

Team 34: Structural Learning in Collegiate Football Game Matchups

Michael Dacus

MWDACUS@STANFORD.EDU

Tyler Weiss

TWEISS19@STANFORD.EDU

Ian HokaJ

IANHOKAJ@STANFORD.EDU

AA228/CS238, Stanford University

1. Introduction

Collegiate Football has been one of the most popular sports in United States for the past century, and considered a personal favorite to many, not in the game itself but also predicting the matchup and outcome between two teams. Individual metrics (wins to losses, player statistics, etc.) from current and previous seasons might determine how a team will perform in upcoming games, but there are many issues that arise from these metrics. In determining the outcome of a future matchup, some metrics may not be as important as others, or vice versa. This is a current issue in combined metrics (AP Poll, CFP Poll, strength of schedule), which can be seen when determining how teams should ultimately be ranked against one another. Currently the College Football Playoff (CFP) Committee determines ranking by subjective overview of individual metrics. Depending on how important each metric is, they will assign a weight (again determined subjectively) to that value.

Another issue that arises is the dependencies that may be related in between individual metrics. This is an incredibly important since – for example – wins and losses may be based on who is on the field (injuries) or the environment that may be playing in. How a metric is determined based on other given variables (conditional probability) is essential in determining how a team might eventually play out in the rest of the season, and should be taken into consideration.

This project seeks to use a Bayesian network approach to determine which team will win in a certain game during the season, based on the aforementioned (and other) variables. From this, we intend to implement the Bayesian network model and compare it to the actual outcome of games during the 2021 College Football regular season to determine model robustness and accuracy.

2. Variables Involved

Like many multi-domain problems, the outcome of a football game based on two teams consists of many variables and layers to the network structure. A core set of variables (so called “individual metrics”) will focus on discrete variables of high level team statistics. These variables include Wins-and-Losses (WL’s), player availability, coaching, depth chart, location of games, and others. Another set of variables will be dedicated to performance statistics, like number of points scored, rushing and passing yards, defensive metrics, and others. This project will focus more on team metrics, then add performance statistics for higher fidelity.

3. Decision Making Involved

As briefly mentioned in 1.Introduction, the motivation behind this structural network is to determine the outcome of Collegiate Football matchups based on individual metrics and the effect it has on other variables. Deciding the outcome of game is a form of decision making, but it cannot be solely based on rankings (which are currently subjective). Given the uncertainty within College Football rankings and other combined metrics, decisions on the outcomes themselves can be made through probabilistic reasoning of our set of variables.