Scaling identity connects human mobility and social interactions

Pierre Deville[a,b], Chaoming Song[c], Nathan Eagle[d], Vincent D. Blondel[a], Albert-László Barabási[b,e,f], and Dashun Wang[g,1]

[a]Department of Applied Mathematics, Université catholique de Louvain, 1348 Louvain-la-Neuve, Belgium; [b]Center for Complex Network Research, Department of Physics, Biology and Computer Science, Northeastern University, Boston, MA 02115; [c]Department of Physics, University of Miami, Coral Gables, FL 33142; [d]College of Computer Science, Northeastern University, Boston, MA 02115; [e]Center for Cancer Systems Biology, Dana Farber Cancer Institute, Boston, MA 02115; [f]Department of Medicine, Brigham and Women's Hospital, Harvard Medical School, Boston, MA 02115; and [g]College of Information Sciences and Technology, Pennsylvania State University, University Park, PA 16802

Edited by Adrian Dobra, University of Washington, Seattle, WA, and accepted by Editorial Board Member Adrian E. Raftery April 20, 2016 (received for review January 21, 2016)

Massive datasets that capture human movements and social interactions have catalyzed rapid advances in our quantitative understanding of human behavior during the past years. One important aspect affecting both areas is the critical role space plays. Indeed, growing evidence suggests both our movements and communication patterns are associated with spatial costs that follow reproducible scaling laws, each characterized by its specific critical exponents. Although human mobility and social networks develop concomitantly as two prolific yet largely separated fields, we lack any known relationships between the critical exponents explored by them, despite the fact that they often study the same datasets. Here, by exploiting three different mobile phone datasets that capture simultaneously these two aspects, we discovered a new scaling relationship, mediated by a universal flux distribution, which links the critical exponents characterizing the spatial dependencies in human mobility and social networks. Therefore, the widely studied scaling laws uncovered in these two areas are not independent but connected through a deeper underlying reality.

human mobility | social interactions | mobile phone data | social networks | spatial networks

Over the past few years, we have witnessed tremendous progress in uncovering patterns behind human mobility (1–7) and social networks (8–10), owing partly to the increasing availability of large-scale datasets capturing human behavior in a new level of detail, resolution, and scale (11, 12). Building on rich, fundamental literature from the social sciences (13–19), these data offer a huge opportunity for research, fueling concomitant advances in areas of both human mobility and social networks with profound consequences in broad domains. One important aspect affecting both areas is the critical role space plays. Indeed, growing evidence suggests both our movements and communication patterns are associated with spatial costs that follow reproducible scaling laws. Indeed, previous studies have shown that human travels adhere to spatial constraints (20), characterized by levy flights and continuous time random walk models (1, 2, 4), a scaling law that has proven to be critical in various phenomena driven by human mobility, from spread of viruses (21–23) to migrations (2, 6) and emergency response (24–26). In another related yet distinct area, there has been much empirical evidence about the geographic effect on communication patterns (20), documenting that the probability for two individuals to communicate decays with distance, following power law distributions (20, 27–30). This robust pattern plays an important role in navigating the social network (31), from routing (32, 33) to search of experts (34, 35) to spread of information (27, 36) and innovations (37). Although human movements and social interactions bear high-level similarities in the role spatial distance plays, and are often referred to as two prominent examples of spatial networks (20), they remain as largely separate lines of inquiry, lacking any known connections between their critical exponents. This is particularly perplexing given the fact that they often exploit the same datasets (5, 20, 38–40) and are treated similarly in most modeling frameworks (6, 41).

In this paper, we test the hypothesis that previously observed spatial dependency captures a convolution of geographical propensity and a popularity-based heterogeneity among locations, by exploiting three large-scale mobile phone datasets from different countries across two continents (see *Datasets* for more details). By separating these two factors, we discovered a scaling relationship linking the critical exponents associated with the spatial effect on movement and communication patterns, effectively reducing the number of independent parameters characterizing human behavior. The uncovered scaling theory not only allows us to derive human movements from communication volumes, or vice versa, it also hints for a deeper connection that may exist among all networked systems where space plays a role, from transportations (2, 6, 42) and communications (27, 29, 30) to the internet (32, 33) and human brains (43).

Results

Mobile communication records, cataloged by mobile phone carriers for billing purposes, provide an extensive proxy of human movements and social interactions at a societal scale. Indeed, by keeping track of each phone call between two users and the spatiotemporal information about the user who initiated the call, mobile phone data offer information on both human mobility and social communication patterns at the same time. In this study, we compiled a uniquely rich database consisting of three

Significance

Both our mobility and communication patterns obey spatial constraints: Most of the time, our trips or communications occur over a short distance, and occasionally, we take longer trips or call a friend who lives far away. These spatial dependencies, best described as power laws, play a consequential role in broad areas ranging from how an epidemic spreads to diffusion of ideas and information. Here we established the first formal link, to our knowledge, between mobility and communication patterns by deriving a scaling relationship connecting them. The uncovered scaling theory not only allows us to derive human movements from communication volumes, or vice versa, but it also documents a new degree of regularity that helps deepen our quantitative understanding of human behavior.

Author contributions: P.D., C.S., N.E., V.D.B., A.-L.B., and D.W. designed research; P.D., C.S., and D.W. performed research; P.D., C.S., and D.W. analyzed data; and P.D., C.S., and D.W. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission. A.D. is a guest editor invited by the Editorial Board.

Freely available online through the PNAS open access option.

[1]To whom correspondence should be addressed. Email: dashunwang@gmail.com.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1525443113/-/DCSupplemental.

different datasets that are of a similar level of detail yet with different demographics, economic status, and scales. The resulting data corpus includes $D1$, which contains 1.3 million users in Portugal and covers a period of 1 mo; $D2$, which is a dataset from an unnamed western European country that covers a 1-y period for about 6 million users; and $D3$, which is collected by the largest mobile phone carrier in Africa, covering a period of 4 y in Rwanda.

To quantify the spatial effect on social communication patterns, we measure the distance distribution of communications using two frequently used distance metrics.

Communication Distance Distribution. The distance r characterizing social communications is the geodesic distance between two individuals u and v, who communicate via phone calls or short message service (SMS). Previous studies suggested that the probability for two individuals to communicate decreases with distance, following a power law distribution (20, 29, 44). Here we recovered previous results (Fig. 1*A*), finding that the distance distribution of each studied system, $P^S(r)$, can be approximated by a power law tail:

$$P^S(r) \sim r^{-\beta_i^r}. \qquad [1]$$

We find the exponents β_i^r to be similar for $D1$ and $D3$ ($\beta_i^r \approx 1.5$) and slightly different for $D2$ ($\beta_i^r \approx 1.35$) (Fig. 1*A* and Table 1).

Rank Distribution. Within a country, the populations are not distributed uniformly in space. To account for such inhomogeneity, previous studies proposed the rank measure as an alternative to quantifying the effective distance between two individuals (27). The rank between two users u and v is the number of people closer to u than v, formally defined as $r' = |w : r(u,w) < r(u,v)|$. We measure the rank distributions for our three datasets (Fig. 1*B*), finding $P^S(r')$ is characterized by a power law tail, consistent with previous studies (20, 27):

$$P^S(r') \sim r'^{-\beta_i^{r'}}. \qquad [2]$$

The exponents $\beta_i^{r'}$ for our three datasets are shown in Table 1.

Similarly, for mobility patterns, the jump size distribution is most commonly used to quantify spatial constraints in human movements. Here we measure this quantity in different distance metrics.

Jump Size Distribution. Jump size measures the displacement in the unit of kilometers between two consecutive sightings of an individual. A fundamental property of human mobility is that the aggregated jump size distribution follows a power law (1, 2, 4),

$$P^M(r) \sim r^{-\alpha_i^r}, \qquad [3]$$

indicating most of the time people travel over short distances, between home and work for example, whereas they occasionally take longer trips. We measured $P^M(r)$ in our data corpus (Fig. 1*C*), finding few variations in α_i^r between datasets $D2$ and $D3$ ($\alpha_i^r \approx 1.75$ and 1.8) but slight differences for dataset $D1$ ($\alpha_i^r \approx 2.02$).

Rank Jump Size Distribution. To account for biases from population density we measure the rank r' of each jump. We find that $P^M(r')$ is also characterized by a power law tail as suggested by previous studies (20, 39),

$$P^M(r') \sim r'^{-\alpha_i^{r'}}. \qquad [4]$$

As shown in Fig. 1*D*, $\alpha_i^{r'}$ is rather similar for $D1$ and $D2$ ($\alpha_i^{r'} \approx 1.22$ and 1.28) but different from $D3$: $\alpha_i^{r'} \approx 1$ (Table 1).

Taken together, the spatial scaling of social interactions [$P^S(r)$ and $P^S(r')$] for dataset i is characterized by exponents β_i^r and $\beta_i^{r'}$, respectively, whereas human movements [$P^M(r)$ and $P^M(r')$] are characterized by exponents α_i^r and $\alpha_i^{r'}$. These quantities were reported previously by independent research groups with different measurement details (1, 2, 29, 44). Here we measure these quantities systematically by using a comprehensive database we compiled. We find that within each of the two categories, the critical exponents (α_i or β_i) in different countries are rather similar to each other. For example, there is little difference between the three α_i^r or β_i^r exponents. For the rank metrics, $D1$ and $D2$ are also very similar to each other, whereas $D3$ is characterized by slightly different exponents. However, most noticeably, we observed substantial and systematic differences between $\alpha_i^{r,r'}$ and $\beta_i^{r,r'}$. Such differences contradict current modeling frameworks from gravity model (45) to radiation model (6) that treat these two classes of problems as similar phenomena given the same population distribution, thus predicting the same scaling exponent within each country. This raises a critical question: What is the origin of the observed differences between exponents α_i and β_i?

$P^S(r')$ [or $P^S(r)$] measures the intensity of social communications as a function of distance, capturing on a population-averaged level the social fluxes between different locations. On the other hand, $P^M(r')$ [or $P^M(r)$] measures the aggregated jumps between places, corresponding to the mobility fluxes from one location to another. Denoting with $T_{i,j}^S$ the social fluxes from location i to j and with $T_{i,j}^M$ the mobility fluxes representing the total number of communications ($T_{i,j}^S$) and jumps ($T_{i,j}^M$) between two locations, we measure $T_{i,j}^S$ and $T_{i,j}^M$ between any two locations over a 1-mo period. We find that both social and mobility fluxes follow fat-tailed

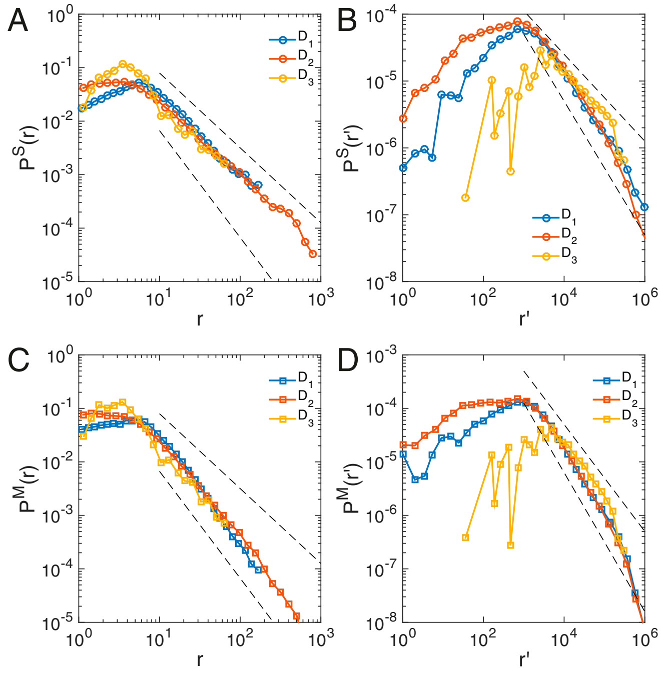

Fig. 1. Communication and jump size distance distributions. (*A*) Communication distance distributions measured in geodesic distance r, $P^S(r)$, for all three datasets. Here, r measures the distance between two users when they communicate with each other via either phone calls or SMS. r is measured in the unit of kilometers. (*B*) Rank distributions $P^S(r')$ for the three datasets follow a power law tail with exponents $\beta^{r'} = 0.89$ for $D1$, $\beta^{r'} = 1.00$ for $D2$, and $\beta^{r'} = 0.64$ for $D3$. (*C*) Jump size distribution $P^M(r)$ measured in geodesic distance r follows a power law distribution. (*D*) Rank jump size distribution $P^M(r')$ for rank r' follows a power law distribution with exponent $\alpha^{r'}$ between 1.2 and 1.3 for $D1$ and $D2$ and $\alpha^{r'} \approx 1$ for $D3$. Here we mainly focus on large r (or r') regime, fitting the tail part of the distributions. For fat-tailed distributions such as power law distributions, the tail part is the most important, determining the convergence/divergence of moments of distributions. The small r (or r') regime before the peak is often referred to as small value saturations. Dashed lines serve as guide to the eye.

Table 1. Critical exponents

Dataset	$\alpha_{r'}$	$\beta_{r'}$	$\theta_{r'}$	$\delta_{r'}$	$\widetilde{\beta_{r'}}$	α_r	β_r	θ_r	δ_r	$\widetilde{\beta_r}$
*D*1	$1.22_{\pm 0.03}$	$0.89_{\pm 0.04}$	$0.89_{\pm 0.02}$	$0.15_{\pm 0.02}$	$0.94_{\pm 0.04}$	$2.02_{\pm 0.08}$	$1.50_{\pm 0.06}$	$0.88_{\pm 0.02}$	$0.16_{\pm 0.04}$	$1.61_{\pm 0.09}$
*D*2	$1.28_{\pm 0.07}$	$1.00_{\pm 0.07}$	$0.94_{\pm 0.02}$	$0.16_{\pm 0.02}$	$1.04_{\pm 0.08}$	$1.75_{\pm 0.05}$	$1.35_{\pm 0.03}$	$0.92_{\pm 0.02}$	$0.17_{\pm 0.03}$	$1.44_{\pm 0.07}$
*D*3	$1.00_{\pm 0.07}$	$0.64_{\pm 0.03}$	$0.67_{\pm 0.03}$	$-0.07_{\pm 0.15}$	$0.70_{\pm 0.16}$	$1.80_{\pm 0.14}$	$1.57_{\pm 0.18}$	$0.83_{\pm 0.04}$	$0.24_{\pm 0.08}$	$1.25_{\pm 0.16}$

We measured $\alpha_{r'}$, $\beta_{r'}$, $\theta_{r'}$, and $\delta_{r'}$ independently for each dataset by using rank as distance metric. We estimate the errors in our measurements based on 95% confidence level. We then compute $\widetilde{\beta_{r'}} = \alpha_{r'}\theta_{r'} - \delta_{r'}$ using Eq. **8**. The error of $\widetilde{\beta_{r'}}$, $\sigma(\widetilde{\beta_{r'}})$, is calculated using error propagations $\sigma(\widetilde{\beta_{r'}}) = \sqrt{\theta_{r'}^2\sigma^2(\alpha_{r'}) + \alpha_{r'}^2\sigma^2(\theta_{r'}) + \sigma^2(\delta_{r'})}$. We find that $\widetilde{\beta_{r'}}$ largely agrees with $\beta_{r'}$ within uncertainties across all datasets. Similarly, we repeated the same measurements by using geodesic distance, obtaining α_r, β_r, θ_r, δ_r, and their corresponding errors, allowing us to compute $\widetilde{\beta}_r$ and its error $\sigma(\widetilde{\beta}_r)$. We find $\widetilde{\beta}_r$ also well approximates β_r. The largest deviations are observed in *D*3, which is characterized by much larger uncertainties in estimations of all exponents. This is due to its much smaller data size. Because both our data size and noninteger nature of distance metrics prevent us from using standard fitting algorithms for power laws (57), we computed all our exponents by using the least-square method.

distributions across our three studied datasets (Fig. 2). This is somewhat expected: Indeed, if we view each location as a node and fluxes as links connecting different locations, the fat-tailed distributions of fluxes are consistent with previous results on link weight distributions (46). Hence, Fig. 2 documents an inherent heterogeneity between locations: There are few fluxes between most locations, yet a nonnegligible fraction of location pairs are characterized by a large number of fluxes. The fat-tailed nature of flux distributions raises an important question: Can distance dependencies (Fig. 1) be accounted for by the observed heterogeneity in fluxes alone (Fig. 2)? To this end, we take *D*1 as an exemplary case and control for spatial effect by choosing location pairs that are of similar distances (r') and measuring the distributions for social [$P_T^S(T|r')$ in Fig. 3*A*] and mobility fluxes [$P_T^M(T|r')$ in Fig. 3*B*], respectively. We find that the fluxes follow a fat-tailed distribution within each group, indicating there still exists much heterogeneity in fluxes even among locations within similar distances. Moreover, locations that are nearby (small r') tend to have higher fluxes, corresponding to higher intensity in both communications (Fig. 3*A*) and movements (Fig. 3*B*). Indeed, the curves in Fig. 3 *A* and *B* shift to the right as r' decreases, indicating the probability for two locations to have large fluxes decays with distance. This is consistent with preceding results (Fig. 1 and Eqs. **2** and **4**) because most communications and movements are associated with short distances, accounting for the majority of the fluxes. However, as shown in Fig. 3 *A* and *B*, not all pairs of nearby locations have large fluxes. To the contrary, most of them have very small fluxes. Rather, it is a small fraction of location pairs in each distance groups, i.e., the tails of $P_T^S(T|r')$ and $P_T^M(T|r')$, that are responsible for generating the majority of fluxes. Most surprisingly, once we rescale the flux distributions with the average fluxes, $\langle T^S(r')\rangle$ or $\langle T^M(r')\rangle$, all curves shown in both Fig. 3 *A* and *B* (10 curves in total) collapse into one single curve, suggesting that a single universal flux distribution characterizes both social interactions and human movements, independent of distance (Fig. 3*C*). To be specific, this data collapse indicates that

$$P_T^{S,M}(T|r') = \left\langle T^{S,M}(r')\right\rangle^{-1} \mathcal{F}\left(T^{S,M} / \left\langle T^{S,M}(r')\right\rangle\right), \qquad [5]$$

where $\mathcal{F}(x)$ is a distance-independent function. The data collapse in Fig. 3*C* is rather remarkable. It indicates that the observed localization in social communications and human movements can be decomposed into two independent factors: one is the universal distribution $\mathcal{F}(x)$, which is distance independent, characterizing the inherent popularity-based heterogeneity among different locations. All of the distance dependencies are now encoded in the average fluxes at a given distance, i.e., $\langle T^S(r')\rangle$ for social and $\langle T^M(r')\rangle$ for mobility fluxes. We repeated our measurements using r as the distance metric, finding again an excellent data collapse (Fig. 3 *D*–*F*).

The uncovered universal function in Eq. **5** indicates that the social and mobility fluxes are important factors to characterize communication and mobility patterns, prompting us to measure correlations between the two quantities. We group location pairs (i and j) based on their distance and measure the relationship between $T^S_{i\to j}(r')$ and $T^M_{i\to j}(r')$ for each group ($r' = 1e3$, $r' = 1e4$, $r' = 5e5$, $r' = 1e6$, and $r' = 2e6$ in Fig. 4 *A*–*E*). In these scatterplots, each gray dot represents a pair of locations, and its x–y coordinates correspond to the mobility [$T^M_{i\to j}(r')$] and social [$T^S_{i\to j}(r')$] fluxes from i to j. We find strong correlations between these two quantities regardless of the separation between these locations. To quantify this correlation, we measure the average social fluxes given the mobility fluxes at a certain distance, $\overline{T^S}(T^M|r')$ (colored symbols in Fig. 4 *A*–*E*), which is formally defined as

$$\overline{T^S}(T^M|r') \equiv \frac{\sum_{i\to j} T^S_{i\to j}\,\delta\left(T - T^M_{i\to j}\right)\delta\left(r' - r'_{ij}\right)}{\sum_{i\to j}\delta\left(T - T^M_{i\to j}\right)\delta\left(r' - r'_{ij}\right)}, \qquad [6]$$

where $\delta(x)$ is the delta function [$\delta(x) = 1$ when $x = 0$, and $\delta(x) = 0$ otherwise]. We find that the average social fluxes $\overline{T^S}(T^M|r')$ follow a power law scaling relationship with T^M, i.e.

$$\overline{T^S}(T^M|r') = A(r')T^M(r')^{\theta_{r'}}, \qquad [7]$$

where the scaling exponents for different r', $\theta_{r'} < 1$, indicating social fluxes scale sublinearly with mobility fluxes, independent of distance. The prefactor in Eq. **7**, $A(r')$, corresponds to the shift along the y axis through Fig. 4 *A*–*E*. We find that as distance increases, the average social fluxes increase given the same

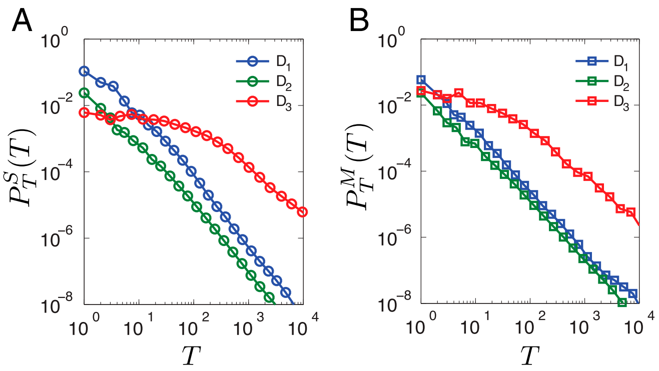

Fig. 2. Flux distributions. Fat-tailed distributions of (*A*) social fluxes T^S and (*B*) mobility fluxes T^M for all three datasets. The fluxes $T^S_{i,j}$ (or $T^M_{i,j}$) are defined as the total number of communications (or jumps) between two locations i and j. The term fat-tailed refers to distributions $p(x)$ whose decay at large x is slower than exponential.

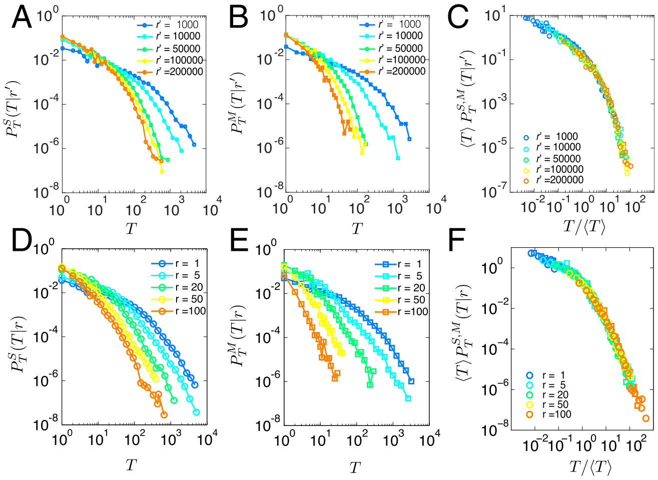

Fig. 3. Flux distributions for different rank and distance groups. (*A*) Flux distributions of social communications for different rank groups, $P^S(T|r')$. (*B*) Same distributions as *A* for mobility fluxes, $P^M(T|r')$. (*C*) Mobility and communication fluxes (ten curves), denoted by circles and squares, respectively, collapse into one single curve after rescaled by the average fluxes in each group $\langle T\rangle$. (*D*) Flux distributions of communications for different distance groups, $P^S(T|r)$ (same as *A* but measured in geodesic distance *r*). (*E*) Same distributions as *D* for mobility fluxes, $P^M(T|r')$. (*F*) Mobility and communication fluxes measured in geodesic distance (ten curves), denoted by circles and squares, respectively, again collapse into one single curve after rescaled by $\langle T\rangle$ for the different geodesic distance groups.

volume of mobility fluxes. Hence, $A(r')$ characterizes the cost tradeoff between phone communications and human movement. Rescaling $\overline{T^S}$ by $r'^{\delta_{r'}}$, we find all curves collapse into a straight line (Fig. 4*F*), indicating that $A(r')\sim r'^{\delta_{r'}}$, where $\delta_{r'}=0.15$. We repeated the same measurement for *D*2 and *D*3, finding that although each dataset is characterized by a different set of $\theta_{r'}$ and $\delta_{r'}$, Eq. **7** holds consistently well across different datasets (Fig. 4 *G* and *H*). We also repeated our analysis by replacing r' with other distance metrics (geodesic distance *r*), finding again consistent results with Eq. **7**. Indeed, each dataset is well described by its characteristic set of θ_r and δ_r exponents, demonstrating the robustness of our findings (*Correlation Between Social and Mobility Fluxes with Geodesic Distance*).

Most important, Eq. **7** together with the data collapses in Fig. 3 *C* and *F* (Eq. **5**) allows us to derive a new scaling relationship,

$$\beta_{r'}=\alpha_{r'}\theta_{r'}-\delta_{r'}, \qquad [8]$$

connecting the exponent that characterizes social communications ($\beta_{r'}$) with the exponent characterizing human movements ($\alpha_{r'}$) (see *Derivation of the Scaling Relationship Between Exponents* for details). Similarly, for geodesic distance metric *r*, we obtain

$$\beta_r=\alpha_r\theta_r-\delta_r. \qquad [9]$$

We measure each exponent in Eqs. **8** and **9** independently for each dataset, finding excellent agreement between the empirical measurements and our theoretical predictions (Table 1). Hence, Eqs. **8** and **9** offer an explicit link between critical exponents characterizing spatial dependencies in human movements and social interactions, showing that the social exponent (β) can be expressed in terms of the mobility exponents (α), a consistently robust result that is independent of the distance metrics used. The uncovered scaling relationship between these two classes of exponents is mediated by a universal flux distribution $[\mathcal{F}(x)]$ we uncovered in this study. This scaling relationship bridges two fields that are traditionally disjoint (12, 20), showing that they represent different facets of a deeper underlying reality, effectively reducing the number of independent parameters characterizing human behavior.

The uncovered relationship offers a powerful framework to derive quantities pertaining to one field from those of the other. Next, we show one practical application in public health domain as an exemplary case. Over the past few years, many computational studies highlighted the importance of social data to tackle public health challenges (10, 47). Among them, epidemic spreading is perhaps one of the most prominent (48–51). To this end, we simulate a virus spreading process using *D*1 to demonstrate how our findings can be used to connect human mobility and social interactions. Of the many ingredients in computational modeling of virus spreading, human mobility is among the most critical (1, 22, 23, 51, 52, 53). To understand how human movements catalyze societal-wide spreading processes, we infect a few randomly selected individuals with some hypothetical germ in a random location at time $t=0$. Denoting with μ the infection rate of this germ, we assume, at each time step, that an infected individual could spread the disease to others within his/her vicinity, i.e., individuals within the same mobile tower. At the same time, any infected individual can recover from the disease at rate ν. This process is known as the susceptible–infectious–susceptible (SIS) model, commonly used in modeling disease spreading (54, 55).

Choosing any set of μ and ν, we can simulate a spatial SIS model by following the real mobility fluxes between locations

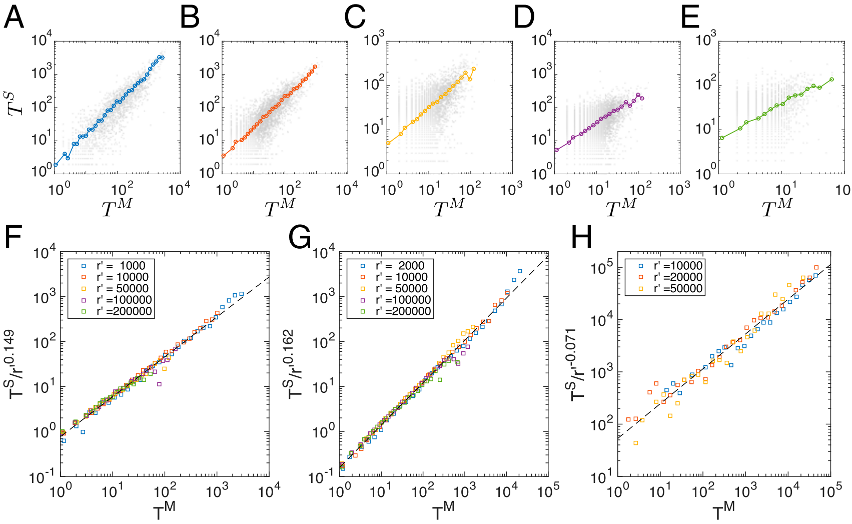

Fig. 4. Correlations between social and mobility fluxes. Correlations between $T^S_{i\to j}(r')$ and $T^M_{i\to j}(r')$ for location pairs (gray dots) separated by a distance of (*A*) $r'=1e3$, (*B*) $r'=1e4$, (*C*) $r'=5e5$, (*D*) $r'=1e6$, and (*E*) $r'=2e6$. We find all curves collapse into a straight line when $\overline{T^S}$ is rescaled by $r'^{\delta_{r'}}$ for (*F*) *D*1, (*G*) *D*2, and (*H*) *D*3.

($T^M_{i,j}$) measured from our dataset (see *Supporting Information* for modeling details). This raises an interesting question: had we not had access to mobility information, how well could we have approximated the observed spreading pattern using social fluxes rescaled by the predicted scaling (Eq. **8**)?

Following Eq. **7** and using the exponents from Eq. **8**, mobility fluxes between a location i and j can be approximated by rescaled social fluxes, $\tilde{T}^S_{i,j}$, defined as $\tilde{T}^S_{i,j} = (r'^{-\delta_{r'}}_{i,j} T^S_{i,j})^{-\theta_{r'}}$, where $\delta_{r'} = 0.15$ and $\theta_{r'} = 0.89$ for $D1$ (Table 1) and $r'_{i,j}$ is the distance between the two locations. We simulate a spreading process in Portugal using the real mobility fluxes T^M and the rescaled social fluxes $\tilde{T}^S$ as well as the mobility fluxes T^M_{GM}, approximated by the widely used gravity model (20, 45) (see *Determination of Gravity Law's Parameters* for more details). To compare these results, we started from the same initial conditions ($\mu = 0.9$, $\nu = 0.3$) and initial infected users located in a major city (Lisbon) for all simulations in this example.

We measure the density of infected users estimated in each location i for the three cases (Fig. 5 *A*–*C*), finding a remarkable agreement between our simulation and the real spreading patterns. Moreover, close up on the city of Porto reveals a superior accuracy of our model comparing with predictions from gravity model. We quantify the differences between the two methods (Fig. 5 *D* and *E*). The drastic difference between Fig. 5 *D* and *E* highlights the superior predictive power of our model.

To systematically assess and compare the accuracy of our results, we simulated 500 independent spreading processes following the same procedure described above but choosing randomly μ and ν parameters as well as the initial infected location and the number of infected users. We find that errors obtained from the 500 simulations are systematically lower than estimations from gravity model across all stages of the spreading processes (Fig. 5*F*), demonstrating the practical relevance of our scaling relationship, effectively predicting mobility patterns using social communication records.

The practical applications are most useful when only one of the two facets of information is available. For example, companies that provide social networking functionalities or services have exploded over the past few years. For many of them, mobility information is essential but difficult to collect. Conversely, companies providing location services such as global positioning system (GPS) have many mobility records yet typically lack social information. For both cases, our method may provide a reasonable estimate to fill the void, which is particularly promising given the fact that it outperforms gravity model, the prevailing framework to predict movements. Therefore, in many cases, our method may serve as a viable alternative, working in unison with or in certain cases even replacing model-based approaches, improving the predictive accuracy of most of the phenomena affected by mobility and transport processes. It could be particularly useful for developing countries where many people still live in data-scarce environments.

Taken together, by analyzing three large-scale mobile phone datasets from three different countries, we uncovered a new scaling relationship between the critical exponents that characterize spatial dependencies in human mobility and social interactions. This scaling relationship is mediated by a universal flux distribution for both movement and communication patterns, indicating the previously observed distance dependencies capture a convolution of geographical propensity and a popularity-based heterogeneity among locations. Separating these two factors allows us to establish a formal connection between different critical exponents that were perceived as independent. Together, our results document a new order of regularity that helps deepen our quantitative understanding of human behavior. Last, our results may reach far beyond communications and transportations studied in this paper because many networked systems are also subject to spatial costs in establishing connections in a very similar fashion as our quoted examples, from routers linked by physical cables to form globally connected internet to axons that connect different regions of human brains. Hence, our results may provide relevant insights to a diverse set of networked systems where space plays a role (20), opening up a promising direction for future investigation.

Finally, our study is not without limitations. Indeed, although the critical exponents we studied here capture macroscopic patterns of mobility and social interactions, both processes are affected by various sociodemographic factors both within and across countries, resulting in population variations that may not be captured adequately by power law exponents alone. It would be fruitful to analyze the degree to which such information affects mobility and social interactions, when more sociodemographic information becomes

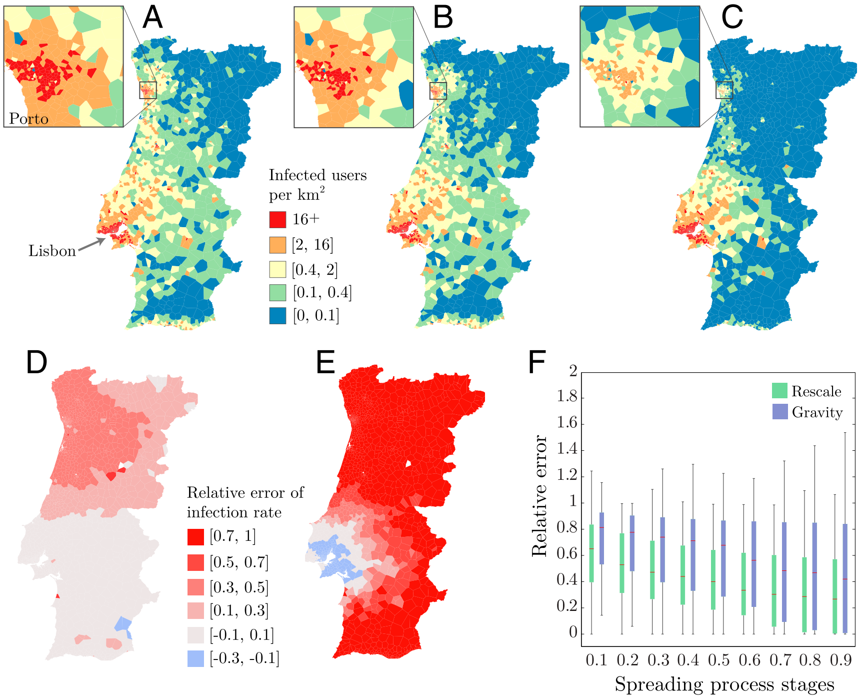

Fig. 5. Epidemic spreading simulations in Portugal. Densities of infected users at time $t = 17$ following a simulation of an SIS spreading process ($\mu = 0.9$, $\nu = 0.3$) originated from Lisbon by using (*A*) real mobility fluxes T^M, (*B*) the rescaled social fluxes $\tilde{T}^S$, and (*C*) the mobility fluxes approximated by gravity model T^M_{gm}. (*D*) Relative errors of infection rate $\tilde{e}_i(t)$ (Eq. S19) and (*E*) $\tilde{e}^{gm}_i(t)$ (Eq. S20) at each location i at time $t = 17$. (*F*) Distributions of mean relative error $\bar{\tilde{e}}(t)$ (green) and $\overline{e^{gm}}(t)$ (blue) over 500 SIS simulations at different stages before reaching the steady state, documenting the superior predictive power of our method comparing with gravity model at all stages of the spreading processes.

available. Among our three datasets, $D3$ seems to be an outlier, having different critical exponents than $D1$ and $D2$. Such information would also help us uncover deeper reasons behind variations across different countries. Furthermore, although our datasets capture people and their interactions, the focus of our paper is on data rather than people. Indeed, the virtue of our results lies in the uncovered statistical regularities revealed by our datasets. As such, our paper focuses on facts that can be measured from the data rather than deeper sociological reasons behind these observations. Last, to what degree are movements and social interactions estimated from mobile phone datasets representative? Although studies that compare self-report surveys and observational data (56) together with results obtained using higher-resolution traces (12) offer additional, convincing assurance that our results are not affected by the peculiarities of call detail records used in our study (*Potential Limitations of Mobile Phone Datasets*), we need further studies to test these assumptions in a more systematic manner.

Materials and Methods

Details of studied datasets are described in *Datasets*. Mathematical derivations of the scaling relationships in Eqs. **8** and **9** are summarized in *Derivation of the Scaling Relationship Between Exponents*. The same measurements as Figs. 2 and 3 obtained by using $D2$ and $D3$ are shown in *Distribution of Social and Mobility Fluxes for D2 and D3*. Data necessary to replicate results of this study ($D1$, $D2$, and $D3$) are available upon request. The use of mobile phone datasets for research purposes was approved by the Northeastern University Institutional Review Board. Informed consent was not necessary because research was based on previously collected anonymous datasets.

ACKNOWLEDGMENTS. This work was supported by the College of Information, Sciences, and Technology at Pennsylvania State University; the Network Science Collaborative Technology Alliance sponsored by the US Army Research Laboratory under Agreement W911NF-09-2-0053; and the Defense Threat Reduction Agency Awards WMD BRBAA07-J-2-0035 and BRBAA08-Per4-C-2-0033. P.D. is supported by the National Fund for Scientific Research (FNRS) and by the Research Department of the Communauté francaise de Belgique (Large Graph Concerted Research Action).

1. Brockmann D, Hufnagel L, Geisel T (2006) The scaling laws of human travel. *Nature* 439(7075):462–465.
2. González MC, Hidalgo CA, Barabási AL (2008) Understanding individual human mobility patterns. *Nature* 453(7196):779–782.
3. Song C, Qu Z, Blumm N, Barabási AL (2010) Limits of predictability in human mobility. *Science* 327(5968):1018–1021.
4. Song C, Koren T, Wang P, Barabási A (2010) Modelling the scaling properties of human mobility. *Nat Phys* 6(10):818–823.
5. Wang D, Pedreschi D, Song C, Giannotti F, Barabasi A-L (2011) Human mobility, social ties, and link prediction. *Proceedings of the 17th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (Association for Computing Machinery, New York), pp 1100–1108.
6. Simini F, González MC, Maritan A, Barabási A-L (2012) A universal model for mobility and migration patterns. *Nature* 484(7392):96–100.
7. de Montjoye Y-A, Hidalgo CA, Verleysen M, Blondel VD (2013) Unique in the Crowd: The privacy bounds of human mobility. *Sci Rep* 3:1376.
8. Watts DJ (2004) *Six Degrees: The Science of a Connected Age* (WW Norton, New York).
9. Barabási A-L (2002) *Linked: The New Science of Networks* (Perseus, Cambridge, MA).
10. Lazer D, et al. (2009) Life in the network: The coming age of computational social science. *Science* 323:721.
11. Blondel VD, et al. (2012) Data for development: The d4d challenge on mobile phone data. arXiv:1210.0137.
12. Blondel VD, Decuyper A, Krings G (2015) A survey of results on mobile phone datasets analysis. arXiv:1502.03406.
13. Wasserman S, Faust K (1994) *Social Network Analysis: Methods and Applications* (Cambridge Univ Press, Cambridge, UK), Vol 8.
14. Milgram S (1967) The small world problem. *Psychol Today* 1(1):61–67.
15. Borgatti SP, Mehra A, Brass DJ, Labianca G (2009) Network analysis in the social sciences. *Science* 323(5916):892–895.
16. Granovetter MS (1973) The strength of weak ties. *Am J Sociol* 78(6):1360–1380.
17. Coleman JS (1988) Social capital in the creation of human capital. *Am J Sociol* 94:(Suppl): S95–S120.
18. Lin N, Cook K, Burt RS (2001) Social capital: Theory and research. *Sociology and Economics: Controversy and Integration Series.* (Aldine de Gruyter, New York), pp 31–56.
19. Fukuyama F (1996) *Trust: The Social Virtues and the Creation of Prosperity* (Free Press, New York), Vol 457.
20. Barthélemy M (2011) Spatial networks. *Phys Rep* 499(1):1–101.
21. Colizza V, Barrat A, Barthélemy M, Vespignani A (2006) The role of the airline transportation network in the prediction and predictability of global epidemics. *Proc Natl Acad Sci USA* 103(7):2015–2020.
22. Colizza V, Pastor-Satorras R, Vespignani A (2007) Reaction–diffusion processes and metapopulation models in heterogeneous networks. *Nat Phys* 3(4):276–282.
23. Balcan D, et al. (2009) Multiscale mobility networks and the spatial spreading of infectious diseases. *Proc Natl Acad Sci USA* 106(51):21484–21489.
24. Bagrow JP, Wang D, Barabási A-L (2011) Collective response of human populations to large-scale emergencies. *PLoS One* 6(3):e17680.
25. Lu X, Bengtsson L, Holme P (2012) Predictability of population displacement after the 2010 Haiti earthquake. *Proc Natl Acad Sci USA* 109(29):11576–11581.
26. Gao L, et al. (2014) Quantifying information flow during emergencies. *Sci Rep* 4:3997.
27. Liben-Nowell D, Novak J, Kumar R, Raghavan P, Tomkins A (2005) Geographic routing in social networks. *Proc Natl Acad Sci USA* 102(33):11623–11628.
28. Wong LH, Pattison P, Robins G (2006) A spatial model for social networks. *Physica A* 360(1):99–120.
29. Lambiotte R, et al. (2008) Geographical dispersal of mobile communication networks. *Physica A* 387(21):5317–5325.
30. Scellato S, Noulas A, Lambiotte R, Mascolo C (2011) Socio-spatial properties of online location-based social networks. *Proceedings of the Fifth International Conference on Weblogs and Social Media* (Association for Advancement of Artificial Intelligence, Menlo Park, CA), pp 329–336.
31. Kleinberg JM (2000) Navigation in a small world. *Nature* 406(6798):845.
32. Boguna M, Krioukov D, Claffy KC (2008) Navigability of complex networks. *Nat Phys* 5(1):74–80.
33. Boguñá M, Papadopoulos F, Krioukov D (2010) Sustaining the Internet with hyperbolic mapping. *Nat Commun* 1:62.
34. Adamic LA, Lukose RM, Puniyani AR, Huberman BA (2001) Search in power-law networks. *Phys Rev E Stat Nonlin Soft Matter Phys* 64(4 Pt 2):046135.
35. Adamic L, Adar E (2005) How to search a social network. *Soc Networks* 27(3):187–203.
36. Wang D, et al. (2011) Information spreading in context. *Proceedings of the 20th International Conference on World Wide Web* (Association for Computing Machinery, New York), pp 735–744.
37. Rogers EM (2010) *Diffusion of Innovations* (Simon and Schuster, New York).
38. Cho E, Myers SA, Leskovec J (2011) Friendship and mobility: User movement in location-based social networks. *Proceedings of the 17th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (Association for Computing Machinery, New York), pp 1082–1090.
39. Noulas A, Scellato S, Lambiotte R, Pontil M, Mascolo C (2012) A tale of many cities: universal patterns in human urban mobility. *PLoS One* 7(5):e37027.
40. Toole JL, Herrera-Yaqüe C, Schneider CM, González MC (2015) Coupling human mobility and social ties. *J R Soc Interface* 12(105):20141128.
41. Zipf GK (1946) The P_1 P_2/D hypothesis: On the intercity movement of persons. *Am Sociol Rev* 11(6):677–686.
42. Rodrigue J-P, Comtois C, Slack B (2013) *The Geography of Transport Systems* (Routledge, New York).
43. Bullmore E, Sporns O (2009) Complex brain networks: Graph theoretical analysis of structural and functional systems. *Nat Rev Neurosci* 10(3):186–198.
44. Krings G, Calabrese F, Ratti C, Blondel VD (2009) Urban gravity: A model for inter-city telecommunication flows. *J Stat Mech* 2009(7):L07003.
45. Erlander S, Stewart NF (1990) *The Gravity Model in Transportation Analysis: Theory and Extensions* (VSP, Zeist, The Netherlands), Vol 3.
46. Barrat A, Barthélemy M, Pastor-Satorras R, Vespignani A (2004) The architecture of complex weighted networks. *Proc Natl Acad Sci USA* 101(11):3747–3752.
47. Giles J (2012) Computational social science: Making the links. *Nature* 488(7412):448–450.
48. Team WER, et al.; WHO Ebola Response Team (2014) Ebola virus disease in West Africa–the first 9 months of the epidemic and forward projections. *N Engl J Med* 371(16):1481–1495.
49. Gomes MFC, et al. (2014) Assessing the international spreading risk associated with the 2014 West African Ebola outbreak. *PLoS Currents Outbreaks* September 2, 2014, Edition 1.
50. Shaman J, Karspeck A, Yang W, Tamerius J, Lipsitch M (2013) Real-time influenza forecasts during the 2012-2013 season. *Nat Commun* 4:2837.
51. Pastor-Satorras R, Castellano C, Van Mieghem P, Vespignani A (2014) Epidemic processes in complex networks. arXiv:1408.2701.
52. Wang P, González MC, Hidalgo CA, Barabási A-L (2009) Understanding the spreading patterns of mobile phone viruses. *Science* 324(5930):1071–1076.
53. Brockmann D, Helbing D (2013) The hidden geometry of complex, network-driven contagion phenomena. *Science* 342(6164):1337–1342.
54. Pastor-Satorras R, Vespignani A (2001) Epidemic spreading in scale-free networks. *Phys Rev Lett* 86(14):3200–3203.
55. Boguñá M, Pastor-Satorras R (2002) Epidemic spreading in correlated complex networks. *Phys Rev E Stat Nonlin Soft Matter Phys* 66(4 Pt 2):047104.
56. Eagle N, Pentland AS, Lazer D (2009) Inferring friendship network structure by using mobile phone data. *Proc Natl Acad Sci USA* 106(36):15274–15278.
57. Clauset A, Shalizi C, Newman M (2007) Power-law distributions in empirical data. arXiv: 0706.1062.
58. Kardar M (2007) *Statistical Physics of Fields* (Cambridge Univ Press, Cambridge, UK).
59. Viboud C, et al. (2006) Synchrony, waves, and spatial hierarchies in the spread of influenza. *Science* 312(5772):447–451.
60. Hufnagel L, Brockmann D, Geisel T (2004) Forecast and control of epidemics in a globalized world. *Proc Natl Acad Sci USA* 101(42):15124–15129.
61. Anderson RM, May RM (1991) *Infectious Diseases of Humans* (Oxford Univ Press, Oxford, UK), Vol 1.
62. Heesterbeek J (2000) *Mathematical Epidemiology of Infectious Diseases: Model Building, Analysis and Interpretation* (Wiley, Chichester, UK), Vol 5.

Supporting Information

Deville et al. 10.1073/pnas.1525443113

Datasets

Mobile communication records, cataloged by mobile phone carriers for billing purposes, provide an extensive proxy of human movements and social interactions at a societal scale. Indeed, by keeping track of each phone call between two users and the spatiotemporal information about the users who initiated and received the call, mobile phone data offer information on both human mobility and social communication patterns at the same time, as we will detail hereunder.

In this project, we compiled a uniquely rich database consisting of three different datasets that are of a similar level of detail yet with different demographics, economic status, and scales:

Dataset $D1$ contains mobile phone calls between 1.3 million users over a period of 1 mo in 2006 from a European country (Portugal). For each phone call, the caller and the callee, both anonymized with a key (hash code); the time; the date; and the phone towers routing the communication are recorded. Only phone calls between users that called each other at least 5 times over a period of 18 mo are known. Furthermore, only the coordinates of the mobile phone towers are known; hence, the position of a user within the range of an antenna is unknown.

Dataset $D2$ covers a 6-mo period of mobile phone calls between 6 million anonymized users from a large European country. For each phone call, the caller, the callee, the time, and the towers routing the communication are recorded. Similarly to $D1$, only the coordinates of the mobile towers are known; hence, the position of a user within the range of an antenna is unknown.

Dataset $D3$ covers a period from 2005 to 2009 and is made of all transaction logs of all mobile phone activity that occurred in an African country (Rwanda) over the 5-y period. The data originate from the largest mobile phone operator in that country and contain about 1.5 million phone calls. The logs include the date, the time, and the mobile phone towers routing the call for each of the phone calls and are again anonymous. Again, only the coordinates of the mobile towers are known; hence, the position of a user within the range of an antenna is unknown.

Inferring Mobility and Social Fluxes

Mobility Fluxes. For each phone call, the position of the tower routing the call is known for the caller. Because we know the location of each tower, we know the location of the user was within the range of the tower's service area. By looking at each consecutive phone call made by a user, we can thus reconstruct the user's jumps between two consecutive locations where his calls were initiated. By aggregating all movements for all users, we can thus obtain the total number of jumps from any tower i to any tower j ($T^M_{i,j}$). All jumps made outside continental territories (i.e., islands) were not taken into account. The jumps do not exceed $\sim 1{,}000$ km, ~ 400 km, and ~ 100 km for datasets $D1$, $D2$, and $D3$, respectively, due to national frontiers and coverage limitations driven by geographical constraints in the country. We consider the number of jumps between two locations as the mobility fluxes between them.

Social Fluxes. For each phone call, the position of the tower routing the call is known for both the caller and the callee. By considering all phone calls, we thus know the total number of calls from a tower i to a tower j ($T^S_{i,j}$). We consider the number of phone calls between two locations as the social fluxes between them.

Jump Size Distribution at Fixed Interevent Time

It is known that the distribution of interevent times between two consecutive calls (locations) from the same user is heterogeneous (2). It is thus important to investigate if the observed displacement statistics (the jumps) are affected by this characteristic.

To simulate location traces left by a phone on a regular basis (instead of those due to calls) using data available to us, we calculate location displacement between a fixed time interval instead of two consecutive phone calls. More specifically, we use our dataset $D1$ and calculate the jump size distribution $P^M(r)$ for displacements separated by a time $\Delta T \pm 0.05\Delta T$. We systematically vary ΔT from 1 h to 1 d (Fig. S1). We find the distributions collapse for different choices of ΔT, suggesting that the use of consecutive calls serves as a good proxy for movements. Also, the curves can be well approximated by a power law consistent with our previous results. Note our results are bounded by the maximum distance a user can travel during ΔT, thus explaining the differences in the tail part of the distribution.

Distribution of Social and Mobility Fluxes for *D*2 and *D*3

In Figs. S2 and S3, we present results obtained for the datasets $D2$ and $D3$ regarding the distributions of the social fluxes $P^S_T(T|r')$ and $P^S_T(T|r)$ (Fig. 3 *A* and *D* for $D1$) and mobility fluxes $P^M_T(T|r')$ and $P^M_T(T|r)$ (Fig. 3 *B* and *E* for $D1$) for pairs of locations that are of similar distances (r and r'). We also show how flux distributions collapse for both datasets when they are rescaled with their average flux, $\langle T(r)\rangle$ or $\langle T(r')\rangle$ (Fig. 3 *C* and *F*). The same procedure for the dataset $D1$ is applied to $D2$ (Fig. S2) and $D3$ (Fig. S3). We again find that the fluxes for each group still follow a fat-tailed distribution, indicating there also exists much heterogeneity in fluxes among locations within similar distances for both $D2$ and $D3$. Locations that are nearby (small r' or r) tend to have higher fluxes, corresponding to higher intensity in both communications (Figs. S2 *A* and *D* and S3 *A* and *D*) and movements (Figs. S2 *B* and *E* and S3 *B* and *E*), corroborating our results for $D1$. Indeed, the curves shift to the right as r' (or r) decreases, indicating the probability for faraway location pairs to have large fluxes is much lower. Once we rescale the flux distributions with the average flux, $\langle T(r')\rangle$ or $\langle T(r)\rangle$, we find all of the curves collapse into a single curve, demonstrating again a single universal flux distribution characterizes both social communication and human movement fluxes, independent of distance (Figs. S2 *C* and *F* and S3 *C* and *F*).

Correlation Between Social and Mobility Fluxes with Geodesic Distance

As developed in the manuscript for the rank-based distance, we here analyze the correlations between the social fluxes $T^S_{i\to j}(r)$ and mobility fluxes $T^M_{i\to j}(r)$ in the case of the geodesic distance. We group location pairs (i and j) based on their distance and measure the relationship between $T^S_{i\to j}(r)$ and $T^M_{i\to j}(r)$ for each group ($r=1$, $r=5$, $r=10$, $r=50$, and $r=100$ in Fig. S4 *A*–*E*). In these scatterplots, each gray dot represents a pair of locations, and its x–y coordinates correspond to the mobility $[T^M_{i\to j}(r)]$ and social $[T^S_{i\to j}(r)]$ fluxes from i to j for dataset $D1$.

Same as for the rank-based measures, we find again strong correlations between these two quantities regardless of how far away these locations are separated. To quantify this correlation, we measure the average social fluxes given the mobility fluxes at a certain distance, $\overline{T^S}(T^M|r)$ (colored symbols in Fig. S4 *A*–*E*), which is formally defined as

$$\overline{T^S}\left(T^M|r\right) \equiv \frac{\sum_{i\to j} T^S_{i\to j}\delta\left(T - T^M_{i\to j}\right)\delta\left(r - r_{ij}\right)}{\sum_{i\to j}\delta\left(T - T^M_{i\to j}\right)\delta\left(r - r_{ij}\right)}, \qquad \textbf{[S1]}$$

where $\delta(x)$ is the delta function [$\delta(x) = 1$ when $x = 0$, and $\delta(x) = 0$ otherwise]. We find that the average social fluxes $\overline{T^S}(T^M|r)$ have again a power law scaling relationship with T^M, following

$$\overline{T^S}\left(T^M|r\right) = A(r)T^M(r)^{\theta_r}, \qquad \textbf{[S2]}$$

where the scaling exponent $\theta_r < 1$ for different r, indicating social fluxes again scale sublinearly with mobility fluxes. The prefactor in Eq. **S2**, $A(r)$, corresponds to the shift along the y axis through Fig. S4 *A–E*. We find, as distance increases, the average social fluxes increase given the same volume of mobility fluxes. Rescaling $\overline{T^S}$ by r^{δ_r}, we find all curves collapse into a straight line (Fig. S4*F*), indicating $A(r) \sim r^{\delta_r}$.We repeated the same measurement for $D2$ and $D3$. We found that although each dataset is again characterized by a different set of θ_r and δ_r, Eq. **S2** (same as Eq. **7** in the main manuscript) holds consistently well across different datasets (Fig. S4 *F–H*), demonstrating the robustness of our findings for both the distance r and r'.

Derivation of the Scaling Relationship Between Exponents

As stated in the main manuscript, we find that the average social fluxes $\overline{T^S}(T^M|r')$ follow a power law scaling relationship with T^M, i.e.,

$$\overline{T^S}\left(T^M|r\right) = A(r')T^M(r')^{\theta_{r'}}, \qquad \textbf{[S3]}$$

where $\theta_{r'} < 1$, indicating social fluxes scale sublinearly with mobility fluxes, independent of distance. As described in *Correlation Between Social and Mobility Fluxes with Geodesic Distance*, a similar result is obtained for geodesic distance metric r (Eq. **S2**).

The data collapses observed in Fig. 3 *C* and *F*, i.e.,

$$P^{S,M}_T(T|r') = \langle T^{S,M}(r')\rangle^{-1}\mathcal{F}\left(T^{S,M}/\langle T^{S,M}(r')\rangle\right), \qquad \textbf{[S4]}$$

together with Eq. **S3** allow us to derive a new scaling relationship between different critical exponents. Indeed, the average social fluxes at distance r', $\overline{T^S}(r')$, can be obtained by integrating $\overline{T^S}(T^M, r')$ over T^M:

$$\overline{T^S}(r') = \int P^M_T\left(T^M|r'\right)\overline{T^S}\left(T^M, r'\right)dT^M. \qquad \textbf{[S5]}$$

Substituting Eqs. **S4** and **S3** into Eq. **S5**, we have

$$\overline{T^S}(r') = \int \mathcal{F}(x)\overline{T^S_M}\left(\overline{T^M}(r')x, r'\right)dx \sim \overline{T^M}(r')^{\theta_{r'}}r'^{\delta_{r'}}\int x^{\theta_{r'}}\mathcal{F}(x)dx, \qquad \textbf{[S6]}$$

where $x \equiv T^M/\overline{T^M}$ as a change of variable. As $\overline{T^S}(r') \sim \sum_{i\to j} T^S_{i\to j}\delta(r' - r'_{ij}) = P^S(r') \sim r'^{-\beta_{r'}}$, and similarly $\overline{T^M}(r') \sim r'^{-\alpha_{r'}}$, we have

$$r'^{-\beta_{r'}} = r'^{-\alpha_{r'}\theta_{r'}}r'^{\delta_{r'}}\int x^{\theta_{r'}}\mathcal{F}(x)dx. \qquad \textbf{[S7]}$$

The tail behavior of $\mathcal{F}(x)$ indicates the integral in Eq. **S7** converges. Hence, Eq. **S7** leads to a scaling relationship,

$$\beta_{r'} = \alpha_{r'}\theta_{r'} - \delta_{r'}, \qquad \textbf{[S8]}$$

connecting the exponent that characterizes social communications ($\beta_{r'}$) with the exponent characterizing human movements ($\alpha_{r'}$). Similarly, for geodesic distance metric r, we obtain

$$\beta_r = \alpha_r\theta_r - \delta_r. \qquad \textbf{[S9]}$$

The scaling analyses performed here have their roots in the canonical statistical physics literature, namely, the scaling identities in phase transitions and critical phenomenon. The power law scaling behavior in the vicinity of a continuous transition is captured by a set of critical exponents $(\alpha, \beta, \gamma, \delta, \sigma, \eta, \ldots)$, characterizing various fundamental quantities such as free energy, specific heat, magnetization, susceptibility, etc. In the beginning, these critical exponents were measured independently and found to vary slightly across different materials. Later, we witnessed a burst of results demonstrating that these critical exponents are not independent but are in fact connected through what we now call scaling identities. The famous examples include Rushbrooke's identity, Widom's identity, Josephson's identity, and Fisher's identity (58).

Determination of Gravity Law's Parameters

The gravity law assumes that the mobility fluxes between a locations i of origin and a location j of destination can be expressed as a function of the two populations at the two locations (m_i and m_j) and the geodesic distance between them ($r_{i,j}$) as

$$T^M_{GM,i,j} = C\frac{m_i^{\mu}m_j^{\kappa}}{f\left(r_{i,j}\right)}, \qquad \textbf{[S10]}$$

where $f(r) = r^{\gamma}$ (6, 20, 45, 59). By taking the logarithm on both sides we obtain

$$log\left(T^M_{GM,i,j}\right) = log(C) + \mu log(m_i) + \kappa log\left(m_j\right) - \gamma log\left(r_{i,j}\right). \qquad \textbf{[S11]}$$

Using the observed mobility fluxes (T^M), we can then estimate the parameters through a least square regression, giving us $[log(C), \mu, \kappa, \gamma] = [-3.42, 0.67, 0.68, 1.32]$.

Epidemic Spreading Simulations

To compare the accuracy and usefulness of our rescaling formula, we simulated an SIS process commonly used in modeling disease spreading (54, 55) by following the observed mobility fluxes T^M and the rescaled social fluxes $\tilde{T}^S$ but also the mobility fluxes T^M_{gm} approximated by the well-known gravity model (20, 45).

We consider the process where each location i (mobile tower) is characterized by a constant population size N_i, equal to the number of distinct users present in the vicinity of the mobile tower over the period covered by the dataset $D1$. The total population in our system is thus given by $\sum_{i=1}^m N_i$, and the system equilibrated as the population is constant. In each location, users are classified according to their infectious state: they can be either infectious (I) or susceptible to be infected (S). The standard generalization of this spatial SIS model is given by

$$S_i + I_i \xrightarrow{\mu} I_i \qquad \textbf{[S12]}$$

$$I_i \xrightarrow{\nu} S_i \qquad \textbf{[S13]}$$

$$S_i \xrightarrow{A_{i,j}} S_j \qquad \textbf{[S14]}$$

$$I_i \xrightarrow{A_{i,j}} I_j, \qquad \textbf{[S15]}$$

where reaction **S5** indicates that susceptible users can become infectious at a rate μ and reaction **S6** corresponds to infected users recovering from the disease at a rate ν. In addition to the standard SIS dynamics, susceptible as well as infected users can randomly move between one location i to another location j as described in reactions **S7** and **S8**. The probability rate of these movements from location i to j is governed by the probability rate $A_{i,j}$ defined as

$$A_{i,j} = \frac{\left(T_{i,j}T_{j,i}\right)^{-2}}{N_i}. \qquad \textbf{[S16]}$$

Because the system is equilibrated, the flux of users from i to j must balance that of j to i (detailed balance condition):

$$A_{i,j}N_i = A_{j,i}N_j, \qquad \textbf{[S17]}$$

which is fulfilled by Eq. **S18**.

In this case, the spatial SIS model can be defined as a set of m coupled ordinary differential equations (ODEs) for the infected people in each location (22, 60):

$$\partial_t I_i = \mu \frac{I_i}{N_i}(N_i - I_i) - \nu I_i + \sum_{j \neq i}\left[A_{j,i}I_j - A_{i,j}I_i\right], \qquad \textbf{[S18]}$$

enabling us to compute the evolution of infected users in each location over time by solving these.

Denoting with n_i the number of users at location i, with a_i the area of location i and $m_i(t)$, $\tilde{m}_i(t)$, and $m_i^{GM}(t)$ the number of infected users at time t in location i when using T^M, $\tilde{T}^S$, and T^M_{GM}, respectively, we measure $m_i(t)/a_i$, $\tilde{m}_i(t)/a_i$, and $m_i^{GM}(t)/a_i$, i.e., the density of infected users estimated in each location i for the three cases (Fig. 5 *A–C* for $t = 17$). We find a remarkable agreement between our simulation and the real spreading patterns. Moreover, close up on the city of Porto reveals a superior accuracy of our model comparing with predictions from gravity model. To quantify the differences between the two methods, we measure

$$\tilde{e}_i(t) = \frac{m_{i,t} - \tilde{m}_{i,t}}{m_{i,t}} \qquad \textbf{[S19]}$$

and

$$e_i^{GM}(t) = \frac{m_{i,t} - m_{i,t}^{GM}}{m_{i,t}}, \qquad \textbf{[S20]}$$

corresponding to the relative error of infection rate in each location i at time t for both methods (Fig. 5 *D* and *E* at $t = 17$). The drastic difference between Fig. 5 *D* and *E* highlights the fact that lower $\tilde{e}_i(t)$ are observed comparing with $e_i^{GM}(t)$ in this particular example, again documenting the superior predictive power of our model.

To systematically assess and compare the accuracy of our results, we simulated 500 independent spreading processes following the same procedure described above but choosing randomly μ and ν parameters as well as the initial infected location and the number of infected users. For each simulation, we compute the mean values $\overline{\tilde{e}}(t)$ and $\overline{e^{GM}}(t)$ from Eqs. **S19** and **S20**, respectively, at different stages (time steps). We find that $\overline{\tilde{e}}(t)$ obtained from the 500 simulations are systematically lower than $\overline{e^{GM}}(t)$ across all stages of the spreading processes (Fig. 5*F*), demonstrating the practical relevance of our scaling relationship, effectively predicting mobility patterns using social communication records.

Normalizing the Time Steps of the Spreading Processes

In this section, we describe the procedure we use to compare spatial spreading processes whose initial conditions differs.

As formulated in Eq. **S18**, each spatial process is characterized by a set of m coupled ODEs. Each ODE corresponds to a spreading subprocess within a location, and each one of them reaches the steady-state after a different number of time steps (61, 62). Here we consider the global process to be at equilibrium when no more changes are observed for any of its subprocesses, i.e., $\max_i \partial_t I_i / N_i < 10^{-5}$.

As described in the main manuscript, we simulated 500 spatial spreading processes, each with parameters μ, $\underline{\nu}$, initial infected location, and initial number of infected users chosen randomly. Each process i will thus reach the equilibrium at a different time t_i^c. To compare their accuracy at different stages of the process, we normalize the time steps t_i of each process i by its time before equilibrium, i.e., t_i/t_i^c. As a result, a time step of $t_i/t_i^c = 0.5$ for any process i would correspond to half the time it takes to reach the equilibrium. This normalization is used in Fig. 5*F* to compare processes at similar stages.

Potential Limitations of Mobile Phone Datasets

For studies on mobility and social interactions, the mobile phone dataset is the most relevant dataset that is currently in existence. Indeed, at present, the most detailed information on human mobility across a large segment of the population is collected by mobile phone carriers. Mobile carriers record the closest mobile tower each time the user uses his or her phone. Other possible data sources include dollar bills, GPS, or check-in datasets from location-based social networking services, all of which suffer from well-known limitations that are resolved by mobile phone datasets. Indeed, dollar bills are carried by various individuals; hence, mobility inferred from them captures population-level aggregated movements instead of individual mobility. GPS tracks individual positions on a continuous basis with high precision, but it operates on a much smaller scale (typically hundreds of people) in contrast to millions of individuals' mobile phone data records. Check-in datasets only record mobility information when users report their positions voluntarily on subset of population who use the service, in contrast to mobile phones that objectively collect mobility information across a societal-scale population. For this reason, research on human mobility has literally exploded following the availability of mobile phone datasets, resulting in a number of rather fundamental papers. Furthermore, mobile phone datasets offer comprehensive information on phone calls and text messages, providing social network information in addition to mobility trajectories of each individual. Therefore, mobile phone datasets are excellent data sources to study simultaneously human mobility and social networks.

However, mobile phone datasets have a number of well-known limitations. Most notably, there are three aspects:

First, as mobile phones approximate a user's location by the tower that routed the call, the spatial resolution of the dataset is limited by the area covered by each tower, which typically ranges between 1 and 3 km. This is a spatial limitation of the data. Luckily, earlier research has extensively focused on this issue and documented that, at least for results we discussed, the results are not affected by this limitation.

Second, a user's position is only recorded when he or she makes a call or sends a text. However, human communications follow bursty patterns. This is the temporal limitation of the data. However, there are ample reasons to believe that our results are not affected by it. Mobility studies that compare mobility patterns obtained through mobile phone data and other continuous tracing technologies consistently find that the two are largely indistinguishable (2). These include GPS traces (2) as well as high-resolution mobile phone records (4, 5). Although we do not have direct access to these datasets, using our own datasets, we further calculated location displacement between a fixed time interval instead of two consecutive phone calls, in doing so artificially creating mobility traces that occur on a continuous basis. We find that results are remarkably stable as we vary the time interval systematically from 1 h to 1 d (*Jump Size Distribution at Fixed Interevent Time*). All these results suggest that although mobility information is obtained from calling patterns in mobile phone datasets, call detailed record (CDR) data offer representative patterns of mobility, offering convincing reassurance that our results are not affected by this limitation.

Third, social network information is inferred based on calling patterns, yet calls using mobile phones can be ambiguous and hence may not represent true social relationships. This is the third limitation of the data. However, results by Eagle et al. (56) compared self-report survey data on mobility and social interactions with observational data obtained using mobile phones, demonstrating a high degree of accuracy (95%) in inferring friendship structures based on observational data alone.

Taken together, mobile phone datasets are the best and largest datasets for the type of study we conducted here. Although they have well-known limitations, extensive studies and results have demonstrated that our study is not affected by these limitations.

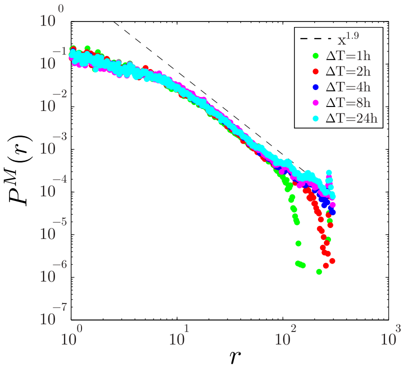

Fig. S1. Jump size distribution $P^M(r)$ for interevent time $\Delta T = 1,\ 2,\ 4,\ 8,$ and 24 h for the $D1$ dataset. A power law with exponent $r^{-1.9}$ provides a guide to the eye. We observe that the curves are bounded by the maximum distance a user can travel during their corresponding interevent time for $\Delta T < 4$ h or the maximum distance enforce by geographical constraints in that country.

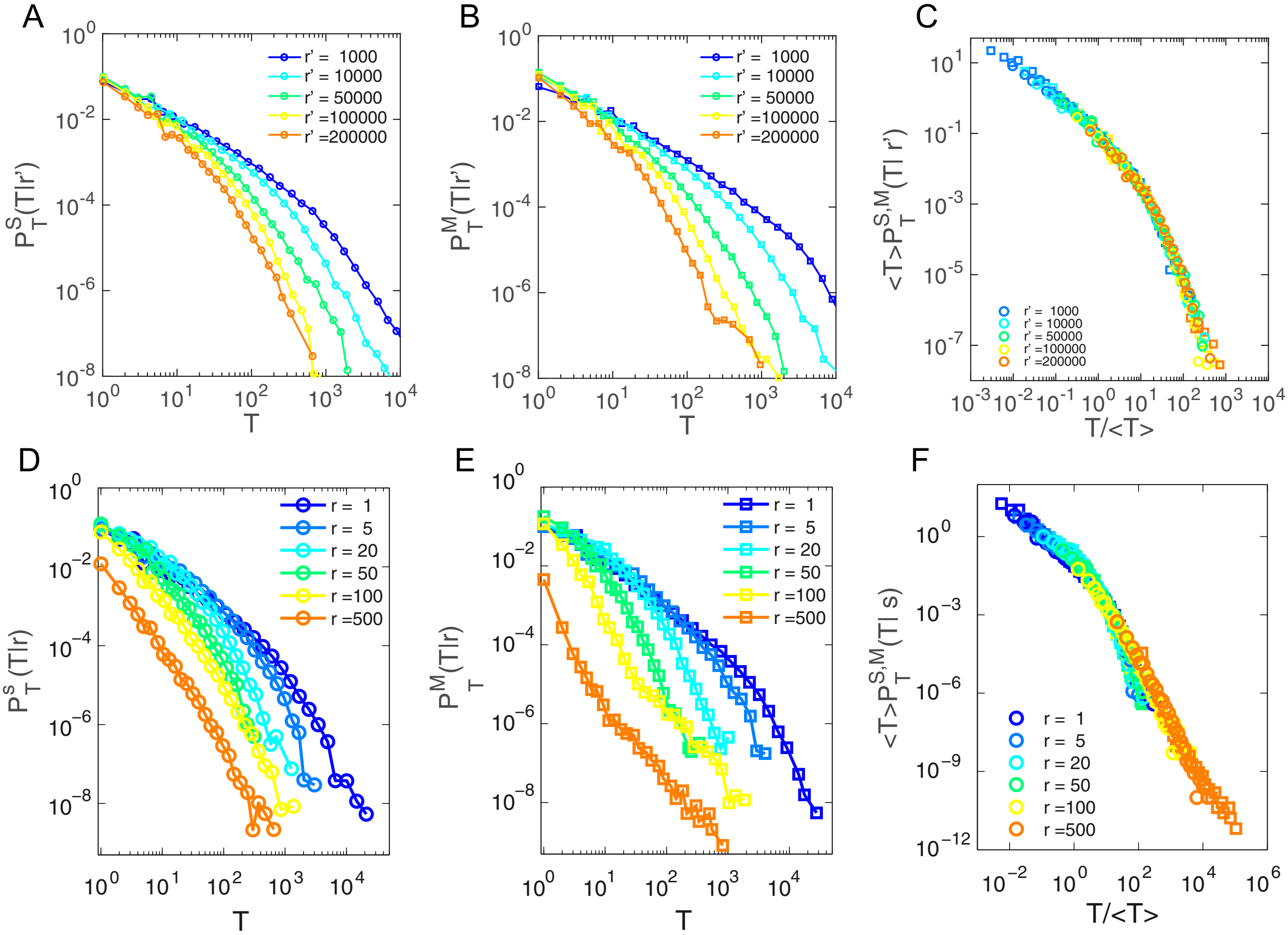

Fig. S2. (*A*) Communication fluxes $P^S(T|r')$ conditional to the rank r' for dataset *D*2. (*B*) Mobility fluxes $P^M(T|r')$ conditional to the rank r' for dataset *D*2. (*C*) Mobility and communication fluxes, denoted by circles and squares, respectively, collapse after rescaled by $\langle T\rangle$ for dataset *D*2 for the rank-based distance. (*D*) Communication fluxes $P^S(T|r')$ conditional to the distance *r* for dataset *D*2. (*E*) Mobility fluxes $P^M(T|r)$ conditional to the distance *r* for dataset *D*2. (*F*) Mobility and communication fluxes, denoted by circles and squares, respectively, collapse after rescaled by $\langle T\rangle$ for dataset *D*2 for the geodesic distance.

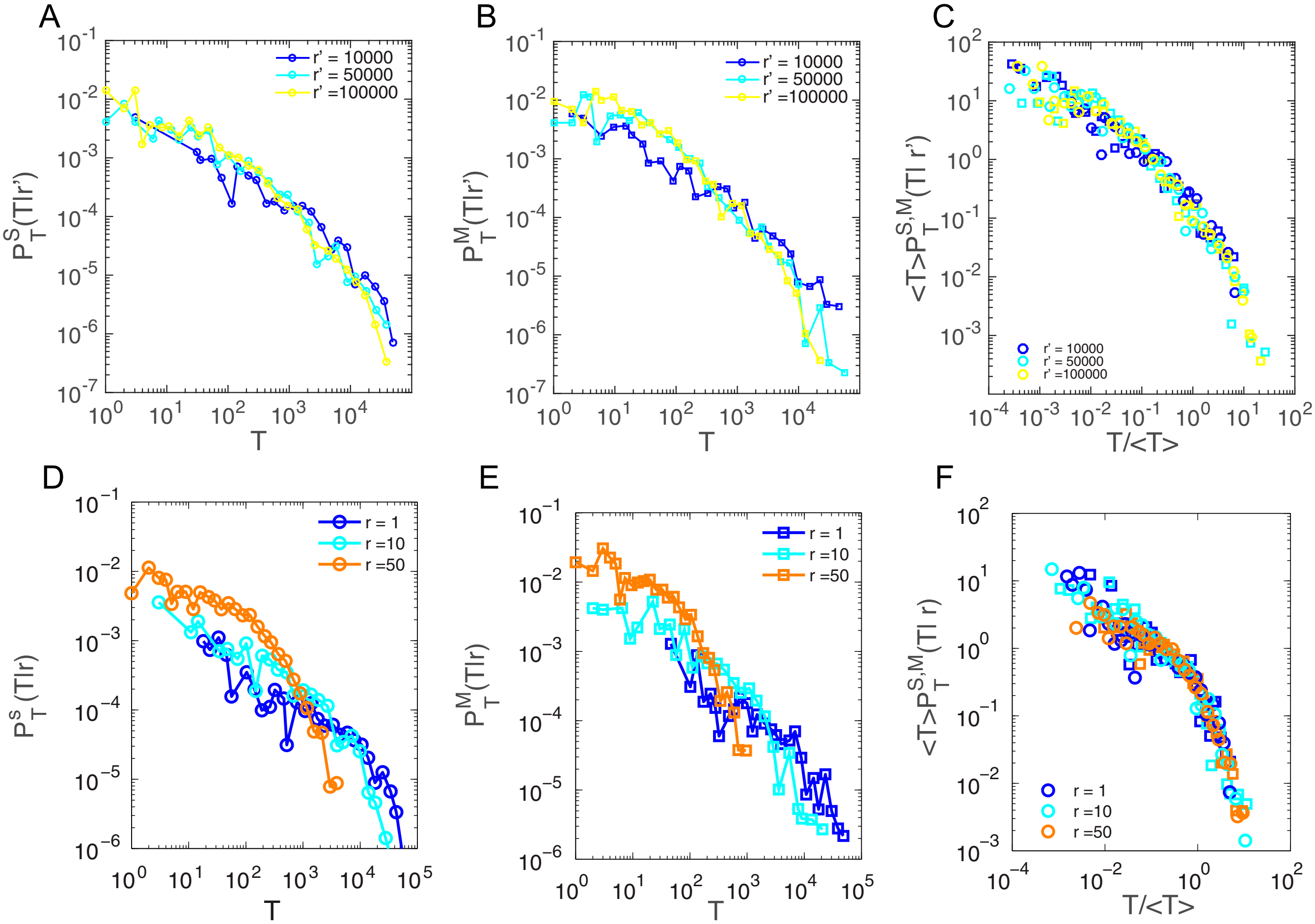

Fig. S3. (*A*) Communication fluxes $P^S(T|r')$ conditional to the rank r' for dataset *D*3. (*B*) Mobility fluxes $P^M(T|r')$ conditional to the rank r' for dataset *D*3. (*C*) Mobility and communication fluxes, denoted by circles and squares, respectively, collapse after rescaled by $\langle T \rangle$ for dataset *D*3 for the rank-based distance. (*D*) Communication fluxes $P^S(T|r)$ conditional to the distance *r* for dataset *D*3. (*E*) Mobility fluxes $P^M(T|r)$ conditional to the distance *r* for dataset *D*3. (*F*) Mobility and communication fluxes, denoted by circles and squares, respectively, collapse after rescaled by $\langle T \rangle$ for dataset *D*3 for the geodesic distance.

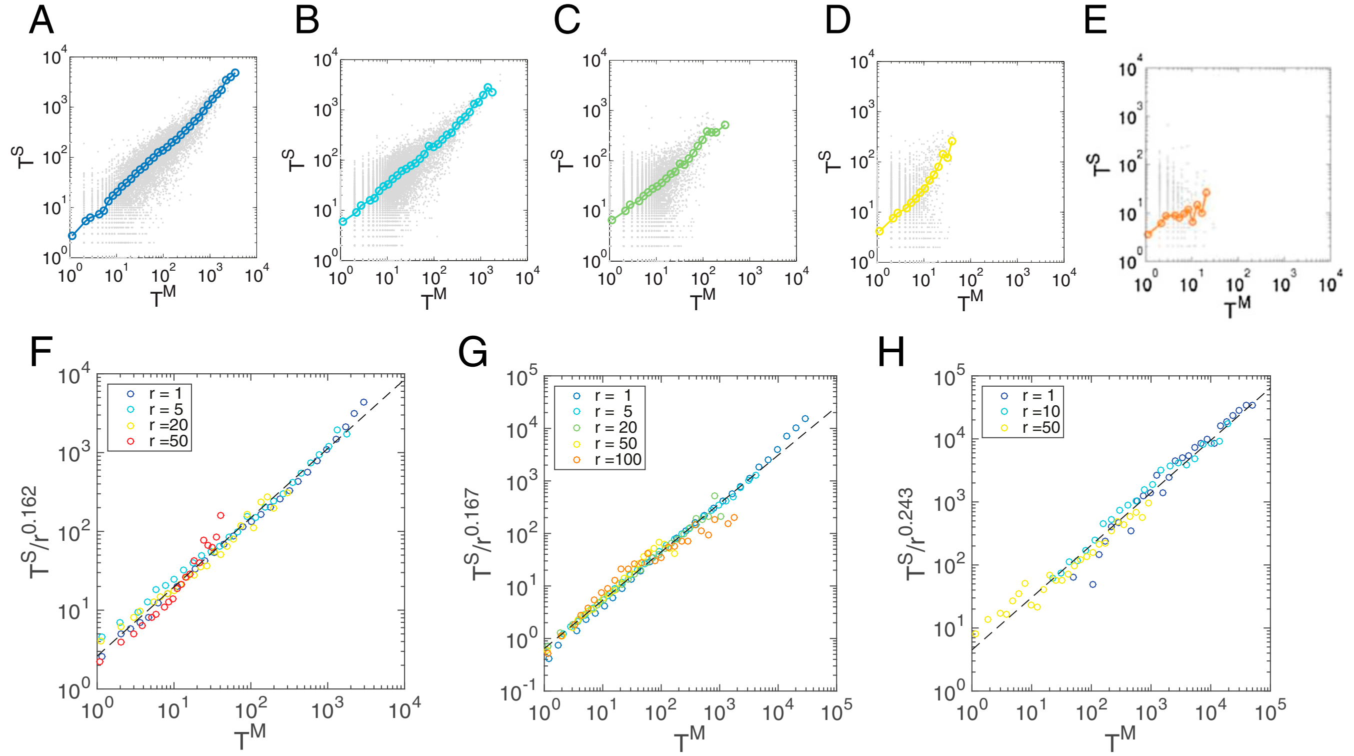

Fig. S4. Correlation between $T^S_{i\to j}(r)$ and $T^M_{i\to j}(r)$ for pair of locations (gray dots) separated by a distance of (*A*) $r = 1$ km, (*B*) $r = 5$ km, (*C*) $r = 20$ km, (*D*) $r = 50$ km, and (*E*) $r = 100$ km for *D*1. We find all curves collapse into a straight line when $\overline{T^S}$ is rescaled by r^{δ_r} for (*F*) *D*1, (*G*) *D*2, and (*H*) *D*3.