432	0.341	0.459667

Conclusion

- He transferred in the same year as Neymar’s (Summer 2017)
- Tolliso is undervalued (his transfer fee was only $47m, while Neymar’s was $600m)
- Even FIFA is biased against his ability

Reference Papers

- “Exploring Team Passing Networks and Player Movement Dynamics in Youth Association Football”
- “Statistical Analysis of Weighted Networks”
- “A network-based approach to evaluate the performance of football teams”