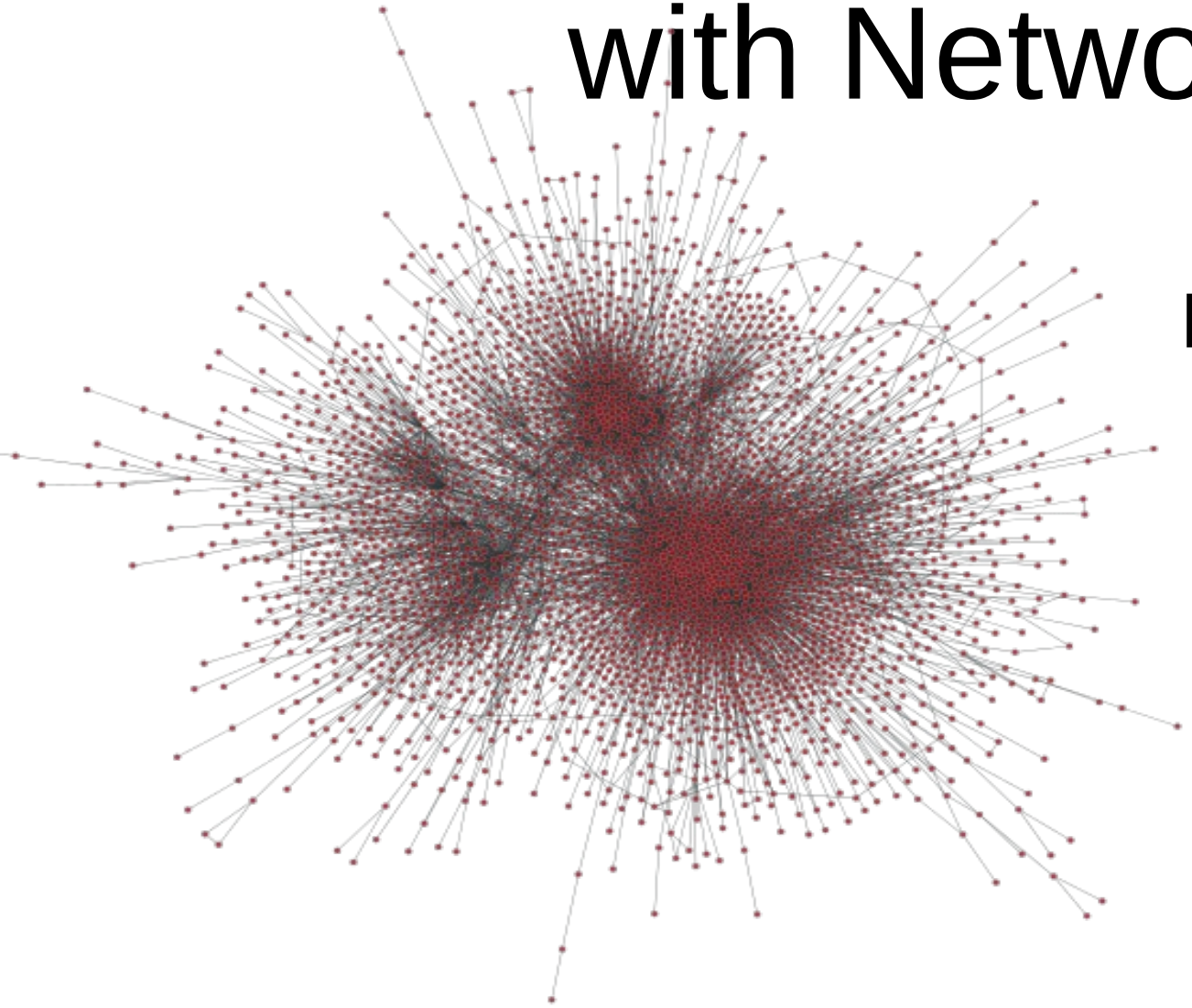
Exploring Recommenders with Networks



By Suhan Ree

07/01/2015

Can networks improve our predictions in machine learning algorithms?

If we know friendship relations between users, can it help us to improve recommendation systems?

YES, but how?

Data from Yelp Challenge Dataset

- Reviews: 1.6 M (10/12/2004 ~ 01/08/2015)
- Businesses: 61K (on 10 cities)
- Users: 366K with social network (2.9M edges)

U.K.: Edinburgh

Germany: Karlsruhe

Canada: Montreal and Waterloo

U.S.: Pittsburgh, Charlotte, Urbana-Champaign, Phoenix, Las Vegas, Madison

(Image from Yelp)

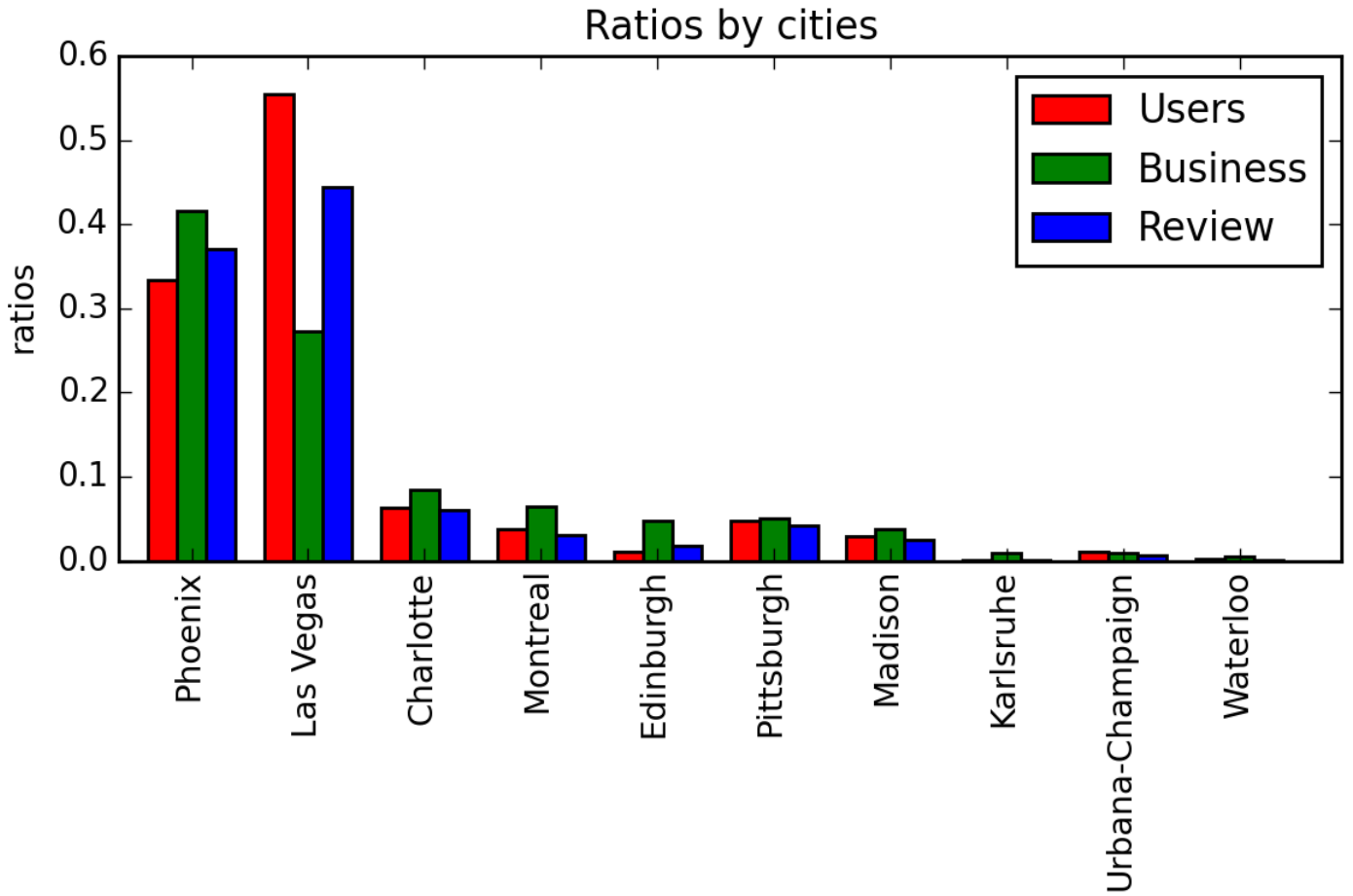
Finding communities

(1) Businesses are classified as 10 cities based on their locations.

(2) Cities are assigned to users based on where they left their reviews.

(3) Only 5% of users left reviews on multiple cities.
→ Look at their friends; and assign cities where the majority of their friends reside (if tied, choose at random among tied cities).

Counts per city

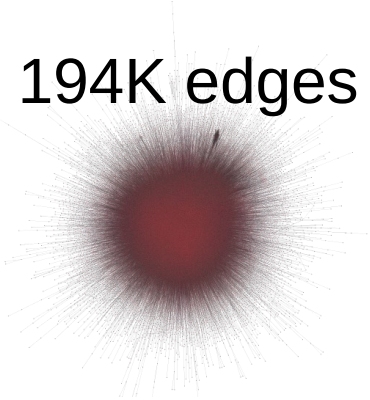


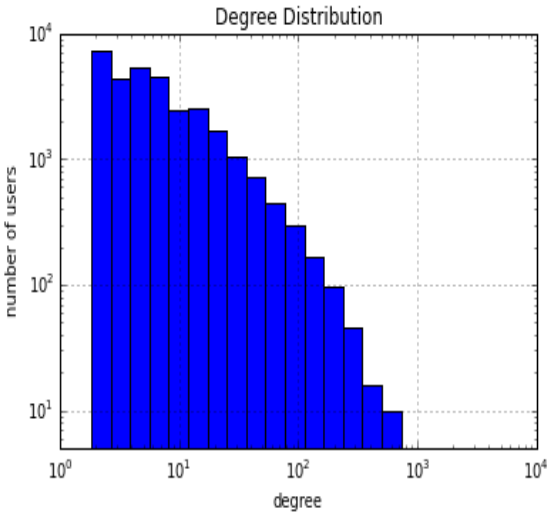
Looking at subnetworks of 3 cities

Phoenix

43K nodes

194K edges

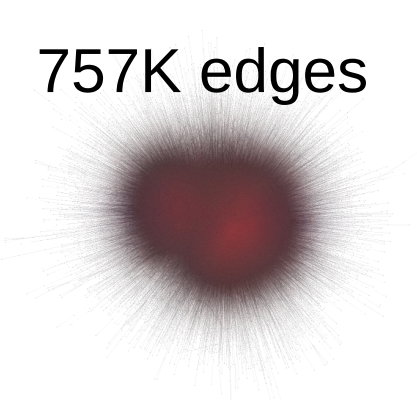


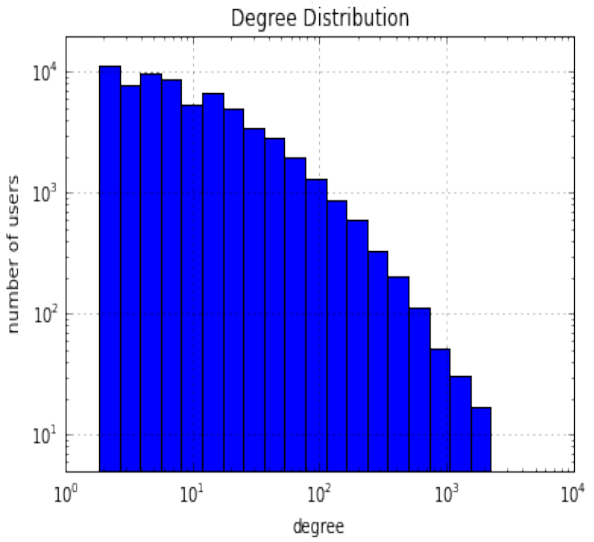


Las Vegas

85K nodes

757K edges

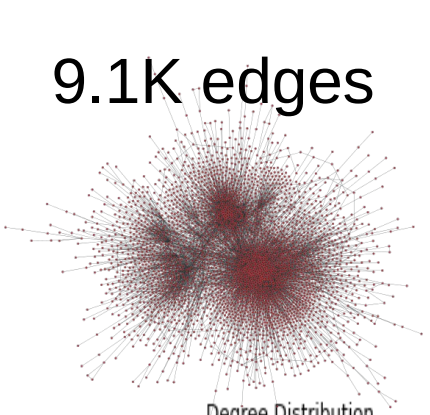


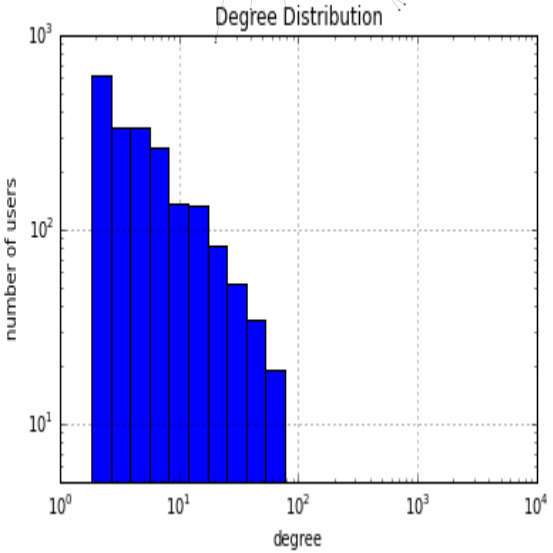


Montreal

3.1K nodes

9.1K edges



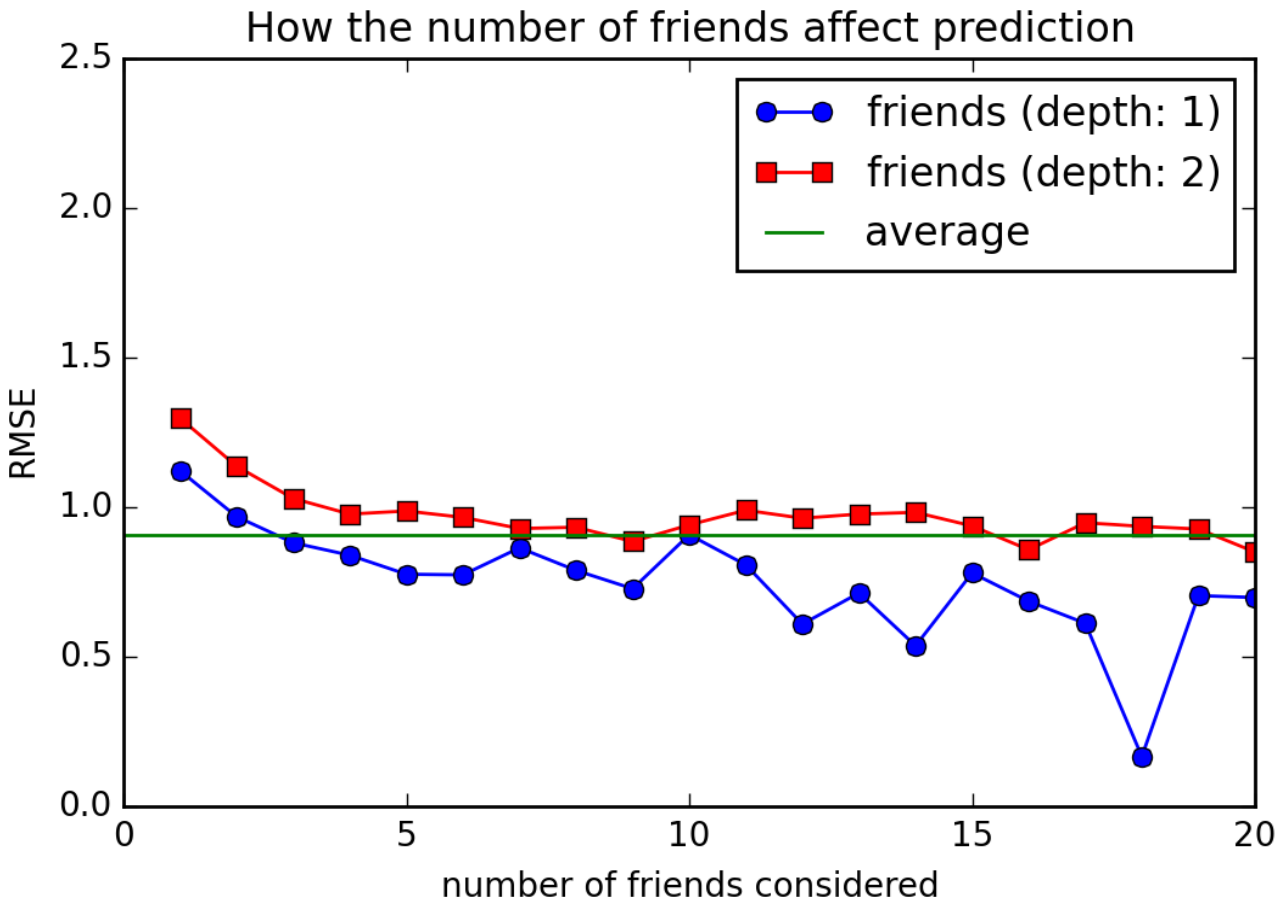


Recommender with a Network: the simplest approach

- Predict ratings for (user, business) pairs using the past ratings of the user's friends (and friends of friends, if necessary).
  - Similar method has been used before.

    "A social network-based recommender system,"
    by Chu and He, Dissertation, UCLA (2010).
- Will be compared with the baseline model using just the average, and the one using the latent-factor Collaborative Filtering.

Why **friend-based** network model?

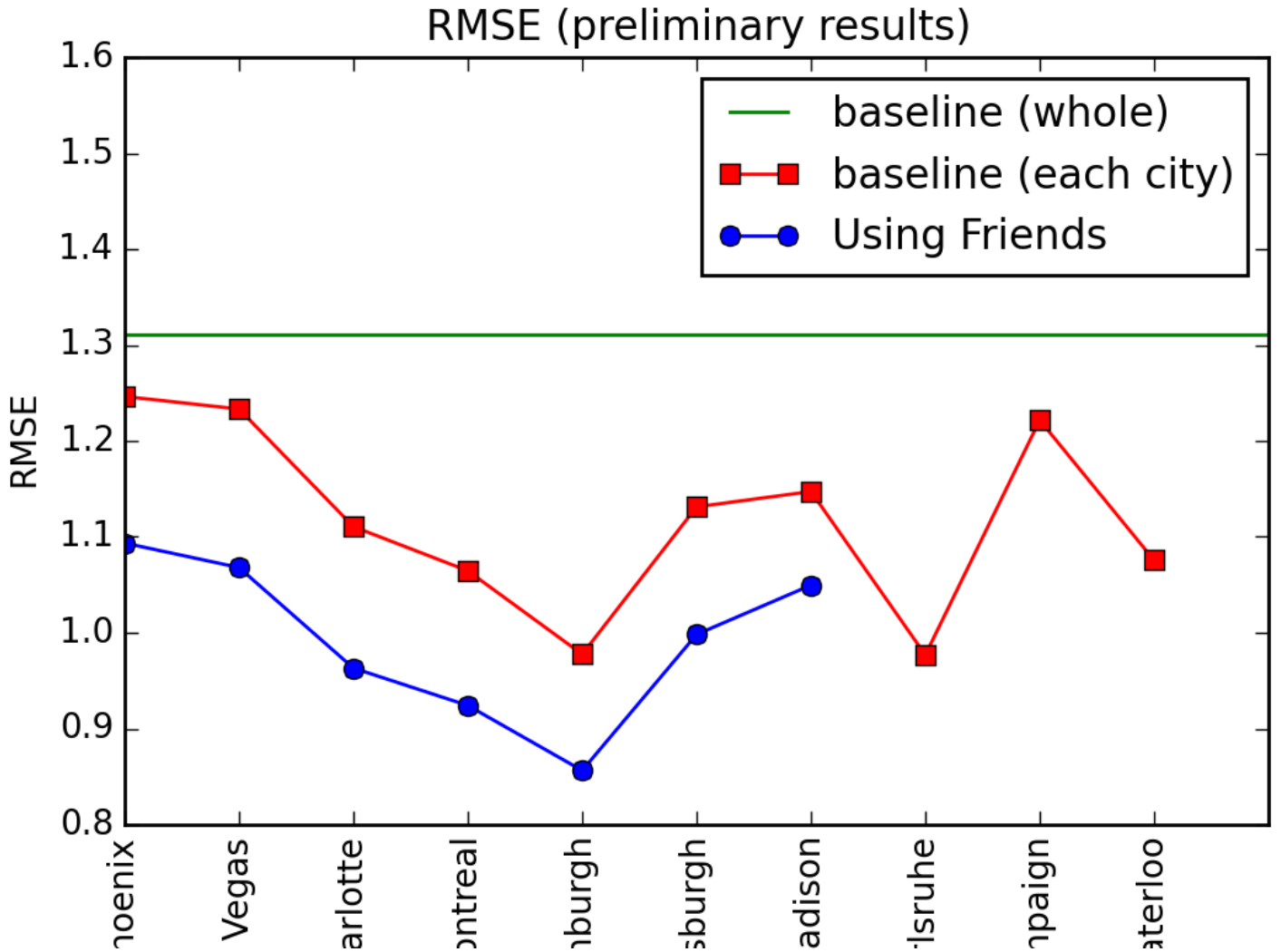


How to evaluate models?

Use K-fold cross validation

- The average RMSE (Root Mean Squared Error) will be found for each model.
- The lower, the more accurate!

Results: RMSE for models used



Conclusions

- Social networks can be useful.
- There are many ways to incorporate them; more to explore!
- Hybrid models should be considered to have better results:
  - Content-based recommender.
  - Collaborative filtering (user-user, or item-item)
  - Collaborative filtering (matrix factorization)
  - Demographic filtering
  - Network-related models, and so on.