RecMax: Exploiting Recommender Systems for Fun and Profit

Amit Goyal

Laks V. S. Lakshmanan

University of British Columbia

http://cs.ubc.ca/~goyal

Recommender Systems

Movies

Products

Music

Videos

News

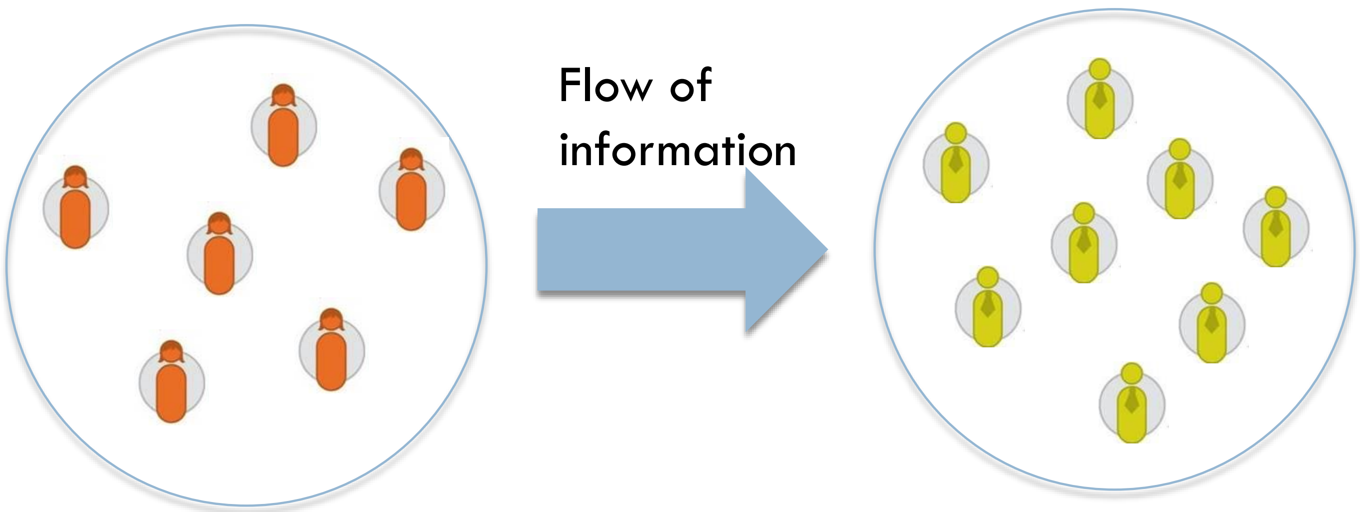
Websites

RecMax – Recommendation Maximization

- Previous research mostly focused on improving accuracy of recommendations.
- In this paper, we propose a novel problem RecMax (short for Recommendation Maximization).

Can we launch a targeted marketing campaign over an existing operational Recommender System?

Consider an item in a Recommender System

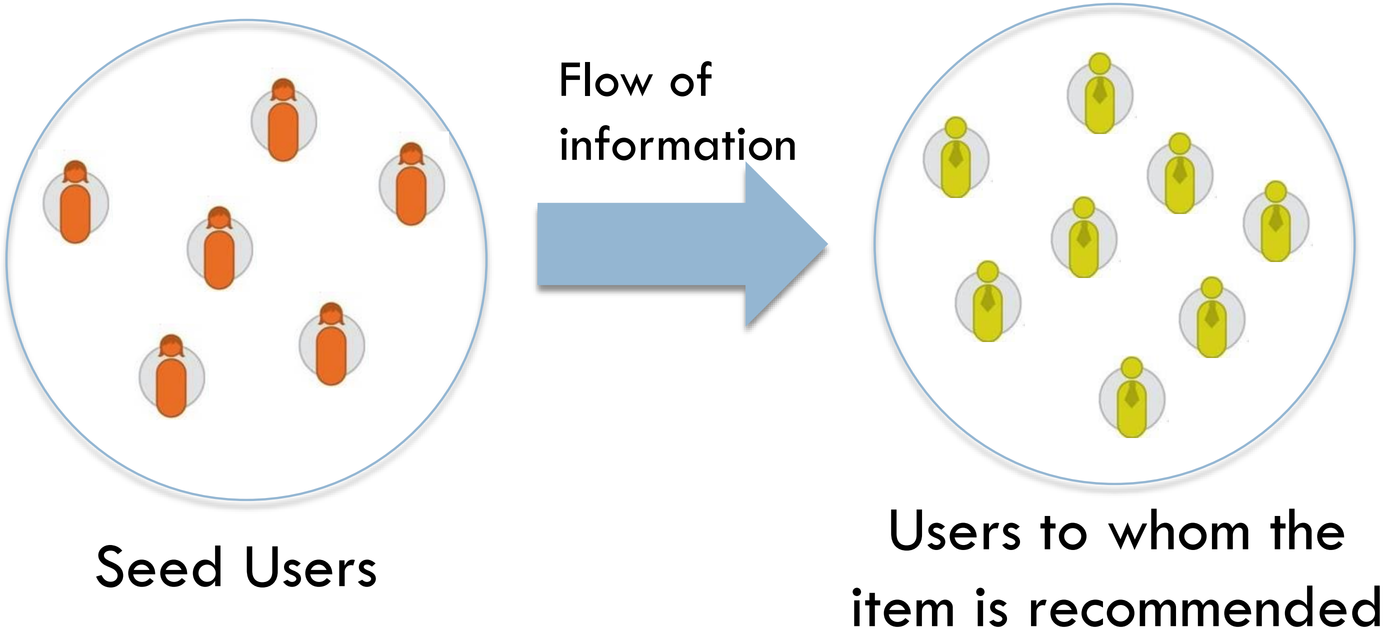


Some users rate the item (seed users)

Because of these ratings, the item may be recommended to some other users.

RecMax: Can we strategically select the seed users?

RecMax



Select k seed users such that
if they provide high ratings to a new product,
then the number of other users to whom the product is recommended (hit score) by the underlying recommender system algorithm is maximum.

RecMax – Problem Formulation

Recommendation List for user *v*

Number of recommendations are l

Recommendations	Expected Rating
Harry Potter	4.8
American Pie	4.3
….	
…	
The Dark Knight	3.2

rating threshold of user v (denoted by θ_v)

For a new item i, if expected rating $R(v,i) > \theta_v$, then the new item is recommended to v

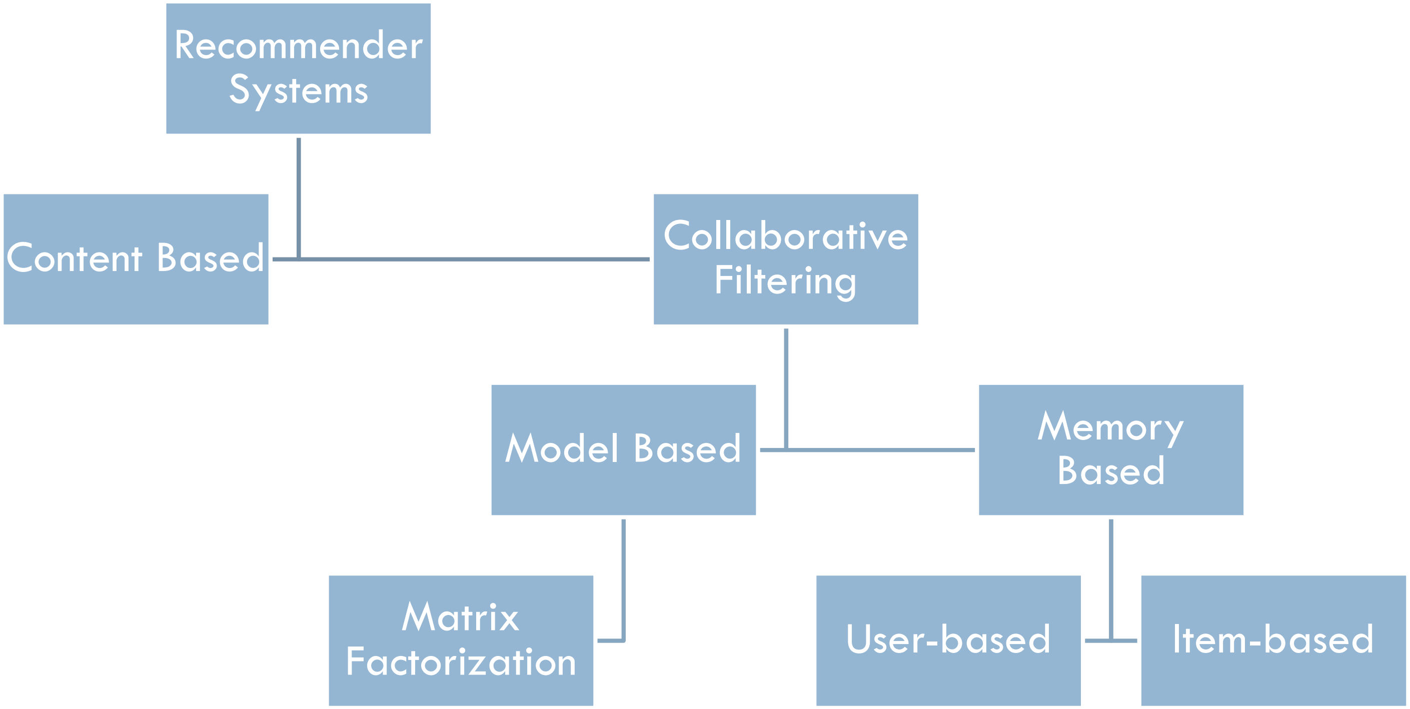
$$f(S) = \sum_{v \in V-S} I(R(v,i) > \theta_v)$$

The goal of RecMax is to find a seed set S such that hit score f(S) is maximized.

Benefits of RecMax

- Targeted marketing in Recommender Systems
  - Marketers can effectively advertise new products on a Recommender System platform.
  - Business opportunity to Recommender System platform.
  - Similar to Influence Maximization problem in spirit.
- Beneficial to seed users
  - They get free/discounted samples of a new product.
- Helpful to other users
  - They receive recommendations of new products – solution to cold start problem.

A key Challenge – Wide diversity of Recommender Systems



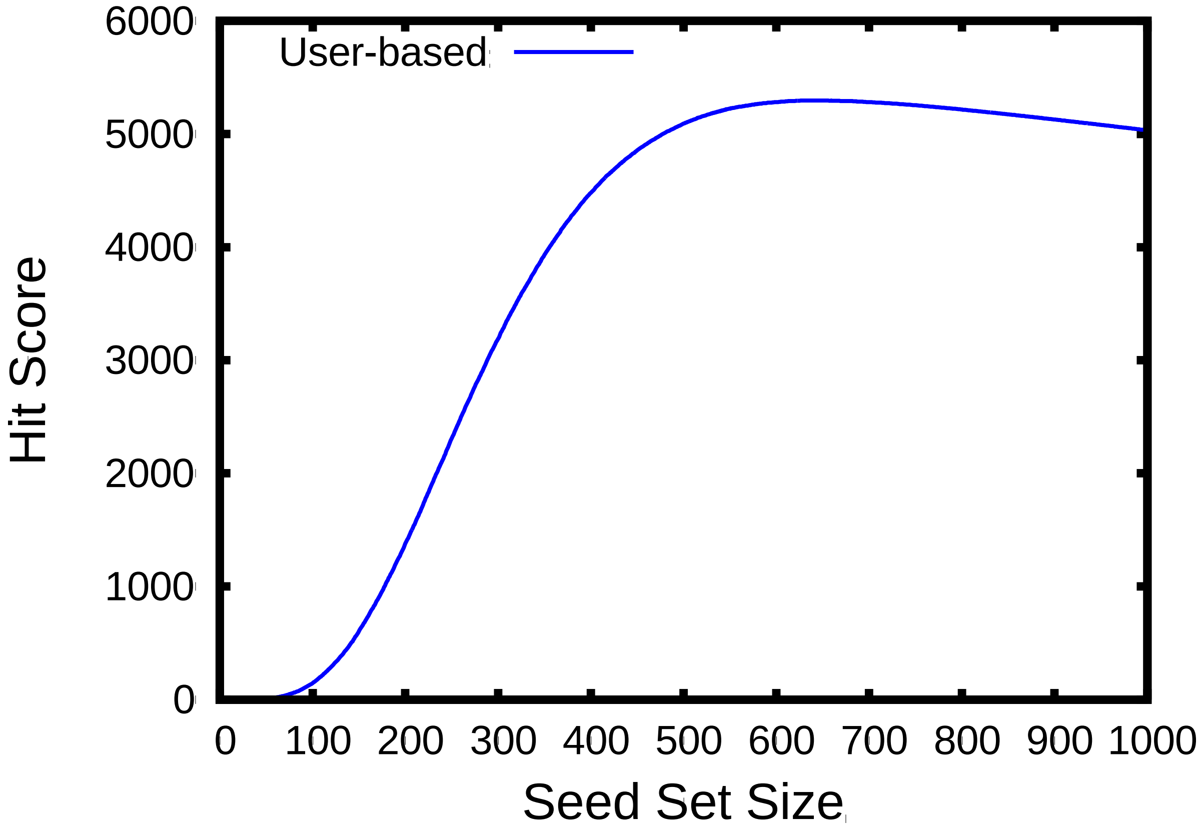
Similarity functions: Cosine, Pearson, Adjusted Cosine etc

Due to this wide diversity, it is very difficult to study RecMax

Outline

- What is RecMax?
- Does Seeding Help? – Preliminary Experiments.
- Theoretical Analysis of RecMax.
- Experiments.
- Conclusions and Future Work.

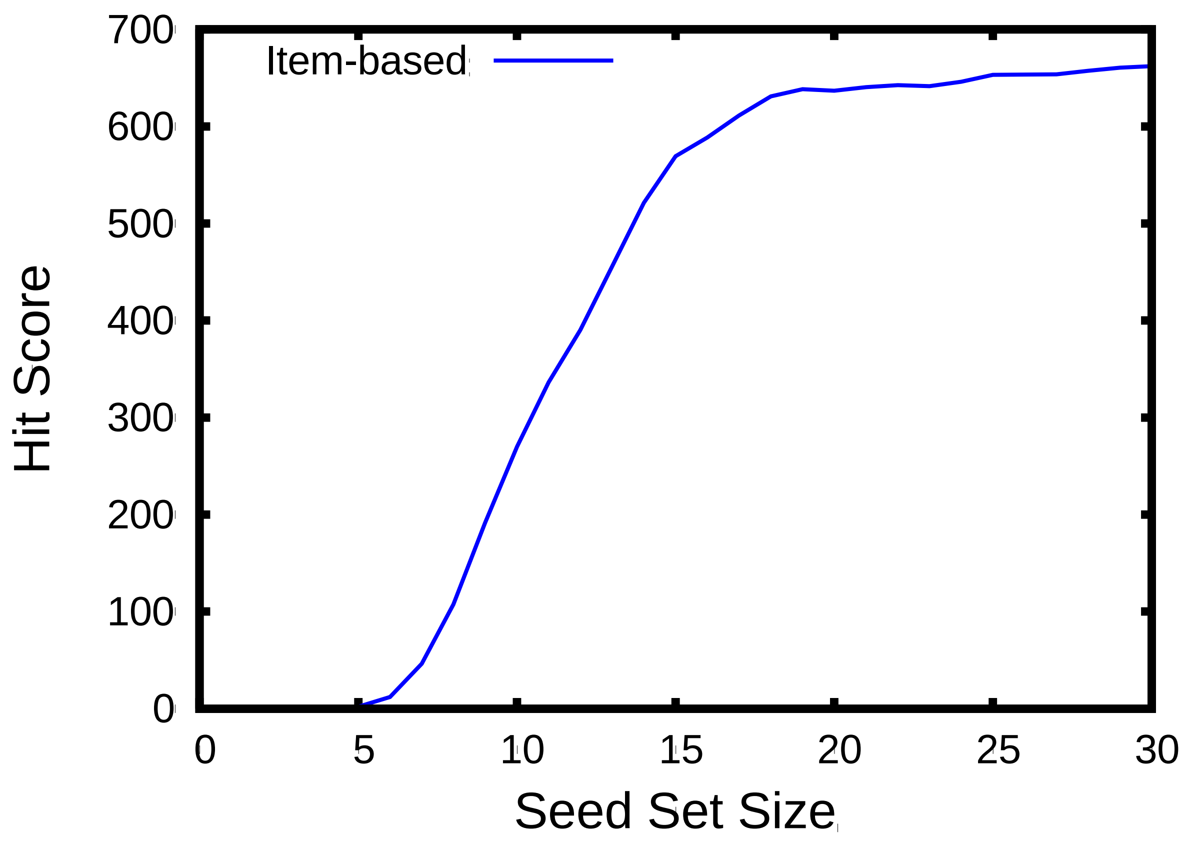
Does Seeding Help?



- Dataset: Movielens
- Recommender System: User-based
- Seeds are picked randomly.
- Recall that Hit Score is the number of users to whom the product is recommended.

A budget of 500 can get a hit score of 5091 (10x) (User-based)

Does Seeding Help?



- Dataset: Movielens
- Recommender System: Item-based
- Seeds are picked randomly.
- Recall that Hit Score is the number of users to whom the product is recommended.

A budget of 20 can get a hit score of 636 (30x) (Item-based)

Outline

- What is RecMax?
- Does Seeding Help? – Preliminary Experiments.
- Theoretical Analysis of RecMax.
- Experiments.
- Conclusions and Future Work.

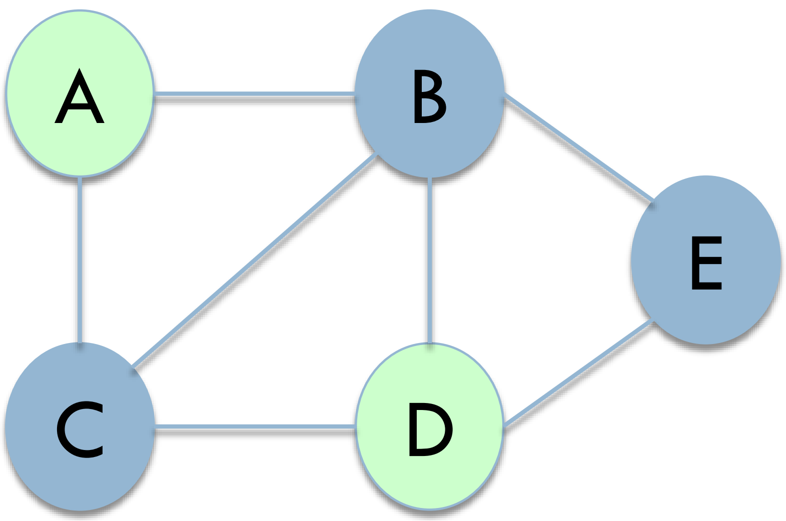
Key Theoretical Results

- RecMax is NP-hard to solve exactly.
- RecMax is NP-hard to approximate within a factor to $1/|V|^{(1-\varepsilon)}$ for any $\varepsilon > 0$.
  - No reasonable approximation algorithm can be developed.

- RecMax is as hard as Maximum Independent Set Problem.
- Under both User-based and under Item-based.

Why is RecMax so hard? (1/2)

- We introduce a helper problem – Maximum Encirclement Problem
  - find a set S of size k such that it encircles maximum number of nodes in the graph.

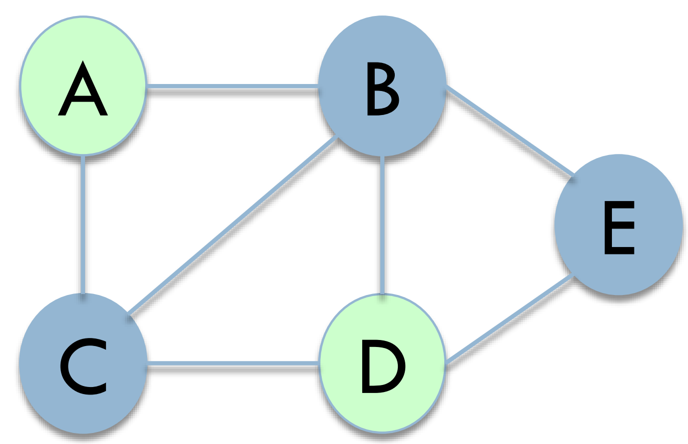


- Nodes {B,C} encircle node A.
- Nodes {B,C,E} encircle node D.
- Thus, {B,C,E} encircle A and D.

Why is RecMax so hard? (2/2)

- Set {B,C,E} is a solution to Maximum Encirclement Problem (for k=3).
- Nodes {A,D} form Maximum Independent Set.

- Reduction: Nodes {B,C,E} must rate the new item highly for the item to be recommended to A and D.
- RecMax is as hard as Maximum Independent Set, and hence NP-hard to approximate within a factor to $1/|V|^{(1-\varepsilon)}$



Discussion (1/2)

- We show hardness for User-based and Item-based methods.
  - What about Matrix Factorization?
  - Most likely hardness would remain (future work).

Discussion (2/2)

- Since the problem is hard to approximate, does it make sense to study?
  - YES, as we saw earlier, even a random heuristic fetches impressive gains.
- We explore several natural heuristics and compare them.
  - What about sophisticated heuristics (future work).

Outline

- What is RecMax?
- Does Seeding Help? – Preliminary Experiments.
- Theoretical Analysis of RecMax.
- Experiments.
- Conclusions and Future Work.

Datasets

	Movielens	Yahoo	Jester
#Users	6040	10K	25K
#Items	3706	5069	100
#Ratings	1M	1M	1.8M
Mean	3.58	51.9	10.88
Std. Dev.	1.12	39.9	5.23

Heuristics

- Random: Seed set is selected randomly. The process is repeated several times and average is taken.
- Most-Active: Top-k users with most number of ratings.
- Most-Positive: Top-k users with most positive average ratings.
- Most-Critical: Top-k users with most critical average ratings.

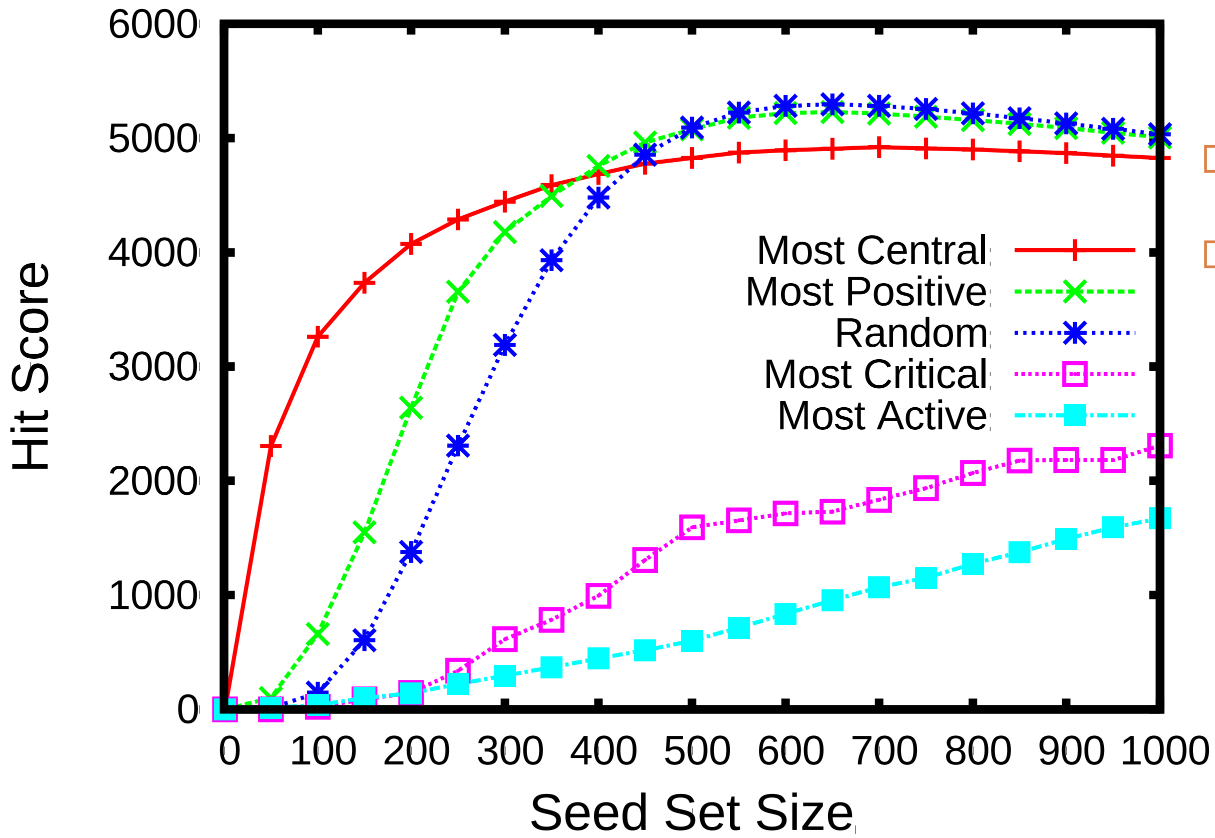
Heuristics

- Most-Central: Top-k central users.

$$agg(u) = \sum_{v \in V - u} sim(u, v)$$

User-Based Recommender Systems

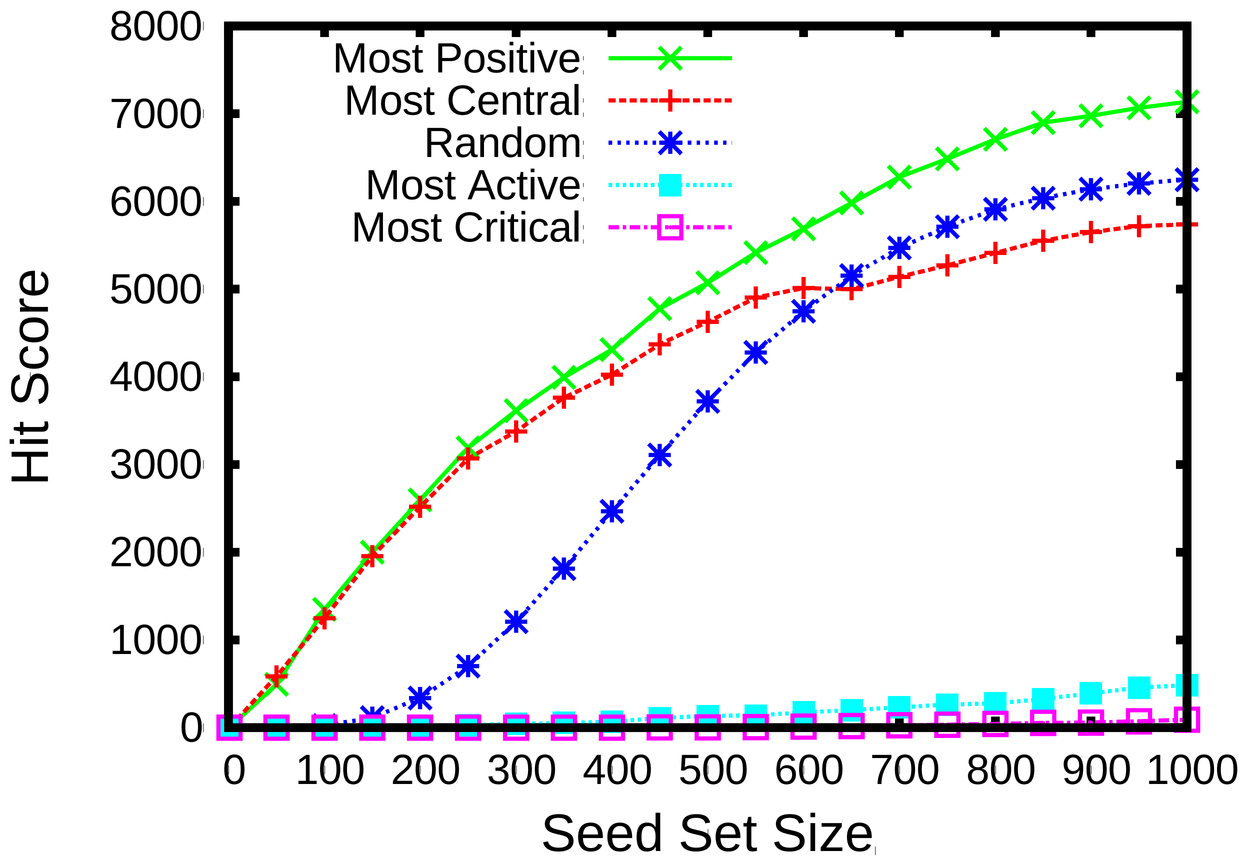
Comparison – Hit Score achieved



- Dataset: Movielens
- Recommender System: User-based

Most Central, Most Positive and Random perform good here.

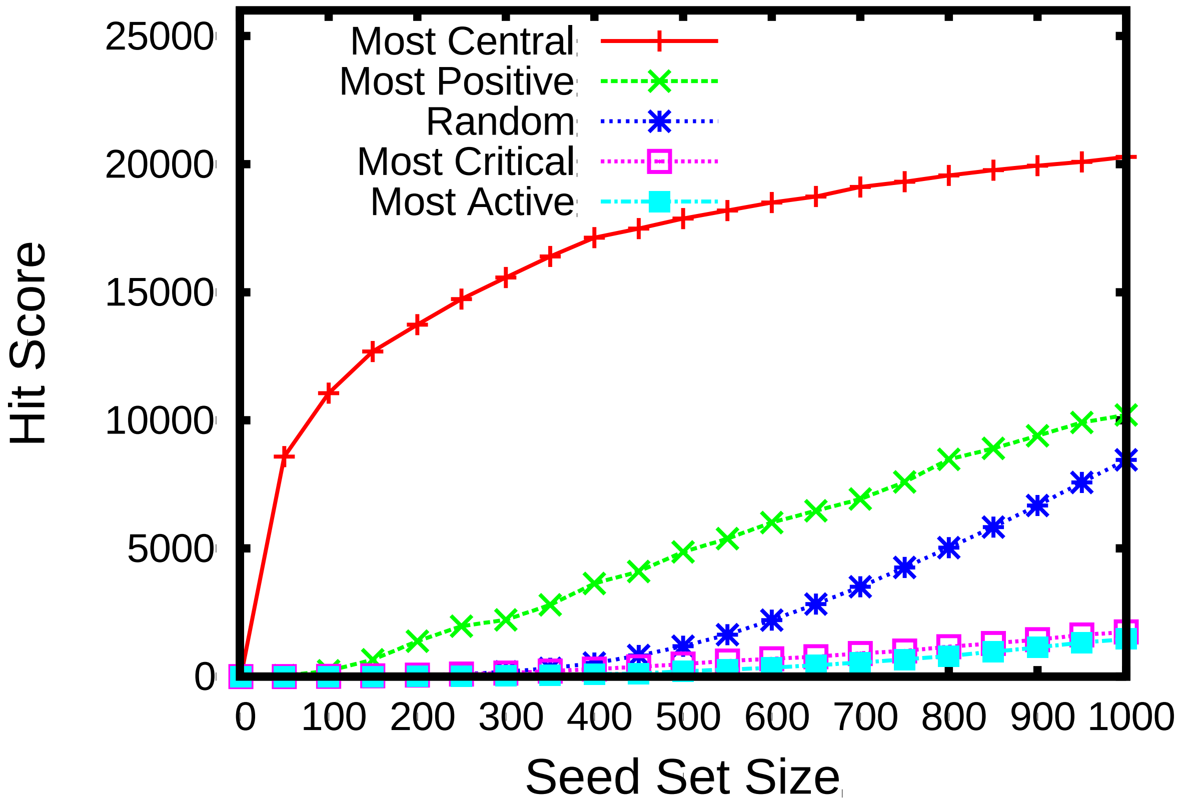
Comparison – Hit Score achieved



- Dataset: Yahoo! Music
- Recommender System: User-based

Most Positive, Most Central perform good here.

Comparison – Hit Score achieved



- Dataset: Jester Joke
- Recommender System: User-based

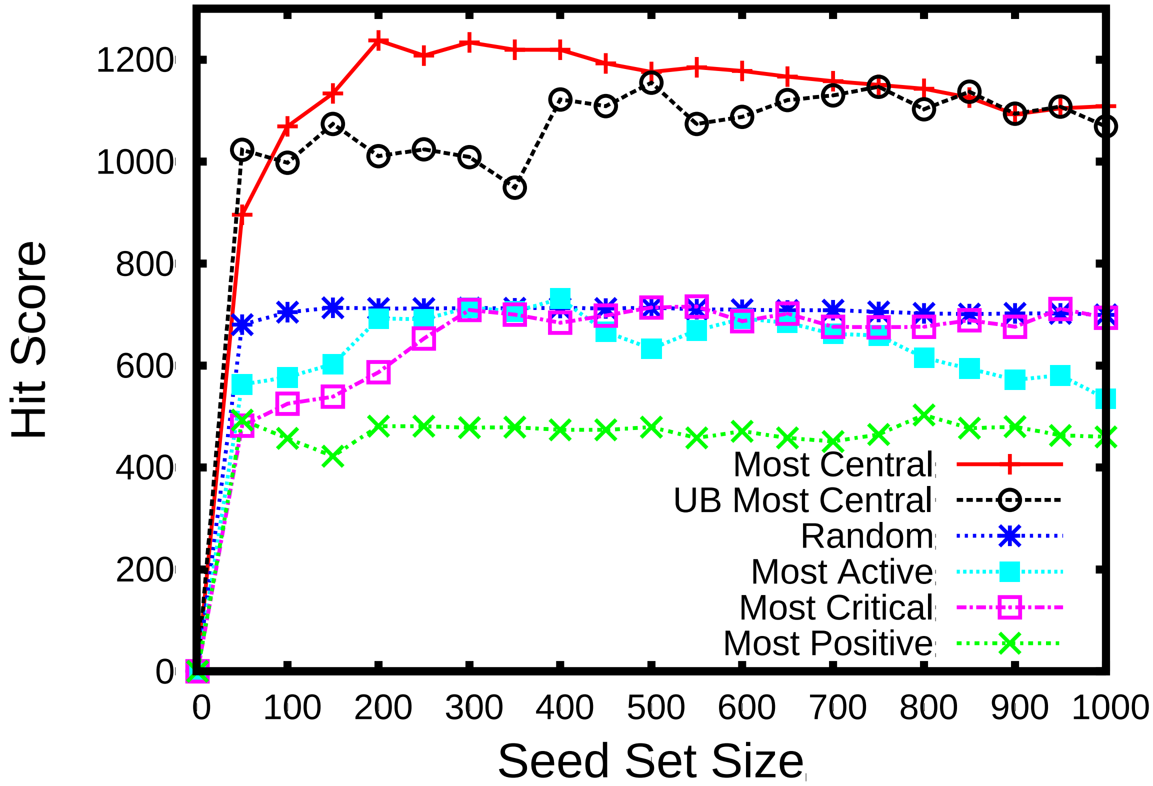
Most Central out-performs all other heuristics.

Key Takeaways

- Even the simple heuristics perform well.
- With a budget of 300, Most-Central heuristic achieves hit score of 4.4K, 3.4K and 15.6K on Movielens, Yahoo! and Jester respectively.
- Depending on the data set, we may encounter a "tipping point" – a minimum seeding is needed for the results to be impressive.

Item-Based Recommender Systems

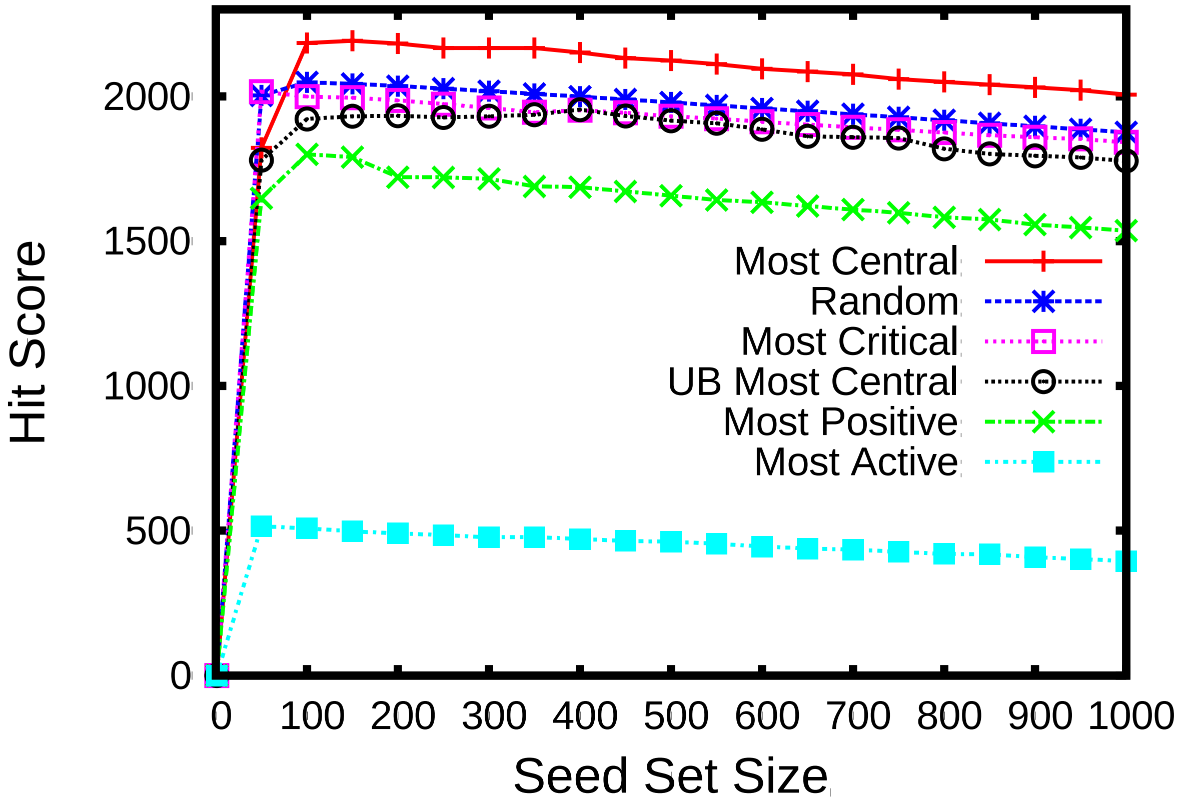
Comparison – Hit Score achieved



- Dataset: Movielens
- Recommender System: Item-based

Most Central performs good here.

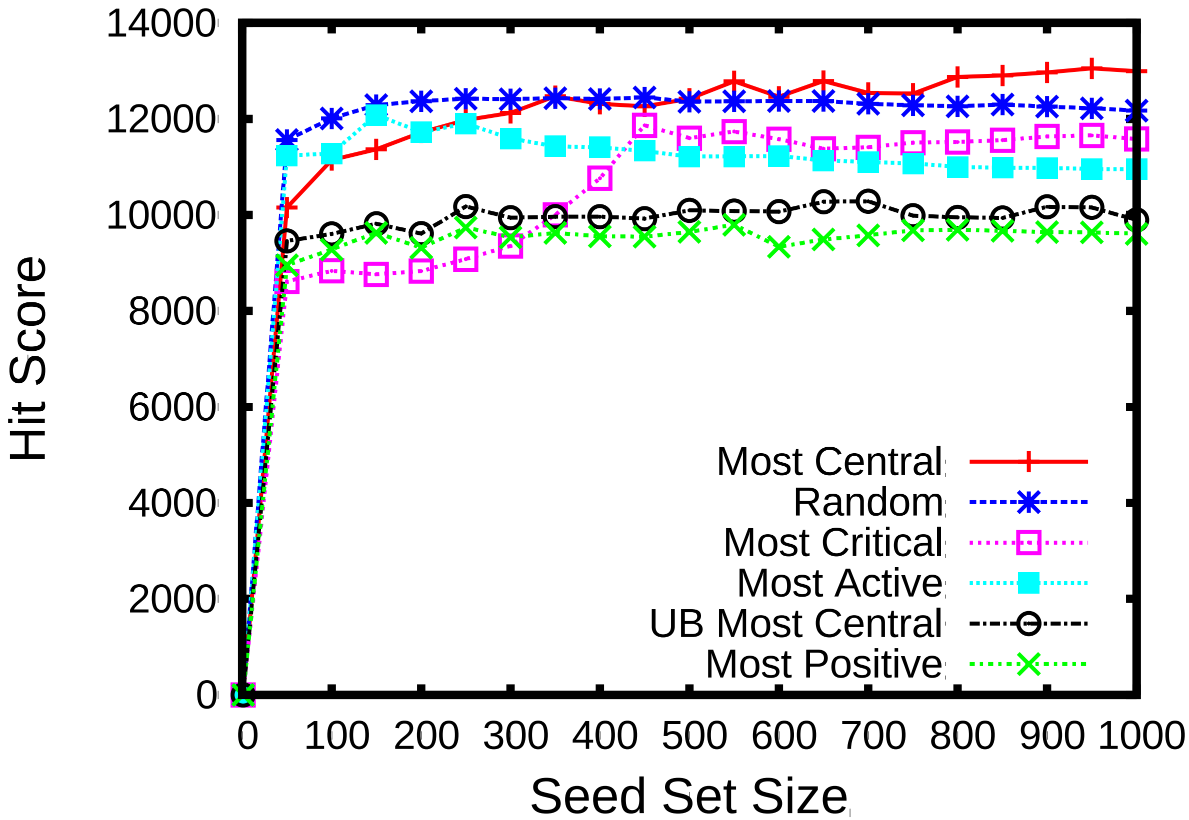
Comparison – Hit Score achieved



- Dataset: Yahoo! Music
- Recommender System: Item-based

Most Central performs good here.

Comparison – Hit Score achieved



- Dataset: Jester Joke
- Recommender System: Item-based

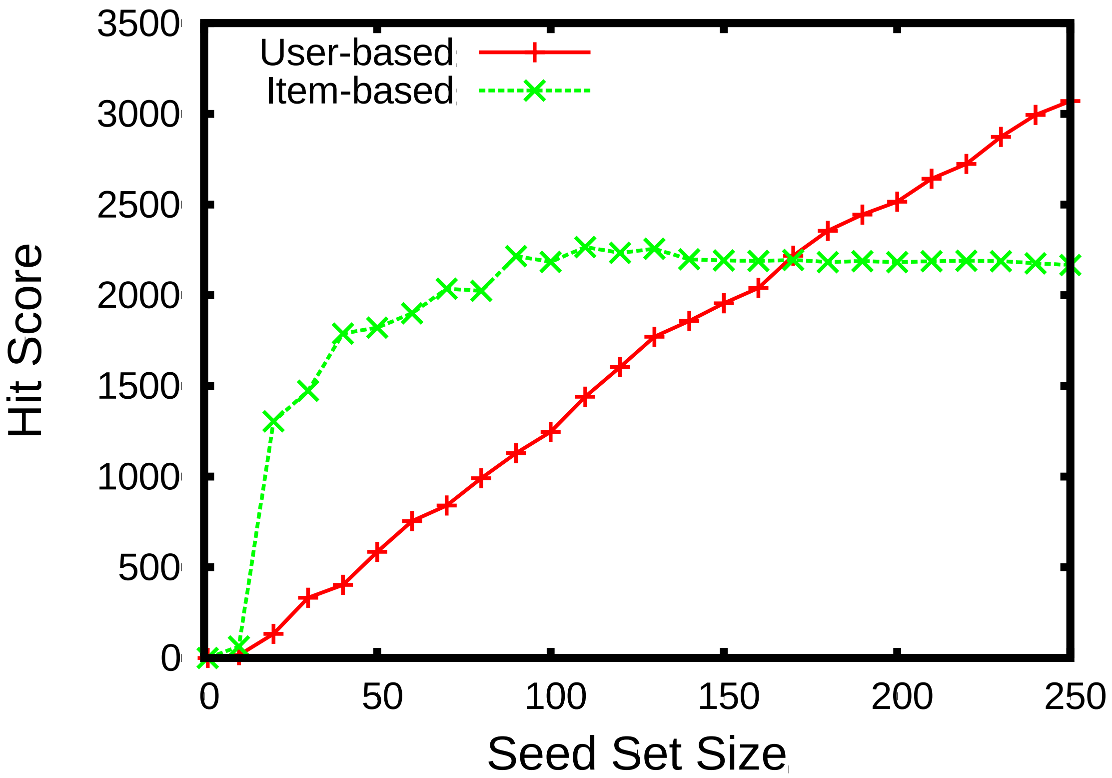
Most Central, Random and Most-Active performs good here.

Key Takeaways

- Again, the simple heuristics perform well.
- Hit score achieved in Item-based is much lower than in User-based.
- Thus, much less seeding is required to achieve maximum possible hit score.
- Overall, Most-Central performs well.
- The difference of Most-Central with baseline Random is not much.
  - We need better heuristics (future work).

User-Based vs Item-Based

User-based vs Item-based



- Dataset: Yahoo! Music
- Initial rise of hit score is steeper in Item-based.
- Hit score saturates much earlier in Item-based.

Eventual hit score that can be achieved is much more in User-based.

User-based vs Item-based

	Common Seeds (out of 1000 seeds)
Movielens	103 (10.3%)
Yahoo! Music	219 (21.9 %)
Jester Joke	62 (0.62 %)

Seed Sets are different in both methods.

Outline

- What is RecMax?
- Does Seeding Help? – Preliminary Experiments.
- Theoretical Analysis of RecMax.
- Experiments.
- Conclusions and Future Work.

Our Contributions

- The main goal of the paper is to propose and study a novel problem that we call RecMax
  - Select k seed users such that if they endorse a new product by providing relatively high ratings, the number of users to whom the product is recommended (hit score) is maximum.
  - We focus on User-based and Item-based recommender systems.
- We offer empirical evidence that seeding does help in boosting the number of recommendations

Our Contributions

- We perform a thorough theoretical analysis of RecMax.
  - RecMax is NP-hard to solve exactly.
  - RecMax is NP-hard to approximate within any reasonable factor.
- Given this hardness, we explore several natural heuristics on 3 real world datasets and report our findings.
- Even simple heuristics like Most-Central provide impressive gains
  - This makes RecMax an interesting problem for targeted marketing in recommender systems.

Future Work

- RecMax is a new problem and has real applications – our work is just the first work.
- Developing better heuristics.
- Studying RecMax on more sophisticated recommender systems algorithms
  - Matrix Factorization.

Thanks and Questions

RecMax: Exploiting Recommender Systems for Fun and Profit

Amit Goyal

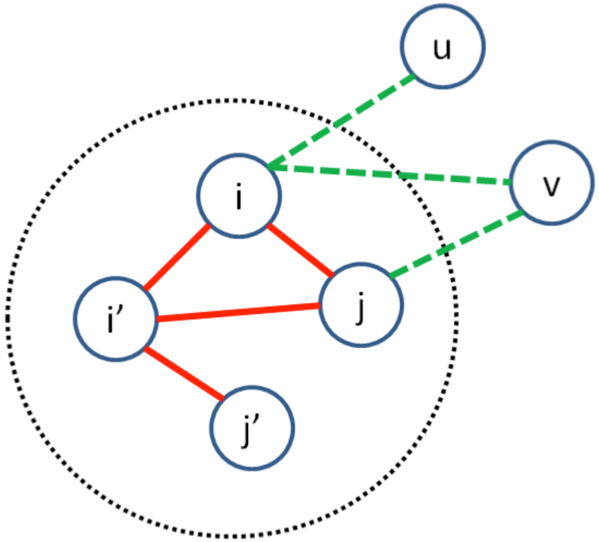
Laks V. S. Lakshmanan

University of British Columbia

http://cs.ubc.ca/~goyal

Heuristics

- Most-Central: Top-k central users.



$$sim(u,v) = \frac{\sum_{i \in I(u), j \in I(v)} w(i,j) \cdot sim(R(u,i), R(v,j))}{\sum_{i \in I(u), j \in I(v)} w(i,j)}$$