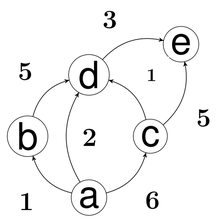
Network theory and analysis of football strategies

Javier López Peña
Department of Mathematics. University College London

Summary

In this work we showcase the use of some tools from network theory in order to describe the strategy of football teams.
Each team gets associated a directed network where nodes correspond to players and arrows to passes providing a direct visual inspection of the team strategy.
The network allows to visualize a team's strategy, identify play pattern, determine hot-spots on the play and localize potential weaknesses.
Different centrality measures are used to determine the relative importance of each player in the game, the "popularity" of a player, and how affected would the team be if each player was removed from the game.

Networks



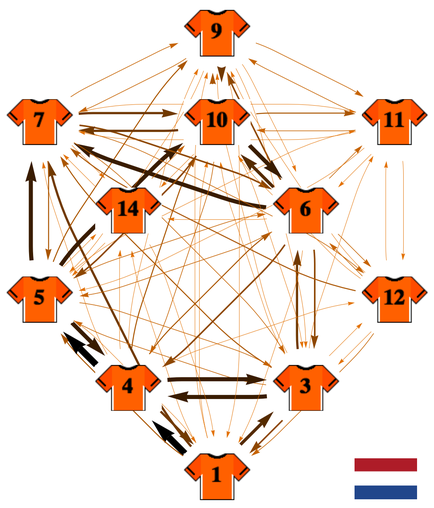
A ***network*** consists of:
- A collection of ***nodes***
- Some ***arrows*** connecting the nodes
- Some ***weights*** associated to the arrows

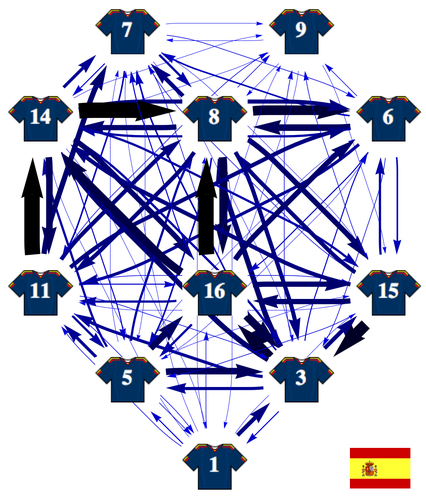
The network of a football team

- **Nodes** represent the **team players**
- **Arrows** represent **passes**
- **Weights** are the **number of passes**
- **Adjacency matrix of the network**:

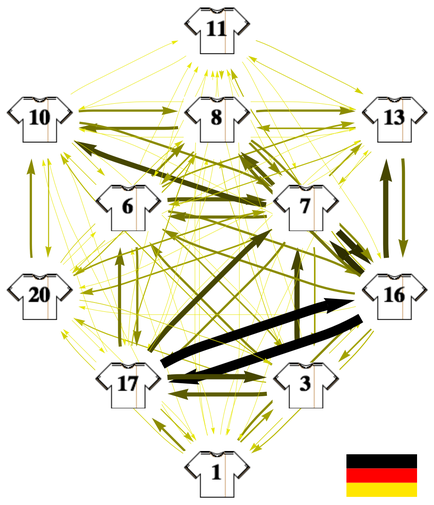
$$A_{ij} = \text{number of passes from } i \text{ to } j$$

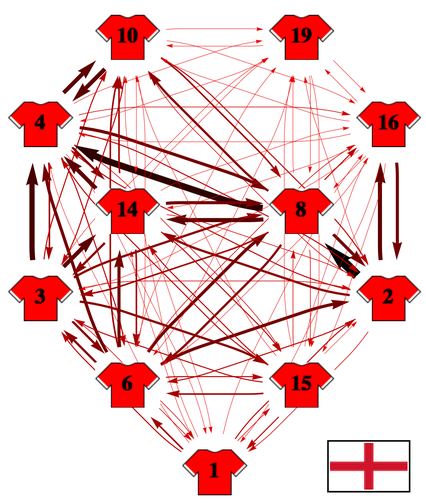
Netherlands vs Spain





Germany vs England





The fundamental theorem of football

"*Good football teams have a recognizable style*"

Our network helps to visualize that style!

Who is the most relevant player?

- ***Centrality measures*** give us relative notions of "importance" of a node in a network
- ***Closeness centrality:*** Mean distance from a node to the other ones

$$C_i = \frac{20}{\sum_{j\neq i} \frac{1}{A_{ij}+1} + \sum_{j\neq i} \frac{1}{A_{ji}+1}} - 1$$

- ***Pagerank centrality:*** Recursive notion of "popularity". A player is popular if receives passes from other popular players

$$x_i = p \sum_j A_{ji} \frac{x_j}{k_j^{\text{out}}} + (1-p)$$

  - $k_j^{\text{out}} = \sum_i A_{ji} =$ total number of passes made by j
  - $p =$ probability of passing the ball ($p \simeq 0.85$ works well)
- ***Betweenness centrality:*** How much the team suffers when a player is removed

$$C_B(i) = \frac{1}{10^2} \sum_{j,k\neq i} \frac{d_{jk}(i)}{d_{jk}}$$

$d_{jk} =$ distance from j to k, $d_{jk}(i) =$ distance after removing i

Centralities for Dutch players

	Player	Closeness	Pagerank	Betweenness
1	Stekelenburg	0.842	8.22%	2.042
3	Heitinga	1.296	9.27%	2.669
4	Mathijsen	1.046	6.34%	1.233
5	Van Bronkhorst	1.578	11.12%	4.159
6	Van Bommel	1.749	9.55%	3.585
7	Kuyt	1.655	11.67%	4.835
9	Van Persie	0.961	10.13%	1.402
10	Sneijder	1.724	11.67%	4.219
11	Robben	0.589	8.55%	0.792
12	Boulahrouz	0.529	6.10%	0.726
14	De Zeeuw	0.348	7.30%	0.333

Centralities for Spanish players

	Player	Closeness	Pagerank	Betweenness
1	Casillas	0.672	5.47%	0
3	Piqué	3.347	8.96%	1.19
5	Puyol	1.849	8.89%	0.92
6	Iniesta	1.889	8.35%	0.12
7	Villa	1.798	10.17%	1.19
8	Xavi	4.358	10.26%	2.49
9	Torres	0.578	8.30%	0
11	Capdevilla	2.975	8.96%	1.19
14	Alonso	3.742	10.26%	2.49
15	Ramos	2.251	10.17%	1.19
16	Busquets	3.239	10.17%	1.19

What do network tell us?

- Different teams have very different networks
- Quick overview of a team style
  - Most used areas of the court
  - Short distance or long distance passes
  - Players not participating enough
  - Problems between players
- Centrality measures give information about players
- Plenty of useful information for a coach!

Future work

- Keep track of unsuccessful passes
- Add an extra node for the gate to keep track of shoots
- Keep data consistent when a player gets changed