ROYAL SOCIETY OPEN SCIENCE

royalsocietypublishing.org/journal/rsos

Research

Cite this article: Poudyal B, Pacheco D, Oliveira M, Chen Z, Barbosa HS, Menezes R, Ghoshal G. 2024 Dynamic predictability and activity-location contexts in human mobility. *R. Soc. Open Sci.* **11**: 240115.
https://doi.org/10.1098/rsos.240115

Received: 18 January 2024
Accepted: 22 July 2024

Subject Category:
Computer science and artificial intelligence

Subject Areas:
complexity, algorithmic information theory

Keywords:
complex systems, information theory, human mobility

Authors for correspondence:
Ronaldo Menezes
e-mail: r.menezes@exeter.ac.uk
Gourab Ghoshal
e-mail: gghoshal@pas.rochester.edu

Electronic supplementary material is available online at https://doi.org/10.6084/m9.figshare.c.7410574.

THE ROYAL SOCIETY PUBLISHING

Dynamic predictability and activity-location contexts in human mobility

Bibandhan Poudyal[1], Diogo Pacheco[2], Marcos Oliveira[2,3], Zexun Chen[4], Hugo S. Barbosa[2], Ronaldo Menezes[2,5] and Gourab Ghoshal[1]

[1]Department of Physics & Astronomy, University of Rochester, Rochester, NY, USA
[2]Department of Computer Science, University of Exeter, Exeter, UK
[3]GESIS—Leibniz Institute for the Social Sciences, Cologne, Germany
[4]The University of Edinburgh Business School, Edinburgh, UK
[5]Department of Computer Science, Federal University of Ceará, Fortaleza, Brazil

BP, 0000-0001-6437-7251; GG, 0000-0001-7593-5626

Human travelling behaviours are markedly regular, to a large extent predictable, and mostly driven by biological necessities and social constructs. Not surprisingly, such predictability is influenced by an array of factors ranging in scale from individual preferences and choices, through social groups and households, all the way to the global scale, such as mobility restrictions in response to external shocks such as pandemics. In this work, we explore how temporal, activity and location variations in individual-level mobility—referred to as *predictability states*—carry a large degree of information regarding the nature of mobility regularities at the population level. Our findings indicate the existence of contextual and activity signatures in predictability states, suggesting the potential for a more nuanced approach to estimating both short-term and higher-order mobility predictions. The existence of location contexts, in particular, serves as a parsimonious estimator for predictability patterns even in the case of low resolution and missing data.

1. Introduction

The understanding of the mechanisms governing human travel behaviour is crucial to a variety of domains such as epidemic modelling [1–5], transportation [6–9], national security [10], urban planning [11–18] and a host of other applications [19]. Human mobility trajectories have been shown to exhibit statistical regularities at multiple scales [20,21], despite the inherent complexity that exists in the available choices for the

routes of their daily travels [22,23]. These regularities are rooted in an array of social, spatial and temporal mechanisms, leading to visitation patterns being highly regular [24]. Indeed, it has been shown that a perfect algorithm can predict, with between 70% and 90% certainty, an individual's future location given their prior location visits [25], depending upon the spatio-temporal granularity of observations [26,27]. In practice, however, the predictive task has proven to be more challenging than what the high predictability could suggest. Current predictive models have been achieving performances ranging from 10% to 15% on similar mobility datasets [28–30].

Moreover, human beings tend to be routine-oriented. For instance, lack of regularity in daily mobility is linked to high levels of stress [31,32]. This change-averse behaviour leads people to favouring well-defined routines, which, in combination with stationarity [26], makes mobility trajectories quite regular [33]. Several factors such as work schedules and physiological processes, influence mobility-related decisions; for instance, daily necessities such as sleeping and eating influence activity schedules [34–39]. Conversely, disruptions to such routines can completely alter the mobility trajectories (as in the recent COVID-19 lockdown measures [40]), which can significantly alter mobility trajectories and therefore the associated levels of predictability [41].

Missing from extant measures of predictability are spatio-temporal constraints and the social embedding behind mobility regularities [42]—guessing that a person will be at home on Tuesday at 4.00 will most likely be a correct prediction. However, it may be much harder to know a person's whereabouts during times in which they might not be bound by typical daily rhythms, on weekends, for example [43]. Additionally, predictability is computed from an asymptotic approximation of entropy based on a non-parametric estimator [44,45] that does not account for the periodic and rhythmic nature of travelling behaviours [46,47]. Finally, the metric does not provide insight into the generative mechanisms that underpin the observed regularities in mobility [19].

To overcome these limitations, we conduct studies on three location-based social networks (LBSNs) that contain location trajectories for all the users that participate in such networks. By location trajectories, we mean visitations made by individuals throughout the history of the data, represented as a vector whose elements correspond to a temporal sequence of alphabets with each alphabet corresponding to a unique labelling for the given location. Our work proposes a new perspective to mobility predictability that accounts for these missing factors, particularly connecting the observed regularities in mobility patterns to human circadian/semi-circadian routines and the types of locations they visit. Furthermore, we study the temporal variations of predictability, unveiling structural patterns in their frequency and time components, which we refer to as *predictability states*. Our results suggest that in addition to the daily routines, regularities in mobility predictability are also marked by periods of approximately 12 and 6 h, which could correspond to the second and fourth harmonics of internal circadian rhythm. These findings could suggest that factors governing mobility-related decisions are also influenced by internal biological cycles beyond sleeping and feeding needs. This is corroborated by the predominance of 12 h periods over other cycles governed by sleep/work/study routines, such as the 8 and 4 h cycles. Additionally, we show the role of a location-based context, demonstrating heterogeneities in mobility profiles based on the types of locations that people visit. The importance of this is demonstrated through agreement between the true distribution—calculated from the full set of visitation trajectories—and a simple linear model that takes only the frequency of visiting a location type as input. Taken together our results indicate that uncertainty and predictability in mobility patterns should be considered as a transient *state* of the individuals that is heavily influenced by their temporal and activity-location contexts.

2. Results

2.1. Datasets

We use data from three different LBSN services. Brightkite and Gowalla were two popular social networking sites that existed from 2007 until 2012. Weeplaces was a website where users could upload their check-in activities from other social network services (e.g. Facebook Places, Foursquare). These datasets contain users' check-in activity including user identification, location coordinates (i.e. latitude and longitude) and the time-stamp of the logged activity. Additionally, the Weeplaces dataset contains a description (i.e. the category) of the locations (e.g. nightlife, outdoors). The details of the data are listed in table 1. For each individual in the dataset, we convert their check-in activity to trajectories described as a time series of the form

Table 1. Data from location-based social networking sites.

dataset	users	records	period
Brightkite [48]	58 228	4 491 143	April 2008–October 2010
Gowalla [49]	107 002	6 405 492	February 2009–October 2010
Weeplaces [50]	15 799	7 658 368	November 2003–June 2011

$$X = \{x(1), x(2), \ldots, x(T)\},$$

where $x(t) \in \mathcal{V}$ is a location and $\mathcal{V}$ is the set of all visited locations by that individual. For more details on the data, refer to electronic supplementary materials, §S1, table S1 and figures S1– S6.

To compute the time-dependent predictability, we segment each user's trajectory into non-overlapping windows of a specified size (1 h, for instance). The window size can be adjusted based on the research question. For each window (e.g. Monday at 1.00–2.00), we calculate the predictability based on the user's location sequence within that window. This process is repeated for all windows across all users. Finally, we average the predictability scores across users for each specific time window (e.g. all users on Monday at 1.00) to obtain the population-level predictability for that time frame.

To analyse the effect of temporal window of observations, we segment each user's entire trajectory for a defined period (e.g. a week with 168 h) into non-overlapping windows of the chosen size. This creates multiple 'bins' per user—168 bins for a 1 h window, 84 bins for 2 h windows and so on. We calculate a predictability score for each individual bin within a user's trajectory. Averaging the individual bin scores for each user, results in a single predictability score that reflects their overall predictability across the chosen window size. For more details of the procedure, cf. electronic supplementary materials, §S1.3 and figure S7.

2.2. Time-independent uncertainty and predictability

We begin our analysis by examining the information contained in the location trajectories of all individuals in each of the datasets. When accounting for only the frequency of location visitations, the degree of uncertainty in capturing the future locations of a trajectory given past observations is encoded in the Shannon entropy (measured in bits)

$$H_u = -\sum_{i \in \mathcal{V}} p(x)\log_2 p(x), \tag{2.1}$$

where $p(x)$ is the probability to visit the location x. The subscript refers to the fact that this is the uncorrelated entropy, given that no information on the sequence of location visits is considered. Accounting for both the visitation frequency and the temporal sequence of location visits, we use a non-parametric estimator [44,45] termed the entropy rate (Lempel–Ziv algorithm), given by the expression

$$H_c = \frac{N\log_2 N}{\sum_{i=1}^{N} \Lambda_i}, \tag{2.2}$$

where N is the number of moves made by an individual (not necessarily to distinct locations), Λ_i is the length of the shortest trajectory sub-sequence beginning at position i not seen previously. In the absence of any structure in the sequence, equation (2.2) reduces to equation (2.1). In figure 1*a*, we plot the results for the datasets finding that in all cases $H_u \approx 6$ bits and $H_c \approx 5$ bits. This indicates that accounting for the temporal sequence reduces the possible space of future location visits from 2^6 to 2^5 possible locations indicating the presence of temporal correlations in visits between locations.

The entropy rate can be converted to a measure of predictability Π using Fano's inequality [51] to define the upper bound of how often an ideal predictive algorithm can correctly guess the next location visit, given prior history. This is calculated by inverting

$$H_{u,c} \leq B(\Pi_{u,c}) + (1 - \Pi_{u,c})\log_2(S-1), \tag{2.3}$$

where S is the number of *distinct locations visited* and $B(x)$ is the binary entropy function capturing the entropy of a simple Bernoulli trial (in this case achieving maximal predictability or not). The metric

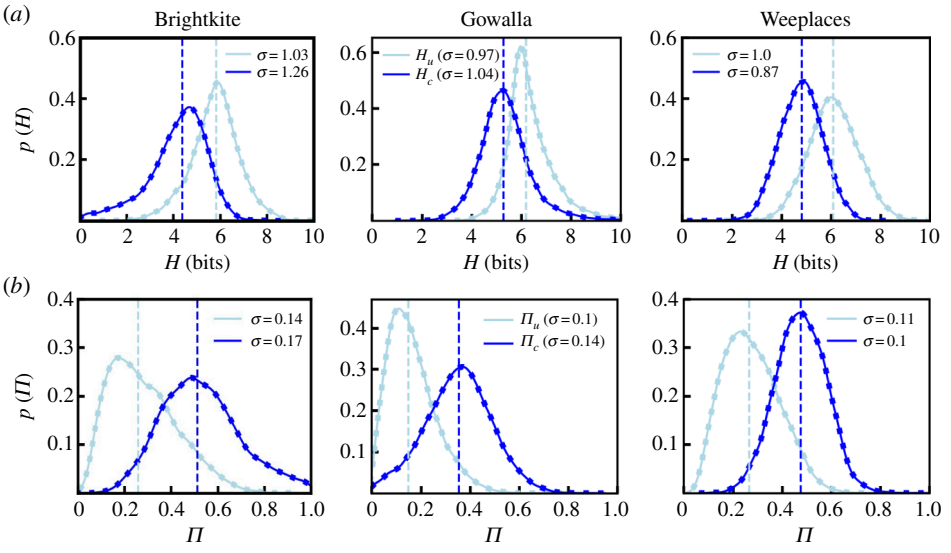

Figure 1. Time-independent entropy and predictability. (*a*) The entropy accounting for only visitation frequencies H_u and incorporating the temporal sequence of visitations H_c. (*b*) The corresponding predictability values were calculated through equation (2.3). In all cases, accounting for the sequence reduces uncertainty from 6 to 5 bits and increases predictability from 20% to 40%. The vertical dashed lines represent the medians for each distribution while the values for the standard deviation σ are shown in the legends.

mathematically bounds the performance of all real predictive methods given an information sources inferred uncertainty. The corresponding results are shown in figure 1*b* and mirror the trends seen for the entropy; including information about the sequence of location visits, over and above the frequency of visiting those locations, increases the predictability from approximately 20% to 40%. In what is to follow we will interpolate between Π_u and Π_c based on the nature of the analysis.

2.3. Time-dependent predictability

Human activity routines are characterized by temporal regularities with time and frequency components. For instance, usual working hours tend to recur every 24 h (i.e. frequency) with changes during the weekends (i.e. time). This temporal regularity is reflected in the user check-in activity patterns observed in electronic supplementary material, figure S4. Thus, it is reasonable to believe that mobility regularities in both time and frequency domains should manifest themselves in the predictability profile of an individual. Thus, in order to extract the temporal variation of the predictability, instead of considering the complete visitation sequence of individuals, we analyse individuals at different moments of the week. Specifically, we split the trajectory of each individual into time slots representing the time of the week. We create 168 slots (i.e. 24 h × 7 days of the week) and define $X_{t=t_0}$ as a random variable representing the places that an individual visits at the time slot $t = t_0 \in [1, \ldots, 168]$.

Given the shorter trajectories in a single-hour window, the temporal correlations are less important, and, consequently, we plot the uncorrelated predictability Π_u. The results are shown in figure 2*a*, where we plot the time-series of $\Pi_u(t)$ for Weeplaces—the corresponding plots for Brightkite and Gowalla are shown in electronic supplementary materials, figures S8 and S9. As the figure indicates, when disaggregating with respect to time, the predictability exhibits considerable variation both from day to day as well as within a given day. For instance, we see a progressive decrease in predictability from Monday to Friday, which then picks up again and peaks on the weekend. Within a day there is considerable variation, with the predictability of an individual fluctuating between 50% and 85%, with individuals being more predictable at night (the standard deviation is shown as the shaded region in figure 2*a* and is consistent with the range shown in figure 1*b*). Finally, we see a clear 24 h periodicity in the predictability profile with high predictability periods during evenings and at night (peaking around 4.00 to 5.00) and troughs at midday. This feature is seen across all three datasets.

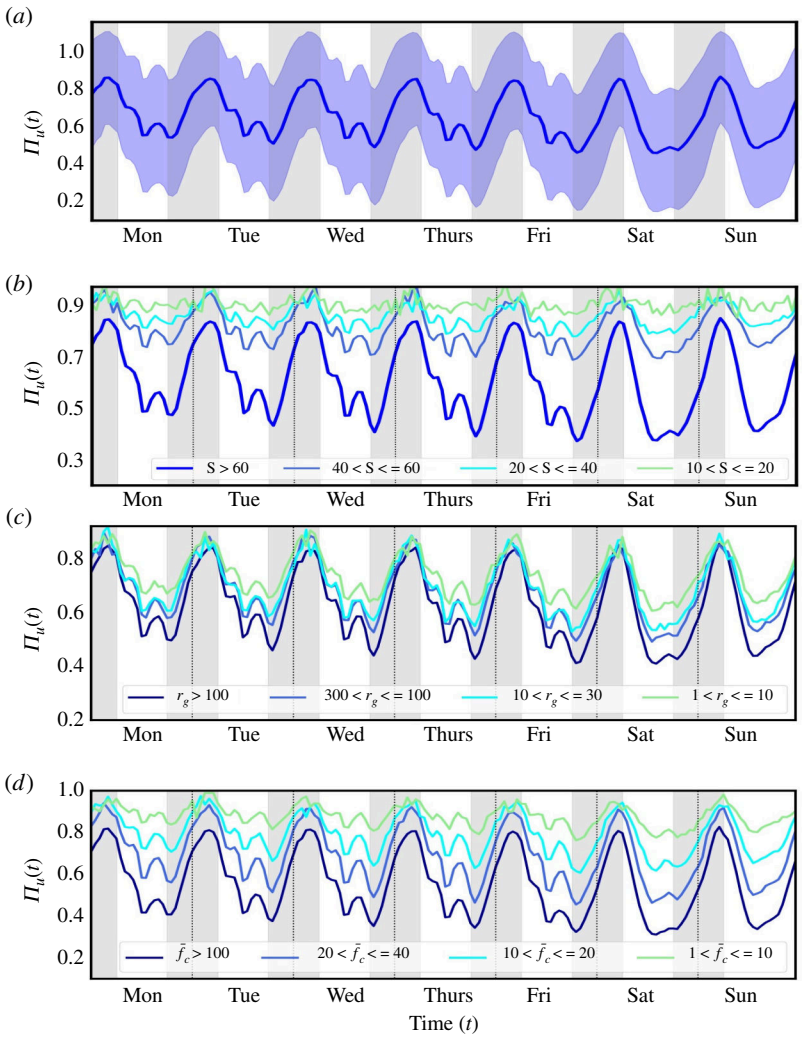

Figure 2. Time-dependent predictabilities. (*a*) The uncorrelated predictability Π_u, disaggregated with respect to time for Weeplaces. The trajectory of each individual is split into time slots representing the time of the week. We create 168 slots (i.e. 24 h × 7 days of the week) and define $X_{t=t_0}$ as a random variable representing the places that an individual visits at the time slot $t = t_0 \in [1, \ldots, 168]$. The shaded area depicts the standard deviation ($\pm\sigma$) around the mean value, which is calculated for each hour. (*b*) The temporal predictability disaggregated with respect to the number of unique locations visited S. (*c*) Now disaggregated with respect to geographical coverage as measured by the radius of gyration r_g in units of kilometres. (*d*) Finally, disaggregated with respect to the average frequency of monthly check-ins $\bar{f}_c$. Across all users, we see daily peaks (4.00 to 5.00) and secondary peaks (12.00 to 17.00) of predictability throughout the time series.

We note that as in other human facets, human mobility and social interactions tend to exhibit a high degree of heterogeneity. If on one hand, most people use their mobile phones a few times a day, some users, on the other hand, make hundreds of calls a day. Some individuals are more active in the sense of travelling more frequently, to a diversity of locations, as well as travelling longer distances. One would expect this activity context to influence the predictability profiles in different ways. To test for this effect we disaggregate the data in terms of three different metrics of activity: number of unique locations visited S; geographical coverage measured by the *radius of gyration* r_g [20]; and monthly recurrent frequency by averaging the number of check-ins per month $\bar{f}_c$. The electronic supplementary material, figure S1 shows the distribution of these quantities for all the datasets, indicating a right-skewed heavy-tailed distribution for all three measures in line with previous observations [20,25].

The influence of the observed heterogeneity in the activity metrics on the predictability is plotted in figure 2*b–d*. First, we note that the temporal trends are maintained even while disaggregating with respect to the activity metrics. However, the range in predictability varies with respect to the levels of activity. For instance, those who visit between 10 and 20 unique locations have an effectively flat

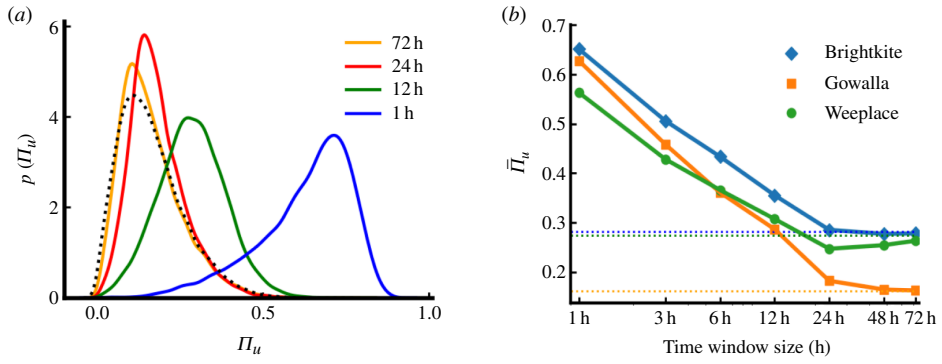

Figure 3. The effect of temporal windows of observation. (*a*) The distribution of Π_u as a function of window size for Gowalla. The dashed curve represents the baseline distribution of Π_u when taking into account the full trajectory of individuals. (*b*) The modes of the distributions are plotted as a function of window size for all three datasets. Horizontal dashed lines indicate the saturated value of Π_u.

temporal profile across the days of the week, as well as within a day ($\Pi_u \approx 90\%$). In contrast, those who visit more than 60 unique locations show much more temporal variability with lower levels of predictability ($50\% \leq \Pi_u \leq 80\%$). Additionally one sees a sharp drop in predictability when comparing populations visiting less than 60 locations with those visiting more than this number. A similar trend is seen for geographical coverage, where once again a flat trend is exhibited by populations venturing not more than 10 km from their residence, whereas those travelling more than 100 km, experience the same variability as those visiting more than 60 unique locations. Now one sees a sharp drop in Π_u when comparing populations travelling more than 100 km with those that travel less. Finally, the trend is also mirrored when measuring the frequency of check-ins $\bar{f}_c$, the difference being the absence of any sharp drops in Π_u with a more smooth decrease between the segmented populations. Electronic supplementary material, figures S8 and S9 indicate the same trends for Brightkite and Gowalla. Across all datasets, the predictability decreases with more diversity in activity (as measured by S, r_g and $\bar{f}_c$). That is, it is harder to predict the mobility of users with more *diverse* routes irrespective of whether the diversity is measured in terms of the number of unique locations visited, the geographical area explored, or the amount of monthly data (traces). Taken together the results indicate an unexplored facet of uncertainty and predictability; stating that 'an individual is 80% predictable' must be interpreted in an *averaged* sense. Missing from this is the instantaneous changes in a person's *predictability state* over time.

Indeed, one would also expect this to vary with respect to the observation window of an individual's trajectory. For instance, if we observe a user only for an hour, they are likely to visit only a few locations. Given the lack of diversity in routes, the predictability should be high (as indicated by figure 2). Conversely, the longer the observation window, the more the number of locations visited, and therefore the predictability necessarily should decrease. It is interesting to consider whether there is a saturation in this decrease in Π_u. In figure 3*a*, we plot the distribution of Π_u as a function of the temporal window of observations for Gowalla, ranging from 1 to 72 h windows. As a baseline, the curve for the predictability considering the full set of trajectories (absent any window) is shown as a dashed line. As the time window gets smaller, the average predictability increases as expected. This is because we move from daily bins (each bin representing a full day of 24 h) to hourly bins (each bin representing a single hour within a day). In figure 3*b*, we plot the mode of the distributions as a function of the window size finding a saturation in the curve at 24 h for all three datasets.

2.4. Temporal and frequency modes in predictability

The temporal variation of the predictability seen in figure 2 suggests the possible presence of additional frequency modes in addition to the clear 24 h periodicity. To uncover this, one can use the continuous wavelet transform to describe the regularities in the time series of the individuals [52,53]. Wavelet analysis reveals the frequency components of signals just like the Fourier transform, but it also identifies where a certain frequency exists in the temporal or spatial domain (see electronic supplementary material, §S4 for details of the method). Figure 4*a* shows the wavelet power spectrum for each of

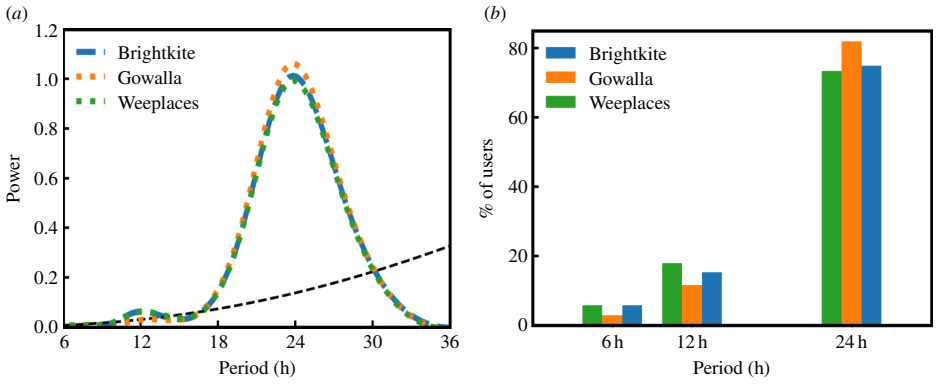

Figure 4. Temporal modes of predictability. (*a*) Estimated global wavelet power spectrum showing peaks at 24 hours (circadian) and 12 h (circasemidian) as well as a non-significant peak at 6 h. The dashed black line represents the statistical significance of the mode as measured by electronic supplementary material, equation (S3). (*b*) Stacked bar chart for the most strong component period (24, 12 and 6 hours) of each individual.

the datasets (the dashed line indicates the statistical significance, electronic supplementary material, equation S3). Not surprisingly, the figure reveals the circadian period (approximately 24 h) as the most prominent component of the predictability regularity. Additionally, and somewhat surprisingly, the 12 h component is the second-strongest component (i.e. the circasemidian period). Given working schedules and sleeping cycles, one might have expected a mode around the 8 h period, however, the third-strongest component is centred approximately around the 6 h regime during the day, even though the signal is not statistically significant at the population level. We note that the power spectrum is essentially identical across all three datasets, indicating the robustness of these modes in determining the observed temporal variation in predictability. Conducting the wavelet analysis on individual-level data, one can represent the percentage of users in each dataset that have either the 24, 12 or 6 h mode as their strongest components. The results are shown in figure 4*b* indicating that across all datasets between 73% and 81% of users have the 24 h mode as their strongest component, 12–18% of users have instead the 12 h mode as their strongest component, whereas a negligible fraction have the 6 h mode.

2.5. Location context of predictability

Thus far we have investigated the temporal context of uncertainty in mobility behaviour. It stands to reason that there is also a location-based context that influences movement patterns. For example, it is plausible that people are more predictable about their workplace or residence, as compared with locations that represent leisure and entertainment activities such as visiting restaurants and museums. This is a function of those types of locations being less frequently visited as compared with those that are driven by the daily work schedules.

To investigate this, one can scan the sequence of trajectory visits and group each location into categories. Now that we are once again analysing the full sequence of trajectories, temporal correlations in the sequence come into play and we can do a comparative analysis of both the correlated and uncorrelated predictabilities. The Weeplaces dataset contains a standardized set of eight location tags: *Home/Work*, *Education*, *Entertainment*, *Food*, *Travel*, *Shops*, *Outdoors* and *Nightlife*. Restricting the analysis of trajectories to each location type, one can compare it with the baseline distributions for $\Pi_{u,c}$ when considering all types of locations. For instance, a context-dependent trajectory for food could be: *breakfast-place-A*, *lunch-place-B*, *coffee-place-C*, etc.; while a full baseline trajectory could be: *home-place-X*, *breakfast-place-A*, *work-place-Y*, *lunch-place-B*, *coffee-place-C*, *gym-place-Z*, etc.

The results are shown in figure 5 with the baseline distribution shown as a dashed line. The figure clearly indicates the role of a spatial context in the distributions for both Π_c and Π_u. For instance, as expected, activities related to *Home/Work*, *Travel*, *Shops* are more predictable than the baseline; *Education*, *Outdoors* essentially mirror the baseline, whereas activities corresponding to *Food*, *Entertainment*, *Nightlife* have peaks markedly below the baseline.

The observed difference in predictability as a function of location-context behoves us to investigate the extent to which the frequency of which particular location types are visited has a bearing on the overall distribution of Π_c. Note this is different from merely considering the frequency distribution of

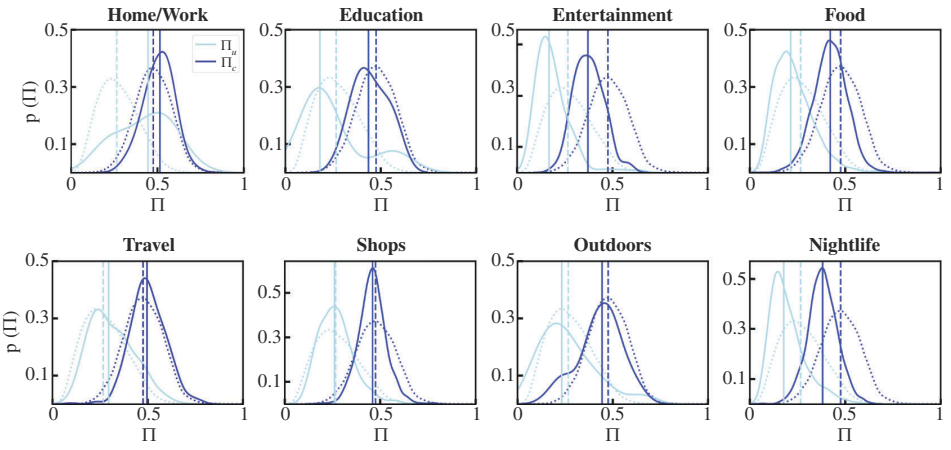

Figure 5. Location context of predictability. In all subplots, the dotted lines represent the distribution of $\Pi_{u,c}$ given their full trajectories for Weeplaces. Solid lines represent distributions for the same individuals but when limiting their trajectories to locations of distinct types. The vertical dashed lines correspond to the median values.

visiting a *particular location* which is the primary input to equation (2.1). Instead, here we coarse-grain these locations into categories and investigate whether this feature can be a good estimator for $P(\Pi_c)$. To that effect, we consider a linear regression model by using the relative frequency that an individual stayed at places from each category as an input. For instance, we investigate whether knowing that a person goes shopping often, informs us about this person's overall predictability. We train the model with the calculated predictability as an independent variable and the relative frequencies of the categories as the dependent variable.

Using this linear model, we estimate the predictability denoted by $P(\hat{\Pi}_c)$ (calculated from coarse-grained trajectories, e.g. 80% home/work, 18% food, 1% nightlife, etc.) in figure 6*a* and plot it against the true distribution $P(\Pi_c)$ (calculated from the full trajectories over multiple days, e.g. home, restaurant, work, restaurant, etc.). As the figure indicates, this simple model does a reasonable job of estimating the true distribution ($R^2 = 0.419$) with the residuals being normally distributed and centred around zero (figure 6*b*). (The coefficients of the model are shown in electronic supplementary material, table S2). Our results highlight two key points. First, the linear model's effectiveness as an approximation to the true distribution suggests the importance of location contexts. Second, this implies that capturing coarse-grained information about individuals, such as location types and visit frequencies, is sufficient to understand the general trends in mobility uncertainty, without resort to the full set of fine-grained mobility trajectories.

3. Discussion

Meeting a citizenry's needs require governments, industries and other stakeholders to be able to plan for demands (e.g. hospital admissions, public transportation and store opening times). The predictability of human movement is at the heart of planning, hence more accurate modelling should inform better decision-making in terms of public policy. Extant research models uncertainty and predictability in mobility trajectories as aggregated value for individual considering their full set of mobility trajectories as an input. This neglects the spatio-temporal factors influencing mobility, such as the temporal variations for different times of the day and days of the week, the levels of activity as well as the types of locations visited. As our results indicate, the predictability is greatly influenced by the window of observation, the length of the observation, the diversity of activities undertaken by an individual, as well as the location of the individual. Consequently, a more accurate way of modelling predictability is to consider it as a transient state, while taking into account the activity context.

The role of location context has some practical considerations. As our findings demonstrate, using coarse-grained location types and measuring visit frequencies provides a reasonable proxy for the predictability extracted from mobility trajectories. While traditional methods require finely grained data with frequent sampling, this approach offers a simpler and potentially more efficient alternative. Indeed, typically high-resolution mobility data are hard to come by and are restricted to only a few

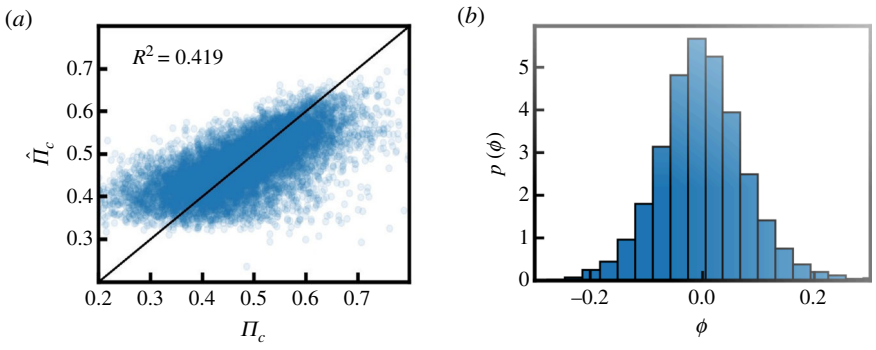

Figure 6. Estimating predictability from location context. (*a*) The estimated predictability $\hat{\Pi}_c$ using a linear model with frequency of visiting location types as input, plotted against the true distribution of Π_c in the Weeplaces dataset ($R^2 = 0.419$). (*b*) The residuals are normally distributed and centred around zero.

regions in the world. Our results show that much can be learnt even with low-resolution information. On the other hand, there is increasing access to higher-resolution data, such as the ones being collected as part of 'track-and-trace' systems in certain countries. In such instances, our framework of combining mobility traces with activity context can lead to more accurate characterizations. Indeed, many governments and companies (as part of 'data-for-good' efforts) are starting to open their datasets to scientists, which will naturally lead to better urban analytics, including human dynamics modelling.

The main implication of this work is that planning of activities related to human mobility (e.g. city events, epidemic modelling and road maintenance) needs to consider time–space variations of individual activities. Furthermore, during periods of restrictions, such as the COVID-19 pandemic, the understanding and characterization of these time–spatial variations can aid governments to make the correct decisions. In 2020 and 2021, many governments imposed curfew/lockdown measures on citizens after certain hours (e.g. Spain, Colombia) in a 'blanket' way. Effective curfews depend on the time and the location, and the consideration of such variations can lead to a better approach where not all areas are treated equally. Using predictability as a state and using the states for planning could lead to more just/equitable outcomes.

3.1. Limitations

Our work of course has limitations; given the analysis was conducted on a particular type of dataset (LBSNs) that vary in their spatio-temporal resolution, geographical coverage and sampling rates. It will be instructive to conduct this analysis in different regions of the world and using other sources of mobility information such as censuses, GPS data and call data records. Nevertheless, given that we conducted the analysis on three different datasets, finding similar trends points towards the results holding up to scrutiny in different settings.

Our analysis relies on the assumption that check-in data comprehensively represents all user activities throughout the day or week. In reality, limitations in data collection methods and potential user participation bias may affect this assumption. Additionally, while our datasets capture spatial information (latitude and longitude), the data reflects app usage rather than detailed activity types. Check-ins may not distinguish between specific activities at a given location, such as a work meeting, lunch break, or shopping trip. These factors call for cautious interpretation of the results and highlight the need for further research that incorporates additional data sources or methods to capture a more nuanced picture of user activity.

Ethics. This work did not require ethical approval from a human subject or animal welfare committee.

Data accessibility. *Brightkite and Gowalla data*: These LBSN datasets are publicly available from the Stanford Network Analysis Project (SNAP) database. We accessed them using their respective accession codes: Brightkite [48] and Gowalla [49]. Both datasets are cited Brightkite and Gowalla within the paper. *Weeplaces data*: This LBSN dataset is owned by Yong Liu and can be downloaded from their website: [50]. We have cited this data source Weeplaces within the paper.

Supplementary material is available online [54].

Declaration of AI use. We have not used AI-assisted technologies in creating this article.

Authors' contributions. B.P.: data curation, formal analysis, investigation, writing—review and editing; D.P.: formal analysis, project administration, writing—original draft; M.O.: data curation, formal analysis, writing—original draft; Z.C.: formal analysis, investigation, methodology; H.S.B.: data curation, formal analysis, investigation, methodology, project administration, writing—original draft; R.M.: conceptualization, funding acquisition, project administration, resources, supervision, writing—original draft, writing—review and editing; G.G.: conceptualization, funding acquisition, methodology, project administration, resources, supervision, writing—original draft, writing—review and editing.

All authors gave final approval for publication and agreed to be held accountable for the work performed therein.

Conflict of interest declaration. We declare we have no competing interests.

Funding. This work was supported by the US Army Research Office under Agreement Number W911NF-17-1-0127.

Acknowledgements. We thank Brooke Foucault Welles for useful discussions.

References

1. Balcan D, Colizza V, Gonçalves B, Hu H, Ramasco JJ, Vespignani A. 2009 Multiscale mobility networks and the spatial spreading of infectious diseases. *Proc. Natl. Acad. Sci. USA* **106**, 21484–21489. (doi:10.1073/pnas.0906910106)
2. Soriano-Paños D, Ghoshal G, Arenas A, Gómez-Gardeñes J. 2020 Impact of temporal scales and recurrent mobility patterns on the unfolding of epidemics. *J. Stat. Mech*. **2020**, 024006. (doi:10.1088/1742-5468/ab6a04)
3. Hazarie S, Soriano-Paños D, Arenas A, Gómez-Gardeñes J, Ghoshal G. 2021 Interplay between population density and mobility in determining the spread of epidemics in cities. *Commun. Phys*. **4**, 191. (doi:10.1038/s42005-021-00679-0)
4. Ansari M, Soriano-Paños D, Ghoshal G, White AD. 2022 Inferring spatial source of disease outbreaks using maximum entropy. *Phys. Rev. E* **106**, 014306. (doi:10.1103/PhysRevE.106.014306)
5. Soriano-Paños D, Cota W, Ferreira SC, Ghoshal G, Arenas A, Gómez-Gardeñes J. 2022 Modeling communicable diseases, human mobility, and epidemics: a review. *Ann. Phys*. **534**, 2100482. (doi:10.1002/andp.202100482)
6. Wang P, Hunter T, Bayen AM, Schechtner K, González MC. 2012 Understanding road usage patterns in urban areas. *Sci. Rep.* **2**, 1001. (doi:10.1038/srep01001)
7. Louf R, Barthelemy M. 2014 How congestion shapes cities: from mobility patterns to scaling. *Sci. Rep.* **4**, 5561. (doi:10.1038/srep05561)
8. Lee M, Barbosa H, Youn H, Holme P, Ghoshal G. 2017 Morphology of travel routes and the organization of cities. *Nat. Commun*. **8**, 2229. (doi:10.1038/s41467-017-02374-7)
9. Kirkley A, Barbosa H, Barthelemy M, Ghoshal G. 2018 From the betweenness centrality in street networks to structural invariants in random planar graphs. *Nat. Commun.* **9**, 2501. (doi:10.1038/s41467-018-04978-z)
10. Laxhammar R. 2014 Conformal anomaly detection: detecting abnormal trajectories in surveillance applications. PhD thesis, University of Skovde, Sweden. (doi:10.1016/B978-0-12-398537-8.00004-3)
11. Roth C, Kang SM, Batty M, Barthélemy M. 2011 Structure of urban movements: polycentric activity and entangled hierarchical flows. *PLoS One* **6**, e15923. (doi:10.1371/journal.pone.0015923)
12. Zhong C, Arisona SM, Huang X, Batty M, Schmitt G. 2014 Detecting the dynamics of urban structure through spatial network analysis. *Int. J. Geogr. Inf. Sci*. **28**, 2178–2199. (doi:10.1080/13658816.2014.914521)
13. Batty M. 2013 *The new science of cities*. Cambridge MA: MIT Press.
14. Pan W, Ghoshal G, Krumme C, Cebrian M, Pentland A. 2013 Urban characteristics attributable to density-driven tie formation. *Nat. Commun.* **4**, 1–7. (doi:10.1038/ncomms2961)
15. Barthelemy M. 2016 *The structure and dynamics of cities*. Cambridge, UK: Cambridge University Press.
16. Bassolas A *et al*. 2019 Hierarchical organization of urban mobility and its connection with city livability. *Nat. Commun.* **10**, 1–10. (doi:10.1038/s41467-019-12809-y)
17. Mimar S, Soriano-Paños D, Kirkley A, Barbosa H, Sadilek A, Arenas A, Gómez-Gardeñes J, Ghoshal G. 2022 Connecting intercity mobility with urban welfare. *PNAS. Nexus.* **1**, pgac178. (doi:10.1093/pnasnexus/pgac178)
18. Poudyal B, Ghoshal G, Kirkley A. 2023 Characterizing network circuity among heterogeneous urban amenities. *J. R. Soc. Interface* **20**, 20230296. (doi:10.1098/rsif.2023.0296)
19. Barbosa H *et al*. 2018 Human mobility: models and applications. *Phys. Rep*. **734**, 1–74. (doi:10.1016/j.physrep.2018.01.001)
20. González MC, Hidalgo CA, Barabási AL. 2008 Understanding individual human mobility patterns. *Nature* **453**, 479–482. (doi:10.1038/nature06958)
21. Hazarie S, Barbosa H, Frank A, Menezes R, Ghoshal G. 2020 Uncovering the differences and similarities between physical and virtual mobility. *J. R. Soc. Interface* **17**, 20200250. (doi:10.1098/rsif.2020.0250)
22. Chen S, Huang W, Cattani C, Altieri G. 2012 Traffic dynamics on complex networks: a survey. *Math. Probl. Eng.* **2012**, 732698. (doi:10.1155/2012/732698)
23. Lima A, Stanojevic R, Papagiannaki D, Rodriguez P, González MC. 2016 Understanding individual routing behaviour. *J. R. Soc. Interface* **13**, 20160021. (doi:10.1098/rsif.2016.0021)

24. Ghoshal G, Mangioni G, Menezes R, Poncela-Casanovas J. 2014 Social system as complex networks. *Soc. Netw. Anal. Min*. **4**, 238. (doi:10.1007/s13278-014-0238-9)
25. Song C, Qu Z, Blumm N, Barabási AL. 2010 Limits of predictability in human mobility. *Science* **327**, 1018–1021. (doi:10.1126/science.1177170)
26. Ikanovic EL, Mollgaard A. 2017 An alternative approach to the limits of predictability in human mobility. *EPJ Data Sci.* **6**, 12. (doi:10.1140/epjds/s13688-017-0107-7)
27. Smolak K, Siła-Nowicka K, Delvenne JC, Wierzbiński M, Rohm W. 2021 The impact of human mobility data scales and processing on movement predictability. *Sci. Rep.* **11**, 15177. (doi:10.1038/s41598-021-94102-x)
28. Sun K, Qian T, Chen T, Liang Y, Nguyen QVH, Yin H. 2020 Where to go next: modeling long- and short-term user preferences for point-of-interest recommendation. In *Proc. of the AAAI Conf. on Artificial Intelligence*, vol. **34**, pp. 214–221, NY, USA. (doi:(10.1609/aaai.v34i01.5353)
29. Rao X, Chen L, Liu Y, Shang S, Yao B, Han P. 2022 Graph-Flashback Network for Next Location Recommendation. In *Proc. of the 28th ACMSIGKDD Conf. on Knowledge Discovery and Data Mining*, Washington, DC, pp. 1463–1471. New York, NY. (doi:10.1145/3534678.3539383)
30. Long W, Xiao Z, Jiang H, Xiong Y, Qin Z, Li Y, Dustdar S. 2024 Learning semantic behavior for human mobility trajectory recovery. *IEEE trans. Intell. Transp. Syst*. 1–16. (doi:10.1109/TITS.2024.3350234)
31. Evans GW, Wener RE, Phillips D. 2002 The morning rush hour: predictability and commuter stress. *Environ. Behav.* **34**, 521–530. (doi:10.1177/0011652034004007)
32. Quelhas Martins A, McIntyre D, Ring C. 2015 Aversive event unpredictability causes stress-induced hypoalgesia. *Psychophysiology* **52**, 1066–1070. (doi:10.1111/psyp.12427)
33. Chen Z, Kelty S, Evsukoff AG, Welles BF, Bagrow J, Menezes R, Ghoshal G. 2022 Contrasting social and non-social sources of predictability in human mobility. *Nat. Commun.* **13**, 1922. (doi:10.1038/s41467-022-29592-y)
34. Stupfel M, Pavely A. 1990 Ultradian, circahoral and circadian structures in endothermic vertebrates and humans. *Comp. Biochem. Physiol. A Physiol.* **96**, 1–11. (doi:10.1016/0300-9629(90)90034-P)
35. Scheer FA, Wright KP, Kronauer RE, Czeisler CA. 2007 Plasticity of the intrinsic period of the human circadian timing system. *PLoS One* **2**, e721. (doi:10.1371/journal.pone.0000721)
36. Toole JL, Herrera-Yaqüe C, Schneider CM, González MC. 2015 Coupling human mobility and social ties. *J. R. Soc. Interface* **12**, 20141128. (doi:10.1098/rsif.2014.1128)
37. Schneider CM, Belik V, Couronné T, Smoreda Z, González MC. 2013 Unravelling daily human mobility motifs. *J. R. Soc. Interface* **10**, 20130246. (doi:10.1098/rsif.2013.0246)
38. Hasan. S, Schneider CM, Ukkusuri SV, González, MC. 2012 Spatiotemporal patterns of urban human mobility. *J. Stat. Phys.* **151**, 304–318. (doi:10.1007/s10955-012-0645-0)
39. Barbosa H, Hazarie S, Dickinson B, Bassolas A, Frank A, Kautz H, Sadilek A, Ramasco JJ, Ghoshal G. 2021 Uncovering the socioeconomic facets of human mobility. *Sci. Rep*. **11**, 8616. (doi:10.1038/s41598-021-87407-4)
40. Santana C *et al*. 2020 *Analysis of socioeconomic aspects related to mobility patterns in the UK during the COVID-19 pandemic (2020)*. See https://covid19-uk-mobility.github.io/Second-report (accessed 1 June 2020).
41. Santana C, Botta F, Barbosa H, Privitera F, Menezes R, Di Clemente R. 2023 COVID-19 is linked to changes in the time-space dimension of human mobility. *Nat. Hum. Behav.* **7**, 1729–1739. (doi:10.1038/s41562-023-01660-3)
42. De Domenico M, Lima A, Musolesi M. 2013 Interdependence and predictability of human mobility and social interactions. *Perv. Mob. Comput.* **9**, 798–807. (doi:10.1016/j.pmcj.2013.07.008)
43. Zhang HT, Zhu T, Fu D, Xu B, Han XP, Chen D. 2018 Spatiotemporal property and predictability of large-scale human mobility. *Physica A. Stat. Mech. Appl.* **495**, 40–48. (doi:10.1016/j.physa.2017.12.024)
44. Lempel A, Ziv J. 1976 On the complexity of finite sequences. *IEEE Trans. Inf. Theory* **22**, 75–81. (doi:10.1109/TIT.1976.1055501)
45. Kontoyiannis I, Algoet PH, Suhov YuM, Wyner AJ. 1998 Nonparametric entropy estimation for stationary processes and random fields, with applications to English text. *IEEE Trans. Inf. Theory* **44**, 1319–1327. (doi:10.1109/18.669425)
46. Jiang S, Yang Y, Gupta S, Veneziano D, Athavale S, González MC. 2016 The TimeGeo modeling framework for urban mobility without travel surveys. *Proc. Natl Acad. Sci. USA* **113**, E5370–E5378. (doi:10.1073/pnas.1524261113)
47. Cornacchia G, Pappalardo L. 2021 A mechanistic data-driven approach to synthesize human mobility considering the spatial, temporal, and social dimensions together. *ISPRS Int. J. Geoinf.* **10**, 599. (doi:10.3390/ijgi10090599)
48. Brightkite dataset. See https://snap.stanford.edu/data/loc-brightkite.html (accessed 15 September 2023).
49. Gowalla dataset. See https://snap.stanford.edu/data/loc-gowalla.html (accessed 15 September 2023).
50. Weeplaces dataset. See https://www.yongliu.org/datasets/ (accessed 15 September 2023)
51. Cover TM, Thomas JA. 2006 *Elements of information theory*, 2nd edn. Dordrecht, The Netherlands: John Wiley & Sons.
52. Percival DP. 1995 On estimation of the wavelet variance. *Biometrika* **82**, 619–631. (doi:10.1093/biomet/82.3.619)
53. Torrence C, Compo GP. 1998 A practical guide to wavelet analysis. *Bull. Am. Meteorol. Soc.* **79**, 61–78. (doi:10.1175/1520-0477(1998)079<0061:APGTWA>2.0.CO;2)
54. Poudyal B, Oliveira MA, Pacheco D, Chen Z, Barbosa H, Menezes R *et al*. 2024 Supplementary material from: Dynamic predictability and activity-location contexts in human mobility. Figshare (doi:10.6084/m9.figshare.c.7410574)