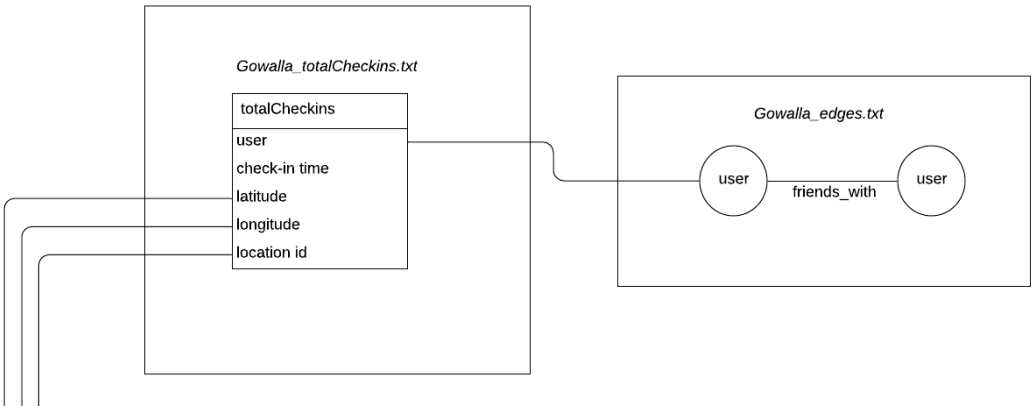
Project Model

Ayan Patel, Robert Hensley

We will be using the Gowalla Graph Dataset provided by the Stanford SNAP project. Our full Project Model can be found here.

Provided Data



Gowalla_totalCheckins.txt

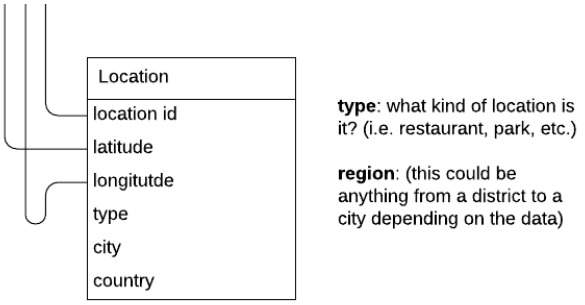
- User
- Check-in time
- Latitude
- Longitude
- Location

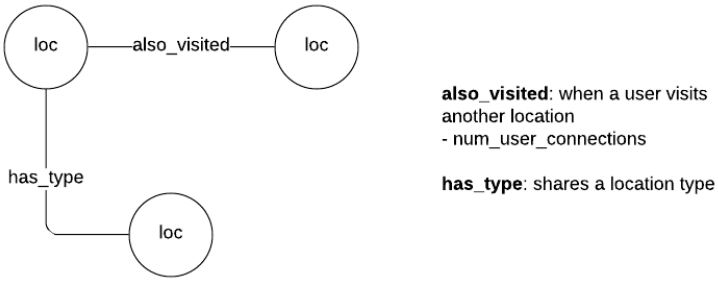
Gowalla_edges.txt

The edges of users who are friends on Gowalla. Nodes are user ID’s.

Generated Data

Location Data

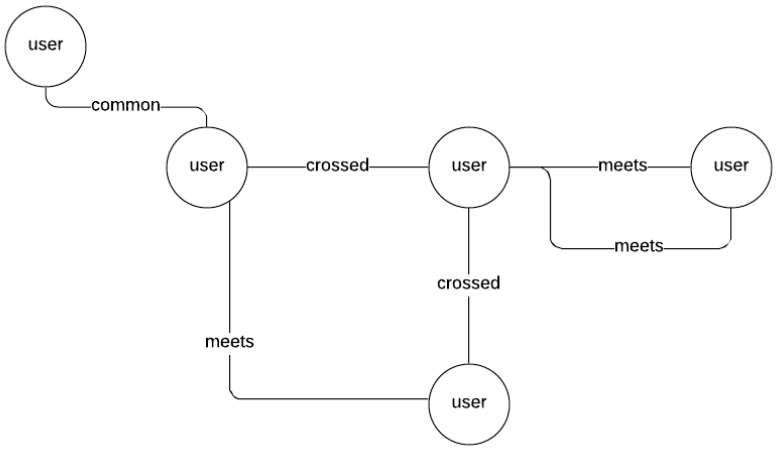




We will create location nodes from the totalCheckins dataframe. Using a map API, we will gather more information about the location id based on it's latitude and longitude information. Information that we will query includes what type of location the node is (i.e. is it a landmark, a restaurant, a park), the city of the location and the country of the location.

With this information, we will create a new graph of locations. This will include how locations are connected by visitors, through the *also_visited* relationship. This will show the number of people who have also visited another location. We will also show the relationships of common attributes, with the *has_type* relationship.

User Data



meets: friends who meet up at the same place at roughly the same time (~30 minutes +/-)
- average time / date
- location (location id)

crossed: users who are not friends are at the same location at the same time
- average time / date (~ +/- 30 minutes)
- location (location id)

common: users who have visited the same location, although not at the same time
- location id

We will create a graph of user interactions based on the check-in and friendship data provided. We will make three different types of relationships between users:

1. Meets: friends who meet at approximately the same time (~30 minutes apart)
2. Crossed: people who aren't friends who were in the same location at approximately the same time (~30 minutes apart)
3. Common: people who have been at the same place at different times

How can we use this model to answer our questions?

<u>Popularity</u>

In terms of popularity we can use the user's friends to identify what user is connected to the most amount of users. We can also look at friends of friends to identify if they have similar interests based on checkin locations

Location

We can use the locations to identify the most popular locations based on the number of users that visited a particular location. We can also use our graph to determine if a user has visited multiple locations and we can check the type of location to infer what places a user likes. This can show us if a group of people prefer a certain type of location to meet up at.

Time

We can use the timestamp to identify at what time of day, or even what season, different users prefer to go to different locations. We can also compare if users that are friends go to the same locations at the same time of day or at different times of day.