Improving Recommendation Accuracy by Clustering Social Networks with Trust

Thomas DuBois, Jennifer Golbeck, John Kleint, and Aravind Srinivasan

University of Maryland

Overview

- We use trust clustering as follows to aid recommender systems:
  - Infer trust between all pairs of users.
  - Find clusters of users with high trust.
  - Give more weight to users in the same cluster.

Talk Outline

- Trust in social networks.
- Clustering in recommender systems.
- Uses of trust in recommender systems.
- An algorithm for clustering networks by trust.
- Trust clusters in recommender systems.
- Experimental results.

Trust and Trust Inference

- Users rate their friends.
- High trust often means similar opinions.
  - Trust correlated to similarity.[a]
  - Most useful for controversial items.[b]
  - Trust can be directly applied to improve recommendation accuracy.[c]
- Sparse networks require trust inference.

[a] Ziegler and Lausen, Analyzing correlation between trust and user similarity in online communities.

[b] Golbeck, Trust and nuanced profile similarity in online social networks.

[c] Massa and Bhattacharjee, Using trust in recommender systems: an experimental analysis.

Clustering

- Bayesian clustering methods:
  - Cluster users by their ratings.[a][b]
  - Mixed results in recommender systems.
- Graph theoretic clusterings.[c]
  - Did not evaluate for rec sys.
- Clustering to speed up recommendations without much accuracy degradation.[d]

---

[a] Breese, Heckerman, and Kadie, Empirical analysis of predictive algorithms for collaborative filtering.

[b] Ungar and Foster, Clustering methods for collaborative filtering.

[c] Mirza, Keller, and Ramakrishnan. Studying recommendation algorithms by graph analysis.

[d] Sarwar et. al, Recommender systems for large-scale e-commerce: Scalable neighborhood formation using clustering.

Trust Inference Algorithm

- Our recent trust inference algorithm takes a random graph view of trust:
  - Direct trust as edge probability.
  - Indirect trust as path probability (approximated).
  - The log of the inverse of path probability defines a metric space.

Random Graph Theory

- Started in the 1950's with the question: in a random graph when is there an $O(n)$ sized connected component.
- Generalized to many types of graphs and many propertied.
- By using random graphs, we can make use of probabilitiy theory as well:
  - We can quickly find "less important" edges in the graph.

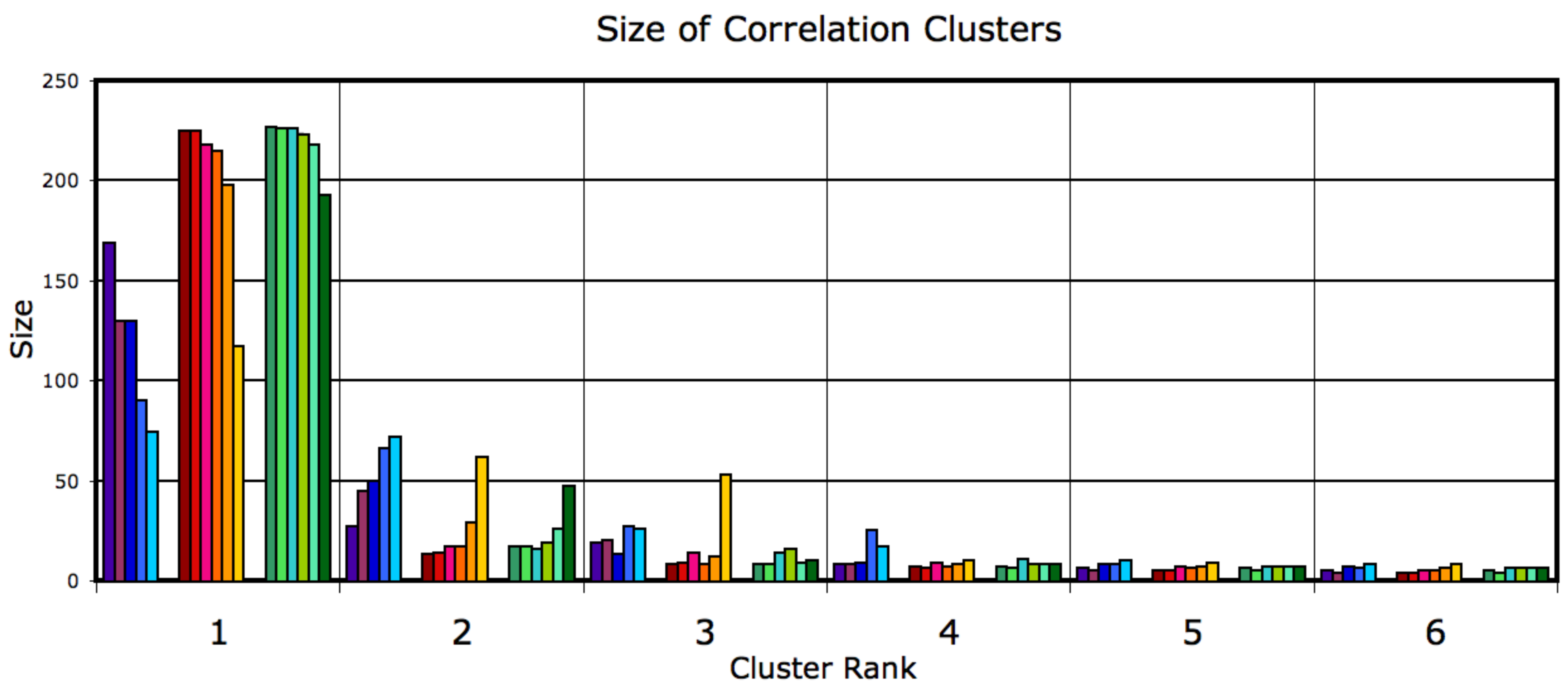
Trust Clustering

- Many clustering algorithms work on metric spaces:
  - k-centers, k-means, and others.
- We use correlation clustering.
  - Maximize closeness within groups and minimize closeness between them.
  - This seems well suited.

FilmTrust

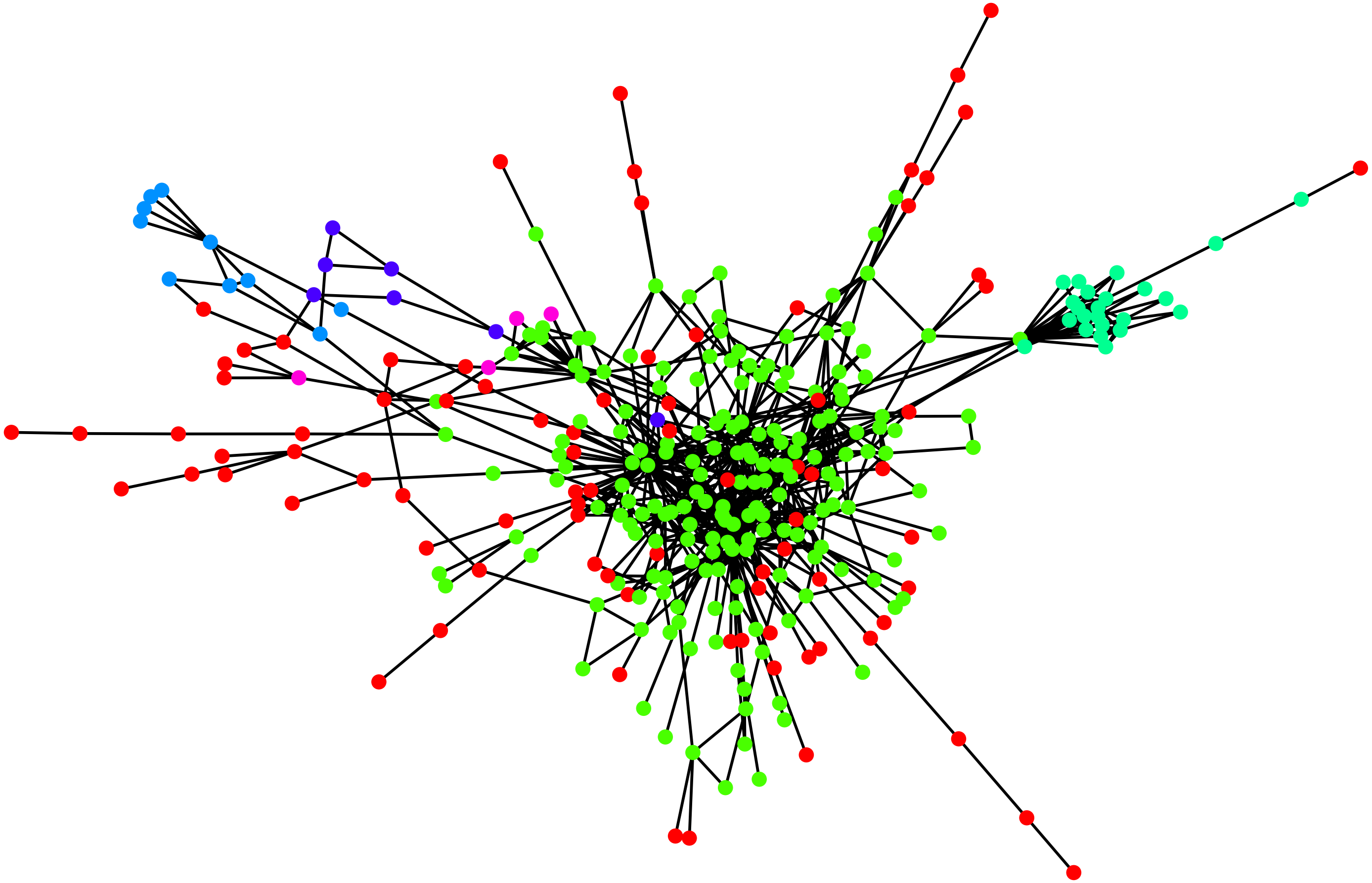
- We used the FilmTrust dataset.
  - Users rate movies
  - And their trust in friend's recommendations.
  - Small enough, ≈ 500 users in the largest component.
- Working to scale up to the epinions dataset.

FilmTrust Clustered



Six largest cluster sizes from iterations of correlation clustering. Purple/blue bars from a radius of 1, red/orange a radius of 2, and green a radius of 3.

FilmTrust Clustered



Using Trust Clusterings

- We modified two algorithms
  - Basic ACF (automated collaborative filtering) using Pearson coefficients.
  - A trust based ACF using trust values as weights.
- We compare the base results to those when users in the same cluster have more weight.

Experimental Results

Method	Cluster Radius	MAE method	MAE control	RMSE method	RMSE control
ACF	1	0.53454	0.53460	0.70199	0.70204
ACF	2	0.53452	0.53460	0.70196	0.70204
ACF	3	0.53453	0.53460	0.70194	0.70204
Trust	1	0.63495	0.63501	0.82614	0.82620
Trust	2	0.63496	0.63501	0.82615	0.82620
Trust	3	0.63496	0.63501	0.82614	0.82620

Improvements were small but noticeable and consistent *even over controls that already use trust.*

Conclusion

- Given a social trust component in the data, recommendations can be improved by clustering users based on trust.
- These improvements are small but consistent.
- These improvements exist even when trust is already considered.