Recommending Businesses on Yelp
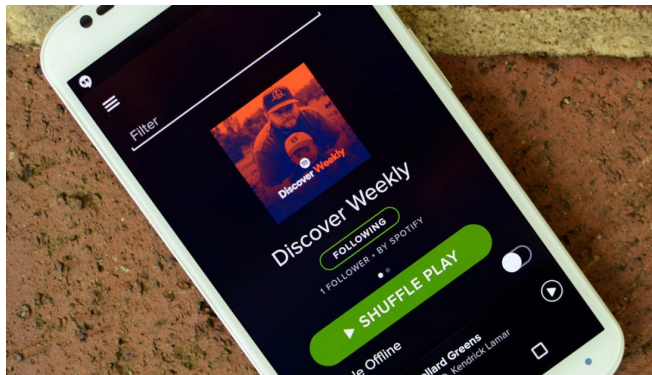
Will Sundstrom
Flatiron School Data Science Fellowship

October 24, 2019

Recommendations

Projected $4.41 billion
industry by 2022

Image Source

Image Source

Can I predict your Yelp ratings?

Better predictions, better recommendations
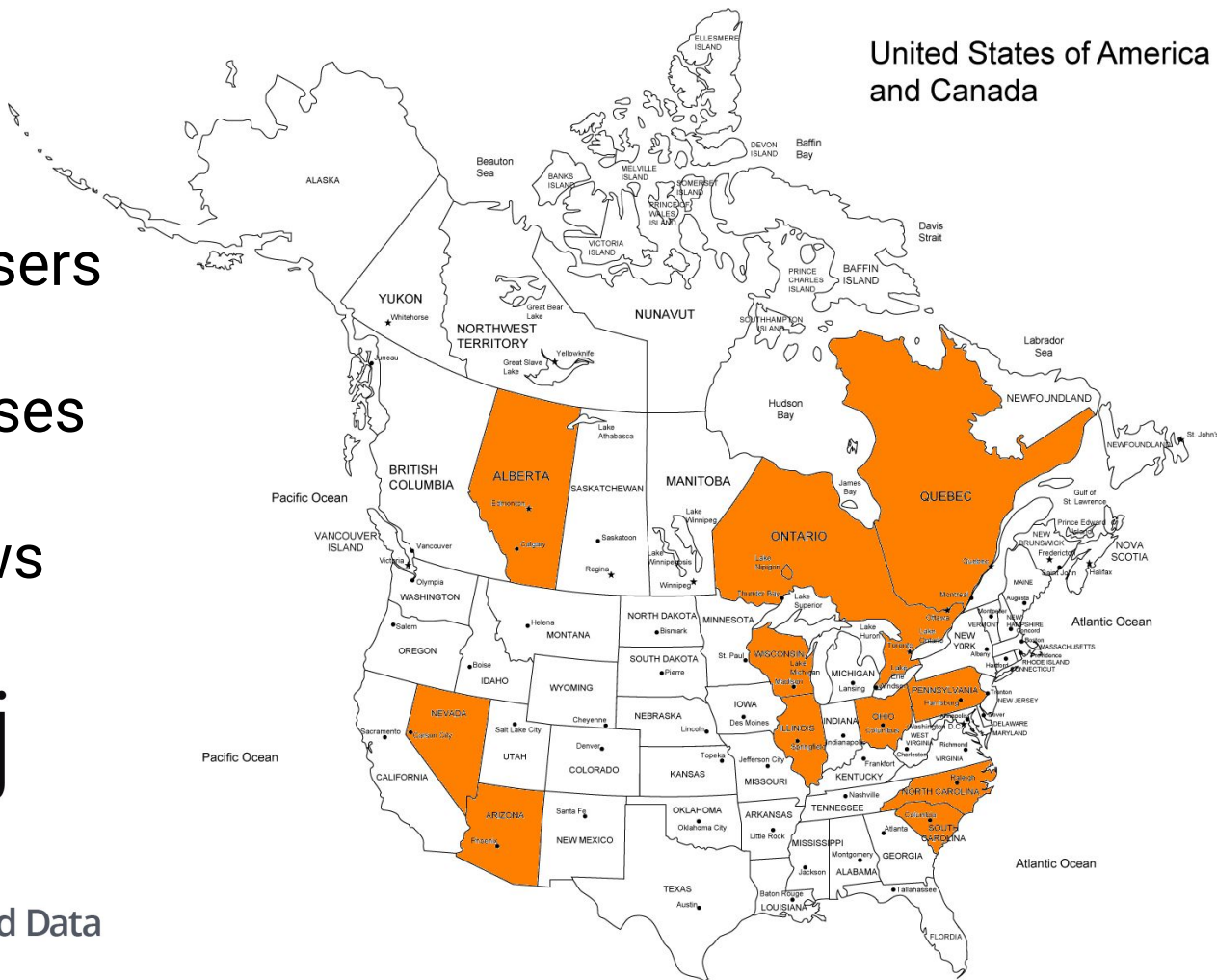
Recommending in real time with graphs
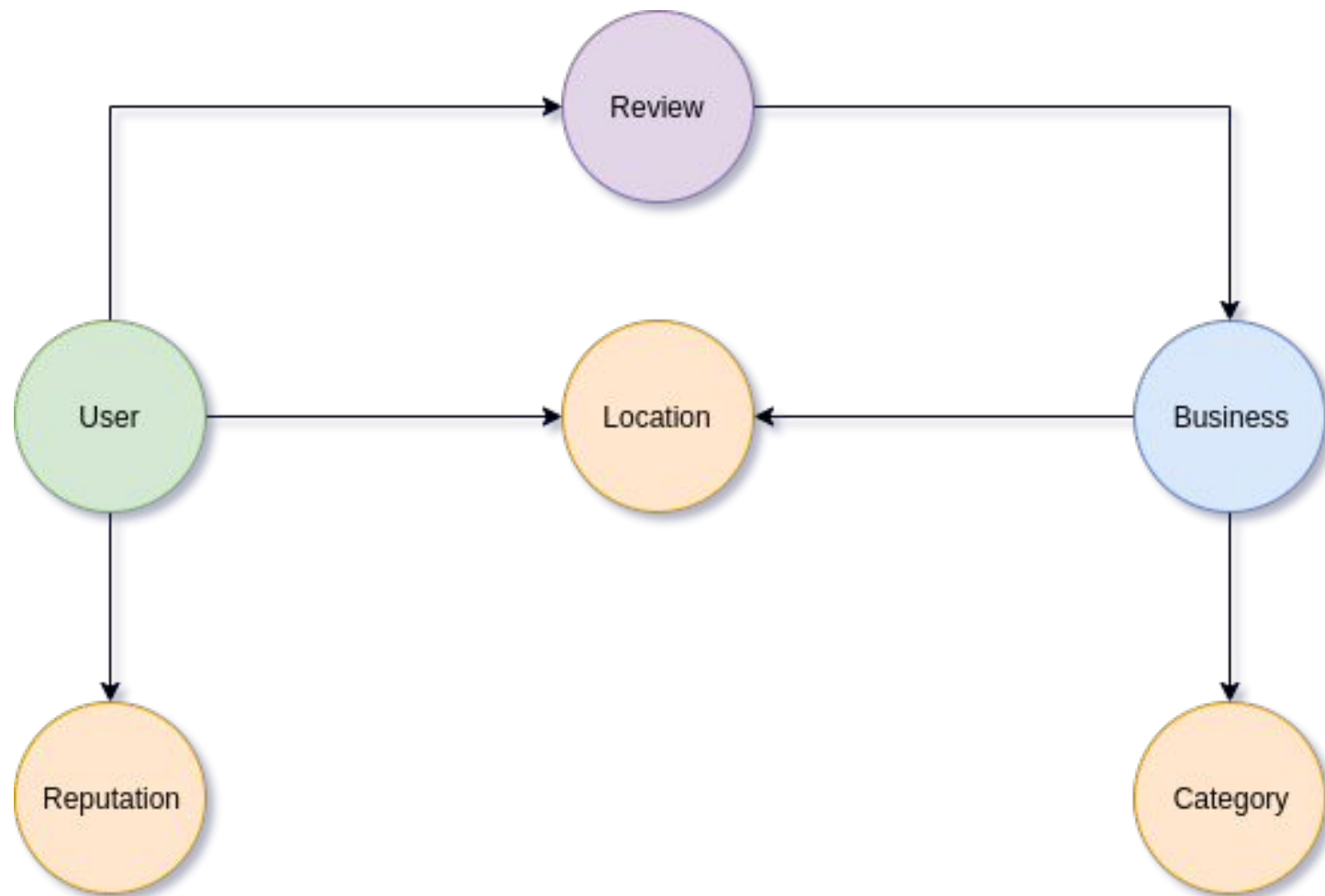
Data

- 1,637,138 Yelp users

- 192,609 Businesses

- 6,685,900 Reviews

neo4j

The #1 Database for Connected Data
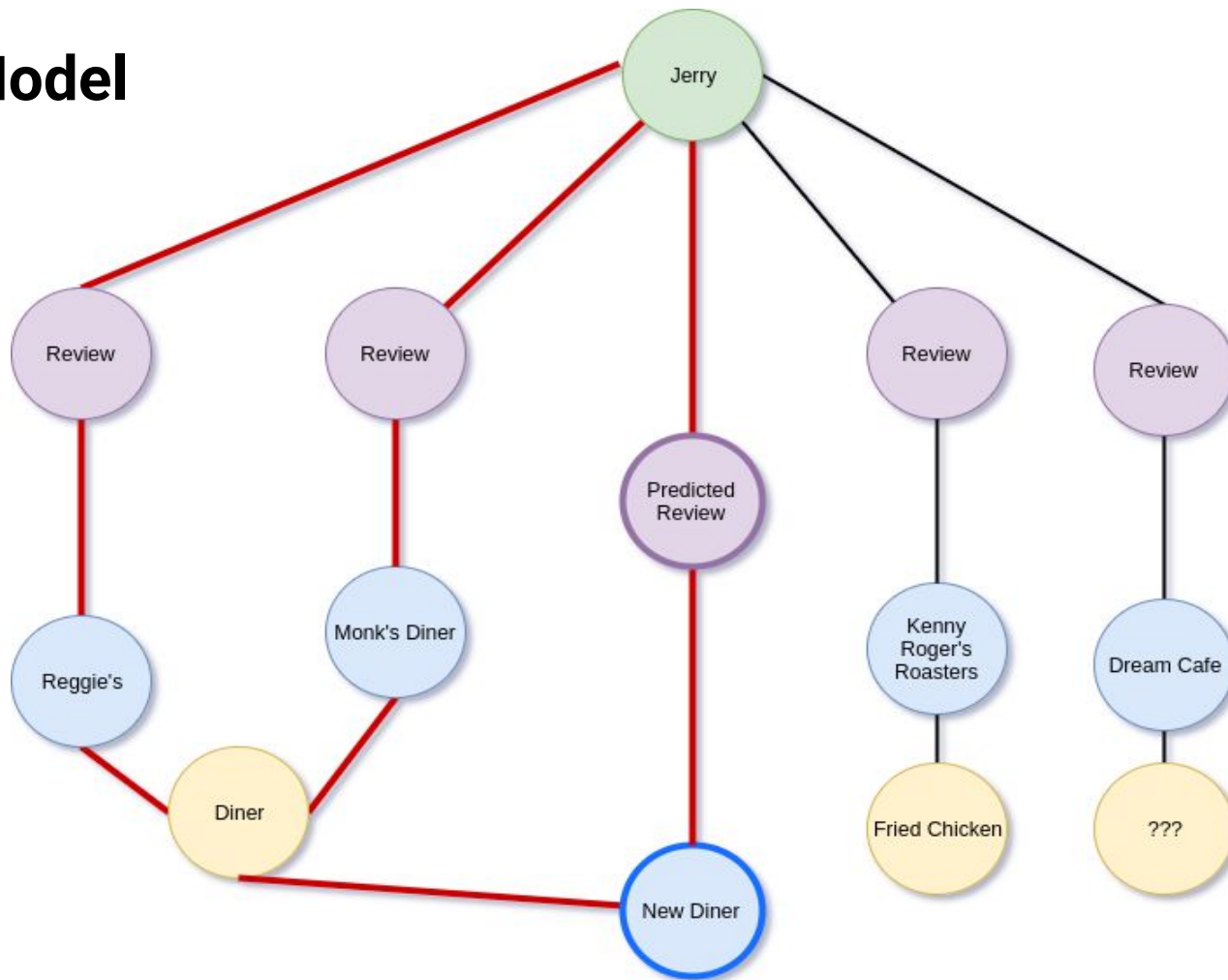
5

Results

Models Tested on Businesses in Quebec	Average Rating Error (RMSE)	Average Time Taken
Collaborative Filtering (on Businesses)	1.1284 Stars	7.52 Seconds
Latent Factor	1.1219 Stars	47.11 Seconds
Graph Model, Tested on Sample of 4,000 Businesses in Quebec	1.0172 Stars	

Collaborative Filtering & Latent Factors
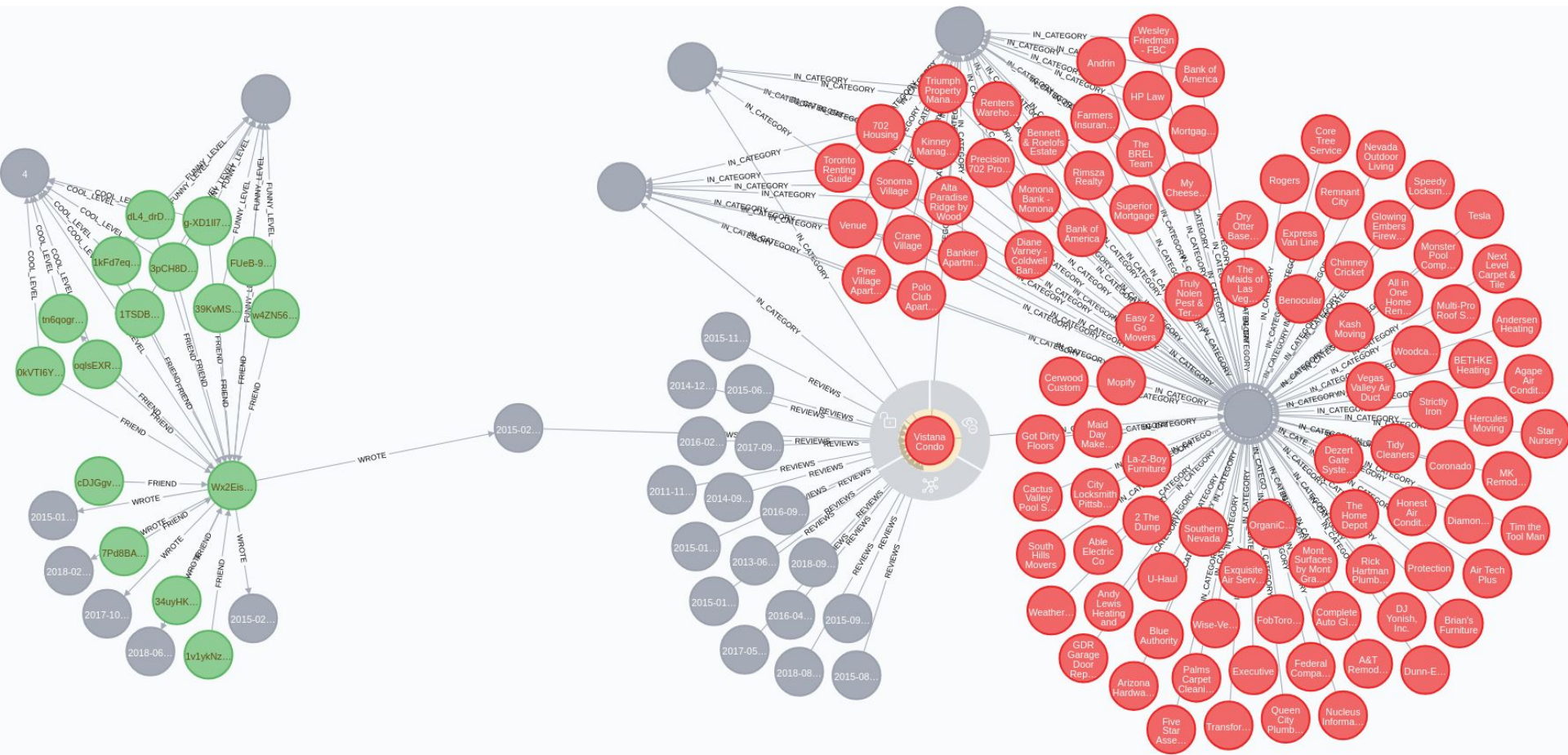
Graph Model

Recommender Demo

Conclusions and Further Work

- Collaborative filtering had good, was fastest

- Latent Factor had best performance, was very slow

- Graph algorithm was promising, too slow to fully test

Going Forward:
- Review text analysis
- Temporal analysis of reviews
- Social network
- Improve and expand graph algorithm

Thank You

Bibliography

- ➢ He, Jianming, and Wesley W. Chu. "A social network-based recommender system (SNRS)." *Data mining for social network data*. Springer, Boston, MA, 2010. 47-74.
  - ○ This is the source for the graph based algorithm, It was essential to my project.
- ➢ Koren, Yehuda. "Factorization meets the neighborhood: a multifaceted collaborative filtering model." *Proceedings of the 14th ACM SIGKDD international conference on Knowledge discovery and data mining*. ACM, 2008.
  - ○ Used for info about SVD++
- ➢ Koren, Yehuda. "Factor in the neighbors: Scalable and accurate collaborative filtering." *ACM Transactions on Knowledge Discovery from Data (TKDD)* 4.1 (2010): 1.
  - ○ Used for info about KNN baseline
- ➢ Robinson, Ian, Jim Webber, and Emil Eifrem. *Graph databases*. " O'Reilly Media, Inc.", 2013.
- ➢ https://neo4j.com/graphacademy/ and https://neo4j.com/docs/cypher-manual/current/
  - ○ These websites and the book above were used as resources for running and operating the neo4j server
- ➢ Images in the presentation are cited, original or come from yelp