

# Networks & Choice Modeling

Andrey Risukhin

University of Washington

June 8, 2021

# How to Spend Resources

- People have limited resources
  - Time
  - Budget
- How do you choose what to spend on?
  - Rational decision maker
  - “Maximizing Utility” for individual
- Example scenarios
  - Whether to join the DRP program
  - Which item on the lunch menu

# How to Hire Faculty

- Universities have limited resources
  - Budget
- How does one choose who to hire?
  - “Maximizing Utility” for group interests
    - Consistency
  - Use a proxy?
    - USN Rank
    - Publication count
  - An algorithm?

# Outline

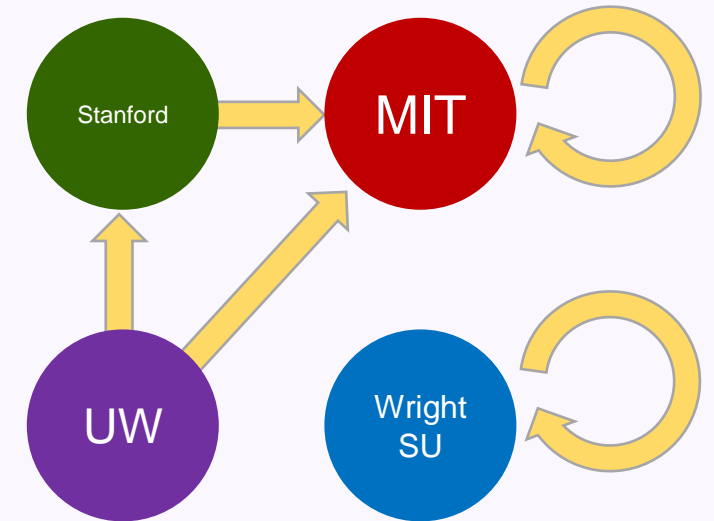
- Inter-university faculty hiring data
- Algorithms to discover inherent ranking
  - Discrete Logit
  - PageRank
  - SpringRank
- Assess the models

# Faculty Hiring

- 206 universities (vertices)
  - Features (region, USN)
  - USN rank may be duplicate or missing
- 4989 hires (edges)
  - Directed (u hires from v)
  - Can have self loops
  - Features (rank, gender)
- Choices & Network Ranks represented

#	u	pi	USN2010	NRC95	Region	institution
1	2.23	1	1	West	Stanford University	
2	2.31	1	3	West	UC Berkeley	
3	3.52	1	2	Northeast	MIT	
4	5.24	11	12	West	California Institute of Technology	
5	6.12	17	11	Northeast	Harvard University	
6	8.29	5	5	Northeast	Cornell University	
7	9.28	1	4	Northeast	Carnegie Mellon University	
8	9.32	8	6	Northeast	Princeton University	
9	9.98	20	14	Northeast	Yale University	
10	11.06	7	9	West	University of Washington	

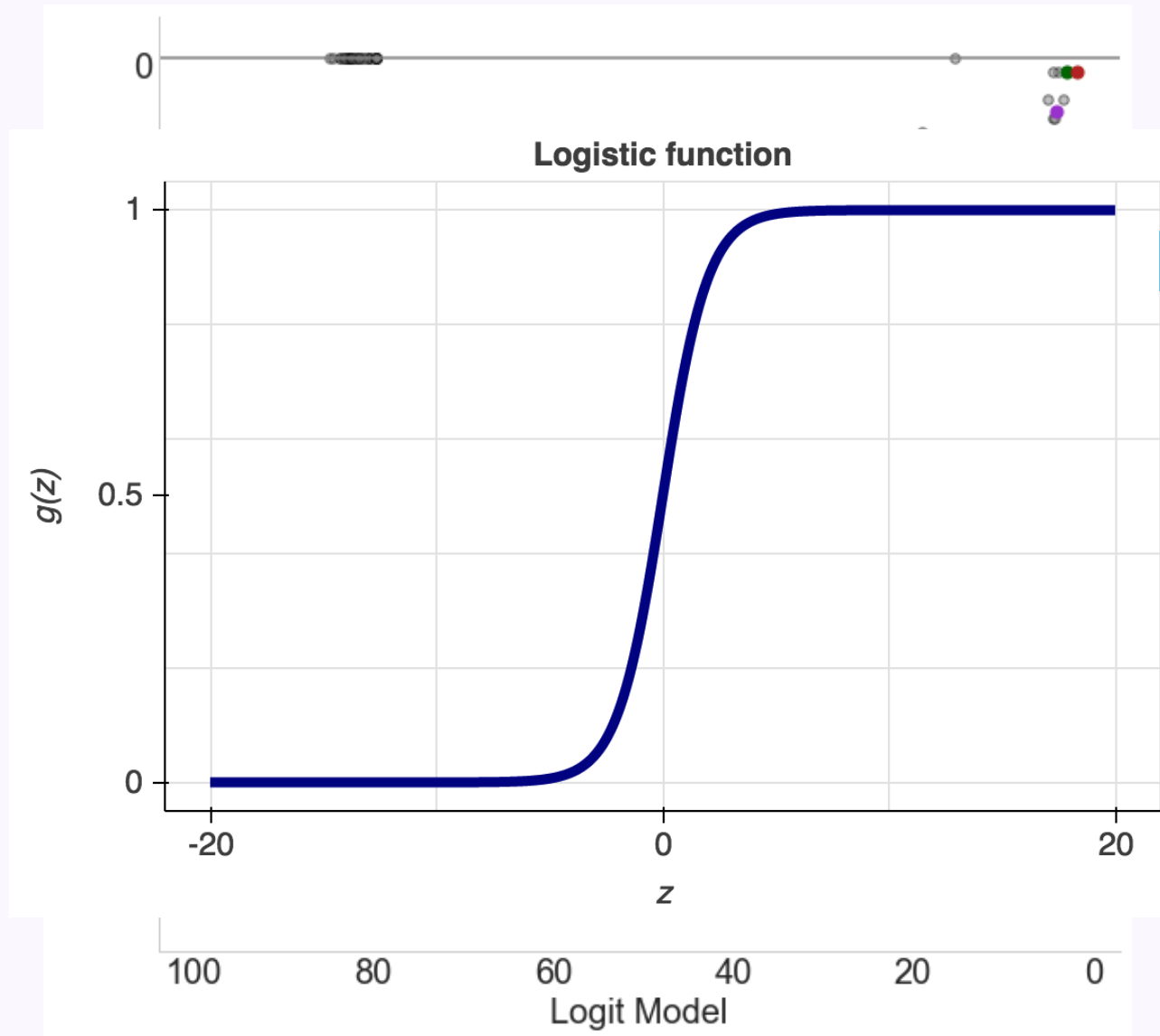
#	u	v	rank	gender
1	1	1	Asst	M
1	1	1	Assoc	M
1	1	1	Assoc	M
1	1	1	Full	M
1	1	1	Full	M
1	1	1	Full	M
1	1	1	Full	F
1	1	1	Full	M
1	1	1	Full	M



# Discrete Logit

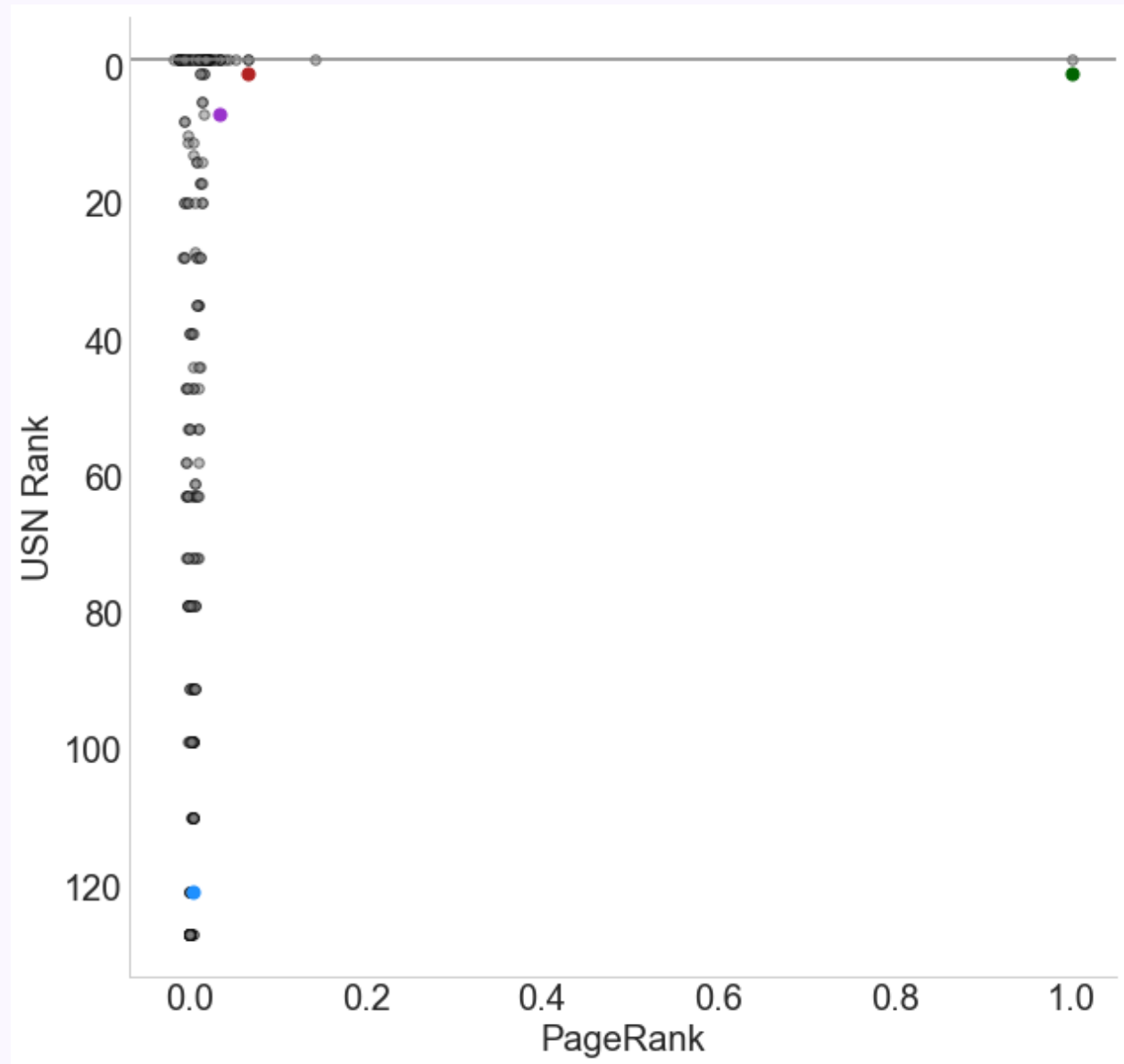
- Utility = observable utility (features) + unknown utility (?)
- Assume unknown utility is “independently, identically distributed extreme value”
  - Derives discrete logit (popular)
- Differences in utility matter ( $\varepsilon^*$ )
  - More important when choices are similar

$$F(\varepsilon_{nji}^*) = \frac{e^{\varepsilon_{nji}^*}}{1 + e^{\varepsilon_{nji}^*}}$$



# PageRank

- Use connections between universities as ranking
  - Works for the internet, Google PageRank uses links
- Linear algebra extracts intrinsic ranks (eigenvalues) from the graph
- Eigenvalues are not interpretable



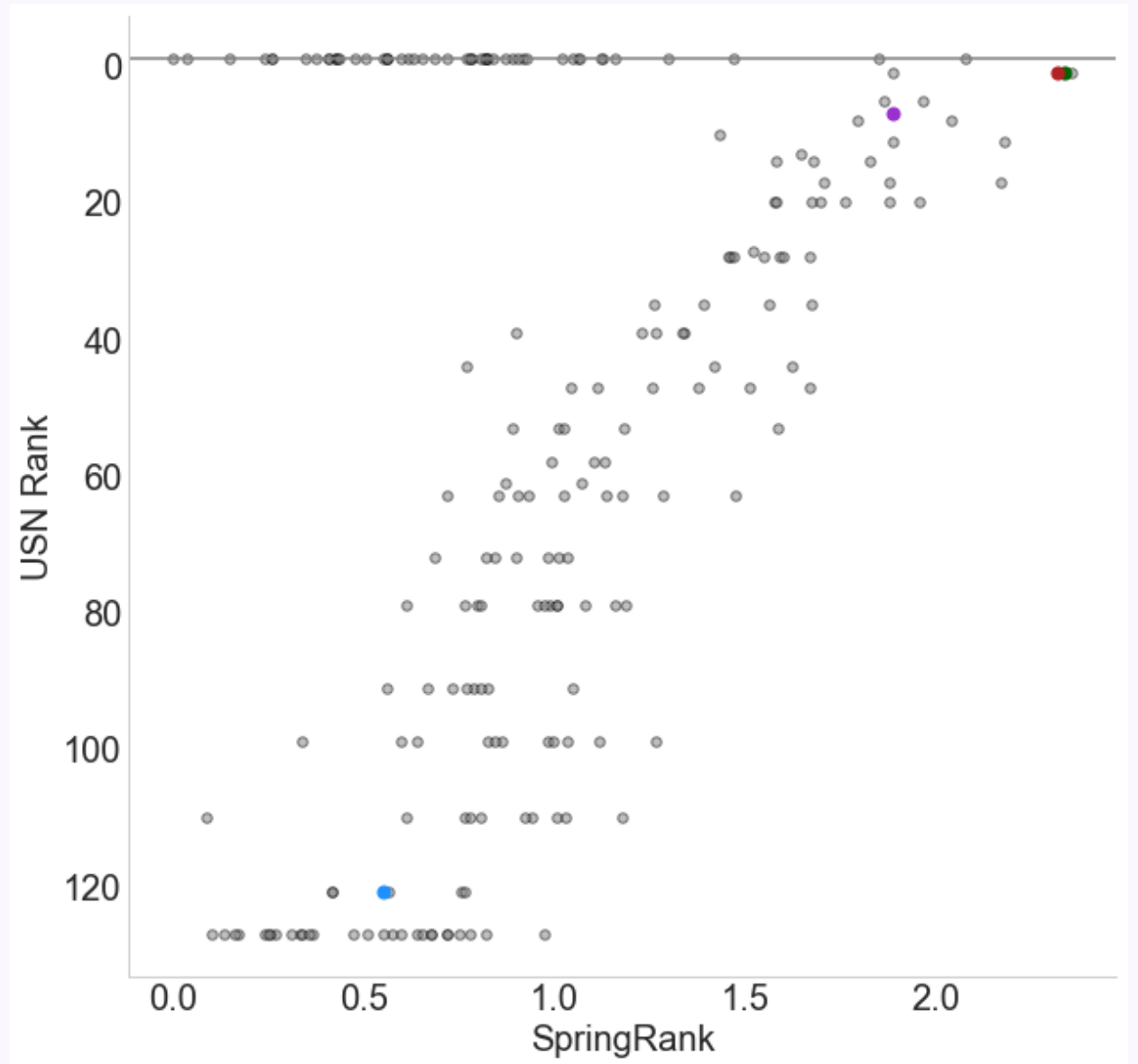
# SpringRank

- Use connections between universities as “springs”, minimize “energy” of network
- Nice interpretation in terms of log(odds)

Probability that an edge goes from  $i$  to  $j$  ( $i$  beats  $j$  in competition)

$$P_{ij}(\beta) = \frac{e^{-\beta H_{ij}}}{e^{-\beta H_{ij}} + e^{-\beta H_{ji}}} = \frac{1}{1 + e^{-2\beta(s_i - s_j)}}.$$

(nodes embedded in real-valued position/rank)

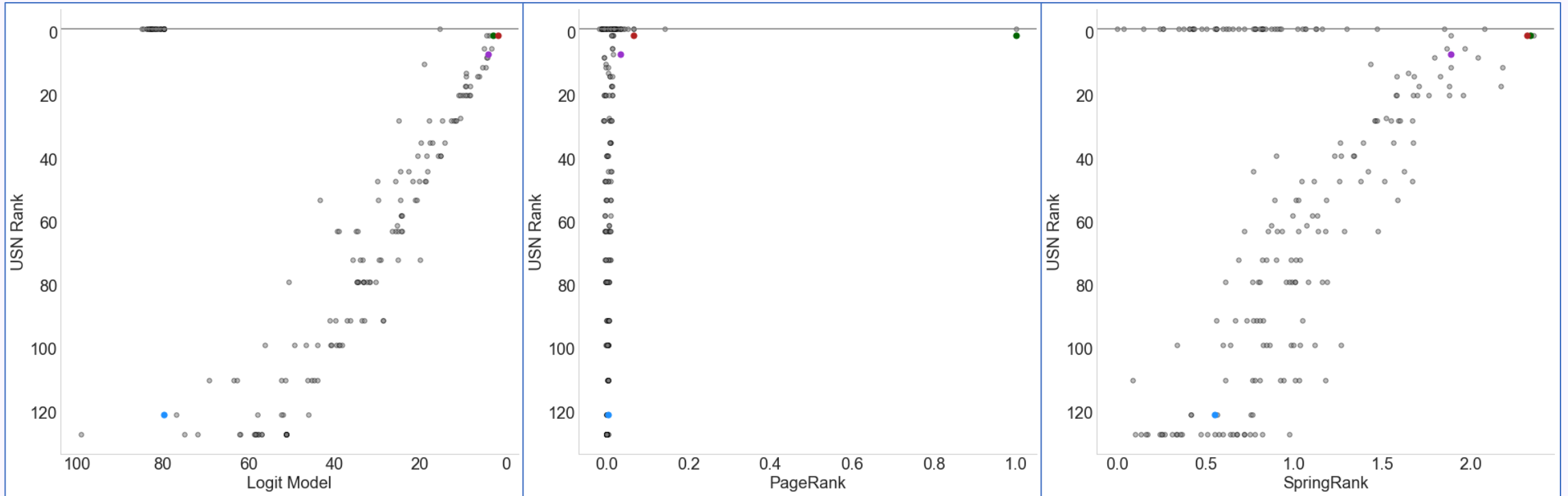




# Summary

## Algorithm Rankings

- 1. Stanford
- 2. MIT
- 3. UW
- 4. Wright State University



Discrete Logit	PageRank	SpringRank
Needs features & edges	Uses only edges	Uses only edges
Interpretable (log odds)	Not interpretable	Interpretable (log odds)

# Acknowledgements

- Mentor Aparajithan (Apara) Venkateswaran
- SPA DRP for opportunity
- Larremore Lab for SpringRank code
- Aaron Clauset, Samuel Arbesman, and Daniel B. Larremore for faculty hiring dataset

# Extra: Logit IID Extreme Value assumption

- assumption that  $\varepsilon$  is distributed iid extreme value for all  $i$ .  
“unobserved factors are uncorrelated over alternatives, as well as having the same variance for all alternatives”
- This assumption, while restrictive, provides a very convenient form for the choice probability (logistic)
- assumption of independence can be inappropriate in some situations. Unobserved factors related to one alternative might be similar to those related to another alternative.
  - For example, a person who dislikes travel by bus because of the presence of other riders might have a similar reaction to rail travel; if so, then the unobserved factors affecting bus and rail are correlated rather than independent.
  - Logit model is applied to sequences of choices over time