The Power of Proximity: Using Network Science to Analyze the Link Between Coaching Communities and Team Success in the NFL

Atul Venkatesh, Siddharth Singh, and Charlie Betts

Background

The National Football League has continued to place a premium emphasis on coaching. While anecdotal evidence reveals the impact of a cohesive coaching unit on team success, little research has been done on optimizing a coaching community. With the help of *ProFootballHistory.com*, we constructed a **comprehensive coaching dataset** of head coaches, coordinators, positional coaches, and quality control coaches. We used this data to create modern coaching communities and identify relations across the league, calculating the effect of coaching chemistry on team success.

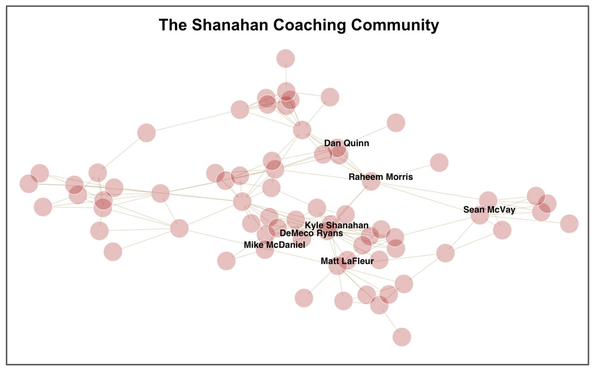
NFL Separation Theory

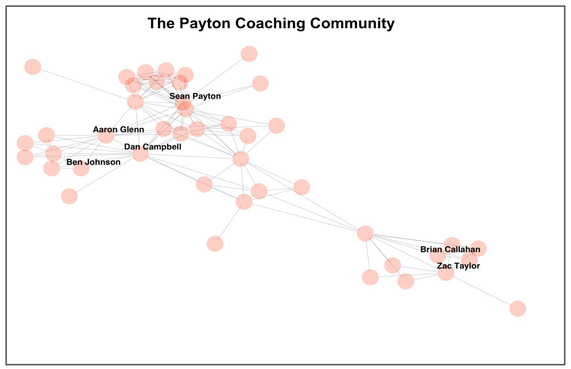
The Six Degrees of Separation theory posits that all people are six or fewer social connections away from each other. We examined this theory among NFL coaches using Kyle Shanahan as the center. Each coach's distance from Shanahan was calculated using a distance formula: $D = 1/(\alpha + \gamma)$, where α is the aggregated closeness score[1] and γ is years coached together. A lower distance score indicates a closer relationship.

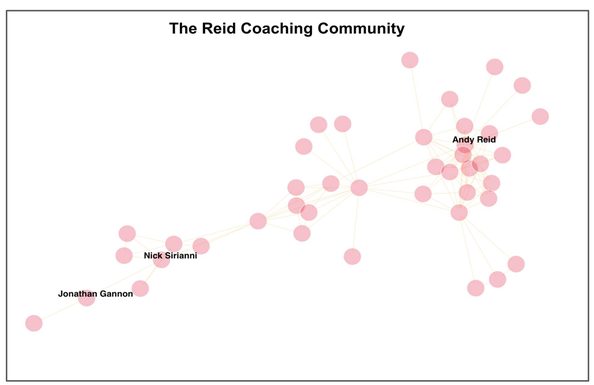
	Coach Name	Path	Score
1	Matt LaFleur	Kyle Shanahan - Matt LaFleur	0.18
2	DeMeco Ryans	Kyle Shanahan - DeMeco Ryans	0.213
3	Mike McDaniel	Kyle Shanahan - Mike McDaniel	0.213
4	Raheem Morris	Kyle Shanahan - Raheem Morris	0.27
5	Sean McVay	Kyle Shanahan - Bobby Turner - Sean McVay	0.35
6	Mike Tomlin	Kyle Shanahan - Danny Smith - Mike Tomlin	0.444
7	Pete Carroll	Kyle Shanahan - Brian Schneider - Pete Carroll	0.448
8	Shane Steichen	Kyle Shanahan - Robert Saleh - Gus Bradley - Shane Steichen	0.658
9	Kevin O'Connell	Kyle Shanahan - Bobby Turner - Sean McVay - Wes Phillips - Kevin O'Connell	0.682
10	Sean McDermott	Kyle Shanahan - Matt LaFleur - Ben Sirmans - Rob Boras - Sean McDermott	0.765
11	Dan Campbell	Kyle Shanahan - John Benton - Darren Rizzi - Dan Campbell	0.781
12	Mike Vrabel	Kyle Shanahan - Jim Haslett - Mike Vrabel	0.8
13	Andy Reid	Kyle Shanahan - Jon Embree - Joe Cullen - Andy Heck - Andy Reid	0.817
14	Mike Macdonald	Kyle Shanahan - Jon Embree - Joe Cullen - Mike Macdonald	0.82

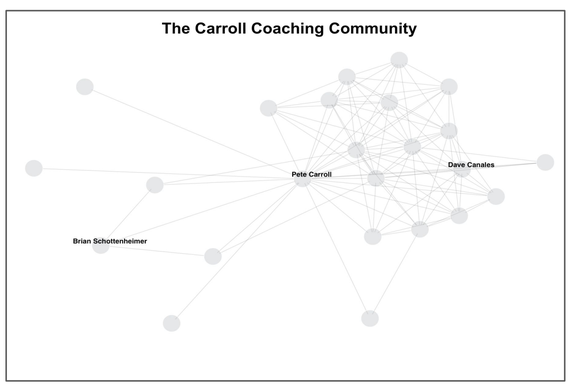
Coaching Communities

Numerical values were assigned to each coaching role to mirror their level of authority and impact. Head coaches were given the largest value of 10 while quality control coaches were given the lowest value of 2.5. A connection between two coaches would occur if both coaches coached for the same team in the same year. In a given year, the connection strength was the sum of the two values. To weigh this tie over the course of a coaching career, an **aggregated closeness score**[1] was created. For the purpose of creating communities, the minimum connection strength was 50. With this, the number of significant coaches shrunk from 6695 to 552. The number of connections shrunk from 44435 to 1877. The network was split into distinct coaching communities using the fast-greedy modularity-maximizing algorithm. The algorithm produced 19 communities. Each community was qualitatively evaluated based on how many important names were in the community and whether there was a shared coaching philosophy amongst coaches. The 14 significant communities are shown below.









Other coaching communities: **Belichick community** (1 active HC), **John Harbaugh community** (2 active HC), **Fangio community** (1 active HC), **Mike McCarthy community** (1 active HC), **Rivera community** (2 active HC), **Tomlin community** (1 active HC), **Arians community** (1 active HC), **Garrett Gruden community** (0 active HC), **Lewis Zimmer community** (2 active HC), **Reich community** (1 active HC)

Coaching Staff Cohesion

Does coaching staff closeness impact team success? To examine this, we created three regression models. For all models, **team closeness score**[2] was the independent variable. Strength of schedule and the previous year's win percentage were constants. Model 1 was a linear regression predicting win percentage. Models 2 and 3 were logistic regressions predicting if a team made the playoffs or won their division.

	Estimate	Std. Error	t Value	Pr(> |t|)
Coefficients				
Intercept	1.6557	0.103874	15.940	<2.0e-16***
Closeness Score	0.0038	0.0004558	8.402	5.15e-13***
Lag Win %	0.2309	0.0369745	6.245	9.43e-10***
Strength of Schedule	-2.8151	0.2014151	-13.977	<2.0e-16***

Notes: ***/**/* denotes significance at the 5/1/.1 percent.
Residual standard error: 0.1392 on 474 degrees of freedom
Multiple R-squared: 0.473, Adjusted R-squared: 0.4697
F-statistic: 141.8 on 3 and 474 DF, p-value: < 2e-16

An increase of 1 in team closeness score is associated with a **0.0038 percentage point increase in seasonal win percentage** (winning *7 percent* of a game)

	Estimate	Std. Error	z Value	Pr(> |z|)
Coefficients				
Intercept	13.3461	2.0616	6.474	$< 2.0 \times 10^{-16}$***
Closeness Score	0.0525	0.0092	5.726	1.03×10^{-7}***
Lag Win %	2.3590	0.6689	3.527	4.21e-4***
Strength of Schedule	-33.9313	4.3035	-7.885	$< 2.0 \times 10^{-16}$***

Notes: ***/**/* denotes significance at the 5/1/.1 percent.
Null deviance: 642.41 on 477 degrees of freedom
Residual deviance: 472.73 on 474 degrees of freedom
AIC: 480.73

An increase of 1 in Team Closeness Score leads to a **1.05% Increase in Playoff Probability**

	Estimate	Std. Error	z Value	Pr(> |z|)
Coefficients				
Intercept	13.3461	2.0616	6.474	$< 2.0 \times 10^{-16}$***
Closeness Score	0.0525	0.0092	5.726	1.03×10^{-7}***
Lag Win %	2.3590	0.6689	3.527	4.21e-4***
Strength of Schedule	-33.9313	4.3035	-7.885	$< 2.0 \times 10^{-16}$***

Notes: ***/**/* denotes significance at the 5/1/.1 percent.
Null deviance: 642.41 on 477 degrees of freedom
Residual deviance: 472.73 on 474 degrees of freedom
AIC: 480.73

An increase of 1 in Team Closeness Score leads to a **1.03% Increase in Division Winning Probability**

Limitations and Future Directions

Limitations: 1) Data was from 2010-2024 2) Equivalent connections might not accurately reflect relationship strength.

Future: 1) Investigate team-specific hiring preferences 2) incorporate college coaching networks

[1] Aggregated closeness score: Cumulative sum of a coaching pair's connection rating

[2] Team closeness score: Average aggregated closeness score + years coached together between all coaches on a given team