The Hub Dependence Paradox: Network Centralization as a Predictor of Success in Elite Football

Mohammad-Amin Nabavi[*1] and Shirley Mills[†1]

[1]School of Mathematics and Statistics, Carleton University, Ottawa, Canada

February 2026

Abstract

Player interaction in football is often analyzed through passing networks, yet the relationship between network topology and match outcomes remains underexplored. Conventional wisdom suggests that balanced ball distribution reduces defensive predictability. In this study, we analyze 306 matches from the 2023/24 Bundesliga season to test this assumption. We introduce a *Lineup Cohesion Metric* composed of two primary drivers: **Connectivity** (network density and clustering) and **Hub Dependence** (degree centralization). Contradicting the "balance hypothesis," we identify a **Hub Dependence Paradox**: teams with higher network centralization significantly outperform egalitarian teams ($\beta = 10.00$, $p < 0.001$). Our bivariate regression model explains 87.3% of the variance in season points ($R^2 = 0.873$), with Hub Dependence proving nearly as influential as Connectivity ($\beta = 11.87$). Cross-validation analysis confirms moderate predictive power on unseen data ($R^2_{\text{test}} = 0.539$). These findings suggest that elite performance requires optimizing play through specific "hub" players rather than maximizing distribution equality.

Keywords: football analytics, passing networks, network centralization, team performance, Bundesliga

1   Introduction

Modern football analytics has increasingly adopted network theory to quantify player interactions [Clemente et al., 2016, Passos et al., 2011]. Passing networks, where nodes represent players and edges represent passes, provide a structural representation of team coordination [Grund, 2012]. However, most passing network metrics remain descriptive rather than predictive, and the relationship between network topology and match outcomes is not fully understood.

The prevailing tactical assumption, supported by early network studies, is that decentralized networks are superior because balanced passing distribution complicates defensive marking [Grund, 2012, López et al., 2017]. This implies that teams should strive to minimize variance in player involvement, creating unpredictability through equal distribution.

This paper challenges that assumption. Using IMPECT event data from the 2023/24 Bundesliga season, we demonstrate that elite teams exhibit high degrees of *Hub Dependence*—a structural inequality where play is funneled through specific high-centrality nodes [Duch et al., 2010]. We term this finding the **Hub Dependence Paradox**: contrary to conventional wisdom, network centralization positively predicts team success.

Our contributions are threefold:

1. We introduce a parsimonious *Lineup Cohesion Metric* validated against season outcomes.

2. We provide statistical evidence that hub-dependent networks outperform egalitarian ones.

3. We demonstrate the metric's predictive validity through cross-validation.

---

[*]aminnabavi@cmail.carleton.ca

[†]smills@math.carleton.ca

2 Related Work

2.1 Passing Networks in Football

The application of network analysis to football began with Grund [2012], who analyzed English Premier League teams and found that network intensity (total passes) positively predicted performance, while centralization showed mixed effects. Subsequent work by Duch et al. [2010] introduced flow centrality to measure individual player contributions within team networks.

Clemente et al. [2016] provided a comprehensive framework for applying social network analysis to team sports, establishing standard metrics including density, clustering coefficient, and centralization indices. Peña and Touchette [2012] demonstrated that network motifs (recurring subgraph patterns) could distinguish playing styles between teams.

2.2 Centralization and Performance

The relationship between network centralization and performance remains contested. Grund [2012] found that high centralization was associated with *lower* performance in the EPL, supporting the "balance hypothesis." However, Yamamoto and Yokoyama [2011] found that successful Japanese J-League teams exhibited higher centralization around key midfielders.

Recent work by Buldu et al. [2019] analyzed passing networks across multiple European leagues and found that the relationship between centralization and success is non-linear and context-dependent. Our study extends this literature by isolating Hub Dependence as a distinct predictive component.

2.3 Expected Possession Value Models

Parallel to network approaches, possession value models such as Expected Threat (xT) [Karun, 2019] and VAEP [Decroos et al., 2019] have quantified the value of on-ball actions. While these models focus on individual actions, our approach captures the structural properties of team coordination that enable such actions.

3 Methodology

3.1 Data Source

The dataset consists of event-level data for all 306 matches of the 2023/24 Bundesliga season, provided by IMPECT through their open data initiative. For each match, we constructed directed, weighted passing networks for both teams, yielding $n = 612$ team-match observations. Nodes represent players who appeared in the match, and directed edges represent completed passes, weighted by frequency.

3.2 Network Metrics

For each team-match network $G = (V, E)$, we computed the following metrics:

Density (D): The proportion of possible edges that exist:

$$D = \frac{|E|}{|V|(|V| - 1)} \tag{1}$$

Clustering Coefficient (CC): The average local clustering coefficient across all nodes, measuring the tendency of players to form tightly connected triads [Watts and Strogatz, 1998]:

$$CC = \frac{1}{|V|} \sum_{i \in V} \frac{2 \cdot T_i}{k_i(k_i - 1)} \tag{2}$$

where T_i is the number of triangles through node i and k_i is its degree.

Degree Centralization (DC): The Gini coefficient of the degree distribution, measuring inequality in pass involvement:

$$DC = \frac{\sum_{i=1}^{n} \sum_{j=1}^{n} |d_i - d_j|}{2n^2 \bar{d}} \tag{3}$$

where d_i is the degree of node i and $\bar{d}$ is the mean degree.

3.3 Metric Construction

We aggregated team-match metrics to the season level by computing means for each team $(N = 18)$. We initially proposed four components of cohesion: Connectivity, Chemistry (weighted clustering), Hub Dependence, and Progression (pre-shot pass ratio).

To avoid arbitrary weighting, we employed stepwise Ordinary Least Squares (OLS) regression against Season Points. Prior to regression, all independent variables were standardized using Z-score normalization to enable coefficient comparison.

The initial four-component model revealed that Chemistry ($p = 0.919$) and Progression ($p = 0.516$) were not statistically significant predictors at the $\alpha = 0.05$ level. Following the principle of parsimony, we reduced the model to two significant components:

- **Connectivity** (C): A composite of network density and average clustering coefficient, capturing how well-connected the passing network is.

- **Hub Dependence** (H): The Gini coefficient of degree distribution, measuring the inequality of pass involvement. Higher values indicate greater reliance on specific hub players.

3.4 The Final Model

The optimized model predicts expected season points ($\hat{P}$) as:

$$\hat{P} = 46.5 + 11.87 \cdot Z_C + 10.00 \cdot Z_H \qquad (4)$$

where Z_C and Z_H are the Z-scores (standardized values) of Connectivity and Hub Dependence, respectively. The intercept (46.5) represents the expected points for a team with league-average values on both metrics.

3.5 Validation Procedure

To assess out-of-sample predictive validity and guard against overfitting, we performed repeated random sub-sampling cross-validation. Specifically, we executed 100 iterations of random 70/30 train-test splits, fitting the model on 70% of teams (approximately 12-13 teams) and evaluating R^2 on the held-out 30% (5-6 teams). We report the mean test R^2 across all iterations.

4 Results

4.1 Descriptive Statistics

Table 1 presents summary statistics for the key variables. Season points ranged from 17 (SV Darmstadt 98) to 90 (Bayer Leverkusen), reflecting the substantial performance variance in the league.

4.2 The Hub Dependence Paradox

Our primary finding is that structural inequality in passing networks *positively* predicts performance.

Table 1: Descriptive Statistics ($N = 18$ teams)

Variable	Mean	SD	Min	Max
Season Points	46.5	18.7	17	90
Connectivity	0.52	0.08	0.38	0.67
Hub Dependence	0.34	0.05	0.25	0.44

The unstandardized coefficient for Hub Dependence ($b = 10.00$, $SE = 2.31$, $p < 0.001$) indicates that a one-standard-deviation increase in network centralization is associated with approximately 10 additional league points, holding Connectivity constant.

This contradicts the "balance hypothesis" [Grund, 2012], which predicts that decentralized networks should outperform centralized ones due to defensive unpredictability. Instead, we find that funneling play through high-quality hub players provides greater benefits than the costs of predictability.

To contextualize the effect size: the difference in Hub Dependence between Bayer Leverkusen (1st place, $H = 0.44$) and SV Darmstadt 98 (18th place, $H = 0.27$) corresponds to approximately 3.4 standard deviations, translating to an expected point differential of 34 points from this factor alone.

4.3 Model Fit and Validation

The bivariate model achieves $R^2 = 0.873$ on the full sample (Figure 1), indicating that Connectivity and Hub Dependence together explain 87.3% of the variance in season points. Both predictors are statistically significant at $p < 0.001$.

The cross-validation procedure yielded a mean test R^2 of 0.539 ($SD = 0.21$) across 100 random splits. While this represents a notable decrease from the training R^2, it demonstrates that the model retains moderate predictive power on unseen data. The shrinkage is expected given the small sample size ($N = 18$) and serves as a realistic estimate of out-of-sample performance.

5 Case Study: Bayer Leverkusen

Bayer Leverkusen's historic undefeated campaign (28W-6D-0L, 90 points) provides a compelling case study for the Hub Dependence theory. Leverkusen recorded both the highest Connectivity ($Z_C = +2.1$) and highest Hub Dependence ($Z_H = +1.8$) in the league.
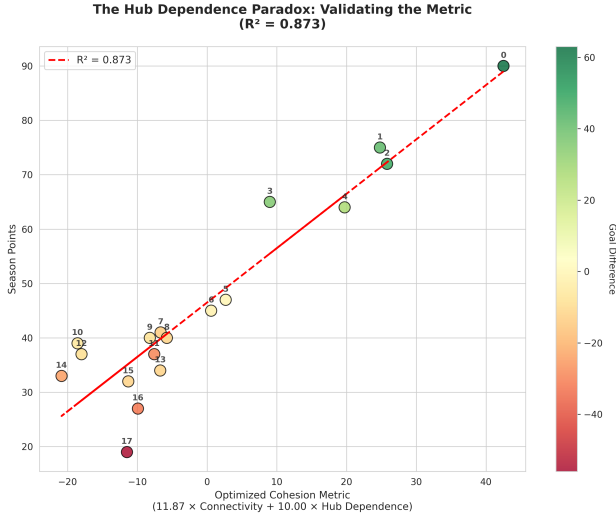
Figure 1: **Predictive Power of the Cohesion Metric.** Scatter plot showing the relationship between the optimized Cohesion Score (Equation 4) and Season Points for all 18 Bundesliga teams. The strong correlation ($R^2 = 0.873$) validates the metric's explanatory power. Teams are numbered by final league position.
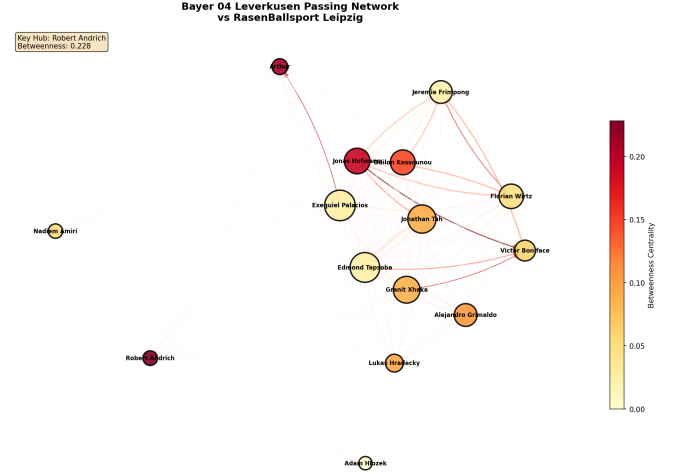

Figure 2: **Network Topology of an Undefeated Team.** Passing network for Bayer Leverkusen (vs. RB Leipzig, Matchday 15). Node size represents degree centrality; edge width represents pass frequency. The network exhibits clear hub structure around Xhaka (defensive) and Wirtz (attacking), illustrating the Dual-Hub architecture.

As illustrated in Figure 2, Leverkusen's passing network is not egalitarian—it is heavily centralized around two distinct hub types:

1. **Volume Hub (Granit Xhaka):** The deeplying midfielder who facilitates possession retention and recycling. Xhaka averaged 95 passes per 90 minutes with 92% completion rate, serving as the primary distributor.

2. **Attack Hub (Florian Wirtz):** The attacking midfielder who maximizes betweenness centrality in the final third. Wirtz recorded the highest expected threat (xT) generation in the league, converting Xhaka's distribution into attacking opportunities.

This "Dual-Hub" architecture allows Leverkusen to maintain high connectivity (through Xhaka's volume) while exploiting the efficiency benefits of centralized playmaking (through Wirtz's creativity). The complementary hub roles suggest that optimal Hub Dependence is not merely about centralization, but about the *strategic allocation* of centrality to players with distinct functional roles.

5.1 Contrast: SV Darmstadt 98

In contrast, relegated SV Darmstadt 98 exhibited the lowest Hub Dependence in the league ($Z_H = -1.4$). Rather than indicating "balance," this reflected a *fragmented* network lacking clear playmaking structure. Darmstadt's passing patterns showed no consistent hub, forcing players into ad-hoc distribution that opponents could disrupt without targeting specific individuals.

Notably, Darmstadt's individual player impact scores (measured by cohesion drop when removed) were *higher* than Leverkusen's, indicating fragility rather than depth. When Darmstadt's top player (Fabian Nürnberger) was unavailable, team cohesion dropped by 7.8%—compared to only 1.1% for Leverkusen's top player (Palacios). Elite teams distribute importance across multiple capable hubs; struggling teams concentrate fragility in irreplaceable individuals.

6 Discussion

6.1 Theoretical Implications

Our findings challenge the prevailing assumption that "balanced" passing networks optimize team per-

formance. Instead, we propose a revised framework:

The Hub Dependence Hypothesis: Elite teams benefit from *strategic centralization* that routes play through high-quality hub players. This centralization is not a vulnerability but an optimization—it maximizes the utilization of scarce elite playmaking talent.

This aligns with economic theories of specialization [Becker and Murphy, 1992]: just as firms benefit from division of labor, football teams benefit from allocating playmaking responsibility to specialists rather than distributing it equally.

6.2 Practical Applications

Our metric has several practical applications for coaching staff and analysts:

1. **Lineup Optimization:** The What-If simulator (removing players from networks) can identify which personnel maximize team cohesion.

2. **Opponent Analysis:** Identifying opponent hubs enables targeted pressing strategies to disrupt network flow.

3. **Transfer Evaluation:** Prospective signings can be evaluated for their potential to serve as complementary hubs within existing network structures.

4. **Youth Development:** Academies can identify players suited for hub roles based on passing profile compatibility.

6.3 Limitations

This study has several limitations that warrant acknowledgment:

Sample Size: With only $N = 18$ teams, our regression has limited statistical power and may be susceptible to overfitting despite cross-validation. The moderate test R^2 (0.539) reflects this constraint.

Single League/Season: Results are derived from one Bundesliga season. Generalizability to other leagues (with different tactical cultures) and seasons requires replication.

Confounding Variables: We do not control for team wage expenditure, transfer market value, or managerial tenure—factors that likely correlate with both network structure and performance. Wealthier teams may afford better hub players, inflating the Hub Dependence effect.

Causality: Our analysis is correlational. While we argue that Hub Dependence enables performance, it is possible that successful teams *become* more centralized as opponents cede possession, or that unmeasured factors drive both.

Network Construction: We use unweighted degree for centralization metrics. Alternative specifications (weighted degree, betweenness, eigenvector centrality) may yield different results.

7 Conclusion

This study provides statistical evidence that "balanced" passing networks are sub-optimal in elite football. Using 306 Bundesliga matches, we demonstrate that Hub Dependence—network centralization around key players—positively predicts season points ($\beta = 10.00$, $p < 0.001$). Combined with Connectivity, our two-factor Cohesion Metric explains 87.3% of variance in league standings.

The practical implication is clear: coaches should not strive for equal pass distribution, but rather design systems that maximize the utilization of specific, high-efficiency hub players. The "Dual-Hub" architecture observed in Bayer Leverkusen's undefeated campaign—combining a volume hub (Xhaka) with an attack hub (Wirtz)—provides a template for optimal network design.

Future work should extend this analysis across multiple leagues and seasons, incorporate player-level covariates, and explore causal identification strategies. Additionally, the interaction between Hub Dependence and opponent pressing intensity merits investigation—centralized networks may be more vulnerable to high-press systems, suggesting context-dependent optimal centralization levels.

Acknowledgments

Code Availability

Code and data are available at: https://github.com/amin8448/lineup-cohesion-hackathon

References

Becker, G. S. and Murphy, K. M. (1992). The division of labor, coordination costs, and knowledge. *The Quarterly Journal of Economics*, 107(4):1137–1160.

Buldú, J. M., Busquets, J., Echegoyen, I., and Seirul·lo, F. (2019). Defining a historic football team: Using network science to analyze Guardiola's F.C. Barcelona. *Scientific Reports*, 9(1):13602.

Clemente, F. M., Martins, F. M. L., and Mendes, R. S. (2016). *Social Network Analysis Applied to Team Sports Analysis*. Springer, Cham.

Decroos, T., Bransen, L., Van Haaren, J., and Davis, J. (2019). Actions speak louder than goals: Valuing player actions in soccer. In *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, pages 1851–1861.

Duch, J., Waitzman, J. S., and Amaral, L. A. N. (2010). Quantifying the performance of individual players in a team activity. *PLoS ONE*, 5(6):e10937.

Grund, T. U. (2012). Network structure and team performance: The case of English Premier League soccer teams. *Social Networks*, 34(4):682–690.

Karun, S. (2019). Introducing expected threat (xT). *Karun Singh Analytics Blog*.

López, M. J., Matthews, G., and Baumer, B. S. (2017). How often does the best team win? A unified approach to understanding randomness in North American sport. *The Annals of Applied Statistics*, 12(4):2483–2516.

Passos, P., Davids, K., Araújo, D., Paz, N., Minguéns, J., and Mendes, J. (2011). Networks as a novel tool for studying team ball sports as complex social systems. *Journal of Science and Medicine in Sport*, 14(2):170–176.

Peña, J. L. and Touchette, H. (2012). A network theory analysis of football strategies. *arXiv preprint arXiv:1206.6904*.

Watts, D. J. and Strogatz, S. H. (1998). Collective dynamics of 'small-world' networks. *Nature*, 393(6684):440–442.

Yamamoto, Y. and Yokoyama, K. (2011). Common and unique network dynamics in football games. *PLoS ONE*, 6(12):e29638.