Does Teamwork Really Make the Dreamwork?

Final Project Report

May Mishima

2025-12-14

Introduction

For this project, I aimed to quantify a concept that is frequently referenced but not easily measured. While "teamwork makes the dream work" is a common phrase in sports culture, there is little evidence that directly evaluates its validity. This study examines the relationship between team chemistry and team success to assess whether teamwork is significantly associated with performance.

Baseball provides a strong context for studying the impact that teamwork has on team success because many aspects of the game depend on coordination and familiarity. Defensive plays such as double plays, relay throws, and player positioning require players to anticipate each other's movements and tendencies. Offensive and strategic decisions, including base running and situational hitting, also benefit from shared experience and communication. Although some individual players can influence team outcomes, baseball is ultimately a team sport in which success depends on team culture and player familiarity.

Teams also adopt different roster building philosophies to achieve success. Some organizations trade for established stars, resulting in higher roster turnover, while others emphasize developing their Minor League players creating "homegrown" talent, creating more stable rosters. This project can help analyze each team's approach in building rosters, but also help answer the question of if more interconnectedness among players translate into stronger team performance.

With these motivations, this project analyzes the social networks of each MLB team's roster. Team chemistry is measured through global transitivity, a measure of how interconnected each roster is. By comparing team transitivity scores to average winning percentages across seasons, this study explores whether teams with more cohesive rosters tend to perform better than those with less cohesive rosters. Understanding this relationship provides insight into how roster construction and player continuity may contribute to success in the MLB.

Data

The data used in this study was obtained from Baseball Reference, a publicly accessible baseball statistics database used by researchers, analysts, and fans. Baseball Reference compiles its data from multiple sources, including Sportsradar used during games to track live stats, and archival databases such as the Lahman Baseball Database, which preserves historical seasonal records (Sports Reference). These data are collected continuously and finalized at the conclusion of each Major League Baseball season.

This specific study focused on the 2015-2025 seasons, in order to ensure ample variability in winning percentages and roster continuity without having too much season to season noise. Player data were drawn from Wins Above Replacement (WAR) statistics for both position players and pitchers. WAR was used to identify rosters because any player appearing in at least one game is included, reflecting the idea that even brief roster members contribute to team dynamics both on the field and in the clubhouse. Player records include player name, team, and season; players traded mid-season were recorded separately for each team they appeared on.

Team success was measured using winning percentage, calculated as total wins divided by total games played for each season. Average winning percentage across study period was computed for each team to produce a stable measure of long term success. Team names were standardized across datasets to account for franchise name changes and ensure consistency between player and team data.

Pitcher and position player WAR data were cleaned to retain player name, team, and season. These datasets were combined and were adjusted for franchise name changes. Team winning percentage data were similarly cleaned, and average winning percentage across 2015-2025 was calculated for each team. The player and team datasets were then merged using consistent team abbreviations for analysis. All setup code individual MLB network plots can be found in my wrangling file. (finalprojectwrangling.qmd)

```r
#read in dataset
networks <- readRDS("networks.rds")
transitivity <- readRDS("transitivity.rds") |>
  mutate(
    Transitivity = as.numeric(Transitivity)
  )
slr_data <- readRDS("slr_data.rds")
```

Methods

To quantify team chemistry, this project uses global transitivity, a network measure that captures how interconnected a team's roster is. In the network, each player is represented as a node, and connections between players indicate seasons played together Transitivity reflects the likelihood that if Player A has played with Player B and Player C, then Player B and Player C have also played together (Csardi, 2025). In this context, higher transitivity includes tighter clustering and greater roster cohesion.

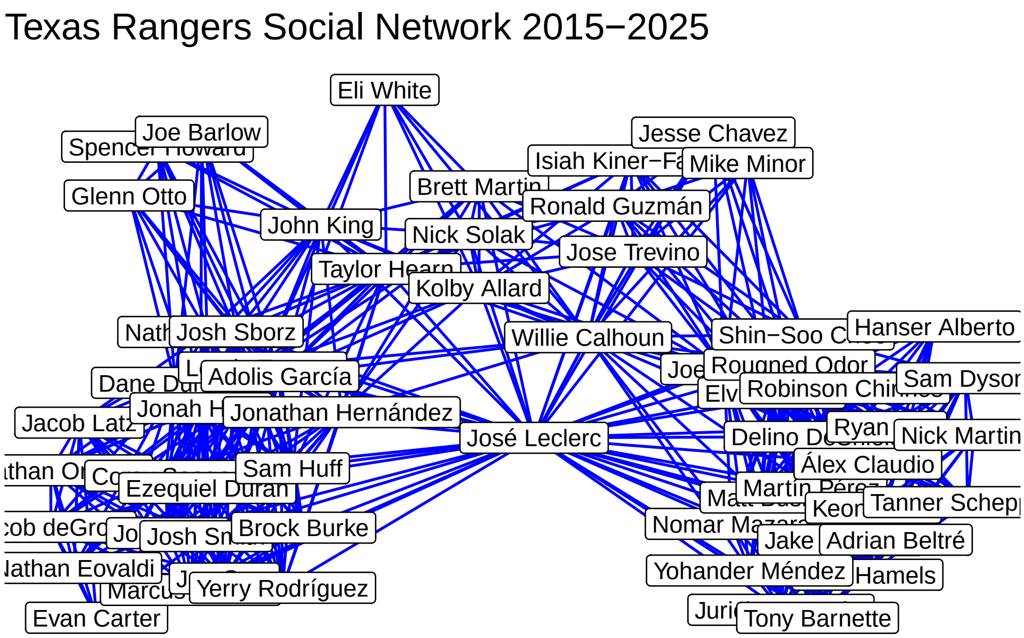
Although players on the same roster might appear connected by definition, this assumption does not hold when multiple seasons are combined. Mid-season trades, short-term call-ups, and roster turnover can produce many player pairs that never overlap directly, resulting in incomplete triads. Global transitivity accounts for this by measuring the proportion of closed triangles relative to all connected triples in the network. This study uses global rather than local transitivity because a single summary value was needed to represent each team's overall level of interconnectedness.

Team networks were constructed by identifying all overlapping player pairs within each team across the 2015–2025 seasons. An edge was defined between two players if they played together for at least three seasons. This threshold was chosen to exclude brief or incidental overlap, such as partial seasons resulting from trades, which are unlikely to reflect meaningful familiarity. Requiring three shared seasons identifies more stable relationships while maintaining sufficient network density for comparison across teams. Nodes represent individual players, and edges represent sustained shared tenure.

Using these networks, global transitivity was calculated for each MLB team. These transitivity scores were then paired with each team's average winning percentage across the study period. A simple regression model was used to explore whether greater roster interconnectedness was associated with higher long-term team success. All 30 MLB networks along with the transitivity table and the scatterplot are compiled in this Shiny App: https://mmishima28.shinyapps.io/MLBSocialNetworks/.

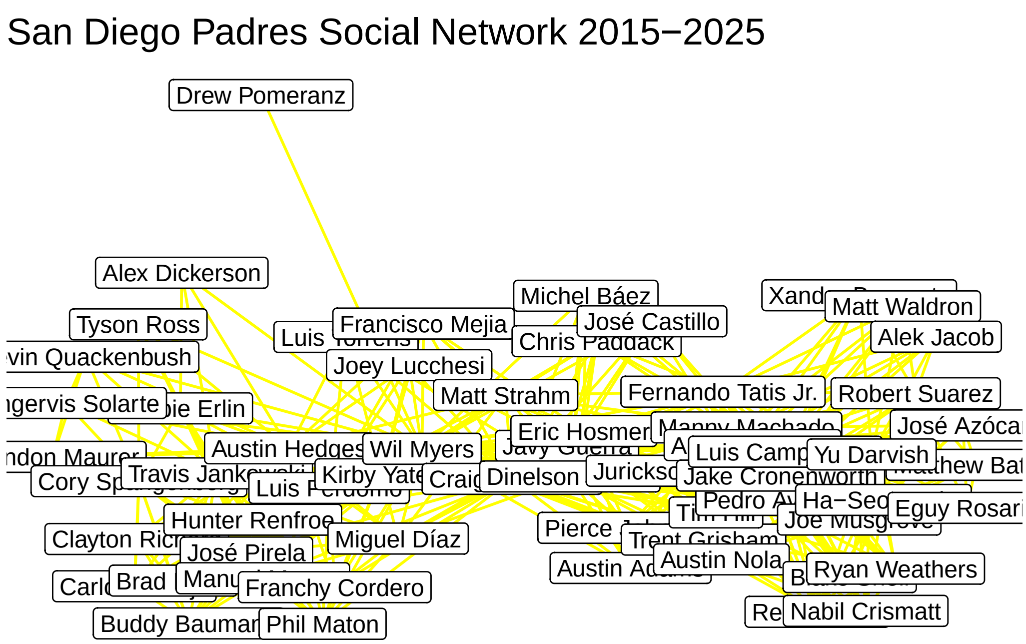
Results / Discussion

```r
#TEX network
#pull Rangers data from RDS
tex_net <- ggnetwork(networks[["TEX"]])

#build network
ggplot(tex_net, aes(x, y, xend = xend, yend = yend)) +
  geom_edges(color = "blue") +
  geom_nodes(size = 3) +
  geom_nodelabel(aes(label = name), size = 3) +
  theme_blank() +
  ggtitle("Texas Rangers Social Network 2015-2025")
```



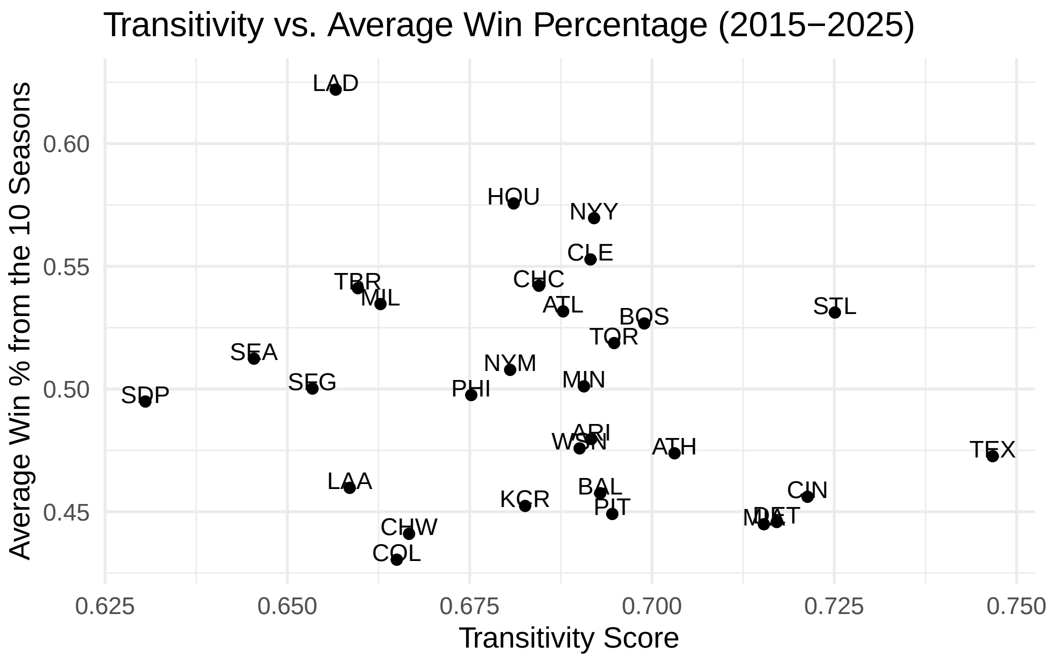
Among all teams, the Texas Rangers exhibited the highest transitivity score (0.75), meaning that approximately three-quarters of all connected player triples formed closed triangles. Based on the network, the largest hub is Jose Leclerc, connecting players from earlier seasons (2015-2019) to those from later seasons (2020-2025). This reflects his long career (2016-2024) as a reliever for the Rangers. The Rangers experienced some postseason success in 2015-2016, but went into a rebuilding phase from 2017-2022, peaked in 2023 winning the World Series, and have missed the postseason ever since. Overall, they are a team with the highest transitivity and a below average team performance.

```r
#SDP network
#pull Padres data from RDS
tex_net <- ggnetwork(networks[["SDP"]])

#network building
ggplot(tex_net, aes(x, y, xend = xend, yend = yend)) +
  geom_edges(color = "yellow") +
  geom_nodes(size = 3) +
  geom_nodelabel(aes(label = name), size = 3) +
  theme_blank() +
  ggtitle("San Diego Padres Social Network 2015-2025")
```



The San Diego Padres have the lowest transitivity score at 0.63, meaning that among all the potential traingles of connected players, 63% of the triangles are complete. The Padres are known for a roster builing philosophy of trading away budding stars such as Juan Soto and acquiring big name stars such as Xander Bogaerts. This leads to more rotational rosters that therefore, lead to a lower transitivity. For the Padres network, we see that the left cluster which represents the building phase of the Padres (2015-2019), the middle represents the Padres most successful era, just shy of a world championship (2020-2022), and the right represents the current Padres who have had two straight postseason appearances. Despite their lower transitivity, the Padres maintained approximately average performance across the study period, including multiple postseason appearances.

```
#scatterplot to show relationship
ggplot(slr_data, aes(x = Transitivity, y = AvgWinPct, label = Team)) +
  geom_point()  +
  geom_text(nudge_y = 0.003, size = 3) +
  theme_minimal() +
  labs(
    title = "Transitivity vs. Average Win Percentage (2015-2025)",
    x = "Transitivity Score",
    y = "Average Win % from the 10 Seasons"
  )
```




```
#transitivity score table
transitivity |>
  arrange(desc(Transitivity)) |>
  kable(
    caption = "Team Transitivity Scores 2015-2025",
    booktabs = TRUE
  )
```

Table 1: Team Transitivity Scores 2015–2025

Team	Transitivity
TEX	0.7467249
STL	0.7250806
CIN	0.7213327
DET	0.7171070
MIA	0.7153488
ATH	0.7030928
BOS	0.6989435
TOR	0.6948120
PIT	0.6945594
BAL	0.6928982
NYY	0.6920599
ARI	0.6916697

Team	Transitivity
CLE	0.6915636
MIN	0.6906720
WSN	0.6900754
ATL	0.6878162
CHC	0.6845020
KCR	0.6826061
HOU	0.6810345
NYM	0.6805505
PHI	0.6752159
CHW	0.6666881
COL	0.6650041
MIL	0.6627702
TBR	0.6596733
LAA	0.6585480
LAD	0.6566315
SFG	0.6534442
SEA	0.6454170
SDP	0.6305270

The analysis found no strong relationship between team transitivity and winning percentage, meaning that team chemistry is not a telling factor for team success. The scatterplot comparing transitivity scores to average winning percentage from 2015-2025 produced a correlation coefficient of -0.2 with a a r^2 of 0.04, meaning that only 4% of the variability in average win percentage is accounted for in the transitivity score.

It is also notable that all 30 MLB teams had transitivity scores above 0.6. This indicates that, across the league, a majority of connected player triples formed closed triangles. Given MLB's 26-player active rosters and the repeated overlap of core players across seasons, this level of interconnectedness is expected. Overall, these findings suggest that while MLB teams generally exhibit high internal cohesion, differences in transitivity alone do not strongly explain differences in team success.

Conclusion

Overall, this study found no strong relationship between team transitivity and team performance for the 2015–2025 MLB seasons. Although all teams exhibited relatively high levels of interconnectedness, variation in transitivity explained very little of the variation in average winning percentage. This suggests that while roster cohesion is common across MLB teams, differences in transitivity alone are not sufficient to predict long-term team success.

Because this study is an observational archival analysis, the findings are limited to the population examined: MLB players on active rosters during the 2015–2025 seasons. The results cannot be generalized beyond professional baseball or interpreted as evidence of causation. However, the data are useful for identifying broad patterns across multiple seasons, particularly in a sport where roster construction strategies and gameplay have evolved substantially over time. Future research could examine whether these patterns differ across shorter or earlier time periods.

One limitation of this study is that the network visualizations did not strongly differentiate teams with higher and lower transitivity scores. Many networks appeared similarly dense, likely because the overall range of transitivity values across teams was relatively narrow (approximately 0.12). This limited variability may have reduced the ability to visually distinguish levels of interconnectedness. Analyzing a smaller time window or focusing on periods of major roster change could increase variability and improve interpretability.

Another limitation is that transitivity does not fully capture team chemistry. Players may share long tenure on a team without meaningful interaction, particularly if they play different positions or occupy distinct roles. Additionally, relief pitchers and closers often emerged as major hubs in the networks due to long careers and frequent movement, despite spending limited time interacting with position players during games. This may inflate transitivity scores without accurately reflecting interpersonal cohesion. A potential improvement would be to construct position-specific networks (e.g., infielders, outfielders, starting pitchers) and examine how cohesion within these subgroups contributes to overall team structure.

Finally, a promising extension of this work would be to incorporate minor league affiliate data into the network analysis. Many players form strong interpersonal bonds during their development in the minor leagues, and these early connections may persist throughout their professional careers. Including minor league rosters could provide a more comprehensive picture of player interconnectedness and offer deeper insight into how long-term relationships shape MLB team dynamics.

References

Csardi, G. (2025). *Igraph R Manual Pages*. igraph. https://igraph.org/r/html/1.2.5/transitivity.html#:~:text=Transitivity%20measures%20the%20probability%20that,also%20called%20the%20clustering%20coefficient. (Accessed December 13).

Nykamp, DQ. "Definition of the transitivity of a graph." From *Math Insight*. http://mathinsight.org/definition/transitivity_graph (Accessed December 13).

Sports Reference LLC. Baseball-Reference.com - Major League Statistics and Information. https://www.baseball-reference.com/. (Accessed December 13).