Maximizing Product Adoption in Social Networks

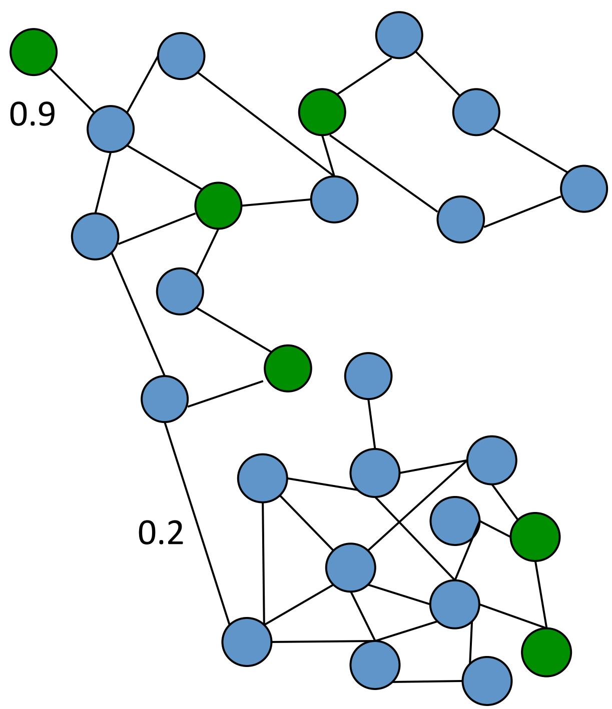
Presented by

Amlan Pradhan, Abhinav Rungta, Yining Zhou, Yingfei Xiang

-Group 1

(Paper appeared in WSDM 2012)

Viral Marketing



Node: User in a social network
(green – seed set)
Edge: Friendship among users
Edge Weight: Influence probability

- Objective: Given a social network, find a small number of individuals (seed set), who when convinced about a product will influence others by word-of-mouth, leading to a large number of adoptions of the product
- Studied as the Influence Maximization Problem

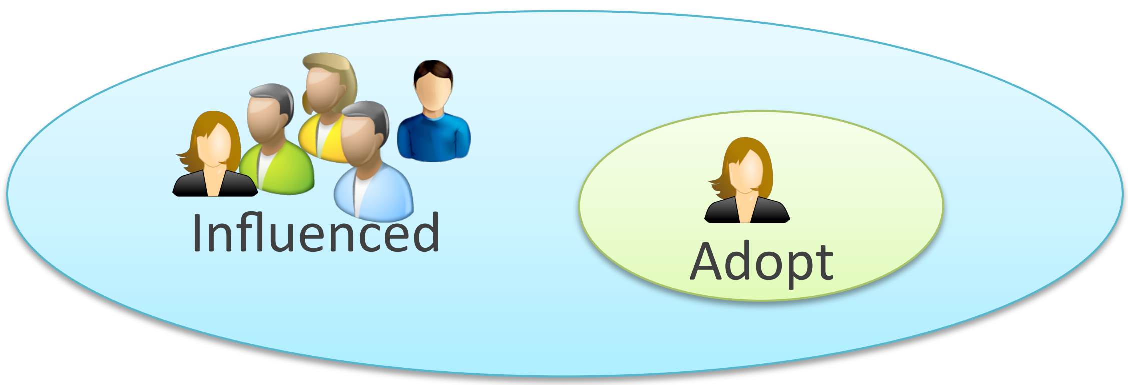
Previous Work

- Two classical influence propagation models:
  - Independent cascades
  - Linear threshold
    - Each user is initially inactive, the seed set is activated (influenced)
    - When the influence from the set of active friends exceeds a threshold for a user *v*, the user activates

- Influence is used as a proxy for adoption

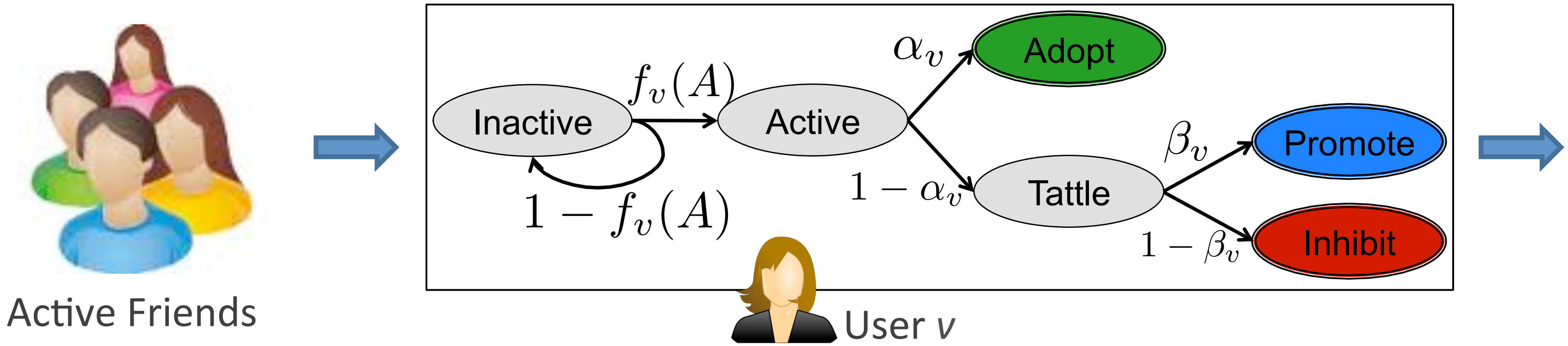
Influence $\nRightarrow$ Adoption

- Observation: Only a subset of influenced users actually adopt the marketed product



- Awareness/information spreads in an epidemic-like manner while adoption depends on factors such as product quality and price
- Moreover there exist users who help in information propagation without actually adoption the product – tattlers.

LT-C Model



- Model Parameters
  - *A* is the set of active friends
  - $f_v(A)$ is the activation function $f_v(A) = \frac{\sum_{u \in A} w_{u,v}(r_{u,i} - r_{\min})}{r_{\max} - r_{\min}}$
  - $r_{u,i}$ is the (predicted) rating for product i given by user u
  - α_v is the probability of user *v* adopting the product
  - β_v is the probability of user *v* promoting the product

Other Models

- LT Ratings Model
  - This is a special case of LT-C Model and does not have the tattle nodes. In this model, all ACTIVE nodes ADOPT the product (i.e. adoption probability $\alpha_v = 1$).

- LT Tattle Model
  - LT-Tattle is again a special case of LT-C Model without the use of ratings. In this, all nodes in the ADOPT or PROMOTE state rate the item as r_{max} and all INHIBIT nodes rate the item as r_{min}.

- Classical LT Model
  - In this model the activation function is just dependent on the influence ratio of its active neighbors. It can be modeled is a special case of LT-C where adoption probability $\alpha_v = 1$ and the rating $r_{u,I}$ is set to r_{max}.

Data

Test Data	Movie Lens	Yahoo
# Users	2113	2309
# Movies	10409	2380

Training Data	Movie Lens	Yahoo	Ciao
# Users	72000	7642	17615
# Movies	10000	11916	16121

In each of the datasets, the ratings are given in the range of 1 to 5 along with their respective timestamps.

Edge Creation

Jaccard Index = $\boldsymbol{(A \cap B)/(A \cup B)}$

- Original Dataset is a bipartite user-movies rating graph.
- To convert it to a directed user graph, we use Jaccard Similarity index to find users that are related.
- If the users rate movies, which are common to them, they are likely to be influenced by each other.
- If the Jaccard Similarity Index (based on the set of movies each user rates) is above a certain threshold value, there is an edge between the two users.
- We are using the threshold value here to be 0.25.

Influence Generation

- How to compute the influence of one user over the other?
- A user who rates a movie before another user is likely to have a stronger influence on the other user.
- We calculate the number of times a user v has rated a product after u. This value normalized over all such neighbors of v gives the influence weight of a user u on v.

$$\sum_{u \in N^{in}(v)} w_{(u,v)} = 1$$

- Here $N^{in}(v)$ is the set of neighbors who have a directed edge towards v.
- $w_{(u,v)}$ is the weight (influence ratio of node u on node v).

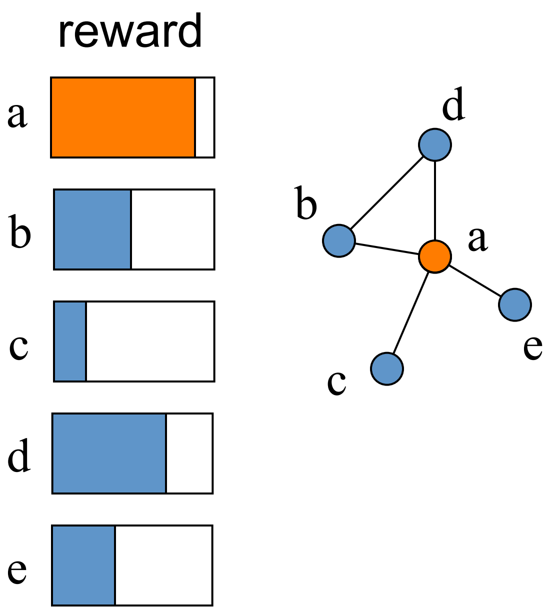
Ratings Matrix Generation

- We employ a simple prediction mechanism to calculate missing ratings value for a user – movie pair.
- Necessary for Calculating the Activation function of a user.
- Take the average of all ratings given by the same user (Lets call it U).
- Take the average of all ratings given to the movie (Lets call it M).
- The rating which the user might have given to the movie, is the average of U & M.

Calculating Seed Set using CELF

- The marginal benefit of adding b_i to the current seed set is the increase in the expected spread of the new seed set.

- CELF algorithm:
  - Keep an ordered list of marginal benefits b_i from previous iteration
  - Re-evaluate b_i only for top sensor
  - Re-sort and prune

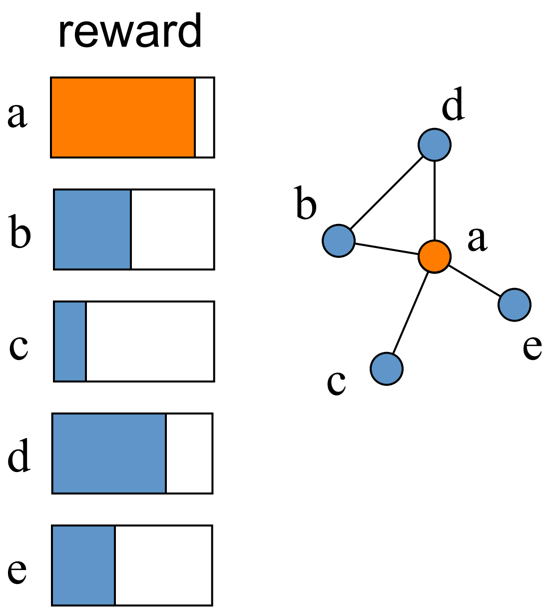


The expected spread needs to be calculated by running the experiment for large number of times.

Calculating Seed Set using CELF

- The marginal benefit of adding b_i to the current seed set is the increase in the expected spread of the new seed set.

- CELF algorithm:
  - Keep an ordered list of marginal benefits b_i from previous iteration
  - Re-evaluate b_i only for top sensor
  - Re-sort and prune

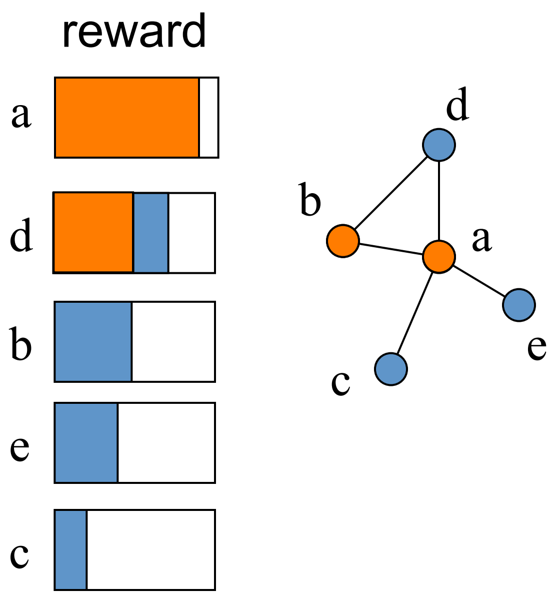


The expected spread needs to be calculated by running the experiment for large number of times.

Calculating Seed Set using CELF

- The marginal benefit of adding b_i to the current seed set is the increase in the expected spread of the new seed set.

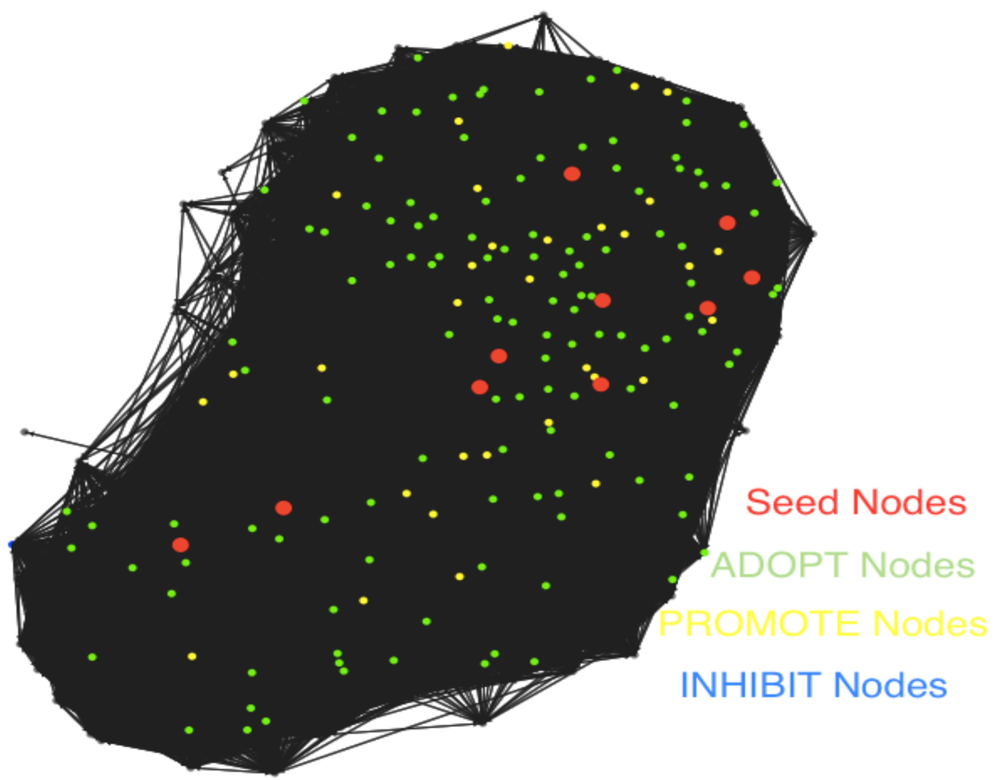
- CELF algorithm:
  - Keep an ordered list of marginal benefits b_i from previous iteration
  - Re-evaluate b_i only for top sensor
  - Re-sort and prune



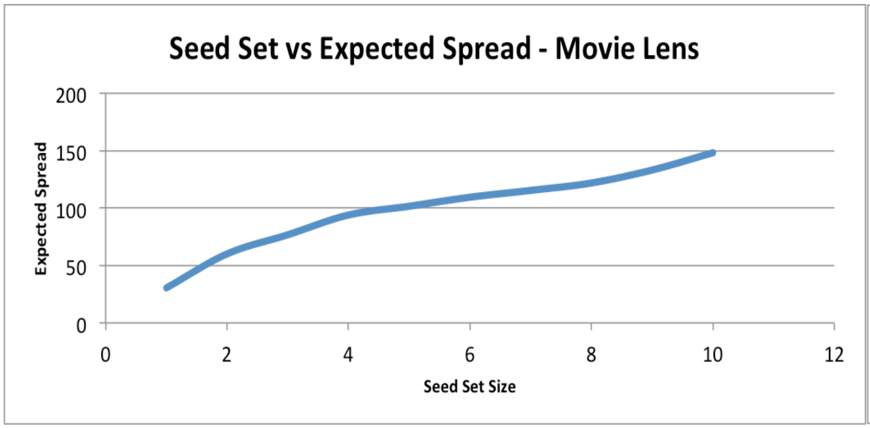
The expected spread needs to be calculated by running the experiment for large number of times.

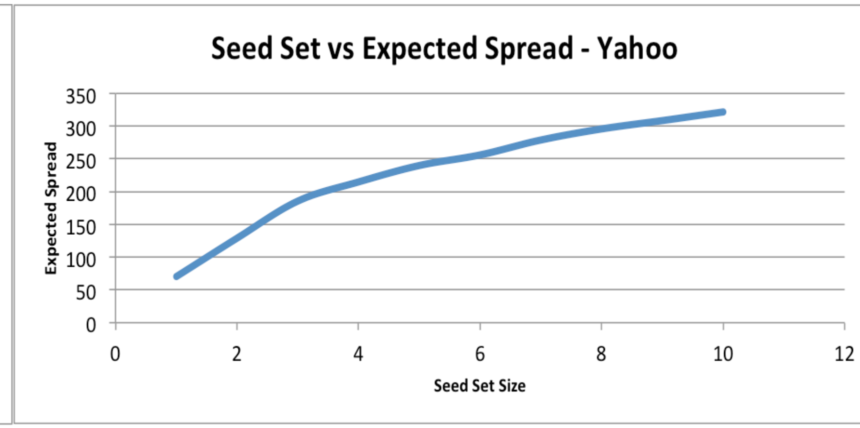
Graph Visualization

- We Use the Graph Stream API to create a visualization of the data set.



Test Run





Total No Of Users : 943
Total No of Movies : 1682
Total No Of Edges : 179950
Total No Of Seeds : 10
No of Users in Adopt State : 151
No of Users in Promote State : 15
No of Users in Inhibit State : 16

Total No Of Users : 2142
Total No Of Movies : 2380
Total No Of Edges : 1411312
Total No Of Seeds : 10
No of Users in Adopt State : 322
No of Users in Promote State : 37
No of Users in Inhibit State : 47

Things to Do

- Implementing other models like LT Ratings, LT-Tattle
- Compare the expected spread with ground truth, average rating vs spread size, etc.
- Code Scale Up so that the models can be run in a reasonable time on large graphs.

References

- [1] D. Kempe, J. M. Kleinberg, and É. Tardos. Maximizing the spread of Influence through a social network. *KDD*, 2003.
- [2] J. Leskovec, A. Krause, C. Guestrin, C. Faloutsos, J. VanBriesen, and N. Glance. Cost-effective outbreak detection in networks. In *KDD*, 2007.
- [3] J. Hartline, V. Mirrokni, and M. Sundararajan. Optimal marketing strategies over social networks. In *WWW*, 2008.
- [4] J. Tang, J. Sun, C. Wang, and Z. Yang. Social influence analysis in large-scale networks. In *KDD*, 2009.
- [5] Michael Levandowsky and David Winter. Distance between Sets. Nature 234, 34 - 35 (05 November 1971)
- [6] Smriti Bhagat, Amit Goyal, and Laks V.S. Lakshmanan. Maximizing product adoption in social networks. *Proceedings of the fifth ACM international conference on Web search and data mining*. In *ACM*, 2012.
- [7] M. Jamali and M. Ester. A matrix factorization technique with trust propagation for recommendation in social networks. In RecSys, 2010.

QUESTIONS ?