Big Flow Data Visual Analytics Through TrajAnalytics

Xinyue **Ye**, Ph.D.,
Associate Professor
Department of Geography & School of Digital Sciences,
Computational Social Science Lab, Kent State University &
Center for Geographical Analysis, Harvard University
ResearchGate: https://goo.gl/udvnKo
Email: xye5@kent.edu

Kent State University
COMPUTATIONAL SOCIALSCIENCE LAB

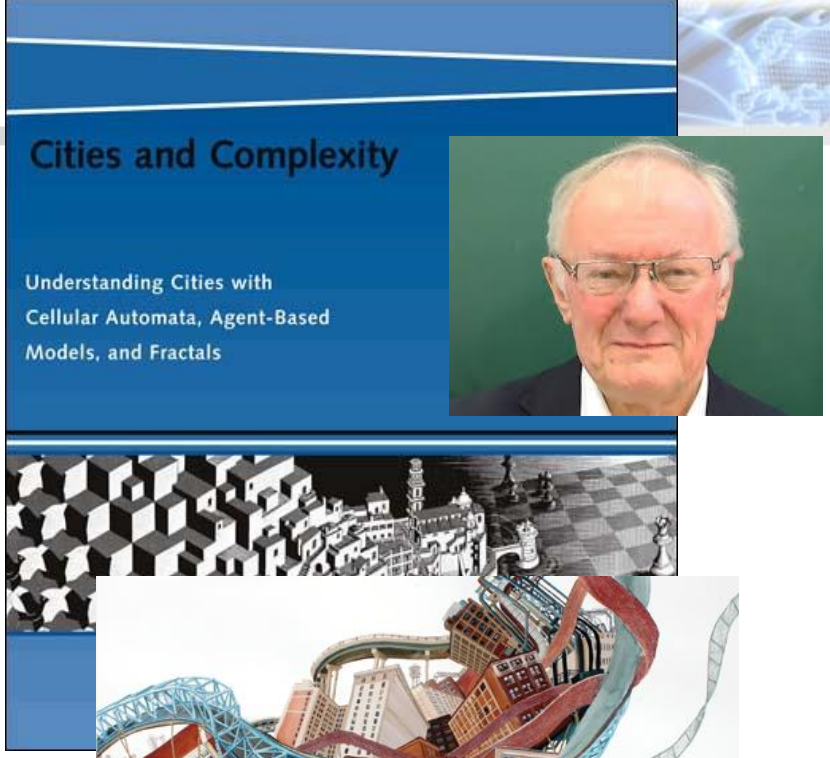
Outline

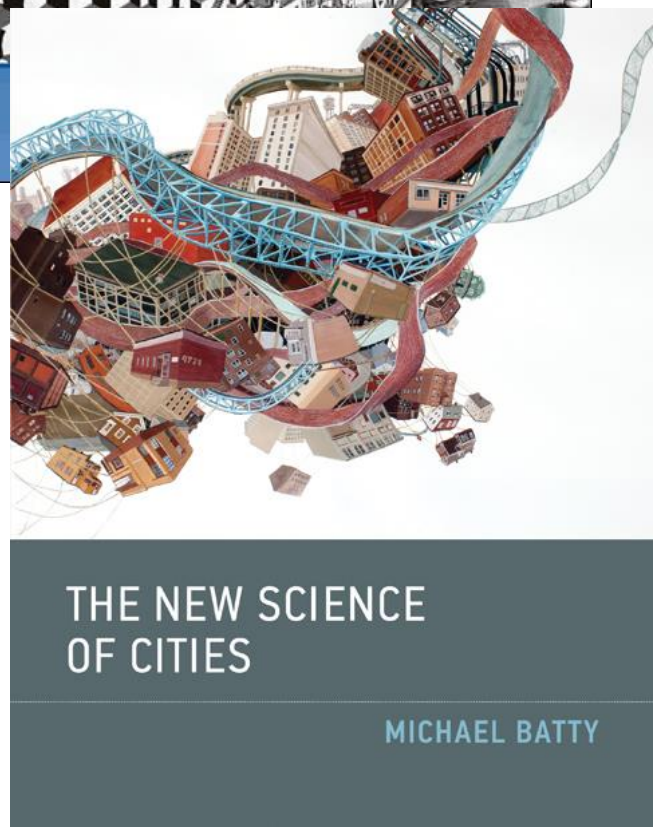
- Overview
- Methodology and Intellectual Merits
- Software Development
- Demos on TrajAnalytics

Human Dynamics (Shaw, Tsou, and Ye, 2016)

- A transdisciplinary research field focusing on the understanding of dynamic patterns, relationships, narratives, changes, and transitions of human activities, behaviors, and communications.

- **Human** as the central element connecting spatial and social networks.





As urban planning moves from a centralized, top-down approach to a **decentralized, bottom-up** perspective, our conception of **urban systems** is changing.

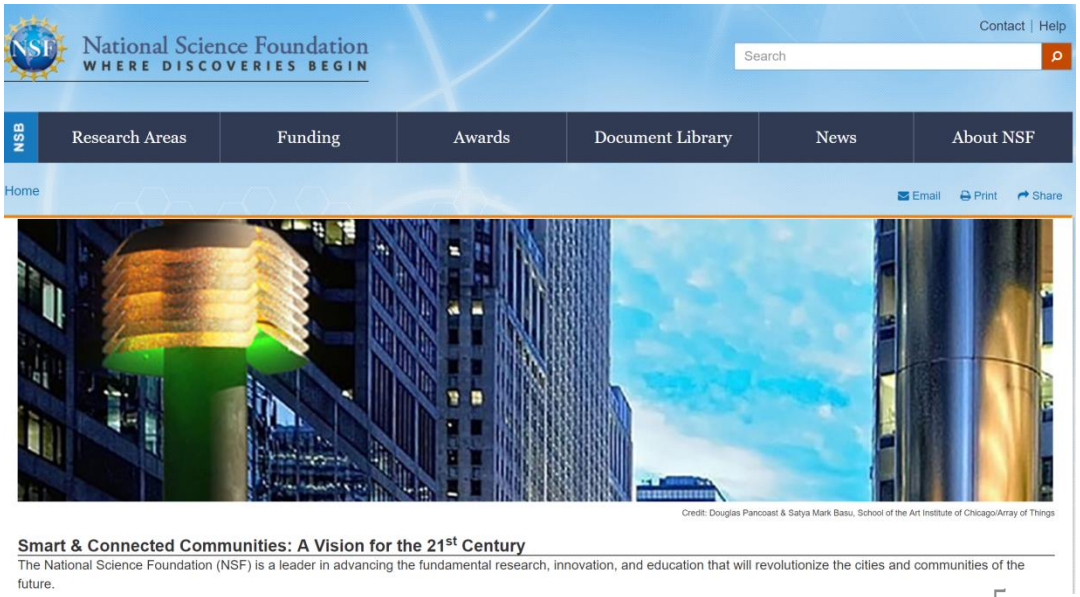
Batty, M. (2005). Cities and complexity: understanding cities with cellular automata, agent-based models, and fractals. The MIT Press.

To **understand cities** we must view them not simply as places in space but as **systems of networks and flows.** To understand **space**, we must understand **flows**, and to understand flows, we must understand **networks**—the relations between objects that comprise the system of the city.

Batty, M. (2013). The new science of cities. The MIT Press.

Cities and communities in the U.S. and around the world are entering a new era of transformational change, in which their inhabitants and the surrounding **built and natural environments** are increasingly **connected** by **smart** technologies, leading to new opportunities for **innovation, improved services, and enhanced quality of life**.

We are living in the *Urban Millennium*. The world's population is increasingly concentrated in urban areas. By 2050 it is projected that 64% of the developing world and 86% of the developed world will be urbanized (The Economist, 2012).



- 2017-2019, SI2-SSE: **GeoVisuals** Software: Capturing, Managing, and Utilizing **GeoSpatial Multimedia Data** for Collaborative Field Research

- 2016-2018, S&CC: Support **Community-**Scale **Intervention** Initiatives by Visually Mining Social Media Trajectory Data

- 2015-2018, SI2-SSE: Collaborative Research: TrajAnalytics: A Cloud-based Visual Analytics Software System to Advance **Transportation** Studies Using Emerging **Urban** Trajectory Data

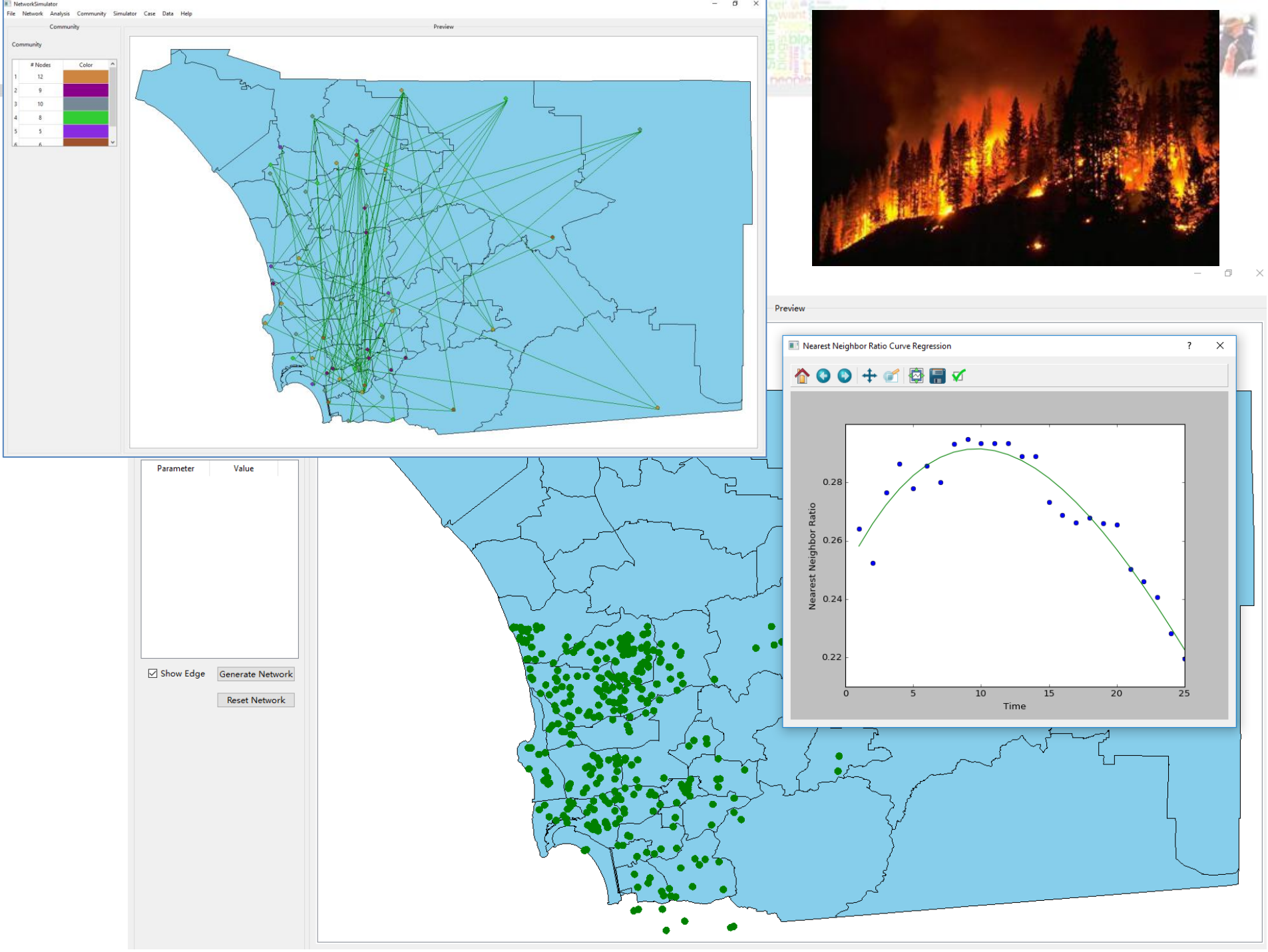
- 2014-2019, IBSS: Spatiotemporal Modeling of **Human Dynamics** across Social Media and Social Networks

Overview

- Emerging urban trajectory data
- Visual analytics software
- Advancing transportation studies using trajectory data

Y. Zheng, L. Capra, O. Wolfson, and H. Yang. Urban computing: Concepts, methodologies, and applications. *ACM Transactions on Intelligent Systems and Technology*, 2014.

"when facing multiple types and huge volume of data, how exploratory visualization can provide an interactive way for people to generating new hypothesis becomes even more difficult. This is calling for an integration of instant data mining techniques into a visualization framework"


NetworkSimulator
File Network Analysis Community Simulator Case Data Help
Community
Preview
Community
Nodes
Color
1 12
2 9
3 10
4 8
5 5
Parameter
Value
Show Edge
Generate Network
Reset Network
Preview
Nearest Neighbor Ratio Curve Regression
Nearest Neighbor Ratio
0.28
0.26
0.24
0.22
0
5
10
15
20
25
Time

Transportation Studies Can Be Transformed by Emerging Urban Trajectory Data I

- With the prevalent GPS, Wi-Fi, Cellular, and RFID devices, population mobility information is recorded as the moving paths of taxis, fleets, public transits, and mobile phones.
- Conventional transportation studies are conducted by (1) identifying the factors that influence transportation and studying their effects through empirical models or survey methods, and (2) using simulation products to evaluate road networks, where users have to specify complex road attributes and trial-and-error processes are demanded. In contrast, the emerging urban trajectory data provides *real* situations from which the statistics of real traffic flow can be extracted and city-wide transport patterns can be discovered.

Transportation Studies Can Be Transformed by Emerging Urban Trajectory Data II

- Nowadays, many trajectory data sets are collected by transportation administrations, companies, and researchers. Some of them are available for public use in research. In the long run, we will see more and more such data with the widespread use of trajectory recording devices and systems.
- Exploiting the emerging data can play a transformative role in transportation-associated research by offering researchers and decision-makers unprecedented capability to conduct data-driven studies based on real-world information.
- Robust, easy-to-use software enabling effective exploration of the data is needed and will contribute to building capacity in seeking solutions for the social, economic, and environmental challenges facing our communities.

Visual Analytics Software is Needed

- 33,000 Beijing taxis for 3 months: trajectories length 400 million kilometers and GPS points 790 million.
- Rich and heterogeneous information can be associated at each position over urban network.
- Integrates scalable data management and interactive visualization with powerful computational capability.

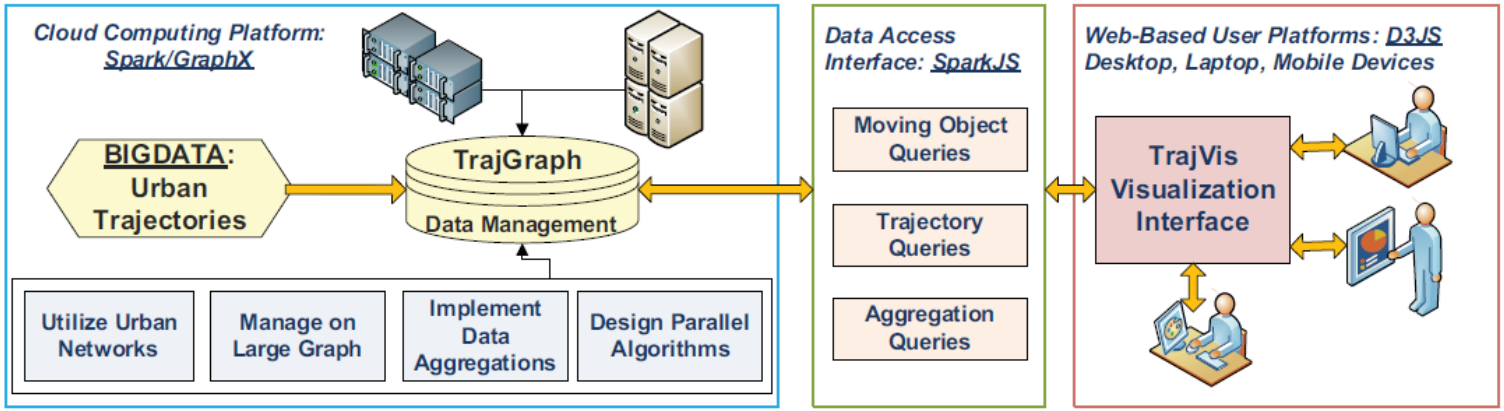
Software requirements

- Powerful computing platform so that domain users are not limited by their computational resources and can complete their tasks over daily-used computers or mobile devices.
- Easy access gateway so that the trajectory data can be retrieved, analyzed and visualized by different transportation researchers, and their results can be shared and leveraged by others.
- Scalable data storage and management which support a variety of data queries with immediate responses.
- Exploratory visualizations that are informative, intuitive, and facilitate efficient interactions.
- A multi-user system which allows simultaneous operations by many users from different places.

TrajAnalytics for Advancing Transportation Studies Using Trajectory Data

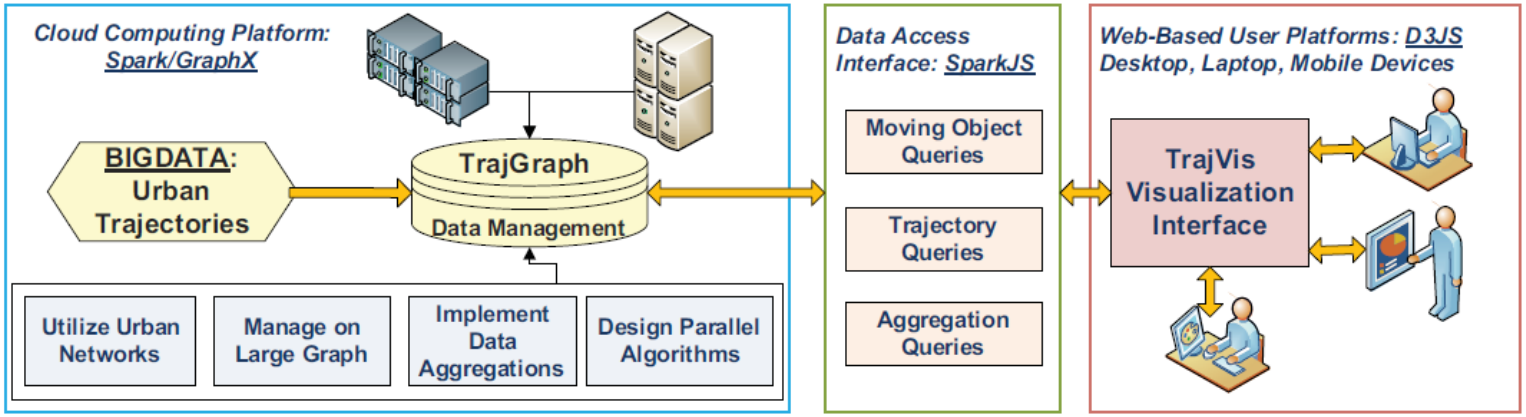
- *TrajGraph*: a scalable *parallel-graph* database designed for big trajectory data management on cloud platforms. Support fast computation over various data queries in a remote and distributed computing environment;

- *TrajVis*: an interactive visualization interface for exploratory data analysis and sharing. Visually query the data stored in TrajGraph, discover and analyze patterns, generate and evaluate hypotheses, and share their insights with others.



Software Engineering Process

- Employ Apache Spark for large-scale data processing on clusters.
- Use Spark's graph processing package, GraphX, in graph-based computation.
- Use D3, the standard visualization library using JavaScript and SVG, to implement visualization tools.
- TrajVis and TrajGraph will be linked through SparkJS, a library built on JavaScript runtime for interacting with the Spark cloud in browsers.
- Utilize the well-known open-source packages of Spark/GraphX, D3, and SparkJS to implement an efficient system for data-intensive real-time tasks that run across distributed devices.
- Create a public-licensed software system freely accessible to domain users under the BSD licenses.



intellectual merits are three-fold

- A cloud-based computing platform where users do not need to specifically store and manage the big data by themselves.
- A parallel graph data model enabling efficient data management of large-scale urban trajectories on a cloud-based database.
- A visualization interface on the database that supports a variety of visual analytics tasks on big trajectory data through interactive visual queries and other interactions.

TrajAnalytics on Clouds

- Very big data requires high-end computational resources for storage and intensive computation. However, domain users usually lack sufficient budget and time to scale up their computational capacity and skills.

- TrajAnalytics will be a software system executed on many compute nodes which can utilize remote cloud platforms to provide users attractive and economical SaaS (Software as a Service).

- Alternatively, the software can also be downloaded for use when local clusters are available. The software will make the computational system transparent to users so that they can complete visual analytics tasks through an interactive online system with desktops, laptops, or even mobile devices.

Parallel TrajGraph Model

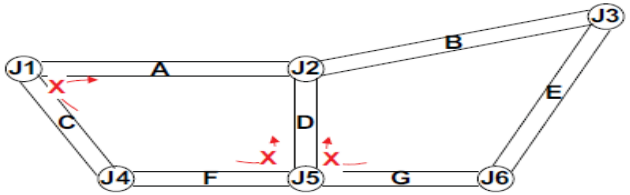
- **Utilize a parallel graph model for trajectory datasets:**

Although a city network usually has a smaller scale of nodes/links compared to web and social graphs, trajectory data becomes very large with continuous recording and associated heterogeneous information. It will be based on the Bulk-Synchronous Programming (BSP) model for large-scale graph computing.

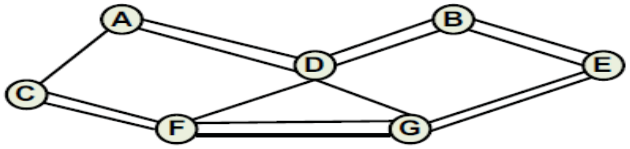
Parallel TrajGraph Model

- **Create vertex centric graphs from urban networks:**

Create TrajGraph from the networks by mapping road segments to vertices and adding edges between two connected vertices. Trajectory data is then stored efficiently over TrajGraph facilitating fast access in the vertex-centric processing.



(a) Road network



(b) Mapped to TrajGraph

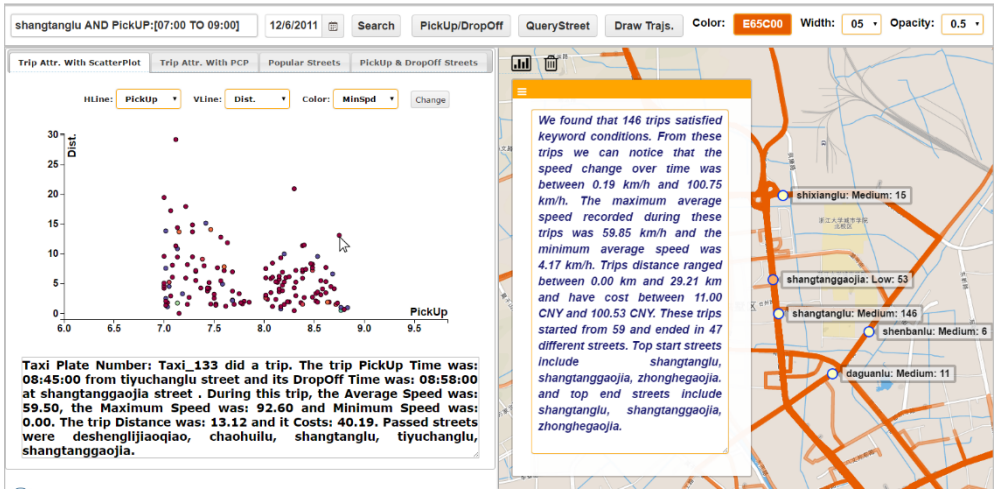
Parallel TrajGraph Model

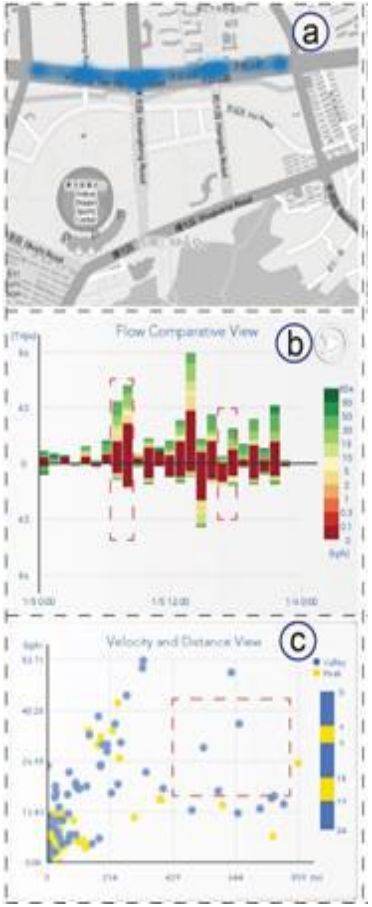
- **Support various and concurrent data queries and aggregations:**

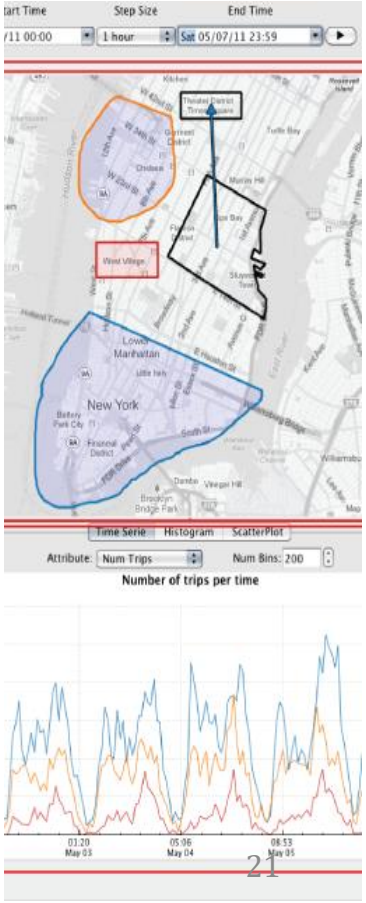
Optimal data structures and algorithms will be developed to efficiently process different types of data queries, including road network queries, moving object queries and trajectory queries. To enable interactive visual exploration, we will further design aggregation techniques over spatial and temporal dimensions to support precomputation and caching of data summaries.

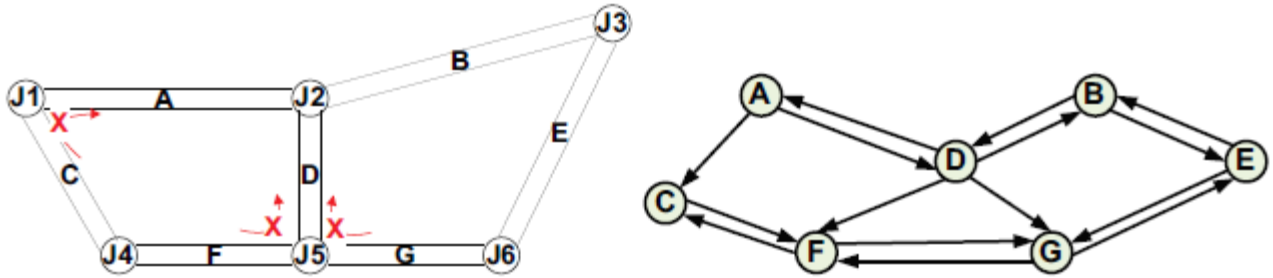
We focus on **how people directly interact with the data – *store*, *query* and *visualize* the results**

- Road Statistics Queries
- Moving Object Queries
- Trajectory Queries



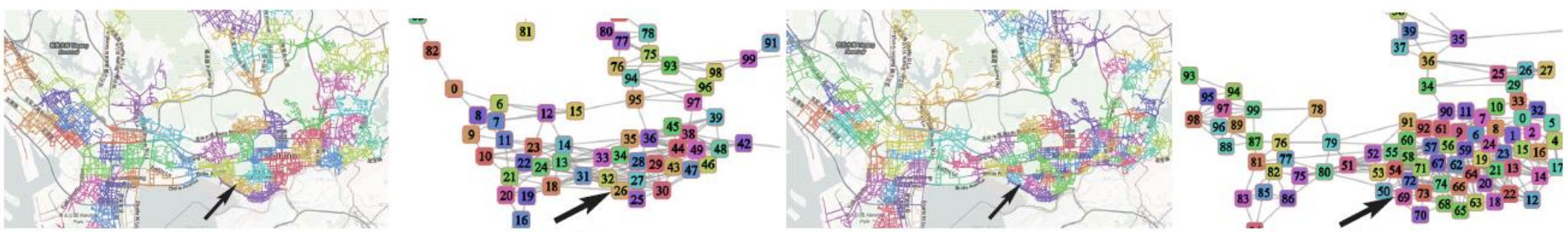






(a) A street network (b) A corresponding graph

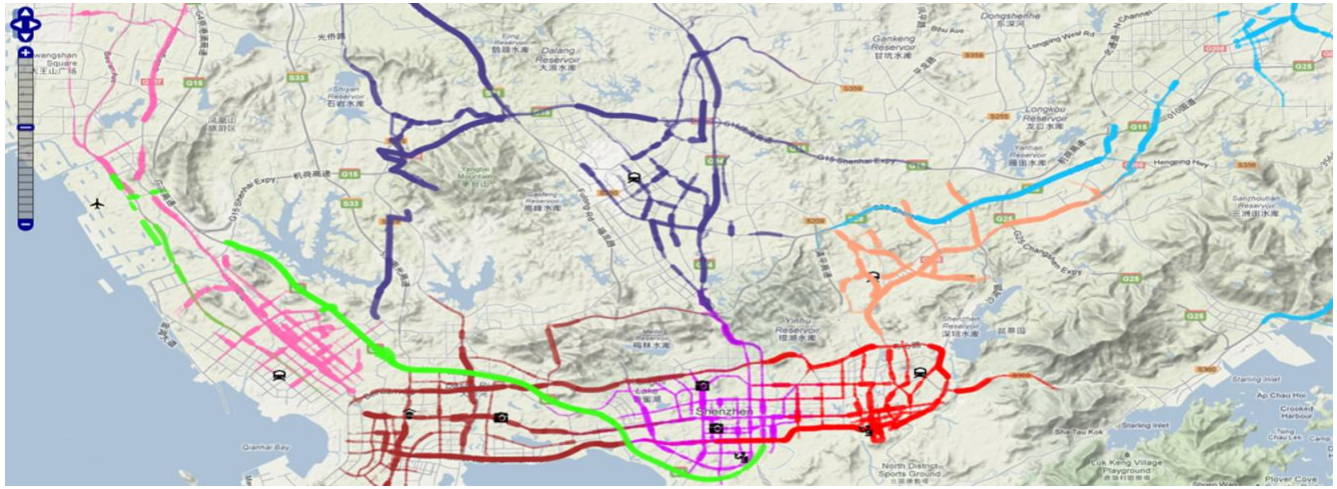
Fig. 1. *Using graph to represent a street network in TrajGraph.*



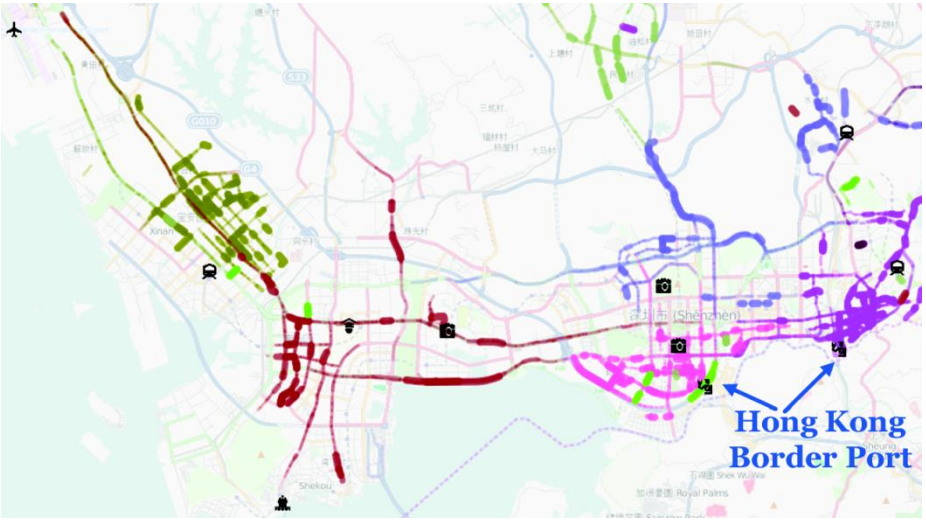
(a) (b) (c) (d)

Fig. 2. *Creating region level graph of ShenZhen by graph partitioning: (a)(b) without traffic information; (c)(d) with traffic information. More nodes (i.e., regions) are generated in the arrowed downtown areas in (c)(d) than in (a)(b). Colors are selected to show different regions on the map.*

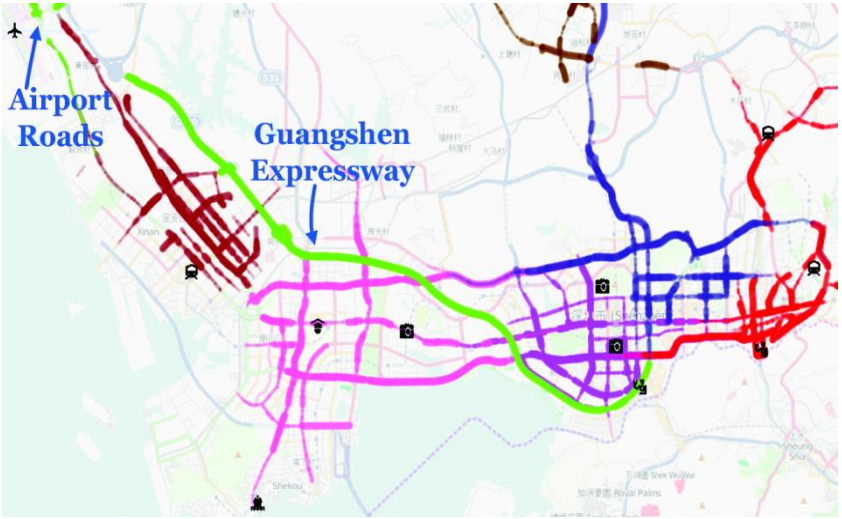
- Reveal typical traveling patterns of city cabs



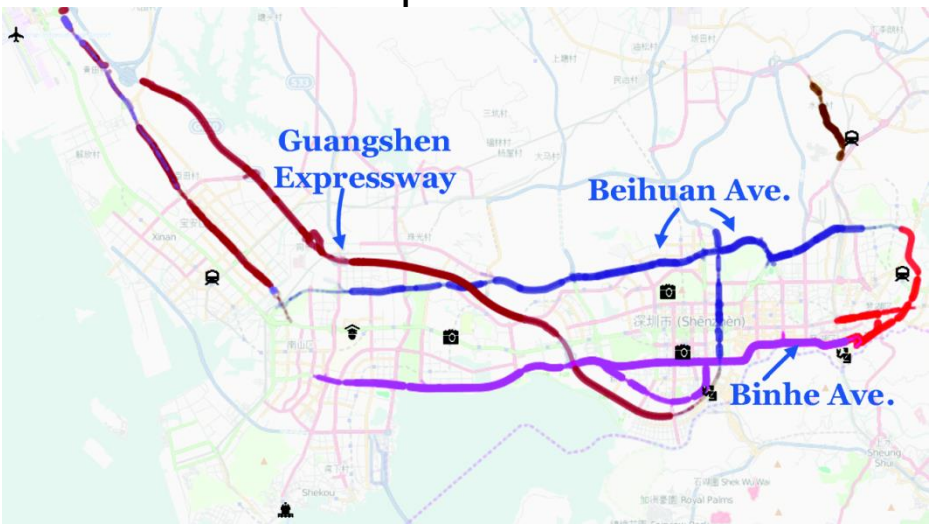
- The topics approximate the city's functional regions
- Topics are more than geometrical divisions
  - An airport highway (Green) is an important component of several topics, connecting different regions



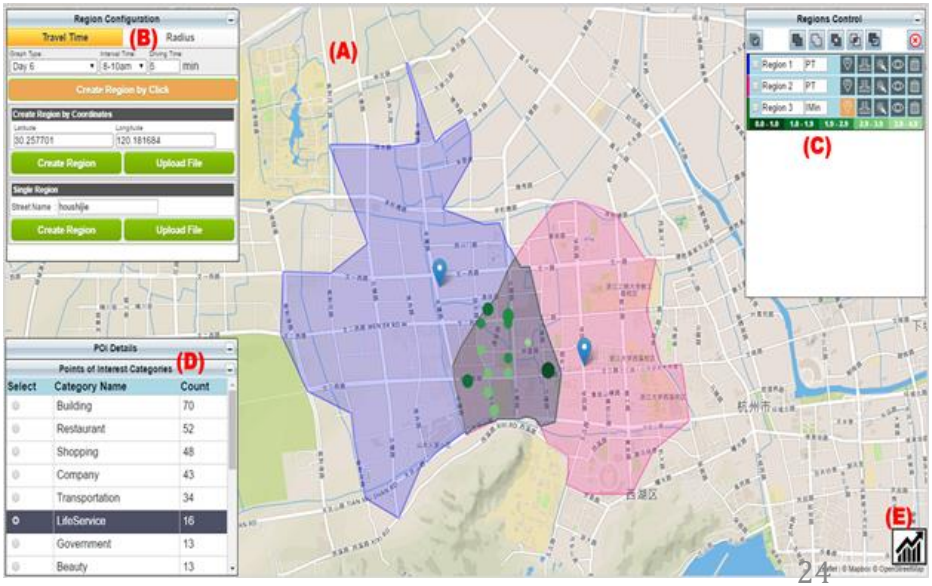
Vacant topics: 3am-6am



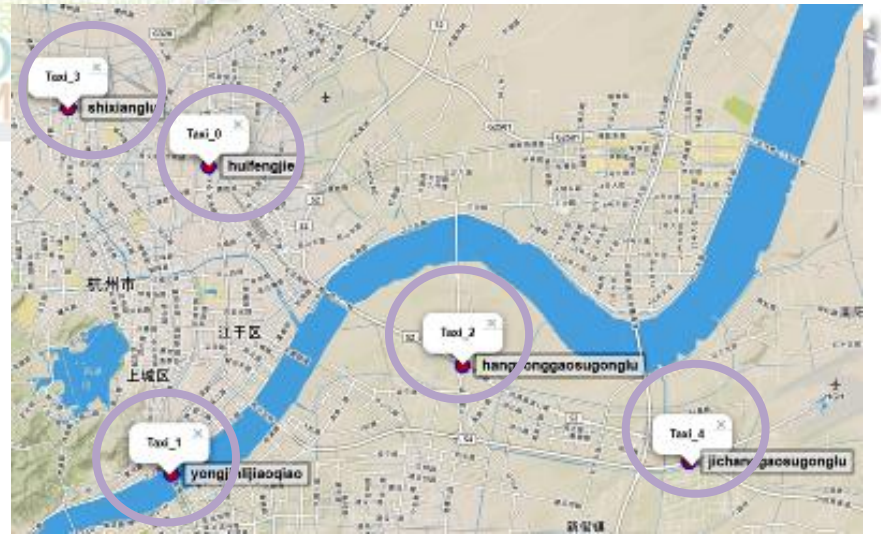
Occupied topics: 6am-9am



Occupied topics: 3am-6am



Latitude	Longitude	Street Name
22.543	113.991	Qiaocheng East Rd
22.546	113.997	Qiaoxiang Rd
22.568	114.066	Beihuan Ave
22.568	114.067	Beihuan Ave



$$
\begin{aligned}
&\text{speed} < 0.01\text{Km/h} \mapsto \textit{Stop},\\
&0.01\text{Km/h} \leq \text{speed} < 20\text{Km/h} \mapsto \textit{Slow},\\
&20\text{Km/h} \leq \text{speed} < 60\text{Km/h} \mapsto \textit{Normal},\\
&60\text{Km/h} \leq \text{speed} < 100\text{Km/h} \mapsto \textit{Fast},\\
&100\text{Km/h} \leq \text{speed} \mapsto \textit{Very Fast}.
\end{aligned}
$$

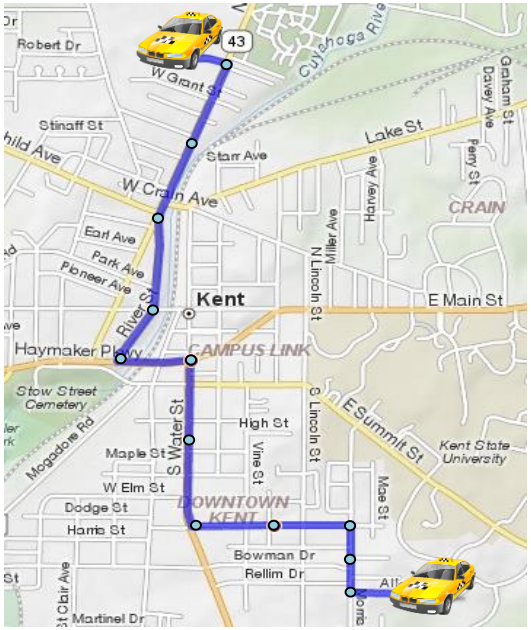
Slow VeryFast

Index of a trip document of a taxi

Taxi Plate Number				
Pick-up Street		Pick-up Time		
Drop-off Street		Drop-off Time		
Travel Distance		Fare		
Street Names → S1	S2	S2	S4	… …
GPS → 22.533 114.044	22.533 114.046	22.532 114.049	22.532 114.050	… …

Index of a trajectory document of a taxi in a given time period

	Taxi Plate Number				
Street Names →	Street1	Street1	Street2	Street2	… …
Speed →	18.2	13.4	70.3	110.1	… …
DSpeed →	Slow	Slow	Fast	Very Fast	… …
Empty or Not →	Y	N	N	N	… …
GPS →	22.533 114.044	22.533 114.046	22.532 114.049	22.532 114.050	… …



Reproductions, replications, or generalizations of seminal or pivotal studies that have served a demonstrably critical role in conceptual or empirical progress in the social, behavioral, and economic sciences, including generalizations that demonstrate validity in atypical or nontraditional populations and samples.

National Science Foundation
WHERE DISCOVERIES BEGIN

Home (/) › Document Library (/publications)

NSF 16-137

Dear Colleague Letter: Robust and Reliable Research in the Social, Behavioral, and Economic Sciences

September 20, 2016

Broader Impacts

- First, TrajAnalytics will provide a platform for researchers in transportation assessment and planning to directly study moving populations in urban spaces. Investigators will be released from the burden of big data management and visualization, allowing them to focus on their research on transport systems.
- Second, the software system will be used by visualization and data mining experts so that they can develop specific technologies quickly.
- Third, more researchers in social and behavioral sciences can use the software platform to understand the complicated mechanisms of urban security, economic activities, behavioral trends, etc.
- Fourth, TrajAnalytics acts as an outreach platform which can help government agencies to communicate more effectively to the public, with the real-world facts and dynamics of city traffic, vehicles, and networks. Therefore, urban transportation issues and policies can be better understood and addressed.

Quick Start

- http://vis.cs.kent.edu
- http://vis.cs.kent.edu/guideline/

This work is supported by National Science Foundation ACI-1535031,1535081.

Overview

Trajectory Datasets and Data Types

Data Loading and Processing Guide

TrajAnalytics Account and Web-based Data Access

Visual Analytics Interface

Result Saving and Report

Procedure

Data Loading

TrajAnalytics provides users with an independent data preprocessing software for users to load their own data to TrajAnalytics.

Start Here!

Road Map Matching

TrajAnalytics automatically fetch corresponding road segments data from OpenStreetMap and match the raw GPS data with road segments.

Start Here!

Region Map Matching

TrajAnalytics automatically match the raw GPS data with Zip code regions (USA Only) or Grid Regions.

Start Here!

Visualization System

TrajAnalytics provides a visual analytics of urban trajectory datasets. It allows users to interactively visualize and analyze trajectories over urban spaces.

Start Here!

Videos

http://vis.cs.kent.edu/videos.html