

IS620 Final Project Proposal

Analyzing Social Networks, Alliances, and Winning Positions in *Survivor*

Brian Chu, Aaron Palumbo - December 2015

Background

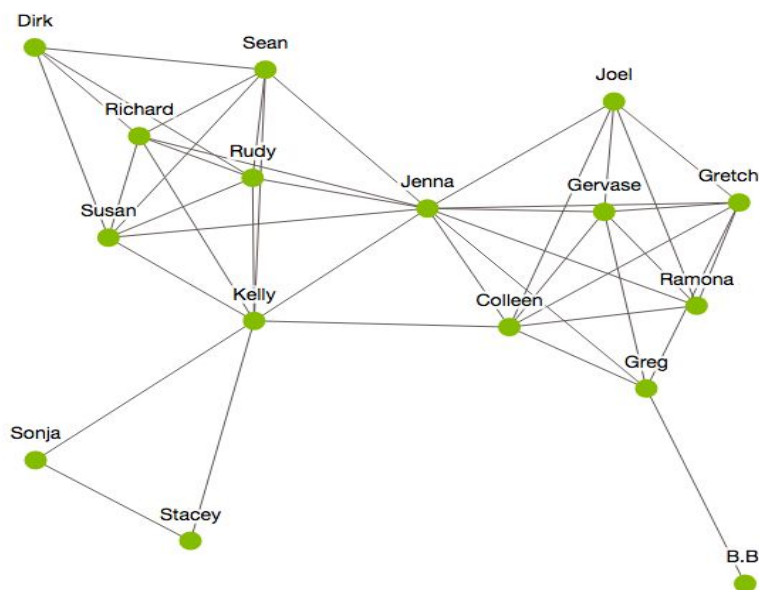
Survivor is a reality game show where 16-20 contestants are put in a remote location to compete in challenges and effectively vote each other off one-by-one until a single winner remains. On the surface, the concept of Survivor seems completely unrelated to web analytics. At its core, however, Survivor is effectively a strategic game where success is often determined by building alliances and connections with other players. Therefore, Survivor can be viewed as a social network that actually exemplifies many of the key concepts studied in this course.

In this project, we aim to use social network analysis to answer some of the following:

- *What do social networks in Survivor look like (shape, centrality)?*
- *Is the general gameplay and alliance strategy of Survivor changing over time?*
- *What is the optimal winning strategy or position in the network?*
- *Can winners be accurately predicted?*
- *Who were the best players of all time?*

Basic setup

The network will consist of a graph with weighted edges. Each contestant will represent a node on the graph. Edges represent two contestants who vote for the same person in the same vote, regardless of the vote outcome. The edge weight will correspond to the number of times both players voted the same way. For example, if Kelly and Rudy both voted for Greg in the same 3 episodes, they will have an edge with a weight of 3. If Kelly and Susan both voted for Greg, Joel, and Colleen in each of those episodes, they would also have an edge weight of 3.



Analysis

Survivor has completed 30 seasons. Since each season has a different cast, we will draw the social network and evaluate each one separately, but use the same analysis template throughout. This will allow us to compare social networks of different seasons and see whether there are any significant longitudinal trends. Here we mostly look at descriptive characteristics of the network such as shape and centrality.

We will also place an emphasis on the winners and high-finishers of each season. This is partly because they have more data available, but we are also interested in determining patterns and features of winning strategies and whether those are consistent over time. We, therefore, aim to train a predictive model to see if the winners and/or order of finish can be predicted with any accuracy for each season.

Potential challenges, considerations

- Edge weight calculation
 - Normalizing edge weights (e.g. # of voting chances, division of teams)
 - Include votes against in the edge-weight calculation (probably as a negative)?
 - Add weight for voting for the eliminated player
- Accounting for (or not knowing about) specific game or season dynamics

Data sources

The contestants and results for each season will be scraped from <http://survivor.wikia.com>

Analysis tools

- Python will be primarily used for analyzing the data (NetworkX, Scikit-learn)
- R may be used for some data transformations or cleaning
- Graph visualizations will use some mix of NetworkX, iGraph, and Gephi

Roles

This will be a collaborative effort with tight communication throughout.

- Idea: The idea for the project was Brian's.
- Exploratory Analysis: Aaron did the initial web scraping work. We will both likely do some amount of exploratory analysis to dial in the strategy.
- Detailed Analysis: We will probably split up the seasons to clean up and finalize results.
- Summary: This will mainly be collaborative, splitting up the work as necessary.

More game-specific rules and details: [http://survivor.wikia.com/wiki/Survivor_\(franchise\)](http://survivor.wikia.com/wiki/Survivor_(franchise))

Github repository: <https://github.com/bchugit/Survivor-Project>