Urban Computing

Dr. Mitra Baratchi

Leiden Institute of Advanced Computer Science - Leiden University

14 February, 2020

Agenda for this session

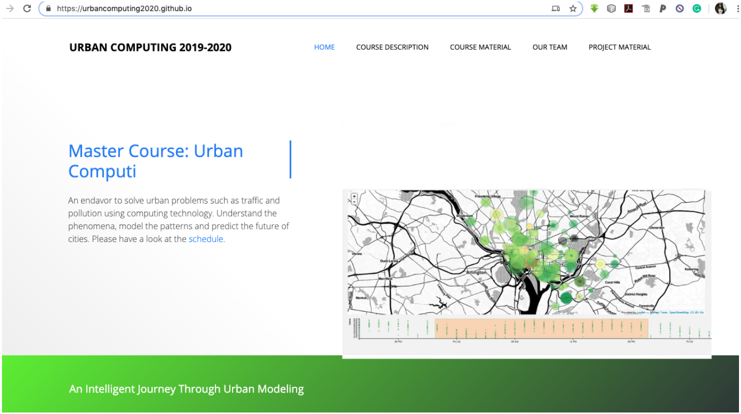
- Part 1: Practical matters
- Part 2: Introduction to the course
  - What is Urban Computing
  - Applications
  - Data sources
- Part 3: Hands-on Lab

Part 1: Practical matters

Teaching assistants

Courseware

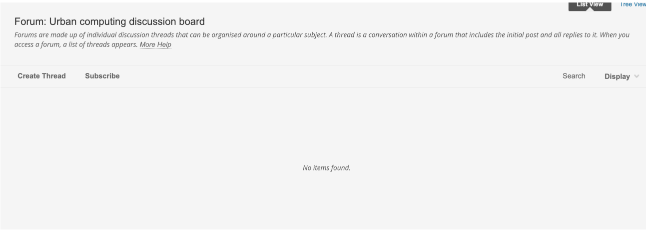
- Courseware access: `https://urbancomputing2020.github.io/`
- Other matters (Announcements, assignment hand-in, discussion forum): Blackboard



Communication

Before sending emails:

- Can you ask the question during the class?
- Can you use blackboard forum?



Course schedule

		Work group session (11:15-13)	Lecture (14:15-16)	16:15-18	Deadlines
	7 Feb 2020	-	No lecture (Dies)	-	-
1	14 Feb 2020	-	**Introduction**	-	
2	21 Feb 2020	-	**Time series data processing, Assignment 1**	-	Deadline selecting a paper
3	28 Feb 2020	Assignment 1	**Spatial data processing**	-	
4	06 Mar 2020	Assignment 1	**Spatio-temporal data processing**	-	Deadline Assignment 1(9 March, 23:59)
5	13 Mar 2020	-	**Machine learning for urban computing, Assignment 2**	-	
-	20 Mar 2020	Assignment 2	**-**	-	
6	27 Mar 2020	Assignment 2	**Machine learning for urban computing 2**	-	Deadline Assignment 2 (30 Match, 23:59)
-	03 Apr 2020	-	No lecture (data science exam)	-	
-	10 Apr 2020	-	No lecture (Good Friday)	-	Deadline proposal (10 April, 23:59)
7	17 Apr 2020	**-**	**Data visualization for urban computing**	-	
-	24 Apr 2020	Meeting with teams (with appointment)	**-**	-	
8	01 May 2020	Meeting with teams (with appointment)	**-**	-	
9	08 May 2020	**Presentation**	**Presentation**	-	
10	15 May 2020	**presentation**	**Presentation**	**Presentation**	Peer review
11	12 June 2020	-	**-**	**-**	Deadline project (15 June 23:59)
12	10 July 2020	-	-	-	Resit project, assignments with penalty (10 July 23:59)

Individual meeting with teams will be scheduled later

Organization of class

Course structure

- Lectures
- Practical sessions for assignments (According to the schedule)
- Practical lab material (self-study material)
- Feedback on projects
- Presentation by students

Grading

- Active participation in class and discussions (10%)
- Assignments (25 %)
- Presentation (15 %)
- Project (50 %)
  - Novelty of the idea
  - Maturity of experiments (considering availability of data sources, or source codes)
  - Results
  - Presentation and documentation

Course rules

- To pass the course both the combined grade and the assignments grade should be over 5.5
- Deadlines are fixed $\rightarrow$ no late submission
- Late submission is considered as resit
- Resit option for assignments and project has maximum grade of 7
- No resit option for presentations
- Submitting project proposal is compulsory (no submission implies deduction of 2 points from the project)

Project

- Select a reference paper as a starting point
  - Bid on the paper bidding list available (select top 5 papers)
  - Alternatively you can bid for papers cited in the Urban computing survey paper [ZCWY14], or previously accepted papers to the annual Urban Computing workshop any year
  - Register your bid (in the form we provide on blackboard)
- Write a brief proposal including (problem statement, research question, methodology, evaluation approach, data sources)
- Proceed with the project
- Write a report (6-8 pages (ACM-proceedings Latex template)) [download]

Part 2: Introduction Urban Computing

What does Urban Computing Mean?


[1]

Figure: Urban Computing

[1] source: http://uctutorial.chinacloudsites.cn

The familiar stranger... [PG04]

The Familiar Stranger: Anxiety, Comfort, and Play in Public Places

Eric Paulos
Intel Research
2150 Shattuck Avenue #1300
Berkeley, CA 94704
paulos@intel-research.net

Elizabeth Goodman
Intel Research
2150 Shattuck Avenue #1300
Berkeley, CA 94704
elizabeth.s.goodman@intel.com

ABSTRACT
As humans we live and interact across a wildly diverse set of physical spaces. We each formulate our own personal meaning of place using a myriad of observable cues such as public-private, large-small, daytime-nighttime, loud-quiet, and crowded-empty. Unsurprisingly, it is the people with which we share such spaces that dominate our perception of place. Sometimes these people are friends, family and colleagues. More often, and particularly in public urban spaces we inhabit, the individuals who affect us are ones that we repeatedly observe and yet do not directly interact with – our *Familiar Strangers*. This paper explores our often ignored yet real relationships with Familiar Strangers. We describe several experiments and studies that lead to a design for a personal, body-worn, wireless device that extends the Familiar Stranger relationship while respecting the delicate, yet important, constraints of our feelings and relationships with strangers in pubic places.

Author Keywords
Strangers, urban space, wireless, wearable, ambient, public place, digital scent, community awareness, ambiguity, *dérive*, *détournement*

ACM Classification Keywords
H.5.3 Group and Organization Interfaces

INTRODUCTION

Figure 1: Familiar Strangers in a typical urban setting

any implications of hostility. A good example is a person that one sees on the subway every morning. If that person fails to appear, we notice.

There are exceptions to the non-interaction rule with Familiar Strangers. The further away from our routine encounter with a Familiar Stranger, the more likely we are to establish direct contact because of a shared knowledge and place. Thus, we are likely to treat our subway Familiar Strangers in San Francisco as close friends if we encounter them in Rome. Similarly, extraordinary events such as an injury, earthquake, *etc.* will also provide the impetus to interact with our Familiar Strangers.

There is a special class of Familiar Strangers called the "socio-metric stars." These are individuals who stand out in a community or group and are readily recognized by an extremely high percentage of people.

A bit of history ....

When was it mentioned first?

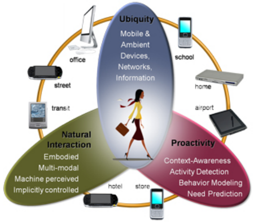
"The term urban computing was first introduced by Eric Paulos at the 2004 UbiComp (Ubiquitous and pervasive computing) conference"

UBICOMP 2018

8-12 October 2018 SINGAPORE

My own story as a member of Pervasive Systems

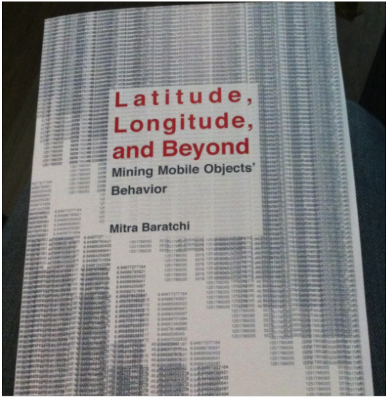
Pervasive and ubiquitous computing (or "ubicomp") is a concept in software engineering and computer science where computing is made to appear anytime and everywhere. In contrast to desktop computing, ubiquitous computing can occur using any device, in any location, and in any format



[2]

[2] source: https://www.parc.com/wp-content/uploads/2010/03/ubicompvenn.jpg

Ubiquitous Computing research with the focus on mobility data



Back to Urban Computing

Urban Computing is the process of acquisition, integration, and analysis of big and heterogeneous data generated by diverse sources in urban spaces, such as sensors, devices, vehicles, buildings, and humans, to tackle the major issues that cities face (e.g., air pollution, increased energy consumption, and traffic congestion). Urban computing connects unobtrusive and ubiquitous sensing technologies, advanced data management and analytic models, and novel visualization methods to create win-win-win solutions that improve urban environment, human life quality, and city operation systems. [ZCWY14]

Mention some urban computing applications ....

Applications


3

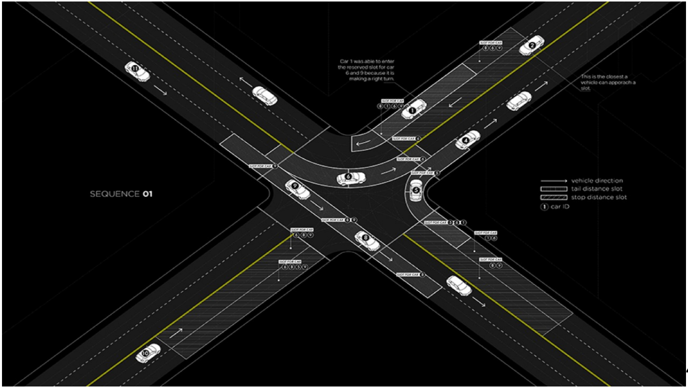
Figure: Traffic management

[3] source: https://www.autoevolution.com/news/the-longest-traffic-jam-in-history-12-days-62-mile-long-47237.html

Applications

Figure: Event management

Applications



[4]

Figure: Autonomous driving

[4] source: http://senseable.mit.edu/light-traffic/

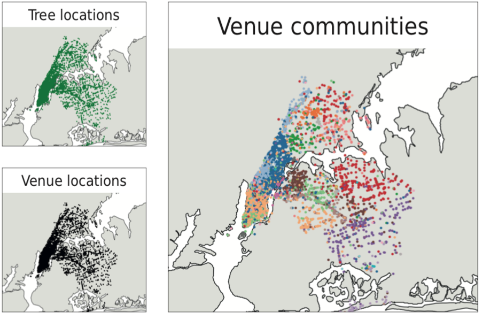


Figure: Urban planning

5

[5] source:[vSTGB]

Rewilding

[6]

Figure: Rewilding

[6] source:https://www.ark.eu/gebieden/buitenland/rewilding-europe

Example of rewilding in the Netherlands

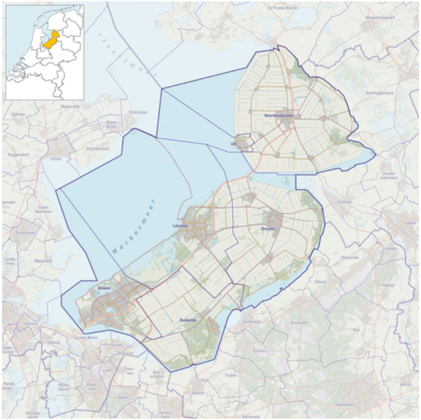


Figure: Rewilding

Main categories of applications in Urban Computing

[7]

---

[7] image source:[ZCWY14]

What do we learn in this course?

What do we learn in this course?

- Things a computer scientist should know when using data to solve urban problems

What we, unfortunately, won't learn in this course:

- Urban planning
- Urban policy making
- Urban ethics
- ...

Topics

- Data sources for Urban Computing research
- Processing time-series data
- Processing spatial data
- Processing spatio-temporal and trajectory data
- Machine learning algorithms for Urban Computing research
- Deep learning for Urban Computing research
- Visualization techniques for Urban Computing research

Why Urban Computing as a new field?

Thinking about urban problems is not new, people have collected data to solve these problems since a long time ago....

Data used for solving urban problems

- Old data sources
- Modern data sources

Old data sources

- Calling people on phones for calculating origin destination matrices (traffic engineering)
- Questionnaires
- Census
- Observations by social scientists

Modern data sources [AB14, AB18]

Modern data sources are categorized into the following three categories based on the origin of the data:

- **Bottom up**: Citizens
- **Intermediate**: Digital companies
- **Top down**: Government

Buttom-up: citizens as sensors

- Data collected through sensing phones (in some manner)

Buttom-up: citizens as sensors

- Data collected through sensing phones (in some manner)
- Data generated as a result of using Apps

Buttom-up: citizens as sensors

- Data collected through sensing phones (in some manner)
- Data generated as a result of using Apps
- Participatory sensing

Buttom-up: citizens as sensors

- Data collected through sensing phones (in some manner)
- Data generated as a result of using Apps
- Participatory sensing: (communities (or other groups of people) contributing sensory information to form a body of knowledge)

Ways to collect data by localizing phones

[8]

Figure: Sensing movement using cellular networks

[8] source: http://unbonmotgroundswell.blogspot.com/2013/07/hybrid-location-technologies-gps.html

Wifi sensing

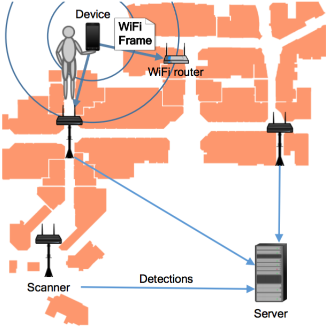


Figure: Sensing movement using Wifi networks [PCB+17]

Wifi sensing, and privacy

[9]

Figure: Wifi sensing

[9] source: https://obj.ca/article/techopia-ottawas-edgewater-wireless-unveils-wi-fi-location-tracking-tech

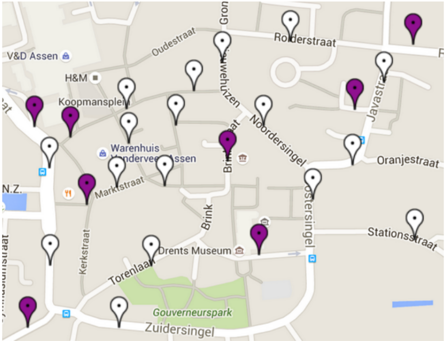


Figure: Assen sensor setup

Intermediate: digital companies

- Free services provided by companies through Internet
- Data generated as a result of the side activity of a digital company
- Companies that aggregate data from local brokers (Funda)

Funda.nl

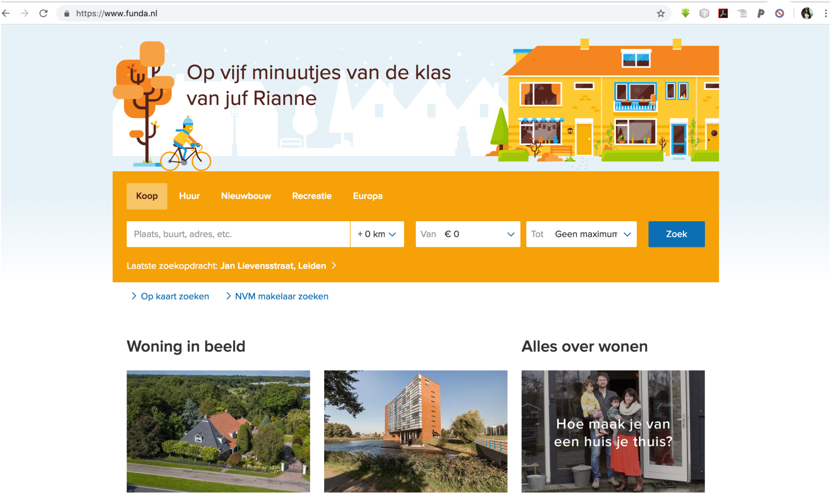


Figure: Funda

Foursquare

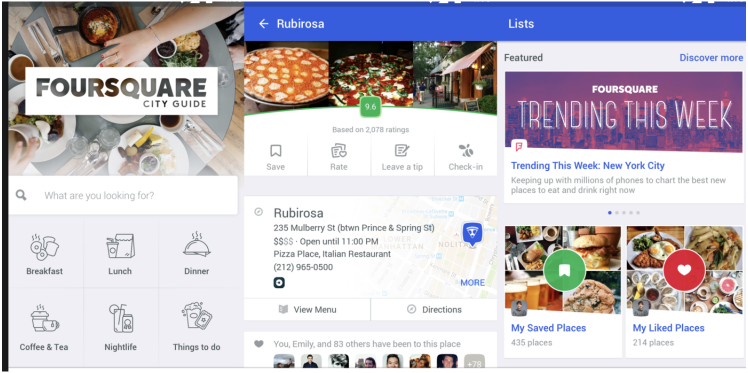


Figure: Foursquare

Top-down: government (Open Data)

Open data is the idea that some data should be freely available to everyone to use and republish as they wish, without restrictions from copyright, patents or other mechanisms of control

Government institutions release (part of) their internal data in open format.

Dutch open data portal

Figure: Dutch open data portal.

A dataset search engine (to be tested)

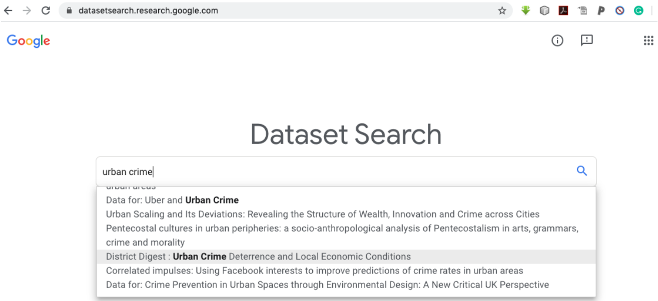


Figure: Google dataset search

Old and modern sources (comparison)

- Where to start from? Collecting data for a specific research question or finding a research question based on data?
- Data collection costs
- Data granularity
- Data quality (noise, error, etc)
- Data sparsity (duration period, missing data)
- Inference quality (one person holding two phones is counted twice,...)

Machine learning/data mining versus statistics

Current approaches to spatio-temporal data handling:

- Statisticians approach
- Machine learning approach

Statisticians approach

Statistics is a branch of mathematics dealing with data collection, organization, analysis, interpretation and presentation [Wikipedia]

- Hypothesis testing
- T-test
- Permutation test
- ...

Machine learning/data mining approach

Machine learning (ML) is the scientific study of algorithms and statistical models that computer systems use to effectively perform a specific task without using explicit instructions, relying on models and inference instead [Wikipedia]

- Design algorithms
- Measure the performance of the algorithm to baselines
  - Classification, clustering, forecasting accuracy
  - Error metrics

Challenges in use of new data sources for machine learning

- Where to get data?
- **How to validate your algorithm?**

Where to get data?

- Collect data by deploying some sensing technology

Where to get data?

- Collect data by deploying some sensing technology (GPS trackers, Wifi scanning, proximity sensing)
- Search for an alternative solution, collect data from a source (Crawl the web, use APIs)

Where to get data?

- Collect data by deploying some sensing technology (GPS trackers, Wifi scanning, proximity sensing)
- Search for an alternative solution, collect data from a source (Crawl the web, use APIs)

How to validate your algorithm?

Case: You are designing an algorithm to find periodic patterns from people's trajectory data

How to validate your algorithm?

Case: You are designing an algorithm to find periodic patterns from people's trajectory data

The recurring issues of ground-truth

Dealing with ground truth issues

- Ask data collectors to label their data (e.g. Lausanne Data Collection Campaign)

Dealing with ground truth issues

- Ask data collectors to label their data (e.g. Lausanne Data Collection Campaign)
- Ask other people to label your data (e.g. Mechanical Turk, Amazon SageMaker Ground Truth)

Dealing with ground truth issues

- Ask data collectors to label their data (e.g. Lausanne Data Collection Campaign)
- Ask other people to label your data (e.g. Mechanical Turk, Amazon SageMaker Ground Truth)
- Validation through defining a higher level machine learning task (e.g. does finding periodic patterns help me predict the future of trajectories better?)

Dealing with ground truth issues

- Ask data collectors to label their data (e.g. Lausanne Data Collection Campaign)
- Ask other people to label your data (e.g. Mechanical Turk, Amazon SageMaker Ground Truth)
- Validation through defining a higher level machine learning task (e.g. does finding periodic patterns help me predict the future of trajectories better?)
- Validation using additional data which is considered highly correlated with the pattern you are looking for (people who work 5 days a week have a strong periodic pattern. Can we distinguish them better from people who do not work?)

Dealing with ground truth issues

- Ask data collectors to label their data (e.g. Lausanne Data Collection Campaign)
- Ask other people to label your data (e.g. Mechanical Turk, Amazon SageMaker Ground Truth)
- Validation through defining a higher level machine learning task (e.g. does finding periodic patterns help me predict the future of trajectories better?)
- Validation using additional data which is considered highly correlated with the pattern you are looking for (people who work 5 days a week have a strong periodic pattern. Can we distinguish them better from people who do not work?)
- Synthetic data generator (e.g. data simulated based on known patterns)

Dealing with ground truth issues

- Ask data collectors to label their data (e.g. Lausanne Data Collection Campaign)
- Ask other people to label your data (e.g. Mechanical Turk, Amazon SageMaker Ground Truth)
- Validation through defining a higher level machine learning task (e.g. does finding periodic patterns help me predict the future of trajectories better?)
- Validation using additional data which is considered highly correlated with the pattern you are looking for (people who work 5 days a week have a strong periodic pattern. Can we distinguish them better from people who do not work?)
- Synthetic data generator (e.g. data simulated based on known patterns)
  - Make synthetic data as close as possible to actual data (add noise, missing data, random patterns,...)
  - Mess with data in all possible ways to make sure your algorithm works all the time

Example generation of synthetic data

5.1 Synthetic Dataset Generation

In order to test the effectiveness of our method under various scenarios, we first use synthetic datasets generated according to a set of parameter. We take the following steps to generate a synthetic test sequence SEQ.

Step 1. We first fix a period T, for example, $T = 24$. The periodic segment SEG is a boolean sequence of length T, with values -1 and 1 indicating negative and positive observations, respectively. For simplicity of presentation, we write $SEG = [s_1 : t_1, s_2 : t_2, \ldots]$ where $[s_i, t_i]$ denote the i-th interval of SEG whose entries are all set to 1.

Step 2. Periodic segment SEG are repeated for TN times to generate the complete observation sequence, denoted as standard sequence SEQ_{std}. SEQ_{std} has length $T \times TN$.

Step 3 (Random sampling η). We sample the standard sequence with sampling rate η. For any element in SEQ_{std}, we set its value to 0 (*i.e.*, unknown) with probability $(1-\eta)$.

Step 4 (Missing segments α). For any segment in standard segment SEQ_{std}, we set all the elements in that segment as 0 (*i.e.*, unknown) with probability $(1 - \alpha)$.

Step 5 (Random noise β). For any remaining observation in SEQ_{std}, we reverse its original values (making -1 as 1 and 1 as -1) with probability β. [10]

[10] [LWH12]

Lessons learned

- Urban Computing field was born due to the adoption ubiquitous and pervasive systems
- Old and new data sources
  - Old $\rightarrow$ questionnaire, census, survey, ...
  - New $\rightarrow$ citizen generated data, open data, digital businesses data, ...
- Urban Computing design computational methods for analysis of new data sources for old urban problems
- Adoption of machine learning methods for new data sources requires careful experiment design and identifying validation approaches (ground truth, labeling, connecting data sources, synthetic data)

End of theory!

Part 3: Hands-on lab

References I

- Daniel Arribas-Bel, *Accidental, open and everywhere: Emerging data sources for the understanding of cities*, Applied Geography **49** (2014), 45 – 53, The New Urban World.
- Dani Arribas-Bel, *Geographic data science'17*, 2018.
- Zhenhui Li, Jingjing Wang, and Jiawei Han, *Mining event periodicity from incomplete observations*, in proc. 18th ACM SIGKDD, ACM, 2012, pp. 444–452.
- Andreea-Cristina Petre, Cristian Chilipirea, Mitra Baratchi, Ciprian Dobre, and Maarten van Steen, *Chapter 14 - wifi tracking of pedestrian behavior*, Smart Sensors Networks, Intelligent Data-Centric Systems, 2017, pp. 309 – 337.
- Eric Paulos and Elizabeth Goodman, *The familiar stranger: anxiety, comfort, and play in public places*, in proc. SIGCHI, ACM, 2004, pp. 223–230.

References II

- JH van Staalduinen, J Tetteroo, D Gawehns, and M Baratchi, *An intelligent tree planning approach using location-based social networks data*.
- Yu Zheng, Licia Capra, Ouri Wolfson, and Hai Yang, *Urban computing: concepts, methodologies, and applications*, ACM TIST **5** (2014), no. 3, 38.