Mapping Human Mobility Using Social Media Data

Group members: Zhenlong Li, Ziqi Liu, Chunzhu, Wei, Weixing Zhang, Tianqi Zou
Mentors: Kiumars Soltani, Hao Hu, Shaowen Wang

CyberGIS Center for Advanced Digital & Spatial Studies

University Consortium for GEOGRAPHIC INFORMATION SCIENCE

Background

With the prevalence of various social media, those Big Geospatial Data sources provide us a different way to study people activities with urban environment. However, it is a great challenge to extract useful information from these unstructured huge volume of data. Our research aims to build a bridge between twitter data and human mobility, traffic pattern specifically, and find out to what extent can social media data be used in geospatial research.

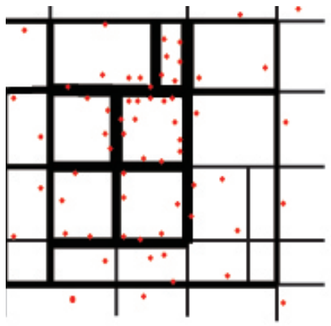
Exposed to three-year tweets data with userID, geo-location and time in Chicago area, we use multiple ways to estimate traffic. In summerizing total trajectories and assigning them to each road segment, super computation enables us to finish this process more efficient.

Social Media Data

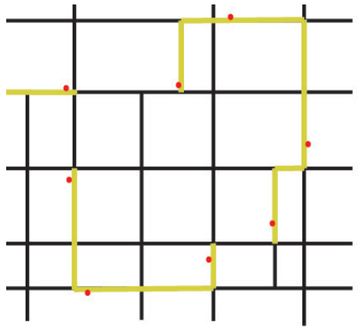
Tweets Data

How to estimate traffic by tweets data

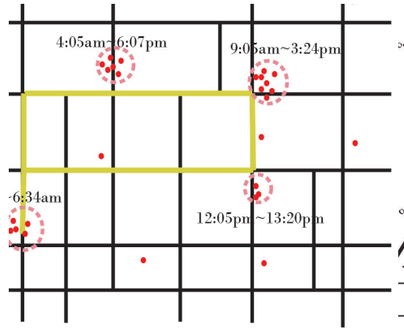
Tweets Density

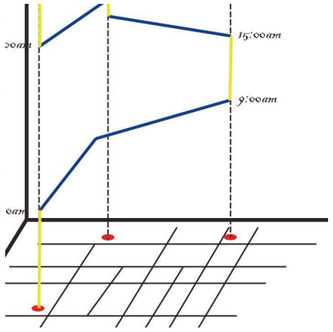


Trajectory Segment



Daily Trajectory (Theoratically)





Human Mobility

Traffic Data

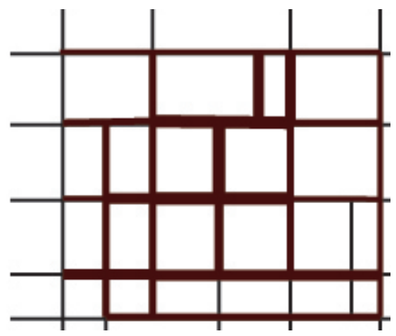
How to find shortest path for each trajectory

Method 1
ArcMap Model Builder

Create road network dataset, solve for the shortest paths for each trajectory, and aggregate the usage of road segments as traffic projection.

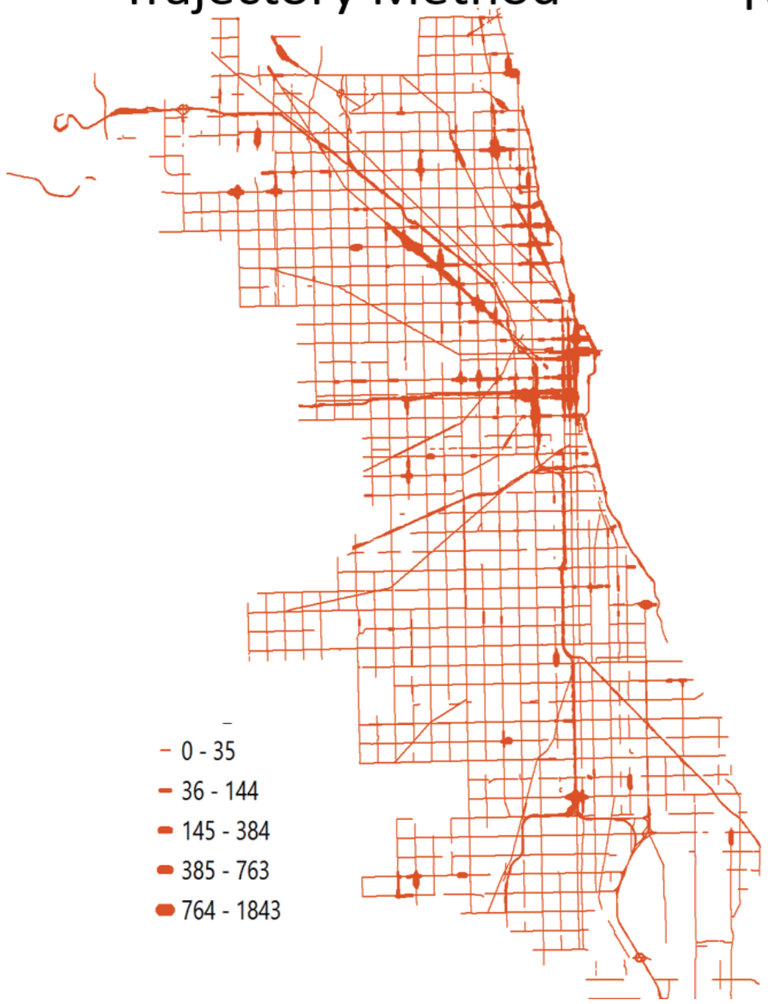
Method 2
Google Map

Write a program in Python to run snap to road_-Google maps automatically, find the the shortest path by inputting starting and ending point and return and coordiante of each road segment.

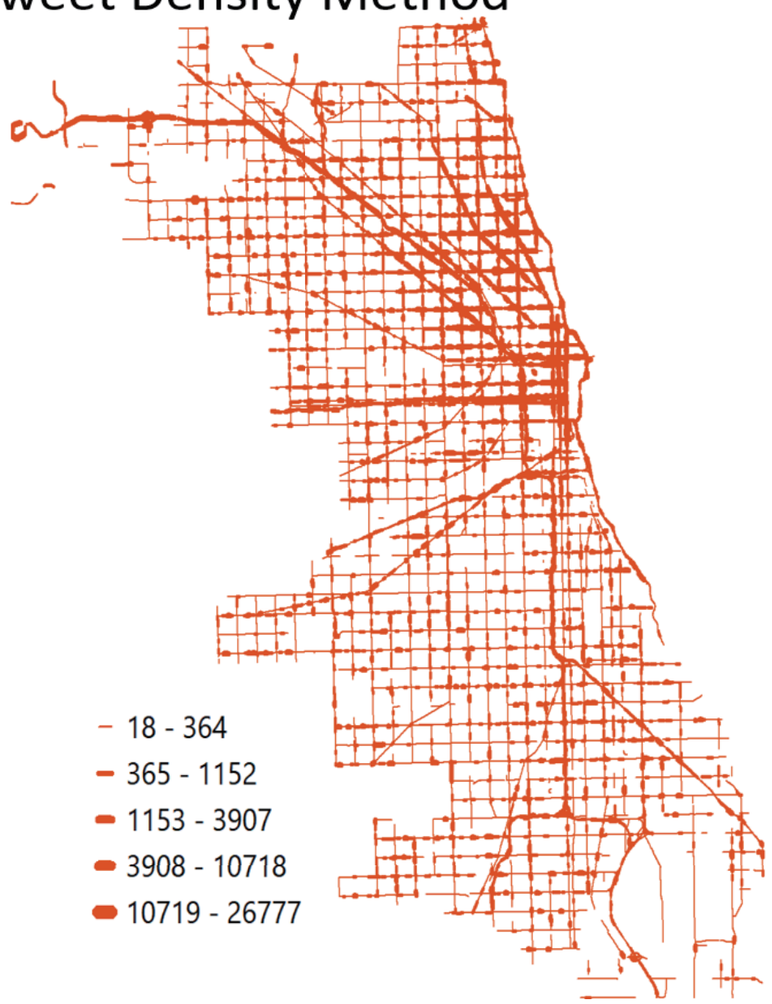


Spatial Correlation

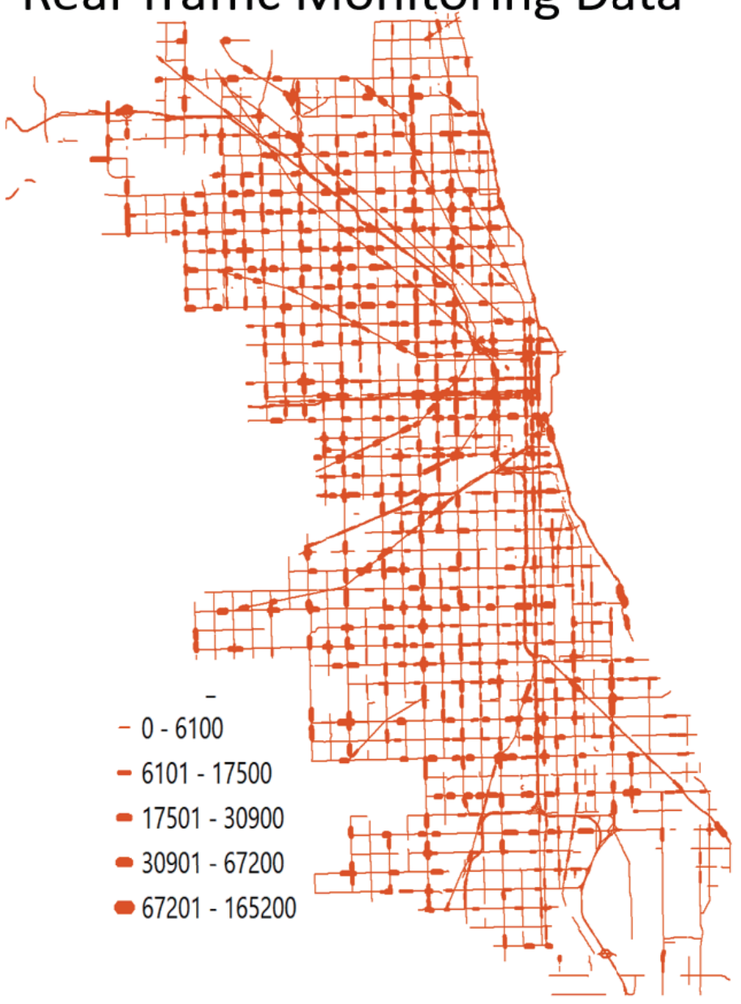
Trajectory Method

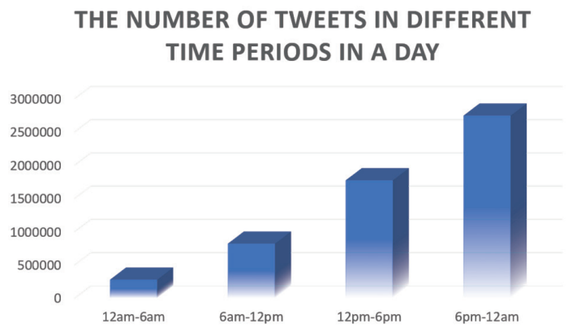


Tweet Density Method



Real Traffic Monitoring Data





Conclusion

(1) Visualization of general features shows that twitter users in Chicago Area are most active in the evening, followed by afternoon, morning and midnight.

(2) The result of auto-correlation shows that traffic estimation by tweets density has a stronger correlation with real traffic data than estimation by trajectory method.

Limitation

(1) In this research, traffic data and twitters data were collected in 2006 and 2013-2016 respectively. The time lag may cause the inconsistency in our result.

(2) Constrained by Google limits of authority, we only run around 10,000 trajectories and get the traffic map by second method. The limited number of trajectories may influence the accuracy of traffic, which leads to the less correlation with real traffic data.