Business Recommendation Using Graph Database

Chandrika Sharma
Prasanna Vikash Peddinti

Motivation and Significance

- In the age where Business are investing time and money in gathering data about about user behaviour patterns and buying trends, it’s important for new age systems to be able to process relationships easily and study these behaviours to recommend exactly what a user desires

- Leveraging Graph databases to make sense of complex relationships between entities which are not a strong suite of Relational databases.

- Walmart uses graph databases to make real-time product recommendations by using information about what users prefer

- Most of the top dating and online job sites use graph to recommend jobs or dates by incorporating the knowledge of the extended network (friends-of-friends and friends-of-friends-of-friends) into the recommendation as people with similar friend circles are more likely to want to date each other

- We wanted to experience how Graph Database was making it easy for huge datasets with complex relationships to be processed easily

Motivation and Significance cont..

- According to a report by industry observer DB-Engines, “Graph DBMSs are gaining in popularity faster than any other database category,” growing 300 percent since January of last year.[6]

- Several of the largest dating sites in the world have shifted toward graph databases in the last nine months.

- LinkedIn has a large team working on a proprietary graph database, which sits at the center of nearly every operation in Linkedin

- Twitter depends on a graph database, and has released FlockDB, a graph database it created, as open source.[6]

Dataset - Yelp Kaggle

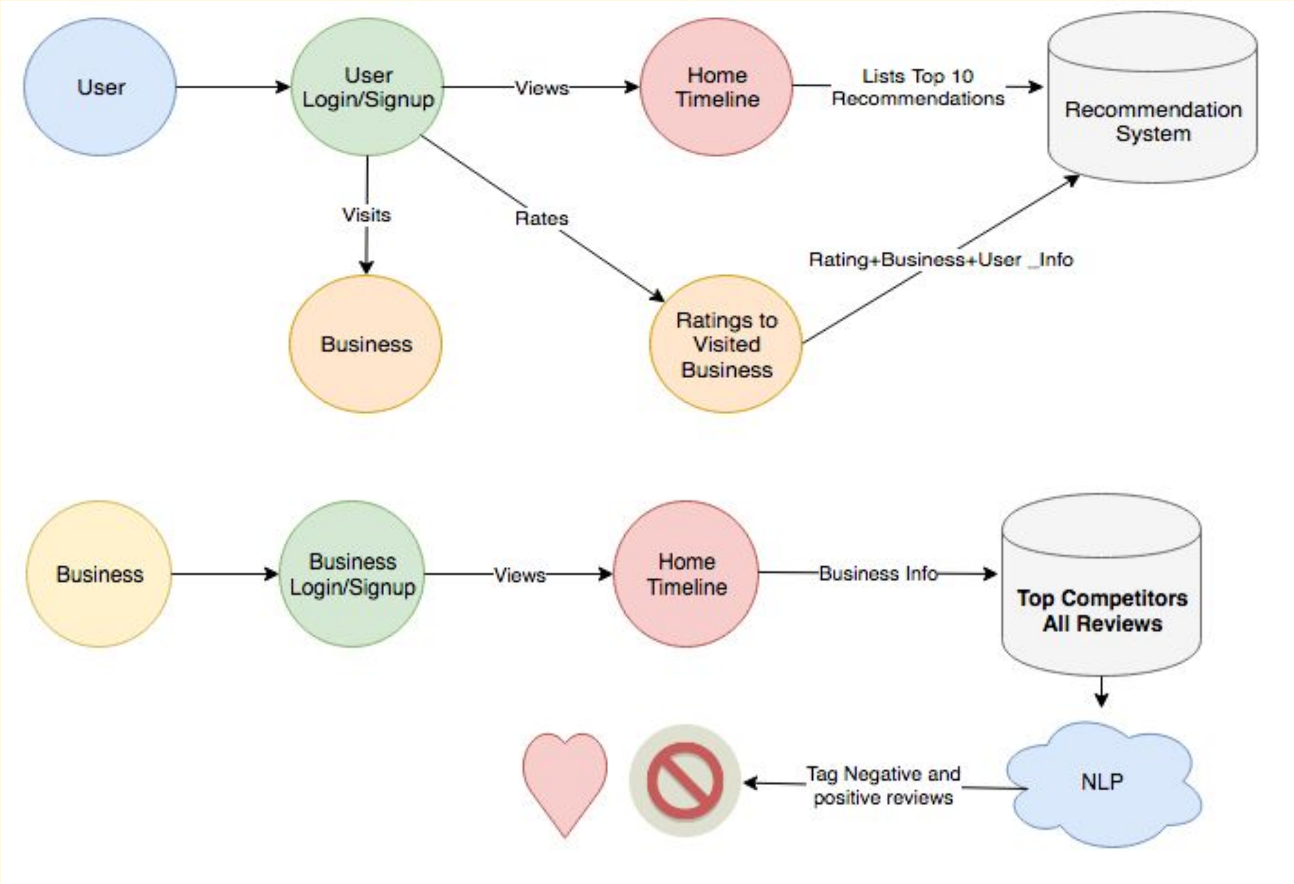
A trove of reviews, businesses, users, tips, and check-in data!

- 5,200,000 user reviews(five million two hundred thousand)
- 174,000 businesses(one hundred seventy-four thousand)
- Spans across 11 metropolitan area

Architecture

- Database: Neo4j
- Querying Language: Cypher
- Command line interface using Py2neo

Use Cases



User
User Login/Signup
Views
Home Timeline
Lists Top 10 Recommendations
Recommendation System
Visits
Rates
Rating+Business+User _Info
Business
Ratings to Visited Business
Business
Business Login/Signup
Views
Home Timeline
Business Info
Top Competitors All Reviews
Tag Negative and positive reviews
NLP

Conceptual / Logical design

Database: NEO4J

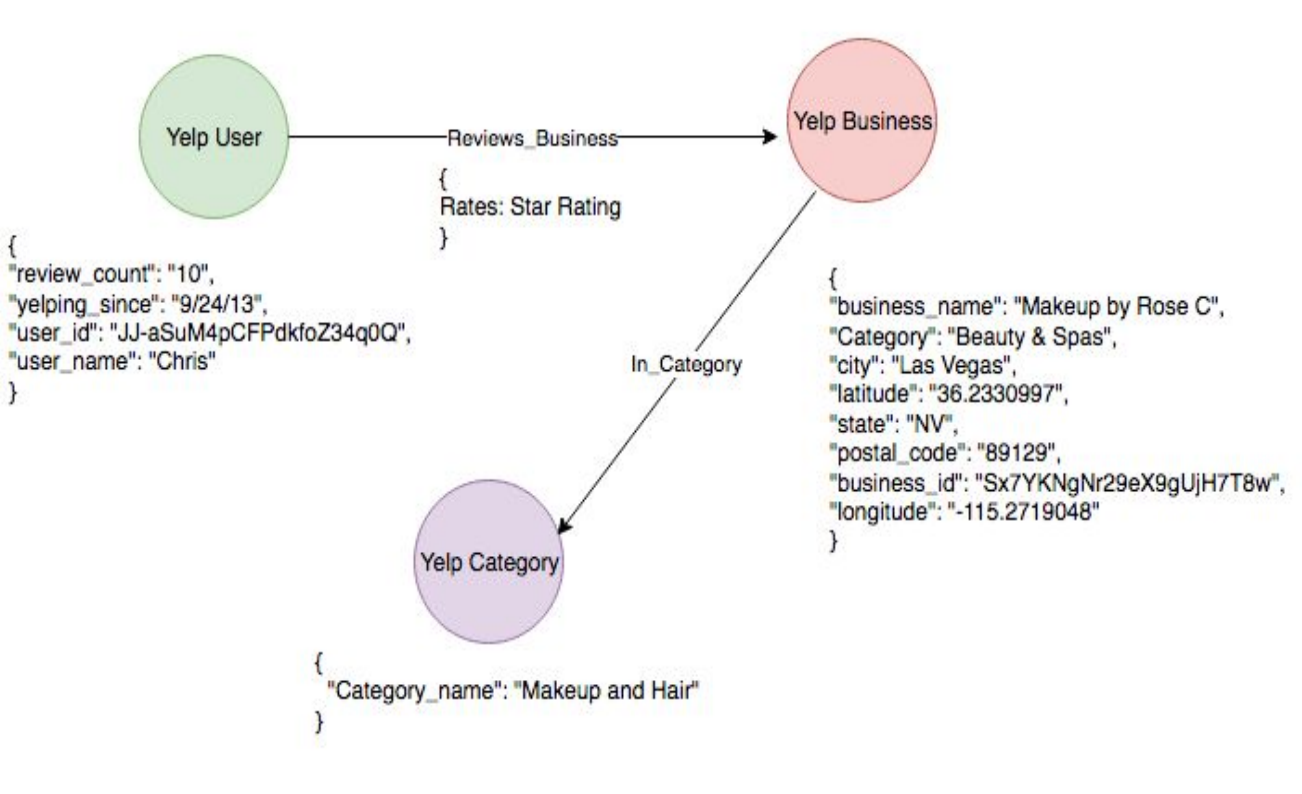
Querying Language: Cypher

Nodes:

1. Yelp User
2. Yelp Business
3. Business Category

Relationships:

1. User Reviewed Business
2. Business belongs to Category

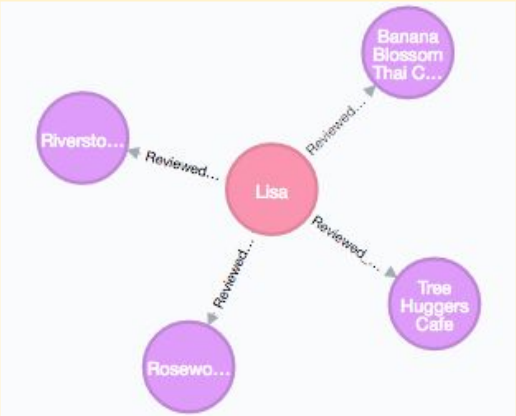


Graph based Recommendations

- <u>Index free adjacency:</u> allows for calculation of recommendations in real time. When you design your graph you explicitly add links between logically related nodes. These links contain fixed addresses to your neighbor nodes
- <u>No complicated joins:</u> Bye Bye Joins!
- <u>Flexibility</u>: Rather than exhaustively modeling a domain ahead of time, data teams can add to the existing graph structure without endangering current functionality
- <u>Performance</u>: With traditional databases, relationship queries will come to a grinding halt as the number and depth of relationships increase. In contrast, graph database performance stays constant even as your data grows year over year.

Basic Types of Recommendations

- Content based filtering (Business attributes)
- Collaborative filtering (user based)



Content-based Recommendations

- Recommend items that are similar to that the user is viewing.
- Similar businesses are found using the attributes of the business. For example in the yelp dataset every business is attributed to a particular category.
- Similarity based on common category.
- **Weighted Content Algorithm**: A business can have many traits similar to category that we can consider to compute similarity.
- The weighted sum score is used for recommendation.

Basic Content-Based Recommendation

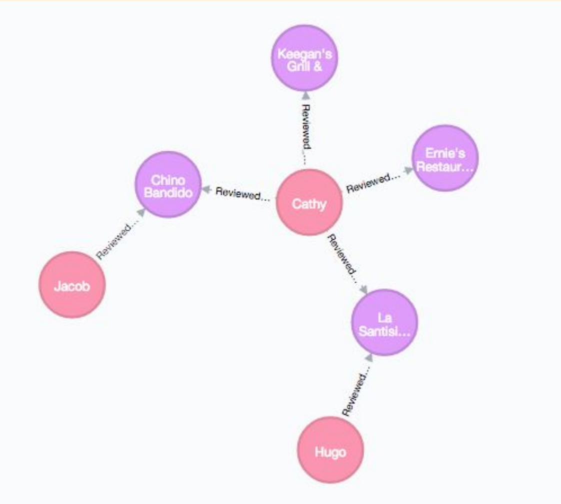
```
// Context Based recommendation based on business reviewed in past
MATCH (u:Yelp_User{user_name:"Mark"})-[rv:Reviewed_Business]->
(b:Yelp_Business)-[:IN_CATEGORY]->(c:Yelp_Category)<-[:IN_CATEGORY]-
(rec:Yelp_Business)
RETURN rec;
```

Collaborative filtering

- Use the actions of other users in the network to recommend
- “Users who bought this also bought this” kind of recommendation.
- Defining a way to evaluate similarity between two users (cosine similarity).
- Step 1: considering similarity between two users as the ratio of number of businesses they visited in common over the total business visited and rank the business as most popular to least popular.
- Step 2: Refining the way businesses are ranked based on the ratings given by the subset of similar users.
- Step 3: Users are considered similar if they visit the same business AND give similar rating to the business

Basic Collaborative filtering

```
//Colaborative Filtering for JACOB outputs recommended Business
MATCH(a:Yelp_User{user_name:"Jacob"})-[:Reviewed_Business]->
(b:Yelp_Business)<-[:Reviewed_Business]-(u:Yelp_User)
MATCH(u)-[:Reviewed_Business]->(rec:Yelp_Business) where not
exists((a)-[:Reviewed_Business]->(rec))
return rec;
```



Natural Language Processing using GraphDB

- Neo4j 3.0+ helped create user defined procedures that can be called from cypher command line.
- Reviews can be mined to know what is actually good at a particular place
- APOC Procedures and GraphAware’s neo4j-nlp-procedures have been used.
- Step 1: Detect language and update each text with that language.
- Step 2: Annotate all texts that is detected as English as the underlying library may not support other languages.
- Step 3: Perform sentiment analysis on the annotated text.

Sentiment analysis

```
$ MATCH (n:Positive) RETURN n.text LIMIT 25
```

Table

Text

</>
Code

n.text
"The Stadium is good Stadium and the atmosphere was amazing the people in Charlotte like to party at the game."
"After the game (Panthers lost) the Seahawks fans were partying and doing chants in front of the Stadium and the nice people of Charlotte just let them enjoy themselves and said good game!"
"Enter: Zizzo's!"
"After a bit of banter around how slim I wanted my sleeves tailored, Phil accommodated my requests and found a good compromise to keep me happy."
"What a hilarious and knowledgeable guy!"
"Fast forward to pick up (it wasn't an emergency so I picked it up a week later, could have been sooner if I needed it to be): I tried on the suit and it fit perfectly from every angle!"
"My wife and I regularly hit up Ah-So for Happy Hour."
"They have fresh and delicious rolls for half off."
"We get seated right away and always have delightful wait staff."
"If you enjoy bargain hunting half as much as I do, this is the place for you!"
"With new shipments in every week, there are plenty of great deals flooding into the store."
"I have found several great Hugo Boss suits at Last Chance, and each of them cost only $120!"
"If you can handle the "crowded garage sale" feel the store has, you can walk out with some really great finds!"
"Good service, good food."

Sentiment Analysis

```
$ MATCH (n:Negative) RETURN n.text LIMIT 25
```

Table

Text

Code

"It sucks that they don't provide oil on the grill anymore like they used to."

"When it was time to pay, I gave the same hostess that I had spoken to both the gift cards and she then told me that they actually don't accept them after running the cards and went to go speak to the manager about it."

"Long story short, this review is more focused on the customer service."

"I ended up paying my debit card which I didn't mind but next time, don't provide me false information... especially if you had another hostess agree with you, when in fact, were both wrong."

"Next time, if you don't know, ask the manager."

"I would've been more understanding."

"I'm not entirely sure if I'll come back."

"I lost some weight from my college days and I needed to find a great tailor to slim down my suits."

"I wear them every day for work and nothing makes you feel like a million bucks like a well-tailored suit."

"It helps to know what you want (length and width of suit sleeves, etc)."

"Phil works quickly and isn't afraid to give you his "honest" opinion."

"It was almost miraculous how much of a custom, handmade feel my newly tailored suit jacket had."

"I have already taken two more jackets to Zizzo's to get altered."

"We usually get the Snow White Roll and the Sweet and Spicy Crunch Roll and shared a bowl of fried rice--all for less than $20 (including tip)."

"Most of them have already been tailored (which is why they can't be resold at a regular Nordstrom) but it usually works out better that way for me."

"But if for some reason it doesn't fit perfect I still might only have to spend $30 getting some minor tailoring."

Started streaming 25 records in less than 1 ms and completed after 1 ms.

References

[1] http://guides.neo4j.com/sandbox/recommendations

[2] https://medium.com/neo4j/using-nlp-in-neo4j-ac40bc92196f

[3] https://github.com/graphaware/neo4j-nlp

[4] https://neo4j.com/docs/developer-manual/current/introduction/

[5]https://github.com/vikash4281/Java-Helps-Java/tree/master/Neo4j/Neo4J%20OGM%20HelloWorld

[6]https://www.forbes.com/sites/danwoods/2014/02/14/50-shades-of-graph-how-graph-databases-are-transforming-online-dating/#6467d8575081

Thank you