Using fast and slow data to unfold city dynamics

James Bagrow
Mathematics & Statistics, University of Vermont, USA
James.bagrow@uvm.edu

Super title: Urban dynamics

Standfirst: Real-time mobility data capturing city-wide human movement characterizes cities, their segregation, and population responses to exogenous events such as pandemics.

To understand the dynamics of a city, we need to understand how people inhabiting the city use it, moving through it day-to-day, and how those movements may change due to external factors. Historically, data on human use of cities has been slow and costly to gather, relying especially on census tracts and travel surveys. Most recently, smartphones and wearable devices have been able to efficiently capture mobility patterns across large numbers of individuals in nearly real-time. Such rich data can describe how individuals move about a city, and with real-time information, we can also understand how a city responds to a sudden crisis, such as an earthquake or a disease outbreak [1,2]. It is worth noting, however, that traditional ('slow') data is still useful: as a matter of fact, the combination of both slow and cutting-edge ('fast') data can be very informative, as demonstrated by Yanyan Xu and colleagues, who – writing in Nature Computational Science [3] – used a wide array of both types of data to shed light on the city dynamics of various urban populations.

Examining cities worldwide, Xu et al. showed a broadly consistent relationship between the population's distribution and their mobility patterns, allowing them to categorize cities in a two-dimensional space, with one dimension describing (with slow data) how dispersed or compact the city population is outside the city center and a second dimension describing (with fast data) the amount of mobility centers that the population uses, meaning, whether the city is 'monocentric' or 'polycentric.' Some cities that appear similar in one dimension are nevertheless disparate in the other. For instance, the San Francisco Bay Area and Shenzhen, China are both strongly polycentric, but Shenzhen's population is far more compact than San Francisco, as noted by the authors. A disease outbreak in one city may proceed quite differently than in the other.

Not everyone experiences the same view of their city, which was another dimension explored by the authors. Urban use patterns are known to reflect systemic racial and economic disparities, and thus, we can anticipate mobility statistics to separate different groups. Using sociodemographic data for a subset of the cities under study, the authors observed, for instance, that in the cities of Boston and Los Angeles in the United States, and Bogotá, Columbia all display racial and income segregation – but the form of segregation varies. In Boston and Los Angeles, lower-income inhabitants generally reside near the city center, while higher-income inhabitants reside in the periphery. In Bogotá, on the other hand, the lower-income residents predominantly live in the periphery. Xu et al. also observed that members of the highest-income groups in Boston and Los Angeles tend to travel the longest distances of any group, whereas in Bogotá, the trend is reversed, with members of the lowest-income group being the widest travelers.

Mobility data allowed the authors to probe these disparities further, asking how the travel patterns differ between diverse racial and economic groups by identifying locations where individuals spend significant portions of time and the typical lengths of their travels. Looking at significant stay locations during the day and the night, the authors revealed that cities are substantially less segregated during the day. Likewise, cross-referencing these data with a subset of individuals who completed travel use surveys allowed the authors to classify work and non-work travel activities. They found that it was non-work activities that showed substantial segregation, leading the authors to conclude that mobility choices during leisure pursuits can explain group segregation. However, whether this association is causal or confounded remains to be investigated.

These datasets overlap with the COVID-19 outbreak, which allowed the authors to study how travel restrictions affected the mobility choices of different groups. Drops in long-distance travel differed for groups of socioeconomic status: travel distances for higher-income individuals dropped more significantly than for lower-income individuals, both during the initial outbreak period of March 2020 and during the “peak season” of October–December 2020. The disparity is probably due to economic opportunities: travel by members of lower-income groups is often more essential, and they are less likely to be able to work from home.

The data used by these and other studies provide important insights, but are also challenging to work with. Typically, it relies on combining the data with statistical or mechanistic models. The LandScan population data used by the authors (https://landscan.ornl.gov/), for example, uses remote imaging data to power a statistical model of the population. Likewise, some mobile phone data used for individual mobility traces are noisy, with many gaps between recorded locations. Xu et al. used the TimeGeo mechanistic model [4] to help fill these gaps. Better assimilating these data and models is an exciting opportunity to further computational research.

At the same time, coupled with the exciting computational and data challenges are critical implications and concerns. Privacy, in particular, is a pressing concern. Many of these mobility data are disaggregated, meaning that individual movement patterns can be identified. While such data are de-identified, research has shown ways to reveal private information from such sensitive data [5,6]. Smartphone makers are now moving away from individual-level data towards aggregated data only. However, doing so effectively without comprising the data’s validity will require further research using tools such as differential privacy [7].

Competing Interests

The author declares no competing interests.

References

1. JP Bagrow, D Wang, and A-L Barabási. *PLOS ONE* 6, no. 3 (2011): e17680.
2. X Lu, L Bengtsson, and P Holme. *Proceedings of the National Academy of Sciences* 109, no. 29 (2012): 11576-11581.
3. Xu et al. *Nature Computational Science* TK TK (2023)
4. Jiang et al. *Proceedings of the National Academy of Sciences* 113, no. 37 (2016): E5370-E5378.
5. R Pellungrini, L Pappalardo, F Pratesi, and A Monreale. *ACM Transactions on Intelligent Systems and Technology* 9, no. 3 (2017): 1-27.
6. Chen et al. *Nature Communications* 13, no. 1 (2022): 1922.
7. B Niu, Q Li, X Zhu, G Cao, and H Li. In *IEEE INFOCOM 2014-IEEE conference on computer communications*, pp. 754-762. IEEE, 2014.