

# Lineup Cohesion

A Network-Based Approach to Starting Eleven Optimization

---

**Amin Nabavi**

NEU Soccer Data Analytics Hackathon | February 2026

Bundesliga 2023/24 • IMPECT Open Data • 306 Matches

# Research Question

**Can we quantify team cohesion from passing networks to predict match outcomes and optimize lineups?**

## Stakeholder Value

Coaches need objective metrics beyond intuition for lineup decisions

## Network Science

Passing patterns encode team chemistry that box scores miss

## Actionable Insight

Identify key hubs, optimal pairings, and substitution strategies

# Cohesion Metric: Four Components

$$\text{Cohesion} = 0.50 \cdot \text{Connectivity} + 0.25 \cdot \text{Chemistry} + 0.15 \cdot \text{Hub} + 0.10 \cdot \text{Progression}$$

## Connectivity

50%

Network density + clustering coefficient

$r = +0.785^{***}$

## Chemistry

25%

Strength of midfield→attack connections

$r = +0.448^*$

## Hub Dependence

15%

Gini coefficient (star-player reliance)

$r = +0.714^{***}$

## Progression

10%

Pre-shot pass ratio in network edges

$r = +0.133$

*Weights empirically optimized from correlation with season points (n=18 teams)*

# Key Insight: The Hub Dependence Paradox

## Initial Expectation

Balanced teams (equal pass distribution) should perform better.

## Actual Finding

Star-dependent networks win more. Elite teams funnel play through hub players (Xhaka, Kimmich).

Original 'Balance' showed  $r = -0.714$  with points.

After inverting → Hub Dependence  $r = +0.714^{***}$

## Metric Improvement

Before (equal weights)

**$r = 0.314$**

$p = 0.204$  (not significant)

After (optimized weights)

**$r = 0.728$**

$p = 0.0006^{***}$  (highly significant)

**+132% improvement in predictive power**

# Validation: Season & Match Level

*[cohesion\_vs\_points.png]*

**Season Level: Cohesion vs Points**

*[cohesion\_by\_result.png]*

**Match Level: Cohesion by Result**

ANOVA:  $F = 36.64$ ,  $p < 0.0001$  — Significant difference between Win/Draw/Loss

# Case Study: Leverkusen's Undefeated Season

28W

6D

0L

90 pts

87 GF

24 GA

## Two Hub Types Discovered

### Granit Xhaka

Volume Hub — 558 passes with Palacios alone  
Recycles possession, dictates tempo

### Florian Wirtz

Attack Hub — Highest max betweenness (0.448)  
Wirtz→Boniface: 22 pre-shot passes (23.4%)

[leverkusen\_network.png]

**The Killer Chain: Xhaka → Wirtz → Boniface → GOAL**

# Application: Lineup Optimization

## Starting XI Selection

Score candidate lineups by predicted cohesion. Select players who maximize connectivity with existing starters.

## Substitution Strategy

Identify which subs preserve hub structure vs. which disrupt it. Avoid removing high-betweenness players.

## Transfer Targets

Project how a new signing integrates: simulate their historical passing patterns into the team's network.

## Opponent Analysis

Identify opponent's hub players to mark/isolate. Disrupting their Xhaka-equivalent drops cohesion.

# Conclusions

## 1. Network cohesion predicts season performance

$r = 0.728$  ( $p < 0.001$ ) — explains 53% of variance in points

## 2. Hub dependence > balance for elite teams

Star-player reliance is a feature, not a bug

## 3. Different hub types serve different functions

Khaka (volume) vs Wirtz (attack) — both essential

## Limitations

- Single season (n=18 teams)
- ~24% passes lack receiver ID
- Position data from metadata only
- No tracking data (spatial context)

Future: Multi-season validation,  
player-level predictions