Supplementary Materials

1 Data description

Our data contains phone records for a six months period in three countries: France, Portugal, and Spain. In total 7 billion interactions are considered. In order to build the social network, only links with at least one communication per direction are included. This is a common technique in the literature [29, 28, 20] to avoid both marketing callers and misdialed numbers. After applying this filter, the network characteristics shown in table S1 are observed.

Regarding degree distribution, our three networks present the common heavy-tail distribution found in previous works [29, 20] with mobile phone data, which could be fitted to a high exponent power-law ($\alpha > 4$) rather than to an exponential function. Degree distributions for all three networks are shown in figure S1. For the scope of this study, the existence of hubs (nodes with very high number of connections) in all three networks is important.

1.1 User location

A key aspect in the creation of a link between two individuals is the geographical distance between them. In our study, users are located in their billing zip code (Spain) or their most used tower (France and Portugal). In total 8928 different locations are available in Spain\(^1\), 17475 in France and 2209 in Portugal. Figure

\(^1\)Spain zip codes are geolocated according to geonames database, available at http://downloads.geonames.org/export/zip, and grouped according to latitude and longitude since some zip codes have identical coordinates. Towers coordinates were provided by the carrier.

<table>
<thead>
<tr>
<th>Country</th>
<th>% GC</th>
<th>Nodes $N$</th>
<th>Links $E$</th>
<th>$\langle k \rangle$</th>
<th>$\langle c \rangle$</th>
<th>$\langle c_r \rangle$</th>
<th>$\langle l \rangle$</th>
<th>$\langle l_r \rangle$</th>
</tr>
</thead>
<tbody>
<tr>
<td>France</td>
<td>99.23</td>
<td>$18.7 \cdot 10^6$</td>
<td>$81.3 \cdot 10^6$</td>
<td>8.73</td>
<td>0.16</td>
<td>$9 \cdot 10^{-7}$</td>
<td>8.52</td>
<td>7.75</td>
</tr>
<tr>
<td>Portugal</td>
<td>96.23</td>
<td>$1.21 \cdot 10^6$</td>
<td>$4.00 \cdot 10^6$</td>
<td>6.57</td>
<td>0.26</td>
<td>$5 \cdot 10^{-7}$</td>
<td>8.35</td>
<td>7.44</td>
</tr>
<tr>
<td>Spain</td>
<td>95.81</td>
<td>$5.92 \cdot 10^6$</td>
<td>$16.1 \cdot 10^6$</td>
<td>5.44</td>
<td>0.21</td>
<td>$48 \cdot 10^{-7}$</td>
<td>10.36</td>
<td>9.20</td>
</tr>
</tbody>
</table>

Tab. S1: Characteristic properties of the social networks in the studied countries like the size of the giant component (%GC), number of users (Nodes) and relationships (Links), average degree $\langle k \rangle$, average clustering coefficient $\langle c \rangle$, as well as the corresponding values for random networks with the same size $\langle c_r \rangle$ and $\langle l_r \rangle$. The observed values are typical for small-world networks.
Fig. S1: Degree distribution for each of the country level networks.

S2 shows the distance distribution to the first, second and third closest zip code or tower in the three datasets. Although towers may provide a slightly more accurate geolocation, both are sufficient for our purposes.

On the other hand, users are not equally distributed among towers and zip codes. Figure S3 shows the cumulative distribution in the three data sets. Most of the towers serve between 100 and a few thousands users, while zip codes’ user count is more heterogeneous (the maximum is a zip code in Madrid with 125,000 users). The explanation for these different results comes from technical reasons: as the demand rises in an area, additional phone towers need to be installed to handle the traffic.

For simplicity, from now on we will refer both to towers and zip codes as towers, unless otherwise mentioned to explain different results among different data sets.

1.2 Sampling effects

Users in the network are not homogeneously distributed, in some regions there is a slightly higher concentration. This variance may come from a higher market share of the mobile phone provider or from a higher usage of mobile phone service in the area (only users who have at least one mutual relationship appear in the network). The differences between different regions are depicted in figure S4.

We refer user density as the ratio $u_i = \frac{\text{Users}_{i}}{\text{Total population}_{i}}$ in a certain region $i$. The main effect of having different $u_i$ seems to be in the average degree of the resulting subnetwork. Figure S5 shows this relationship, which turns out to be close to linear. For a network where all inhabitants are present (i.e., $u_i = 1$), a projection of the resulting linear model would be $\langle k \rangle \simeq 16$. 
Fig. S2: Distance distribution to the first, second and third closest tower or zip codes. Towers (France and Portugal) are slightly closer to each other than zip codes (Spain) are.

Fig. S3: Empirical cumulative distribution function of number of users in each tower or zip code. Due to technical reasons, the range of users per location is smaller when towers are used, while zip codes distribution is more broad.
Fig. S4: Users/Population ratio in the province level. Brighter colors represent a higher ratio.

In any case, the number of contacts in a phone network is relatively small compared to other social networks obtained from online social sites (average degree are in the hundreds [24, 14]) or compared to different figures proposed as average degree for humans: extrapolation from observed correlation between social group size and neocortex volume in primates drove Dunbar to propose 150 [11], while recent statistical estimation methods based on self-reported data range between 290 [22] and 610 [23]. We will show that increasing the average degree has a positive effect on routing, which means any result we get by studying the phone social network can be considered as a lower bound for the real world’s social network. On the other hand, the phone network can be seen as the backbone of the social network, since it contains only interactions the people are willing to pay for.
Fig. S5: Dependence of the average degree $\langle k \rangle$ on the ratio between users and population. Each point represents a province network. It can be appreciated how closely related $\langle k \rangle$ and $u_i = \frac{\text{Users}}{\text{Population}}$ are. Blue line presents a linear fit $\langle k \rangle = 3.16 + 13.24u_i$ where $R^2 = 0.818$. 
2 Link-distance distribution

The probability of finding a social tie decreases with distance, regardless the proxy used to infer the social network: blogs [21], location based social networks [32, 9] or mobile phone data [20, 28, 19]. In all of them the probability $P(d)$ of finding a social tie between two actors who are within distance $d$ from each other decreases (at least during a certain interval) as a power law, with exponents between $-1$ and $-2$. As shown in figure S6, our data fits this behavior for all three networks. Moreover, due to the high number of links considered we are able to observe long-range peaks. The reason for these peaks is the heterogeneity in the population distribution (we observe the same peaks even if we randomize the links while keeping actors in the same location). To support this point, in the figure we highlight how the peaks match the distances between main cities.

3 Social communities’ spatial clustering

Community detection has been an active field in the latest years [13, 3, 6]. Although there are many algorithms published so far, their goal is common: identify dense areas in the social graph. In our routing experiments, we eventually use communities as a proxy for social attributes (school attended, field of work, family and so on). Communities have been reported to be closely related to
geographical distance in every possible scale: triangles (which can be defined as the simplest community) have been proven to be severely geographically driven [20], small k-clique communities in a mobile phone social graph were reported to have an unexpected number of people living in the same zip code [30] and finally, country-wide communities found using multilevel aggregation method (also known as Louvain method) have been reported to be severely affected by geography [6, 28, 12].

In the routing experiments, we have used this Louvain method. We found a fundamental difference between the behavior on urban scale and on the country one: geographical clustering turns out to be more intense in the intercity scenario than in the intricacy one. To reach this conclusion we have calculated the spatial clustering of the communities by the following steps:

• Perform a community detection on the network

• Associate the tower to the most common community among that tower’s users.

• Calculate the average distance $\langle d_{\text{com}} \rangle$ between any two towers belonging to the same community

$$\langle d_{\text{com}} \rangle = \frac{\sum_{c=1}^{C} \sum_{a=2}^{N_c} \sum_{b=1}^{a-1} d(a,b)}{\sum_{c=1}^{C} \sum_{a=2}^{N_c} (a-1)} \quad (1)$$

where $C$ denotes the number of communities found, $N_c$ the number of towers in community $c$ and $d(a,b)$ the distance between towers $a$ and $b$.

• Assign communities with the same sizes randomly to the towers and calculate the average distance (1) of the randomized data $\langle d_r \rangle$.

Figure S7 illustrates the difference between community geoclustering in both local and country level. The geographic clustering is more pronounced in the country network. These results are confirmed by the calculation of $\langle d_r \rangle$ and $\langle d_{\text{com}} \rangle$ in the networks presented in table S2.

<table>
<thead>
<tr>
<th>Network</th>
<th>$\langle d_{\text{com}} \rangle$ (km)</th>
<th>$\langle d_r \rangle$ (km)</th>
<th>$\langle d_r \rangle / \langle d_{\text{com}} \rangle$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Portugal</td>
<td>64.4</td>
<td>240.1</td>
<td>3.72</td>
</tr>
<tr>
<td>France</td>
<td>115.7</td>
<td>410.71</td>
<td>3.54</td>
</tr>
<tr>
<td>Spain</td>
<td>118.5</td>
<td>521.2</td>
<td>4.39</td>
</tr>
<tr>
<td>Lisbon (concelho)</td>
<td>3.4</td>
<td>4.31</td>
<td>1.26</td>
</tr>
<tr>
<td>Paris (department)</td>
<td>4.1</td>
<td>5.7</td>
<td>1.39</td>
</tr>
<tr>
<td>Madrid (municipio)</td>
<td>3.2</td>
<td>3.46</td>
<td>1.08</td>
</tr>
</tbody>
</table>

Tab. S2: Average distance between two towers belonging to the same community. The geographical effect $\langle d_r \rangle / \langle d_{\text{com}} \rangle$ is more pronounced in the nationwide communities.
Fig. S7: Geographical clustering of social communities. On the country scale, towers belonging to the 20 biggest communities are presented in different colours and shapes. On the city scale, towers within each capital city are presented. On the country scale most of communities fit with the administrative boundaries. This has been the motivation of a research line oriented to “redraw” the political maps according to social network features [5, 31, 8]. However, we just show how geographical bias in communities is more pronounced in country-wide networks.
4 Algorithms

In order to deliver the message, several strategies can be used. In the following we describe every criteria used in our experiments.

DFS By adding depth first search (DFS) into a routing algorithm, we effectively avoid the message to get into infinite loops. The application of DFS in the Milgram experiment is quite straightforward: when a participant receives a message, he knows the list of people who already got the message. The participant will never forward to none of these people, unless all of his friends are in the list. In this case, he will send the message back to the person who first sent the message to him. In a tree network, this would be the case of a branch which has been explored without success and the search process continues going backwards. Since our social network is far from being a tree, the number of rolling back events is low (less than $10^{-6}$ in all of our simulations).

GEO This procedure consists of sending the message to the friend geographically closest to the target. In the intercity scenario, locations are considered on the municipality level. In the intracity scenario, tower locations are employed. Note that this discretization produces a number of ties (two or more friends are at the same distance from the target).

DEG In this case, the message is forwarded to the friend with the largest number of friends among the candidates.

COM In order to mimic social attributes (school, work) communities are detected in the network. Specifically, the Louvain method [6] has been employed: each person $i$ has a community vector $c_i$ which contains identifiers of the community for different aggregation levels. Although the number of aggregation levels $L$ depends on the network, in all of our networks the algorithm provided between 3 and 7 aggregation levels. Note that this algorithm provides hierarchical communities. If two nodes $i$ and $j$ have a community of level $l$ in common they will share as well all the communities in higher levels, formally:

$$i, j \in [1, ..., N] / c_{il} = c_{jl} \rightarrow c_{ix} = c_{jx} \forall x \in [l, ..., L]$$

where $N$ is the number of people. A person will send the message to a friend with the lowest possible community level in common with the target.

In our experiments, these criteria are combined, by using several of them to solve ties: this way, we will denote $dfs\text{-}deg$ to a routing scheme where first the already visited nodes are discarded from candidates ($dfs$), and then those with the highest degree are chosen ($deg$). If there is still more than one possible friend after the routing logic is completed, the message is forwarded to one of these candidates at random. In our $dfs\text{-}deg$ example, this happens if two or more friends were not previously visited and have the same degree.
5 Intercity routing experiment

5.1 Assignation of user to cities

Our first experiment consists of, given a random pair of nodes in the network A and B, trying to deliver a message from A to the area where B lives. For systematically delimiting this "area where B lives" we have chosen administrative division over a regular spatial grid, because the resulting modularity in the social networks is significantly higher. Specifically, we will study two levels of aggregation in each of the networks:

- We will generically refer as provinces to the following administrative divisions: départements in France, provincias in Spain and distritos in Portugal. This way we divide the country into 97, 50, and 20 provinces, respectively. According to official census, the population ranges from 77 thousand (Lozère, France) to 6.4 million (Madrid, Spain). A province map for all three countries is depicted in figure S8a.

- We will generically refer as municipalities to the following administrative divisions: cantons in France, municipios in Spain and concelhos in Portugal. Our users are located in 3520, 5446 and 297 different municipalities respectively. A map depicting municipalities in all three countries is presented in figure S8b.

To map the user coordinates into the appropriate divisions we have used Global Administrative Divisions database except for France's cantons, where the GEOFLA database by IGN has been used.

5.2 Experiment conditions

Once we have assigned users to their cities, we ran the experiment in the following setup: in each country we chose 60 thousand random source and targets among all nodes in the network. Next, we try to deliver the message using combinations of techniques described in section 4. Additional to those, we have performed a pure georeedy (passing to the geographically closed friend, and if no one is closer than the current user, consider the chain broken) as well as the modification proposed in [21], which we have denoted georeedy++, and consist of forwarding the message to another user in the same location even if she is not connected to the current user. For each pair and algorithm, up to 1000 hops are simulated before reaching target’s city.

---

2 We have used this division instance of the communes because of the high number and high heterogeneity of the latter (over 36 thousand different communes, ranging from 10 people to 2 million). Most of cantons are composed of several communes, being Paris a special case: Paris city actually fills the whole department 75, and is divided into 20 arrondissements (districts) which are counted as cantons. In any case, when we refer the Paris city in the intracity network experiment, we mean department 75. Some other large French cities are also divided into several cantons.

3 http://www.gadm.org/

4 http://professionnels.ign.fr/geofla
5.3 Experiment results

First of all, in intercity routing, using provinces as target seems to make the routing process trivial (even random routing delivers a significant amount of messages), so we will present the results of the routing trying to reach the right municipality. The main conclusion is that any routing strategy other than random will deliver the messages with a high probability (as we can see in figure 2a in the main text). If we study small differences in error rate after 100 steps between the algorithms (see figure S9) we find statistically significant differences between the algorithms. In general, \textit{geo} methods outperform \textit{com} methods, and solving geographical ties (two people are at the same distance from the target) using degree increases routing performance. Another relevant finding is that these distributed routings reflects the same behavior than the optimal routing: it is harder to route in Spain (due to the smallest average degree) than in France, despite the number of nodes in France is about 4 times larger.

In order to provide a more detailed look of this experiment, we have published the following webapp: www.someurl.com. In the app, the user can pick among 180 thousand routes we have simulated, choosing first target city and then source. To illustrate the difference between distributed an optimal routing, both optimal and best decentralized (\textit{dfs-geo-deg}) routes are plotted, and also the number of nodes explored to find the optimal path is presented. Theoretically, a \textit{dfs} can go “backwards” in the exploration of the network if all friends have been already visited, producing a loop in the sequence of explored nodes. However, we did not find evidence for this in our simulations (overall, over 3.2 million hops were simulated). In figure S10 an snapshot of the app is presented, with one route as an example. On average, in France, the distributed routes found have 18.1 hops, while 7.2 hops are optimal. However, in order to find the optimal routes on average 8.1 million nodes have to be explored.

Besides, we have studied how the size of target’s city influences the length of the distributed route found. Intuitively it is easier to reach a big town like
Fig. 59: Intercity performance for different routing algorithms. Top graph shows the fraction of messages arriving at the target in the first 50 hops \( f(n) \). Since the fraction of messages decreases with the number of hops, one could evaluate the performance by measuring the mean and the integral of this distribution after \( N \) hops, with \( N \) being large enough. The bottom graph presents the average path length of the delivered messages \( \langle l \rangle = \sum_{n=1}^{N} n f(n) \) and the fraction of failed messages \( E = 1 - \sum_{n=1}^{N} f(n) \) for \( N = 100 \).
Fig. S10: Snapshot of the app we developed for visualize our results. The red route is the result of distributed dfs-com-deg while the green one displays the optimal route. In this example, the distributed route needs 17 steps to reach the destination city, while the optimal route uses 7. However, the distributed algorithm explores only 17 nodes, while more than 11 million nodes are checked for finding an optimal route.

Madrid (3.2 million inhabitants in the municipality and half million users in our network) than a small city with just a few hundred inhabitants. However, our results show how the size of the destination city affects only logarithmically to the length of the route found to reach them (see figure S11).

6 Intracity experiment

For the intracity experiment, we have divided the country networks into both provinces and municipalities networks. All provinces have been studied, and the 100 most populated municipalities in each country (300 municipalities and 168 provinces in total). Province networks are almost connected (over 95% nodes in the giant component) and municipalities have also a quite big giant component (over 80%). In any case, these administrative boundaries produce significantly larger connected networks than any regular spatial grid. The reason for this is that the classification of nodes in either provinces or municipalities is indeed a good community classification (modularity\textsuperscript{5} scores over 0.4 and 0.5 respectively) probably due to the high clustering of our networks. For the routing experiment,

\textsuperscript{5}Modularity is a standard metric to evaluate performance of community detection method, defined in [26] as $Q = \frac{1}{2m} \left( \sum_{ij} A_{ij} - k_i k_j \right) \delta(i,j)$ where $A$ is the adjacency matrix of the network, $k_i$ is the degree of vertex $i$ and $\delta(i,j) = 1$ if $i$ and $j$ belong to the same community and $\delta(i,j) = 0$ otherwise.
we take into account the nodes in the giant components (a path between any given two nodes actually exists) just as we did with the country networks.

We repeat the experiment in each of the networks with the same setup we used for the intercity experiment. In this case 100 thousand random pairs are simulated for the algorithms presented in figure 2c in the main text, while for all other algorithms, 10 thousand pairs are considered.

### 6.1 Results analysis

In figure S12 we present the routing results for the three capital cities (in fact these are worst case scenarios, since the networks are the largest). Figure S12 presents both $P(l)$ distributions and their equivalence in the $(\langle l \rangle_{100}, E_{100})$ plane, which we will use for comparison. In figures S13-S18 we include results for the top 20 provinces and municipalities in each country. Careful observation of these graphics allows us to draw the following conclusions:

- Algorithm ranking from best to worst, is almost constant over all studied networks.
- Among dfs methods (algorithms avoiding loops), $\langle l \rangle$ and $E$ are fairly correlated. If an algorithm A outperforms another algorithm B by finding smaller $\langle l \rangle$ it will also provide a smaller error rate. Thus, we can compare algorithms by using only one of the two metrics. In figure S19 we show the relation between these two metrics for the dfs-com-deg algorithm.
• Contrary to what takes place in the intercity scale, using geography to route within the city does not produce efficient routing. Consistently over the network sets we study, community based routing $dfs$-$com$-$deg$ significantly outperforms $dfs$-$geo$-$deg$. Interestingly, having additional geography information besides the community structure (this means there is more information to make the routing decision) seems to be misleading, specially in large networks, as it can be observed in the performance of the $dfs$-$com$-$geo$-$deg$ routing strategy.

• Among all algorithms tested, $dfs$-$com$-$deg$ is the one producing the best results.

6.2 Efficient routing and average degree

Networks are considered to be small-worlds if they have a high clustering coefficient (ratio between closed triangles and connected triples), and at the same time the shortest path length scales with the number of nodes in the network $N$ like $O(\log N)$ [33]. A routing algorithm is considered to be efficient if it is polylogarithmic [18]: i.e. it is able to find, between any two nodes, a path of length $O(\log^a N)$ with a high probability.

Then, we check if our $dfs$-$com$-$deg$ is in fact an efficient routing algorithm. In figure S20(top) we show the relation between network size and error rate. Although in most networks we find that the error rates depend logarithmically on the number of nodes, we see a number of outliers. We find these outliers have small average degree. In fact, the majority of networks that do not lie in the $O(\log N)$ behavior have average degree smaller than 4. Although to the best of our knowledge there is no previous result in the literature to explain this finding, we suggest the following explanation. In a random graph where all nodes have the same degree $k$, $k \geq 3$ is needed to be able to find paths $O(\log N)$[7]. On the other hand, recent work in the effect of clustering in percolation studies show how a growing transitivity implies a higher average degree is needed for the emergence of a giant component[27, 1, 4]. Since having a connected network is a necessary condition to route, we conclude our empirical observation is consistent with previous theoretical results: it is not feasible to route efficiently in networks with an average degree smaller than 4. In fact as we show in figure S20(bottom), networks with low average degree actually have a significantly larger diameter.

6.3 Relation to decentralized routing theory

As mentioned in the paper, a number of approaches have been employed in the literature to explain the capability of humans participating in Milgram-like experiments to find short paths: repetitions of the experiment asking the participants about routing criteria are performed [10, 25], computer simulation of decentralized search strategies are tested on real network data [21, 2], and analytic studies certain properties of networks are conducted [36, 18]. In this last category, lots of attention was attracted by Kleinberg’s work [18, 15] where it is
Fig. S12: Intracity experiment results for the 3 main cities. Top graph shows the fraction of messages arriving at the target in the first 50 hops $f(n)$. Since the fraction of messages decreases with the number of hops, one could evaluate the performance by measuring the mean and the integral of this distribution after $N$ hops, with $N$ being large enough. The bottom graph presents the average path length of the delivered messages $\langle l \rangle = \sum_{n=1}^{N} n f(n)$ and the fraction of failed messages $E = 1 - \sum_{n=1}^{N} f(n)$ for $N = 100$. 
Fig. S13: Intracity results for the 20 biggest provinces in Portugal. \( N \) denotes the number of nodes, and \( \langle k \rangle \) the average degree. Success rates refer to the proportion of messages delivered after 100 steps and \( \langle l \rangle \) to the average path length of successful chains.
Fig. S14: Intracity results for the 20 biggest provinces in Spain. \( N \) denotes the number of nodes, and \( \langle k \rangle \) the average degree. Success rates refer to the proportion of messages delivered after 100 steps and \( \langle l \rangle \) to the average path length of successful chains.
Fig. S15: Intracity results for the 20 biggest provinces in France. $N$ denotes the number of nodes, and $\langle k \rangle$ the average degree. Success rates refer to the proportion of messages delivered after 100 steps and $\langle l \rangle$ to the average path length of successful chains.
Fig. S16: Intracity results for the 20 biggest municipalities in Portugal. \( N \) denotes the number of nodes, and \( \langle k \rangle \) the average degree. Success rates refer to the proportion of messages delivered after 100 steps and \( \langle l \rangle \) to the average path length of successful chains.
**Fig. S17:** Intracity results for the 20 biggest municipalities in Spain. $N$ denotes the number of nodes, and $\langle k \rangle$ the average degree. Success rates refer to the proportion of messages delivered after 100 steps and $\langle l \rangle$ to the average path length of successful chains.
Fig. S18: Intracity results for the 20 biggest municipalities in France. $N$ denotes the number of nodes, and $\langle k \rangle$ the average degree. Success rates refer to the proportion of messages delivered after 100 steps and $\langle l \rangle$ to the average path length of successful chains.
Fig. S19: Correlation between the average shortest path length and the error rate for each province and municipality, using the \texttt{dfs-com-deg} routing strategy. The size of the symbols is correlated with the corresponding population.
Fig. S20: Scaling of error rate with size for the *dfs-com-deg* strategy (top). Colors represent the average degree $\langle k \rangle$. If networks are connected enough $\langle k \rangle > 4$, scaling follows a logarithmic behavior. A similar behavior emerges in the scaling of the average path length $l_{optimal}$ (bottom), where networks with low degree have a diameter relatively large for their size.
proven that a regular two dimensional lattice can obtain small world structure by adding randomly links between nodes. Additionally, only if these links are added with probability $\frac{1}{r^6}$, a decentralized algorithm is able to find these short paths. Even if this is indeed a very interesting finding, we cannot map our phone network on a two dimensional lattice with additional long-range links.

However, in [16, 17] the same author proposes a generalization which we can in fact apply, which it is called the group model. In short, let be a network whose node set is $V$, and a set of groups, $S = \{S_1, S_2, ... S_n\}$ where $S_i = \{v_1, v_2, ..., v_i \mid v_i \in V\}$ and at least one of the groups $S_i$ is the full vertex set $V$. Under these assumptions, for any pair of nodes $(u, v)$, a function $g(u, v)$ can be defined such as $g(u, v)$ is the size of the smallest group $S_i$ containing both $u$ and $v$. If a network is constructed so that $k$ edges are added to each node with probability proportional to $g^{-1}(u, v)$, then a decentralized algorithm can route in polylogarithmic time.

Both our main routing strategies, communities and geography, can be mapped to groups: it is straightforward in the case of communities since the hierarchy resulting of community detection is a valid set of groups $S$. For geography, we can consider $g(u, v)$ as the number of people who are closer from $v$ than $u$, which means $S_i$ are the balls of population centered in a tower with a given radius $r$. A similar model was actually proposed in [21] to explain how a simple greedy technique is capable of sending messages to the right city.

As it can be seen for Paris urban network in figure 2c in the paper, and in S21 for the Lisbon urban network, both geographically determined balls and communities show approximately the reverse linear behavior that theoretically guarantees search efficiency. However, we observe a significant difference be-

\[ r \text{ denotes the Manhattan distance between two given nodes in the lattice} \]

\[ \text{There are some characteristics in our networks which makes them different from the theoretical model: our networks have heterogeneous degree and we need to relax some of the properties of the groups, especially in the case of geographic balls. Concretely, the original model requires that for any group $S_i$ of size $g >= 2$ containing a node $v$, there has to be a group} \]
tween the community and geographic structures.

The explanation relies on the following fact: given a group $S$ where the target belongs (can be a geographic ball or a community), a decentralized algorithm tends to search the whole group before trying nodes in other groups. In this situation, imagine the group is not a connected component on the network, this is, there are no paths between all arbitrary pair of nodes $u, v \in S$ where all the nodes on the path are also in $S$. In this case, our decentralized search fails. In figure 2d of the paper we show the difference between geographic balls and communities: while communities are by definition connected, geographic balls lose connectivity for small radius. This means, within the same tower, there are islands of users. However, as we can see in the figure, if we calculate the giant components of the geographic balls on the country scales (locating users in municipalities) we observe no such breakdown. This finding agrees with the fact that geo strategies are actually efficient on the country scale, as discussed in the previous section.

### 6.4 Connectivity collapse within cities

As we have discussed in the previous section, given a ball of radius $r$ km, if we construct the social network between the people living in municipalities within the ball, this network will have a giant component (figure 2d in the paper with dark colors). However, if we choose a ball within a municipality, and build the network between people living within the same tower, the giant component vanishes. In figure S22 we show the reason for this collapse, by studying the intra-tower networks for the 30 top towers in each capital city and then compare to two randomized versions of the networks. The first randomization keeps average degree (Erdös-Rényi), and the second keeps the whole degree distribution, but both eliminate clustering. Our results demonstrate that clustering is the main responsible for the absence of a giant component.

In summary we have strong evidence that the observed relation between geographic space and social network (connected pieces of land produce connected networks) breaks within cities. Thus, we neither can find a distance $r_{\text{critical}}$ nor a geographical group size $S_{\text{critical}}$ below which there is no connected component in the induced subgraph, because cities have very different extension and population. To support this claim we have studied all intra-tower networks in the capital cities and compared to municipalities networks of the same size. Figure S23 shows the average giant component for towers and municipalities of a certain size. Municipalities with a given population have a larger giant component than a tower in a city with the same population.

Given a fixed number of nodes, a giant component emerges more likely with a higher number of links and with low clustering (a link closing a triangle cannot enlarge any component). As it can be seen in figures S23B and S23C, both effects are present and they can explain the different giant component sizes between $S_i \subseteq S_t$ containing $v$ which is strictly smaller than $S_t$, but contains at least $\min(\lambda g, g - 1)$ nodes, where $\lambda < 1$. To accomplish this in our case, we need to choose a $\lambda$ arbitrary small, at most $1/t_{\text{max}}$, where $t_{\text{max}}$ is the maximum number of users in one tower.
Fig. S22: For each of the top 30 towers in each capital city, the fraction of nodes in the giant component is computed. Additionally, the giant components for randomized versions Rand-ER (keeps average degree) and Rand-Degree (keeps degree distribution) are shown. Each random point in the graph is averaged over hundred realizations of the randomization process.
Fig. S23: A) shows the relation between population size and fraction of nodes in the giant component for all towers in the capital cities (blue) and municipalities in the country within the same range of population (red). Errors bars represent the standard error of the mean $\sqrt{\frac{\sigma}{n}}$. It can be clearly observed how a connectivity within municipalities tends to be higher than within towers of the same size. B) and C) depict the causes of this behavior, smaller average degree (except for Lisbon) and higher clustering are the reasons why the giant component is larger in municipalities.
municipalities and towers. However, clustering seems to be dominant, since in Portugal the average degree in the same tower and municipalities. Moreover, the small average degree does not seem to be lack of data, since the France data with the highest average degree exhibits the smallest average degree on the tower scale.

**Number of social ties within towers**

Our results in figure S6 agree with previous literature [21, 20] finding that the probability of two users within distance \( r \) to be connected decreases similar to \( \frac{1}{r^\alpha} \). However, this finding does not give us any guideline about the number of links between people within the same tower, since in principle they are within \( r = 0 \) distance. In order to be able to apply pure geographical models to our data, we have to randomize the position of the user around the tower’s location.

A common assumption for mobile phone data is considering that if a call is processed by a tower, then that tower is the closest to the user’s location. This assumption implies the geographic space can be divided according to the Voronoi diagram of the towers in that region. This way our randomization assigns a users a position uniformly distributed in the Voronoi cell they belong. Figure S24 shows the randomization process in Paris and Lisbon.\(^8\)

After randomization, the distance \( r \) between any two users is greater than zero, so we can apply \( \frac{1}{r^\alpha} \) models the number of predicted and present intra-
Fig. S25: Number of links within the same tower using several randomization models. Results are averaged over 10 runs. The real network has a bigger number of intra-tower links than a space independent graph (ER) and a $\frac{1}{2}$ model. In the case of Lisbon, the real network has even more links than a $\frac{1}{2}$ model. To explain the high number of intratower links the geographical distance is not sufficient, thus another effect like clustering is needed.

tower links for the same number of links in the whole network. In figure S25 we show how the geographic effect should be more intense than the observed between towers, in order to explain the number of intra tower links just using geography. This implies clustering plays a major effect in this level, producing highly clustered islands within the same tower.

### 6.5 Crossover in geography-based routing

Figure S26 shows the performance of different routing strategies in the intracity scenario considering that a delivery is successful if the message was able to reach the target in less than 50 steps (figure 2b in the paper is analogous to this figure but with 100 steps threshold). One interesting aspect is the crossover behavior between municipality and provinces in the geographic based routing. In this section we explain the emergence of such behavior by using a simplified example.

Figure S27 shows a simplified version of a province with $N$ users and 3 cities. Let’s denote $P(S)$ the probability that a message is successfully delivered. For dfs algorithm it is straightforward to conclude the probability $P_{dfs}(S) = \frac{1}{N}$ being $N$ the number of nodes in the network, no matter if the network represents a province or a city. This conclusion agrees with our results in figures S26 and
Fig. S26: Success rate for different routing strategies in provinces and municipalities with \((k)>4\). This figure is equivalent to figure 2b in the main paper, but considering successful routing if the message was delivered within 50 steps, instead of 100. Pure \(dfs\) routing produces a reverse linear decrease, while the community based routing produces a much slower decay. Geographical routing in the intracity scenario produces a crossover between behavior provinces and municipalities.
However, for geographic routing, we denote $P(c)$ where $c \in \{A, B, C\}$ the probability of reaching the right city $c$ and $P(S|c)$ being the probability that the message is successfully delivered given it is already in the right city $c$. In the intracity experiment scheme (see section 5) we have proven that the geo approach is valid, delivering the vast majority of the messages to the right city, so we consider $P_{geo}(c) = 1^9$. Using results from our intracity experiment we assume $P_{geo}(S|c) = \frac{1}{n^\alpha}$, with $0 < \alpha < 1$. Then

$\begin{align*}
P_{geo}(S) &= \sum_{c \in \{A,B,C\}} \frac{n_c}{N} P_{geo}(c) P_{geo}(S|c) = \\
&= \sum_{c \in \{A,B,C\}} \frac{n_c}{N} \frac{1}{n_c^{\alpha}} = \frac{\sum_{c \in \{A,B,C\}} n_c^{(1-\alpha)}}{N} \geq \frac{(\sum_c n_c^{(1-\alpha)})}{N} = \frac{1}{N^\alpha},
\end{align*}$

which means that using geo approach, a province with a certain population $N$ has a higher success rate than a municipality with the same size. Even if we generalize $P_{geo}(S|c) = f(n_c)$ where $f$ is any decreasing function this result holds: if geo is capable to deliver all messages to the right city, then $P_{geo}(S)$ is a weighted average of the performances in the cities forming the province such that $f(n_{max}) \leq P_{geo}(S) \leq f(n_{min})$ where $n_{min}$ and $n_{max}$ denote the size of the smallest and biggest cities respectively.

References


---

9 This is a fair assumption since experimental results found error rates $E < 10^{-3}$. 
Fig. S27: Simplified version of a province with population $N = 10^5$ and 3 municipalities. Routing process can be divided into 2 steps: reaching the right municipality and then finding the right target within that city. $P(A)$ denotes the probability that a message whose target is in city $A$ actually reaches $A$. $P(S|A)$ denotes the probability that a message reaches its target given it is already in $A$. Geo strategy is efficient to reach the right city so $P(A) = P(B) = P(C) = 1$ which implies the performance on the overall province is actually better than in the major city, producing the crossover observed in the results.


