PNAS RESEARCH ARTICLE | APPLIED PHYSICAL SCIENCES SOCIAL SCIENCES

Mixing individual and collective behaviors to predict out-of-routine mobility

Sebastiano Bontorin[a,b], Simone Centellegher[a], Riccardo Gallotti[c], Luca Pappalardo[d,e], Bruno Lepri[a], and Massimiliano Luca[a,1]

Affiliations are included on p. 8.

Edited by Andrea Rinaldo, Ecole Polytechnique Federale de Lausanne, Lausanne, Switzerland; received July 24, 2024; accepted March 19, 2025

Predicting human displacements is crucial for addressing various societal challenges, including urban design, traffic congestion, epidemic management, and migration dynamics. While predictive models like deep learning and Markov models offer insights into individual mobility, they often struggle with out-of-routine behaviors. Our study introduces an approach that dynamically integrates individual and collective mobility behaviors, leveraging collective intelligence to enhance prediction accuracy. Evaluating the model on millions of privacy-preserving trajectories across five US cities, we demonstrate its superior performance in predicting out-of-routine mobility, surpassing even advanced deep learning methods. The spatial analysis highlights the model's effectiveness near urban areas with a high density of points of interest, where collective behaviors strongly influence mobility. During disruptive events like the COVID-19 pandemic, our model retains predictive capabilities, unlike individual-based models. By bridging the gap between individual and collective behaviors, our approach offers transparent and accurate predictions, which are crucial for addressing contemporary mobility challenges.

human mobility | generalization | collective behavior | GPS Data

Significance

Crowds significantly influence individual decisions, as evidenced by studies on collective intelligence, social psychology, and behavioral economics. Existing models that predict individual patterns struggle to predict unexpected behaviors as they tend to memorize individual preferences and lack generalization capabilities. Our approach mixes collective and individual decisions to overcome these limitations while offering interpretability and stronger performance in out-of-routine mobility prediction. We find collective information is crucial in forecasting out-of-routine movements by capturing broader patterns and trends that individual-based models overlook. In addition, we find that collective information makes the model robust to external shocks, potentially allowing it to adapt to sudden changes in mobility patterns, such as those caused by public events, infrastructure disruptions, or natural disasters.

Understanding human mobility patterns is relevant to many pressing problems in our societies (1), including the design of sustainable and livable cities (2, 3), traffic congestion avoidance (4, 5), epidemics spread mitigation and public health monitoring (6–9), urban and socioeconomic segregation (10, 11), and migration management after natural disasters, economic shocks, and wars (12–16).

The task of predicting individuals' future whereabouts, often referred to as next location prediction (17, 18), has attracted particular interest in light of the growing availability of extensive mobility data and the development of advanced statistical techniques (1, 17, 19). On the one hand, sophisticated deep learning solutions have gained substantial attention given their capacity to uncover complex patterns from extensive datasets (17, 20–23). These models often lack interpretability, functioning as black boxes that obscure the underlying mechanisms driving predictions (1, 17). On the other hand, simple and interpretable models such as Markov models allow the analysis of the mechanisms behind the predictions but often exhibit lower accuracy in forecasting future movements (24–27). Both deep learning and Markov models are trained on individual mobility trajectories, which are typically inherently predictable as people tend to visit previously visited locations at regular times (19, 28–33). However, in some cases, individuals may have a marked preference for exploring new destinations (31, 32, 34–39) or be forced to alter their routine due to external factors such as job loss, health issues, natural disasters, or epidemics (9, 36, 40–48). Predicting such out-of-routine mobility is a challenge for statistical models because, being designed to capture regular patterns in individual trajectories, they often memorize training data rather than learning generalized mobility behaviors (17, 33, 49). A large body of literature on human behavior across various contexts, such as social networks (50–53), financial networks (54–56), and voting and political polarization (57–59), indicates that an individual's decisions are significantly influenced by the behavior of the group or community they are exposed to (60–64). This may also suggest that information about collective behaviors also holds predictive power on individuals' choices. However, the potential of combining individual and collective behaviors to enhance human mobility prediction remains largely unexplored. For instance, deep learning models for human mobility memorize individual preferences and do not use collective preferences explicitly (49), while physics-grounded models integrated individual and collective information thought weighting schema (26). In both approaches, the generalization capabilities and predictability properties of such

The authors declare no competing interest.

This article is a PNAS Direct Submission.

[1]To whom correspondence may be addressed. Email: mluca@fbk.eu.

This article contains supporting information online at https://www.pnas.org/lookup/suppl/doi:10.1073/pnas.2414848122/-/DCSupplemental.

Published April 23, 2025.

models have never been thoroughly investigated. In this manuscript, we bridge this gap by introducing an approach for next-location prediction that dynamically integrates individual and collective mobility behaviors, leveraging the expected predictability of an individual's movement. By explicitly integrating a collective component into the model, we demonstrate through an evaluation framework that our approach effectively generalizes to out-of-routine mobility and remains robust against external shocks that cause behavioral shifts. Additionally, this integration allows us to analyze the interplay between collective and individual patterns and their relationship with the socioeconomic and geographical characteristics of cities.

We evaluate our model's performance using the trajectories of millions of anodized, opted-in individuals across five US cities over eight months and compare its accuracy with approaches relying only on individual or collective data. By offering transparency without compromising predictive accuracy, our model demonstrates considerable generalization capabilities, particularly in out-of-routine mobility behaviors, surpassing more sophisticated deep learning methods. Notably, a detailed spatial analysis reveals that our model particularly benefits from collective information in out-of-routine mobility prediction, especially in areas with a high density of points of interest (POIs). While individual-based models are significantly affected by disruptions to recurrent mobility patterns, our model's dynamic integration of individual and collective behaviors enables it to maintain predictive capabilities during disruptive events such as the COVID-19 pandemic.

Results

A spatiotemporal point is a pair $p = (i, t)$, where i represents a geographic location and t the time of the visit. We define a trajectory $P = \{p_1, p_2, \ldots, p_n\}$ as a daily time-ordered sequence of n spatiotemporal points. Each individual user u has a set of N historical trajectories $\mathcal{H}^{(u)} = \{P_1, \ldots, P_N\}$ from which we compute $I_i^{(u)}$, representing the set of transition probabilities of user u starting from location i (see Fig. 1*A* and *Materials and Methods* for details). By aggregating the trajectories of all individuals, we calculate the collective OD matrix C (Fig. 1*B*), with C_i representing the probability distribution of all transitions made by any individual starting from location i.

Given an individual's trajectory $P \in \mathcal{H}^{(u)}$, next location prediction is the problem of forecasting the next point $p_{n+1} \in P$ (17, 18, 49). $I_i^{(u)}$ and C_i are Markov-based solutions to next location predictors (18, 25).

We define a model $M_i^{(u)}$ that dynamically combines $I_i^{(u)}$ and C_i based on the predictability of u's next location from origin i. When the next location is highly predictable based on $\mathcal{H}^{(u)}$, the model relies more on individual information in $I_i^{(u)}$ for the prediction. Conversely, the model relies more on the collective information in C_i when the next location is hard to predict. To quantify the predictability of u's next location from i, we employ the normalized Shannon's entropy of $I_i^{(u)}$ (28, 65, 66):

$$S_i^{(u)} = -\frac{\sum_{k \in L^{(u)}} I_{i,k}^{(u)} \cdot \log(I_{i,k}^{(u)})}{\log |L^{(u)}|}, \quad [1]$$

where $L^{(u)}$ is the set of distinct locations visited by u, $\log |L^{(u)}|$ is a normalization factor so that $S_i^{(u)} \in [0, 1]$, and $I_{i,k}^{(u)}$ is the probability of u moving from location i to location k. $S_i^{(u)}$ is high when the locations u visited from i have similar visitation probabilities, indicating a diverse range of destinations; $S_i^{(u)}$ is low when u predominantly visits one location from i, indicating a marked individual preference for a specific destination. We use $S_i^{(u)}$ to combine probabilities from $I_i^{(u)}$ and C_i as follows:

$$M_i^{(u)} = (1 - S_i^{(u)}) I_i^{(u)} + S_i^{(u)} C_i, \quad [2]$$

where $1 - S_i^{(u)}$ is the confidence of model $M_i^{(u)}$ in relying on individual information. To derive Markov transition probabilities, we normalize $M_i^{(u)}, \forall i$ using the softmax function:

$$\text{softmax}(l)_i = \frac{e^{l_i}}{\sum_{j=1}^{n} e^{l_j}},$$

where $l_1, l_2, \ldots l_n$ are the transition probabilities of $M_i^{(u)}$ in Eq. **2**. When a transition from location i is not represented in $\mathcal{H}^{(u)}$, the probability distribution $I_i^{(u)}$ is empty, and we set $S_i^{(u)} = 1$ to indicate maximum uncertainty. In such instances, no historical data are available for location i, and the model prediction relies solely on collective information in C_i. Fig. 1 *C–E* illustrates how model $M_i^{(u)}$ works. Additional information on the computational costs of the proposed model can be found in *SI Appendix*, Note S11.

We employ the notations I, C, and M to represent the models exploiting individual, collective, and their combination. To derive I and C for real individuals, we use privacy-enhanced GPS trajectories collected in Boston, New York City (NYC), Phoenix, New Orleans, and Seattle from January 3rd to March 1st, 2020 (*SI Appendix*, Note S1, Table S1, and Fig. S1). We tessellate the cities into rectangular tiles (locations) of 1.2 km × 609.4 m, corresponding to GeoHashes of level 6. These tiles are the basis for mapping user's stops and computing users' trajectories (*Materials and Methods*). To address potential under- and overrepresentation issues, we filter out trajectories with less than four points ($|P^{(u)}| < 4$), users with fewer than two trajectories ($|\mathcal{H}^{(u)}| < 2$), and remove the top 95th percentile of the most represented users (*SI Appendix*, Note S1). Fig. 1 *A* and *B* shows examples of I and C derived from the GPS dataset for the city of Boston.

For each individual u, we allocate 80% of their least recent trajectories for model training, while the 20% most recent trajectories form the test set. During the training phase, for each location i and user u, we compute $I_i^{(u)}$, C_i, and $S_i^{(u)}$. Subsequently, we perform next location predictions on the test set and assess the models' performance using the top-5 accuracy metric (ACC@5). ACC@5 is a standard metric for evaluating next location prediction tasks and represents the percentage of instances where the correct next location is among the top five predicted destinations (17, 49).

Table 1 and Fig. 2*A* show that M significantly enhances accuracy compared to I and C and obtains comparable performances with deep learning baselines like RNNs (17, 67). Further information on RNN implementation can be found in *SI Appendix*, Note S6. M exhibits relative improvements in ACC@5 over I of +14% in NYC, +11% in Seattle, +9% in Boston, +13% in Phoenix and +25% in New Orleans. Furthermore, M outperforms C with improvements of +31% in

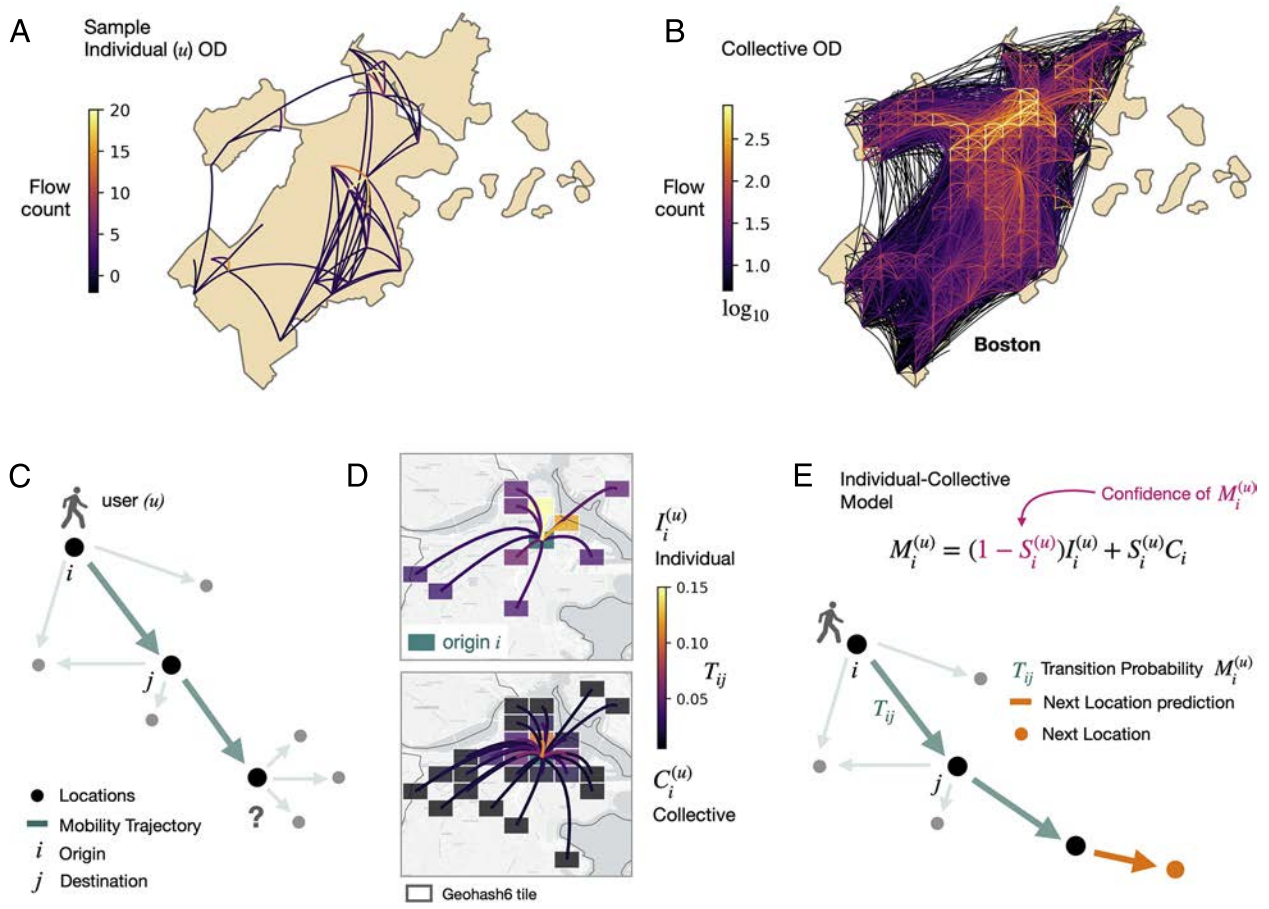

Fig. 1. Dynamic interplay of individual and collective mobility. (*A*) An individual origin–destination (OD) matrix for a synthetic individual u. (*B*) The collective OD matrix computed for Boston using GPS trajectories. (*C*) An individual trajectory for a user u starting from location i. Next location prediction consists of predicting u's next visited location. (*D*) The set of u's historical trajectories (panel *A*) is used to define the transition probabilities $I_i^{(u)}$ from location i. C_i represents the probability distribution of all transitions made by any user starting from location i, generated from the OD matrix (panel *B*). Destinations' locations j are colored based on their visitation probability, T_{ij}, from origin i. (*E*) $M_i^{(u)}$'s prediction of individual u's next location is performed by dynamically combining $I_i^{(u)}$ and C_i, based on the normalized Shannon entropy $S_i^{(u)}$ computed from the mobility trajectories of u. Maps: Stamen Maps. Icons: Fontawesome.

NYC, +23% in Seattle, and +14% in Boston +32% in Phoenix and +8% in New Orleans. Despite relying on a large number of parameters, RNNs exhibit lower performance compared to M, with a relative performance reduction of −4% in NYC, −2% in Seattle, and −4% in Boston, −4% in Phoenix and −4% in New Orleans (*SI Appendix*, Note S7 and Tables S3–S5). Therefore, the dynamic interplay between individual and collective information enables model M to significantly enhance the overall predictive performance compared to models relying exclusively on individual (I) or collective (C) information. Tests on the robustness of M to temporal stratifications of C are presented in *SI Appendix*, Note S7 and Tables S6 and S7.

Models' Generalization Capability. The recurrent patterns in human mobility often result in significant similarity among individual trajectories (19, 49, 68). Consequently, our test set may include a significant portion of trajectories already present in the training set. Recent research emphasizes the importance of considering trajectory overlap between the training and test sets for a comprehensive evaluation of next-location predictors, as it significantly influences the assessment of a model's generalization capability (49).

We quantify trajectory train-test overlap with the LCST (49), which evaluates shared subsequences between two trajectories, considering the order and frequency of visits. LCST ranges between 0 and 1 and measures how much of a test trajectory the model has already observed in the training set. A high LCST indicates the presence of recurrent mobility patterns in an individual's trajectories, while a low LCST indicates instances of an individual's out-of-routine mobility behaviors. We stratify test trajectories into bins based on LCST, spanning overlap ranges of 0 to 20, 20 to 40, 40 to 60, 60 to 80, and 80 to 100% (see *Materials and Methods* for details). For example, a test trajectory falls within the 0 to 20% bin when its maximum overlap with any training trajectory has an LCST $\in [0, 0.2]$. The distribution of test trajectories across the bins is not uniform, as bins 0 to 20% and 80 to 100% contain fewer trajectories than bins 20 to 80% (*SI Appendix*, Notes S1 and S8 and Figs. S18 and S19).

Table 1 and Fig. 2*B* provide models' performance within each trajectory overlap bin. M significantly outperforms I and C for intermediate levels of overlap, with improvements across cities up to +16% (20 to 40% overlap), +13% (40 to 60% overlap), and around +2% (60 to 80% overlap). C surpasses M for 0 to 20% overlap only, with improvements up to 16% across the cities, while I outperforms M by only around 1% for 80 to 100% overlap. Fig. 2*C* shows that the median value of $1 - S_i^{(u)}$ distribution increases with trajectory overlap, i.e., the higher the trajectory overlap, the higher the confidence of model M on I

Table 1. Accuracy of models

	Full test set			0 to 20			20 to 40			40 to 60			60 to 80			80 to 100		
	I	*C*	*M*	*I*	*C*	*M*	*I*	*C*	*M*	*I*	*C*	*M*	*I*	*C*	*M*	*I*	*C*	*M*
NYC	0.588	0.509	**0.669**	0.065	**0.399**	0.350	0.347	0.399	**0.483**	0.616	0.473	**0.671**	0.849	0.581	**0.855**	**0.981**	0.733	0.953
Boston	0.678	0.648	**0.739**	0.052	**0.456**	0.382	0.346	0.455	**0.504**	0.643	0.562	**0.707**	0.872	0.707	**0.872**	**0.988**	0.867	0.918
Seattle	0.617	0.554	**0.682**	0.044	**0.384**	0.320	0.340	0.402	**0.476**	0.618	0.491	**0.675**	0.853	0.627	**0.857**	**0.985**	0.818	0.905
Phoenix	0.581	0.498	**0.655**	0.063	**0.358**	0.323	0.344	0.361	**0.463**	0.610	0.466	**0.656**	0.839	0.584	**0.844**	**0.978**	0.773	0.933
New Orleans	0.545	0.628	**0.681**	0.054	**0.485**	0.410	0.325	0.492	**0.519**	0.620	0.596	**0.706**	0.842	0.723	**0.861**	**0.983**	0.888	0.908

Performance of models *I*, *C*, *M* on all trajectories in the full test set and the trajectories stratified based on their overlap with training trajectories using the longest common sub-trajectory (LCST). Bold value is the best performing model for each city/overlap bin.

and, consequently, the lower the reliance on *C*. In other words, when historical information about the individual is unreliable due to significant differences from the current trajectory (low trajectory overlap), using collective information is crucial for making reasonable predictions. When the current trajectory closely resembles those in the training set (high trajectory overlap), it is best to rely on individual information. Note that RNNs achieve a low accuracy on low trajectory overlap bins and high accuracy on high overlap bins (Fig. 2*B*), confirming that deep learning models struggle to predict out-of-routine mobility. To verify the robustness of our findings, we carry out the same analysis using tessellations of different sizes and shapes. In particular, in *SI Appendix*, Note S7 and Tables S3–S5, we show the results obtained with Geohashes of level 7 (153 m × 153 m) and hexagonal H3 tiles of level 9 (0.105 km^2). In addition, we also explored other metrics and methods to stratify test trajectories

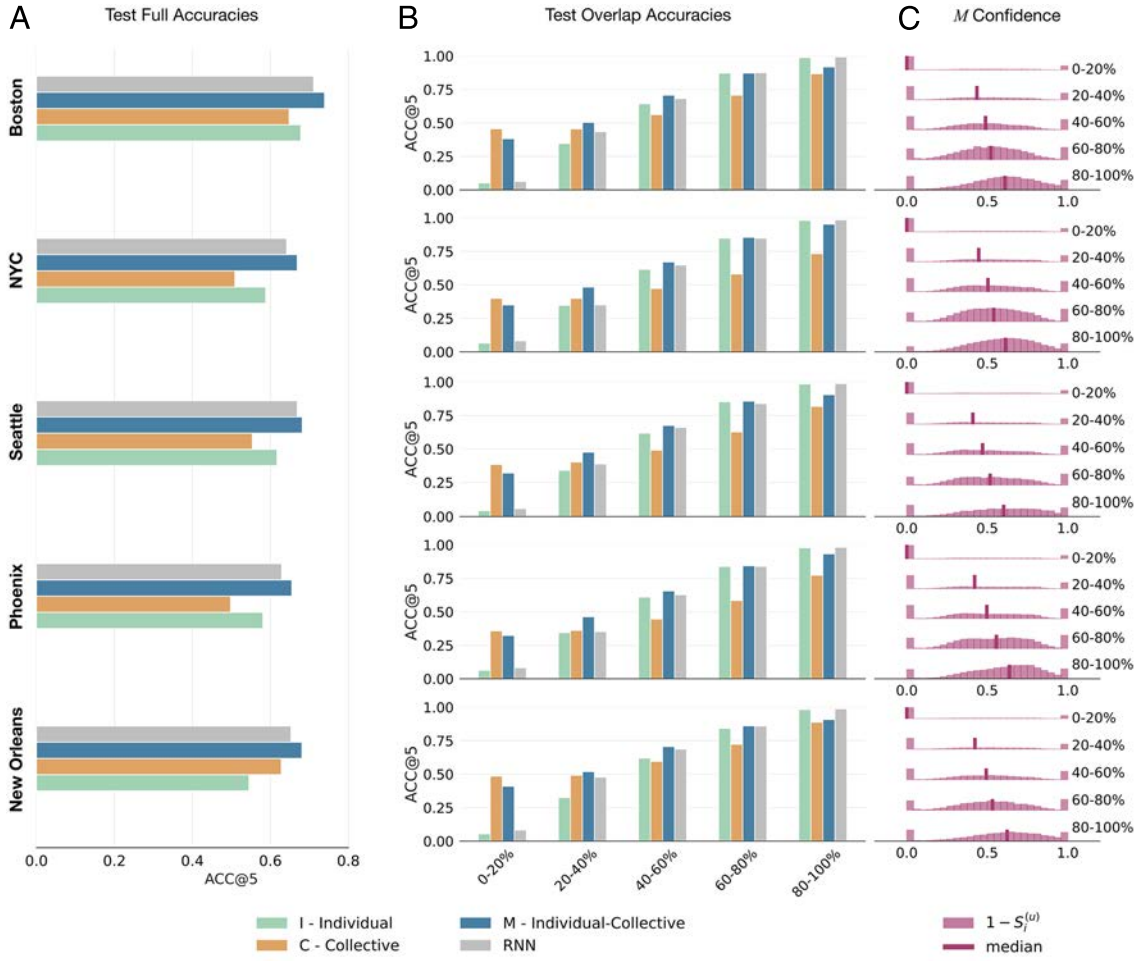

Fig. 2. Accuracy of the models. Top-5 accuracy (ACC@5) for Boston, New York City (NYC), Phoenix, New Orleans, and Seattle using models *I*, *C*, and *M*. (*A*) ACC@5 on the full test set. *M* shows a performance better than *I* and *C* and comparable to recurrent neural networks (RNNs). (*B*) Models are tested against different train-test overlap scenarios (based on LCST), with 0 to 20% describing out-of-routine mobility and 80 to 100% routinary mobility behavior. *M* shows improvements in accuracy over *I* in smaller overlaps, where test trajectories mostly consist of novel transitions never observed during training. (*C*) Distributions of *M*'s confidence, represented by $1 - S_i^{(u)}$, in relying on individual information *I*. In the case of out-of-routine behaviors (low overlaps), the lower median value of $1 - S_i^{(u)}$ indicates less reliance of *M* on individual information *I*. In this scenario, collective behaviors *C* enhance the predictive capabilities of *M*. The peaks observed around $1 - S_i^{(u)} = 0$ result from instances of transitions from a location *i* that is not represented in the training trajectories of user *u*. In such a case, we set $S_i^{(u)} = 1$, forcing *M* to rely only on *C*.

(*Materials and Methods*), finding similar patterns (*SI Appendix*, Notes S7–S9 and Figs. S18–S20).

Spatial Properties of Models' Accuracy. To investigate potential spatial dependencies in the accuracy of I, C, and M, we compute ACC@5 specifically for the subset of test transitions originating from a location i. We refer to it as ACC@5_i. In Fig. 3*A*, we present the spatial distributions of ACC@5_i for the city of Boston. We find similar distributions for Seattle, NYC, Phoenix, and New Orleans (*SI Appendix*, Notes S2 and S4 and Figs. S3–S6 and S12).

In scenarios with low trajectory overlap (0 to 40%), the ACC@5_i distributions for M and C are consistently shifted to higher values compared to I (Fig. 3*A*). While model C performs best for trajectories with 0 to 20% overlap, model M shows the highest performance in the 20 to 40% overlap range, benefiting from collective information from C. As trajectory overlap increases, the ACC@5_i distributions for M and I become more similar, while the distribution for C widens and remains skewed toward lower accuracy (Fig. 3*A*).

Fig. 3*C* shows that when out-of-routine is predominant (0 to 40% overlap), there is considerable spatial heterogeneity in the distribution of ACC@5_i across different locations. This is particularly pronounced when using collective information for predictions, as in C and M. We quantify this spatial property using Moran's index (69, 70) and find that models C and M exhibit a significant and positive spatial autocorrelation, indicating that locations where the models are accurate are spatially close (Fig. 3 *B* and *C* and *SI Appendix*, Figs. S3–S6).

In particular, we observe that locations with high accuracies for C and M are clustered in the proximity of critical urban areas. For instance, in Boston around downtown and Logan International Airport (Fig. 3*C*), in NYC in Manhattan, in Seattle, and New Orleans around downtown (*SI Appendix*, Figs. S3–S5) and Phoenix (*SI Appendix*, Note S3 and Fig. S6).

The observed spatial autocorrelation for C and M suggests a potential relationship between collective mobility behaviors and spatially clustered urban factors when out-of-routine mobility is predominant. To verify this hypothesis, we measure the predictability of collective behaviors from a location i as the entropy of collective mobility:

$$S_i^{(C)} = -\frac{\sum_{k\in L^{(C)}} C_{ik}\cdot \log(C_{ik})}{\log|L^{(C)}|} \in [0,1], \qquad [3]$$

where $L^{(C)}$ is the set of unique locations in C. $S_i^{(C)}$ is high when locations people visited from i have similar visitation probabilities, indicating a diverse range of destinations; $S_i^{(C)}$ is low when people predominantly visit one location from i, indicating a marked collective preference for a specific destination. $S_i^{(C)}$ is strongly negatively correlated with ACC@5_i (Pearson correlation of $\rho = -0.85$), indicating that locations from which C provides the most accurate next location predictions are those with the lowest entropy $S_i^{(C)}$ (Fig. 4*A* and *SI Appendix*, Figs. S10 and S11). Notably, we find that locations with low $S_i^{(C)}$ are clustered

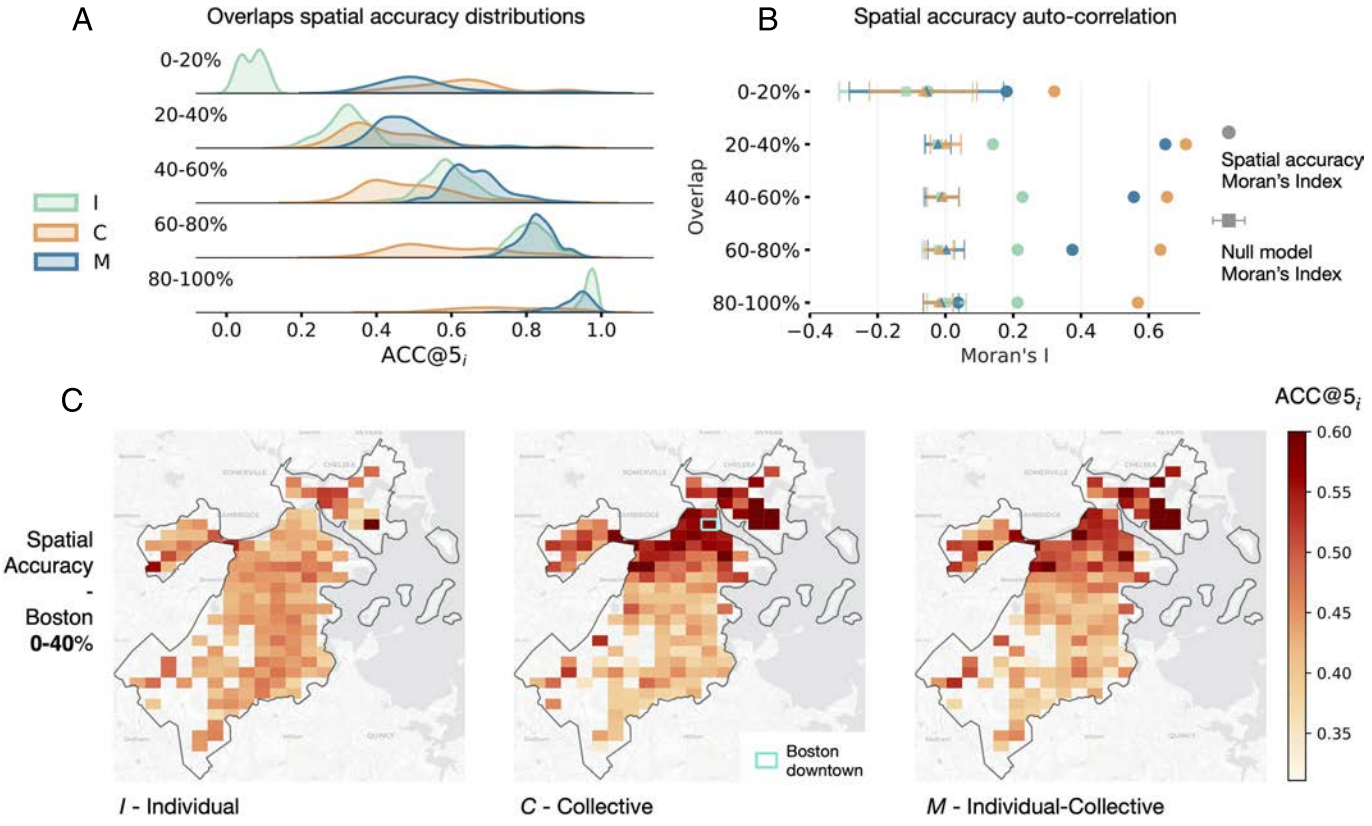

Fig. 3. Spatial distributions of accuracies. (*A*) Distribution of accuracies for I, C, and M in predicting movements from a location i (ACC@5_i). As the test set includes more out-of-routine movements (e.g., 0 to 40% overlap), model M aided by collective information C performs better. Conversely, the individual model I provides better predictions when tested on routinary trajectories (high overlaps). (*B*) Spatial autocorrelation of the models' accuracies in corresponding overlaps quantified via the Moran's index. For low overlaps, such as 0 to 40%, model C exhibits clustered accuracy (large Moran's index). (*C*) Spatial distribution of ACC@5_i in Boston for I, C, and M in the 0 to 40% overlap. Notably, for C and M, areas with higher accuracies are concentrated in proximity to downtown (*Upper Center*) and Boston Logan International Airport (*Upper Right*). Maps: Stamen Maps.

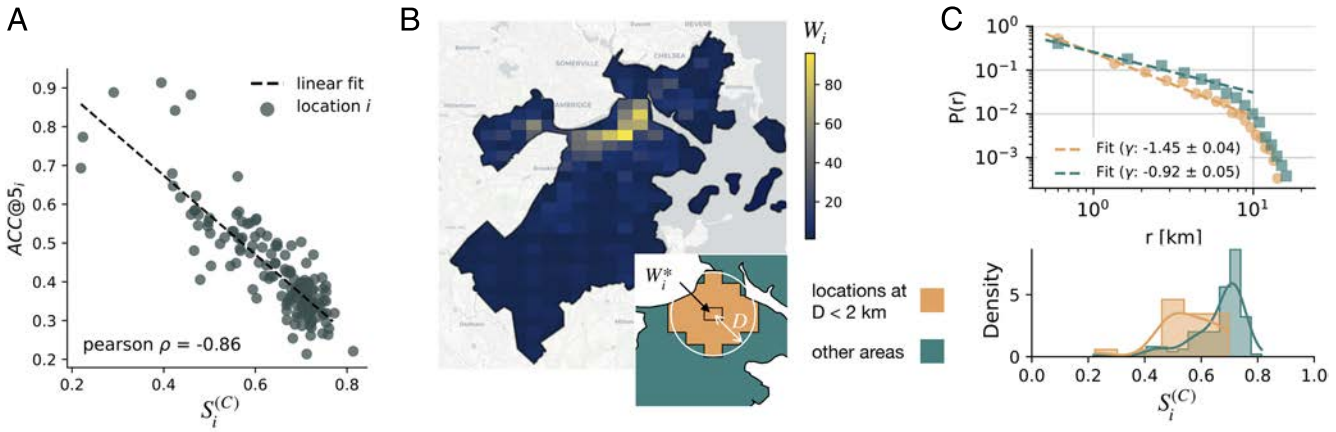

Fig. 4. Statistical properties of collective mobility in Boston. (*A*) Accuracy of *C* (ACC@5$_i$) from a location *i* versus its entropy $S_i^{(C)}$. We find a negative correlation (Pearson $\rho = -0.86$). (*B*) Spatial distribution of the number of POIs per location, W_i (extracted from OpenStreetMap). (*Inset*) In orange, locations within a distance $D = 2$ km from the location i^* with the largest number of POIs (W_i^*). In green, locations farther away from i^* by a distance greater than *D*. (*C*) Distribution of travel distances, $P(r)$, distinguishing between origins within the two aforementioned areas and fitted with a power-law function (19) in the interval of 0 to 10 km. The exponent of $\gamma = -1.45 \pm 0.04$ underscores the prevalence of localized mobility when individuals are in proximity to i^*, while in other areas we have an exponent of $\gamma = -0.92 \pm 0.04$. This result indicates that mobility near POIs tends to be more concentrated and less spatially dispersed toward specific destinations. This behavior is further corroborated by the entropy $S_i^{(C)}$ distribution, which skews toward lower values and indicates mobility directed toward specific tiles. Maps: Stamen Maps.

in proximity to specific urban areas, which we hypothesize to be locations hosting key commercial, financial, and cultural venues.

To verify this hypothesis, we collect from OpenStreetMap (71) the number of POIs in each location *i*, W_i. We then split each city into two areas: one comprising locations within a geographical distance *D* from the location i^* with the maximum number of POIs (W_i^*), and the other consisting of locations farther away from i^* by a distance greater than *D* (see Fig. 4*B* for $D = 2$ km in Boston).

We find that the distribution of $S_i^{(C)}$ skews toward lower values when location *i* is within distance *D* to i^* (orange area in Fig. 4*B*). Individuals' movements that originate within this distance tend to travel shorter distances than those originating farther away (Fig. 4*C*). These results support our hypothesis that mobility in dense areas of POIs is less spatially dispersed, with the probability of traveling to a destination concentrated in a small subset of locations. We find similar results in Seattle, NYC, New Orleans, and Phoenix (*SI Appendix*, Note S3 and Figs. S7–S9). For statistical robustness, we verify that these results are not a consequence of possible biases, such as areas close to POIs having a larger sample size of transitions in the dataset (*SI Appendix*, Note S5 and Figs. S13–S17). Finally, since predictions may potentially vary based on sociodemographic/socioeconomic factors, we tested whether there exists a spatial cross correlation between the accuracies produced by our model *M* and the median income levels (*SI Appendix*, Note S10 and Fig. S21).

Model Reliability Under COVID-19 Restrictions. The COVID-19 pandemic significantly altered people's mobility patterns (9, 44, 45, 72, 73), with nonpharmaceutical interventions inducing a shift in how people moved and visited locations (9, 10, 74). In this context, we investigate the reliability of the models during these behavioral shifts.

We train *I*, *C*, and *M* using trajectories recorded until March 11th, 2020, and evaluate their performance on a test set consisting of 5 mo of data collected between March 11th, 2020 [WHO's pandemic declaration (75)], and August 11th, 2020. For each month in the test set, we report models' ACC@5 in Table 2.

In each city, *I* exhibits a notable decline in performance as the months go by, losing 44.16% of its predictive power by the time of testing trajectories collected between July and August. *C*'s performance also degrades over time but more moderately, with an average drop of 4.51% between the first and last months. *M* exhibits a moderate average decrease of 5.32%, which is slightly worse than *C* but notably better than *I*. These results highlight that models based on individual-level information are less resilient to behavioral shifts than models based on collective information. As a consequence, combining individual information with collective one allows for enhanced resilience against disruptive behavioral changes.

Discussion

Crowds influence individual decisions, a phenomenon extensively documented in studies on collective intelligence, social psychology, and behavioral economics (50, 51, 54, 55, 57, 58, 76, 77). In various scenarios, individuals may face uncertainty or possess limited knowledge, prompting them to defer to collective decisions for guidance. We leverage these insights in a parameter-free approach to tackle the next-location prediction problem. Our solution dynamically integrates individual and collective information. Our model is designed to harness collective behaviors alongside individual patterns, offering a holistic framework for predicting future whereabouts.

We offer a potential solution to the limitations inherent in current state-of-the-art models, including sophisticated deep learning approaches. Indeed, while deep learning models excel in predicting routinary movements (17, 49), they suffer from two primary drawbacks: opacity and limited interpretability (1), and difficulties in forecasting out-of-routine choices because of their tendency to memorize patterns observed in individual trajectories (49). Our approach overcomes these shortcomings by offering full interpretability and strong performance in out-of-routine movement prediction, even during disruptive events like the COVID-19 pandemic. Collective information is particularly crucial in anticipating these out-of-routine choices, enabling us to capture patterns and trends that may not be apparent at the

Table 2. Models accuracy during the COVID-19 pandemics

	March to April			April to May			May to June			June to July			July to August		
	I	*C*	*M*	*I*	*C*	*M*	*I*	*C*	*M*	*I*	*C*	*M*	*I*	*C*	*M*
NYC	0.481	0.545	0.657	0.391	0.544	0.621	0.357	0.539	0.605	0.323	0.528	0.588	0.288	0.519	0.573
Boston	0.489	0.636	0.699	0.395	0.614	0.659	0.349	0.624	0.654	0.306	0.623	0.645	0.254	0.618	0.641
Seattle	0.423	0.584	0.652	0.349	0.589	0.632	0.309	0.588	0.624	0.269	0.562	0.601	0.234	0.548	0.584
Phoenix	0.396	0.529	0.620	0.311	0.510	0.580	0.276	0.490	0.555	0.241	0.483	0.541	0.216	0.476	0.529
New Orleans	0.362	0.618	0.677	0.326	0.569	0.620	0.295	0.559	0.604	0.244	0.556	0.596	0.213	0.556	0.593

ACC@5 of *I*, *C*, and *M* trained on trajectories collected before March 1st, evaluated on trajectories during and after lockdown every month after the pandemic declaration (March 11th). The COVID-19 pandemic acts as a disruptive event that introduces behavioral changes in human mobility. The individual model *I* loses half of its predictive capabilities, while the models relying on collective behaviors (*C* and *M*) generalize even after months, maintaining accuracies comparable to the pre-COVID-19 period.

individual level alone. In the context of the ongoing discourse on mechanistic versus deep learning models for human mobility (17, 19), our study suggests the necessity to integrate dynamic mechanisms like ours into deep learning frameworks. This integration is essential for achieving predictions that are both interpretable and accurate in scenarios involving both routinary and out-of-routine movements.

Our study additionally reveals that predictions relying on collective information, especially the most accurate ones, spatially cluster around urban areas characterized by a dense concentration of POIs. In these areas, the probability of traveling to a destination is highly concentrated within a smaller subset of locations. These findings align with recent research on flow generation, underscoring that the presence and nature of POIs play a significant role in shaping human mobility (78). An intriguing research challenge lies in enhancing our model by incorporating additional mechanisms that account for factors such as the density of POIs, as well as sociodemographic and socioeconomic characteristics. These factors could potentially improve model performance, particularly in areas where signals from collective OD matrices are weaker.

In addition, while our model enhances generalization capabilities and has the potential to address cold-start issues (i.e., predicting the mobility of new individuals), it relies on a substantial amount of data to construct the collective matrix. However, there are regions where sufficient data may not be available to compute an accurate collective matrix. As there are models capable of generating mobility flows based on population and distances (19), it would be interesting to explore whether the collective matrix could be replaced with flow-generated data. Using such models could also potentially mitigate issues of biases and data representativeness. Our findings also motivate the exploration of new loss functions—particularly for deep learning models—that incorporate constraints derived from collective mobility laws (19, 79).

Materials and Methods

Stop Location Detection. For our experiments, we use anonymized and privacy-enhanced GPS dataset provided by Cuebiq as part of the Data for Good COVID-19 Collaborative program. This dataset encompasses privacy-enhanced GPS locations spanning nine months (January to August 2020) in New York City, Seattle, Boston, Phoenix, and New Orleans. The data originate from approximately 2 million anonymous users who willingly opted-in to share their information anonymously for research purposes, adhering to the guidelines of the CCPA (California Consumer Privacy Act) compliant framework. In addition to anonymizing the data, the data provider removes sensitive POIs from the dataset and obfuscates inferred home locations to the Census Block Group level. All researchers were contractually obligated not to share data further or attempt to deidentify data. Studies using the same datasets obtained IRB exemptions (11, 80), and the same conclusion was reached for our study and the dataset we used by the ethical committee of the Human Inspired Technology research center (https://hit.psy.unipd.it/comitato_etico_HIT) of the University of Padua (protocol 2025_267).

We extract individual user stops from the dataset through the following procedures inspired from other studies. Initially, we identify each temporal sequence of GPS coordinates within a 65-m radius, where a user stayed for a minimum of 5 min (81). Subsequently, we apply the DBSCAN algorithm (82) to identify dense clusters of points within a distance of $\epsilon = \Delta_s - 5$. We define these dense clusters as stop locations. For a more detailed explanation of the GPS data processing, refer to ref. 9.

POIs. We download from OpenStreetMap (OSM) data about POIs in New York City, Seattle, Boston, Phoenix and New Orleans. POIs describe public venues such as restaurants, and parks in a city. We employ a dictionary of amenities covering a large set of public venues (see *SI Appendix*, Note S1 for the specific entries used). We compute the number of POIs extracted from OSM in each GeoHash tile i and refer to it as W_i. The maps of POIs for Boston, Seattle, New York City, Phoenix and New Orleans can be seen in *SI Appendix*, Fig. S7.

OD Matrix. Each individual u in our datasets is associated with a set of consecutive trajectories $\mathcal{H}^{(u)} = \{P_1, \ldots, P_N\}$, each capturing the locations visited over a 24-h period. The OD matrix of u captures the transition probability, $T_{ij}^{(u)}/T_i^{(u)}$, between each pair of locations visited by u, where $T_{ij}^{(u)}$ is the total number of transitions in $\mathcal{H}^{(u)}$ from location i to location j, and $T_i^{(u)} = \sum_{j \in L} T_{ij}^{(u)}$ accounts for the total transitions from location i to any other location in the set $L^{(u)}$ of locations visited by u (68).

Trajectory Level Overlap. To quantify the overlap between two trajectories $P^{(u)} = \{p_1, p_2, \ldots, p_n\}$ and $R^{(u)} = \{r_1, r_2, \ldots, r_m\}$, we propose to use two widely adopted metrics: Jaccard Similarity (JS) and LCST (49).

With JS, we can quantify the proportion of locations in the test trajectories that overlap with those in the training trajectories, independent of their order and number of instances in the trajectory. JS varies between 0 and 1. A high JS indicates that the test trajectories share many locations with the training data. In contrast, a low JS suggests greater unpredictability in test trajectories, as they largely consist of locations absent from the training set. Formally, we define the JS between a trajectory $R \in D_{\text{test}}$ and $P \in D_{\text{train}}$ as

$$\text{JS}(P, R) = \frac{|P \cup R| - |P \cap R|}{|P \cup R|}.$$

We quantify the overlap between R and the training set as the maximum JS over all the trajectories in the training set:

$$\max_{P \in D_{\text{train}}} \text{JS}(P, R).$$

Measuring the similarity between trajectories, considering the order of visits and related frequency, can be fundamental. Thus, we also use LCST as a stratification metric as LCST considers both frequency and order when measuring how much novel a trajectory is. To compute the LCST between two trajectories, $P^{(u)} = \{p_1, p_2, \ldots, p_n\}$ and $R^{(u)} = \{r_1, r_2, \ldots, r_m\}$, we introduce the prefix P_i of $P^{(u)}$ as the list of the first i-th locations in $P^{(u)}$, i.e., $P_i = \{p_1, \ldots, p_i\}$ (dropping index $^{(u)}$ for simplicity). The definition extends similarly to R. The size of LCST for two prefixes P_i and R_j is denoted as LCST (f in Eq. **4**) and is defined as follows:

$$f(P_i, R_j) = \begin{cases} 0, & \text{if } i = 0 \text{ or } j = 0 \\ f(P_{i-1}, R_{j-1}) + 1, & \text{if } i, j > 0 \text{ and } p_i = r_j \\ \max(f(P_{i-1}, R_j), f(P_i, R_{j-1})), & \text{if } i, j > 0 \text{ and } p_i \neq r_j. \end{cases} \quad [4]$$

We quantify the overlap between a test trajectory R and the training set as the maximum LCST over all the trajectories in the training set:

$$\max_{P \in \mathcal{H}^{(u)}} \text{LCST}(P, R). \quad [5]$$

We normalize the LCST score within the range $[0, 1]$. Each trajectory, regardless of the metric, is assigned to one of the following five bins: 0 to 20%, 20 to 40%, 40 to 60%, 60 to 80%, and 80 to 100%, based on their overlap score (49). To ensure that results do not depend on how bins are selected, we also show how accuracies and overlap scores relate in *SI Appendix*, Note S7, Tables S3–S5, and Figs. S18 and S19. For example, a test trajectory with an LCST score less than or equal to 0.2 will be assigned to the bin of trajectories with 0-20% LCST. We exclude trajectories with an exact 0% overlap as they primarily consist of individuals remaining in the same location for a small set of locations. Refer to *SI Appendix*, Note S1, Table S2, and Fig. S2 for further details. The distribution of LCST/JS for test trajectories is detailed in *SI Appendix*, Note S8 and Figs. S18 and S19.

Data, Materials, and Software Availability. Code has been deposited in GitHub (https://github.com/ssai-trento/mixing-individual-collective-mobility) (83). Some study data are available: the data supporting the findings of this study are accessible through Cuebiq's Data for Good initiative. For details on how to request access, including conditions and limitations, please visit: https://www.cuebiq.com/about/data-for-good/.

ACKNOWLEDGMENTS. We thank Cuebiq for kindly providing us with the mobility dataset for this research through their Data for Good program. L.P. has been supported by 1) PNRR (Piano Nazionale di Ripresa e Resilienza) in the context of the research program 20224CZ5X4_PE6_PRIN 2022 "URBAI–Urban Artificial Intelligence" (CUP B53D23012770006), funded by European Union–Next Generation EU; 2) EU project H2020 SoBigData++ G.A. 871042; 3) NextGenerationEU-National Recovery and Resilience Plan (Piano Nazionale di Ripresa e Resilienza, PNRR), Project "SoBigData.it-Strengthening the Italian RI for Social Mining and Big Data Analytics," prot. IR0000013, avviso n. 3264 on 28/12/2021. The work of S.C., R.G., B.L., and M.L. has been supported by the PNRR ICSC National Research Centre for High Performance Computing, Big Data and Quantum Computing (CN00000013), under the NRRP MUR program funded by the NextGenerationEU. B.L. also acknowledges the support of the PNRR project FAIR - Future AI Research (PE00000013), under the NRRP MUR program funded by the NextGenerationEU and by the European Union's Horizon Europe research and innovation program under Grant Agreement No. 101120237 (ELIAS). Portions of the paper were developed from the thesis of S.B.

Author affiliations: [a]Mobile and Social Computing Lab, Bruno Kessler Foundation, Povo 38123, TN, Italy; [b]Department of Physics, University of Trento, Povo 38123, TN, Italy; [c]Complex Human Behavior Lab, Bruno Kessler Foundation, Povo 38123, TN, Italy; [d]Istituto di Scienza e Tecnologie dell'Informazione-National Research Council, Pisa 56127, PI, Italy; and [e]Scuola Normale Superiore of Pisa, Pisa 56126, PI, Italy

Author contributions: S.B. and M.L. designed research; S.B., S.C., and M.L. performed research; S.B. and M.L. contributed new reagents/analytic tools; S.B., S.C., and M.L. analyzed data; S.B., S.C., L.P., B.L., and M.L. interpreted results; and S.B., S.C., R.G., L.P., B.L., and M.L. wrote the paper.

1. L. Pappalardo, E. Manley, V. Sekara, L. Alessandretti, Future directions in human mobility science. *Nat. Comput. Sci.* **3**, 588–600 (2023).
2. M. Lee, P. Holme, Relating land use and human intra-city mobility. *PLoS ONE* **10**, e0140152 (2015).
3. M. Böhm, M. Nanni, L. Pappalardo, Gross polluters and vehicle emissions reduction. *Nat. Sustainability* **5**, 699–707 (2022).
4. M. Akhtar, S. Moridpour, A review of traffic congestion prediction using artificial intelligence. *J. Adv. Transp.* **2021**, 1–18 (2021).
5. P. Wang, T. Hunter, A. M. Bayen, K. Schechtner, M. C. González, Understanding road usage patterns in urban areas. *Sci. Rep.* **2**, 1001 (2012).
6. A. Wesolowski *et al.*, Quantifying the impact of human mobility on malaria. *Science* **338**, 267–270 (2012).
7. C. Xiong, S. Hu, M. Yang, W. Luo, L. Zhang, Mobile device data reveal the dynamics in a positive relationship between human mobility and covid-19 infections. *Proc. Natl. Acad. Sci. U.S.A.* **117**, 27087–27089 (2020).
8. M. U. Kraemer *et al.*, The effect of human mobility and control measures on the covid-19 epidemic in china. *Science* **368**, 493–497 (2020).
9. L. Lucchini *et al.*, Living in a pandemic: Changes in mobility routines, social activity and adherence to covid-19 protective measures. *Sci. Rep.* **11**, 1–12 (2021).
10. T. Yabe, B. G. B. Bueno, X. Dong, A. Pentland, E. Moro, Behavioral changes during the covid-19 pandemic decreased income diversity of urban encounters. *Nat. Commun.* **14**, 2310 (2023).
11. E. Moro, D. Calacci, X. Dong, A. Pentland, Mobility patterns are associated with experienced income segregation in large US cities. *Nat. Commun.* **12**, 4633 (2021).
12. X. Lu, L. Bengtsson, P. Holme, Predictability of population displacement after the 2010 Haiti earthquake. *Proc. Natl. Acad. Sci. U.S.A.* **109**, 11576–11581 (2012).
13. P. Deville *et al.*, Dynamic population mapping using mobile phone data. *Proc. Natl. Acad. Sci. U.S.A.* **111**, 15888–15893 (2014).
14. M. C. Gonzalez, C. A. Hidalgo, A. L. Barabasi, Understanding individual human mobility patterns. *Nature* **453**, 779–782 (2008).
15. P. Bosetti *et al.*, Heterogeneity in social and epidemiological factors determines the risk of measles outbreaks. *Proc. Natl. Acad. Sci. U.S.A.* **117**, 30118–30125 (2020).
16. P. P. Klamser *et al.*, Enhancing global preparedness during an ongoing pandemic from partial and noisy data. *PNAS Nexus* **2**, pgad192 (2023).
17. M. Luca, G. Barlacchi, B. Lepri, L. Pappalardo, A survey on deep learning for human mobility. *ACM Comput. Surv. (CSUR)* **55**, 1–44 (2021).
18. A. G. Chekol, M. S. Fufa, A survey on next location prediction techniques, applications, and challenges. *EURASIP J. Wirel. Commun. Netw.* **2022**, 29 (2022).
19. H. Barbosa *et al.*, Human mobility: Models and applications. *Phys. Rep.* **734**, 1–74 (2018).
20. A. De Brébisson, É. Simon, A. Auvolat, P. Vincent, Y. Bengio, "Artificial neural networks applied to taxi destination prediction" in *Proceedings of the 2015th International Conference on ECML PKDD Discovery Challenge* (2015), pp. 40–51.
21. Q. Liu, S. Wu, L. Wang, T. Tan, "Predicting the next location: A recurrent model with spatial and temporal contexts" in *Thirtieth AAAI Conference on Artificial Intelligence* (2016).
22. J. Feng *et al.*, "DeepMove: Predicting human mobility with attentional recurrent networks" in *Proceedings of the 2018 World Wide Web Conference* (2018), pp. 1459–1468.
23. H. Xue, F. Salim, Y. Ren, N. Oliver, MobTCast: Leveraging auxiliary trajectory forecasting for human mobility prediction. *Adv. Neural Inf. Process. Syst.* **34**, 30380–30391 (2021).
24. D. Ashbrook, T. Starner, "Learning significant locations and predicting user movement with GPS" in *Proceedings of the Sixth International Symposium on Wearable Computers* (IEEE, 2002), pp. 101–108.
25. S. Gambs, M. O. Killijian, M. N. del Prado Cortez, "Next place prediction using mobility Markov chains" in *Proceedings of the First Workshop on Measurement, Privacy, and Mobility* (2012), pp. 1–6.
26. F. Calabrese, G. Di Lorenzo, C. Ratti, "Human mobility prediction based on individual and collective geographical preferences" in *13th International IEEE Conference on Intelligent Transportation Systems* (IEEE, 2010), pp. 312–317.
27. C. Zhao, A. Zeng, C. H. Yeung, Characteristics of human mobility patterns revealed by high-frequency cell-phone position data. *EPJ Data Sci.* **10**, 5 (2021).
28. C. Song, Z. Qu, N. Blumm, A. L. Barabási, Limits of predictability in human mobility. *Science* **327**, 1018–1021 (2010).
29. M. De Domenico, A. Lima, M. Musolesi, Interdependence and predictability of human mobility and social interactions. *Pervasive Mobile Comput.* **9**, 798–807 (2013).
30. X. Lu, E. Wetter, N. Bharti, A. J. Tatem, L. Bengtsson, Approaching the limit of predictability in human mobility. *Sci. Rep.* **3**, 2923 (2013).
31. L. Pappalardo *et al.*, Returners and explorers dichotomy in human mobility. *Nat. Commun.* **6**, 8166 (2015).
32. A. Cuttone, S. Lehmann, M. C. González, Understanding predictability and exploration in human mobility. *EPJ Data Sci.* **7**, 1–17 (2018).
33. K. Smolak, K. Siła-Nowicka, J. C. Delvenne, M. Wierzbiński, W. Rohm, The impact of human mobility data scales and processing on movement predictability. *Sci. Rep.* **11**, 15177 (2021).
34. L. Scherrer, M. Tomko, P. Ranacher, R. Weibel, Travelers or locals? Identifying meaningful sub-populations from human movement data in the absence of ground truth. *EPJ Data Sci.* **7**, 1–21 (2018).

35. L. Amichi, A. C. Viana, M. Crovella, A. A. Loureiro, "Understanding individuals' proclivity for novelty seeking" in *Proceedings of the 28th International Conference on Advances in Geographic Information Systems, SIGSPATIAL '20* (Association for Computing Machinery, New York, NY, 2020), pp. 314–324.
36. Z. He, Y. Hu, L. L. Duan, G. Michailidis, Returners and explorers dichotomy in the face of natural hazards. *Sci. Rep.* **14**, 13184 (2024).
37. R. Wang, N. Li, Y. Wang, Does the returners and explorers dichotomy in urban human mobility depend on the observation duration? An empirical study in Guangzhou, China *Sustain. Cities Soc.* **69**, 102862 (2021).
38. M. Schläpfer *et al*., The universal visitation law of human mobility. *Nature* **593**, 522–527 (2021).
39. C. Krumme, A. Llorente, M. Cebrian, A. Pentland, E. Moro, The predictability of consumer visitation patterns. *Sci. Rep.* **3**, 1645 (2013).
40. J. L. Toole *et al*., Tracking employment shocks using mobile phone data. *J. R. Soc. Interface* **12**, 20150185 (2015).
41. A. Almaatouq, F. Prieto-Castrillo, A. Pentland, "Mobile communication signatures of unemployment" in *Social Informatics: 8th International Conference (SocInfo 2016), Bellevue, WA, November 11–14, 2016, Proceedings, Part I 8* (Springer, 2016), pp. 407–418.
42. H. Barbosa *et al*., Uncovering the socioeconomic facets of human mobility. *Sci. Rep.* **11**, 8616 (2021).
43. S. Centellegher, M. De Nadai, M. Tonin, B. Lepri, L. Lucchini, The long-term and disparate impact of job loss on individual mobility behaviour. arXiv [Preprint] (2024). https://arxiv.org/abs/2403.10276 (Accessed 10 October 2024).
44. N. Haug *et al*., Ranking the effectiveness of worldwide covid-19 government interventions. *Nat. Hum. Behav.* **4**, 1303–1312 (2020).
45. J. Zhang *et al*., The impact of relaxing interventions on human contact patterns and SARS-COV-2 transmission in China. *Sci. Adv.* **7**, eabe2584 (2021).
46. R. Wilson *et al*., Rapid and near real-time assessments of population displacement using mobile phone data following disasters: The 2015 Nepal Earthquake. *PLoS Curr.* **8**, 1–24 (2016).
47. B. Hong, B. J. Bonczak, A. Gupta, C. E. Kontokosta, Measuring inequality in community resilience to natural disasters using large-scale mobility data. *Nat. Commun.* **12**, 1870 (2021).
48. C. L. Gray, V. Mueller, Natural disasters and population mobility in Bangladesh. *Proc. Natl. Acad. Sci. U.S.A.* **109**, 6000–6005 (2012).
49. M. Luca, L. Pappalardo, B. Lepri, G. Barlacchi, Trajectory test-train overlap in next-location prediction datasets. *Mach. Learn.* **112**, 4597–4634 (2023).
50. W. Mason, D. J. Watts, Collaborative learning in networks. *Proc. Natl. Acad. Sci. U.S.A.* **109**, 764–769 (2012).
51. B. Jayles *et al*., How social information can improve estimation accuracy in human groups. *Proc. Natl. Acad. Sci. U.S.A.* **114**, 12620–12625 (2017).
52. J. Becker, D. Brackbill, D. Centola, Network dynamics of social influence in the wisdom of crowds. *Proc. Natl. Acad. Sci. U.S.A.* **114**, E5070–E5076 (2017).
53. A. Almaatouq *et al*., Adaptive social networks promote the wisdom of crowds. *Proc. Natl. Acad. Sci. U.S.A.* **117**, 11379–11386 (2020).
54. M. Kaustia, S. Knüpfer, Peer performance and stock market entry. *J. Fin. Econ.* **104**, 321–338 (2012).
55. W. Pan, Y. Altshuler, A. Pentland, "Decoding social influence and the wisdom of the crowd in financial trading network" in *2012 International Conference on Privacy, Security* (2012), pp. 203–209.
56. R. Z. Heimer, Peer pressure: Social interaction and the disposition effect. *Rev. Fin. Stud.* **29**, 3177–3209 (2016).
57. S. Coleman, The effect of social conformity on collective voting behavior. *Polit. Anal.* **12**, 76–96 (2004).
58. J. Becker, E. Porter, D. Centola, The wisdom of partisan crowds. *Proc. Natl. Acad. Sci. U.S.A.* **116**, 10717–10722 (2019).
59. F. Shi, M. Teplitskiy, E. Duede, J. A. Evans, The wisdom of polarized crowds. *Nat. Hum. Behav.* **3**, 329–336 (2019).
60. J. Surowiecki, *The Wisdom of Crowds* (Anchor, 2005).
61. J. R. Dyer, A. Johansson, D. Helbing, I. D. Couzin, J. Krause, Leadership, consensus decision making and collective behaviour in humans. *Philos. Trans. R. Soc. Lond. B, Biol. Sci.* **364**, 781–789 (2009).
62. A. W. Woolley, C. F. Chabris, A. Pentland, N. Hashmi, T. W. Malone, Evidence for a collective intelligence factor in the performance of human groups. *Science* **330**, 686–688 (2010).
63. M. A. Hogg, K. D. Williams, From I to We: Social identity and the collective self. *Group Dyn. Theory Res. Pract.* **4**, 81 (2000).
64. H. Tajfel, Individuals and groups in social psychology. *Br. J. Soc. Clin. Psychol.* **18**, 183–190 (1979).
65. N. Eagle, A. S. Pentland, Eigenbehaviors: Identifying structure in routine. *Behav. Ecol. Sociobiol.* **63**, 1057–1066 (2009).
66. L. Pappalardo *et al*., An analytical framework to nowcast well-being using mobile phone data. *Int. J. Data Sci. Anal.* **2**, 75–92 (2016).
67. I. Goodfellow, Y. Bengio, A. Courville, *Deep Learning* (MIT Press, 2016).
68. C. M. Schneider, V. Belik, T. Couronné, Z. Smoreda, M. C. González, Unravelling daily human mobility motifs. *J. R. Soc. Interface.* **10**, 20130246 (2013).
69. L. Anselin, Local indicators of spatial association—LISA. *Geogr. Anal.* **27**, 93–115 (1995).
70. Y. Chen, A new methodology of spatial cross-correlation analysis. *PLoS ONE* **10**, e0126158 (2015).
71. OpenStreetMap contributors, Planet dump retrieved from https://planet.osm.org (2017). https://www.openstreetmap.org. Accessed 10 October 2024.
72. S. Flaxman *et al*., Estimating the effects of non-pharmaceutical interventions on covid-19 in Europe. *Nature* **584**, 257–261 (2020).
73. M. Chinazzi *et al*., The effect of travel restrictions on the spread of the 2019 novel coronavirus (covid-19) outbreak. *Science* **368**, 395–400 (2020).
74. L. Napoli, V. Sekara, M. García-Herranz, M. Karsai, Socioeconomic reorganization of communication and mobility networks in response to external shocks. *Proc. Natl. Acad. Sci. U.S.A.* **120**, e2305285120 (2023).
75. World Health Organization *et al.*, WHO director-general's opening remarks at the media briefing on covid-19 (January 30, 2022). https://www.who.int/director-general/speeches/detail/who-director-general-s-opening-remarks-at-the-media-briefing-on-covid-19---11-march-2020. Accessed 10 October 2024.
76. R. B. Cialdini, R. B. Cialdini, *Influence: The Psychology of Persuasion* (Collins, New York, NY, 2007), vol. 55.
77. T. A. Venema, F. M. Kroese, J. S. Benjamins, D. T. De Ridder, When in doubt, follow the crowd? Responsiveness to social proof nudges in the absence of clear preferences. *Front. Psychol.* **11**, 499433 (2020).
78. F. Simini, G. Barlacchi, M. Luca, L. Pappalardo, A deep gravity model for mobility flows generation. *Nat. Commun.* **12**, 6576 (2021).
79. S. Cuomo *et al*., Scientific machine learning through physics-informed neural networks: Where we are and what's next. *J. Sci. Comput.* **92**, 88 (2022).
80. S. Wang *et al*., Infrequent activities predict economic outcomes in major American cities. *Nat. Cities* **1**, 305–314 (2024).
81. R. Hariharan, K. Toyama, "Project Lachesis: Parsing and modeling location histories" in *Geographic Information Science*, M. J. Egenhofer, C. Freksa, H. J. Miller, Eds. (Springer, Berlin/Heidelberg, Germany, 2004), pp. 106–124.
82. M. Ester *et al*., "A density-based algorithm for discovering clusters in large spatial databases with noise" in *Proceedings of the Second International Conference on Knowledge Discovery and Data Mining, KDD'96* (1996), pp. 226–231.
83. S. Bontorin *et al*., Mixing individual and collective behaviours to predict out-of-routine mobility - Repository. GitHub. https://github.com/ssai-trento/mixing-individual-collective-mobility. Deposited 10 February 2025.