

Northeastern University
Network Science Institute

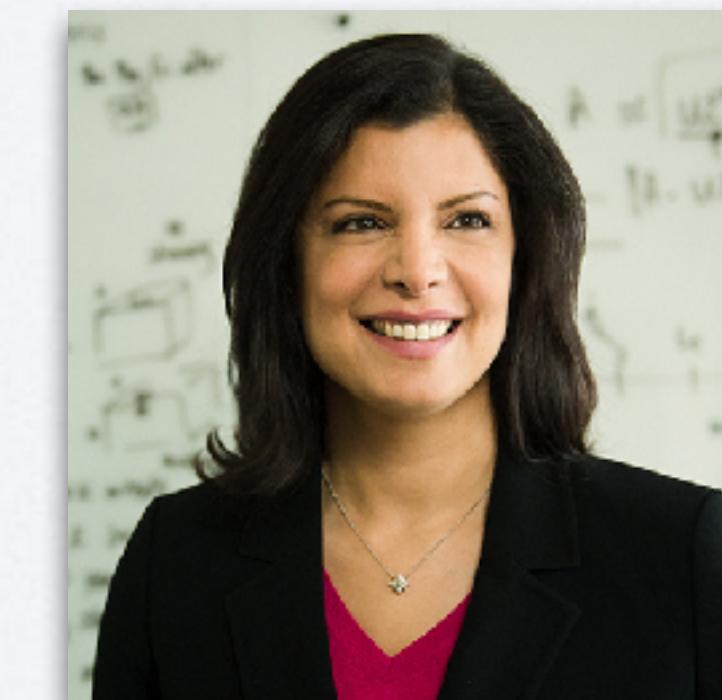
LEARNING TO PLACE OBJECTS: A NETWORK-BASED APPROACH

Xindi Wang

Onur Varol



Tina Eliassi-Rad

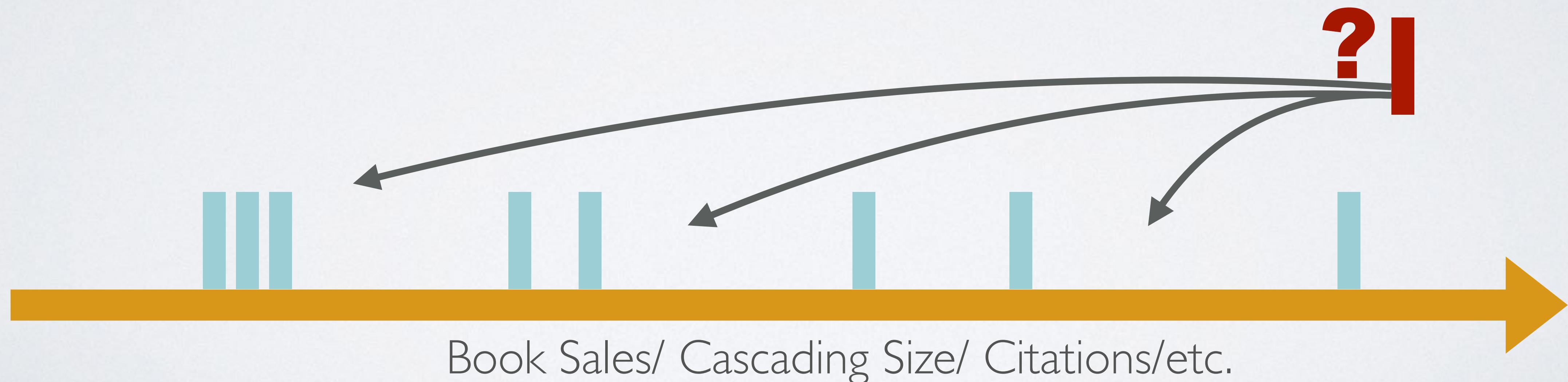


Albert-László Barabási



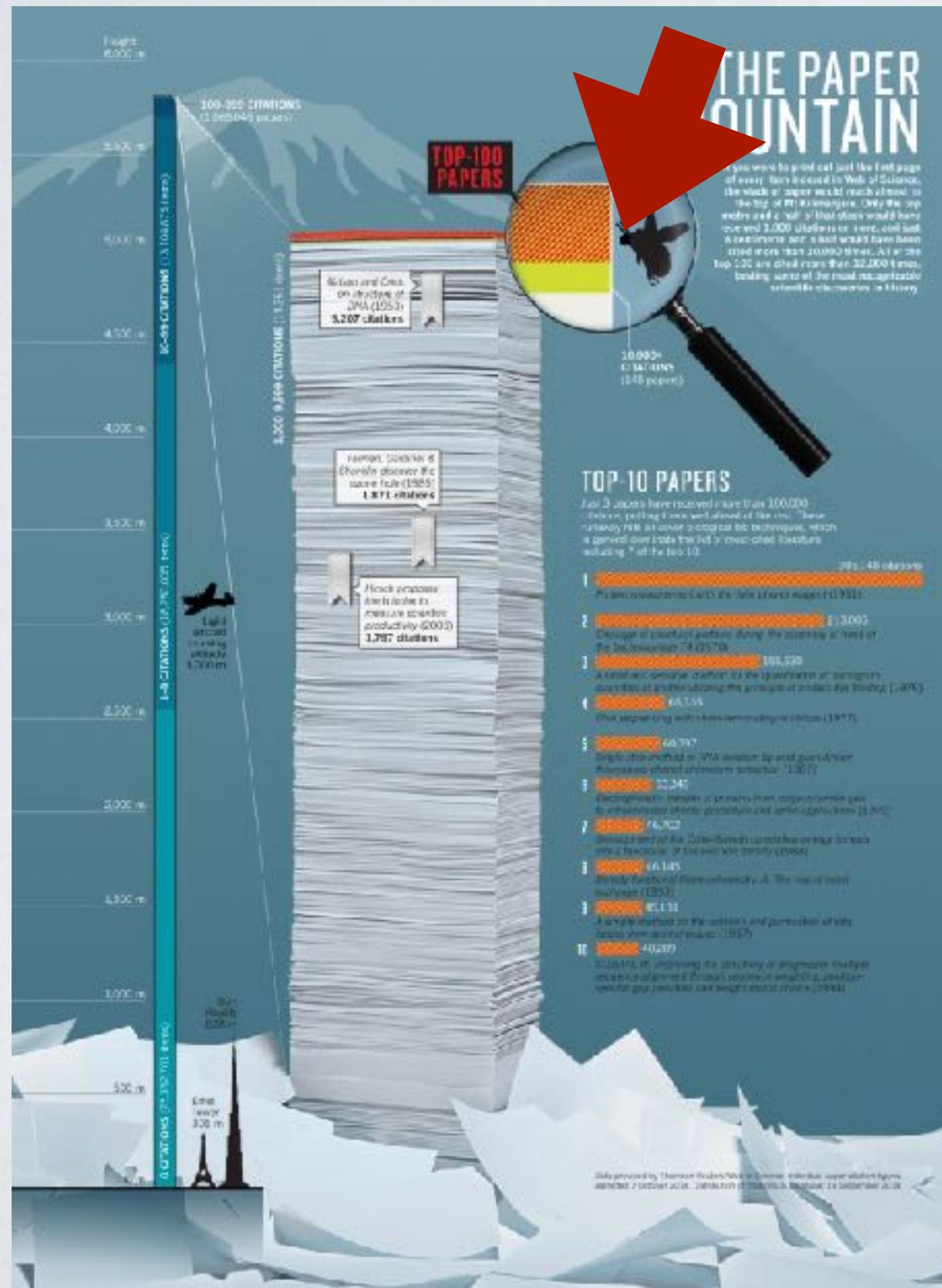
PROBLEM DEFINITION

Learning to place: Given a sequence of ordered items, where should we place a new item?

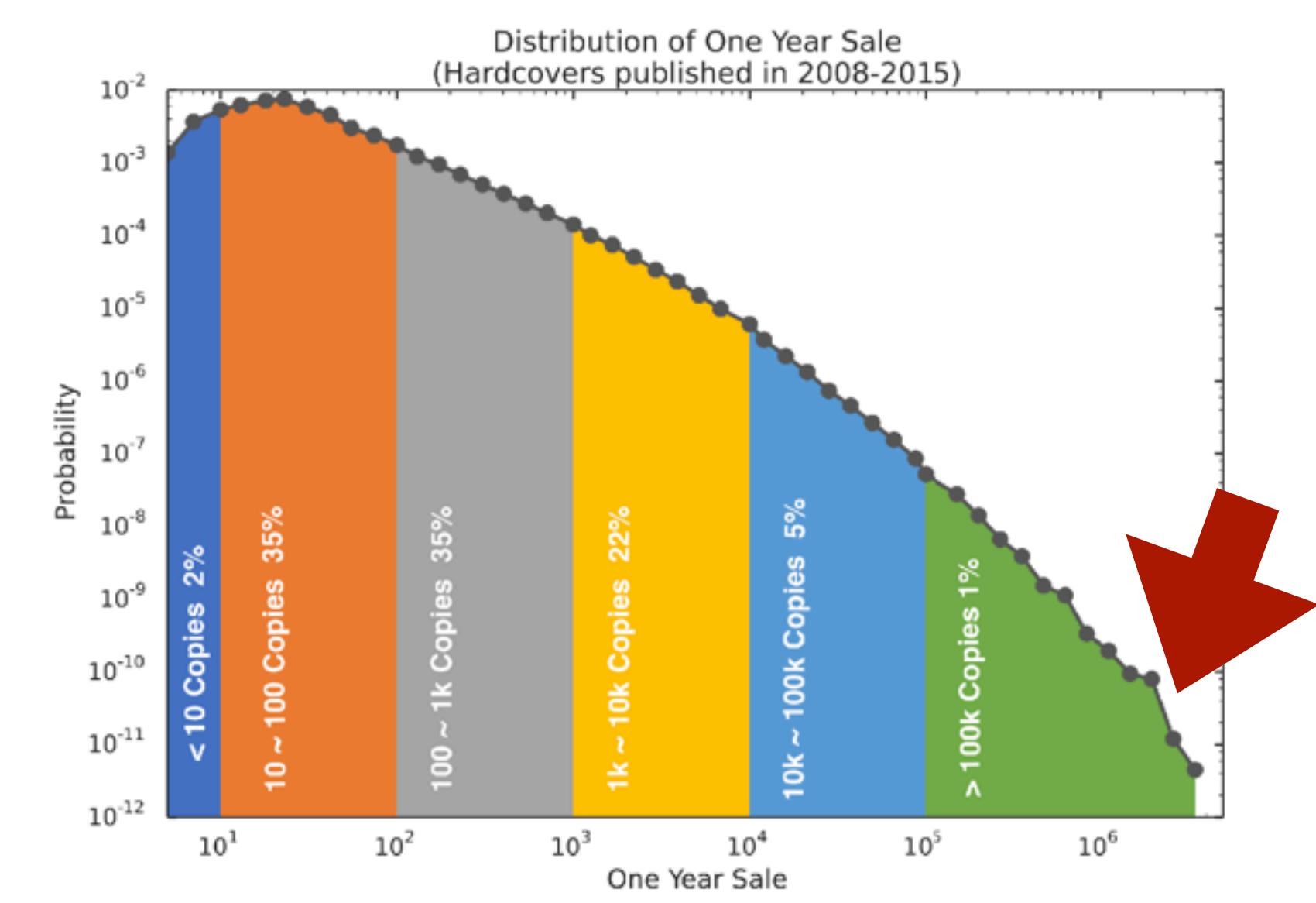


PREDICT HEAVY-TAILED ATTRIBUTES OF AN ITEM

Citation of Papers



Book Sales



Cascade Size of Memes

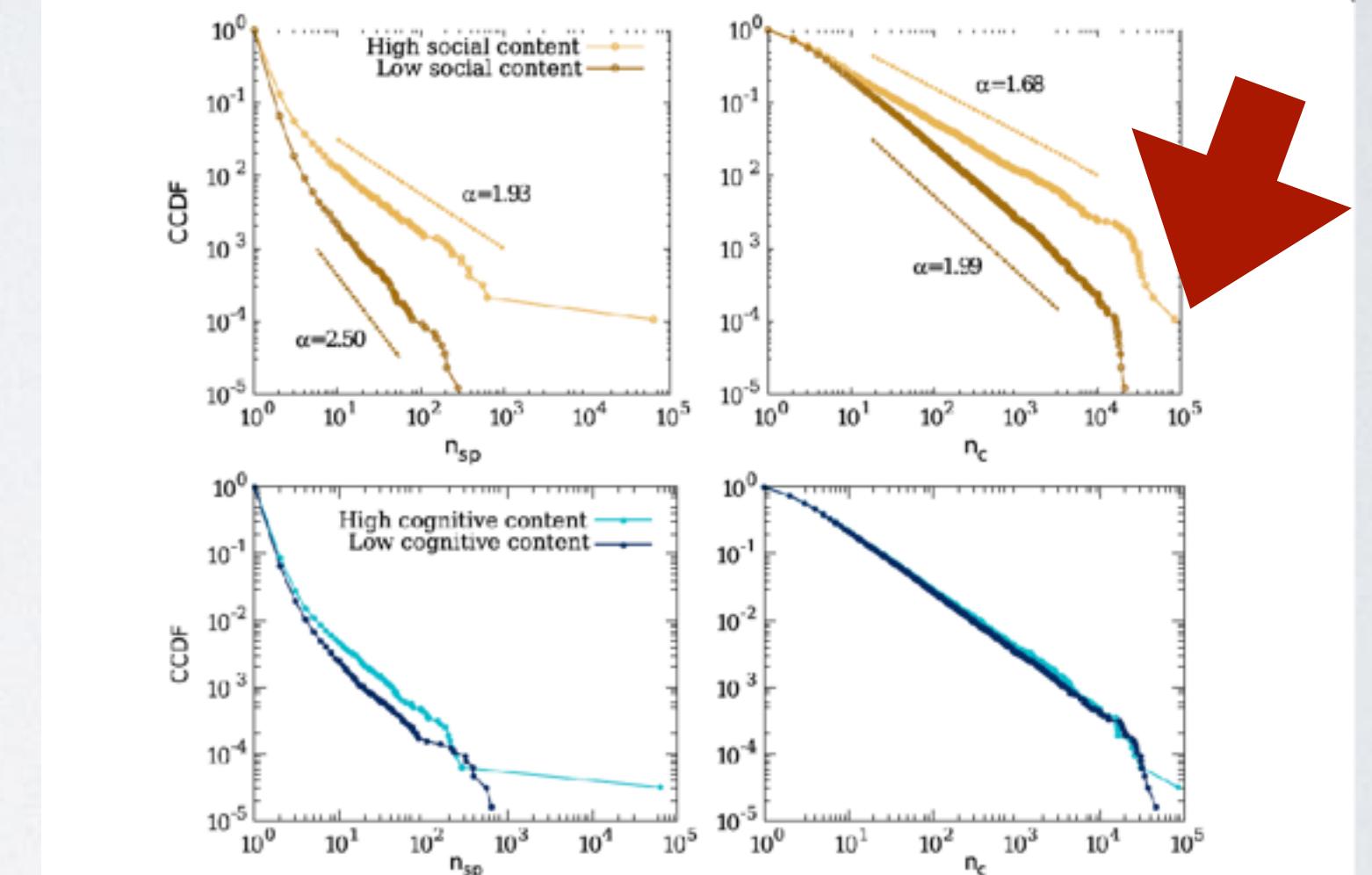


Figure 2 CCDF of activity (left) and information (right) cascade sizes for cascades of high and low social content (top) and high and low cognitive content (bottom). Dashed lines show the result of power-law fits.

Image: The Paper Mountain (Van Noorden, Richard, Brendan Maher, and Regina Nuzzo. "The top 100 papers." *Nature News* 514.7524 (2014): 550.)

Figure: Cascade Size Distribution (Alvarez, Raquel, et al. "Sentiment cascades in the 15M movement." *EPJ Data Science* 4.1 (2015): 6.)

CHALLENGES

- Heavy-tailed property leads to a huge class imbalance problem
- Methods like linear regression heavily underpredict the “big and rare” instances

CURRENT APPROACHES

- SMOTE: Synthetic Minority Over-sampling Technique
(Chawla et al., 2002)
 - Over-sample minority class and under-sample majority class
 - Redefine the problem to a binary classification one
 - Would this cascade exceeds this threshold or not?
- (Subbian et al., 2017)

Chawla, Nitesh V., et al. "SMOTE: synthetic minority over-sampling technique." *Journal of artificial intelligence research* 16 (2002): 321-357.

Subbian, Karthik, B. Aditya Prakash, and Lada Adamic. "Detecting large reshare cascades in social networks." *Proceedings of the 26th International Conference on World Wide Web*. International World Wide Web Conferences Steering Committee, 2017.

LEARNING TO PLACE

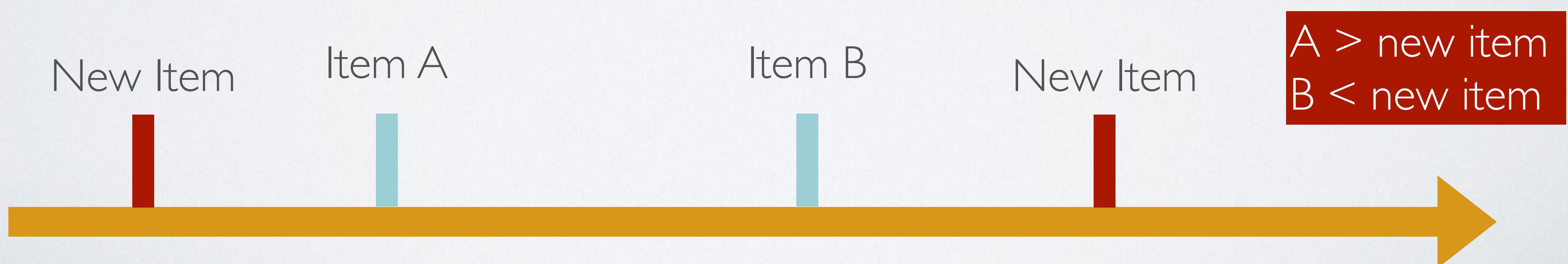
- Phase I: Build a classifier for pairwise comparison (based on a given training set)



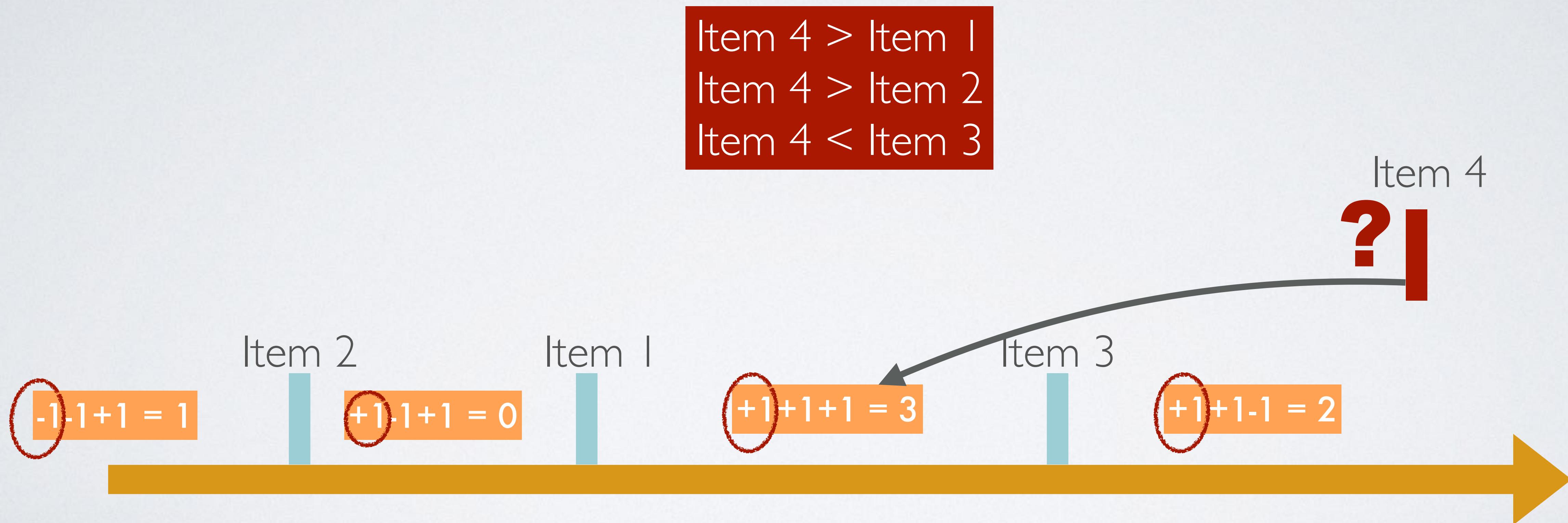
A > B: +1

A < B: -1

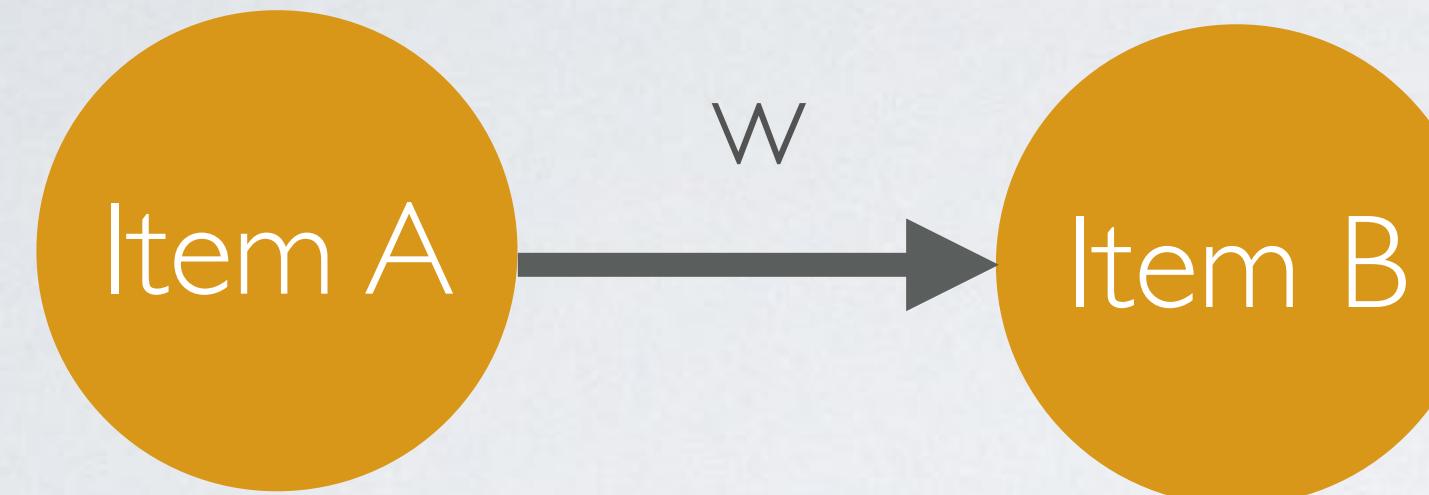
- Phase II: Find places for new items



ONE APPROACH FOR PHASE II: VOTE



ANOTHER APPROACH FOR PHASE II: USE A GRAPH



A beats B with confidence w

Weighted Tournament Graph (WTG)



High Out-degree, Low In-degree



High In-degree, Low Out-degree

THREE HEURISTICS ON WTG TO PLACE NEW ITEMS

- Cohen et al.
- Our approach: WTG wave
- FAS-PIVOT (Ailon et al., 2008)

Cohen, William W., Robert E. Schapire, and Yoram Singer. "Learning to order things." *Advances in Neural Information Processing Systems*. 1998.

Ailon, Nir, Moses Charikar, and Alantha Newman. "Aggregating inconsistent information: ranking and clustering." *Journal of the ACM (JACM)* 55.5 (2008): 23.

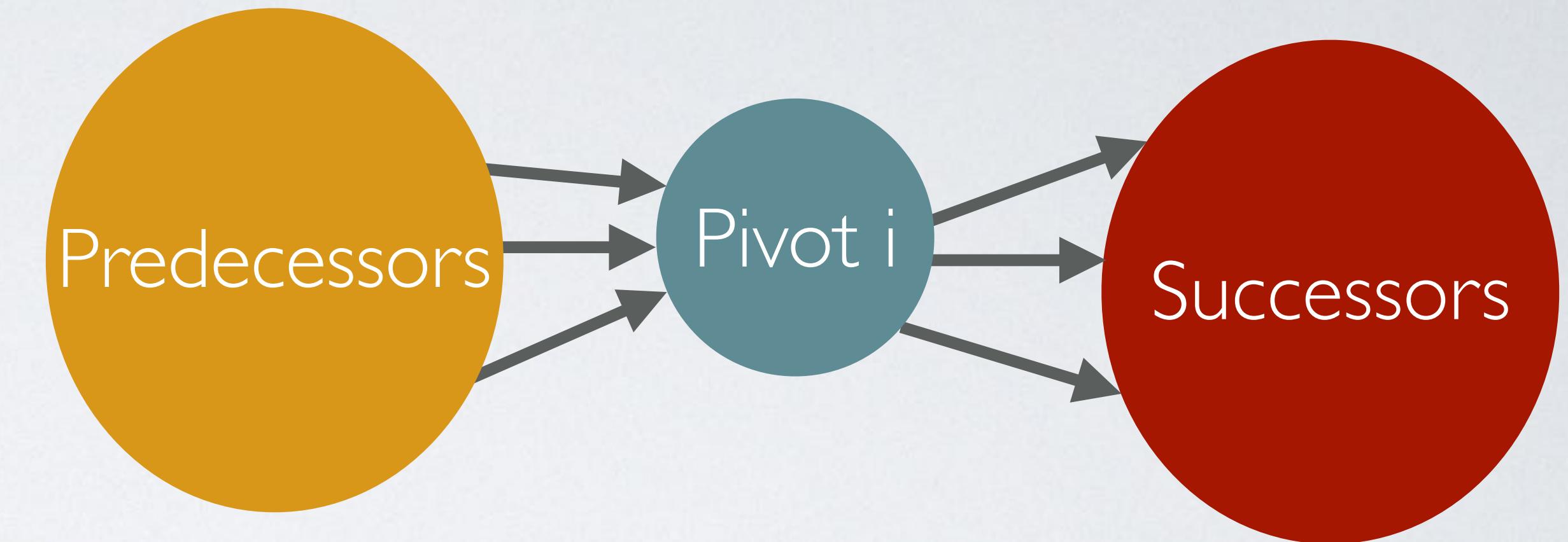
COHEN ET AL. AND WTG WAVE

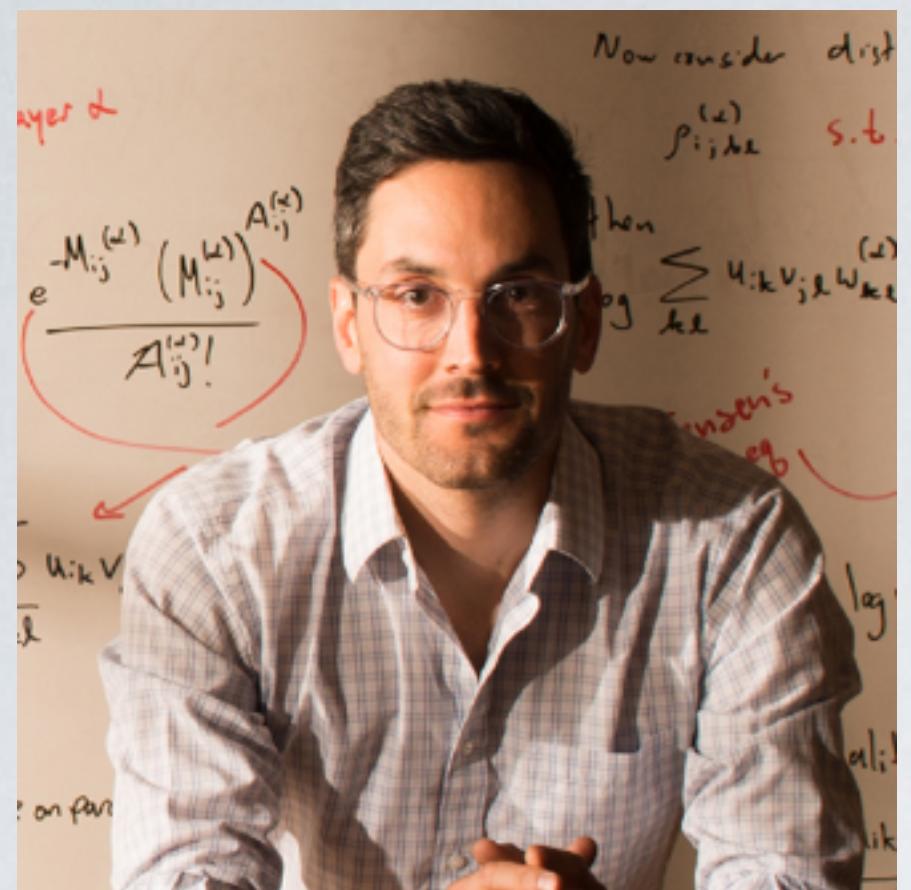
- Cohen et al.
 - $\text{potential}(i) = \text{outdegree}(i) - \text{indegree}(i)$
 - Iteratively remove the node with highest potential and recalculate potential scores
- Our approach:WTG wave
 - Cohen is greedy and localized
 - Extension: look at each node's successors' potential as well
 - $\text{score}(i) = \alpha \cdot \text{potential}(i) + (1 - \alpha) \cdot \sum_{j \in N_i} \text{potential}(i)$

FAS-PIVOT

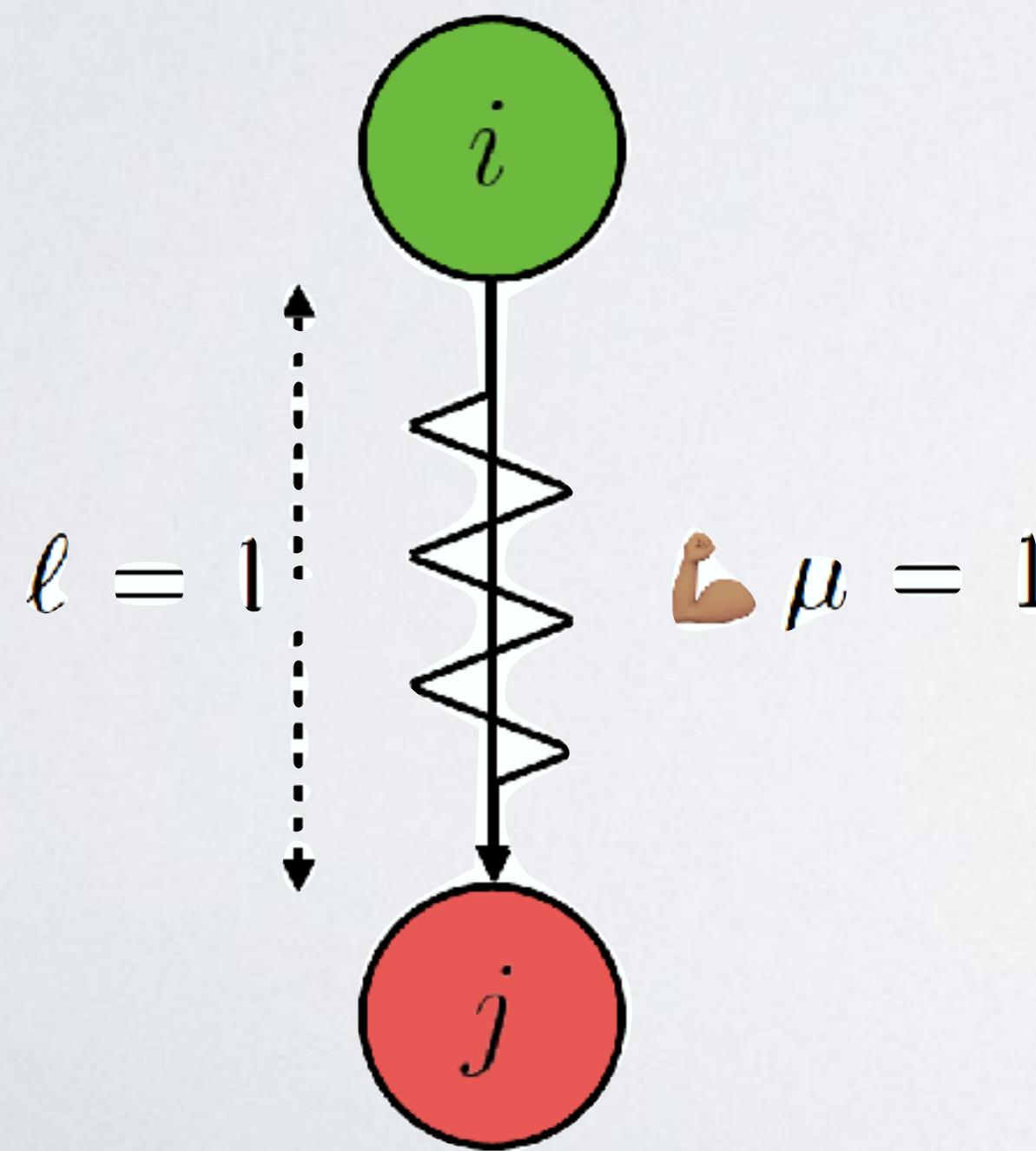
minimum feedback arc set

- Randomly pick a pivot node i
- Put i 's predecessors in V_L ; put i 's successors in V_R
- Recursively do FAS-PIVOT on induced subgraph of V_L and V_R

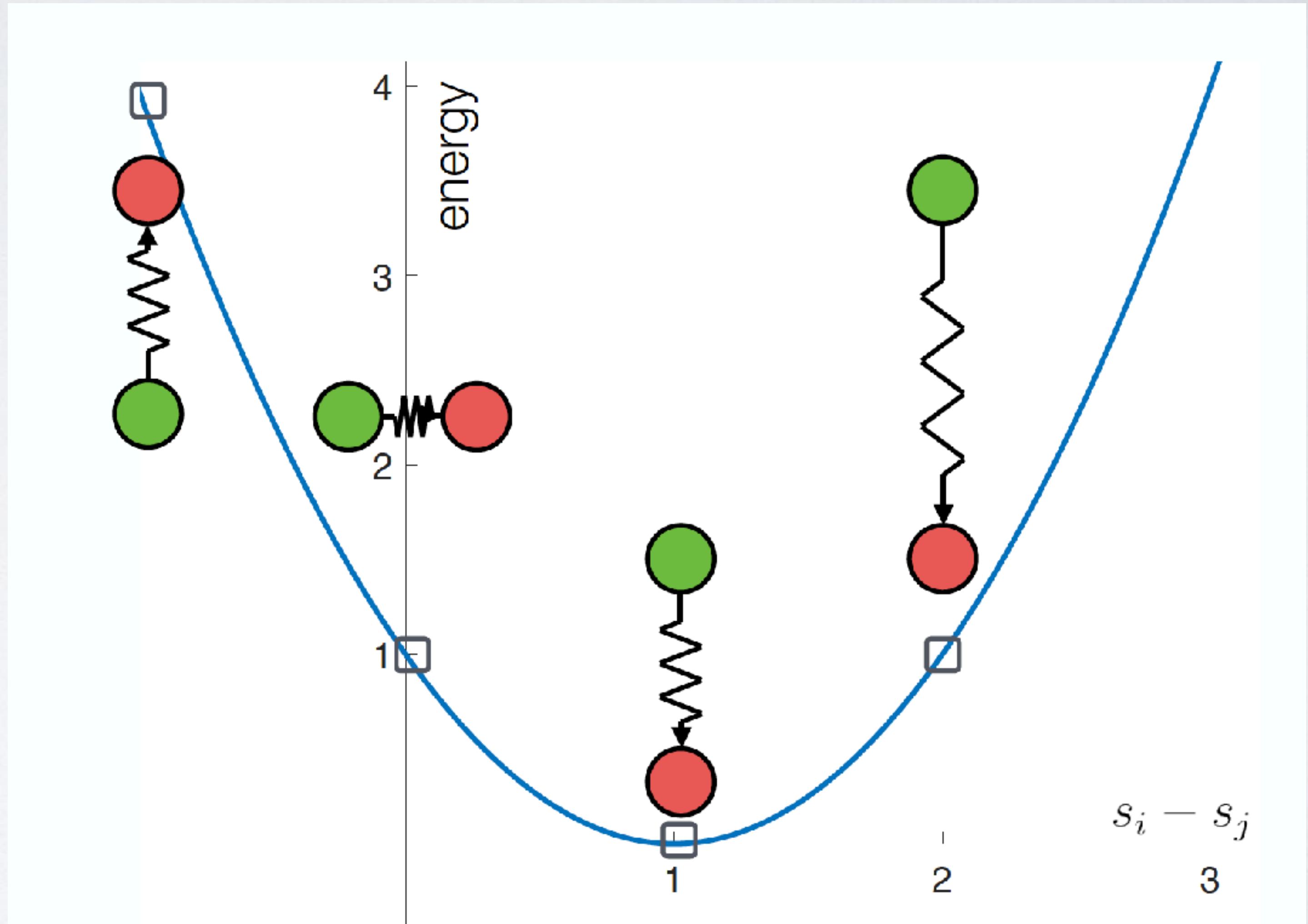




Daniel B. Larremore



SPRINGRANK

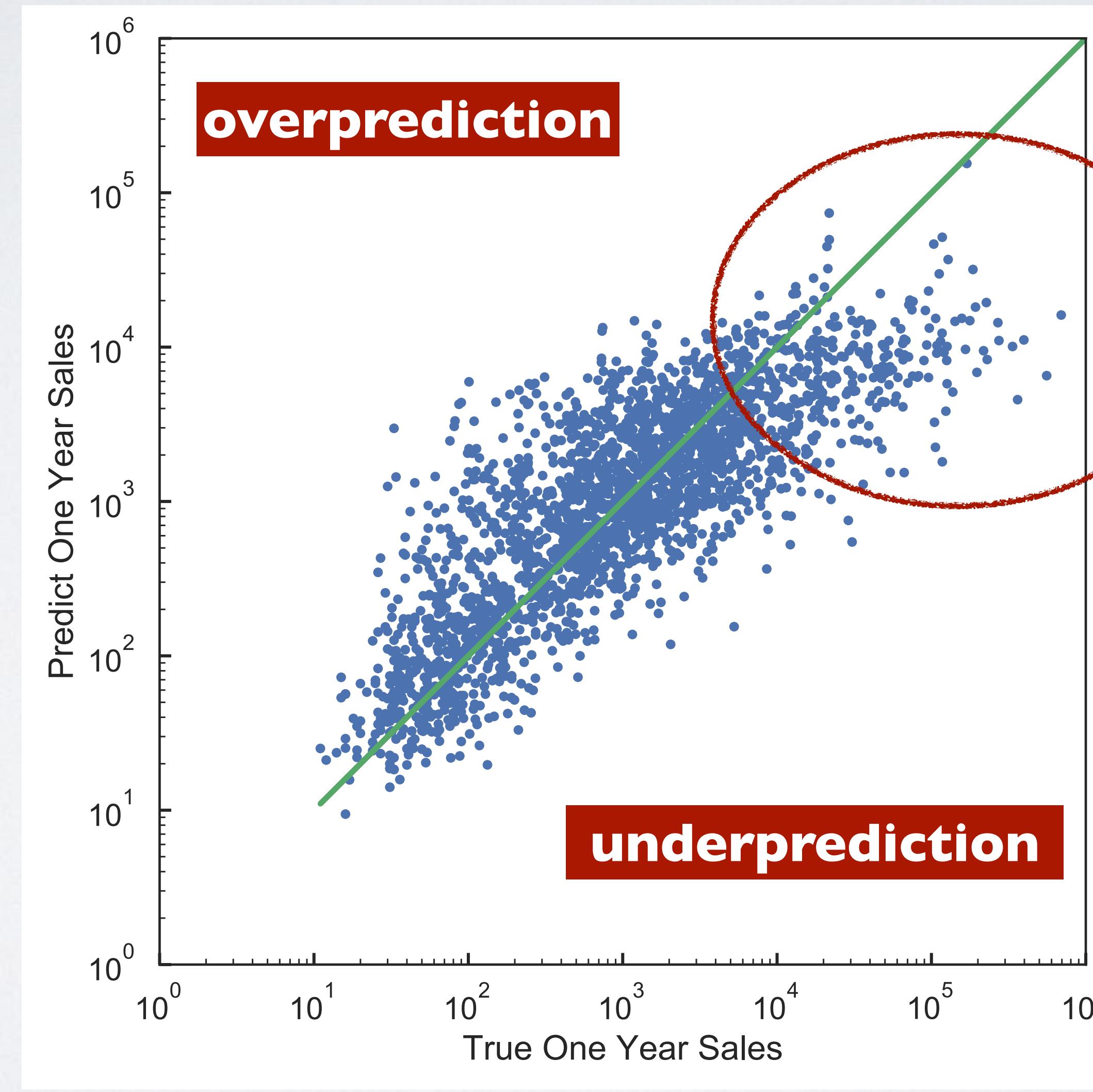


EXPERIMENTS

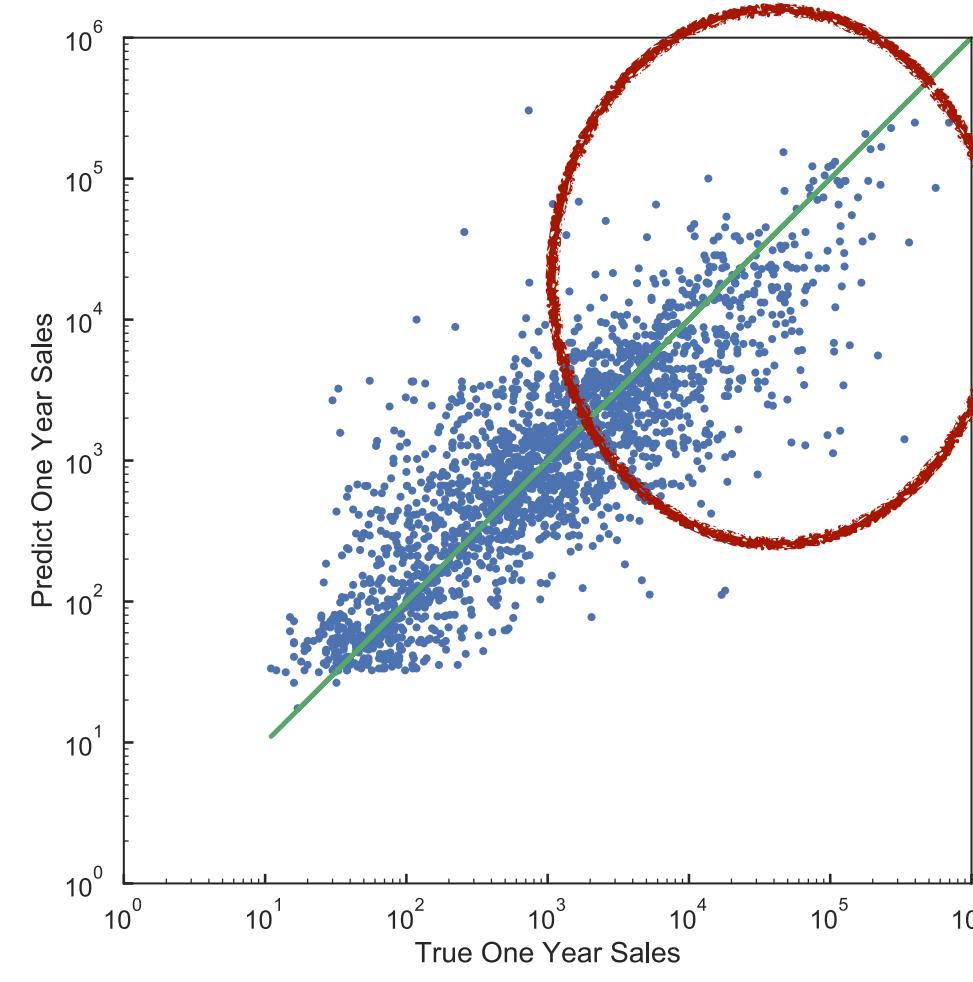
- Data: Book sales data from NPD bookscan, features from various sources including NPD bookscan, Wikipedia, GoodReads, etc.
- Task: Predict the one year sales of the book based on books' features
- Methodology: 5-fold cross validation
- Measurement
 - Sales prediction vs actual sales
 - ROC and AUC on the ranking

UNDERPREDICTION IN LINEAR REGRESSION

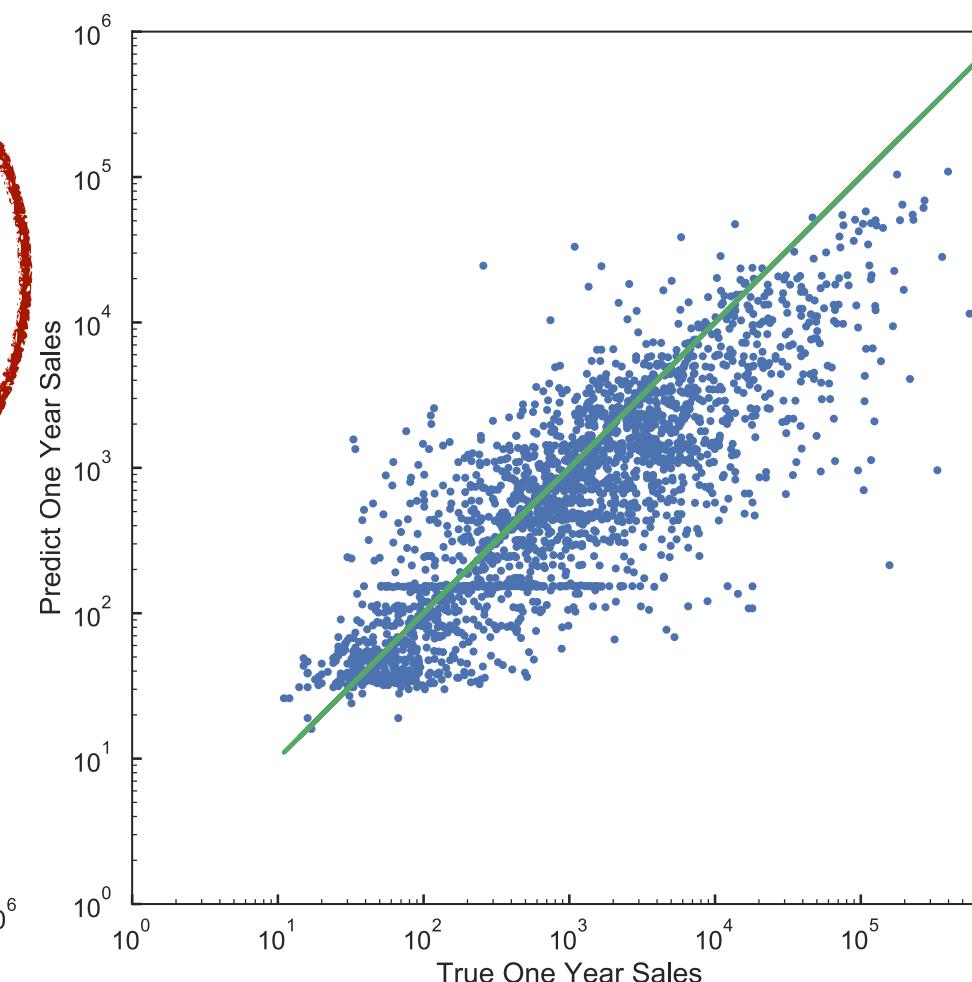
Fiction



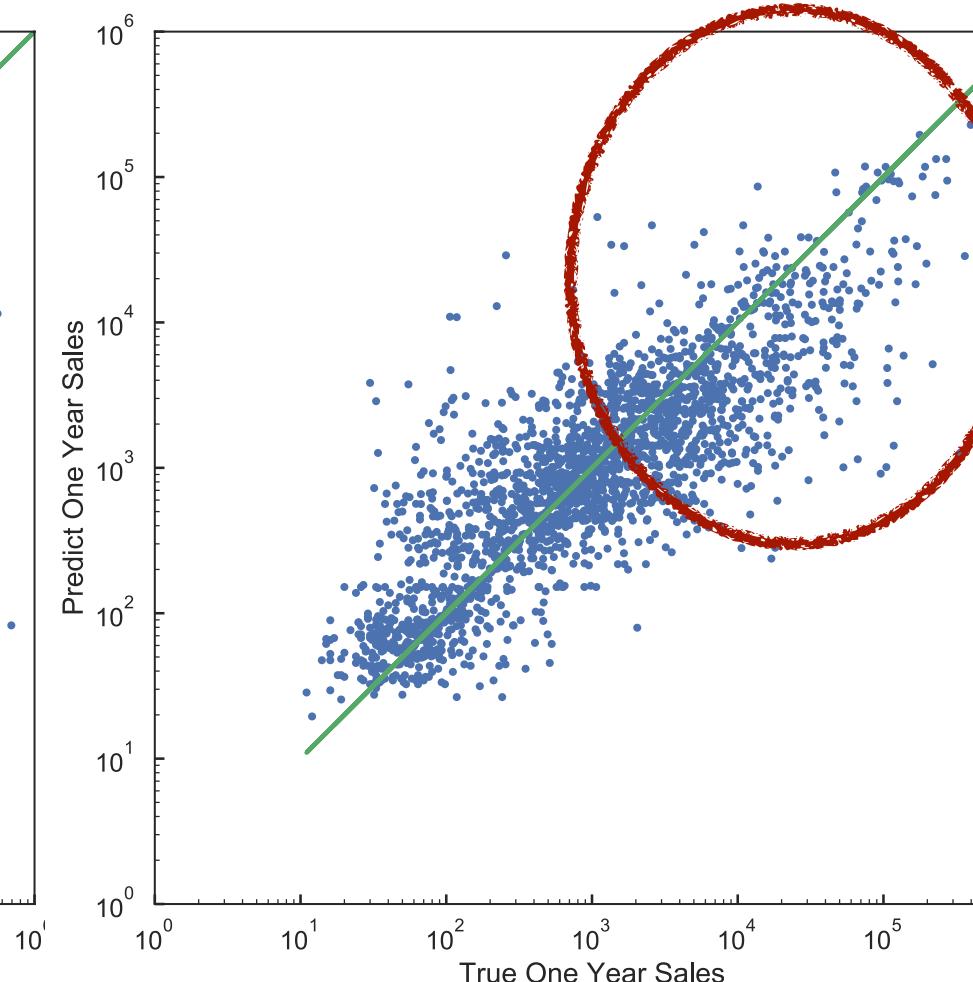
Voting



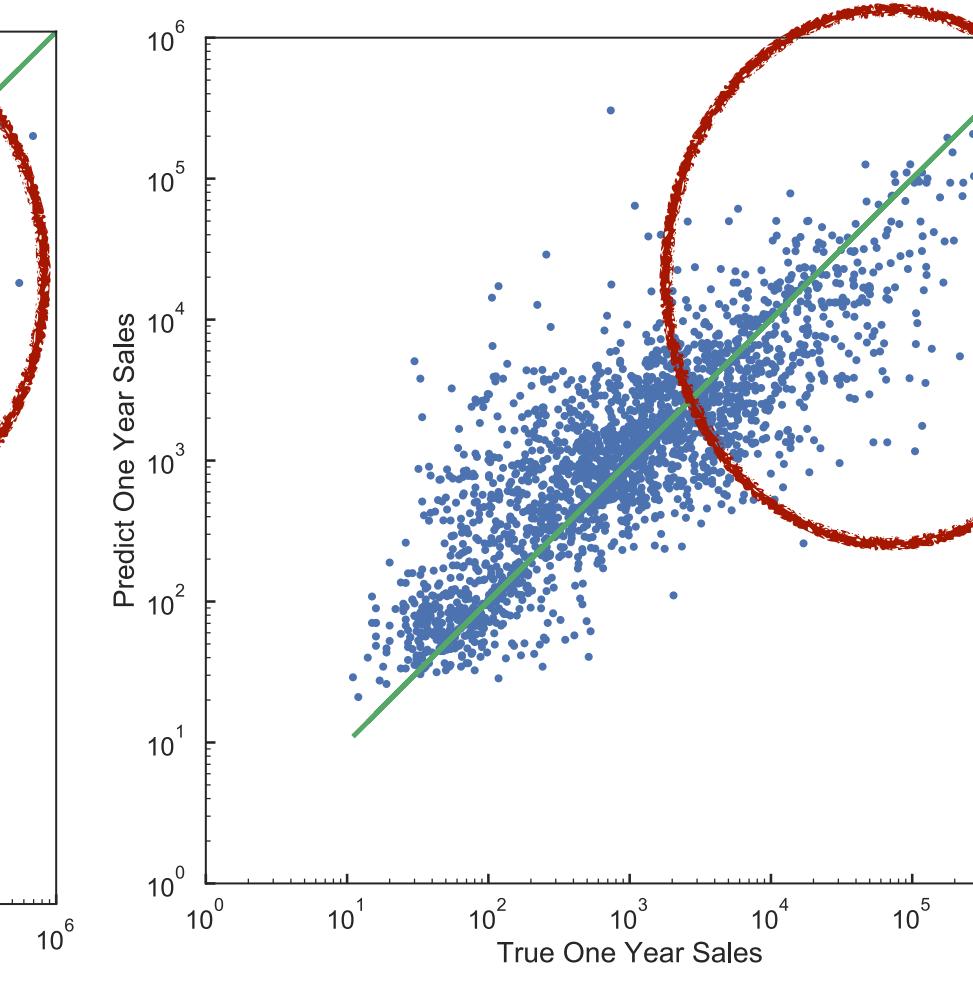
Cohen



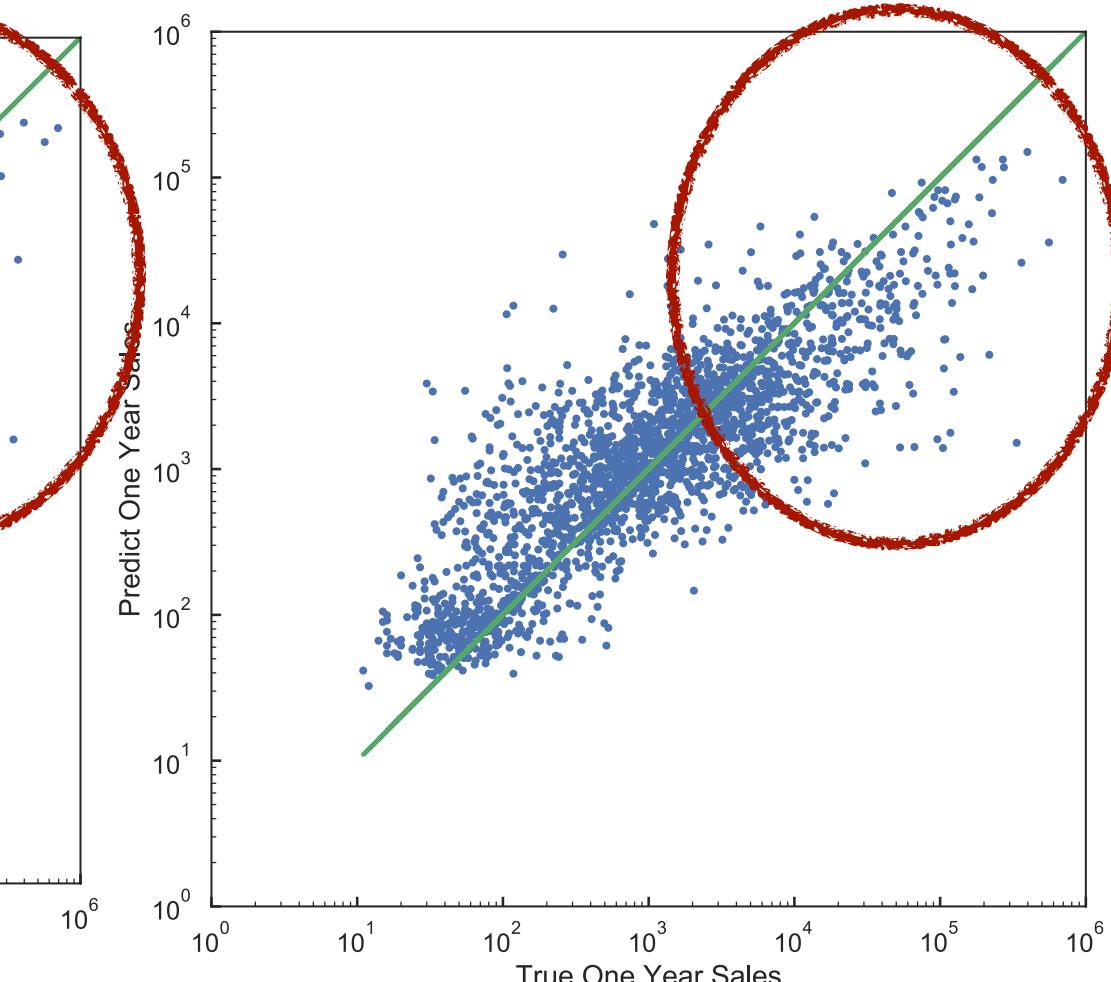
WTG wave



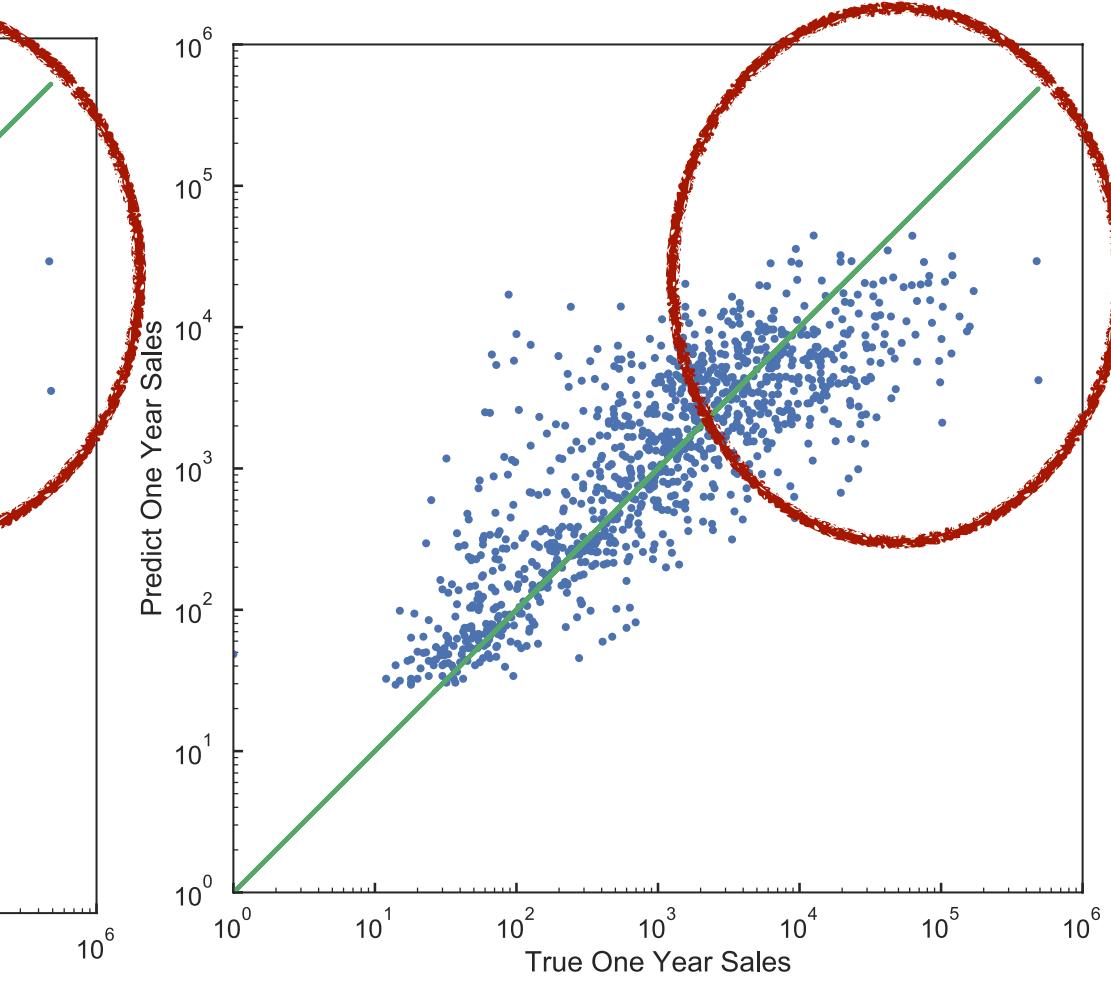
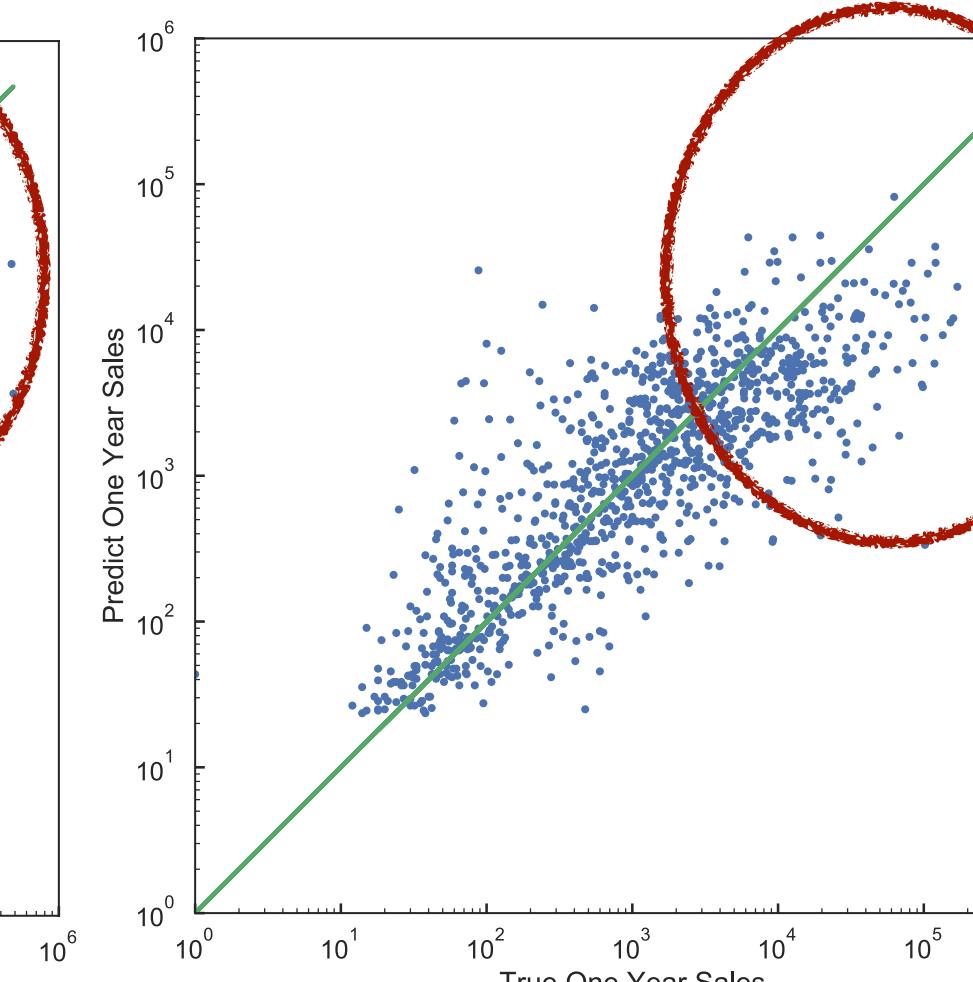
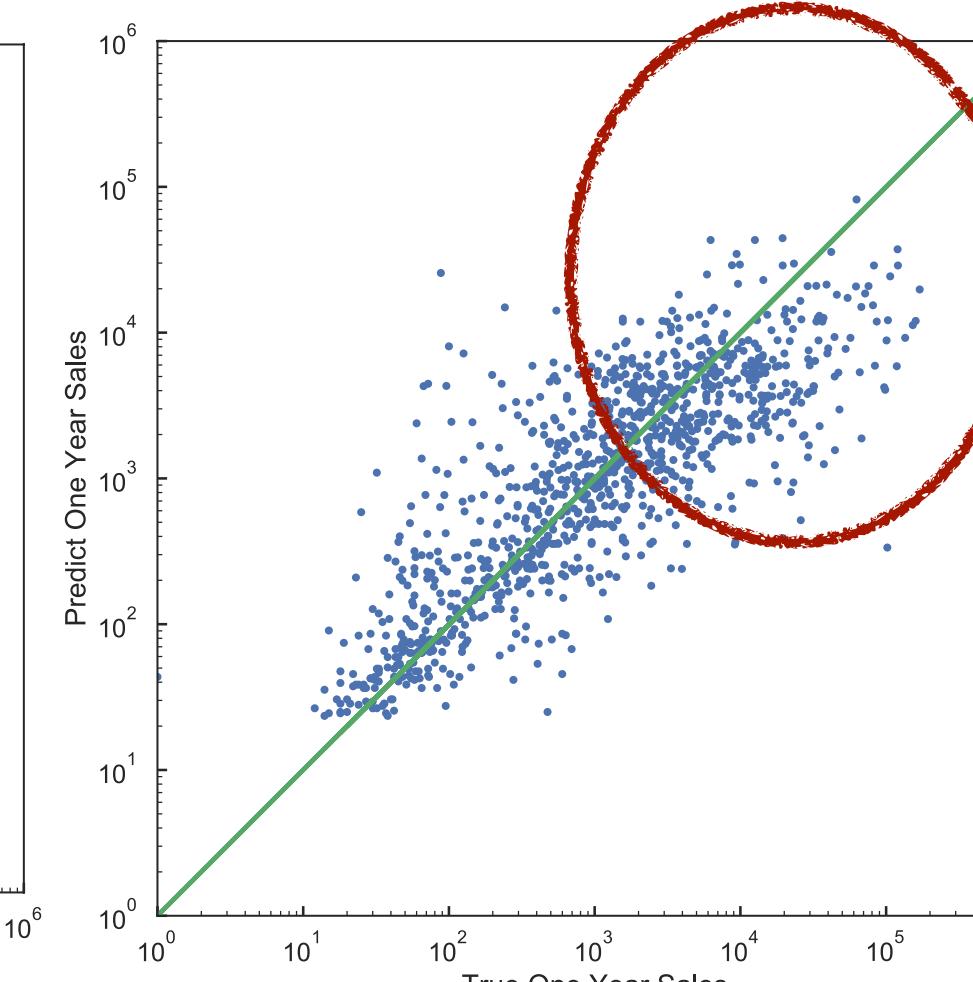
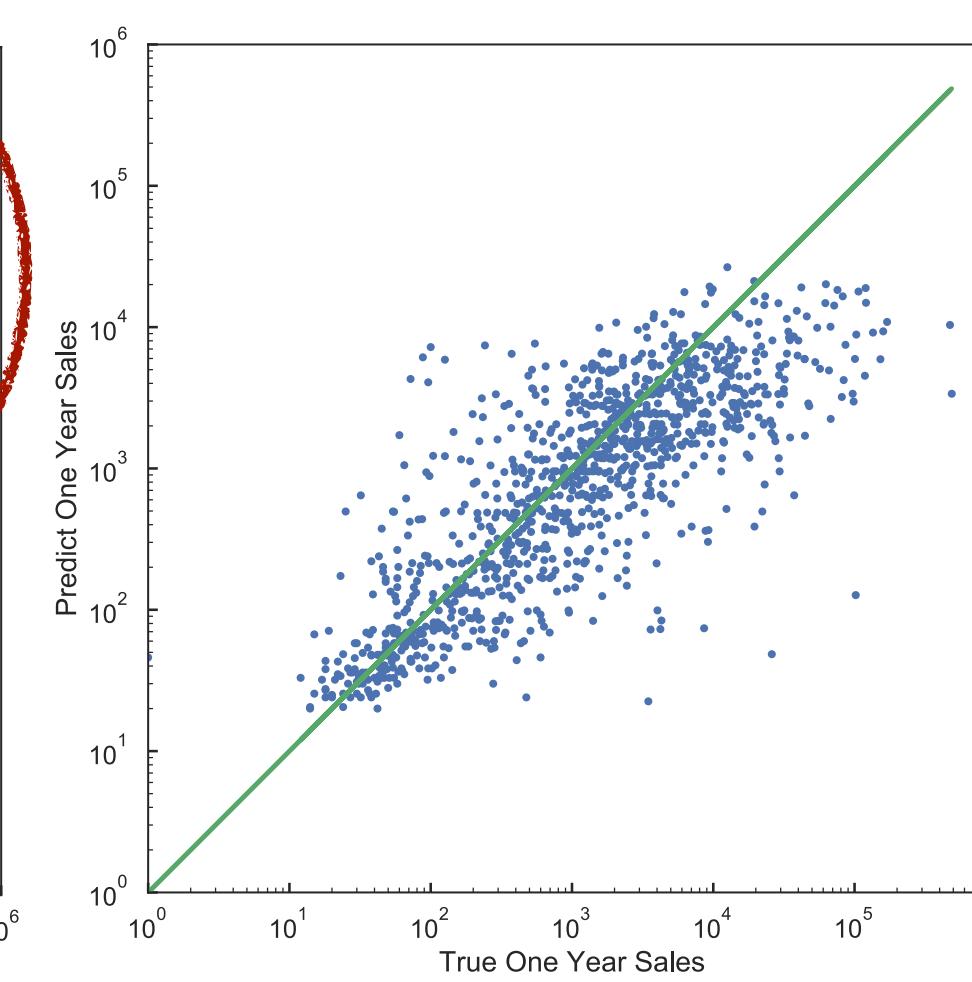
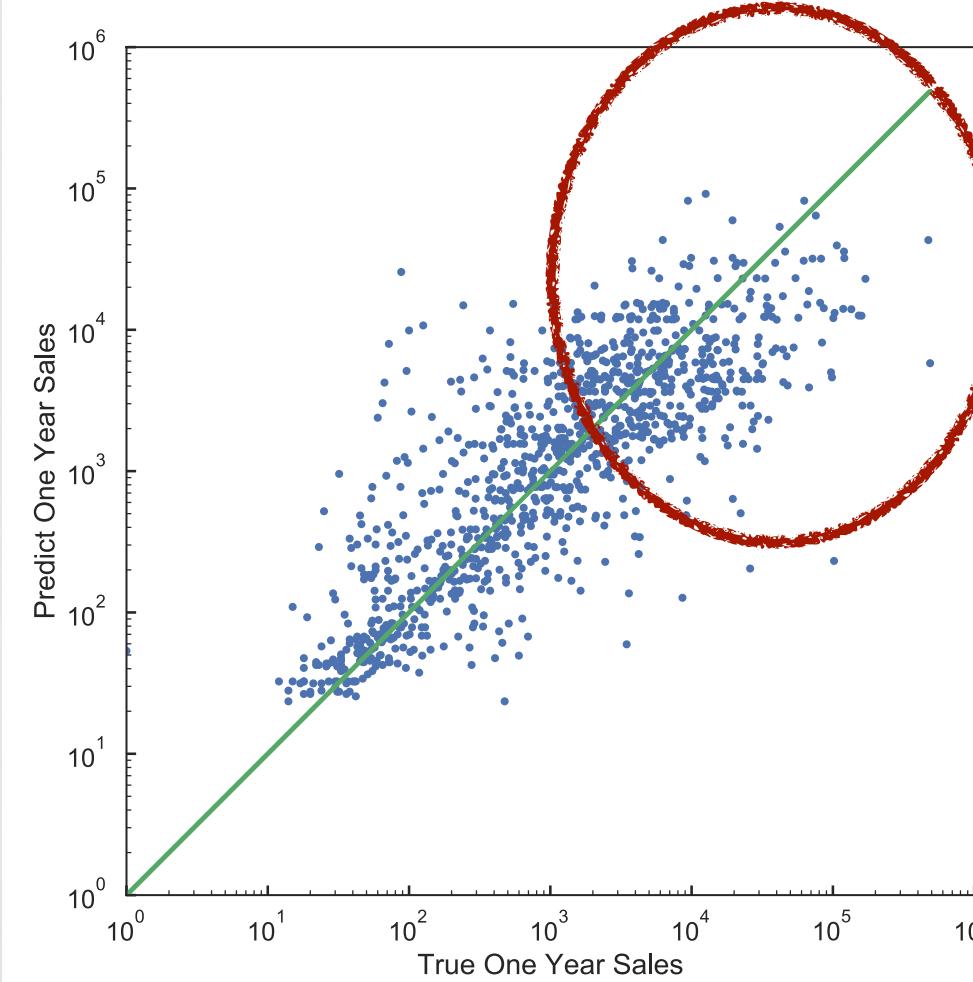
FAS-PIVOT



SpringRank

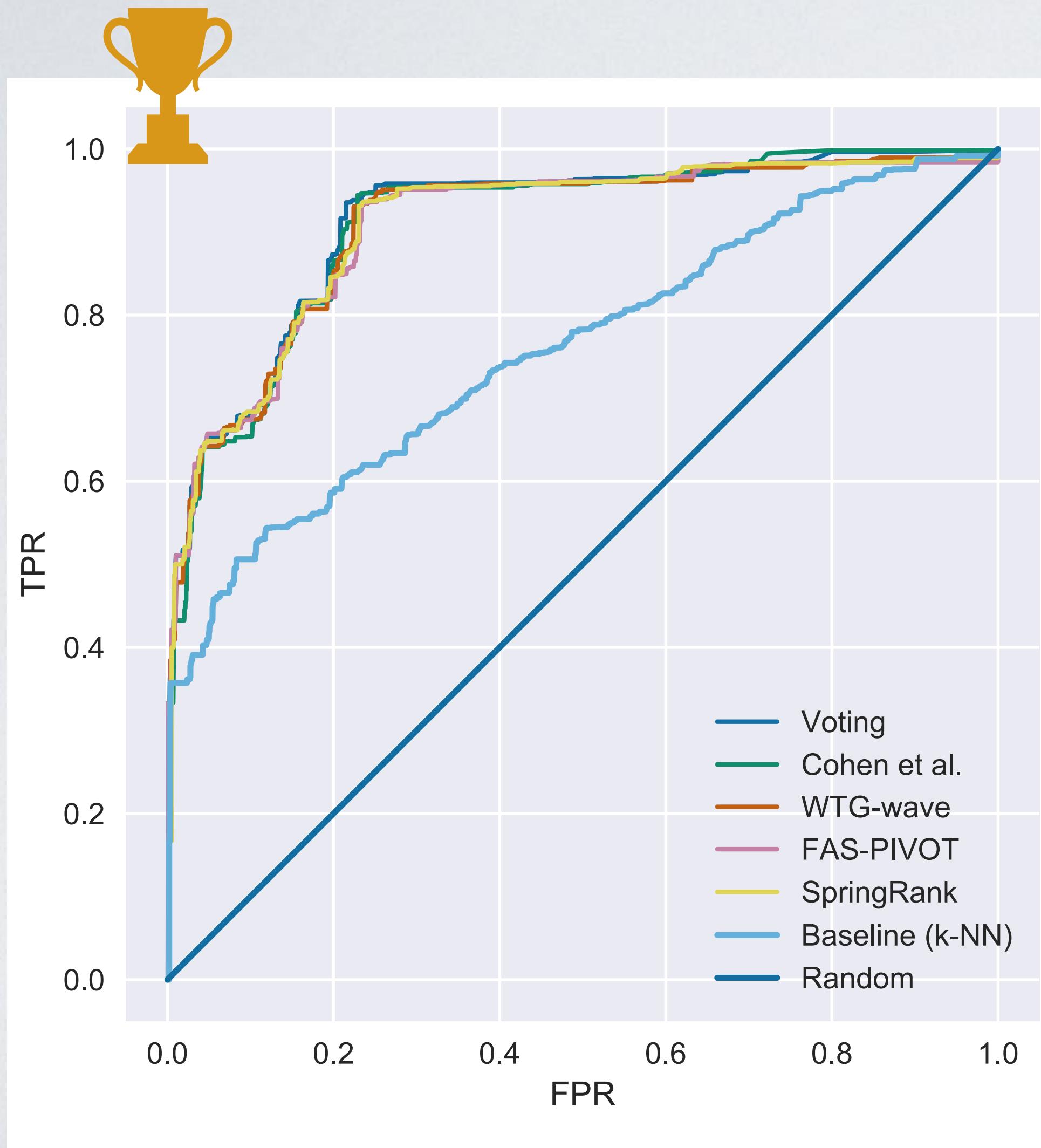


Fiction

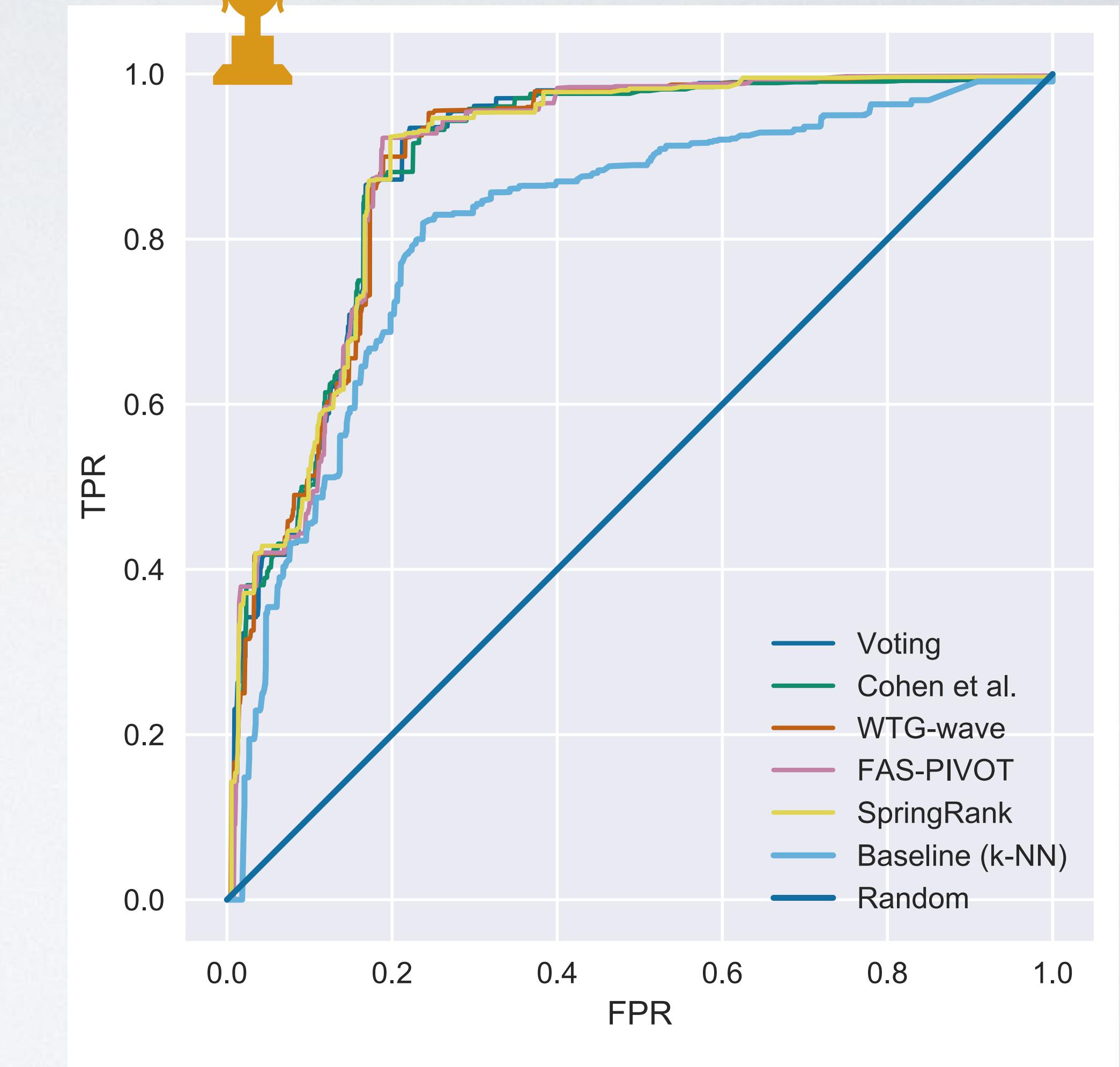


Biography

ROC



Fiction



Biography

AUC

Method	AUC on Fiction	AUC on Biography
Pairwise +	KNN	0.759
	Cohen et al.	0.892
	WTG wave	0.910
	Voting	0.915
	FAS-PIVOT	0.907
	SpringRank	0.908

CONCLUSION AND FUTURE WORK

- Introduced a two phase algorithm for learning to place: pairwise-comparison classifier + ranker
 - Reduce underprediction for heavy-tailed distributed attributes
- In phase II, Voting, WTG wave, FAS-PIVOT, and SpringRank perform similarly,
 - All are better than Cohen et al.
- Future work
 - Would higher order comparison make things better?

A photograph of a large crowd of people from a low angle, looking up. Many people have their hands raised, reaching towards a bright, overexposed light source at the top of the frame. The scene is dark, with the light creating a dramatic effect.

Questions?