

DiRec: Diversified Recommendations for Semantic-less Collaborative Filtering

Rubi Boim

Tova Milo

Slava Novgorodov

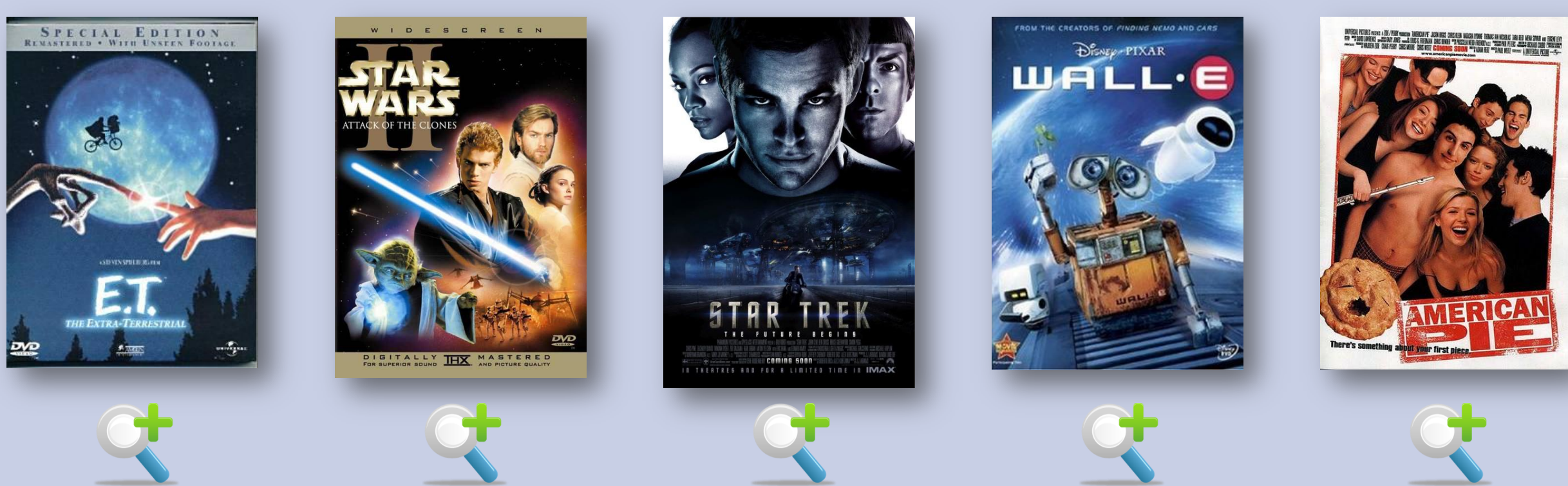
Goal

Diversify recommendations of CF systems

Top-K relevance (default)



A more Diverse and Wider list with "Zoom-in"



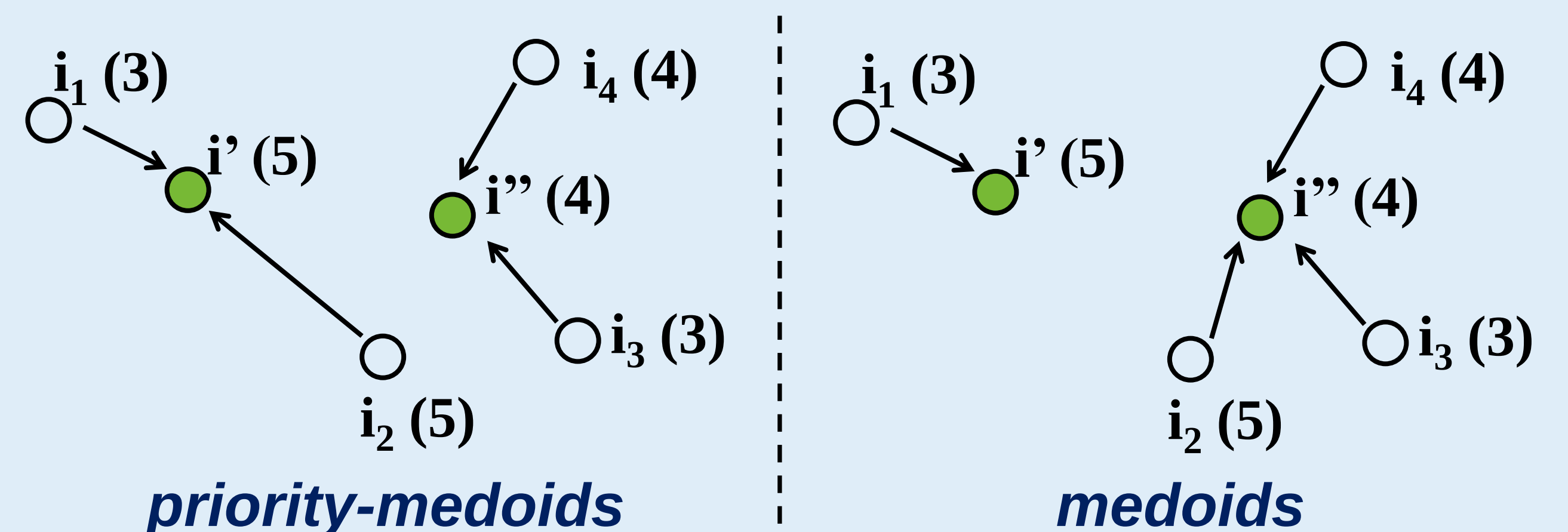
Priority Medoids

A generalization of the standard medoids:

→ Minimize the distances between the items and their medoids



→ Medoids must have higher priorities than their cluster members



→ Proven to be NP-Hard

Collaborative Filtering

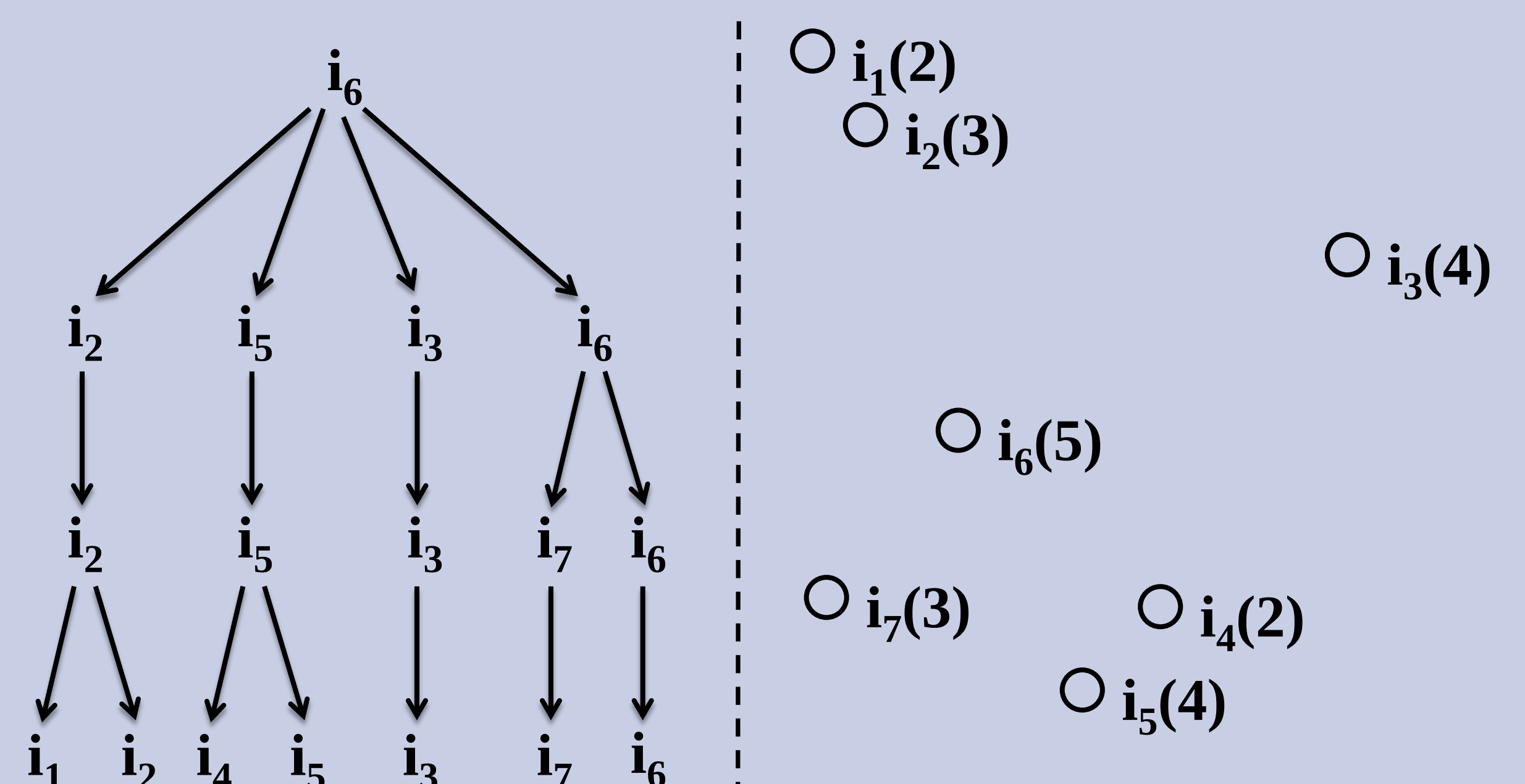
Predicts ratings based on how similar users rated similar items

- User U rated item I "4 stars"
- Each item is viewed as a vector of ratings
- Pearson's correlation measure the distances
- Naïve approach presets the top-k predictions

Approx. using Priority Cover-Tree

A generalization of the standard cover-tree:

- Nesting
- Separation (all nodes are at least $1/2^l$)
- Covering (children are within $1/2^l$)
- NEW** → Priority (children have lower rating)



Difficulties

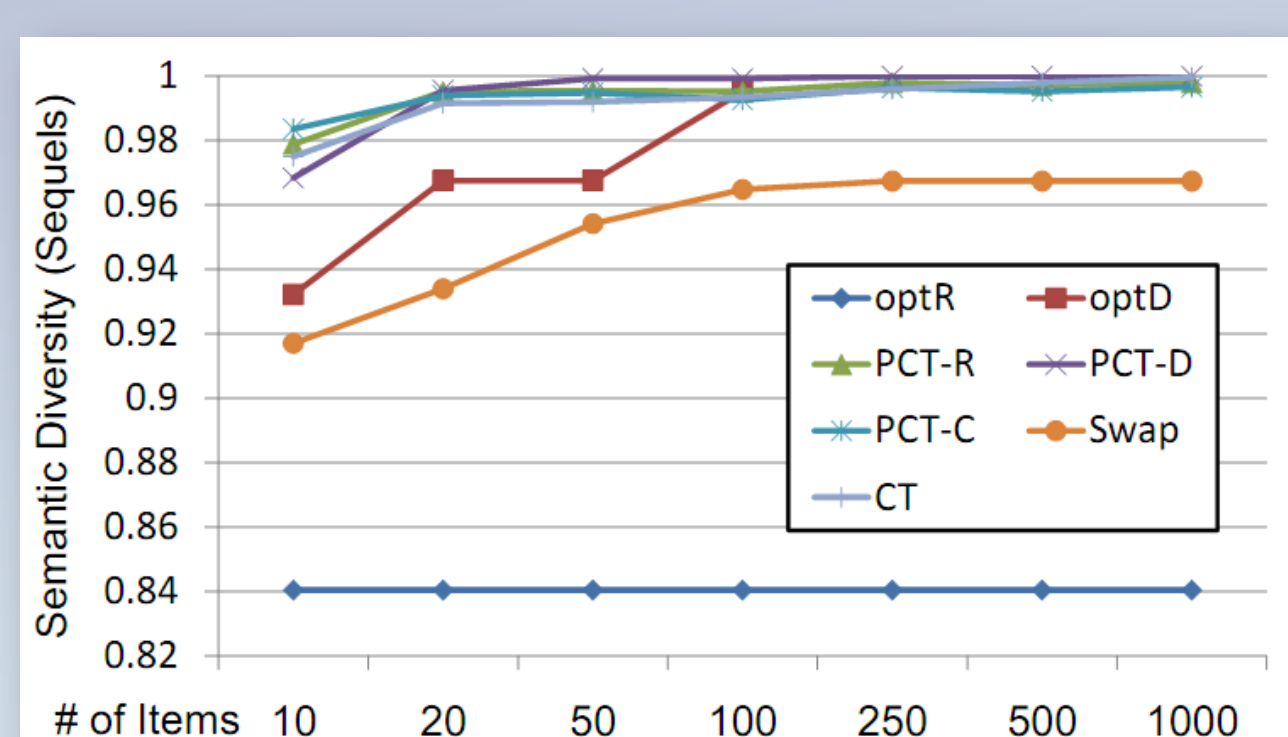
Measuring diversity

- Semantic-less environment
- When exists, choosing the right features

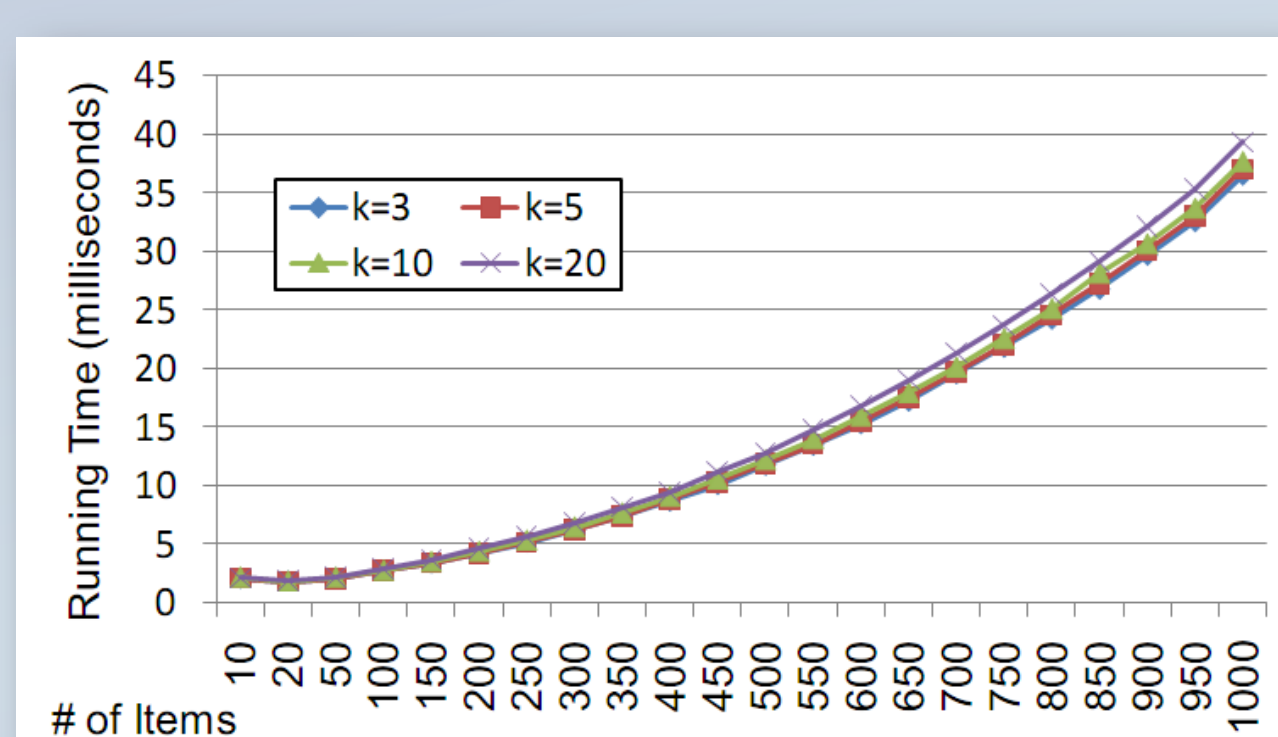
Balancing ranking and diversity

- Previous solutions use threshold/weights
- Problematic to choose (context dependent)
- May need to be refined when data changes

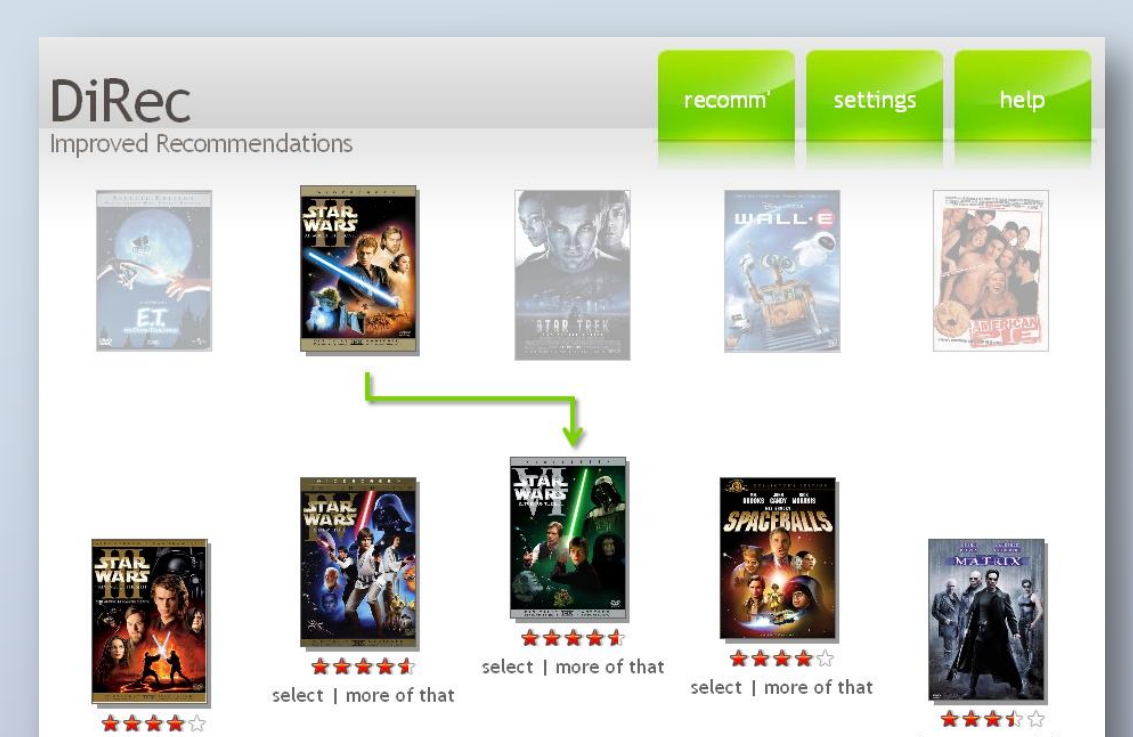
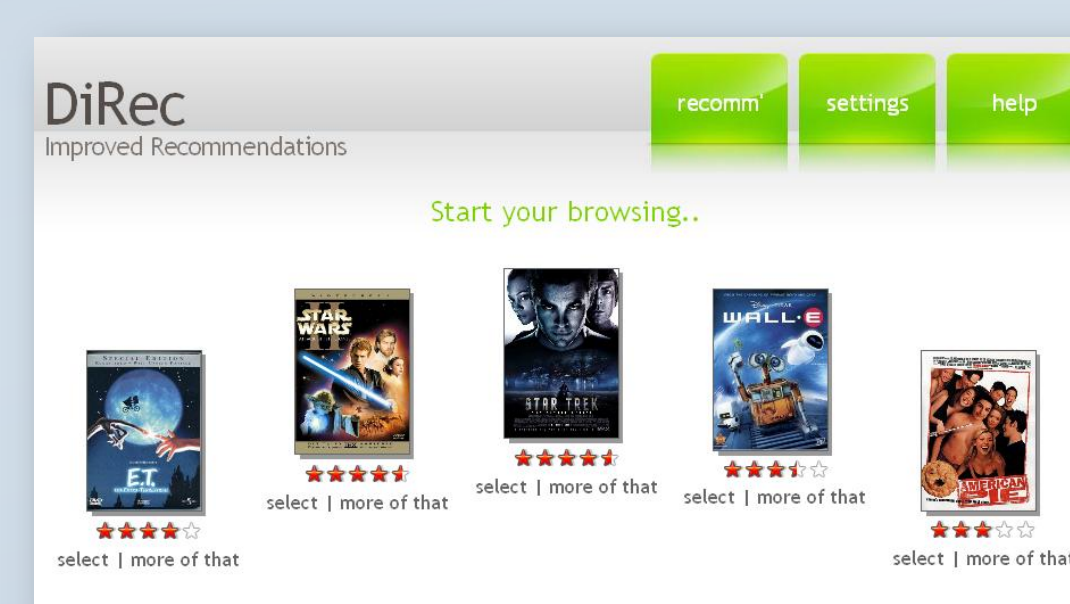
Experimental Results



Semantic (sequel) Diversity



Running Time



DiRec Screenshots