Analyzing Team Scale-Consistency Using Football Passing Networks: **The Premiere League 2015/16 Case.**

David Mateo Carpio R.

January, 2025

EXCELENCIA
MARÍA
DE MAEZTU
2023 - 2027

- Acquisition of **event** and tracking **data** has boosted in-depth analysis of football.

- Opens a path for the definition of **new metrics** to characterize the players and team performances.

Event data collection for a typical Football match

Introduction

- Acquisition of **event** and tracking **data** has boosted in-depth analysis of football.

- Opens a path for the definition of **new metrics** to characterize the players and team performances.

- *Buldú et al.* defined a network-based **metric** to characterize a team's **consistency** during a season.

Event data collection for a typical Football match

OPEN

Consistency and identifiability of football teams: a network science perspective

D. Garrido[1,2], D. R. Antequera[1,2], J. Busquets[3], R. López del Campo[4], R. Resta Serra[4], S. Jos Vielcazat[4] & J. M. Buldú[1,2,5]

[1]D. Garrido: *Scientific reports* ***10*****(1)**, 19735 (2020)

EXCELENCIA MARÍA DE MAEZTU 2023 - 2027

Introduction

- Acquisition of **event** and tracking **data** has boosted in-depth analysis of football.
- Opens a path for the definition of **new metrics** to characterize the players and team performances.
- *Buldú et al.* defined a network-based **metric** to characterize a team's **consistency** during a season.

Event data collection for a typical Football match

OPEN **Consistency and identifiability of football teams: a network science perspective**

D. Garrido[1,2], D. R. Antequera[1,2], J. Busquets[3], R. López del Campo[4], R. Resta Serra[4], S. Jos Vielcazat[4] & J. M. Buldú[1,2,5]

[1]D. Garrido: *Scientific reports* ***10*****(1)**, 19735 (2020)

> Research Question:
>
> *Is consistency a reliable measure for inferring the success of a team's season?*

(1) Extract passing data from a given football match, including initial and final spatial positions.

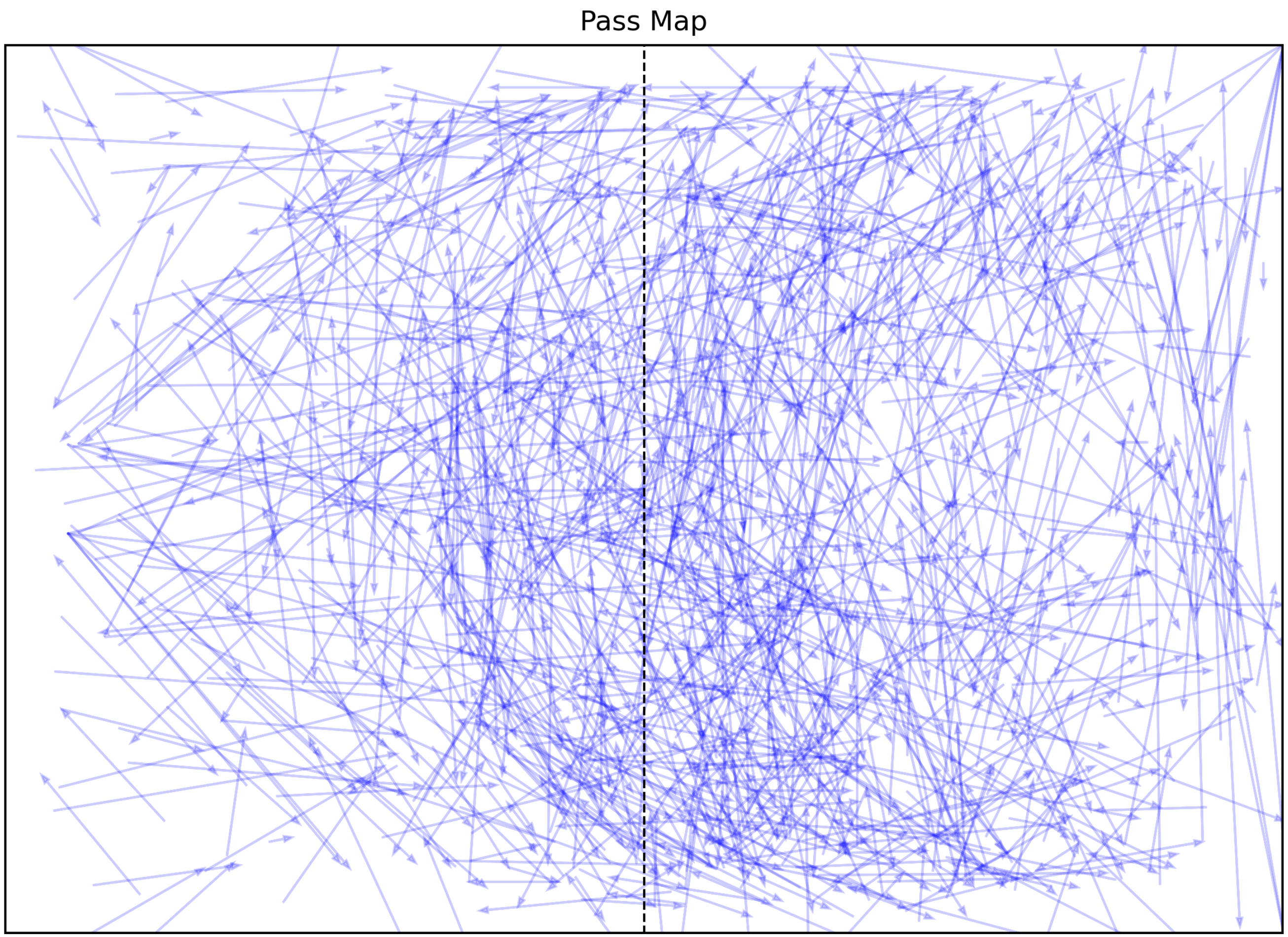


Figure 1: Passing data from a typical football match. This example corresponds to the match Arsenal vs Aston Villa (15/05/16).

(2) Divide the pitch into a $p \times q$ grid.

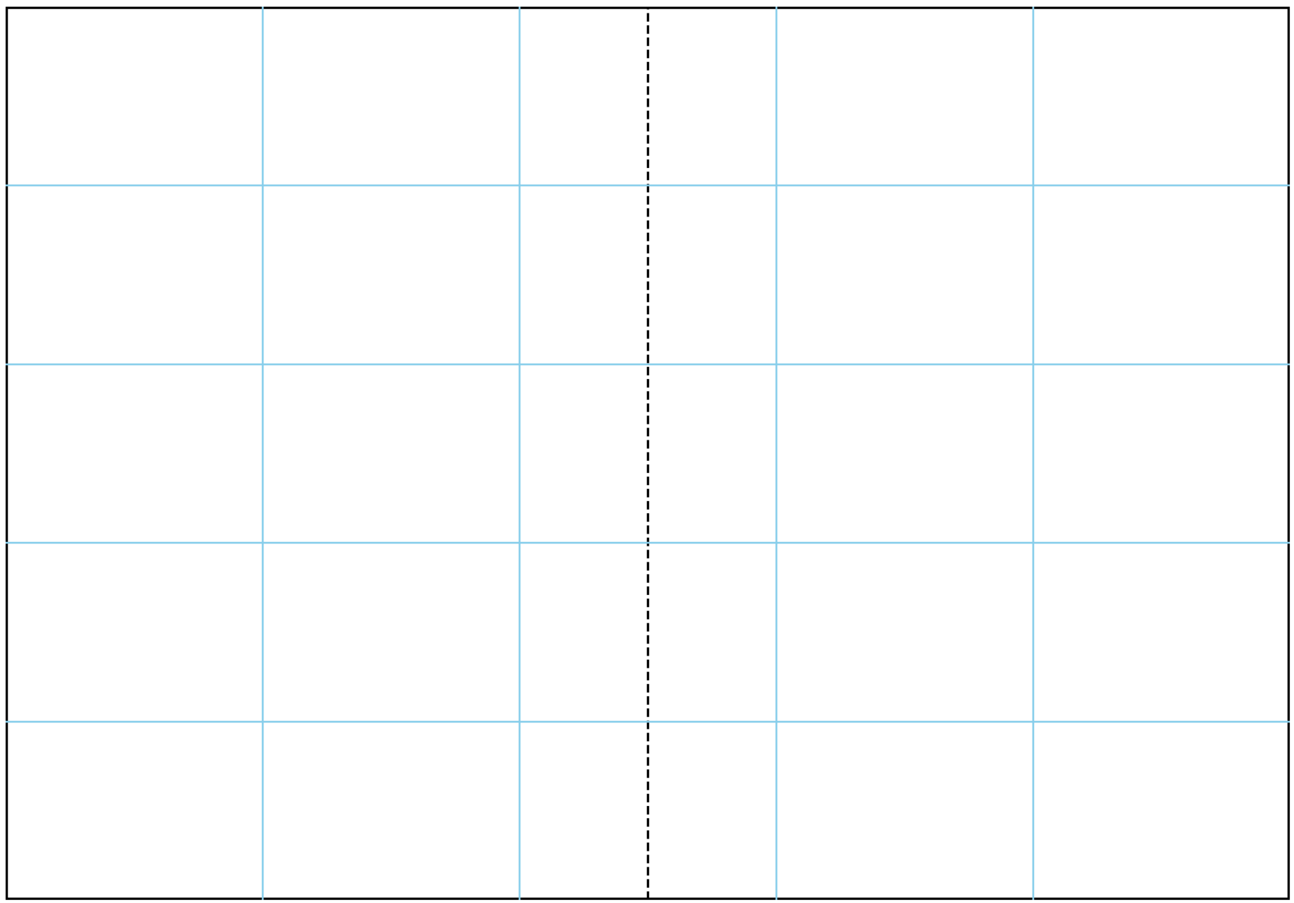


Figure 2: Pitch divided into 5×5 regions.

(3) Assign a **node** to each **region** of the pitch. We obtain $m = p \times q$ nodes.

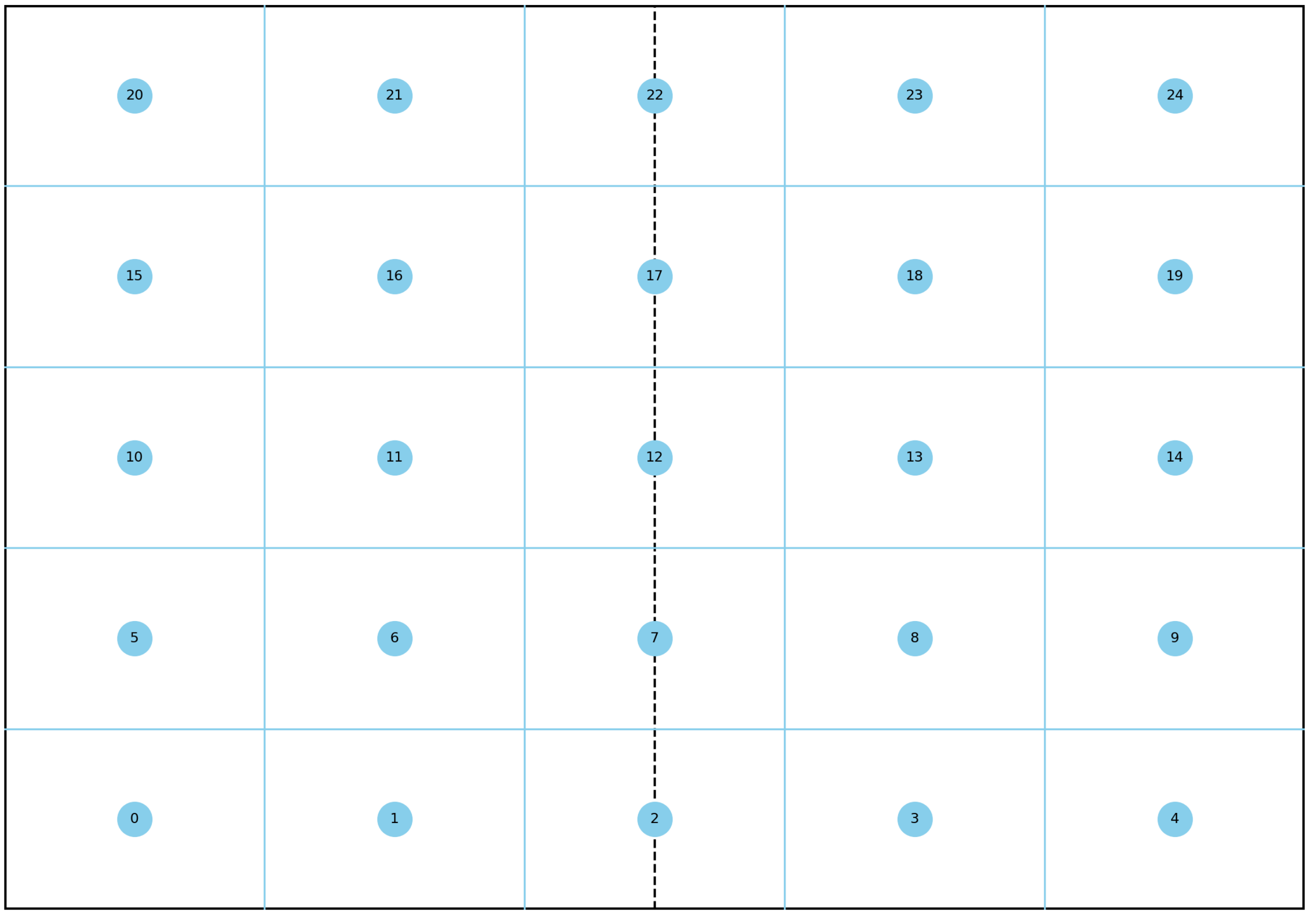


Figure 3: Pitch divided into 25 nodes.

(4) Each pass represents a **directed** link between two nodes.

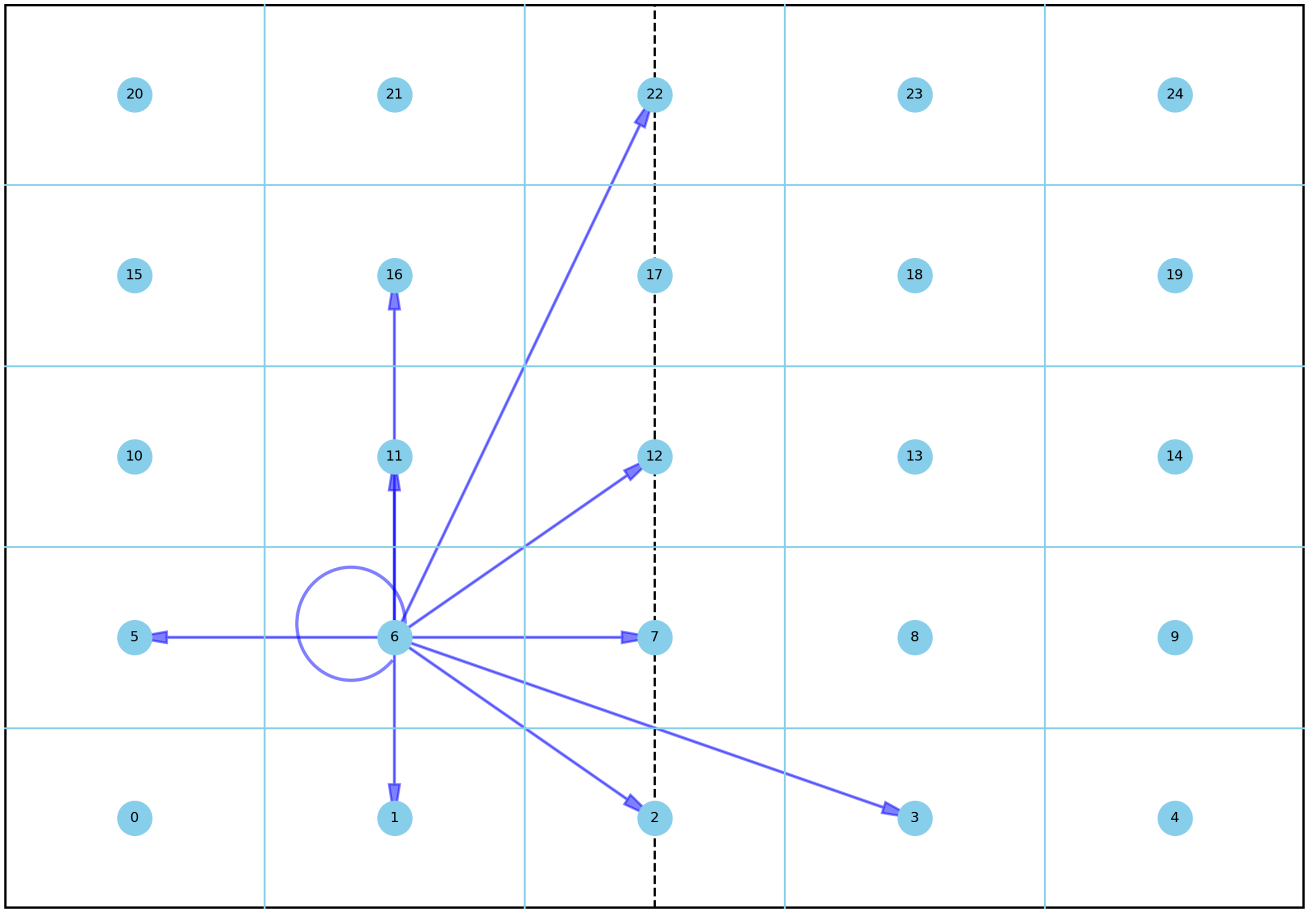


Figure 4: Links created for the node 6 (Arsenal passes) in the match Arsenal vs Aston Villa (15/05/16).

(5) The total number of passes between two regions represents the **weight** of the links.

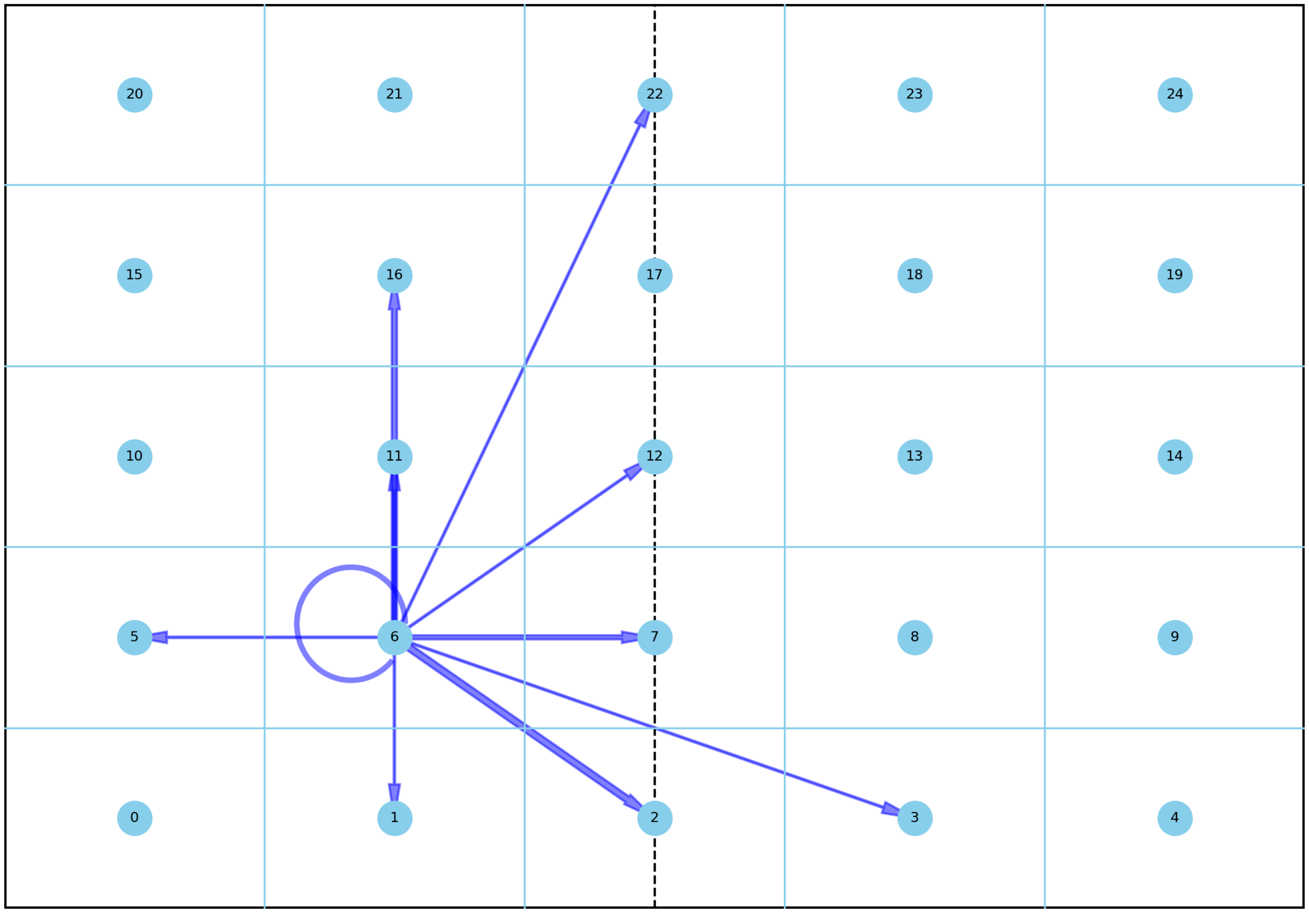


Figure 5: Weighted links created for the node 6 (Arsenal passes) in the match Arsenal vs Aston Villa (15/05/16).

(6) Then, we can construct a weighted directed network for each team.

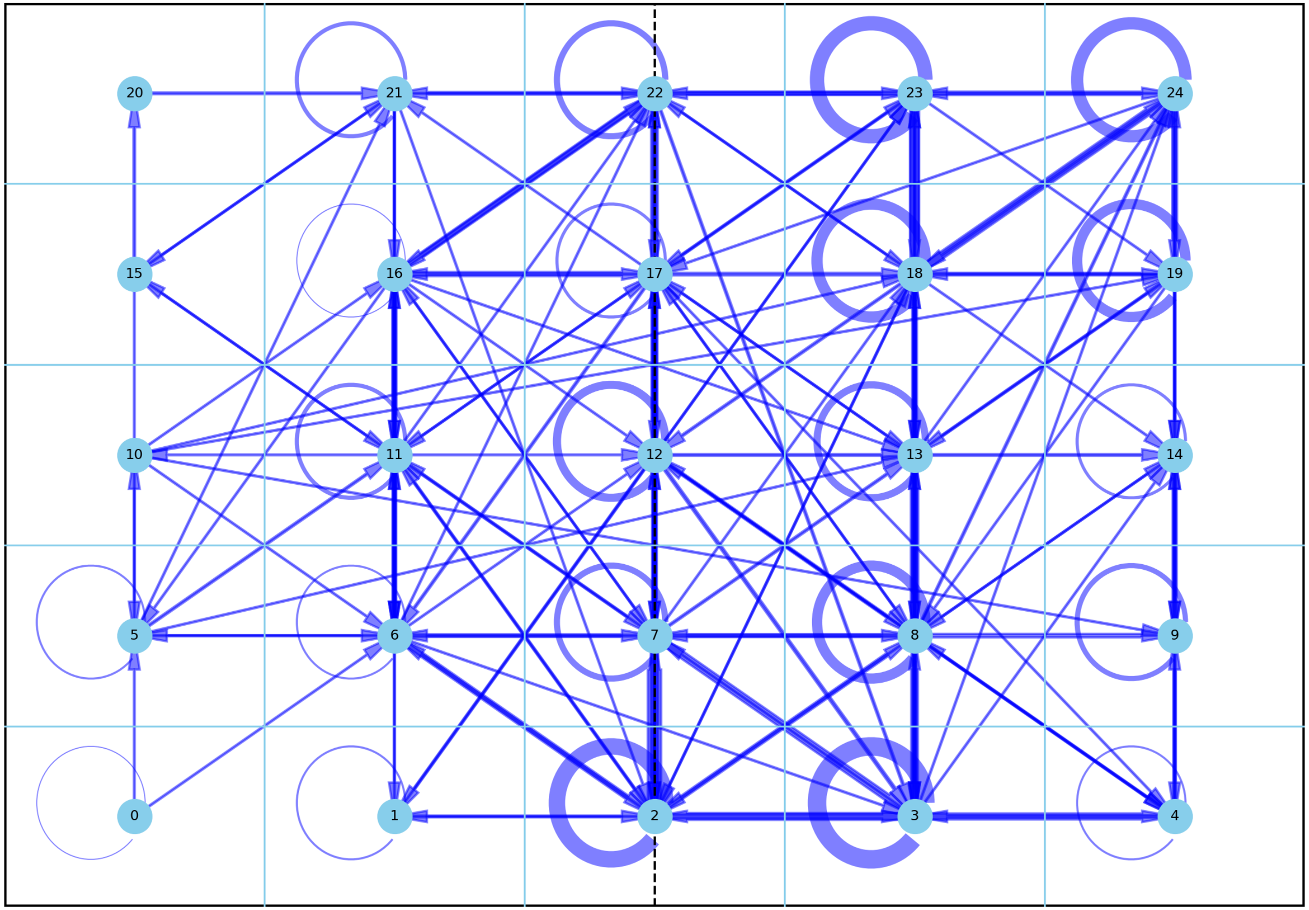


Figure 6: Weighted pitch passing network of **Arsenal** in its match against Aston Villa (15/05/16).

(6) Then, we can construct a weighted directed network for each team.

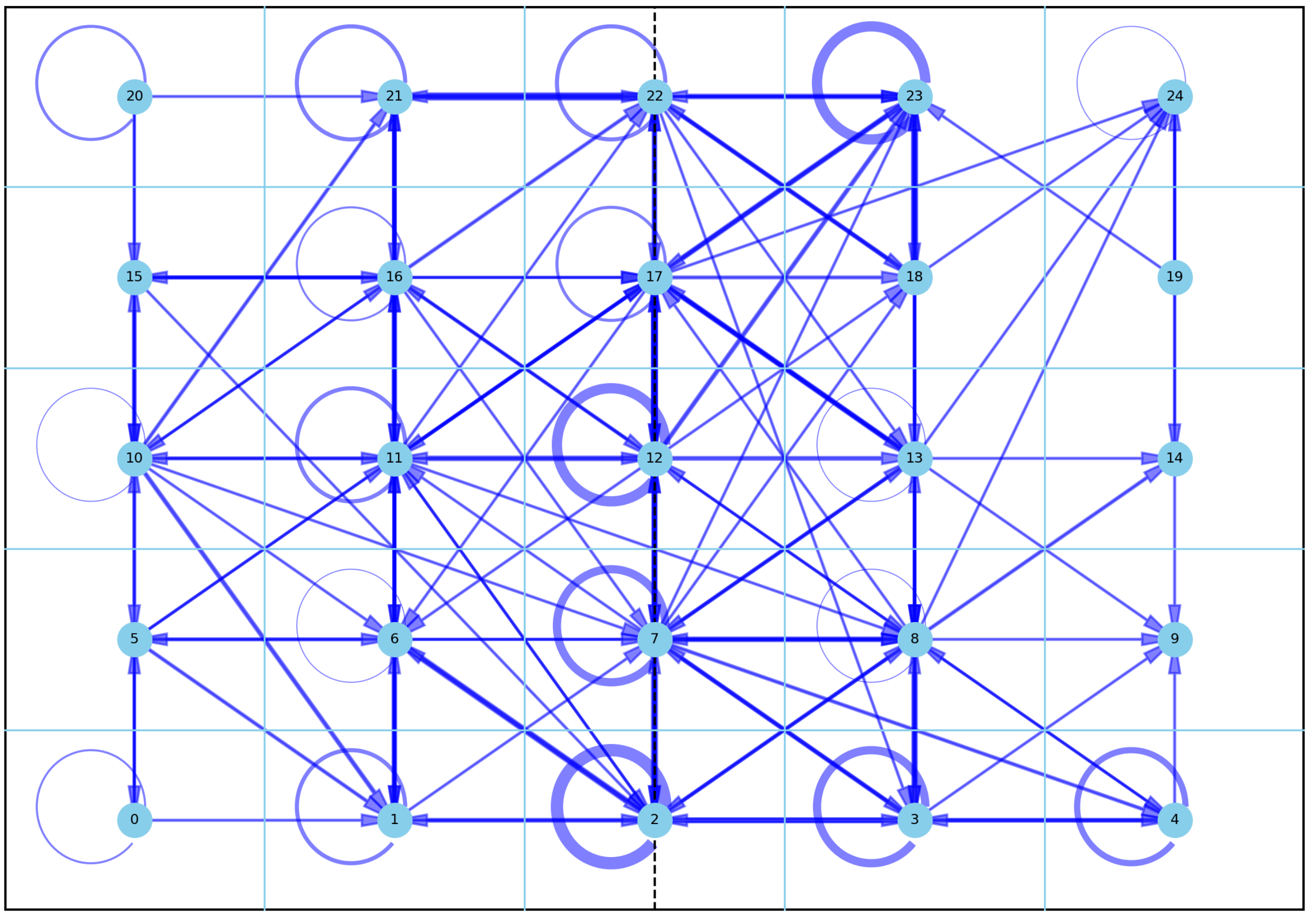


Figure 7: Weighted pitch passing network of **Aston Villa** in its match against Arsenal (15/05/16).

(7) This allows the construction of adjacency matrices for each team in a given match.

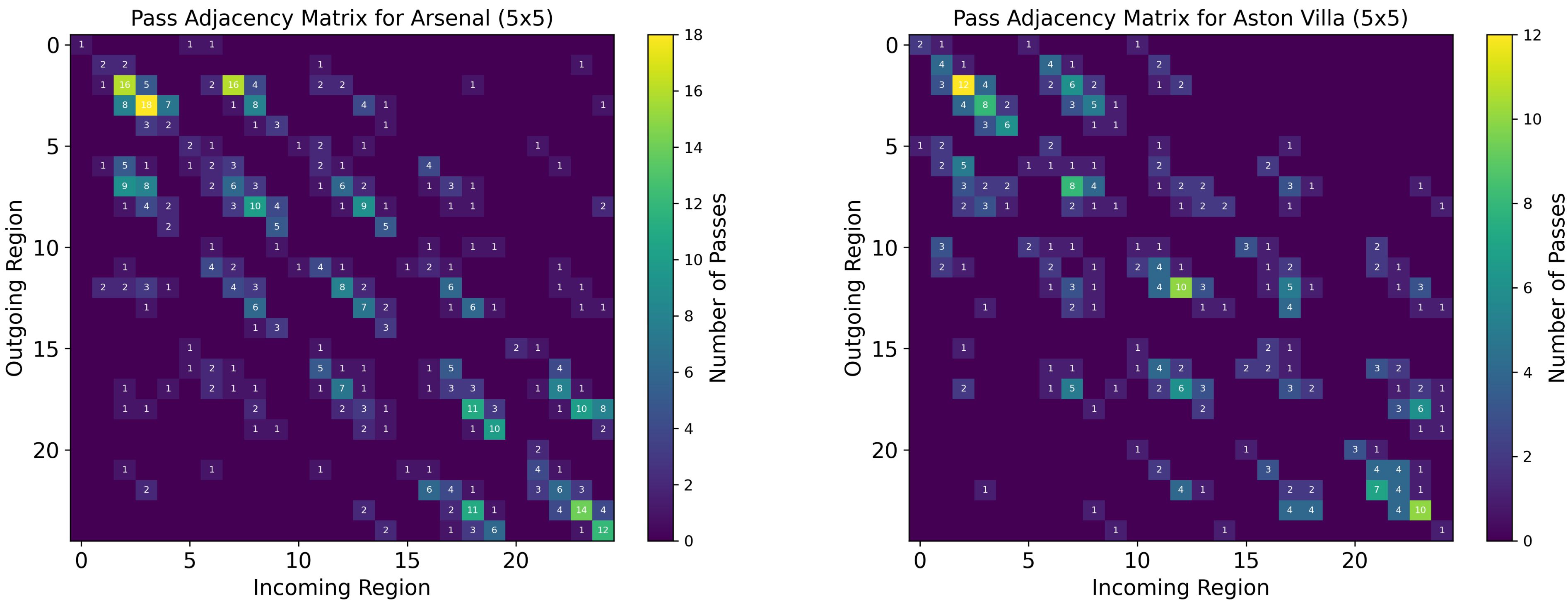


Figure 8: Weighted adjacency matrices for the pitch passing network for the match: Arsenal vs Aston Villa (15/05/16). For (a) **Arsenal** and (b) **Aston Villa**.

(8) By varying the divisions $p \times q$ in the pitch, we can build different adjacency matrices.

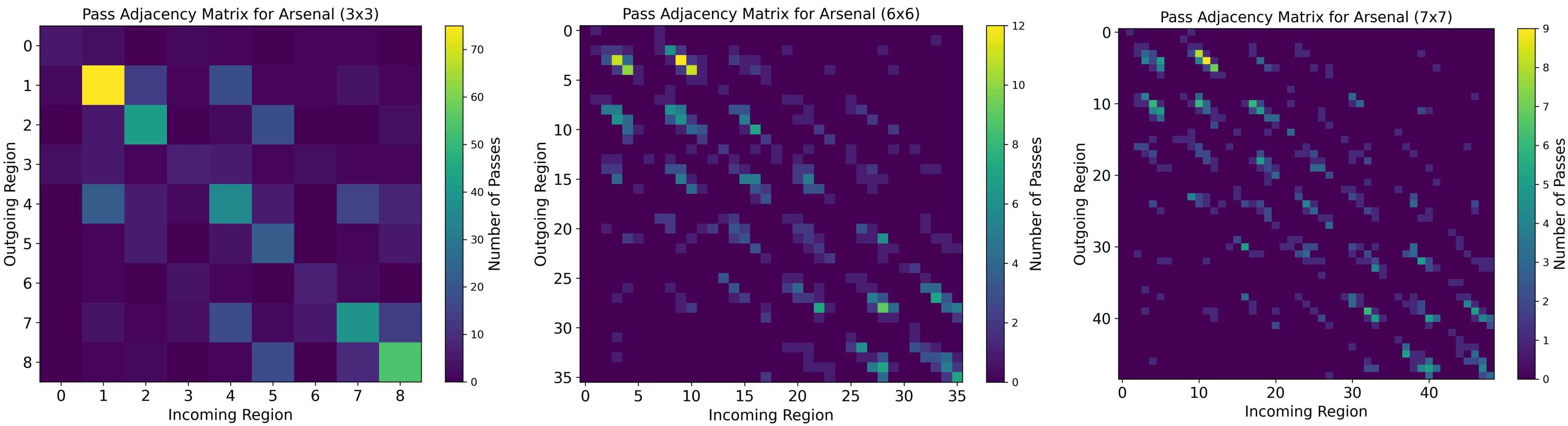


Figure 9: Different weighted adjacency matrices for **Arsenal** calculated for its match against Aston Villa as pitch divisions $p \times q$ vary. (a) 3×3, (b) 6×6, and (c) 7×7.

(9) We vary the divisions of the pitch by:

$$1 \leq p \leq 20, \quad 2 \leq q \leq 20$$

This process results in 380 adjacency matrices for each team in a single match.

(9) We vary the divisions of the pitch by:

$$1 \leq p \leq 20, \quad 2 \leq q \leq 20$$

This process results in 380 adjacency matrices for each team in a single match.

(10) We analyze all 380 matches from the Premier League 2015-2016 season. This amounts to a total of:

$$2 \times 380^2 = 288800$$

adjacency matrices analyzed. A comprehensive dataset for computations.

(9) We vary the divisions of the pitch by:

$$1 \leq p \leq 20, \quad 2 \leq q \leq 20$$

This process results in 380 adjacency matrices for each team in a single match.

(10) We analyze all 380 matches from the Premier League 2015-2016 season. This amounts to a total of:

$$2 \times 380^2 = 288800$$

adjacency matrices analyzed. A comprehensive dataset for computations.

(11) We calculate the ***correlation*** of matrices at <u>a given scale</u> **across a team's matches** during the season.

At a specific scale $m = p \times q$, the **correlation** between two adjacency matrices A and B, representing **two** different **matches of a team**, is calculated as:

$$\mathrm{corr(A,B)} = \frac{\sum_i \sum_j \left(A_{ij} - \bar{A}\right)\left(B_{ij} - \bar{B}\right)}{\sqrt{\left(\sum_i \sum_j \left(A_{ij} - \bar{A}\right)^2\right)\left(\sum_i \sum_j \left(B_{ij} - \bar{B}\right)^2\right)}}$$

At a specific scale $m = p \times q$, the **correlation** between two adjacency matrices A and B, representing **two** different **matches of a team**, is calculated as:

$$\text{corr(A,B)} = \frac{\sum_i \sum_j \left(A_{ij} - \bar{A}\right)\left(B_{ij} - \bar{B}\right)}{\sqrt{\left(\sum_i \sum_j \left(A_{ij} - \bar{A}\right)^2\right)\left(\sum_i \sum_j \left(B_{ij} - \bar{B}\right)^2\right)}}$$

Scale-consistency, $C(m)$***,*** is defined as the average correlation between matrices of the same team at a given scale m:

$$C(m) = \frac{1}{N(N-1)} \sum_{l=1}^{N} \sum_{k=l+1}^{N} \text{corr}(A_l(m), A_k(m)),$$

where the indexes runs over the matches and N is the total number of matches for a team in the season.

Scale consistency, $C(m)$, quantifies the stability of a football team's passing behavior across matches, evaluated at a specific spatial resolution m.

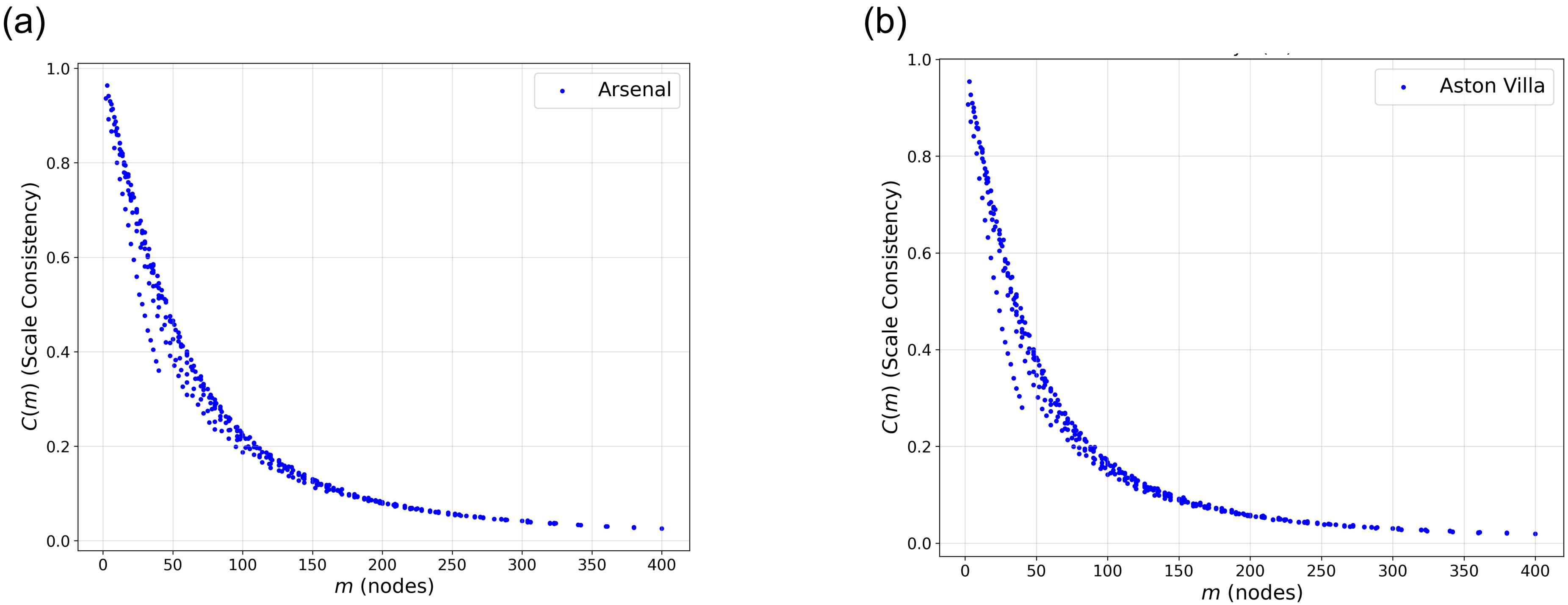


Figure 10: Scale-consistency $C(m)$ of a football team, where m is the number of divisions in the pitch (a) Arsenal, (b) Aston Villa.

The consistencies of two teams can be **compared** at each scale.

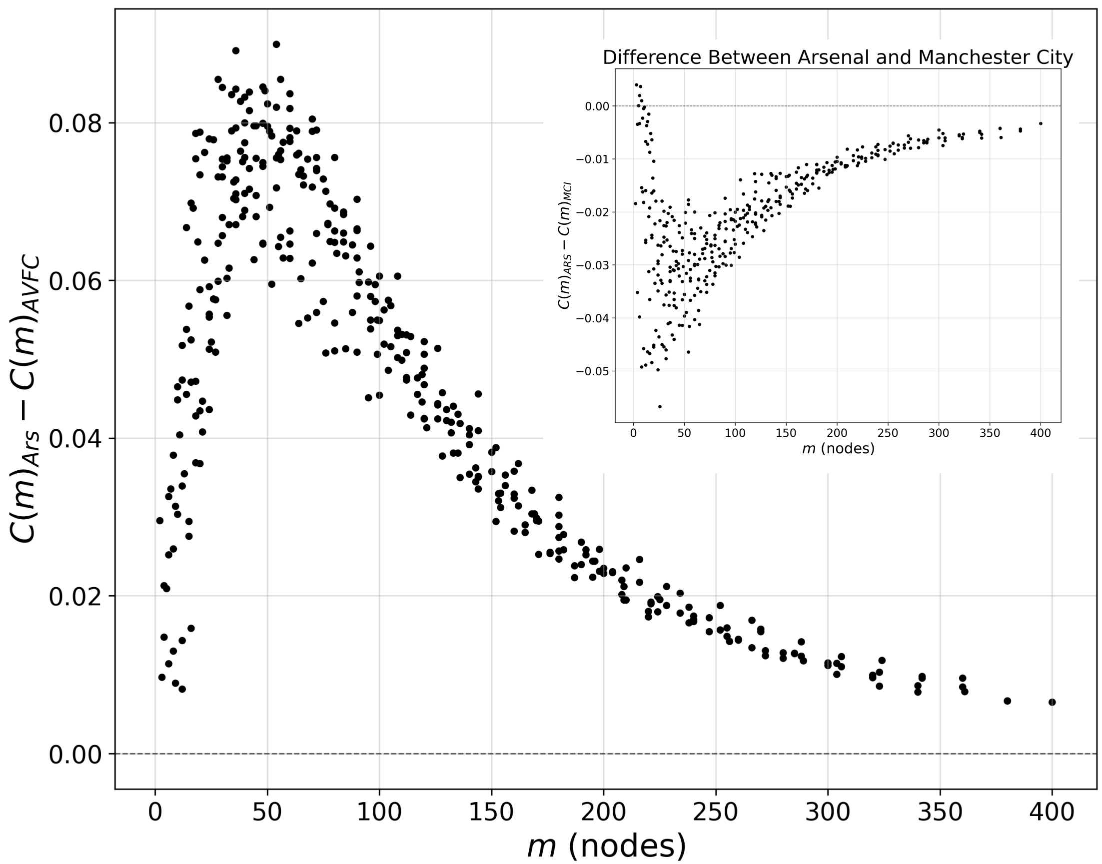


Figure 11: Difference in scale-consistency between Arsenal and Aston Villa as a function of the number of nodes. Inset: Difference in scale-consistency between Manchester City and Arsenal.

- Scale-consistency is higher for Arsenal (which ranked 2^{nd}) than Aston Villa (which ranked 20^{nd}), regardless of the number of divisions in the pitch. The difference peaks around $m \sim 50$.
- Inset: The only team with higher consistency than Arsenal in most of the scales is Manchester City (which ranked 4^{th}).

We define the **average consistency** as the average of the scale-consistency:

$$|C| = \langle C(m) \rangle \,.$$

We define the **average consistency** as the average of the scale-consistency:

$$|C| = \langle C(m) \rangle \,.$$

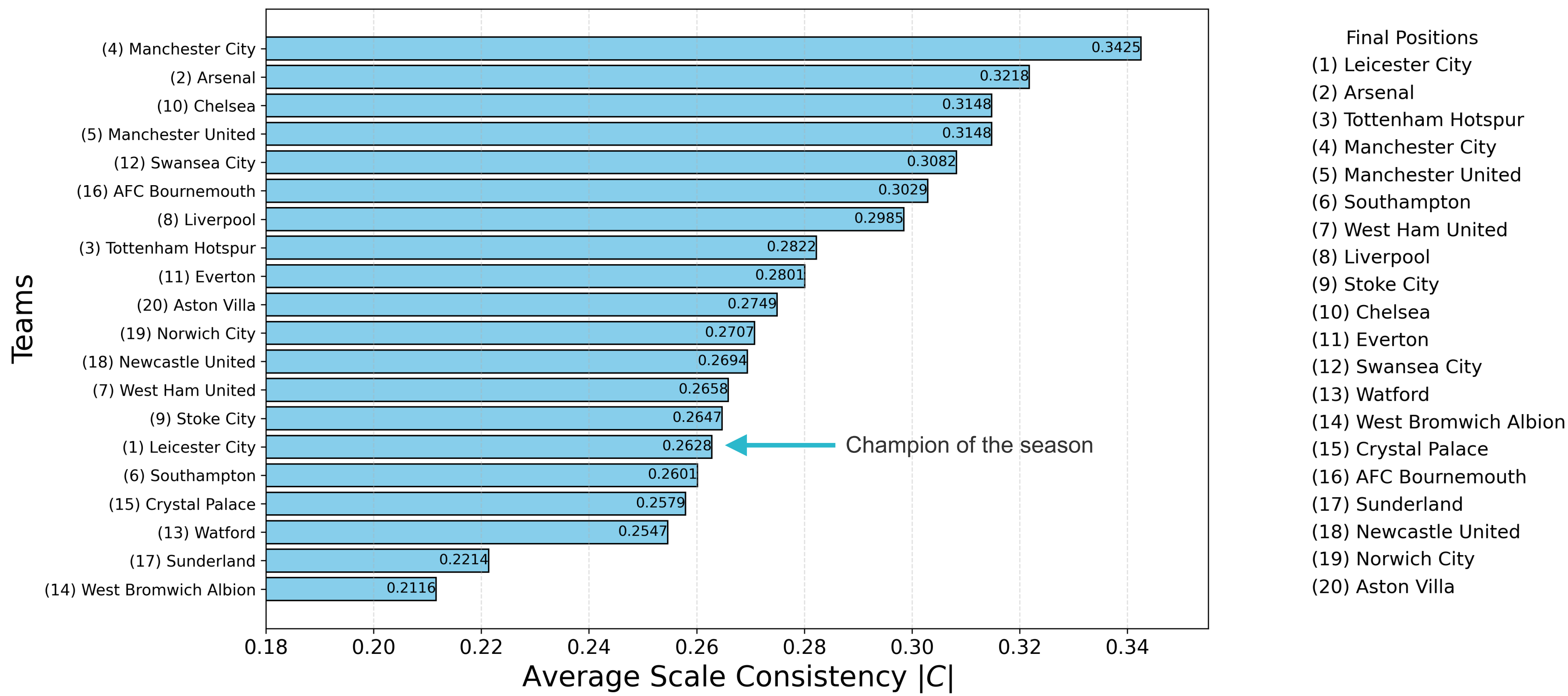
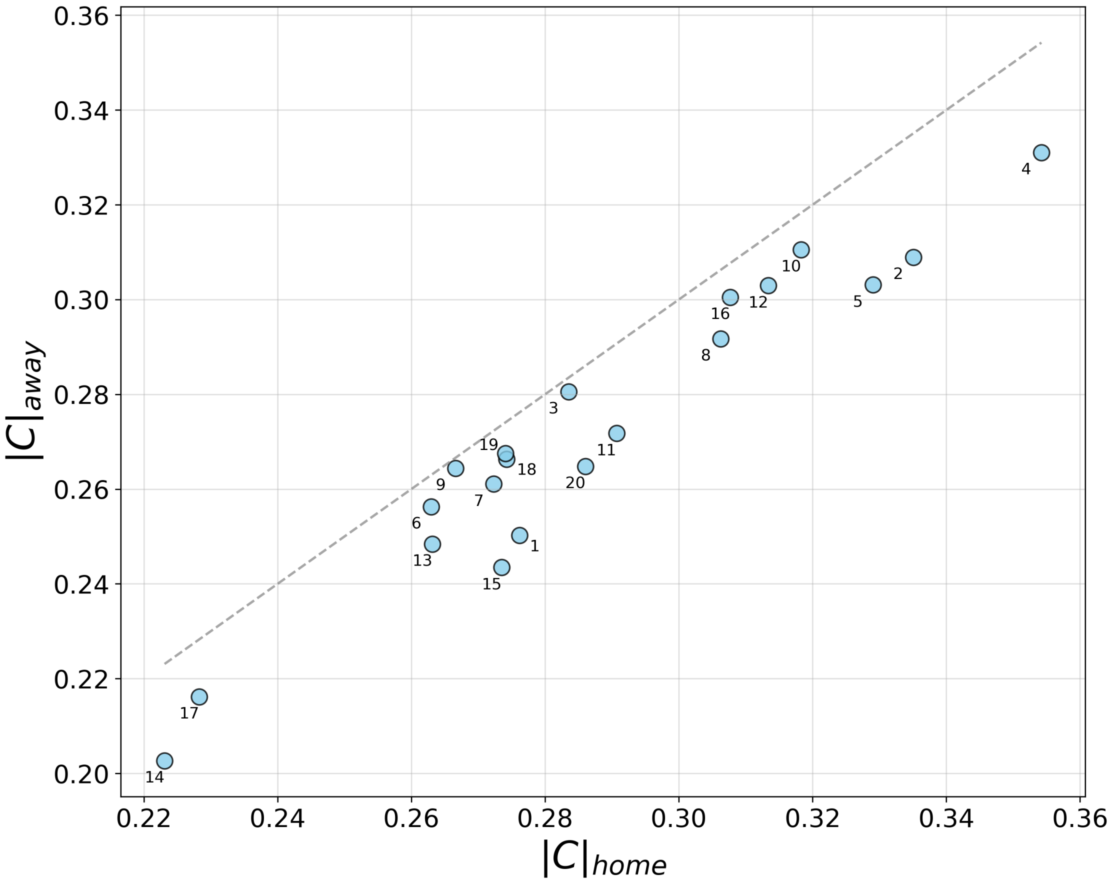


Figure 12: Average consistency $|C|$ for each team in the season. Labels indicate each team's final position at the end of the season. Interestingly, the season's champion displays a low consistency value, highlighting that consistency is not a necessary condition for success in a single season.

We can restrict the calculation of the scale-consistency $C(m)$ to matches played just at home or just away. Thus, we can calculate $|C|_{home}$ and $|C|_{away}$.



Final Positions
(1) Leicester City
(2) Arsenal
(3) Tottenham Hotspur
(4) Manchester City
(5) Manchester United
(6) Southampton
(7) West Ham United
(8) Liverpool
(9) Stoke City
(10) Chelsea
(11) Everton
(12) Swansea City
(13) Watford
(14) West Bromwich Albion
(15) Crystal Palace
(16) AFC Bournemouth
(17) Sunderland
(18) Newcastle United
(19) Norwich City
(20) Aston Villa

It appears that:

- Chelsea ranked 10^{th} yet displayed a high consistency value
- Leicester ranked 1^{st} shows low values of consistency.

Figure 13: Average consistency when playing at home $|C|_{home}$ or away $|C|_{home}$. Labels refer to the final position of the team at the end of the season.

We can categorize teams into three groups using the k-means algorithm.

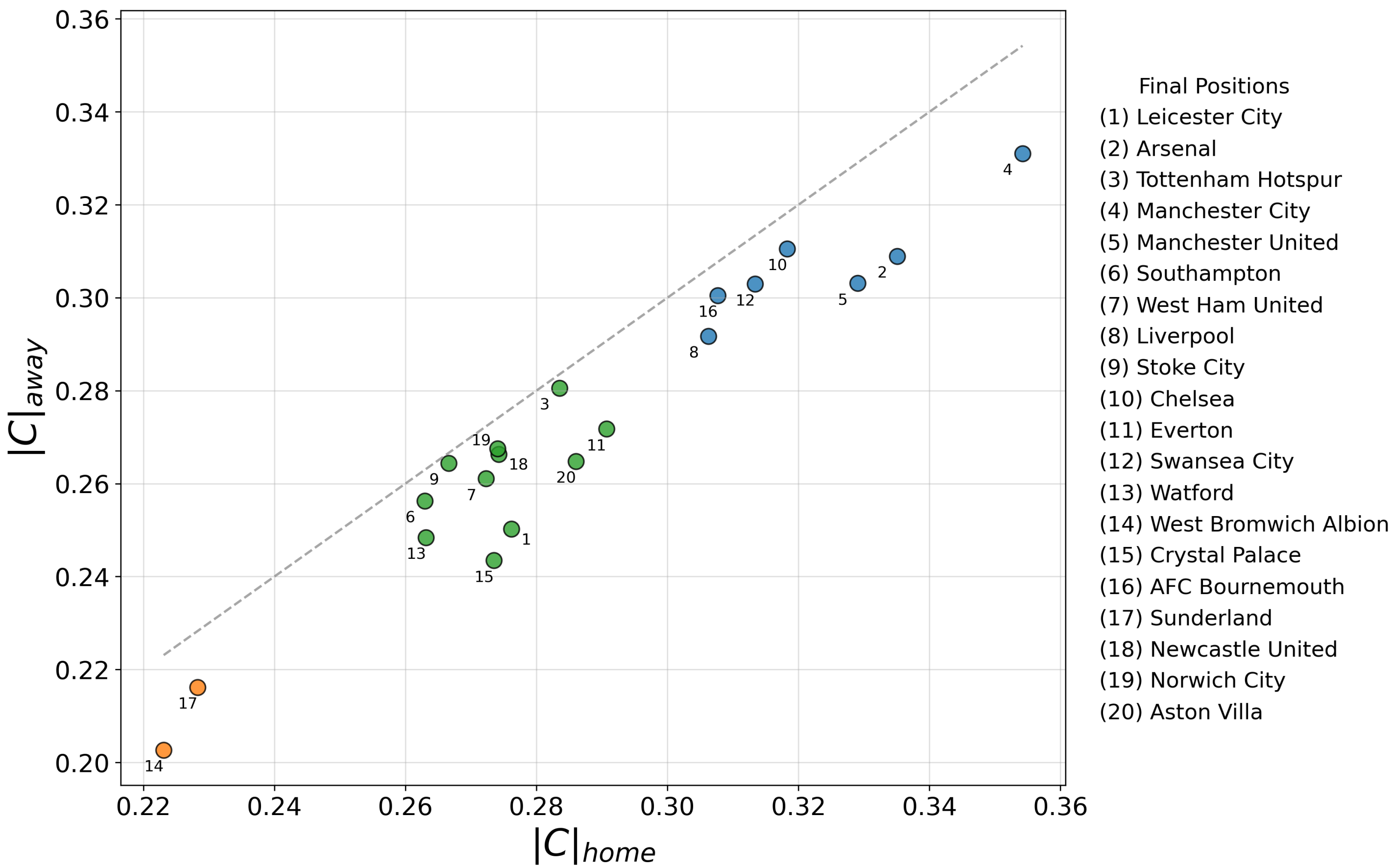


Figure 14: Clusters detected by the k-means algorithm. The optimal number of clusters $k = 3$ was determined using both the Elbow Method and Silhouette Scores.

Clustering categorizes the team as:

- High consistent
- Intermediate consistent
- Low consistent

No clear relationship is observed between these categories and the final rankings at the end of the season.

Now, consider the ranking of teams based on the percentage of points earned by each team over the last five seasons.

Rank	Team	Percentage Points Last 5 Seasons (%)
1	Manchester City	69.8
2	Manchester United	66.3
3	Arsenal	64.6
4	Chelsea	62.8
5	Tottenham Hotspur	60.4
6	Liverpool	56.0
7	Leicester City	54.9
8	AFC Bournemouth	51.6
9	Southampton	51.4
10	Everton	50.0
11	Watford	49.1
12	West Ham United	46.7
13	Stoke City	42.5
14	Crystal Palace	42.2
15	Swansea City	41.8
16	Newcastle United	40.5
17	Norwich City	40.2
18	West Bromwich Albion	38.4
19	Sunderland	34.9
20	Aston Villa	30.2

Table 1: Average points earned over the last five regular seasons (2011-2016) by teams participating in the Premier League (2015-2016), expressed as percentages.

The results become more meaningful when teams are ranked based on their performance over the last five seasons.

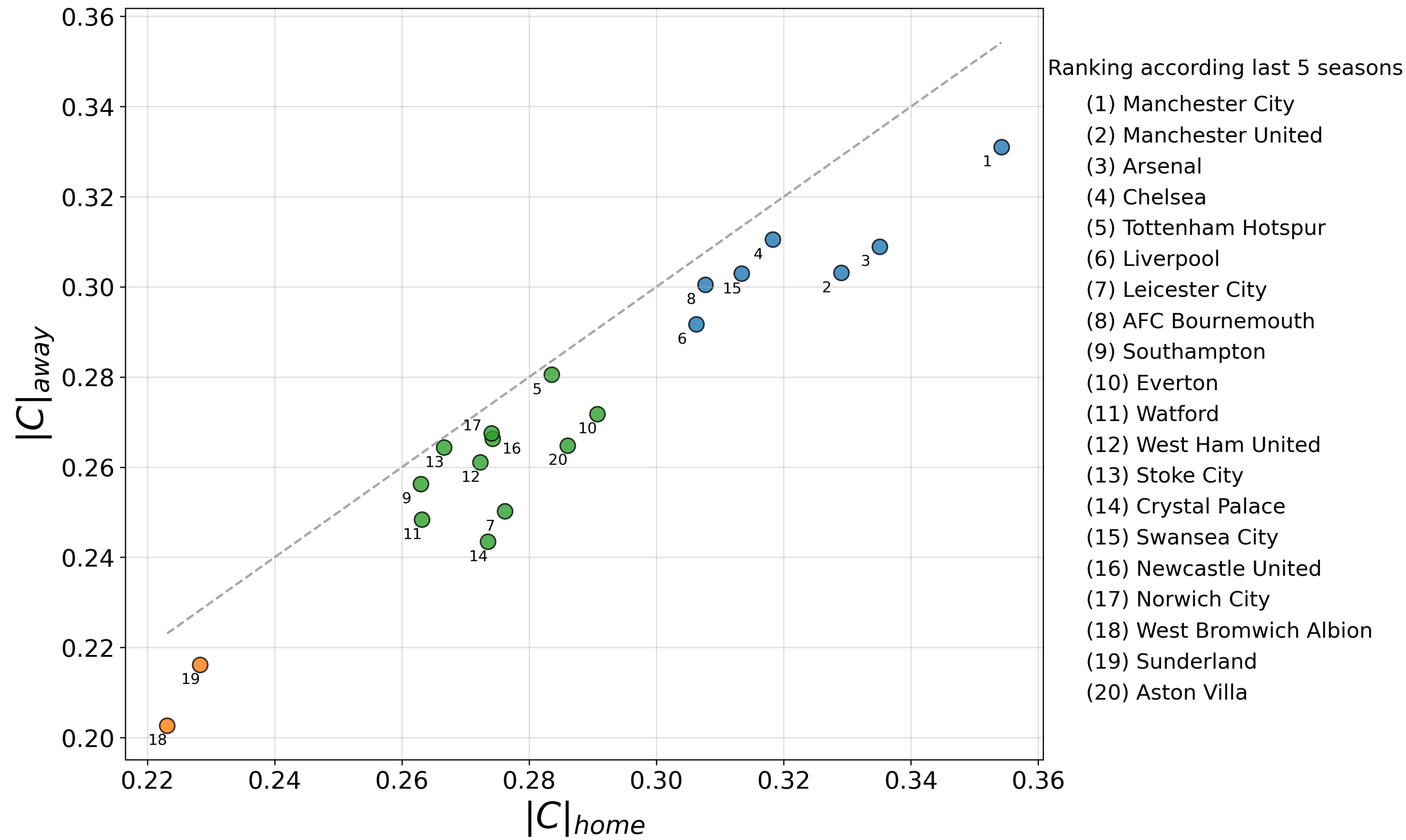


Figure 15: Labels are ranked based on the percentage of points earned by each team over the last five seasons.

Now clusters include:

- 6 out 8 best teams
- 8 teams ranging from 9^{th} to 17^{th}
- 2 out 3 worst teams.

- **Leicester's** low consistency value demonstrates that **consistency** is **not** a **sufficient** condition for achieving **success** in a single season.

- **Leicester's** low consistency value demonstrates that **consistency** is **not** a **sufficient** condition for achieving **success** in a single season.
- **Consistency** in a season correlates with the **team's long-term performance.** This is evidenced by the identified clusters when teams are ranked based on the their performance over the last five seasons.

Summary & Conclusions

- **Leicester's** low consistency value demonstrates that **consistency** is **not** a **sufficient** condition for achieving **success** in a single season.
- **Consistency** in a season correlates with the **team's long-term performance.** This is evidenced by the identified clusters when teams are ranked based on the their performance over the last five seasons.
- **Consistency** alone is **not a reliable measure** for **inferring** the **success** of a team's season. It appears to be more **indicative** of how a team has been **performing** over the **long term.**

Summary & Conclusions

- **Leicester's** low consistency value demonstrates that **consistency** is **not** a **sufficient** condition for achieving **success** in a single season.
- **Consistency** in a season correlates with the **team's long-term performance.** This is evidenced by the identified clusters when teams are ranked based on the their performance over the last five seasons.
- **Consistency** alone is **not a reliable measure** for **inferring** the **success** of a team's season. It appears to be more **indicative** of how a team has been **performing** over the **long term.**
- **Future Work:** Integrating consistency with additional variables, such as **identifiability** (e.g., correlating matrices between matches of different teams) could yield deeper insights into team performance.

References

[1] Garrido, D., Antequera, D. R., Busquets, J., López Del Campo, R., Resta Serra, R., Jos Vielcazat, S., & Buldú, J. M. (2020). Consistency and identifiability of football teams: a network science perspective. Scientific reports, 10(1), 19735.

[2] Gong, B., Zhou, C., Gómez, M. Á., & Buldú, J. M. (2023). Identifiability of Chinese football teams: A complex networks approach. Chaos, Solitons & Fractals, 166, 112922.

[3] Buldú, Javier M., et al. "Defining a historic football team: Using Network Science to analyze Guardiola's FC Barcelona." *Scientific reports* 9.1 (2019): 13602.

[4] StatsBomb Open Data: https://github.com/statsbomb/open-data.

Nice repository of **open data** for football matches.

THANK YOU

for your attention

MINISTERIO DE CIENCIA, INNOVACIÓN Y UNIVERSIDADES

Cofinanciado por la Unión Europea

AGENCIA ESTATAL DE INVESTIGACIÓN