Computing Choice

Devavrat Shah

Massachusetts Institute of Technology

Collaborators

A Ammar (NYC), G Chen(CMU/MIT), G Bresler (MIT), S Jagabathula (NYU), V Farias (MIT)
S Negahban (Yale), S Oh (UIUC), L Voloch (MIT)

Some scenarios

Winners (and losers)

ATP tennis, ICC cricket, Online gaming, EURO cup …

Using outcome of games between players / teams

Election

Primaries, Graduate admissions, Faculty hiring, Conference acceptance, …

Using votes / opinions of a population of …

Rank aggregation

Compute ranked order (with scores) of objects

Using partial / limited revealed preference data

Revealed preferences

Sports: outcome of pairwise games

Conference: star rating or ordering of papers

Election: top (few) candidate(s)

Some scenarios

Social recommendations

Amazon, Netflix, YouTube, E-Harmony (Tinder?!), ...

Using likes / dislikes of users / items, browse logs, ...

Resource planning

Transportation, Retail inventory, ...

Using information such as surveys, transactions, activity logs, ...

Multiple rankings

Compute dominant rankings of objects with their prominence

Using partial / limited revealed preference data

Revealed preferences

Amazon: purchases, reviews and ratings of products

Retail inventory: browses, purchases, . . .

Transportation: commute logs, survey responses, . . .

Data, Decision

Revealed preference data

Bag of pair-wise comparisons

Portugal defeated France : Portugal > France

Oleanna *****, Cuchi Cuchi ****: Oleanna > Cuchi Cuchi

Browsed A, B and bought A : A > B

Decision

Rank aggregation

Rank order all objects along with their scores

Multiple rankings

Dominant rankings along with their prominence

(Choice) Model

Model of preference or choice

Distribution over rankings / permutations of objects

Revealed preferences are partial ordering

generated per this model

0.25 A > B > C

0.75 B > C > A

A > C

C > A

B > A

B > C

Data, Model and Decision

Data

Bag of pair-wise comparisons

Learn the (choice) model

From observations

(Implicitly or explicitly)

Make decision

Rank aggregation:

Find top ranking for population

Personalization:

Find top ranking for an individual given her/his preference history

Rank Aggregation

A. Ammar and D. Shah (Sigmetrics 2011)

S. Negahban, S. Oh and D. Shah (NIPS 2012, OR 2016)

A Brief History

If choice model is known

- Axiomatic impossibility
  - Celebrated result of Ken Arrow (1951)

Some algorithms using choice model

- Kemeny optimal: minimizes dis-agreements
  - Satisfies extended Condorcet criteria
  - NP-hard + 2 - approx via network flow (Dwork et al 2001)

Useful notion, but only approximation is computationally feasible

A Brief History

If choice model is known

- Axiomatic impossibility
  - Celebrated result of Ken Arrow (1951)

Some algorithms using choice model

- Borda count: average position as the score
  - Simple (adding numbers)
  - Has useful axiomatic properties (Young 1974)

Can we adopt it when pair-wise comparison as observations?

Borda Count from Pair-wise Comparisons

Some formalism

N objects: teams, players, candidates, students, papers, …

Data: collection of pair-wise comparisons between N objects

C(i,j): number of comparisons with i > j

Algorithm (Ammar and Shah 2011)

Compute p(i,j) = C(i,j) / (C(i,j) + C(j,i)) for all pairs (i, j)

fraction of times i defeats j

Score of object i: $\mathrm{S(i)} = \sum_{\mathrm{j}} \mathrm{p(i,j)}$

how often i defeat others

Borda Count from Pair-wise Comparisons

Algorithm (Ammar and Shah 2011)

Compute p(i,j) = A(i,j) / (A(i,j) + A(j,i)) for all pairs (i, j)

Score of object i: $\mathrm{S(i)} = \sum_{\mathrm{j}} \mathrm{p(i,j)}$

how often i defeats others ?

Theorem (informal statement) (Ammar and Shah 2011)

If p(i,j) is exactly known for all pairs i,j then it is equivalent to

Borda count with the entire choice model known!

In reality, we may know p(i,j) for a small fraction of pairs, and in a noisy form

Noisy, Partial Pair-wise Comparisons

Various approaches have been proposed from different perspectives

… Saaty '01, Dwork et al '01, Caramanis '11, Nowak et al '11, '13, '15, Fernoud '11, Duchi et al '12, …

We'll discuss

Rank centrality: a "natural" algorithm

An experiment

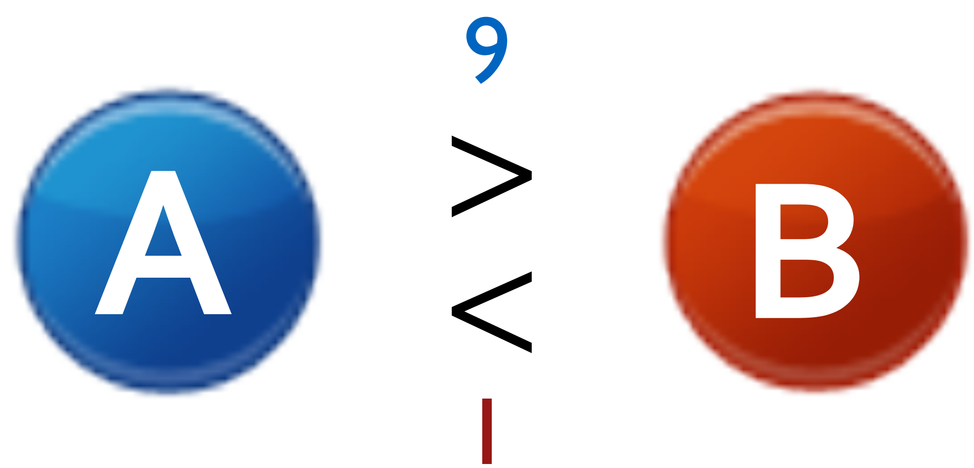
Properties

Rank Centrality

(Negahban, Oh, Shah '12, '16)

Let N = 2: Two Players

Consider out-come of 10 matches



score 0.9

score 0.1

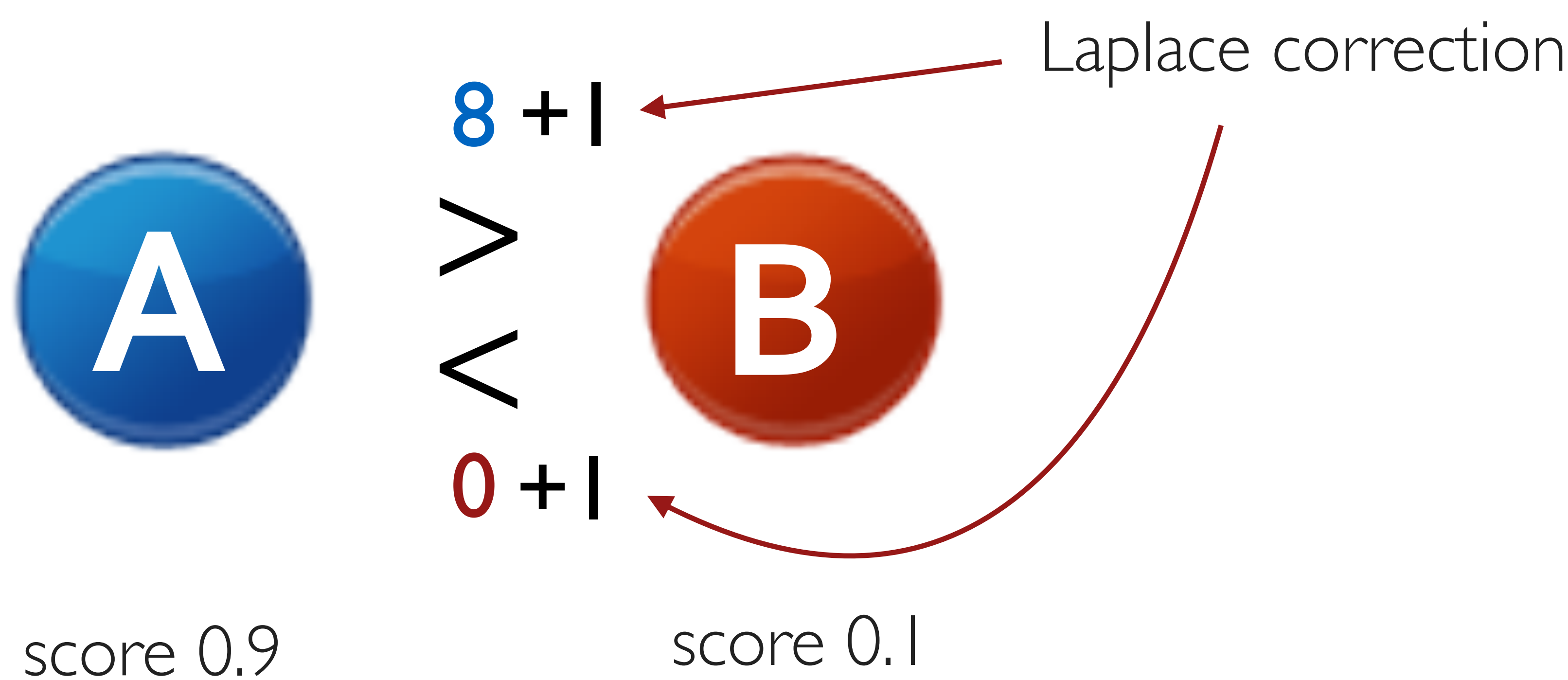
aren't we taking observations too seriously?

Rank Centrality

(Negahban, Oh, Shah '12, '16)

Let N = 2: Two Players

What if one team has lost it all - does it deserve score of 0?

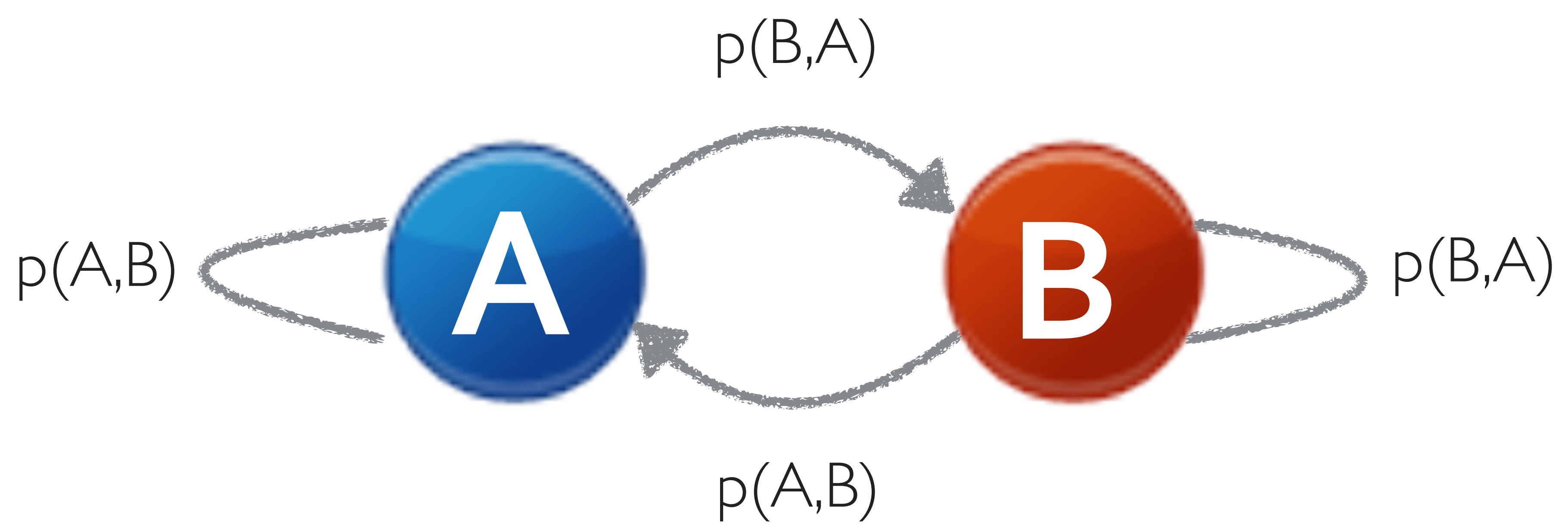


Rank Centrality

(Negahban, Oh, Shah '12, '16)

Let N = 2: Two Players

Algorithm: scores are stationary distribution of the Markov Chain



where, p(A,B) = (C(A,B)**+1**) / (C(A,B) + C(B,A)**+2**)

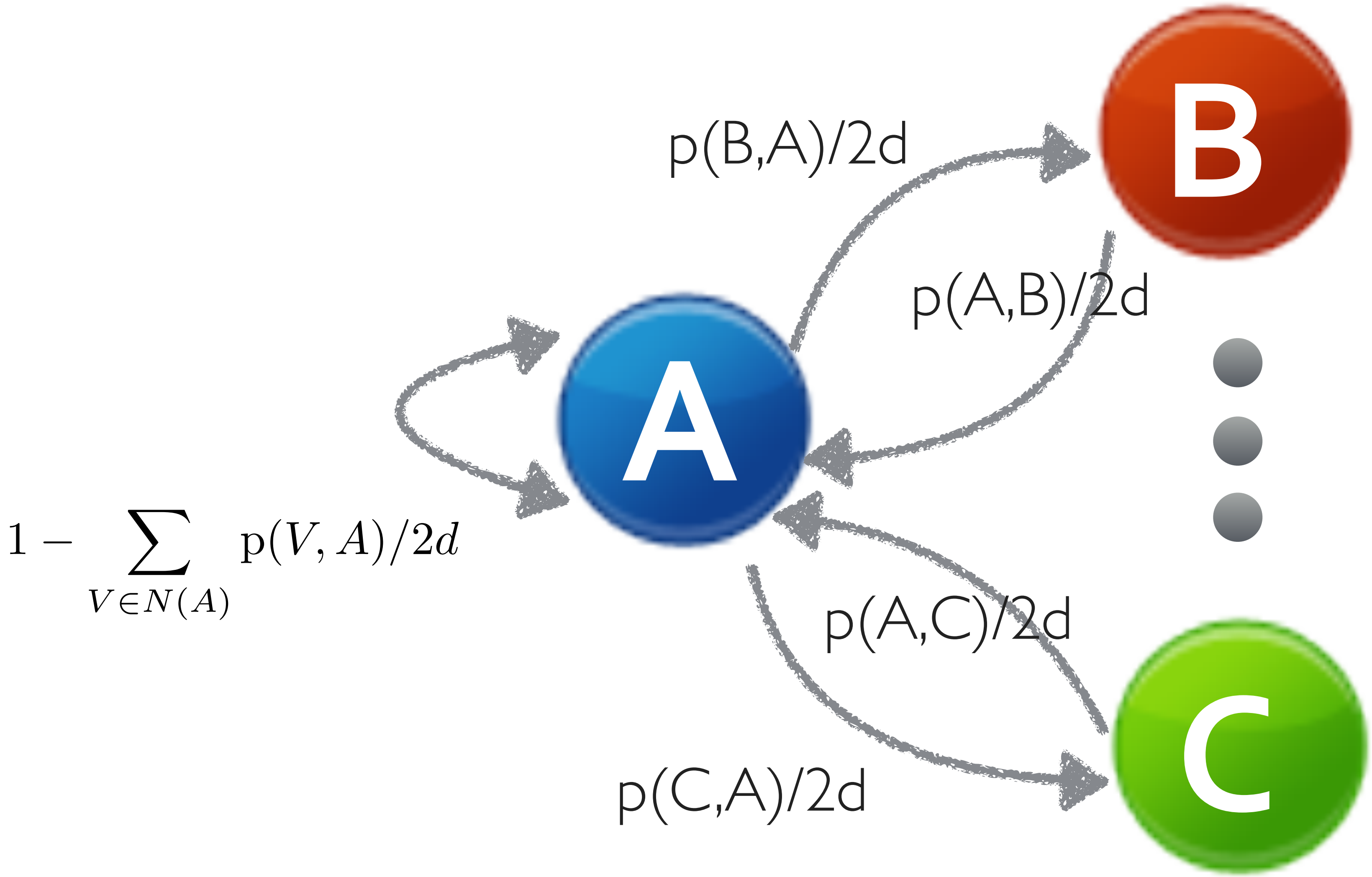
p(B,A) = 1-p(A,B) = (C(B,A)**+1**) / (C(A,B) + C(B,A)**+2**)

Rank Centrality

(Negahban, Oh, Shah '12, '16)

For general N

Algorithm: scores are stationary distribution of the Markov Chain



where, d = max vertex degree of "comparison graph"

Rank Centrality: ICC Cricket Ranking

ODI ranking	Team	Win	Loss	Tie	deg	Rank Centrality $\epsilon^a = 0$		Rank Centrality $\epsilon = 1$	
1	England	35	23	2	10	0.1526	3	0.0957	1
2	South Africa	29	14		11	0.1794	2	0.0943	2
3	India	47	26	3	11	0.1317	4	0.0911	3
4	Australia	45	26	1	13	0.1798	1	0.0900	4
5	Sri Lanka	43	34	1	12	0.1243	5	0.0801	5
6	Parkistan	35	30		13	0.0762	6	0.0715	6
7	West Indies	22	32	1	12	0.0396	7	0.0546	9
8	Bangladesh	17	34		11	0.0320	9	0.0500	10
9	New Zealand	19	31		10	0.0354	8	0.0466	12
10	Zimbabwe	13	27		11	0.0307	10	0.0481	11
11	Ireland	17	15		13	0.0124	11	0.0561	8
12	Netherlands	8	16		10	0.0017	13	0.0432	14
13	Kenya	4	17		10	0.0007	14	0.0367	15
14	Afghanistan	10	10		7	0.0005	15	0.0435	13
15	Scotland	9	6		7	0.0029	12	0.0620	7
16	Canada	5	17		11	0.0003	16	0.0365	16

Rank Centrality: Property

(Negahban, Oh, Shah '12, '16)

Learns parameters of a popular choice model!

And does so utilizing minimal samples

Multinomial Logit Model

Thurstone 1927, Luce 1945, Bradley-Terry 1960s, McFadden 1965, …

Each object has as associated positive parameter

w(i) for object i

Ranking is sampled by sampling elements in positions 1, 2, …iteratively

proportional to their parameter values

Multinomial Logit Model

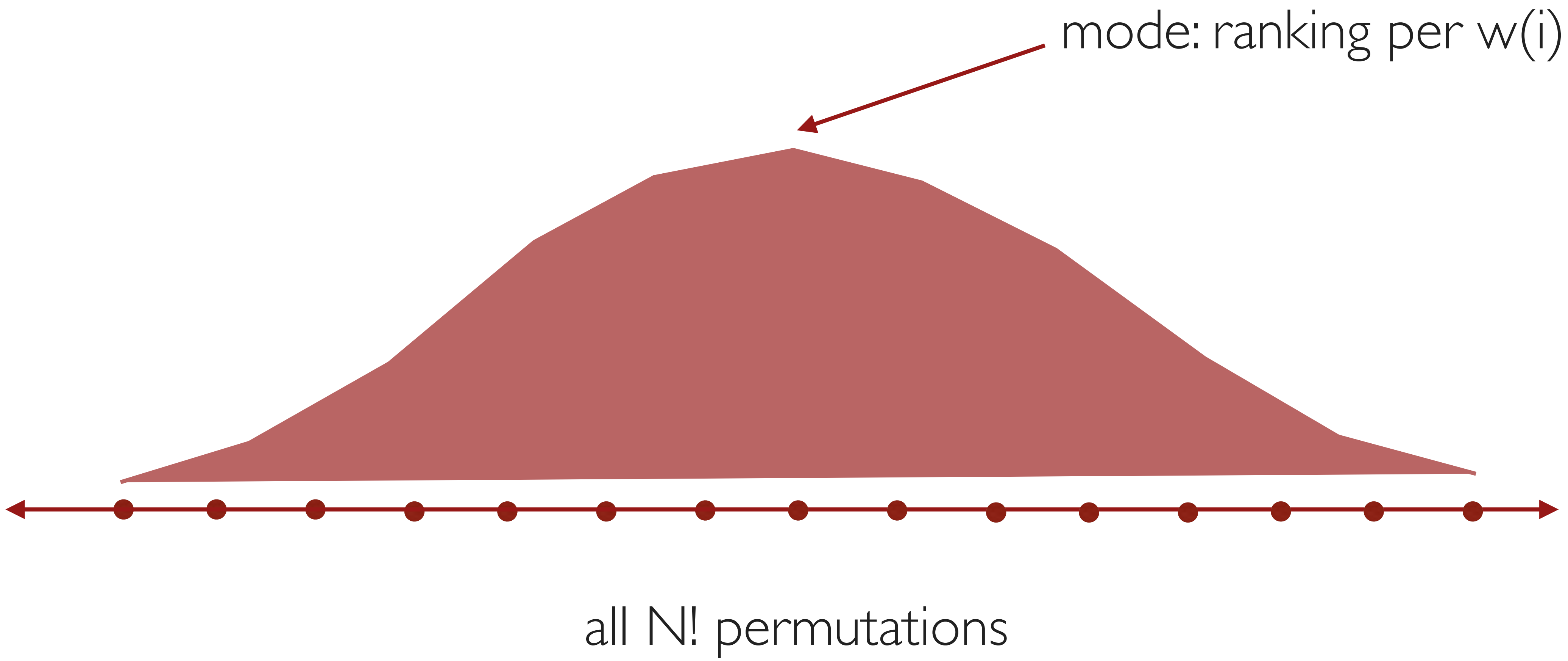
Examples

uniform distribution

w(i) = 1/N for all i

delta distribution (single permutation)

w(1) >> w(2) >> w(3) … (1 > 2 > 3 …)



Rank Centrality: Property

(Negahban, Oh, Shah '12, '16)

Theorem (informal statement)

Rank centrality learns parameters w(i)

as long as comparison graph is connected

Normalized error in norm of learnt parameter vs true parameter

$$\frac{\|\hat{\mathrm{w}} - \mathrm{w}\|}{\mathrm{w}} \leq C \sqrt{\frac{\log N}{k\, d}}$$

max degree of graph (d)

comparisons per pair (k)

where C depends on

inverse of spectral gap of normalized Laplacian of comparison graph

If graph can be designed, use a *spectral* expander

will require O(N log N) comparisons

Personalization

S. Jagabathula and D. Shah (Info Th Trans 2011)

A. Ammar, S. Oh, D. Shah and L. Voloch (Sigmetrics 2014)

S. Oh and D. Shah (NIPS 2014)

G. Bresler, D. Shah and L. Voloch (Sigmetrics 2016)

Personalization

Data

Partial preferences across population

And a specific individual's partial preferences

Personalization for the specific individual

Find individual's (choice) model

given her/his history and population preference data

Compute ranking of objects using it

It's personalized rank aggregation

Personalization

Compute individual's (choice) model: A Bayesian view

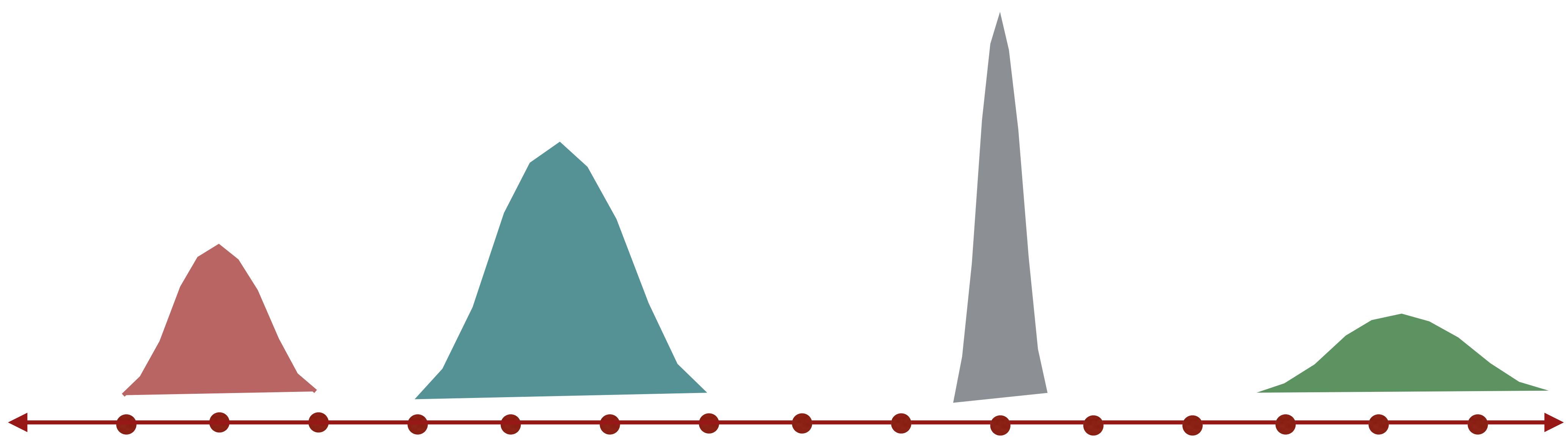
Use population data to learn (choice) model

Individual's model is conditional distribution (over permutations)

given her/his history

Population (choice) model

A generic view: mixture of MNL models



Personalization

Compute individual's (choice) model: A Bayesian view

Use population data to learn (choice) model

Individual's model is conditional distribution (over permutations)

given her/his history

Individual's (choice) model

Identify the mixture component (it's an MNL model)

use it to solve rank aggregation

Equivalently, estimate

$P(i > j \mid \text{history})$ for all pairs of i, j

Personalization

Key problems

Learn mixture MNL from data

Identify individual's mixture component

Or directly estimate

$P(i > j \mid \text{history})$ for all pairs of i, j

Learning mixture of MNL

Data

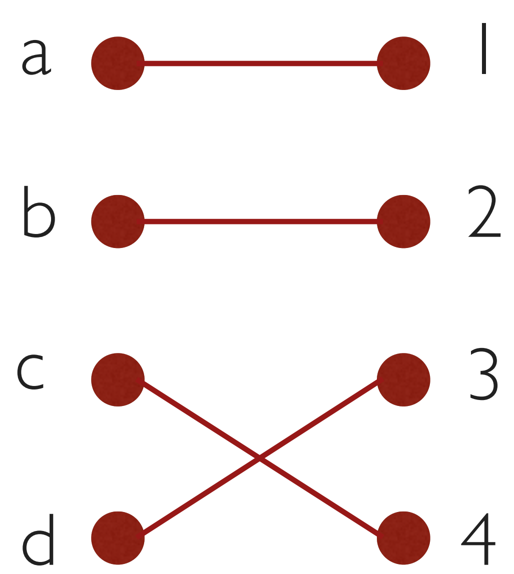
pair-wise comparisons

ideally, exact pair-wise marginals p(i,j) *probability that i defeats j*

Learning mixture of MNL

Impossibility to learn mixture of 2 components with N = 4

even from exact pair-wise marginals

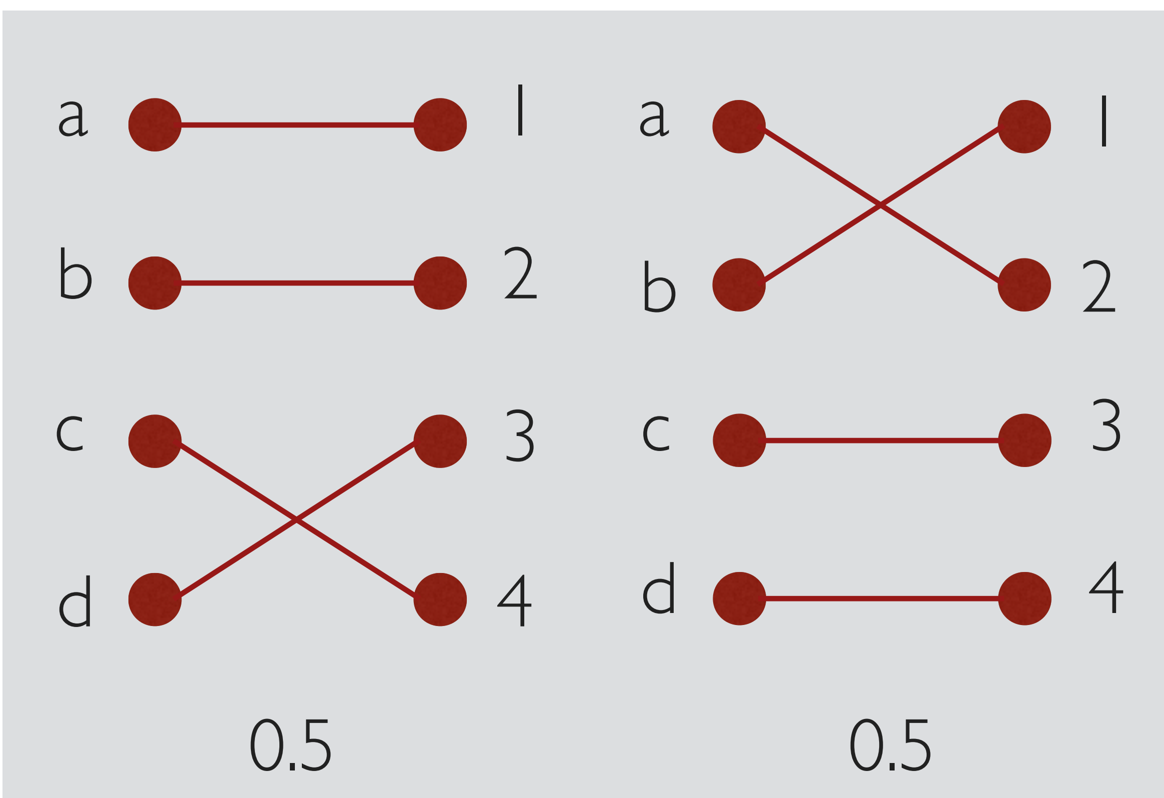


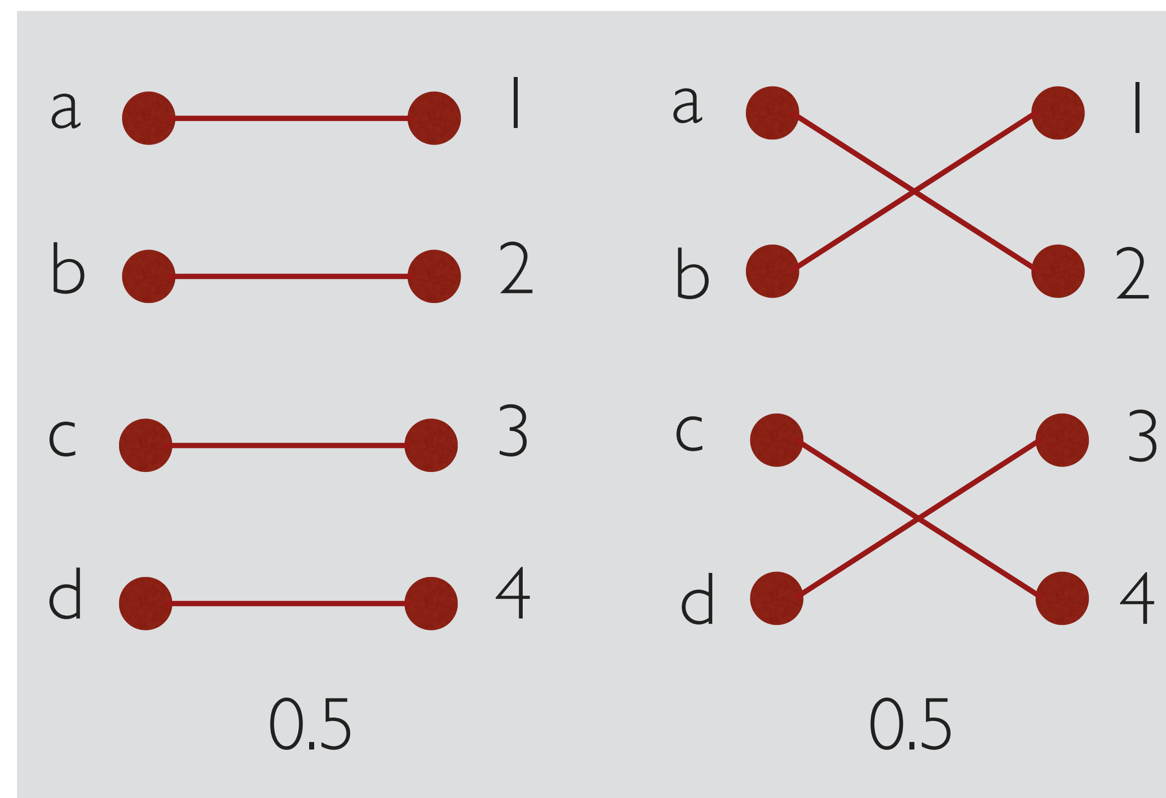
$=$ a > b > d > c

Learning mixture of MNL

Impossibility to learn mixture of 2 components with N = 4

even from exact pair-wise marginals





Can not differentiate from the above two cases using pair-wise marginals

Learning mixture of MNL

Theorem (informal statement) (Ammar, Oh, Shah, Voloch '14)

All mixture distributions of N/2 components can not be learnt accurately even from all the log N-wise marginals.

Is there any hope with additional assumption?

with pair-wise marginals?

Learning mixture of MNL

Theorem (informal statement) (Jagabathula, Shah '11)

When mixture components are drawn at random, mixtures with up to log N components can be learnt using exact pair-wise marginals

extends for k-wise marginals

but requires exact knowledge, what about noise?

Learning mixture of MNL

Theorem (informal statement) (Oh, Shah '14)

When there is sufficient "gap" between mixture components, then using noisy pair-wise marginals, mixture model can be learnt with fixed number of mixture components using poly(N) comparisons.

spectral algorithm: uses higher order moments to learn the parameters

but "sufficient gap" is essential

is "gap" really needed?

well, yes (or s'thing like it) for learning mixture MNL

Personalized Ranking

Recall

Goal is to estimate

P(i > j | history) for all pairs of i, j

If two components are "too close", both might have similar

P(i > j | history) for all pairs of i, j

Therefore

it may be possible to do without explicit "gap" condition

Personalized Ranking

A recent result (Bresler, Shah, Voloch '16)

To estimate binary preferences (likes / dislikes) accurately

length of history need to scale with the "dimension" of the distribution

"dimension" is less stringent than the "gap" requirement

e.g. "dimension", for k mixture components, is log k independent of "gap"

when "dimension" is constant

history required depends only on accuracy desired

extending it for estimating P(i > j)?

Personalized Ranking

Data:

Yelp Boston Data of ~1K restaurants

Task:

For an individual predict whether i > j or i < j

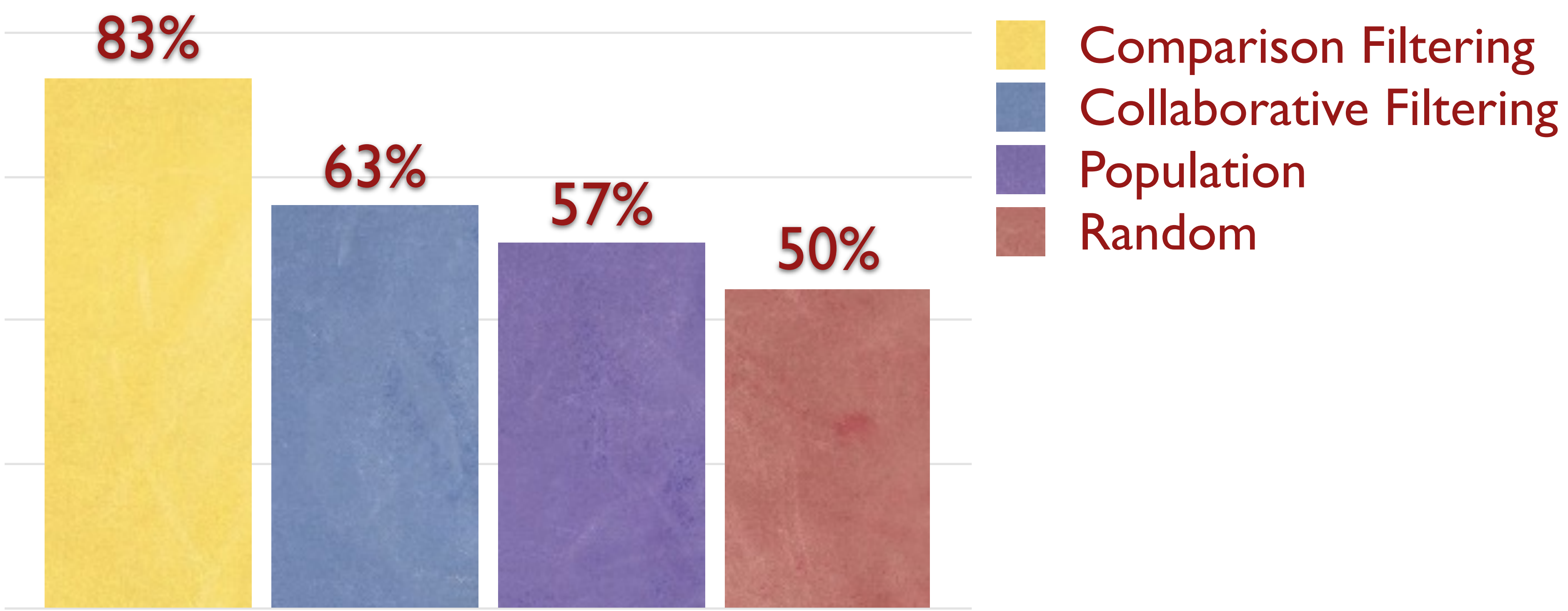
Personalized Ranking

Data:

Yelp Boston Data of ~1K restaurants

Task:

For an individual predict whether i > j or i < j



Related Works

Discrete Choice Modeling

Thurstone (1927)

Samuelson (1938)

Luce (1946)

William-Daly-Zachary (1970s)

McFadden (1970s), McFadden-Train (2001)

Dueling Bandits

Yue-Broder-Kleinberg-Joachim (2012)

Zoghi et al (2013), Urvoy et al (2013), Ailon et al (2014)

Jamison-Kataria-Deshpande-Nowak (2015)

Related Works

Noisy sorting, Rank aggregation and Tournaments

Adler-Gemmell-HarcholBalter-Karp-Kenyon (1994)

Braverman-Mossel (2008)

Ailon-Charikar-Newmann (2004), Alon (2006)

Rajkumar-Agarwal (2014)

Soufiani-Parkes-Xia (2014)

Shah-Wainwright (2015, 2016)

Learning mixtures

Avasthi-Blum-Sheffet-Vijayaraghavan (2014)

Oh-Thekumparampil-Xu (2015), Negahban (2015)

In Practice: Few Examples

Rankings, Winners

TrueSkill rating

Collecting Opinion

allourideas.org

Policy / Transportation

McFadden (1970s-), Ben Akiva (1980s-)

Pricing, Revenue Management and Assortment Optimization

Talluri-Van Ryzin (2004)

Farias-Jagabathula-Shah (2011) →

celect

http://celect.com

Summary

Choice and computation

- Ubiquitous in social scenarios
- Distribution over permutation is a powerful model of choice
- However, historically it has been largely ignored because
  - computational difficulties
  - lack of data

This talk

- Addressing computational challenge for decision making with choice model
- Glimpse of exciting work across disciplines of CS / EE / Stats / OR / Econ