



# Social network analysis in football: passing network analysis of Manchester City in the season 2017/2018

Luca D'Ambrosio - 0000890761

Progetto di “Social Network Analysis”

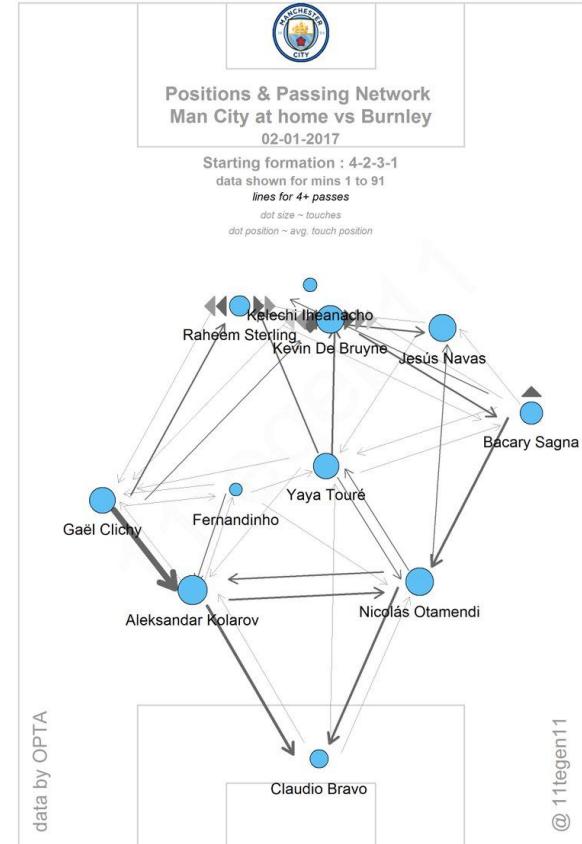
A.A. 2019/2020

# Introduzione

L'obiettivo di questo progetto è stato quello di utilizzare alcuni strumenti di network theory per analizzare la stagione calcistica 2017/2018 della squadra Manchester City.

L'analisi è stata effettuata su 3 livelli differenti:

-  **Player analysis**
-  **Team analysis**
-  **Match analysis**





# Networks



Per raggiungere l'obiettivo prefissato bisogna focalizzarsi su uno degli elementi base del calcio.

Una match è caratterizzato da circa 5000 eventi, tra cui:

- Tiri
- Falli
- Sostituzioni
- **Passaggi**

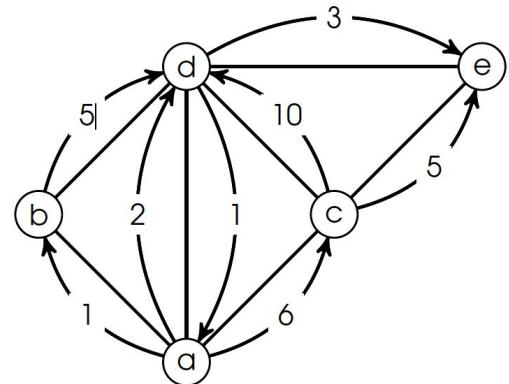
In questo studio si è focalizzata l'attenzione su quest'ultimo aspetto.

# La passing network di una squadra di calcio



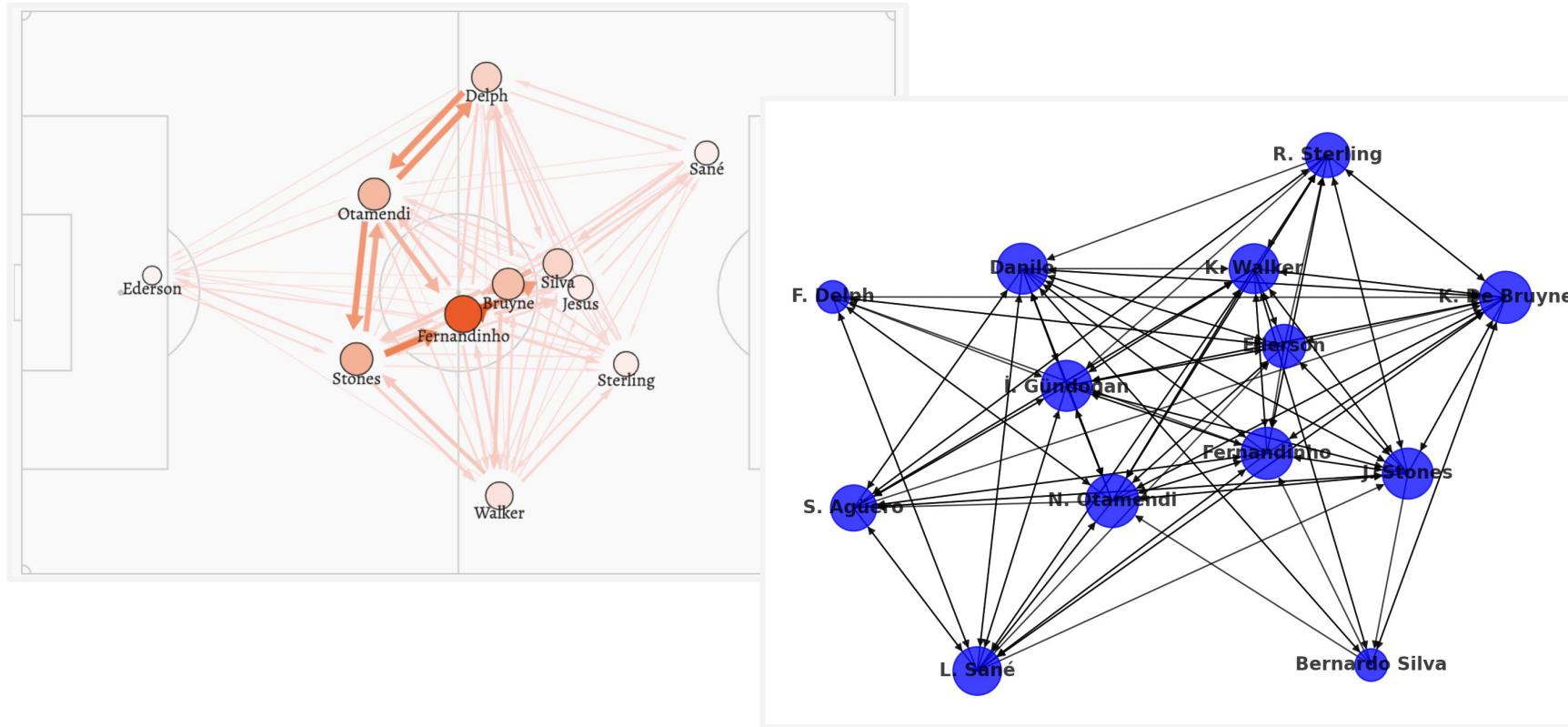
Definiamo la **passing network** di un team come la rete in cui:

- I **nodi** corrispondono ai giocatori della squadra
- Gli **archi** rappresentano i passaggi tra i giocatori
- I **pesi** che caratterizzano gli archi sono rappresentati dal numero di passaggi che intercorrono tra due nodi





# La passing network di una squadra di calcio





# Dataset

I dataset utilizzati in questo progetto sono stati raccolti e vengono forniti da **Wyscout**, una società italiana che supporta lo scouting calcistico, l'analisi delle partite e le dinamiche di trasferimento.

**wyscout**



# Dataset

I dati si descrivono gli **eventi** accaduti nei **match** della stagione 2017/2018 di cinque competizioni calcistiche nazionali in Europa. Sono disponibili anche i dati della Coppa del mondo 2018 e dell'Europeo 2016.

A completare i set di dati precedenti (**Events**, **Match**), sono presenti altri set di dati:

- Competitions
- Players
- Teams
- Referees
- Coaches



BUNDESLIGA



Premier  
League



LIGUE 1



LaLiga  
Santander



UEFA  
EURO2016  
FRANCE



FIFA WORLD CUP  
RUSSIA 2018



# Dataset

Ogni set di dati fornito è codificato in formato **JSON (JavaScript Object Notation)**.

```
{"birthArea":  
{"alpha2code": "AR",  
"alpha3code": "ARG",  
"id": "32",  
"name": "Argentina"},  
"birthDate": "1987-06-24",  
"currentNationalTeamId": 12274,  
"currentTeamId": 676,  
"firstName": "Lionel Andres",  
"foot": "left",  
"height": 170,  
"lastName": "Messi Cuccittini",  
"middleName": "",  
"passportArea": {"alpha2code": "ES",  
"alpha3code": "ESP",  
"id": "724",  
"name": "Spain"},  
"role": {"code2": "FW", "code3": "FWD", "name": "Forward"},  
"shortName2": "L. Messi",  
"weight": 72,  
"wyId": 3359}
```

```
{"city": "Torino",  
"name": "Juventus",  
"area": {"alpha2code": "IT", "name": "Italy", "alpha3code": "ITA", "id": "380"},  
"wyId": 3159,  
"officialName": "Juventus FC",  
"type": "club"}
```

Tali documenti descrivono il giocatore Lionel Messi e il team Juventus. Essi fanno parte rispettivamente dei dataset “Players” e “Teams”.



# Dataset - Events

```
{"eventId": 8,
"eventName": "Pass",
"eventSec": 2.4175,
"id": 253668302,
"matchId": 2576335,
"matchPeriod": "1H",
"playerId": 3344,
"positions":
[{"x": 49, "y": 50}, {"x": 38, "y": 58}],
"subEventId": 85,
"subEventName": "Simple pass",
"tags":
[{"id": 1801}],
"teamId": 3161}
```

---

Informazioni su un **evento** di tipo **Pass** contenuto nel set di dati **Events**.



# Strumenti utilizzati



## NetworkX - Python

Pacchetto disponibile per il linguaggio di programmazione Python utile per l'analisi delle reti.

Attraverso questo strumento sono state svolte le seguenti analisi:

- Player analysis
- Team analysis
- Match analysis



## Gephi

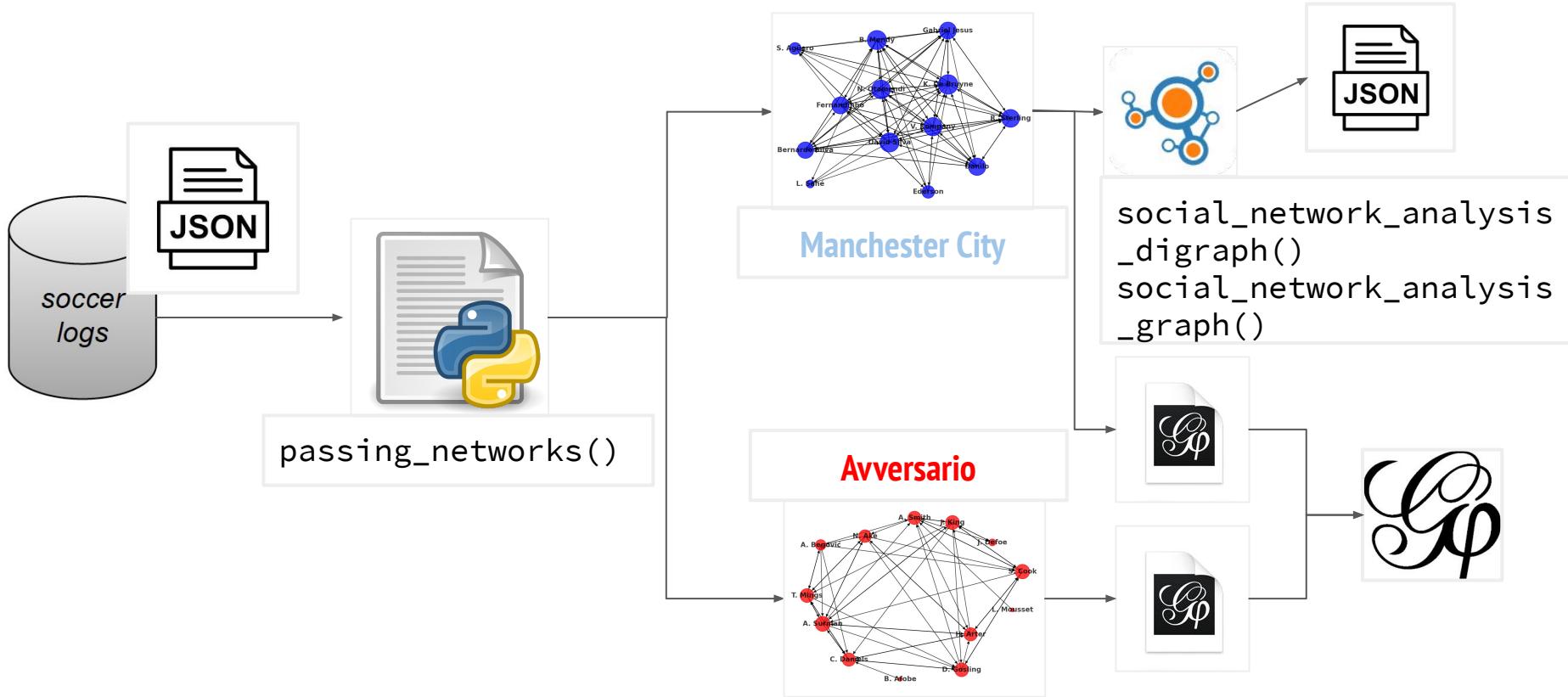
Software open source per la visualizzazione grafica e l'analisi delle reti sociali.

Tramite tale software è stata svolta una match analysis e la visualizzazione grafica della rete secondo il modulo di partenza del team.





# Social network analysis





La player analysis è stata effettuata per misurare il **contributo individuale** di un giocatore.

Tale risultato è stato raggiunto:

- esaminando la passing network locale di ogni nodo;
- attraverso il calcolo di misure di centralità.

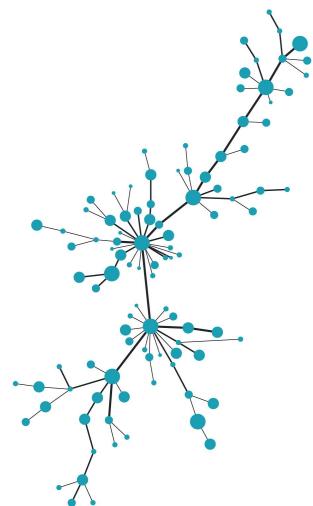
## 1. Player analysis





# 1. Player analysis

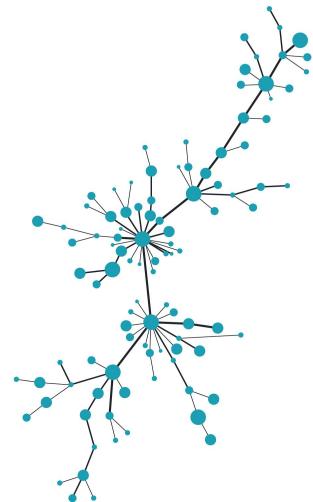
INDICE	DESCRIZIONE
Degree	Nel caso di un digrafo il suo valore è pari alla somma del valore di indegree e outdegree.
Indegree	Misura la popolarità di un attore, il suo valore è pari al numero di archi entranti nel nodo.
Outdegree	Misura l'espansività di un attore, il suo valore è pari al numero di archi uscenti dal nodo.
Degree centrality	Per un nodo $v$ è la frazione tra degree e il numero di nodi -1.
Indegree centrality	Per un nodo $v$ è la frazione tra indegree e il numero di nodi -1.
Outdegree centrality	Per un nodo $v$ è la frazione tra outdegree e il numero di nodi -1.





# 1. Player analysis

INDICE	DESCRIZIONE
<b>Closeness centrality</b>	Fornisce una misurazione diretta di quanto sia facile raggiungere un determinato giocatore all'interno di una squadra.
<b>Betweenness centrality</b>	Misura come il flusso di palla tra altri giocatori dipende dal calciatore analizzato.
<b>PageRank</b>	Nozione ricorsiva di "popolarità". L'idea di base è che "un giocatore è popolare se ottiene passaggi da altri giocatori popolari".
<b>Clustering coefficient</b>	Misura che indica quanto i vicini di un nodo tendano a formare una cricca.
<b>Triangles</b>	Numero di triangoli che un nodo può creare nella rete.





# 1. Player analysis

Le misure presentate precedentemente sono state sottoposte ad un ulteriore computazione, con lo scopo di descrivere dei valori che possano avere una valenza generale nella stagione.

In particolare, tale calcolo si compone dei seguenti passaggi:

1. Per ogni partita, una volta calcolati gli indici, da questi sono stati estratti valori minimi e massimi,
2. Tali valori sono stati inseriti in delle liste separate per indice e grandezza,
3. Per ogni elemento della lista (calciatore) è stata calcolato una valore che sia in grado di esprimere l'incidenza media del giocatore durante la stagione,
4. I risultati ottenuti sono stati ordinati e salvati in un file JSON.



# 1. Player analysis

## Player Analysis (Manchester City - Season 17/18) - NetworkX

Max Degree Centrality	0.708 (Fernandinho - C)
Min Degree Centrality	0.008 (S. Agüero - A)
Max Indegree Centrality	0.384 (Fernandinho- C)
Min Indegree Centrality	0.004 (Danilo - D)
Max Outdegree Centrality	0.376 (Fernandinho - C)
Min Outdegree Centrality	0.004 (P. Foden - C)





# 1. Player analysis

## Player Analysis (Manchester City - Season 17/18) - NetworkX

Max Clustering coefficient	0.251 (Ederson - P)
Min Clustering coefficient	0.013 (B. Diaz - A)
Max PageRank	0.028 (Otamendi- D)
Min PageRank	0.0003 (E. Mangala - D)





# 1. Player analysis

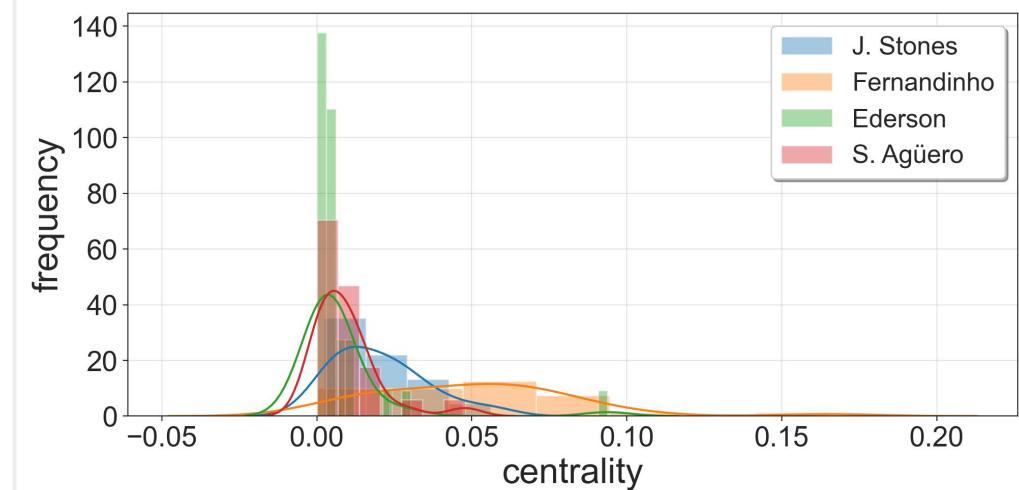
## Player Analysis (Manchester City - Season 17/18) - NetworkX

Max Betweenness centrality	0.018 (Fernandinho - C)
Min Betweenness centrality	-0.15 (Bernardo Silva - A)

Stones mean b: 0.02018875732111026  
Fernandinho meanb : 0.050858548759618274  
Ederson mean b: 0.007457055547333325  
Aguero mean b: 0.009277727827727828

```
"min_betweenness_centrality_avg": [
    [
        "Bernardo Silva",
        -0.1578055168515695
    ],
    [
        "E. Mangala",
        -0.05263157894736842
    ],
    [
        "S. Ag\u00fcero",
        -0.052612835507572356
    ]
]
```

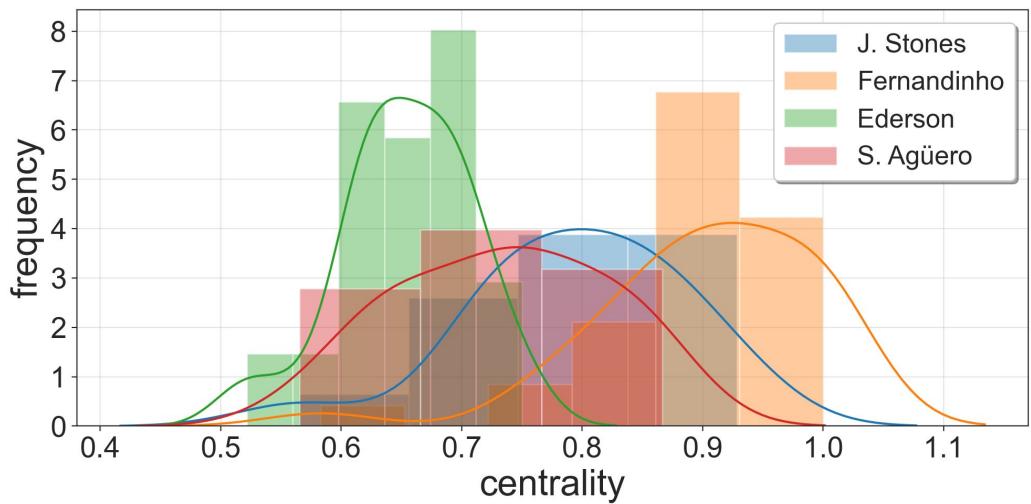
```
"max_betweenness_centrality_avg": [
    [
        "Fernandinho",
        0.018057358090252825
    ],
    [
        "K. De Bruyne",
        0.015947621734104987
    ],
    [
        "David Silva",
        0.013941738418413058
    ]
]
```





# 1. Player analysis

```
Stones mean: 0.7951358384523722
Fernandinho mean: 0.9035565941070403
Ederson mean: 0.6538303886337493
Aguero mean: 0.7339598249150678
```



## Player Analysis (Manchester City - Season 17/18) - NetworkX

Max Closeness centrality	0.385 (Fernandinho - C)
Min Closeness centrality	0.014 (E. Mangala - D)

```
"max_closeness_centrality_avg": [
    [ {
        "Fernandinho",
        0.3850491613649509
    },
    [ {
        "K. De Bruyne",
        0.35103142471563525
    },
    [ {
        "David Silva",
        0.22678637152321363
    },
], [
    [ {
        "E. Mangala",
        0.01425438596491228
    },
    [ {
        "S. Ag\u00f3cero",
        0.014874141876430205
    },
]
```



## 2. Team analysis



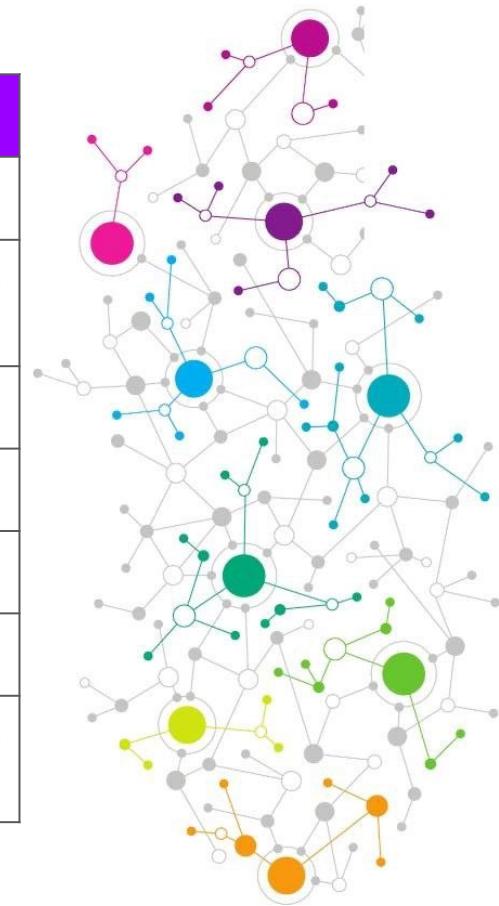
La **team analysis** ha lo scopo di capire quanto una squadra sia connessa e come ciò influisce nel modo di giocare di un team.





## 2. Team analysis

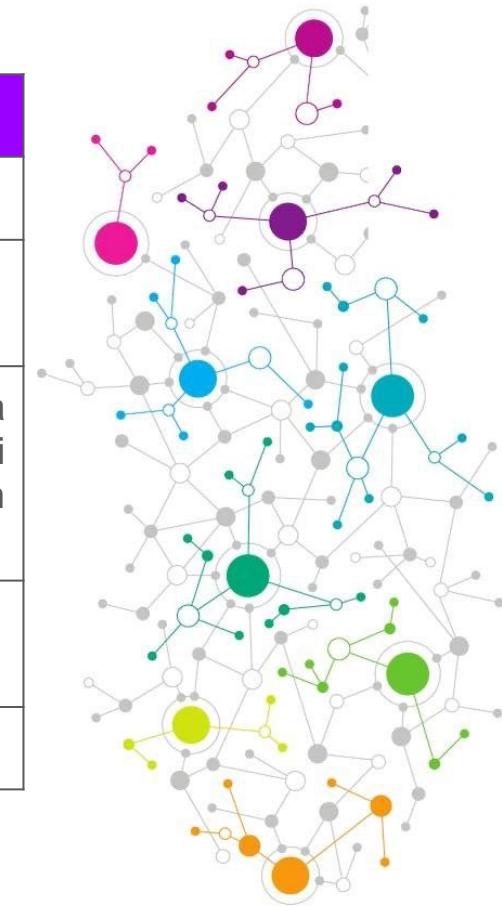
INDICE	DESCRIZIONE
Edges	Numero degli archi presenti nella rete
Density	Indica quanto le entità di un grafo sono interconnesse tra di loro.
Average degree	Valore medio del degree del grafo.
Average indegree	Valore medio dell'indegree del grafo.
Average outdegree	Valore medio dell'outdegree del grafo.
Average degree centrality	Valore medio del degree centrality del grafo.
Edge connectivity	Numero minimo di edges che è necessario rimuovere per disconnettere la rete.





## 2. Team analysis

INDICE	DESCRIZIONE
Average clustering coefficient	Media dei valori dei clustering coefficient
Max K-core	Numero k massimo per la formazione di un k core.
Max clique number (graph)	Numero di cricca massimo. Una cricca in una squadra rappresenta un sottoinsieme di giocatori che sono collegati in coppia con passaggi diretti
Number of triangles (graph)	Numero totale di triangoli presenti all'interno della rete.
Transitivity (graph)	$T = \frac{\# \text{ triangles}}{\# \text{ triads}}$





## 2. Team analysis

Team Analysis (Manchester City - Season 17/18) - NetworkX	
Average density	0.687
Average degree	17.38
Average edge connectivity	2.73
Average max clique number	8.34
Average clustering coefficient	0.807
Average transitivity	0.834

Come nella fase precedente, anche in quest'analisi, i valori calcolati per ogni partita sono sottoposti al calcolo della media, utile per avere una panoramica sull'intera stagione.



- ⚽ **Match analysis** è una fusione tra le due precedenti analisi (player + team).
- ⚽ Viene analizzata una partita specifica (Manchester City - Stoke City 7-2) al fine di individuare il contributo dei singoli calciatori in relazione alla connessione del team.

### 3. Match analysis





### 3. Match analysis

 Analisi del match considerando  
**la formazione iniziale e sostituzioni**

- Analisi dei risultati ottenuti con Gephi,
- Comparazione di tali risultati con i dati (NetworkX) contenuti nel file JSON che si riferisce al match in questione.

 Analisi del match considerando  
**la formazione iniziale**

- Analisi di rete con relativa discussione dei risultati e visualizzazione grafica di entrambe le formazioni che hanno preso parte al Match
- Comparazione risultati analisi Man. City con un articolo di StatsBomb che descrive il match



### 3. Match analysis

INDICE	DESCRIZIONE
<b>Modularity</b>	Misura utile per determinare se all'interno della rete sono presenti moduli o comunità.
<b>Network diameter</b>	Diametro del grafo, la lunghezza della più grande distanza geodetica tra ogni coppia di nodi.
<b>Shortest path length</b>	è quel percorso che collega i vertici e che minimizza la somma dei costi associati all'attraversamento di ciascun arco.



### 3. Match analysis

#### Formazione iniziale + sostituzioni

Match analysis (Manchester City - Stoke City) - NetworkX - Player	
Nodes	14
Edges	144
Max Degree	25 (J. Stones)
Min Degree	13 (Ederson)
Max indegree	12 (F. Delph - D. Silva - J. Stones- R. Sterling)
Min indegree	7 (Ederson)

Match analysis (Manchester City - Stoke City) - Gephi - Player	
Nodes	14
Edges	144
Max Degree	25 (J. Stones)
Min Degree	13 (Ederson)
Max indegree	12 (F. Delph - D. Silva - J. Stones- R. Sterling)
Min indegree	7 (Ederson)



### 3. Match analysis

#### Formazione iniziale + sostituzioni



Match analysis (Manchester City - Stoke City) - NetworkX - Player	
<b>Max outdegree</b>	13 (J. Stones - N. Otamendi)
<b>Min outdegree</b>	6 (Ederson)
<b>Max closeness c.</b>	0.928 (F. Delph - J. Stones - R. Sterling - D. Silva)
<b>Min closeness c.</b>	0.68 (Ederson)
<b>Max betweenness c.</b>	0.035 (J. Stones)
<b>Min betweenness c.</b>	0.0012 (G. Jesus - A)

Match analysis (Manchester City - Stoke City) - Gephi - Player	
<b>Max outdegree</b>	13 (J. Stones - N. Otamendi)
<b>Min outdegree</b>	6 (Ederson)
<b>Max closeness c.</b>	1.0 (N. Otamendi - J. Stones)
<b>Min closeness c.</b>	0.65 (Ederson)
<b>Max betweenness c.</b>	0.035 (J. Stones)
<b>Min betweenness c.</b>	0.0012 (G. Jesus - A)



### 3. Match analysis

#### Formazione iniziale + sostituzioni



##### Match analysis (Manchester City - Stoke City) - NetworkX

<b>Max c. coefficient</b>	0.964 (G. Jesus)
<b>Min c. coefficient</b>	0.762 (J. Stones)
<b>Max PageRank</b>	0.113 (J. Stones)
<b>Min PageRank</b>	0.033 (Bernardo Silva)

##### Match analysis (Manchester City - Stoke City) - Gephi

<b>Max c. coefficient</b>	0.964 (G. Jesus)
<b>Min c. coefficient</b>	0.762 (J. Stones)
<b>Max PageRank</b>	0.084 (J. Stones)
<b>Min PageRank</b>	0.051 (Ederson)



### 3. Match analysis

#### Formazione iniziale + sostituzioni

Match analysis (Manchester City - Stoke City) - NetworkX - Teams

Density	0.791
Avg. clustering coefficient	0.837
Average degree	10.286
Edge connectivity	6
Max K-Core	16
Max clique number	6
Transitivity	0.879

Match analysis (Manchester City - Stoke City) - Gephi - Team

Density	0.791
Avg. clustering coefficient	0.831
Average degree	10.826

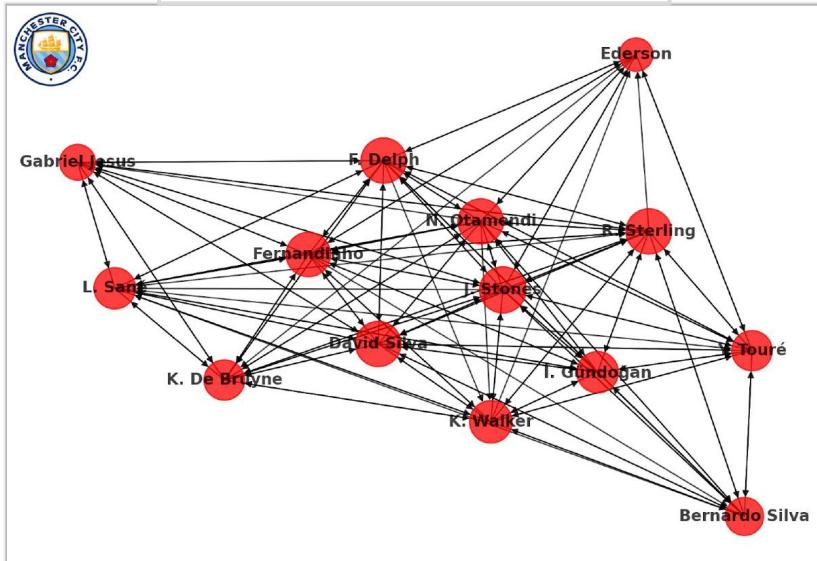


# 3. Match analysis

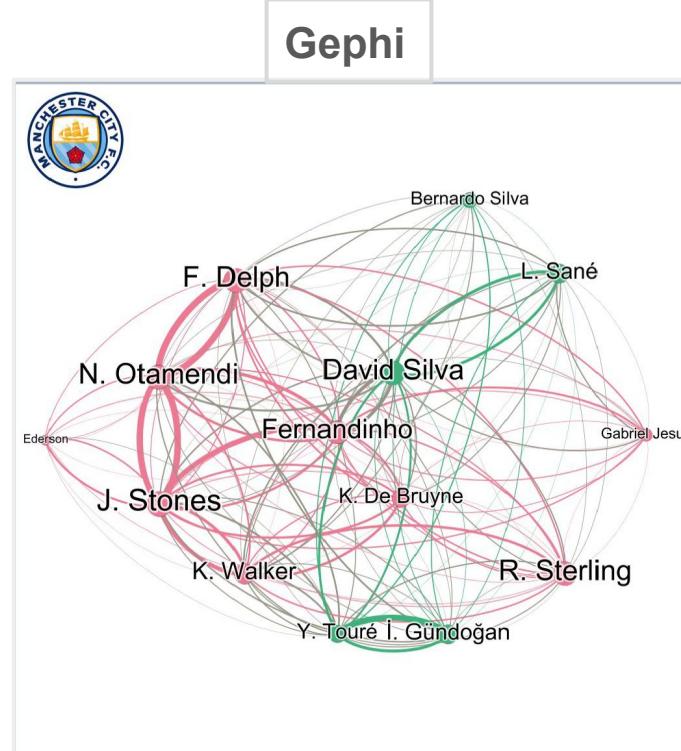
## Formazione iniziale + sostituzioni



NetworkX Digraph



Gephi





# 3. Match analysis

## Formazione iniziale



Manchester City - Gephi - Team



<b>Node</b>	11
<b>Edges</b>	94
<b>Density</b>	0,855
<b>Average Degree</b>	8,545
<b>Max Degree</b>	20 (Fernandinho)
<b>Max Indegree</b>	10 ( Fernandinho - K. De Bruyne)
<b>Max Outdegree</b>	14 (J. Stones - N. Otamendi - F. Delph - Fernandinho)

Stoke City - Gephi - Team



<b>Node</b>	11
<b>Edges</b>	47
<b>Density</b>	0.427
<b>Average Degree</b>	4.273
<b>Max Degree</b>	13 (Zouma)
<b>Max Indegree</b>	7 (E. Choupo-Moting)
<b>Max Outdegree</b>	7 (Zouma)



# 3. Match analysis

## Formazione iniziale



Manchester City - Gephi - Team	
Min degree	11 (Ederson)
Min Indegree	6 (Ederson)
Min Outdegree	5 (Ederson)
Avg clustering coefficient	0.877
Max k-cores	14
Modularity	0.048
Network diameter	2

Stoke City - Gephi - Team	
Min degree	4 (Diouf)
Min Indegree	2 (Diouf)
Min Outdegree	2 (Diouf)
Avg clustering coefficient	0.505
Max k-cores	7
Modularity	0.136
Network diameter	3



### 3. Match analysis

#### Formazione iniziale



Player	Closeness	Betweenness	Pagerank %	Clustering coefficient
Ederson	<b>0.66</b>	0.28	<b>6.5</b>	0.95
Walker	0.76	0.78	8.3	0.88
Stones	<b>1.0</b>	2.44	9.7	0.83
Otamendi	<b>1.0</b>	1.78	8.7	0.84
Delph	<b>1.0</b>	2.31	9.6	0.83
De Bruyne	0.90	1.98	10.4	0.83
Fernandinho	<b>1.0</b>	<b>2.77</b>	<b>10.5</b>	<b>0.82</b>
D. Silva	0.90	0.98	9.4	0.90
Sané	0.83	0.73	8.5	0.94
G. Jesus	0.76	<b>0.25</b>	8.3	<b>0.96</b>
Sterling	0.90	1.64	9.4	0.84



# 3. Match analysis

## Formazione iniziale

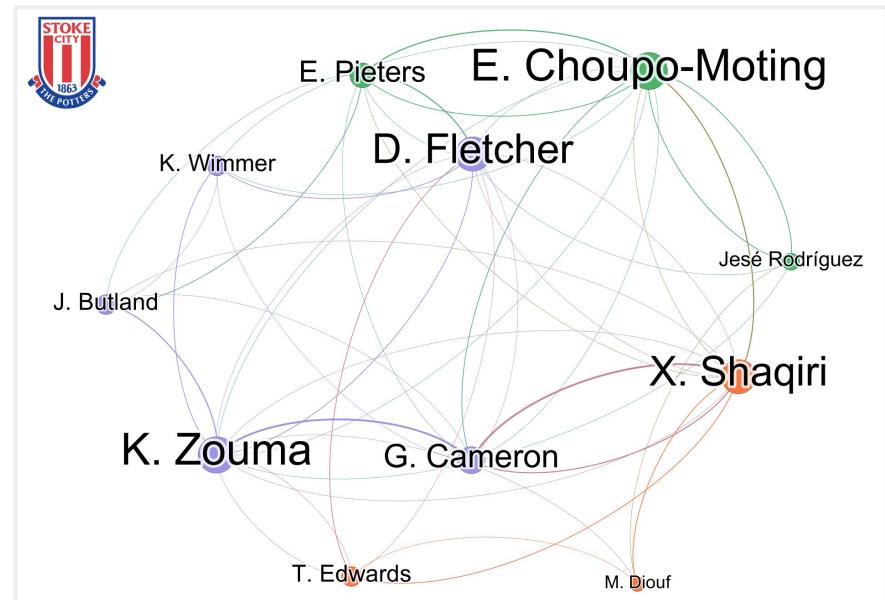
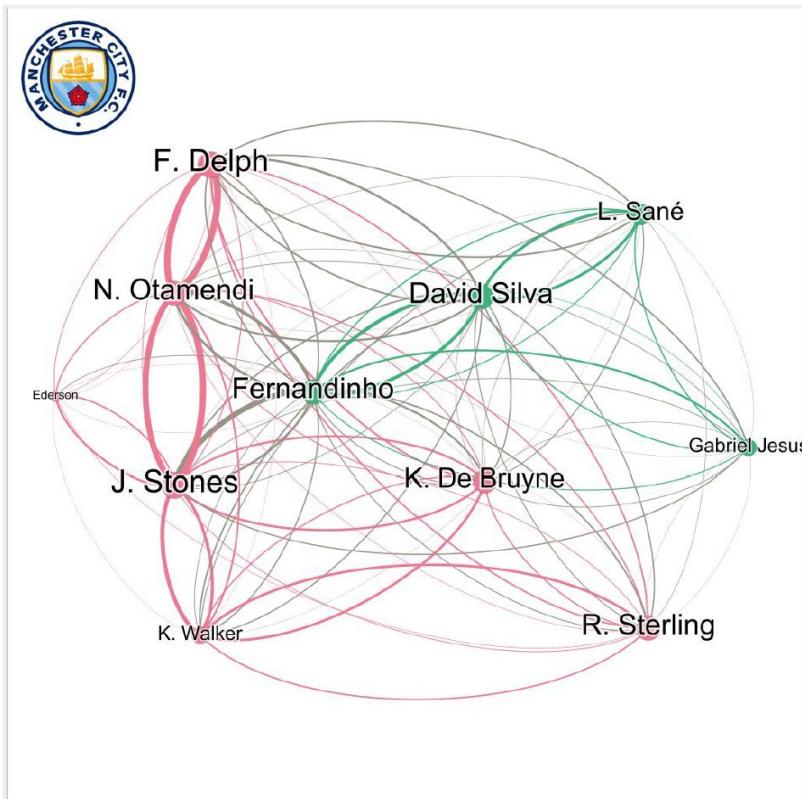


Player	Closeness	Betweenness	Pagerank %	Clustering coefficient
J. Butland	0.58	4.67	7.7	0.45
Edwards	0.55	<b>1.58</b>	5.6	0.6
Zouma	<b>0.76</b>	14.33	11.2	<b>0.36</b>
Wimmer	<b>0.47</b>	3.76	8.2	0.55
Pieters	0.55	6.05	10.1	<b>0.67</b>
Fletcher	<b>0.76</b>	9.68	10.4	0.44
Cameron	0.58	1.59	9.5	0.59
Choupo - Moting	0.71	<b>15.14</b>	<b>15.1</b>	0.452
Diouf	0.5	2.03	<b>4.3</b>	0.5
Jesé Rodriguez	0.55	2.32	5.4	0.5
Shaqiri	0.71	11.80	11.7	0.42



# 3. Match analysis

## Formazione iniziale

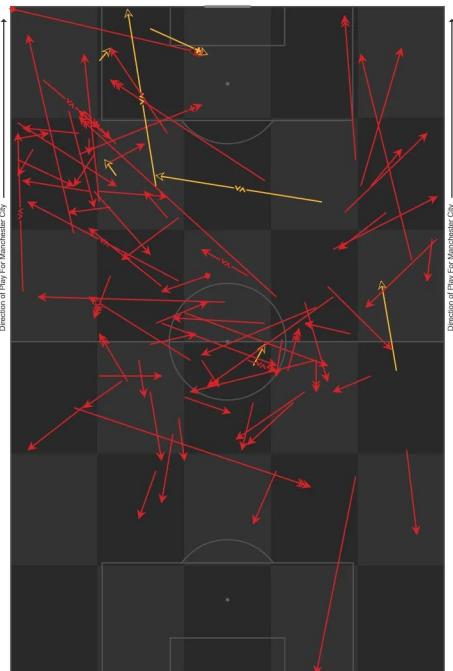




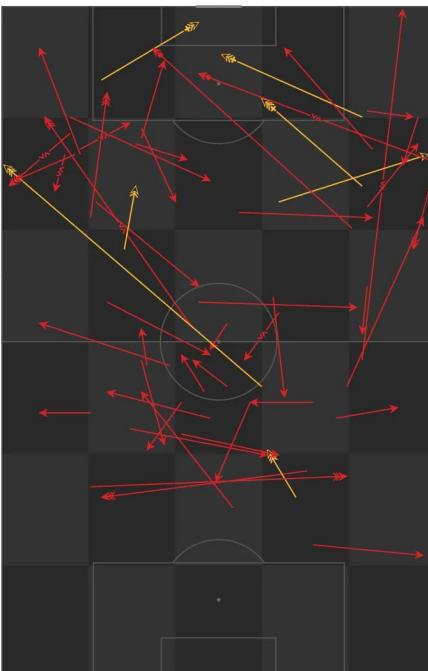
# 3. Match analysis

## Formazione iniziale

David Silva  
Manchester City



Events Kevin De Bruyne  
Manchester City



L'analisi redatta da StatsBomb mette in maggiore evidenza la performance di D. Silva e De Bruyne, tale considerazione viene confermata dai risultati ottenuti nella nostra analisi anche se dai nostri risultati emerge una maggiore centralità da parte di Fernandinho.

**STATSBOMB**



# Conclusioni e sviluppi futuri

Il lavoro svolto in questo progetto ha avuto lo scopo di evidenziare come anche il calcio, nonostante non sia una scienza esatta, possa essere guidato da numeri.

L'obiettivo è stato quello di fornire delle metriche utili ad allenatori o staff per valutare le performance di squadra e dei giocatori.

## Limiti del progetto:

- Tiene traccia solo dei passaggi riusciti
  - Aggiungere una probabilità al peso!
  - Tenere conto dell'importanza del passaggio (Assist)
- Non tiene conto di tiri in porta e gol
- Cosa succede quando un giocatore viene cambiato?

## Sviluppi futuri:

- aggiungere un nodo alla rete che identifichi la porta avversaria, e quindi analizzare altri tipi di edges, i tiri verso la porta avversaria
- analisi di reti two mode, che veda coinvolti due gruppi di attori, collegati tra di loro in base ai duelli in campo



Grazie per  
l'attenzione





# Bibliografia

- ⚽ Peña, López, J., & Touchette. (2012, June 28). *A network theory analysis of football strategies*. Retrieved from <https://arxiv.org/abs/1206.6904>
- ⚽ Korte, Florian, Daniel, Groll, Johannes, & Martin. (2019, July 12). *Play-by-Play Network Analysis in Football*. Retrieved from <https://www.frontiersin.org/articles/10.3389/fpsyg.2019.01738/full>
- ⚽ Cintia, Paolo & Rinzivillo, Salvatore & Pappalardo, Luca. (2015). *A network-based approach to evaluate the performance of football teams*.
- ⚽ Pappalardo, Luca, & Emanuele. (2020, January 28). *Soccer match event dataset*. Retrieved from <https://doi.org/10.6084/m9.figshare.c.4415000.v5>



# Sitografia

- ⚽ <https://wyscout.com/>
- ⚽ *Classic Game Rewind: Manchester City 7-2 Stoke City, October 2017.* (2020, March 25). Retrieved from <https://statsbomb.com/2020/03/classic-game-rewind-manchester-city-7-2-stoke-city-october-2017/>
- ⚽ Mói, R. (n.d.). *Football Passing Networks*. Retrieved from <https://grafos-da-bola.netlify.app/>
- ⚽ Tag: Soccermetrics. (n.d.). Retrieved from <http://www.11tegen11.com/tag/soccermetrics/>