Network theory and analysis of football strategies

Javier López Peña

Department of Mathematics
University College London

Physics of Sports, Paris 2012

Disclaimer

- Joint work with Hugo Touchette
- ((Very) Pure) Mathematician speaking
- For any Americans in the audience:

$$\text{Football} = \text{Soccer}$$

What can maths say about football?

Mathematicians are good at two things:

- Finding patterns
- Turning easy things intro abstract nonsense

(Normally we do it the other way around)

Question

Can the abstract nonsense tell us something useful?

The Fundamental Theorem of football

UCL

Theorem (Fundamental Theorem of football)

Good football teams have a recognizable style

- But not necessarily *the same* for all teams!

Question

- *Can we describe the "style" mathematically?*
- *And then say something about the team?*

What to focus on?

Many aspects of football one might look at!

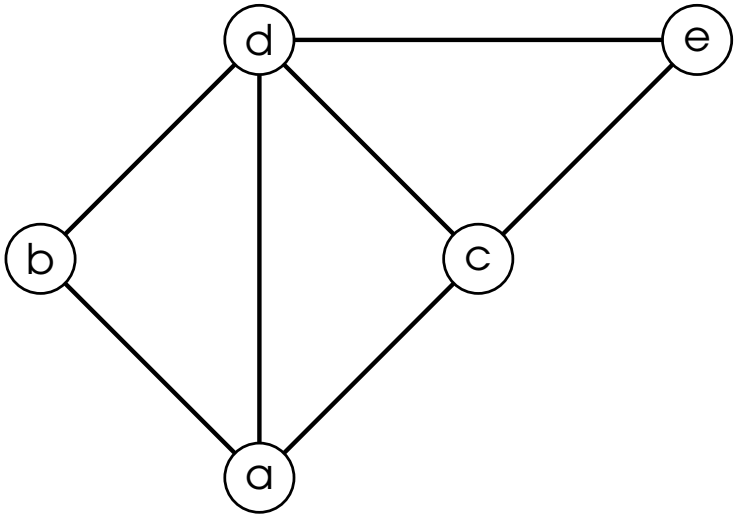
- Goals
- Fouls
- Percentage of victories
- Ball possession
- Passing information

We'll focus on the last one

A bit of abstract nonsense: Networks

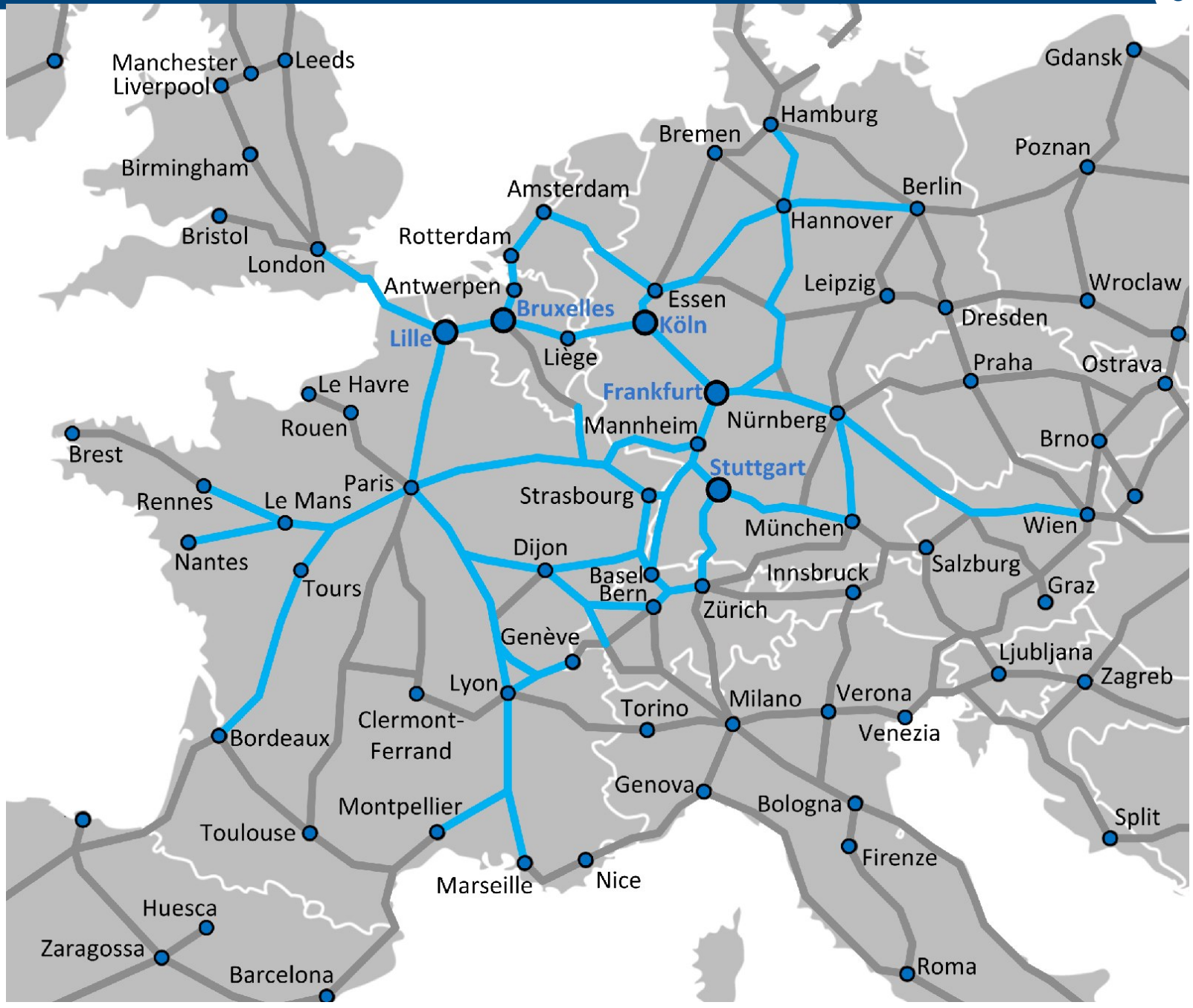
A network consists of:

- A collection of nodes (or vertices)
- Some edges connecting the nodes



- Nodes can have a clear physical meaning.
- But they don't have to.

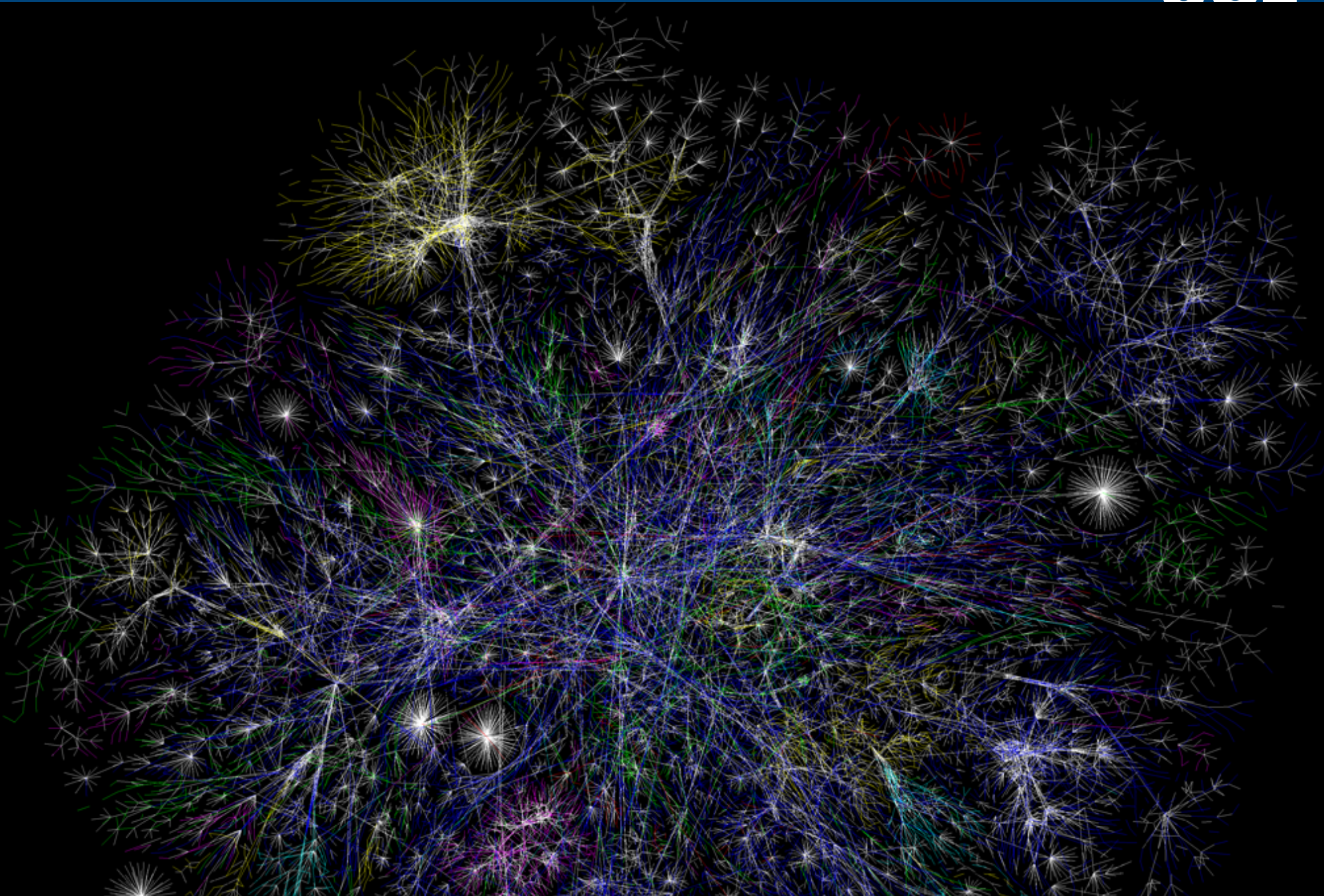
Example: High speed train network



Example: North America power grid

Example: The Internet

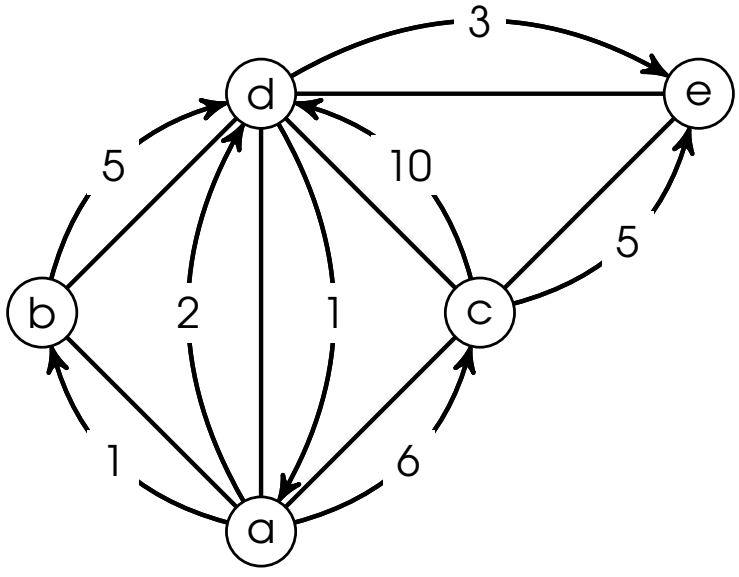
UCL



Oriented networks

Not all edges are created equal!

- We can use directed edges (or arrows)
- Perhaps pointing in both directions
- Or attach weights to them



The passing network of a football team

We associate a network to each football team

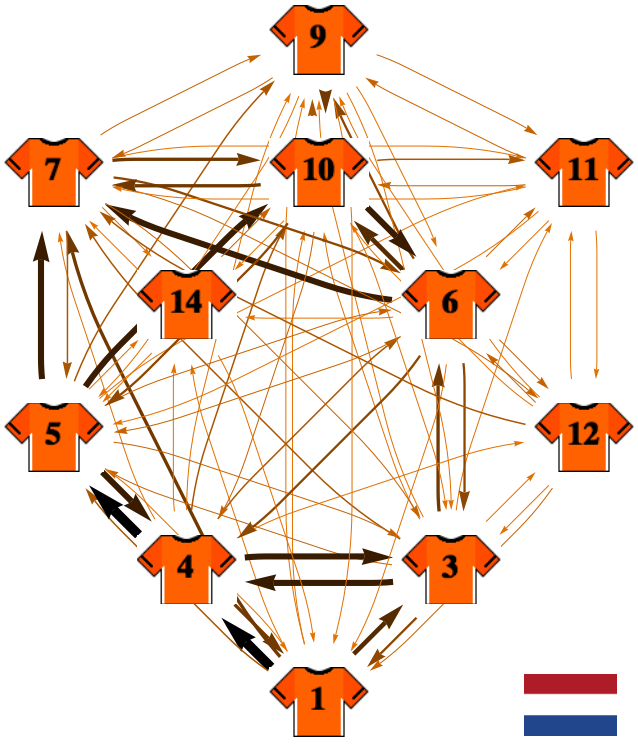
- Nodes are the team players
- Arrows represent passes between the players
- Weights given by the number of passes

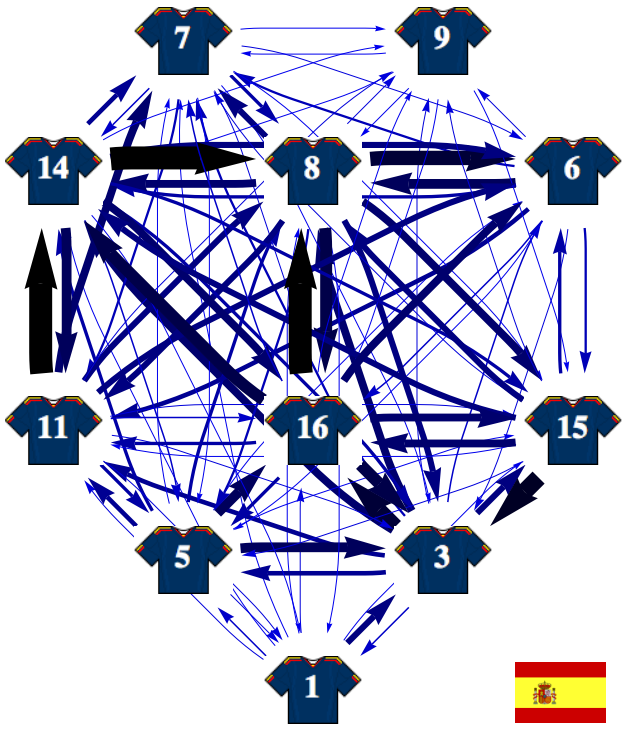
In the drawing, represent the weight as arrow thickness

The passing network of a football team

UCL

Netherlands vs. Spain

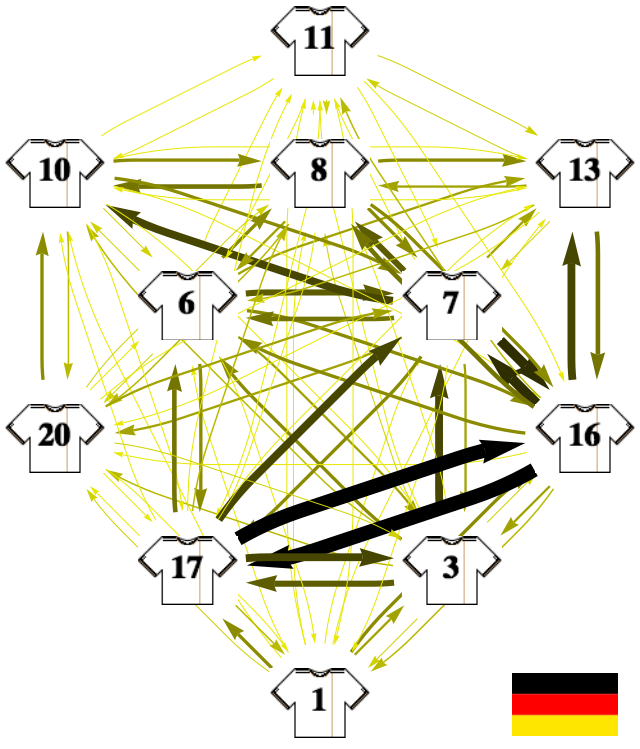


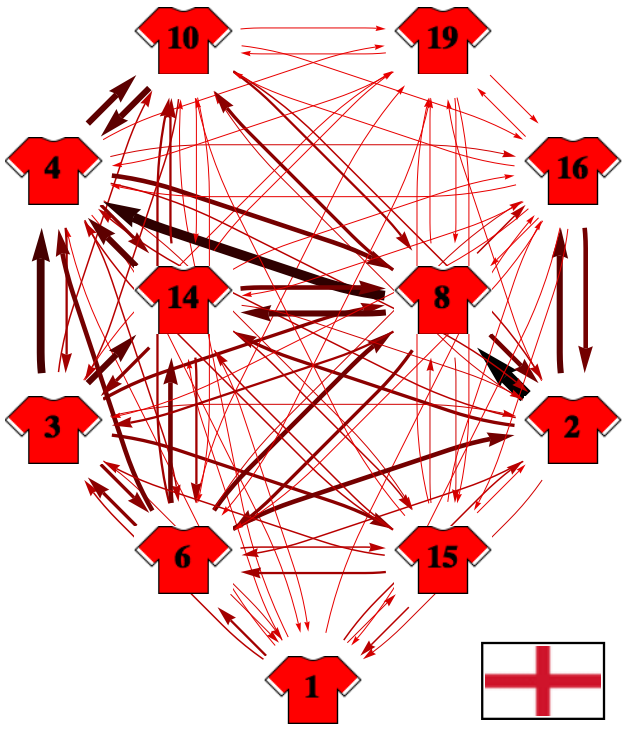


The passing network of a football team

UCL

Germany vs. England





Extracting information from the network

Mathematical representation of the network

- Use the adjacency matrix (A_{ij})

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

- Matrix is bad for visualization
- But good for computations

How an adjacency matrix looks like

UCL

$$
\text{Spain} \rightarrow \begin{pmatrix}
0 & 7 & 3 & 0 & 1 & 1 & 0 & 2 & 1 & 1 & 2 \\
3 & 0 & 6 & 3 & 4 & 8 & 2 & 6 & 8 & 7 & 11 \\
2 & 9 & 0 & 1 & 1 & 5 & 0 & 6 & 5 & 1 & 9 \\
0 & 2 & 1 & 0 & 4 & 9 & 2 & 6 & 5 & 4 & 3 \\
0 & 1 & 1 & 2 & 0 & 5 & 1 & 4 & 3 & 1 & 2 \\
1 & 7 & 4 & 11 & 7 & 0 & 2 & 7 & 8 & 8 & 10 \\
0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & 1 & 1 & 1 \\
2 & 2 & 5 & 7 & 8 & 7 & 2 & 0 & 14 & 2 & 3 \\
1 & 5 & 3 & 7 & 6 & 14 & 1 & 9 & 0 & 5 & 8 \\
0 & 12 & 2 & 5 & 2 & 7 & 2 & 1 & 7 & 0 & 8 \\
0 & 10 & 5 & 8 & 3 & 14 & 1 & 5 & 10 & 8 & 0
\end{pmatrix}
$$

About the players: centrality

UCL

Question

How to measure the importance of a node in a network?

Answer: Centrality measures

- There are different ways of measuring importance
- Different types of centrality to address them!

Closeness centrality

- Mean distance from a node to the other ones
- Distance is the inverse of the number of passes

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

- w and $1 - w$ are weights to passing/receiving
- There is some normalization going on
- Actual value is not important
- Just focus on the relative order

Pagerank centrality

- Recursive notion of "popularity"
- A node is popular if linked by other popular nodes

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

- $k_j^{\text{out}} = \sum_i A_{ji} =$ total number of passes made by j
- p is the (estimated) probability of passing the ball
- Estimate made by heuristics
- $p = 0.85$ normally works well

Betweenness centrality

- How the network suffers when a node is removed
- A node is popular if linked by other popular nodes

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

- d_{jk} = distance from j to k
- $d_{jk}(i)$ = distance without going through i
- Nodes with high C_B are dangerous for the network

Centralities for Spanish players

UCL

	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!

The limits of the tool

Network analysis is not a silver bullet

- Not for all sports
- Only tracks *successful* passes
  - Add a probability to the weight!
- Doesn't account for shots and goals
  - Add an extra node for the opponent's gate!
- What happens when a player gets changed?
- Passing data is hard to obtain!

Thanks for your attention!