9^{th} International Conference on Computational Social Science IC^2S^2
July 17-20, 2023, Copenhagen, Denmark

Contrasting social and non-social sources of predictability in human mobility

Keywords: Human mobility, information theory, social networks, privacy, predictability

Extended Abstract

The recent availability of extensive geolocated datasets related to human movement, has enabled the quantitative study of human movement at an unprecedented level [1], contributing greatly to insights in estimating migratory flows, traffic forecasting, urban planning, mitigating pollution, uncovering socioeconomic inequalities, and epidemic modeling among other applications [2, 3, 4, 5]. An important finding in this context, is the ability to predict the future locations of individuals or groups using their prior history of travel. Indeed, it has been shown that a perfect algorithm can predict, with between 70-90% certainty, an individual's future location given their prior location visits [6], depending upon the spatiotemporal granularity of observations [7].

Human beings are typically highly social creatures and social structures can influence behavior in a variety of human activities including movement patterns. In fact, it has been shown that social relationships statistically account for between 10% to 30% of all human movement [8]. Social structures inherently encode information flow between parties, such that residual information about an individual can be inferred from their social ties. Coupled with the observation that movement patterns in the virtual and physical domains are strikingly similar [9], this leads to the intriguing question as to whether one can leverage a person's social network to predict their future mobility patterns, absent any information on their own history.

Here we apply non-parametric information-theoretic estimators to study human mobility extracted from three Location-Based Social Networks (LBSNs), that contain sequences of location trajectories and the (reported) social network of a subset of users. In addition, we also analyze CDRs from Rio de Janeiro in Brazil. Each type of dataset has its own limitations: in the case of LBSNs, the mobility patterns being biased by the types of individuals using the platform; in the case of CDRs, the lower-spatial resolution. Despite this, we demonstrate the existence of information transfer in all four networks, finding that a given ego's future location visits can be predicted, with between 80–100% of the ego's own accuracy (Fig. 1), by studying the historical patterns of just 10 of their alters (ranked by number of common locations visited). Remarkably, colocators who are not part of an individual's social network, while individually providing less information than social ties, can in the aggregate provide similar levels of predictability (Fig. 1F). The information flow provided by colocators is also surprisingly robust to temporal-displaced colocations, implying users that never physically colocate can still provide comparable information to social ties. Indeed, the information transfer appears to be driven by the overlap of unique locations visited by the ego and alters, in both social ties and non-social colocators.

The presence of predictive information, both socially and otherwise, has crucial implications. Privacy protections regarding social data are important to protect sensitive information about a user and their social ties. Social information flow suggests that an individual's future movements can be predicted by studying the mobility patterns of a few acquaintances. On the

other hand, our study also demonstrates that social ties are not the only source of predictive mobility information, and measures of colocation are enough to uncover novel sources of mobility information. This means that locations monitoring individual visits, for example, a grocery store tracking the smartphones of shoppers [10], may in principle be collecting the building blocks of mobility profiles, and individuals providing access to their mobility data may also be providing information about both social and non-social ties [11].

While these data can inform important applications such as contact tracing in the early stages of a disease outbreak, significant ethical concerns surrounding such information sources make it critical to place strong access constraints on mobility information. Indeed, the results presented here provide further impetus to the ongoing debate on best practices for privacy protection, both in terms of legislation and ethical algorithmic development.

References

[1] Barbosa, H. *et al.* Human mobility: Models and applications. *Physics Reports* **734**, 1–74 (2018).

[2] Batty, M. *The new science of cities* (MIT press, 2013).

[3] Simini, F., Gonzalez, M. C., Maritan, A. & Barabási, A.-L. A universal model for mobility and migration patterns. *Nature* **484**, 96–100 (2012).

[4] Pan, W., Ghoshal, G., Krumme, C., Cebrian, M. & Pentland, A. S. Urban characteristics attributable to density-driven tie formation. *Nature Communications* **4**, 1961 (2013).

[5] Toole, J. L. *et al.* The path most traveled: Travel demand estimation using big data resources. *Transportation Research Part C: Emerging Technologies* **58, Part B**, 162–177 (2015).

[6] Song, C., Qu, Z., Blumm, N. & Barabasi, A.-L. Limits of Predictability in Human Mobility. *Science* **327**, 1018–1021 (2010).

[7] Ikanovic, E. L. & Mollgaard, A. An alternative approach to the limits of predictability in human mobility. *EPJ Data Science* **6**, 12 (2017).

[8] Cho, E., Myers, S. A. & Leskovec, J. Friendship and mobility: User movement in location-based social networks. In *Proceedings of International Conference on Knowledge Discovery and Data Mining*, KDD '11, 1082–1090 (Association for Computing Machinery, New York, NY, USA, 2011).

[9] Hazarie, S., Barbosa, H., Frank, A., Menezes, R. & Ghoshal, G. Uncovering the differences and similarities between physical and virtual mobility. *Journal of the Royal Society Interface* **17**, 20200250 (2020).

[10] Enck, W. *et al.* Taintdroid: an information-flow tracking system for realtime privacy monitoring on smartphones. *ACM Transactions on Computer Systems (TOCS)* **32**, 1–29 (2014).

[11] Garcia, D. Leaking privacy and shadow profiles in online social networks. *Science advances* **3**, e1701172 (2017).

...

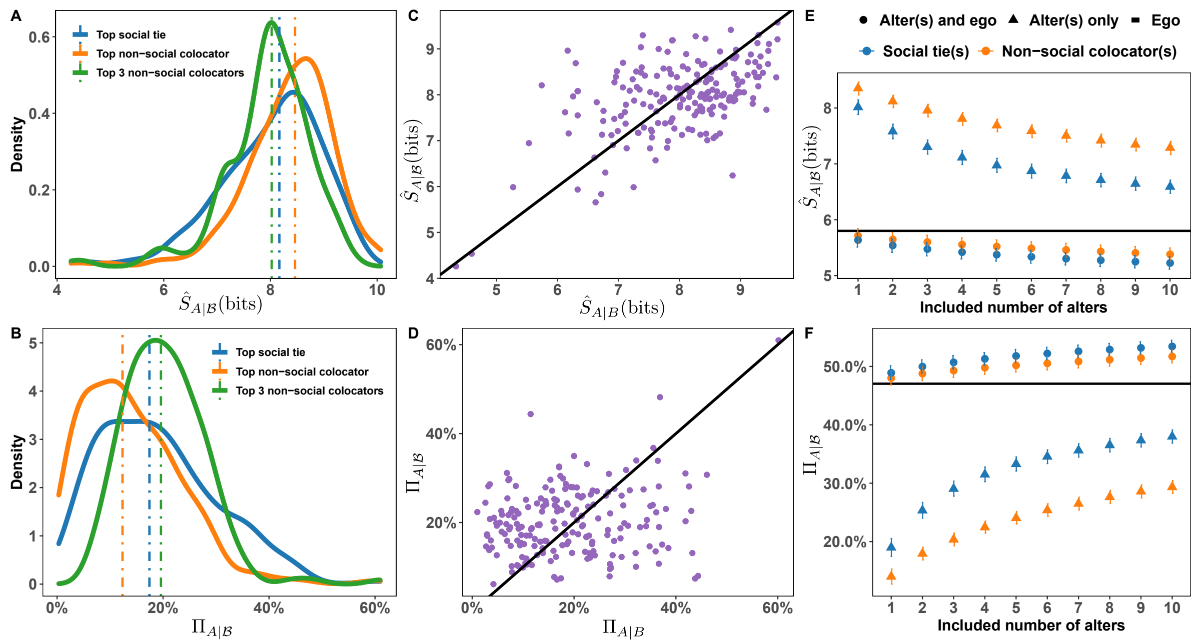

Figure 1: **Cross-entropy and predictability in social ties and non-social colocators**. **A** Distributions of the cross-entropy $\hat{S}_{A|B}$ for the rank-1 social tie (median 8.17 bits), non-social colocator (median 8.46 bits), and cumulative cross-predictability $\hat{S}_{A|\mathscr{B}}$ for the top-3 non-social colocators (median 8.02 bits) in Weeplaces. **B** The corresponding cross-predictability $\Pi_{A|B}$ for the social (median17.43%), and non-social colocators (median 12.35%), and cumulative cross-predictability $\Pi_{A|\mathscr{B}}$ for the top-3 non-social colocators (median 19.60%). **C** $\hat{S}_{A|B}$ encoded in the top-social tie as a function of $\hat{S}_{A|\mathscr{B}}$ for the top-3 non-social colocators. Each point corresponds to a single ego and the solid line denotes $y = x$. **D** As in panel **C** but with predictability instead of cross-entropy. **E, F** $\hat{S}_{A|\mathscr{B}}$ and $\Pi_{A|\mathscr{B}}$ after accumulating the top-ten social alters and non-social colocators. Horizontal lines denote the average entropy (5.80 bits) of egos and their self-predictability (47.05%). Shapes indicate whether the past trajectory of the ego was included in the sequence (circles) or excluded (triangles). Error bars denote 95% CI.