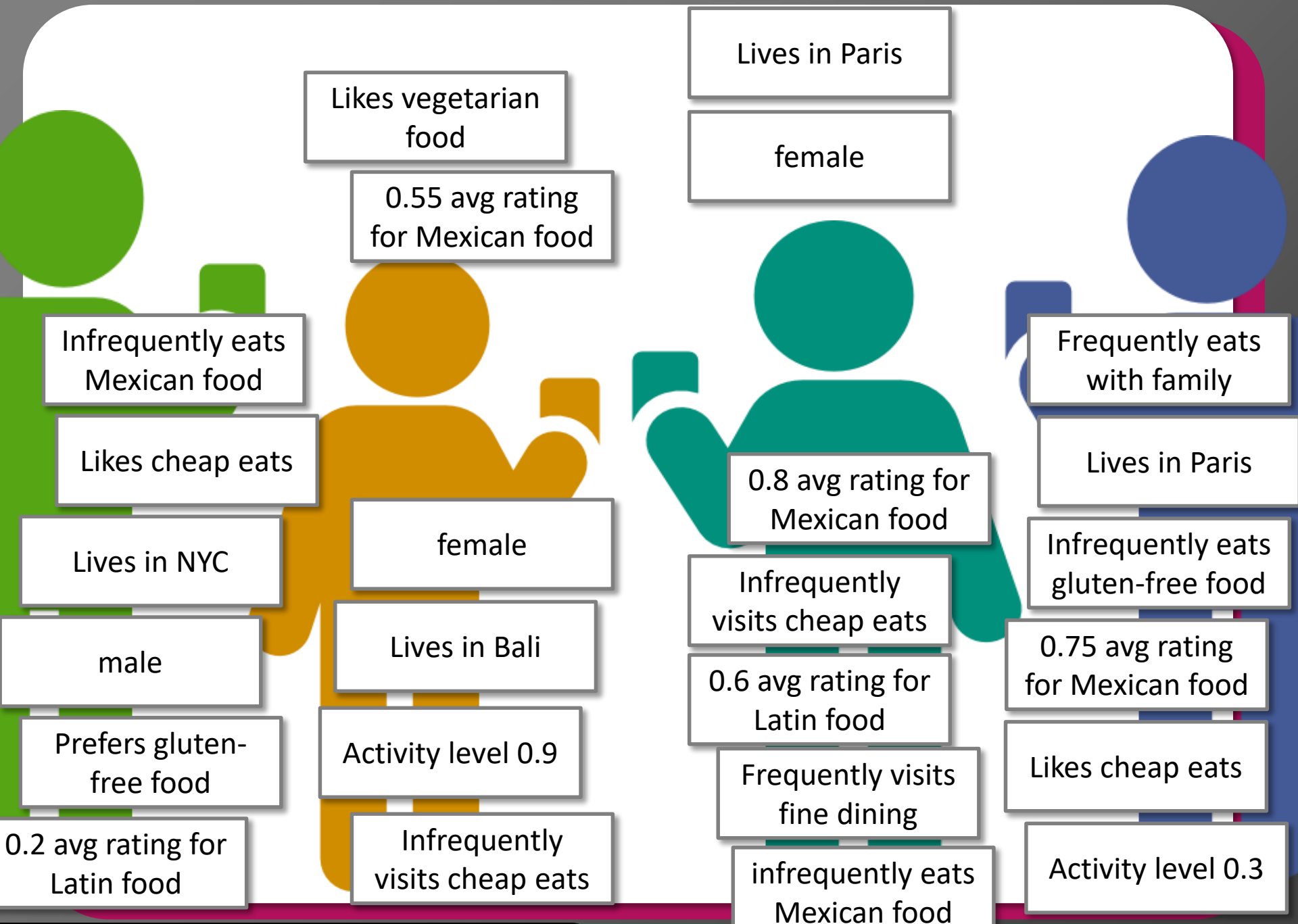
Diverse User Selection for Opinion Procurement

Yael Amsterdamer **Oded Goldreich**

Department of Computer Science

Tradition of Excellence

Bar-Ilan University


Likes vegetarian food
0.55 avg rating for Mexican food
Infrequently eats Mexican food
Likes cheap eats
Lives in NYC
male
Prefers gluten-free food
0.2 avg rating for Latin food
female
Lives in Bali
Activity level 0.9
Infrequently visits cheap eats
Lives in Paris
female
0.8 avg rating for Mexican food
Infrequently visits cheap eats
0.6 avg rating for Latin food
Frequently visits fine dining
infrequently eats Mexican food
Frequently eats with family
Lives in Paris
Infrequently eats gluten-free food
0.75 avg rating for Mexican food
Likes cheap eats
Activity level 0.3

Goals

1. Define user groups given high-dimensional data (on a scale)
2. Define “representativeness” measure of a user subset
   - Every user represents many groups (overlaps)
   - Many groups of different importance
3. Efficient computation of a good user subset
   - Minimizing manual effort
4. Optional customization

Previous work

- Distance-based versus coverage-based diversity
- Diversity contexts:
  - Survey research — Statistical soundness; Heavy manual load; Reduced dimensionality
  - Crowdsourcing — Reduced dimensionality; (Distance-based)
    - Team/group formation
  - Recommender systems — Distance-based
  - Information retrieval — Reduced dimensionality

Our Solution – Brief Overview

Profile Model

User properties with scores in [0,1]

Grouping System

1-d clustering per property
users are assigned to multiple groups

Weighted Prioritization Framework

Each group is assigned a **cover rate** and a **weight score**
Natural options: uniform, proportional to size

Group cover scores are aggregated to a total score

Explanations and Customizations

Explanations: groups, selected users, group coverage

Customization: must(not) have groups, priority coverage, by weight manipulation

Max-Diversity Problem

Maximize $\sum_{G \in \mathcal{G}} \text{wei}(G) \cdot \min\{|U \cap G|, \text{cov}(G)\}$ for a selected subset U such that $|U| \leq B$

Not penalizing over-representation, allowing under-representation

- **NP-complete in B**
- A greedy algorithm achieves a $\left(\mathbf{1} - \frac{\mathbf{1}}{\boldsymbol{e}}\right)$**-approximation** in time

$$O\left(B \cdot \max_{G \in \mathcal{G}} |G| \cdot \max_{u \in \mathcal{U}} \text{groups}(u)\right)$$

- Further practical optimizations via linked data structures

Explanations

- Why user X was selected?
  - Top-k weighted groups represented by this user
- Importance of group Y?
  - Group name is indicative – derived from profile data
  - Group weights
- Visualizations:

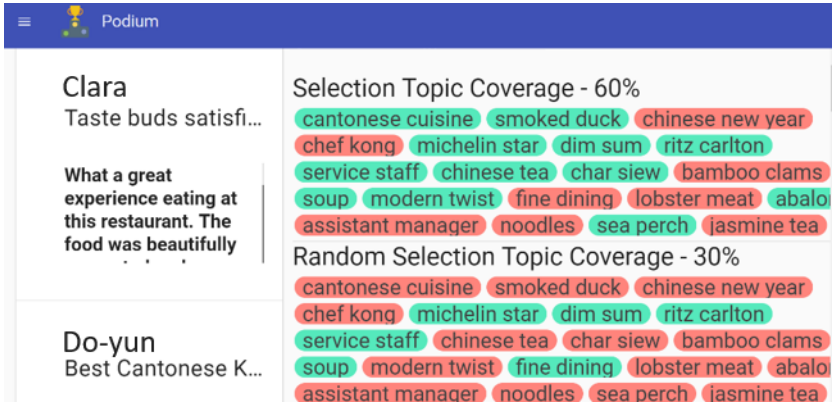
On hover – menu for customization

Explanation-based Cutomization

- Prioritize covering selected groups
  - Manually selected 
  - Ties are broken by other properties
- Target a better fit of a specific dimension
  - I.e., prioritize covering its groups
    - E.g., low, medium and high rating for Mexican food
- Focus on/filter out groups
  - Hard constraints

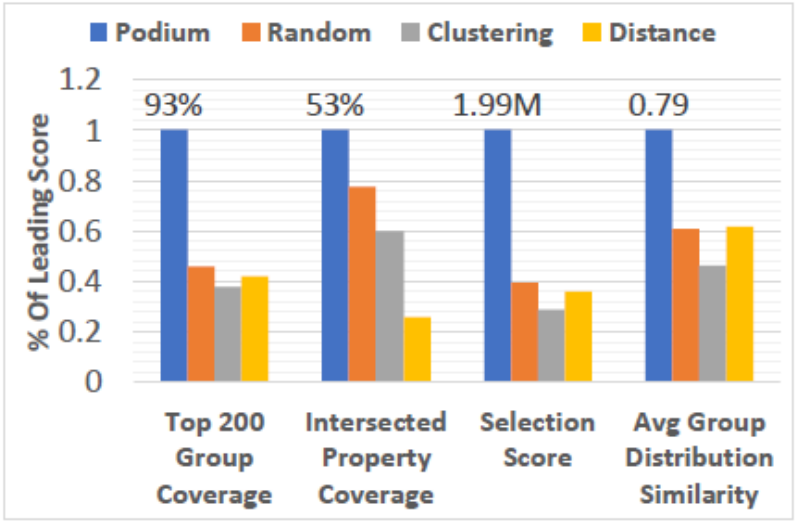
Evaluation

- Comparison to methods from IR, unsupervised learning
- Metrics:
  - Inherent quality of selected subset
    - Coverage of underlying (top-k) groups
    - Coverage of multi-dimensional meaningful groups, identified by mining
    - Similarity of distribution to source population
  - Diversity of task results
  - Client satisfaction

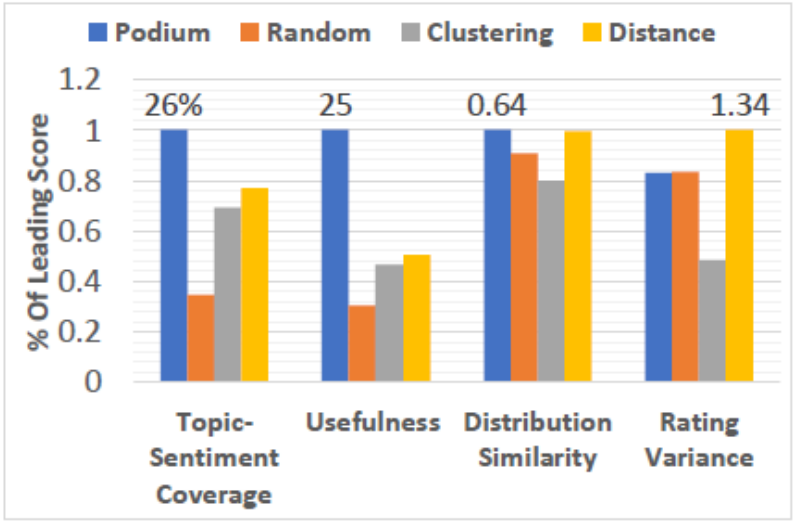


Experimental study

User repositories: yelp, TripAdvisor



Intrinsic diversity (yelp)



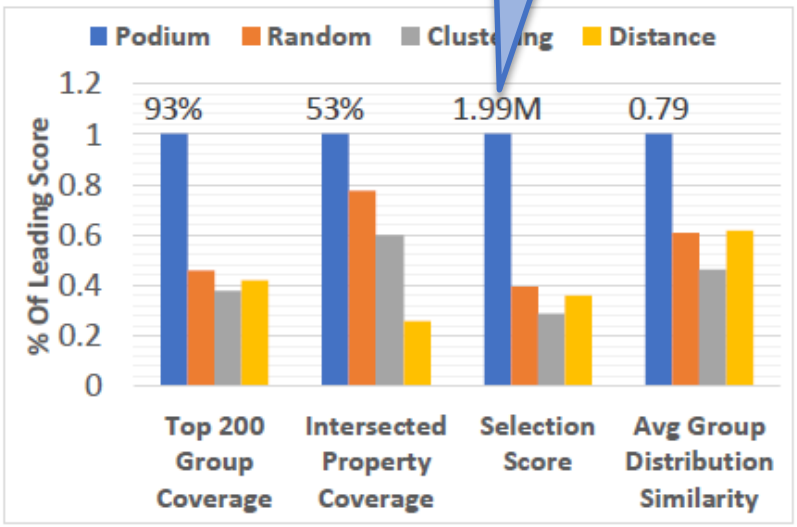
Opinion diversity (yelp)

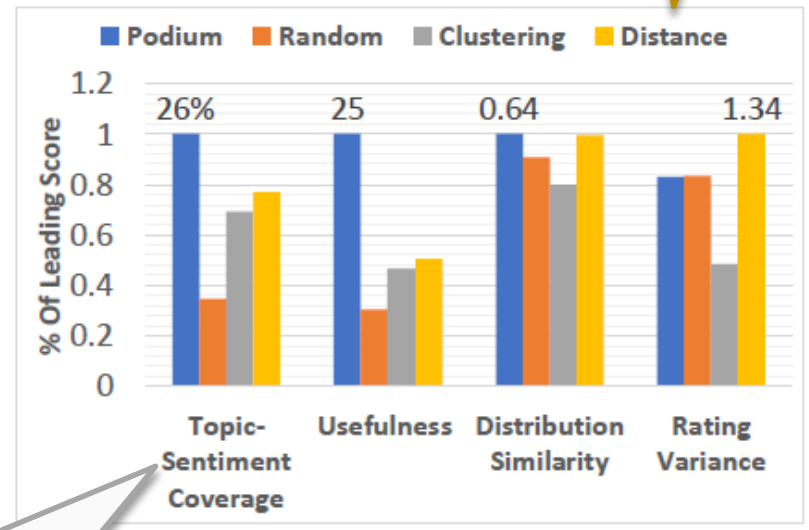
Experimental study

Near-optimal in practice

Wu et al 2015

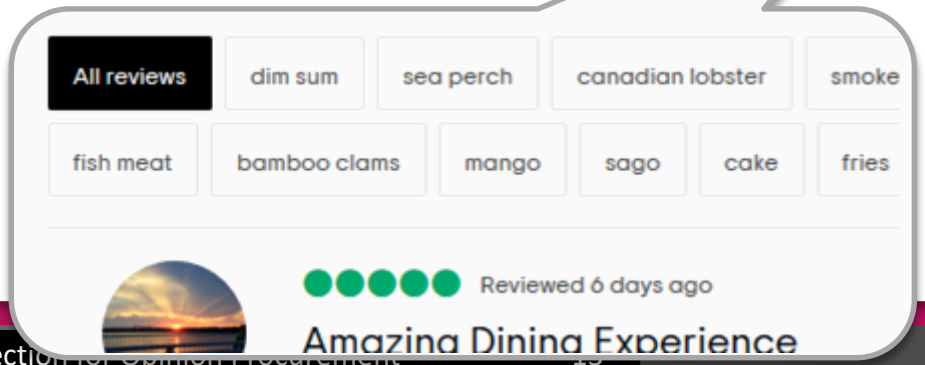
User reposito es: yelp, TripAdvisor





Intrinsic diversity (yelp)

Opinion diversity (yelp)

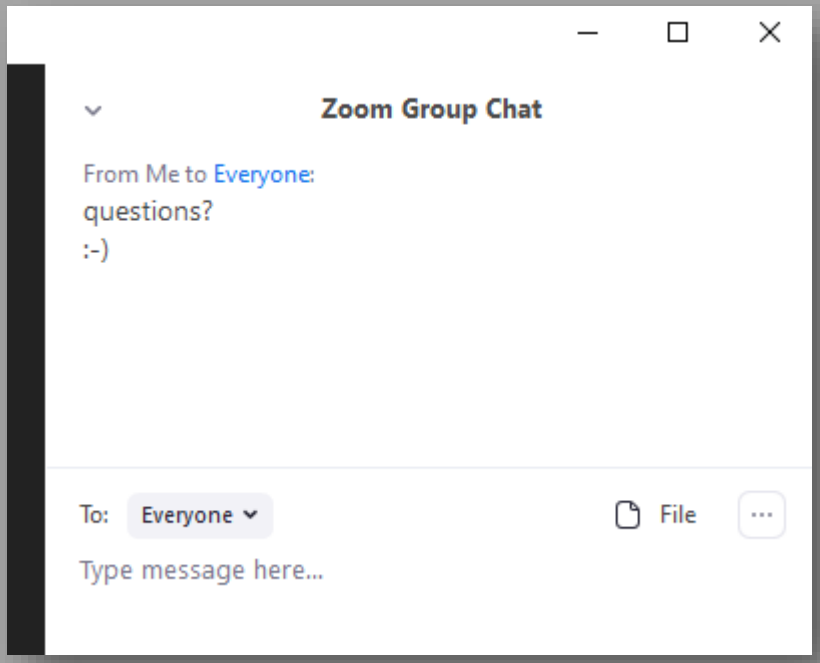


Measured by real opinions of selected users, not used in selection

Summary

- ✓ Weighted Framework for diverse user selection
- ✓ Population grouping system
- ✓ Diversity definition and efficient algorithm
- ✓ Explanations and customization
- ✓ Experimental studies indicate good balance between variance and coverage

Thank you



Bar-Ilan University