SightWalk

Enjoyable Walking Paths in New York City with Social Media Data

Sebastian Brarda, Felipe Ducau, Ma. Elena Villalobos Ponte

NYU CENTER FOR DATA SCIENCE

ABSTRACT

Efficiency has become a cult in our society. We are always striving to do things faster, quicker, and with less cost. In this context, GPS and map systems such as Google Maps, Apple maps and Waze, are not the exception. When providing directions to a place, these systems suggest the shortest route in terms of distance and/or time. However**, the fastest route is not always the most enjoyable**, or the one that maximizes utility for users. The shortest path is not always the one that connects us with nature, makes us meet new people, or help us discover new places.

The **goal** of this project is to create a system to automatically suggest alternative routes which might take marginally more time but **go through more attractive spots for the user**. The main application (although not the only one) would be for tourists visiting the city. Visitors, when exploring the city, eventually need to go from point A to point B, but they do not necessarily want this trip to be optimized with respect to time, they probably want to maximize their experience while walking.

PROBLEM FORMULATION

We tackled the problem in two steps:

- First defining an "**interestingness**" or "**attractiveness**" **score**. By extracting features from social data and based on the user input we will compute new edge weights.
- Second, we created **two graph based optimization algorithms** that will compute alternative paths based on these weights.

DATA PREPROCESSING

We downloaded the metadata from ~200k photos from Flickr taken in the last two years. Later, we scrapped the amount of "likes" per photo and the static photo URL that were not available through the API. Regarding the Graph of NYC, we use coordinate nodes from OpenStreetMap

flickr API INTEGRATION

OpenStreetMap

Later, we used GeoPandas to map each photo to the closest "edge" (square) in the city. Finally, we computed aggregated metrics for the main social metrics at the edge level.

These metrics are "Likes", "Views", "Number of photos", and "comments". All these metrics were computed both for NY Flickr accounts and Tourists accounts which were identified upon.

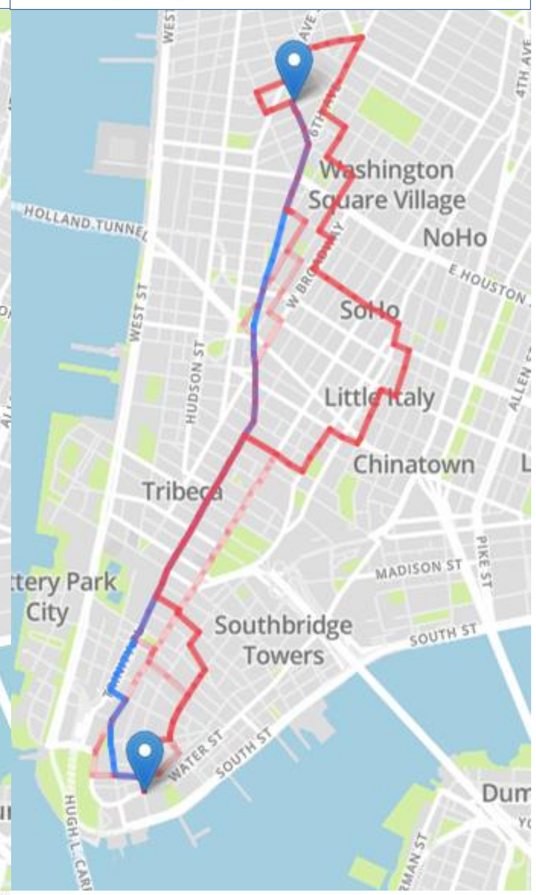
SEARCH ALGORITHMS

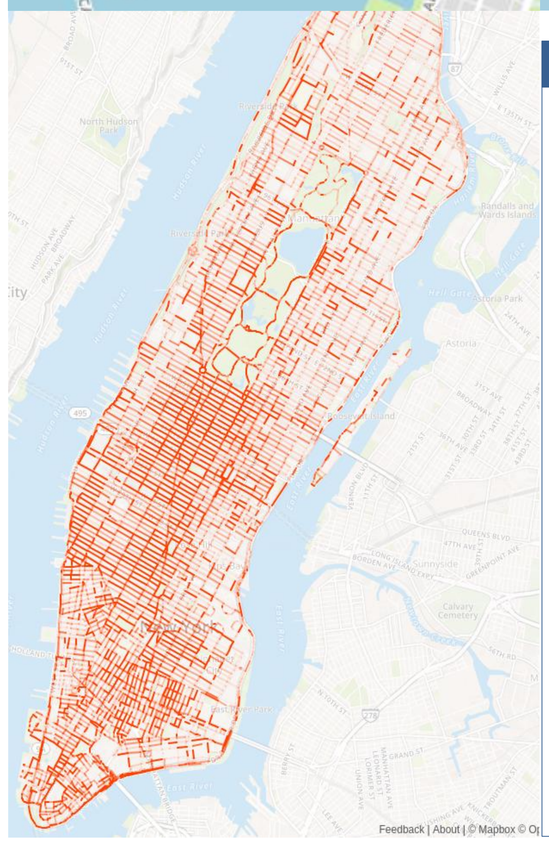
Randomized Dijkstra

Eliminate edges randomly

A*

Mapping weights to node potentials





SCORING FUNCTION

Defining a function to determine the level of "interestingness" is a difficult task, since each person have **different preferences**.

However, given a certain interestingness metric, we can evaluate how efficient Graph algorithms are, and Human Evaluation can determine whether the function is working well.

$$\textbf{Uniqueness} = \frac{\#\,\textit{Comments} + \#\,\textit{Likes}}{(\#\,\textit{Photos})\text{^}2}$$

$$\textbf{Popularity} = \#\,\textit{Photos}$$

The axis Touristic vs. Local is based on the origin of the user account..

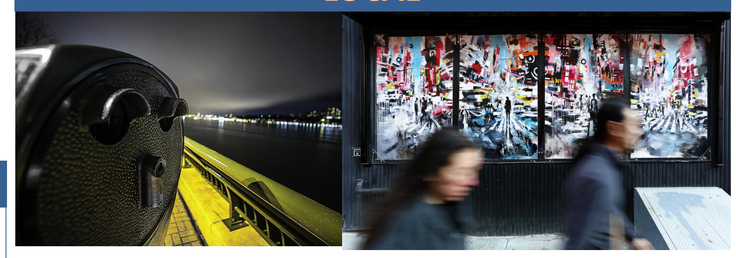
All data was normalized so that total interestingness of the city sums to 1, but the weight by edge will change depending on user parameters.

POPULAR

UNIQUE

TOURISTIC

LOCAL



EVALUATION

So far we have evaluated the performance of the algorithms given the interestingness function, showing that the A* algorithm works better overall.