

Preferential Return Behaviors in People's Daily Life: Evidence from semantic enriched geo-located Twitter data

Junjun Yin^{1,+}, Myeonghun Jeong^{1,2,+}, and Zhenhong Du^{2,3,*}

¹Department of Geography and Geographic Information Science

²CyberGIS Center for Advanced Digital and Spatial Studies
University of Illinois at Urbana-Champaign, IL, 61801, USA

³Zhejiang University

*duzhenhong.zju@gmail.com

[†]these authors contributed equally to this work

ABSTRACT

[illegible]

Introduction

Understanding detailed human mobility patterns in the urban settings is critical for a wide range applications, such as urban planning, traffic management, and even the spatial spread of disease.

On the other hand, to get However, the geo-located Twitter Data contains a variety of uncertainties, such spatial sparseness To address these issue, this study implements a novel approach for getting the semantic trajectories by integrating detailed land use maps with geo-located Twitter data.

Emerging as a new source for mobility data, today's pervasive Location Based Social Media (LBSM) platforms (e.g., Twitter and Foursquare) offer continuous spatial Big Data streams with massive amount of detailed and frequently updated user digital traces in the form of real-world trails and footprints[?]. One significant advantage of LBSM data streams is the large spatial coverage, for example, researchers have used geo-located Twitter data for studying global mobility patterns¹, which is otherwise impossible by using other mobility datasets (e.g., GPS traces and mobile phone call records). In addition, the publicly available LBSM data streams offer unique opportunities for conducting reproducible scientific findings regarding the concerns of infringement on individual privacy, such as using mobile phone call records^{?,?,?}.

Related Work

Methods and Materials

Geo-located Twitter Data

In this study, the geo-located tweets were downloaded using the Twitter Streaming API, where we specified a geographical bounding box as an area-of-interest to retrieve all the geo-located tweets that fall within it. To ensure the complete coverage over Great Britain, we have set the bounding box to British Isles (lower left coordinates in (latitude, longitude): (49.497, -14.854) and upper right coordinates: (61.186, 2.637)), which also includes the whole area of Ireland and a part of France. We have implemented a data crawler and continuously collected 7-month data (1st June – 31st December, 2014) with over 101.8 million tweets and 60 GB in size. During the data collection phase, the data crawler did not encounter any issue regarding whether it exceeds the data quota by the 1% policy mentioned in (Hawelka et al., 2014). It means we have managed to download all the geo-located tweets for the given bounding box. In particular, to showcase the overall spatial coverage of the collected geo-located tweets, a density map of the Twitter user locations in British Isles for July 2014 is shown in Figure 4, where the

collected points alone visually reveal the geography of cities and countries, e.g. the clusters with high density of tweets (in color red) correspond to the skeleton of major cities.

The shape of human mobility

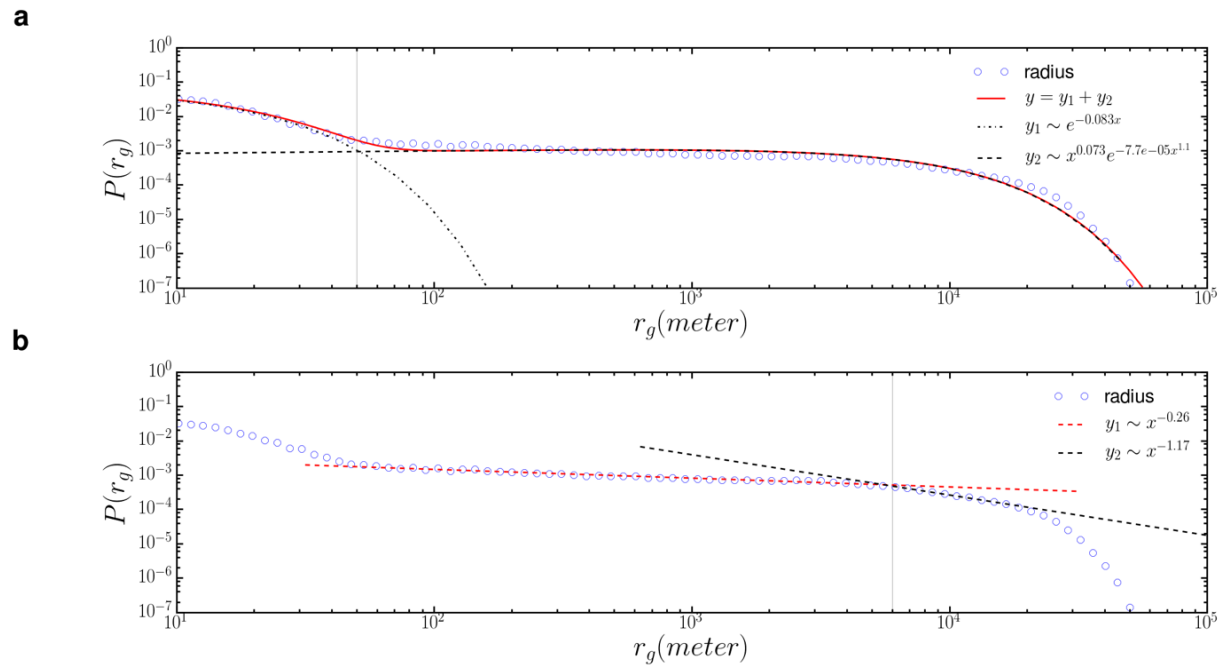


Figure 1. The distribution of the radius of gyrations for Twitter users in Chicago

Results

Discussion and Conclusion

Author contributions statement

J.Y. conceived the experiment(s), J.Y. and M.J. conducted the experiment(s), J.Y., M.J. and D.Z. analysed the results. All authors reviewed the manuscript.

References

1. Hawelka, B. *et al.* Geo-located twitter as proxy for global mobility patterns. *Cartography and Geographic Information Science* **41**, 260–271 (2014).

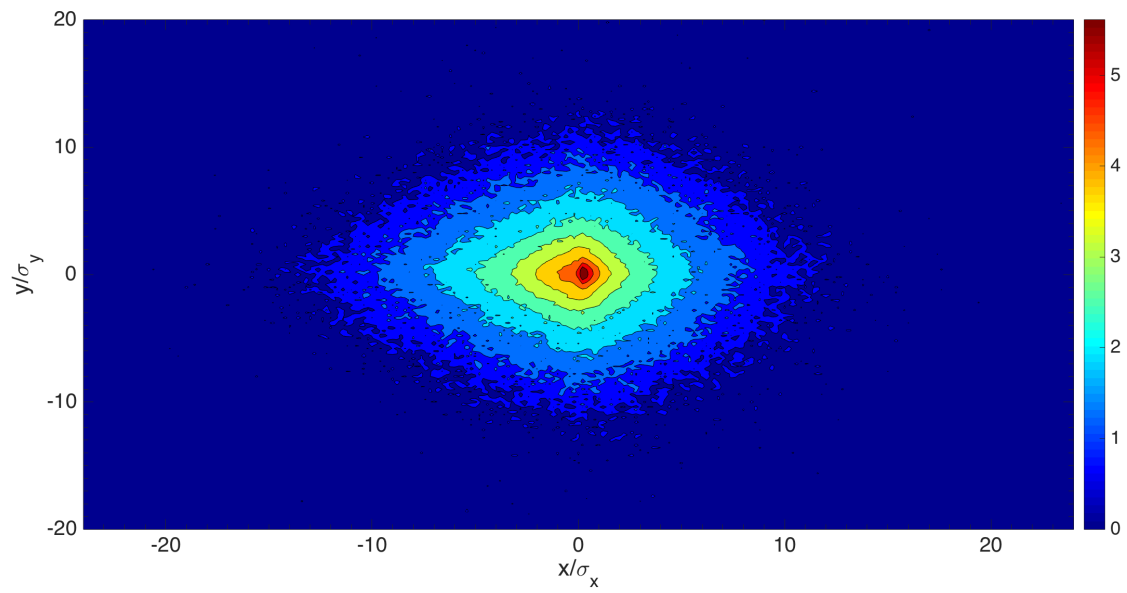


Figure 2. The shape of Twitter user mobility in Chicago for year 2014

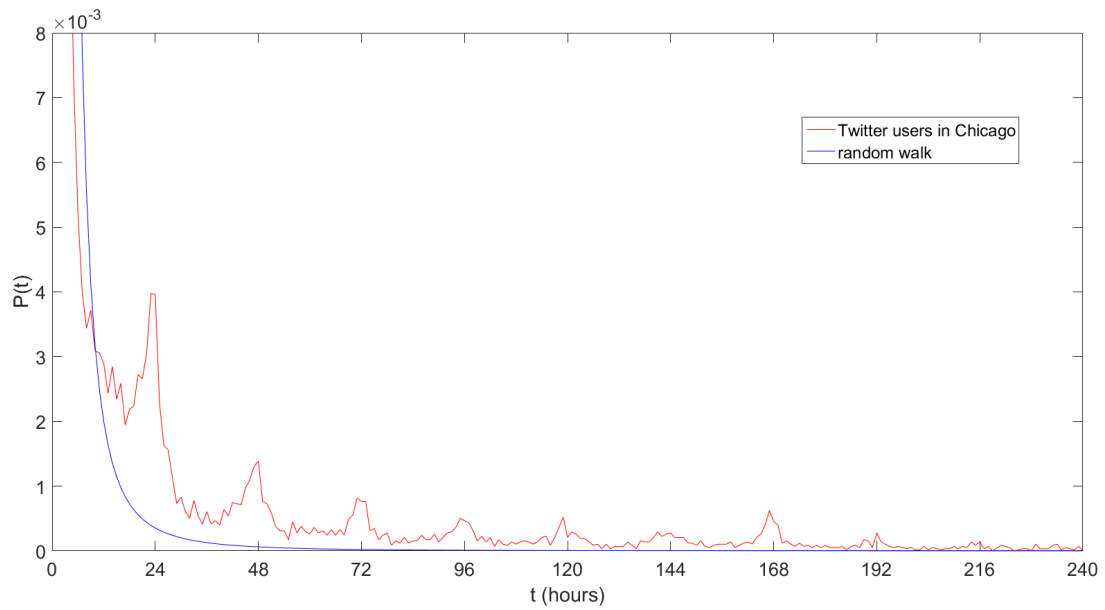


Figure 3. The shape of Twitter user mobility in Chicago for year 2014