Author: Andrey Risukhin
Mentor: Apara Venkat
SPA DRP Spring 2021

Networks and Choice Modeling

Motivation

People must make decisions every day, calculating tradeoffs of their choices. A rational decision maker is interested in maximizing the usefulness ("utility") they extract from a choice. Through this project, we examine three algorithmic approaches to determine which option a choice-maker should pursue, in the context of a university deciding where to hire faculty from.

Approaches

Multinomial Logit

The utility of an item can be decomposed into two elements: observable utility and unknown utility. The unknown utility cannot be measured by researchers, so they make assumptions about its distributions. The most popular assumption is that unknown utility follows "independently, identically distributed extreme value". This is a popular assumption because it allows the Logit model to be derived, which has nice properties such as a log(odds of winning) interpretation.

Because universities may hire faculty from more than one university, we used the Multinomial Logit to fit to the data. The Logit model depended on features available about the data to train on, so data problems in machine learning such as missing data and irrelevant information are important to consider. The rankings were consistent with the US News journal rankings.

PageRank

PageRank is a well-known network ranking algorithm, famously attributed to Google. It compares nodes in a directed network based on their inbound and outbound connections. An interpretation is of a webcrawler teleporting between pages. All pages have a nonzero probability of being teleported to, and the probability is affected by inbound and outbound links.

By modeling the hiring behavior between universities as university v hiring from university u, we frame our data as a directed network. We then applied PageRank to this network. Due to the linear algebra performed to find the eigenvalues, the rankings appear strongly confident about the top university, with decreasing confidence (difference in rank) for every additional university. These rankings are not interpretable, and PageRank did not rely on any features to reach these rankings (which seemed consistent with the US News journal rankings).

SpringRank

SpringRank determines the ranking of nodes in a directed network by modeling each edge as a directed spring, with some potential energy function, and minimizing the potential energy over the entire network by choosing positions (ranks) for the nodes.

By repeating the representation of hires as directed edges from hirer to faculty, we modeled the data as a directed network. Applying SpringRank yielded rankings that were more evenly spaced than PageRank. There was also a set of universities whose US News rank was unknown, and SpringRank assigned more separated ranks to these than the other models (capturing more information about them based on their connections to other nodes and ranks withing the network). Without relying on features, SpringRank generates a ranking interpretable as the log(odds of winning), where "winning" is the direction of an edge between two universities.

Author: Andrey Risukhin
Mentor: Apara Venkat
SPA DRP Spring 2021

Conclusion

I would recommend using SpringRank for ranking these universities, because it outputs interpretable rankings without relying on features about the nodes in a network. These rankings can assist universities in determining which universities are most prestigious, and therefore most worth hiring from.

Resources

The slides presented on June 8, 2021 and the code used in this project will be published to my github: https://github.com/AndreyRisukhin/networks-choice-modeling

Acknowledgements

I am grateful to my mentor Apara for helping me learn about approaches to solve ranking and choice making problems; to the Statistics and Probability Association for hosting the Directed Reading Program; and to the Larremore Lab for creating the code I used for SpringRank.

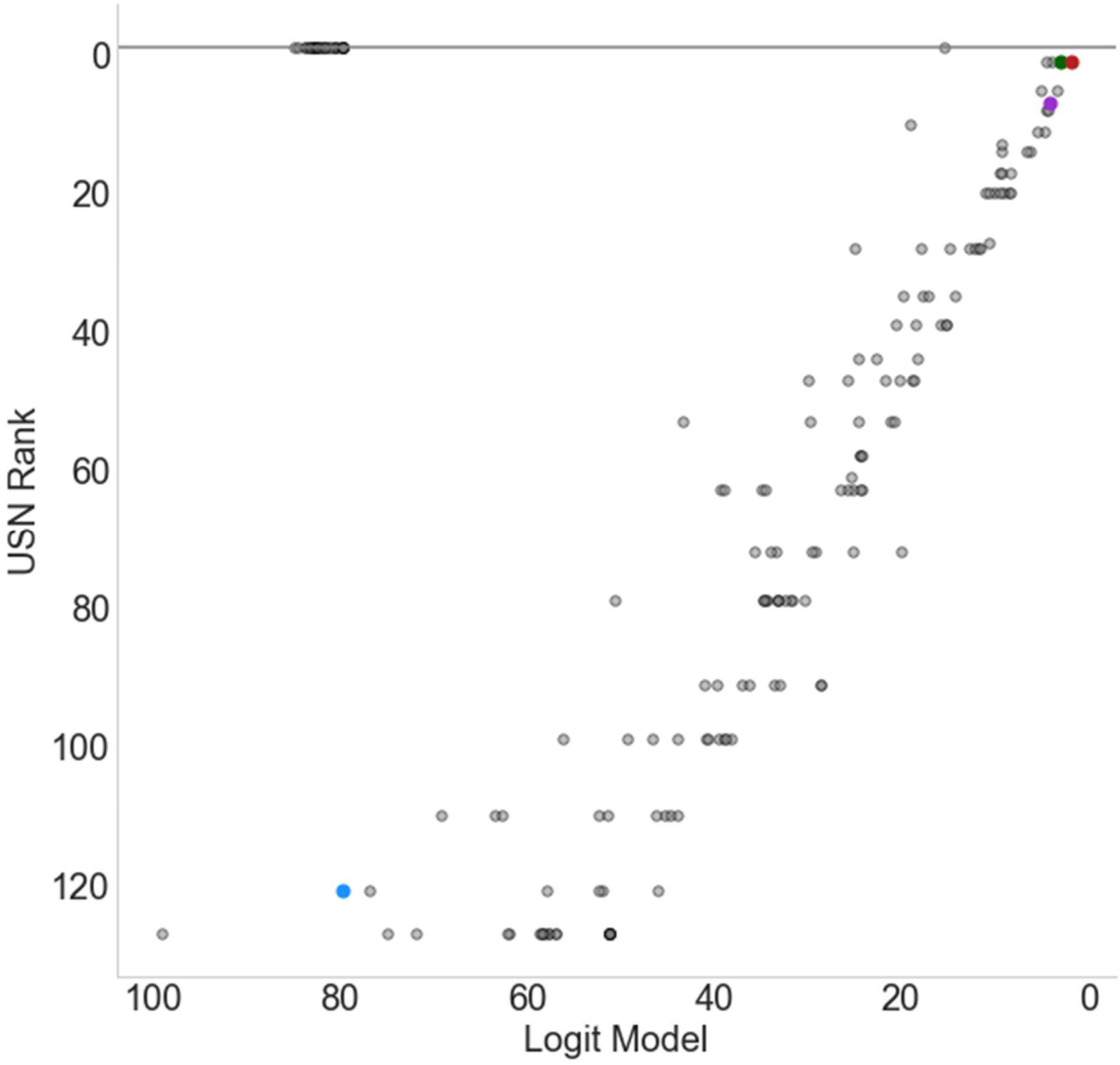
Plots

Below are the plots of rankings generated by each approach, compared to the US News ranks. Note: it is unlikely that the US News ranks are the ground truth ranking.

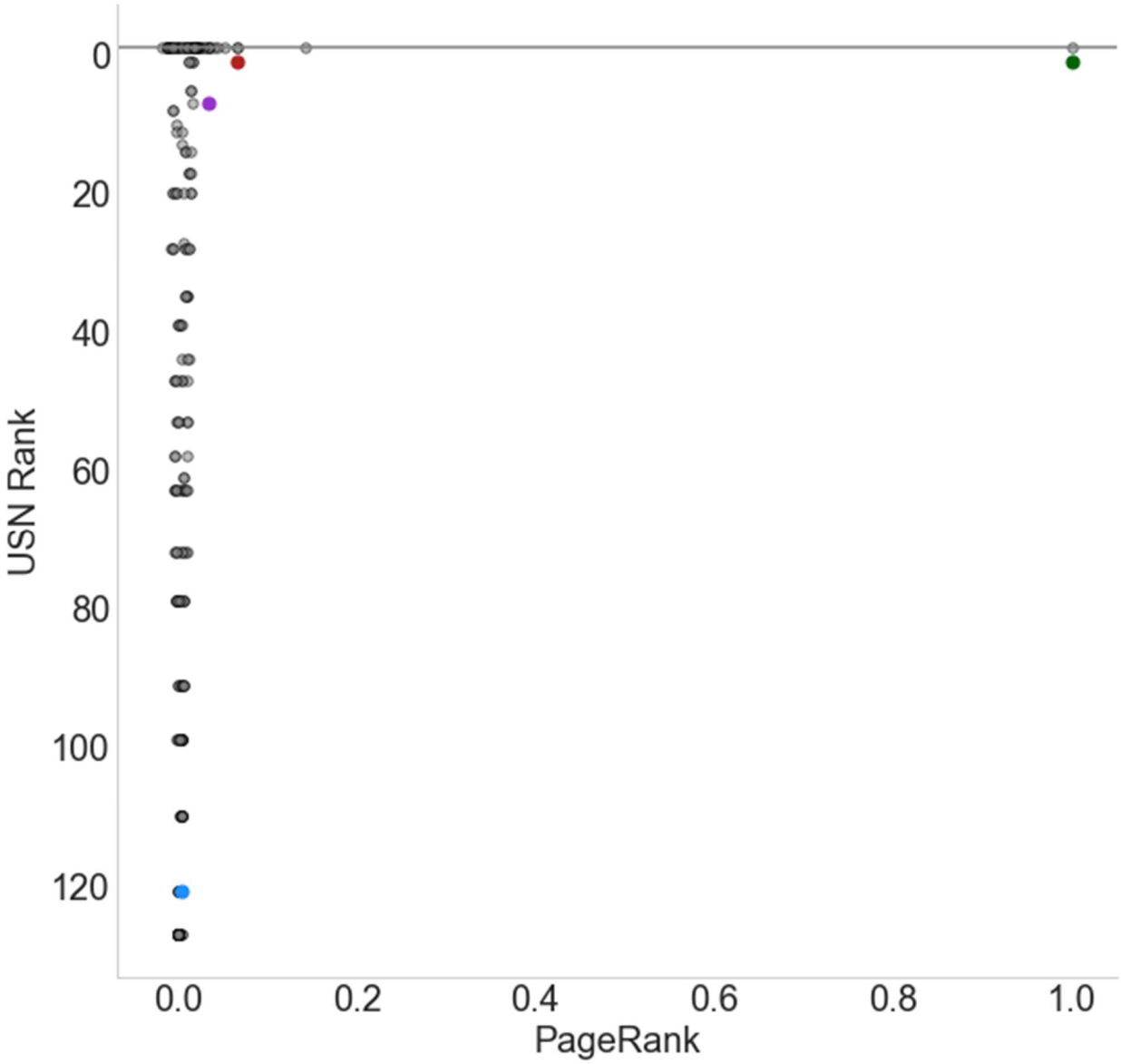
Four universities are highlighted in the plots: Stanford (green), MIT (red), University of Washington (purple), and Wright State University (blue).

The horizontal line at the top represents universities in the dataset without USN ranks. Notable, SpringRank captured more information about the difference in rank between these universities than the other methods, likely due to optimizing the entire network when determining rankings.

Author: Andrey Risukhin
Mentor: Apara Venkat
SPA DRP Spring 2021



Author: Andrey Risukhin
Mentor: Apara Venkat
SPA DRP Spring 2021



Author: Andrey Risukhin
Mentor: Apara Venkat
SPA DRP Spring 2021