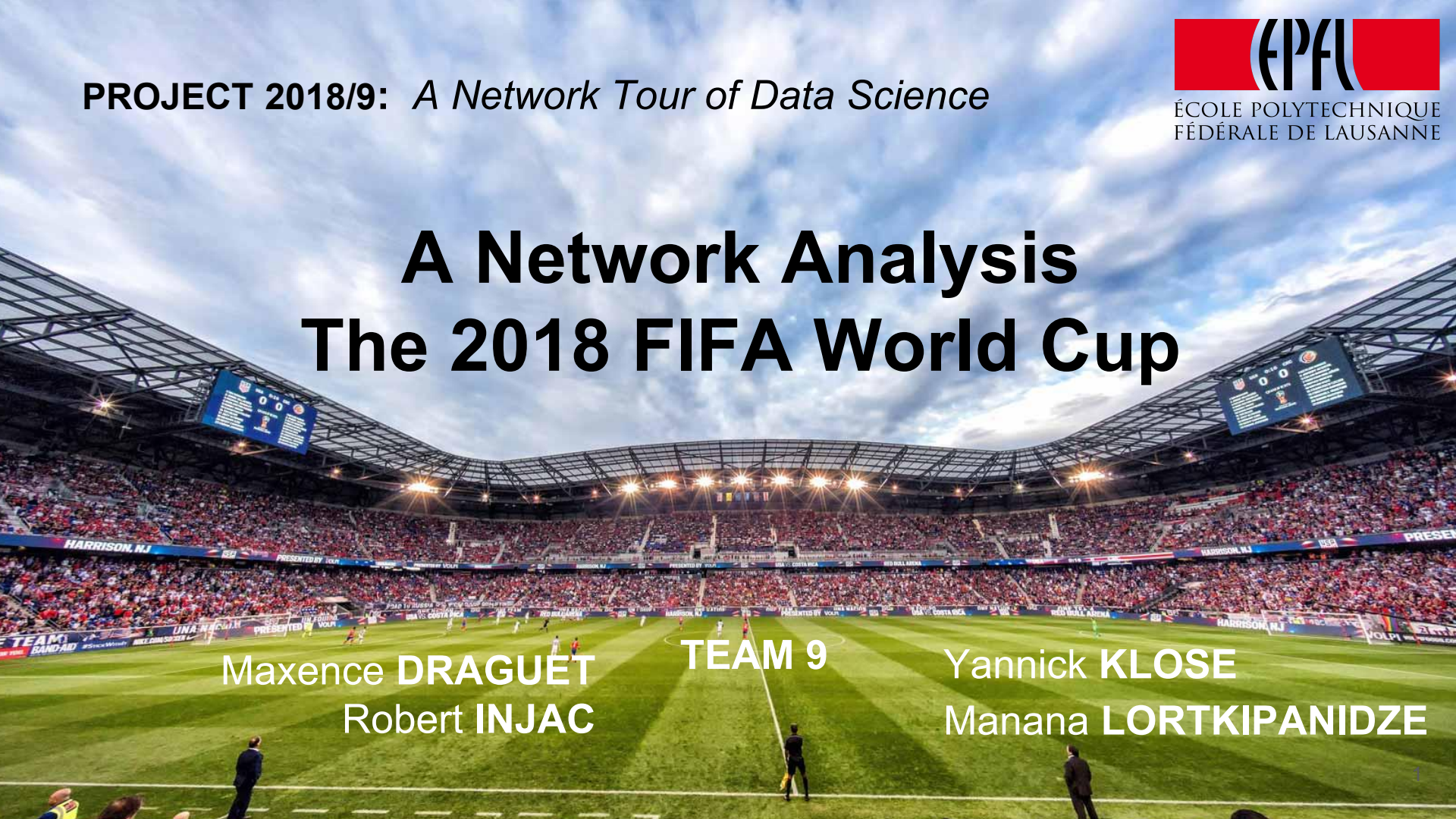
PROJECT 2018/9: *A Network Tour of Data Science*

ÉCOLE POLYTECHNIQUE
FÉDÉRALE DE LAUSANNE

A Network Analysis
The 2018 FIFA World Cup



Maxence **DRAGUET**
Robert **INJAC**

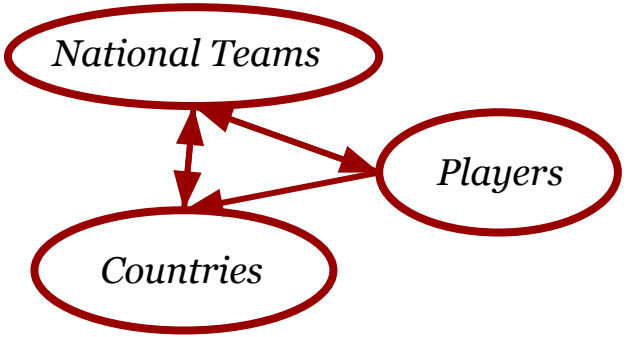
TEAM 9

Yannick **KLOSE**
Manana **LORTKIPANIDZE**

Introduction & Data

One of the most hyped events, lasting a month with activities occurring at discretised dates.

Natural structure echoed on Wikipedia



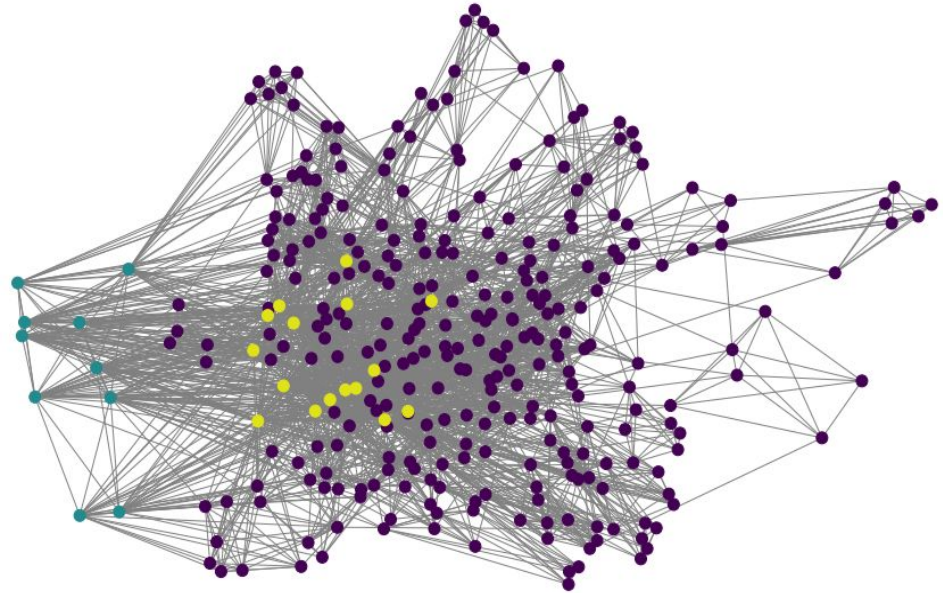
➢ Hierarchy between node types ⟹ Hyperlinks

➢ Fluctuating popularity correlating to events ⟹ Number of Visits

The HyperLink Network

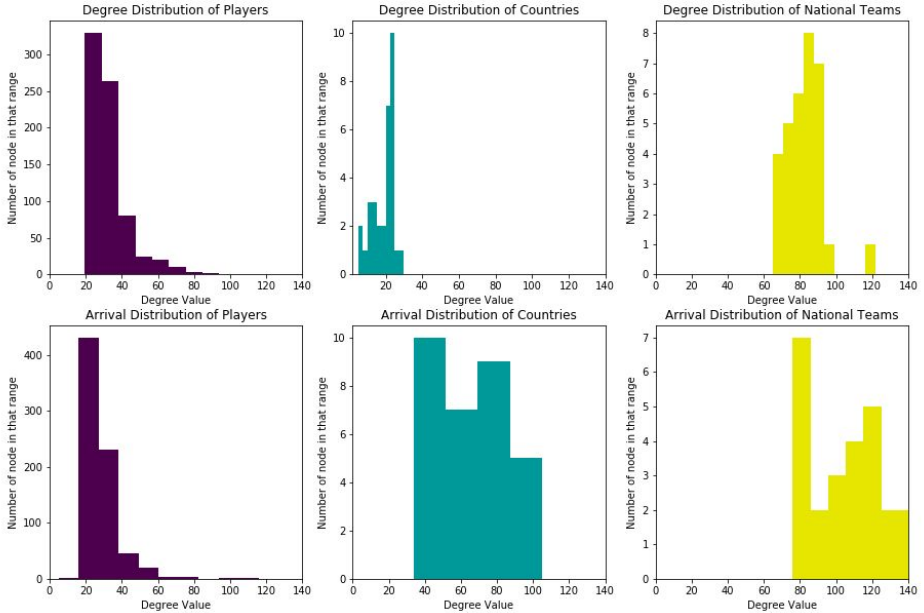
NODES: *Players*, *Countries* and *National Teams* **EDGES:** Hyperlinks

A **Binarised Directed Graph** …



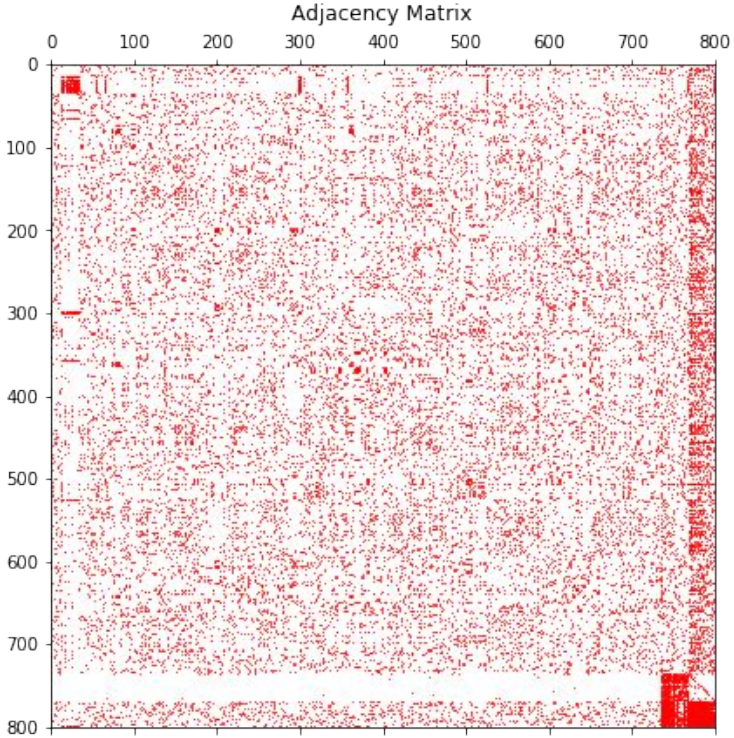
This is a subsample !

… with a **Structure**



The HyperLink Network

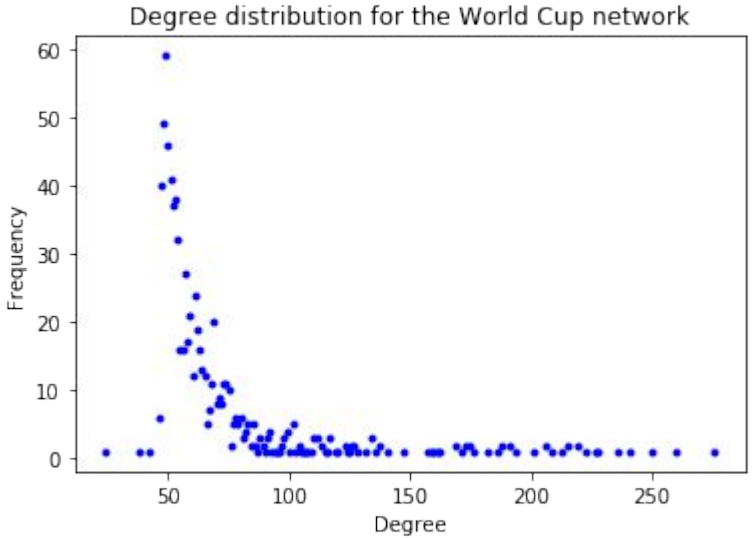
Sparse, Scale Free and Small World



Sparsity: 0.0432

Diameter: 5

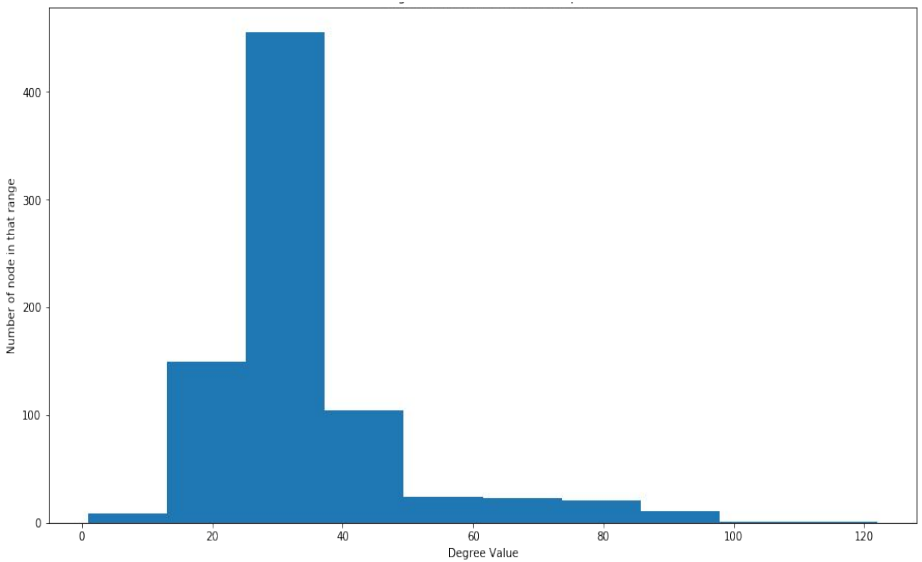
Transitivity: 0.508



The Importance of Players

Measure - Degree Distribution

Degree Distribution



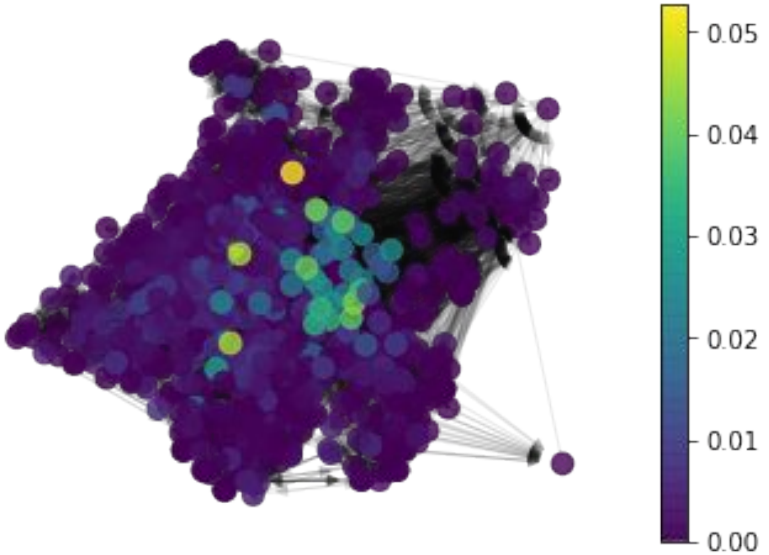
Top Five

Rank	Player	Degree
1	Lionel Messi	157
2	Eden Hazard	149
3	Cristiano Ronaldo	145
4	Luis Suárez	140
5	Neymar	127

The Importance of Players

Measure - Betweenness Centrality

Betweenness Centrality

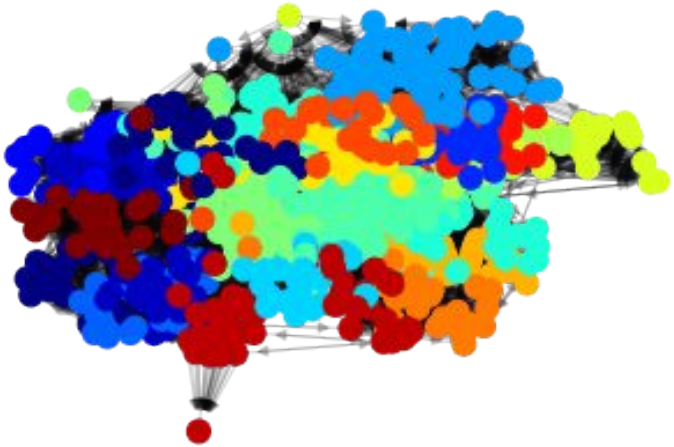


Top Five

Rank	Player	Betweenness
1	Keylor Navas	0.053
2	Luis Suárez	0.052
3	Lionel Messi	0.046
4	Mohamed Salah	0.046
5	Son Heung-min	0.034

Finding Teams - The Louvain Method

Community/Cluster detection



Teams:	32
Number of Clusters:	18
Teams as a Cluster:	11
Team Splits:	0

***Teams* in Isolated Clusters**

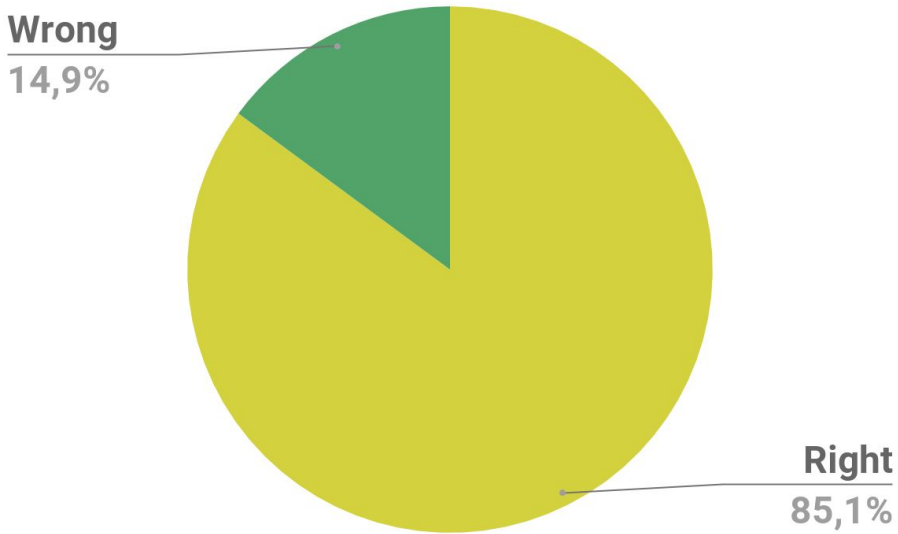
Nigeria, Russia, Iran, Serbia, Colombia, Peru, Iceland, Sweden, Denmark, Morocco, Senegal

Finding Teams - Heat Transfer

Quality of label prediction - Croatia

Correct | Wrong | Unrelated to label

Quality of label prediction - All teams



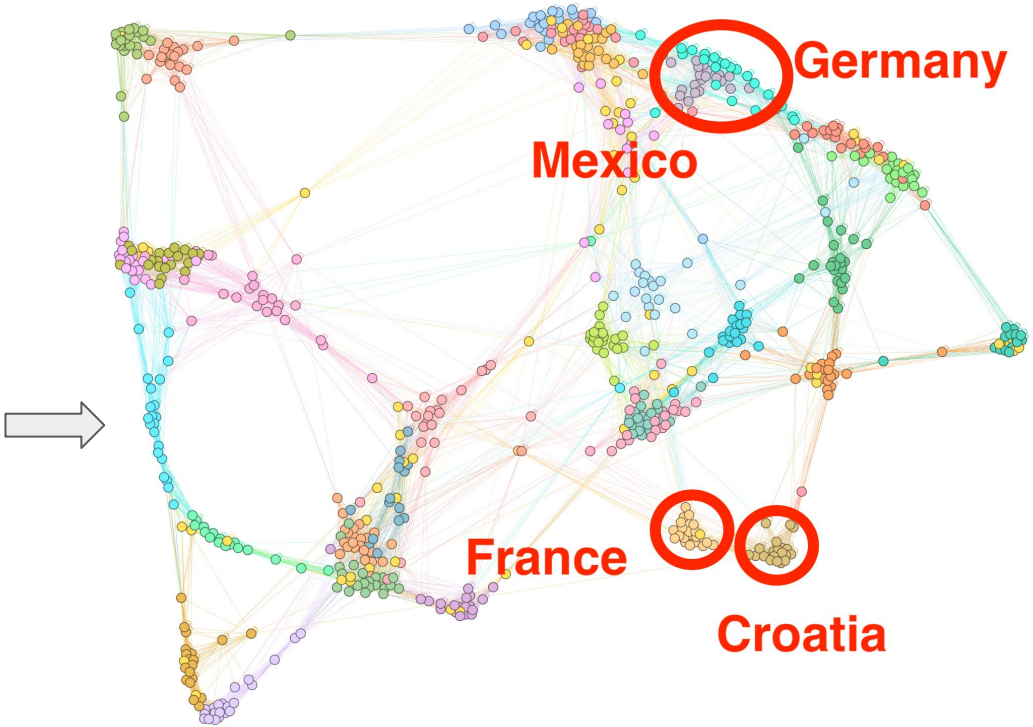
Finding Teams - The Correlation Network

NODES: *Players*

WEIGHTS: Pearson Correlations of Feature Vectors

A **Weighted Undirected Graph**

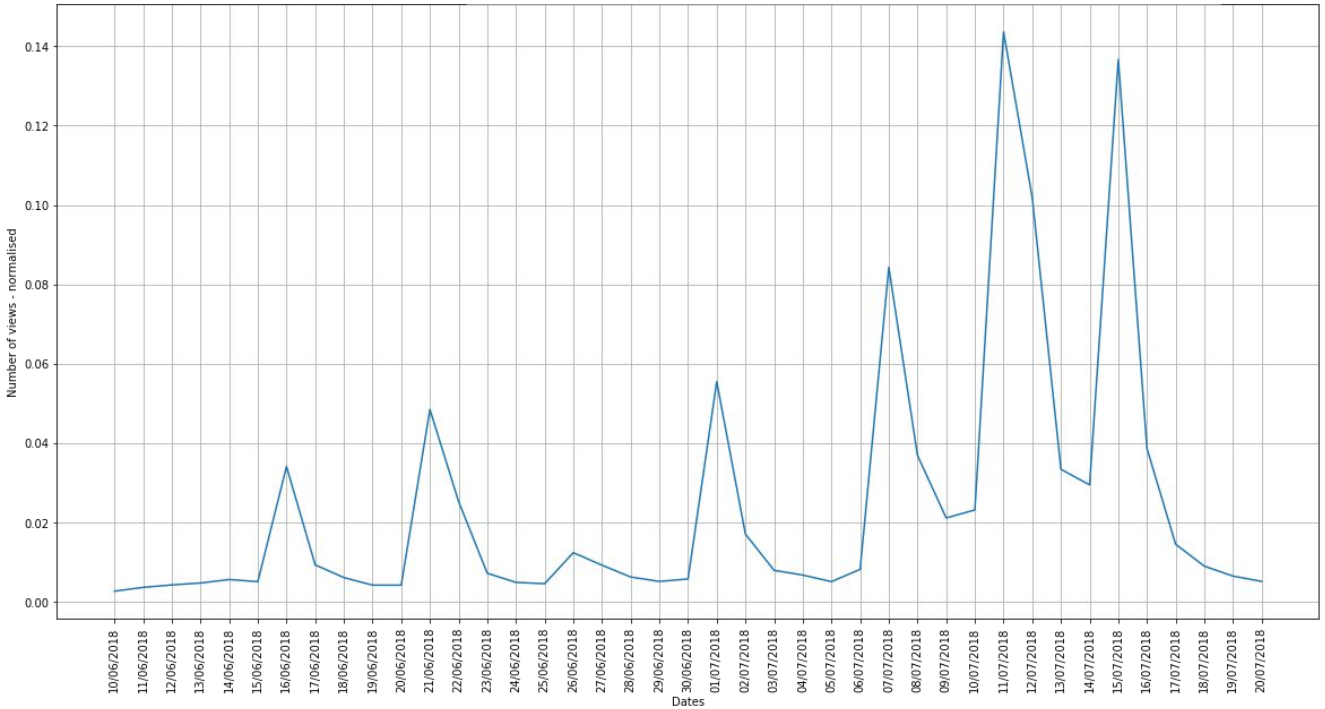
Binarising the 21* most intense correlations per *player*



* *or less*

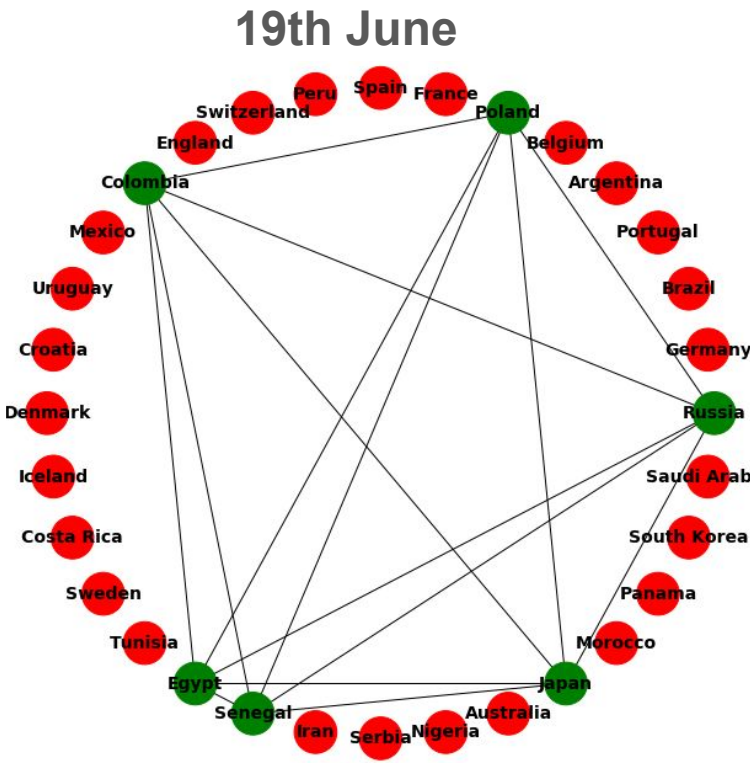
Finding Matches - Signal

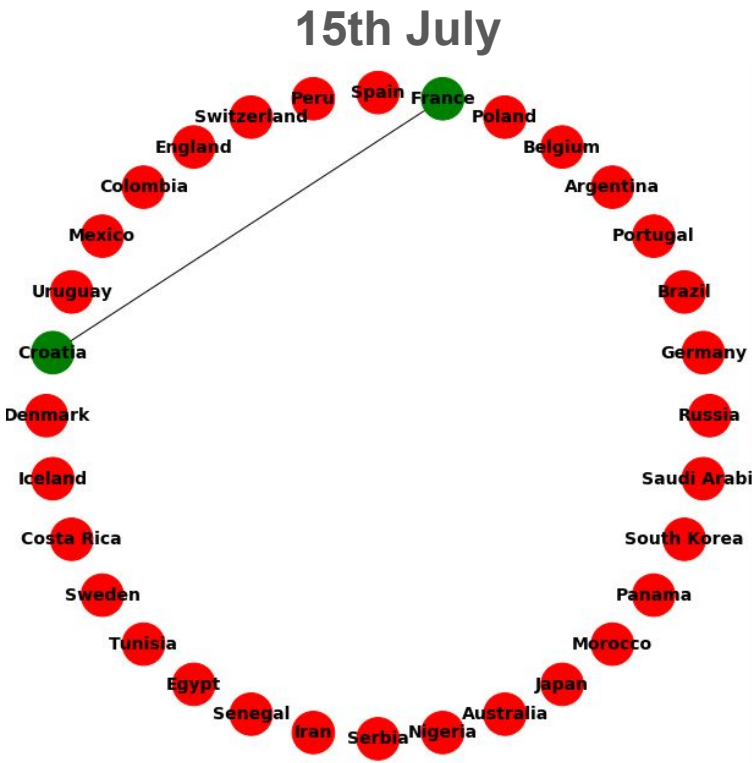
Normalised number of views per day - Croatian team



Finding Matches - Dynamic Visualisation Network

NODES: *National Teams* **EDGES:** both numbers of views peak on selected day



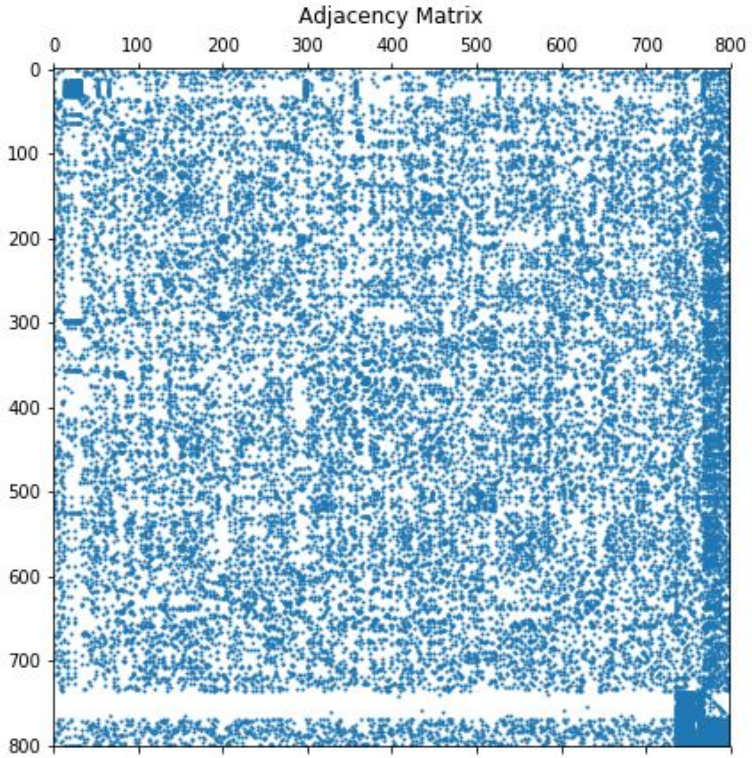


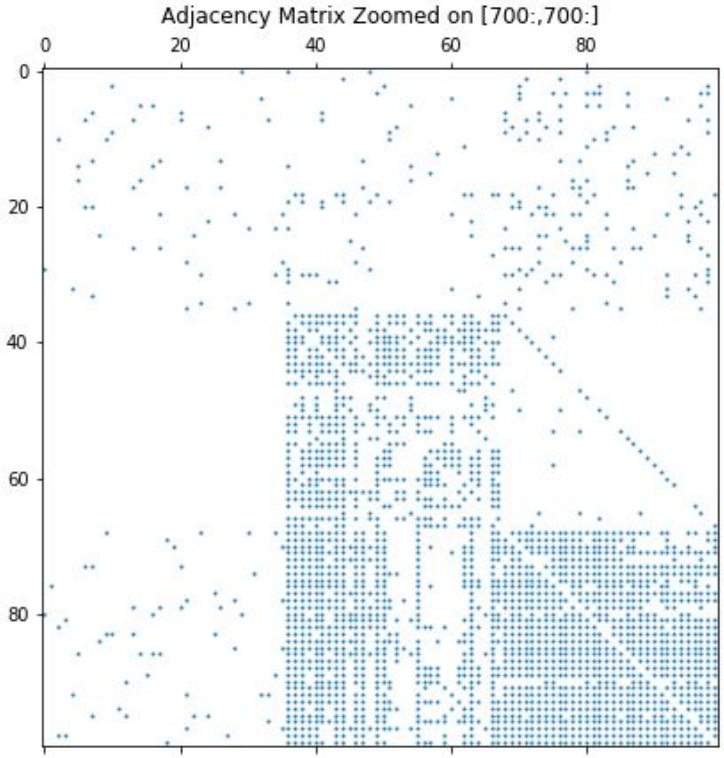
The End

Thank you for your attention!

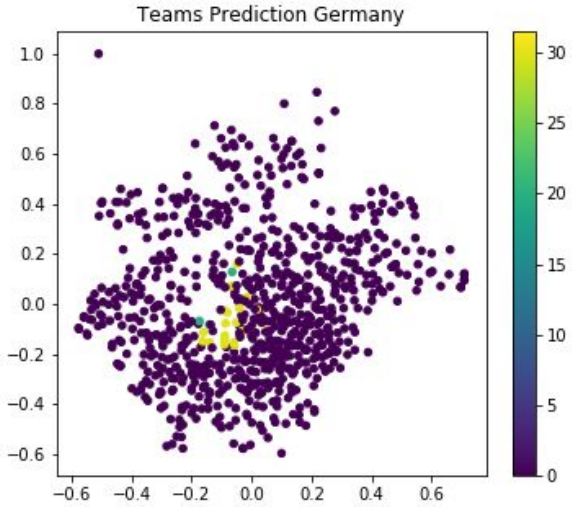
Additional Information

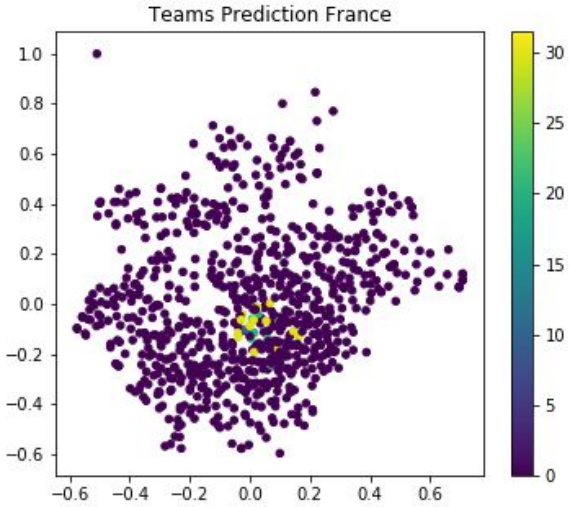
Other Plots - HyperLink Adjacency

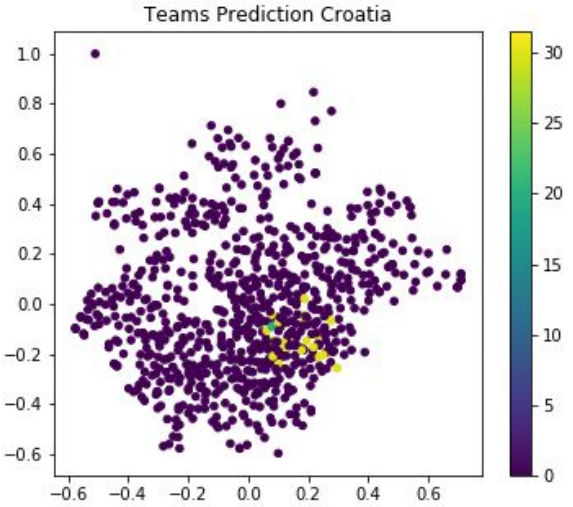




Other Plots - Heat Transfer

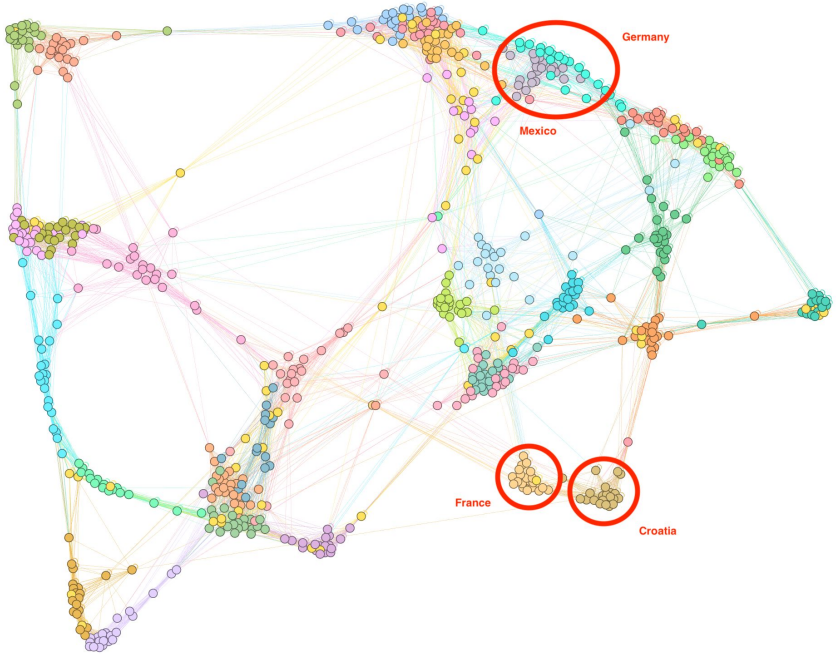




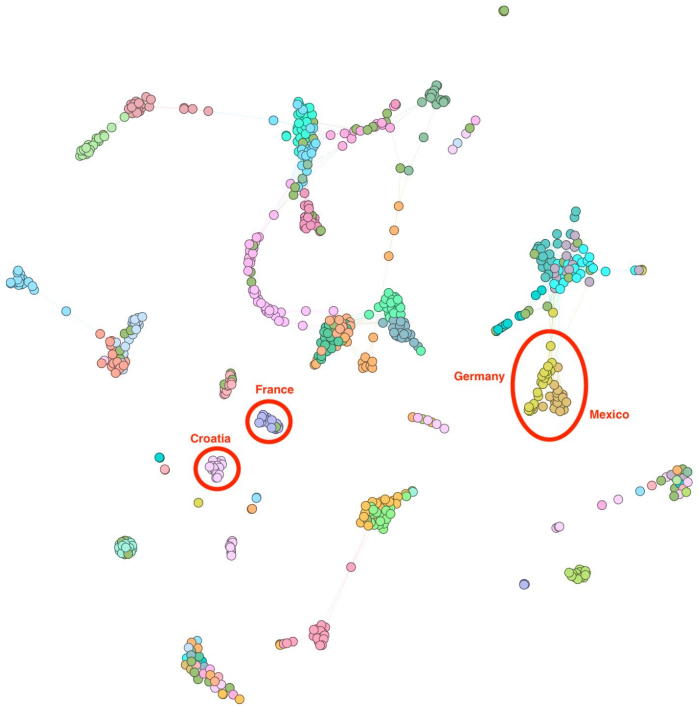


Other Plots - Correlation Network

Selecting 21 edges

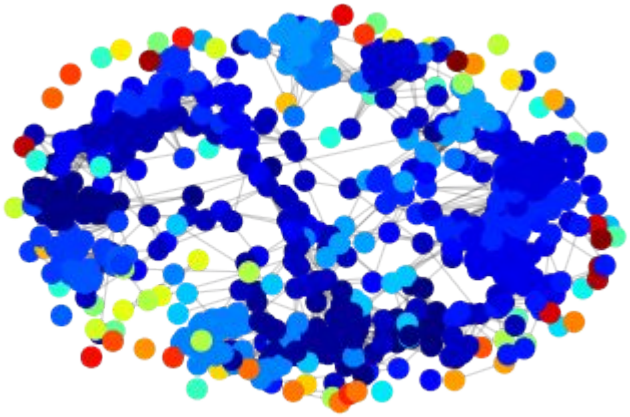


Thresholding the Matrix



Finding Teams - The Louvain Method

Community/Cluster detection using correlation matrix



Teams:	32
Number of Clusters:	29
Team Splits:	265

Even though number of clusters after convergence is closer to the number of teams, clustering is not very precise ⟹ members of the same team belong to the different clusters