Exploring the NFL

Alison Smith  <alison dot smith dot m at gmail dot com>

Introduction
The National Football league began in 1920 with 12 teams and has grown to 32 teams broken into 2 leagues and 8 divisions. In this time, a total of 14,657 games have been played [1]. Current scheduling requirements lead to a particular distribution of matchups, which when represented as an adjacency matrix should lead to a clustering representative of the league structure. In order to analyze the network and produce the clustering, I used NodeXL, an “open-source template for Microsoft® Excel® 2007 and 2010 that makes it easy to explore network graphs.” [3]

Data
In order to produce the adjacency matrix, I needed the total count of games played between each team in the NFL. I was able to find an online database which provided the ability to query by pairs of teams [1]. Filling in the symmetric 32 x 32 adjacency matrix by hand would require querying the database 528 times. A first idea was to perform a scraping of the website to programmatically amass a map of the matchup counts and then write this to csv file as an adjacency matrix. I instead decided to make use of the Mechanical Turk [2] to accumulate the entries. I broke the tasks down by teams and filled in the adjacency matrix with the results from the completed assignments.

Analysis

Overview
After initial import of the adjacency matrix into NodeXL, groups were created for each of the current divisions and the nodes were colored and shaped appropriately. The first attempt at visualization produced a visually intriguing representation of the obvious fact, “all teams have played each other at least once.”
The Fruchterman-Reingold algorithm was used to layout the network, and the visualization was annotated with the conference structure, divisions prior to 2002, and the current division structure.

**Conferences**

The National and American Conferences were formed in 1966 [5]. The algorithm separated teams from the two conferences, pushing the NFC teams to the left of the visualization and the AFC teams to the right.
Divisions prior to 2002

Prior to 2002 restructuring, the NFL was made up of 6 divisions comprised of 31 total teams. The clustering algorithm aligned the teams of these divisions together quite accurately.
Current Divisions

The restructuring and the addition of the Houston Texans in 2002, led to a football league with 8 divisions comprised of 32 total teams. This lead to some overlapping clusters in our layout. The effect is strongest where new divisions were formed or existing divisions were changed drastically, which is the case for the AFC South and NFC West, respectively.
Headlines

Newer teams are the “outliers” of the network

The clustering of the network graph pushed a number of “outlier” nodes to the edges. Upon further review, it appears that these outliers represent newer teams to the league. In 1976, the Seattle Seahawks and Tampa Bay Buccaneers joined the league. The Carolina Panthers, Baltimore Ravens, and Jacksonville Jaguars joined in 1995 and 1996, followed by the newest team, the Houston Texans, which joined in 2002 [7].
In most cases, teams that have played more games against each other are considered “rivals.”

In order to quickly visualize the teams that have played many games against each other, I correlated the edge widths of the network graph to the edge weights, and I filtered to show edges for only those teams that have played against each other greater than 90 times. I then took the top 7 of the top 12 rivalries in the NFL as outlined in [4] and labeled them on the graph.
Of these 7 rivalries, 6 are between teams that have played greater than 90 games against each other, and all of the rivalries are within the current divisions. The second ranked rivalry is between the Green Bay Packers and the Chicago Bears, who have played against each other a record 184 times.
Two teams, the Arizona Cardinals and the Green Bay Packers, bridge divisions. After filtering out edges for those pairs of teams with less than 90 games played against each other, two interesting teams appeared.
The Arizona Cardinals team is clustered with the NFC East, which it was a member of prior to the restructuring, but also holds a strong tie to the Chicago Bears. This is due to the fact that from 1920 - 1943, the Arizona Cardinals were actually the Chicago Cardinals and played the Chicago Bears 1-2 times every year [1].

The Green Bay Packers team is clustered with the NFC North, which is it's current division, but also has a strong tie to the St. Louis Rams. The Rams joined the league in 1937 as the Cleveland Rams. The Packers and Rams shared the NFL until the AFL merge usually playing at least 2 games against each other per year. The two then shared the NFC until it was split into divisions and eventually the restructure occurred in 2002 placing the teams in different divisions [5].

Software Critique
NodeXL provided a useful platform for performing network analysis based off of Excel. It would be a nice perk if NodeXL were compatible with other operating systems, but as I have a Dell laptop, this wasn't a problem for me.
I was able to easily import the adjacency matrix from another workbook, create groups, format the nodes, and experiment with a number of different layouts. Although the Fruchterman-Reingold algorithm worked well for my dataset, it would have been nice to have a variety of clustering algorithms to choose from.

Although I chose to use a hand-created adjacency matrix for my dataset, it is nice that NodeXL supports importing from several social media applications - Twitter, Flickr, etc; however, when I attempted to use the Twitter importer I was immediately hit with rate limit issues. I’m unsure if there’s a feasible solution to rate limiting (especially since this is free software), but providing users of an estimate of how long it will take (especially given the rate limit lockout time of 1 hour) to import a network for a particular search/user would be extremely useful.

References