U.S. Social Fragmentation at Multiple Scales

Leila Hedayatifar, Rachel A. Rigg, Yaneer Bar-Yam and Alfredo J. Morales*

New England Complex Systems Institute

277 Broadway, Cambridge, MA, 02139

* E-mail: alfredo@necsi.edu

Abstract

Despite global connectivity, societies seem to be increasingly polarized and fragmented. This phenomenon is rooted in the underlying complex structure and dynamics of social systems. Far from homogeneously mixing or adopting conforming views, individuals self-organize into groups at multiple scales, ranging from families up to cities and cultures. In this paper, we study the fragmented structure of the American society using mobility and communication networks obtained from geo-located social media data. We find self-organized patches with clear geographical borders that are consistent between physical and virtual spaces. The patches have multi-scale structure ranging from parts of a city up to the entire nation. Their significance is reflected in distinct patterns of collective interests and conversations. Finally, we explain the patch emergence by a model of network growth that combines mechanisms of geographical distance gravity, preferential attachment, and spatial growth. Our observations are consistent with the emergence of social groups whose separated association and communication reinforce distinct identities. Rather than eliminating borders, the virtual space reproduces them as people mirror their offline lives online. Understanding the mechanisms driving the emergence of fragmentation in hyper-connected social systems is imperative in the age of the Internet and globalization.

1 Introduction

The increasing polarization of societies is becoming apparent around the world. Despite access to global communication [1], people seem to be splitting into groups that mostly listen to their own members [2, 3, 4]. Individual choices of association due to ideologies [5, 6, 7], occupations [8, 9], or consumer habits [10] can drive the emergence of social polarization or fragmentation [11, 9]. While different social features affect processes of homophily and influence, in this work we study how fundamental geographical factors also affect the large-scale structure of social interactions and communication networks. Previous studies have proposed distance as the driving factor for social interactions [12, 13, 14]. We show that the structure of the emergent social networks is richer than what distance alone can explain and includes the influence of factors like administrative borders and urban structures. It is crucial to understand the structural and geographical properties of collective association and their relationship to the social space.

The social space is defined as the place where people meet and interact [15]. While group cohesion is strongly influenced by internal communication, weaker external ties are necessary for integration at larger scales, providing individuals with information and resources beyond the bor-

ders of their own community [15, 16, 17, 18, 19, 20]. Previous studies have shown that the structure of both strong and weak ties affects the behavior of social systems, including the spread of innovation [21], business and culture [22], crime systems [23], and the development of regional and national events [24]. Social fragmentation affects the way information flows among individuals [25] and consequently their emergent behaviors [5, 26, 9], including political or physical conflict [27, 28, 29, 30].

The recent availability of large-scale datasets obtained from communication or transaction records for landlines, mobile phones, social media, and banknote circulation has considerably improved our ability to study social systems [31, 32, 33, 34]. Geo-located data sources, such as Twitter, enable direct observation of social interactions and collective behaviors with unprecedented detail. While the Twitter user base is known to skew younger and more urban [35, 36], the large size of its user base and high frequency of tweets has enabled new types of studies of networks and geo-located activities. For example, Twitter data has been utilized in studies on a wide range of behavioral phenomena, including human migration, disease outbreaks, and patterns of happiness and lifestyle [37, 38, 39, 40, 41, 42].

Networks of human mobility [31, 34, 43, 44, 45] and communication [10, 34, 46, 47, 48, 49, 37] reveal the existence of geo-located communities or patches. Researchers have used Twitter data on mobility to show where geo-located communities deviate from administrative boundaries in Great Britain [42]. Others have generated networks of Twitter communications and examined community formation in various countries [37] or in a natural disaster [41]. While these studies analyze the structure of mobility or communication networks separately, we show that these two are not independent from one another and rather that networks in physical space are mirrored in the virtual space.

In this work, we utilize geo-located Twitter data to identify two networks in the U.S., human mobility and communication. We show that the specific geographic patches of both networks are very similar. We validate the significance of these patches by analyzing hashtag use by location and find similar patterns of divergence as in the mobility and communication networks. Finally, we build a model of network growth to understand the generic statistical properties of the natural human dynamics observed in the data. Our model combines a distance gravity component for cluster formation with preferential attachment and spatial growth mechanisms to allow clusters to differentiate in geographical space and grow over time. This work provides a comprehensive depiction of network dynamics and social fragmentation in the U.S.

2   Materials and Methods

2.1   Data

We use geo-located Twitter data to generate geographical networks based on where people travel or communicate. The data were obtained using the Twitter Streaming Application Programming Interface (API). We collected tweets from August 22, 2013 to December 25, 2013, totaling over 87 million tweets posted by over 2.8 million users in the U.S. A link to the raw data file is available at: www.necsi.edu/fragmentation/data.

2.2   Networks

We analyze mobility and communication patterns by generating geographical networks. Nodes represent a lattice of $0.1°$ latitude $\times$ $0.1°$ longitude cells overlaid on a map of the U.S. Each cell is approximately 10 km wide. There are about 400,000 cells comprising inhabited areas of the U.S. Network edges reflect two types of data: mobility and communication. In the mobility network, edges are created when a user u tweets consecutively from two locations, i and j. In the communication network, edges are created when a user u at location i mentions another user v that has most recently tweeted at location j. The weight of an edge represents the number of people who either travel or communicate between i and j. These networks aggregate the heterogeneities of human activities in a large-scale representation of social collective behaviors [50].

2.3   Methods

The term network fragmentation is often used in the literature to describe the process of network dismantling [51, 52]. In this work, we use the term "social fragmentation" to represent the modular structure of a social system due to absence of links and nodes. This is in line with terminology from other works that employ community detection methods such as the Girvan-Newman method [53].

We analyze social fragmentation by applying the Louvain method [54] with modularity optimization [55] to the mobility and communication networks obtained from Twitter data. The Louvian algorithm starts by considering each node as a single community. Iteratively, nodes move to the neighboring communities and join them to maximize modularity (M). Modularity is a scalar value $-1 < M < 1$ that quantifies how distant the number of edges inside a community are from those of a random distribution. Negative modularities occur when nodes are assigned to the wrong communities, zero occurs when all the nodes are assigned to a single community, and higher values represent increasingly optimal partitions as the values get closer to 1 [56, 57].

To study communities at multiple scales, we use a generalized version of modularity [54] that includes a resolution parameter γ. In the conventional modularity equation, $\gamma = 1$ and the same weight is given to observed links and expected links from a randomized network. In the generalized form, $\gamma < 1$ gives more weight to the observed links, which generates larger communities, while $\gamma > 1$ puts more weight on the randomized term and generates smaller communities. Because it is a method with multiple maxima, we chose partitions that are robust to multiple runs of the algorithm.

We validate the significance of the patches by observing hashtag use. We create a matrix whose rows represent locations and columns represent hashtags. In order to observe collective behaviors, we consider only those hashtags that were posted at least 500 times and locations with at least 20 tweets. We apply the term frequency-inverse document frequency (TF-IDF) transformation [58, 59] to the matrices in order to normalize the hashtags (columns of the matrix). We then apply principal component analysis (PCA) [60] to the hashtag matrix and retrieve the top 100 components, and then apply t-distributed stochastic neighbor embedding (t-SNE) [61, 62] to the resulting PCA matrix.

3   Results

3.1   Social Fragmentation

We first generated a mobility network of instances in which a user tweets from different locations, representing travel (see Section 2.3). Figure 1 depicts the spatial properties of the mobility network on a map of the U.S. in terms of degree centrality (Figure 1 (a)) and two levels of modular structure (Figure 1 (b) and Figure 1 (c)). The degree centrality shows the density of user movements at each geographical point. The activity is concentrated in large cities (red in Figure 1 (a)) and decreases toward suburban and rural areas (green, blue and gray). In areas of the country with high population density, cities merge into large regions of high activity (e.g., the East Coast corridor). In other areas, roads are also visible, as people tweet when they travel between cities. Highways in rural areas with higher traffic appear in green, and less traveled roads are blue.

The spatial fragmentation of social systems arises when people travel and choose which boundaries not to cross either directly or incidentally. Our results suggest that the U.S. mobility network is fragmented into 20 large communities (Figure 1 (b)) whose boundaries often follow state boundaries but may in particular cases be parts of one state or the combination of multiple states. At a finer scale of subdivision, these large communities of the mobility network are subdivided into patches that typically include individual cities and their surrounding areas. There are 206 such communities that we obtain by applying the same modularity optimization algorithm to each larger community (Figure 1 (c)).

Following the mobility network, we generated a communication network from Twitter mentions, shown in Figure 2. Our modularity analysis on this network shows that it also has structure of social fragmentation that is consistent with the mobility network. Thus, while the Internet and social media have drastically affected the dynamics of communications, the geographic structure of online communication remains fragmented and presents a similar structure to the one obtained from offline interactions. There are some differences as well. In contrast to the 20 modules in the mobility network, there are 15 modules that arise in the communication network.

The borders of some communities in Figure 2 are almost the same as those in the mobility network (Figure 1 (b)), such as the community encompassing states of the Northwest (WA, OR, ID, and MT), the community corresponding to Michigan (MI), and the community corresponding to Florida (FL). Ohio (OH), western Pennsylvania (PA) and West Virginia (WV) are also still in the same patch. Meanwhile, other communities in the mobility network merge into a larger community in the communication network. For example, the six-state region of New England (Maine (ME), Massachusetts (MA), New Hampshire (NH), Vermont (VT), Rhode Island (RI) and Connecticut (CT)) is a separate community in the mobility network but is combined with New York (NY), New Jersey (NJ), and Pennsylvania (PA) in the communications network. The two patches of North and South Carolina (NC and SC) and Virginia (VA) and Maryland (MD) are also combined into one. This demonstrates that certain areas have a broader radius of online communication than physical travel. Finally, Figure 2 (c) represents the smaller communities within each community in Figure 2 (b). These patches show areas connected to urban centers and are very similar to those of the mobility network in Figure 1 (c). Some less populous states are now single communities, such as Montana (MT), Nebraska (NE), Kansas (KS), Oklahoma (OK), Arkansas (AR), and New Mexico (NM), while more densely populated areas are subdivided around urban centers.

To further investigate the role of state boundaries in community formation, we quantified to what extent each state contributes to communities for both networks (Section S1.1 and Figure
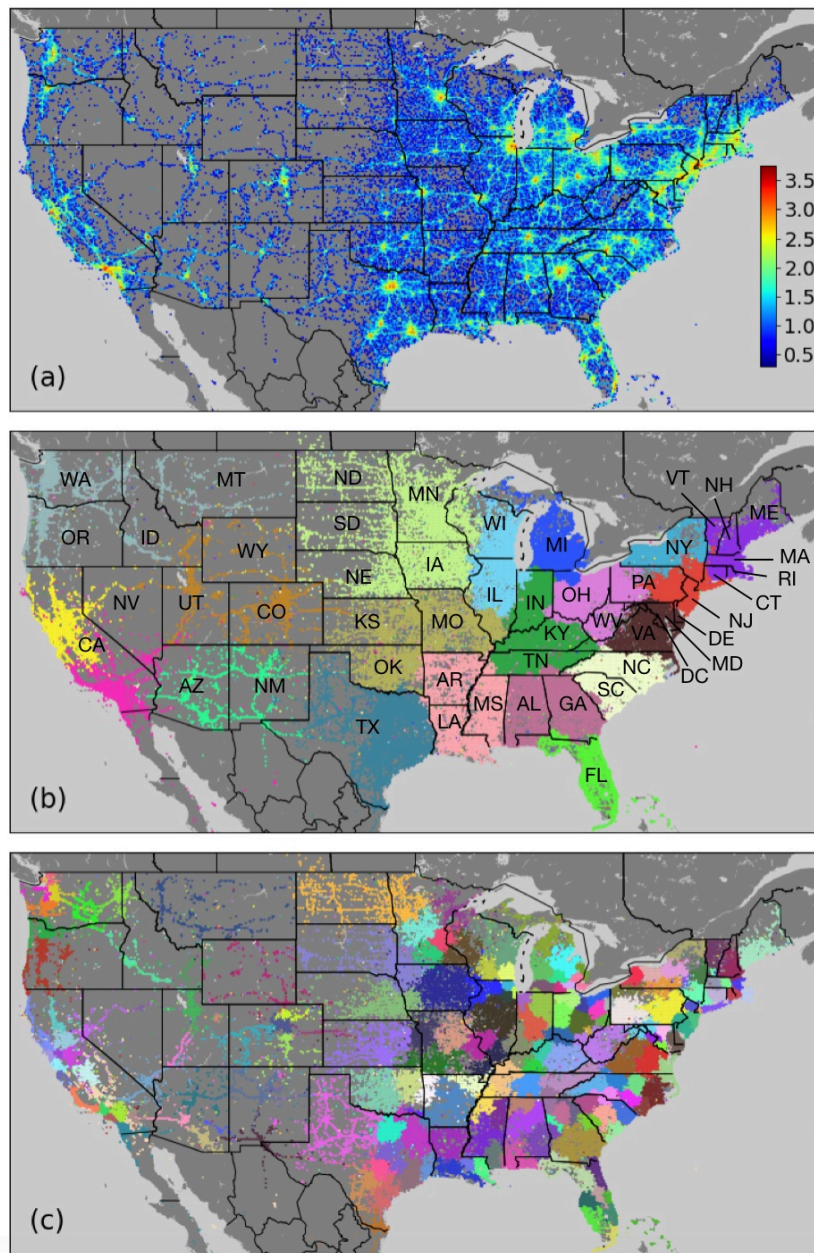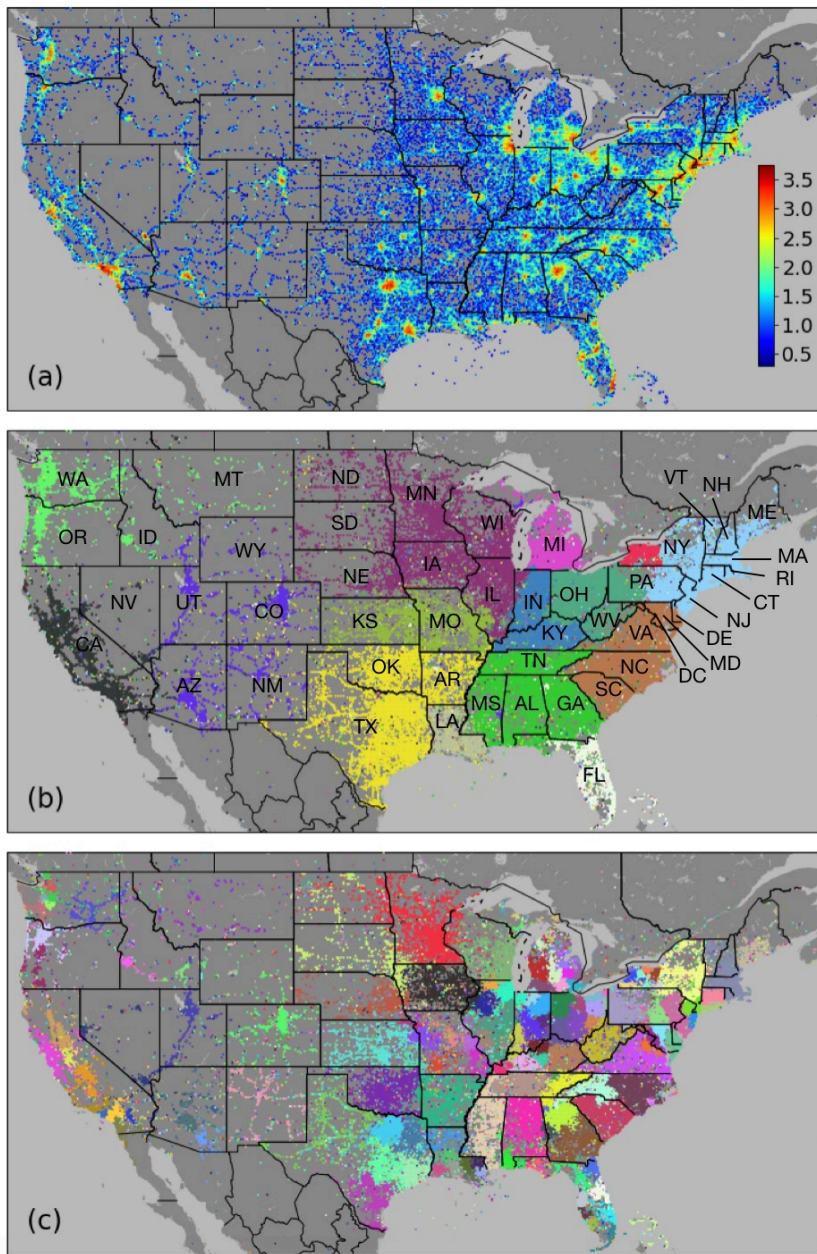
4

Figure 1: Structure and fragmentation patterns of the network associated with human mobility. (a) Spatial degree centrality of the mobility network. Colors indicate the amount of people traveling at each location, measured by the logarithm of the degree centrality of each node (scale inset). The mobility network was used to generate communities using modularity optimization, with distinct colors indicating (b) 20 patches that can be visually associated to states or regions and (c) 206 smaller sub-communities within the communities of panel (b) that can be visually associated to urban centers.

Figure 2: Structure and fragmentation patterns of the network associated with human communication. (a) Spatial degree centrality of the communication network. Colors indicate the amount of communication at each location, measured by the logarithm of the degree centrality of each node (scale inset). The communication network was used to generate communities using modularity optimization, with distinct colors indicating (b) 15 patches that can be visually associated to states or regions and (c) 168 smaller sub-communities within the communities of panel (b) that can be visually associated to urban centers.
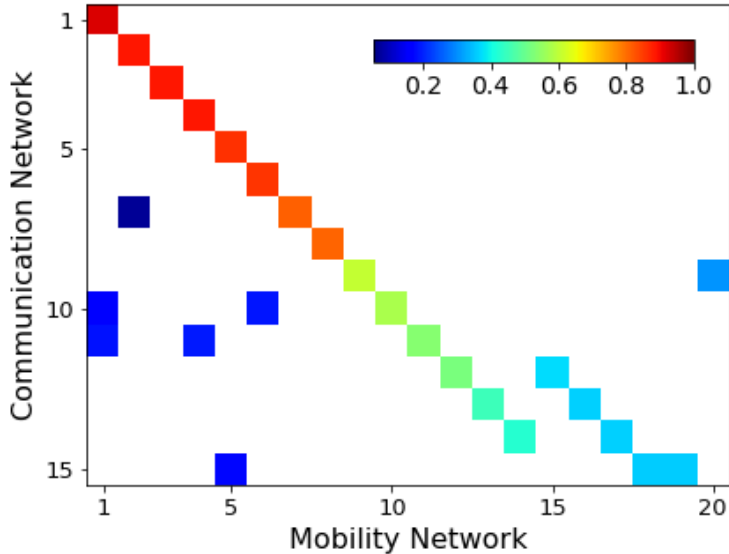
Figure 3: Similarity of communities in the communication and mobility networks. Matrix of the regional communities for the communication network (y-axis, $n = 15$) and mobility network (x-axis, $n = 20$), ordered by decreasing overlap between communities. Cell colors represent the number of nodes overlapping between the two networks in each community, normalized by the size of the communities per row (scale inset), with no overlap indicated in white.

S1). States mostly belong to specific communities. This shows that the structure we observe is not simply due to the effects of distance [13]. To show this, we generated artificial networks with links weighted by only the inverse of distance or distance squared (see Section S1.2). While spatial patches are also present in these artificial networks, the patches do not follow state boundaries and are not consistent across both types of networks (see Figures S2 and S3). We also performed validation of community stability and find that the number of communities and boundaries we show are consistent and stable across multiple realizations of the algorithm (see Section S1.3 and Figure S4). Overlapping regions across realizations can happen either because small locations flip between large communities or because large communities are split into smaller ones.

We quantitatively compared the modular structure of the mobility and communication networks (Figure 3) by creating a matrix where we count the number of overlapping nodes of communities arising from the networks of communication (rows) and mobility (columns). Rows have been normalized by the size of each community in the communication network. Some communities from the communication network are almost identical in the mobility network and therefore show a high overlap (red). Others are similar but not identical. A few communities from the mobility network are merged into communities in the communication network (green and light blue). Despite the observed differences in the networks representing two fundamentally different types of interactions, the modular structure is remarkably consistent, revealing that there is a strong coupling between the way people travel in physical space and communicate with each other online.

In order to further understand similarities between the mobility and communication networks, we performed a multi-scale analysis of community structure using a generalized modularity op-

timization algorithm that introduces a resolution parameter, γ [54]. Smaller values of γ identify progressively larger communities, and vice-versa. The multi-scale analyses of the mobility and communication networks are shown for some examples of γ values in Figures 4 and 5, respectively. Partitions range from a single large module of the entire U.S. (top panels) down to urban scale partitions (bottom panels). Some states like Pennsylvania (PA) are split into multiple communities early in the process ($\gamma \approx 0.4$ in the mobility network), while other states like Texas (TX) emerge as single communities ($\gamma \approx 1$ in the mobility network) and internally fragment later in the process. These differences are directly associated with the internal structure of social ties and their geographical breakpoints, further explored in the Discussion (Section 4). In order to validate these partitions, we compared them with the communities detected by Infomap [63]. This method finds the best partition based on the flow of information in a network. The comparison shows that the patterns obtained using Infomap are very similar to the ones obtained from the multi-scale modularity method at specific values of γ (see Section S1.4 and Figure S5).

We compare the partitions in both networks for different values of γ by using three measures of cluster similarity: Purity [64], Adjusted Rand Index [65] and Fowlkes-Mallows Index [66]. These measures evaluate the overlap of partitions, with values ranging between 0 (no intersection) and 1 (perfect match). Figure 6 shows a matrix whose rows and columns represent the partitions of the mobility and communication networks at different values of resolution (γ-mobility and γ-communication) and whose elements show the average of the three measures of similarity. The highest similarity between the two networks occurs at similar values of resolution (red diagonal), showing that the relative structure of these networks is consistent across scales. Additional comparisons between the two networks can be found in Section S1.5, including measures of degree centrality and edge weight (Figure S6) and an alluvial diagram (Figure S7).

The consistency between the mobility and communication networks reveals that social spaces are not limited to the physical space. Instead, offline interactions seem to condition the structure of online communications. Moreover, the hierarchical multi-scale structure of these networks reveals that smaller communities with cohesive social ties, interactions, and associations belong to progressively larger ones. It may be expected that locations from the same community will have more similarity than locations from different communities.

Locations from the same community show similarity in hashtag use and divergence with locations from different communities for either the mobility or communication networks (Figure 1(b) and 2(b)). Hashtags highlight specific, shared experiences and serve as markers of social interaction [67]. We compared hashtags for locations in the mobility and communication networks at $\gamma = 1$ using principal component analysis (PCA) followed by t-distributed stochastic neighbor embedding (t-SNE) analysis (Figure 7 (a) and (c) for mobility and communication, respectively). See Section 2.3 for more information on the method. We colored each dot by location, matching the colors of the communities in Figures 1 and 2, panel (b). A number of distinct colored clusters emerge, suggesting that hashtag use by location corresponds to communities of the mobility or communication networks. Some clusters appear to separate into smaller clusters near to each other, representing sub-communities inside the communities. These patterns are statistically significant after randomizing locations ($p < 0.001$), detailed in Section S1.6 and Figure S8. To compare communities to each other based on hashtag use, we performed analysis of cosine similarity (Figure 7 (b) and (d) for mobility and communication, respectively). Squares are colored from blue to red for increasing similarity (color bar, right). About half of community pairs have less than 50% similarity, while the rest have $50-90\%$ similarity. Communities are distinct at some scales and form larger communities
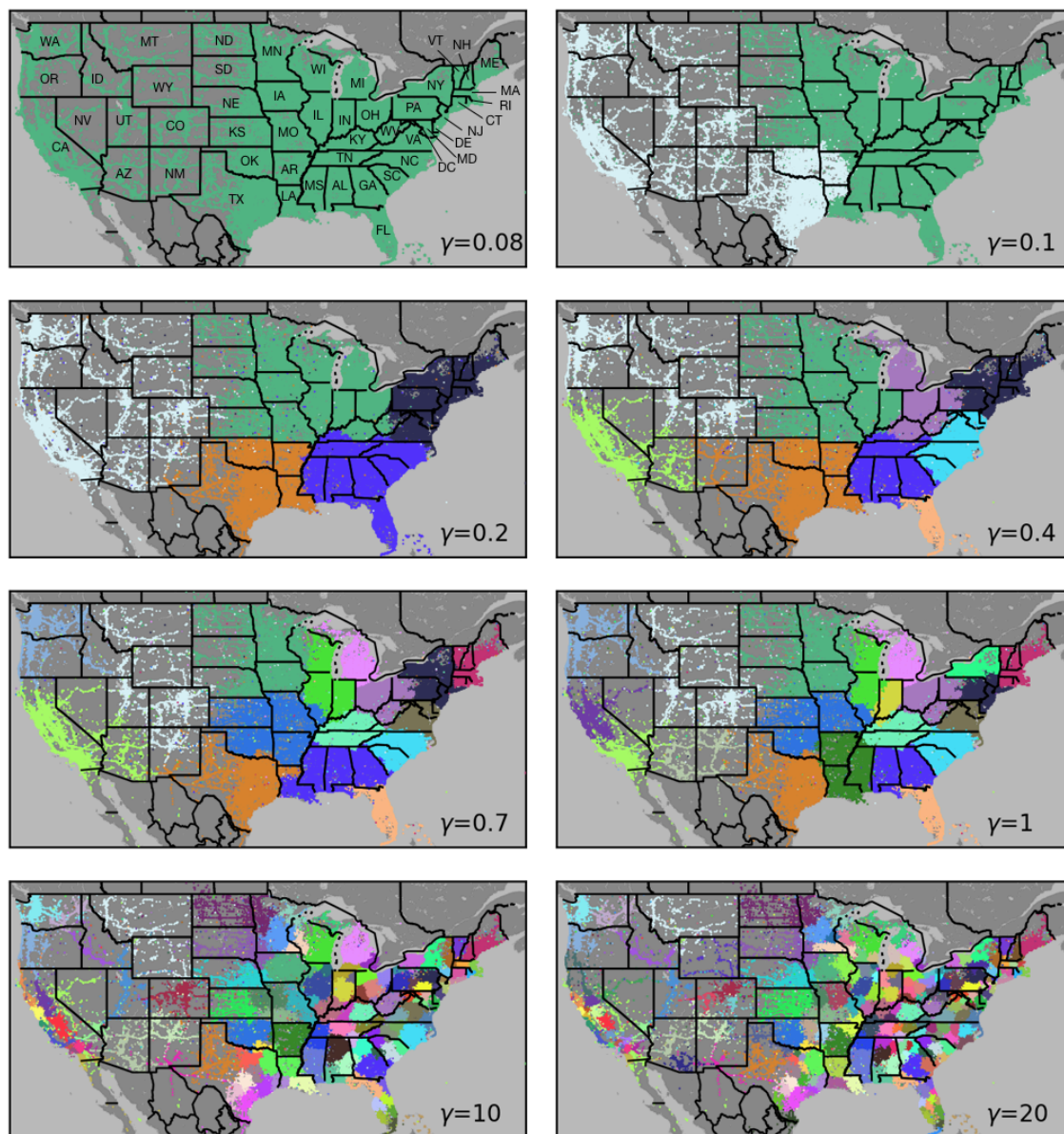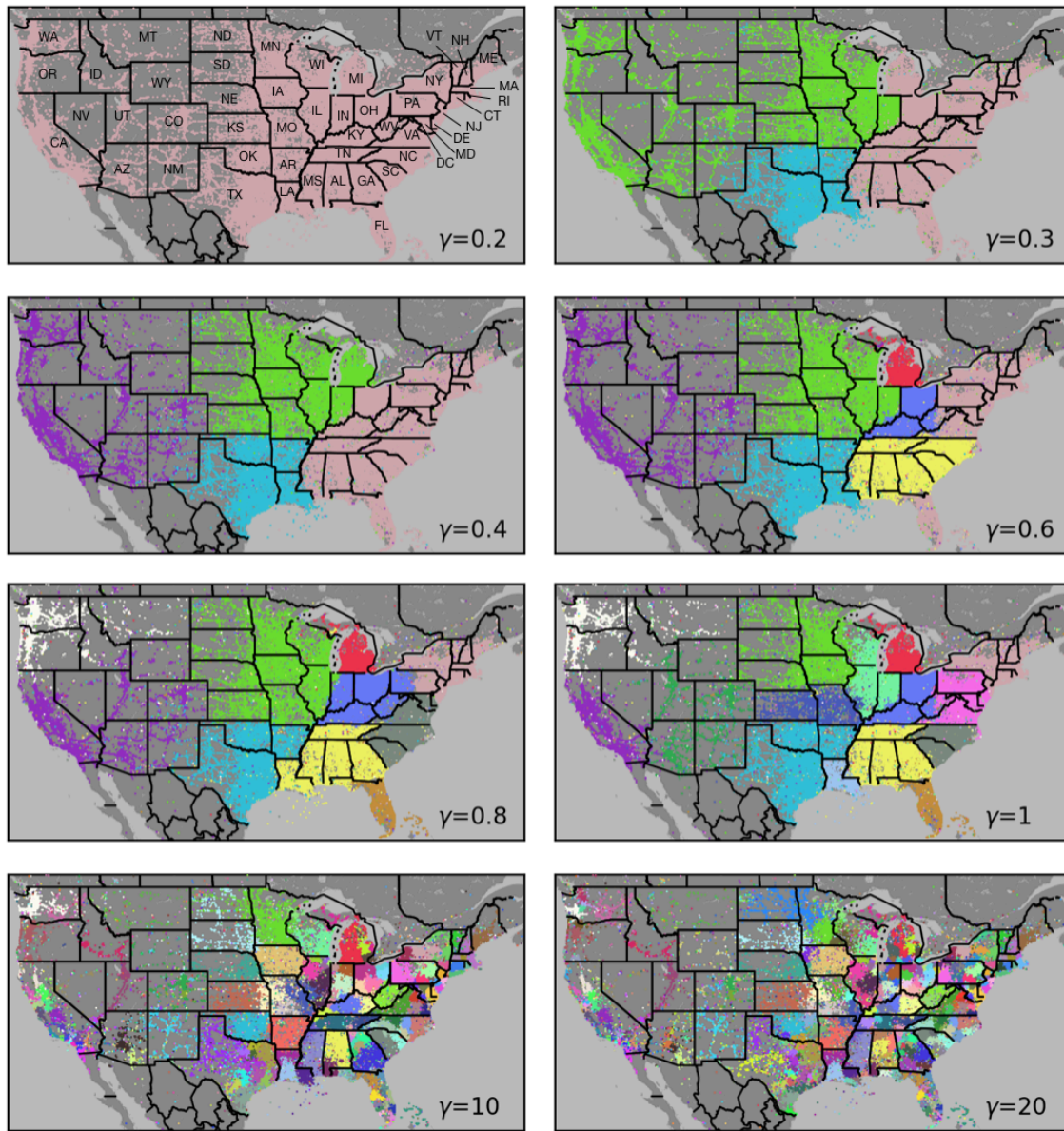
Figure 4: Multi-scale decomposition of the mobility network. Colors indicate geographical patches detected in the mobility network for values of the resolution parameter γ varied from $0.08 - 20$ (upper left to bottom right). Colors are retained across panels by the following rule: when a community is divided into multiple sub-communities, the sub-community that is the most connected to the original ("parent") community retains the color of the parent community; the other sub-communities are assigned new colors. The modularity for all of the panels is over 0.8.

Figure 5: Multi-scale decomposition of the communication network. Colors indicate geographical patches detected in the communication network for values of the resolution parameter γ varied from $0.2 - 20$ (upper left to bottom right). Colors are retained across panels by the following rule: when a community is divided into multiple sub-communities, the sub-community that is the most connected to the original ("parent") community retains the color of the parent community; the other sub-communities are assigned new colors. The modularity for all of the panels is over 0.8.
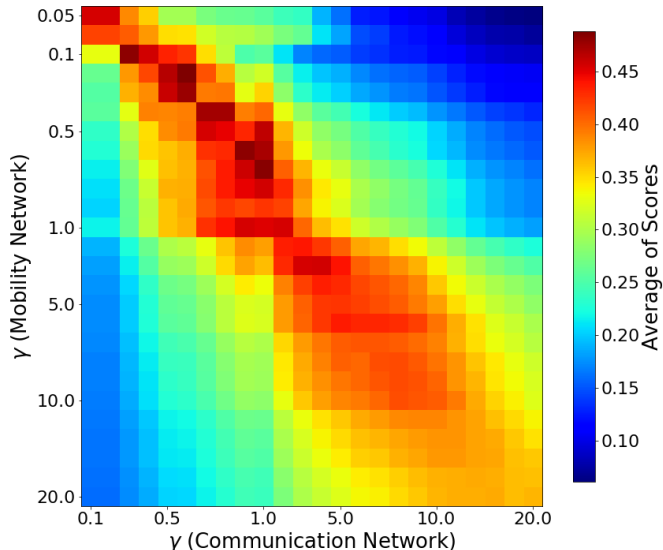
Figure 6: Similarity between the mobility and communication networks across multiple scales. Similarity is measured by the average of the Purity, Adjusted Rank, and Fowlkes-Mallows Indexes (color scale shown). Scale is defined by the different values of the resolution parameter γ (horizontal and vertical axes).

at higher scales.

3.2 Model

We constructed a network growth model that combines aspects of network dynamics and human mobility in order to show the emergence of social fragmentation. Our model combines geographical distance gravity [14], preferential attachment to allow creation of hubs (cities), and spatial growth to allow the growth of cities [68]. We begin with a lattice representing geographical locations, and grow connections among them simulating the way people travel. The probability of creating an edge between locations i and j in each time step is:

$$P_{ij} \sim <k_{nn}>_i^{\nu} \frac{k_j^{\alpha}}{d_{ij}^{\beta}} \tag{1}$$

where i represents the origin of the interaction, j indicates the destination, $<k_{nn}>_i$ indicates i's nearest neighbors' average degree, k_j represents j's degree, and d_{ij} represents the distance between i and j. The exponents α, β and ν control the effects of the preferential attachment mechanism, geographical distance gravity and spatial growth, respectively. The model reproduces the growth of geographical clusters similar to cities (ν), their degree of attractiveness (α) and the linkage between urban centers and surrounding areas, including neighboring cities (β). We introduce the preferential attachment mechanism to break the symmetry of spatial connections over time and the spatial growth mechanism to allow the city-like structures to grow.

Each location in the lattice has 4 nearest neighbors, except for locations in corners and on edges, which have 2 and 3 neighbors, respectively. Simulations start with a random seed of three
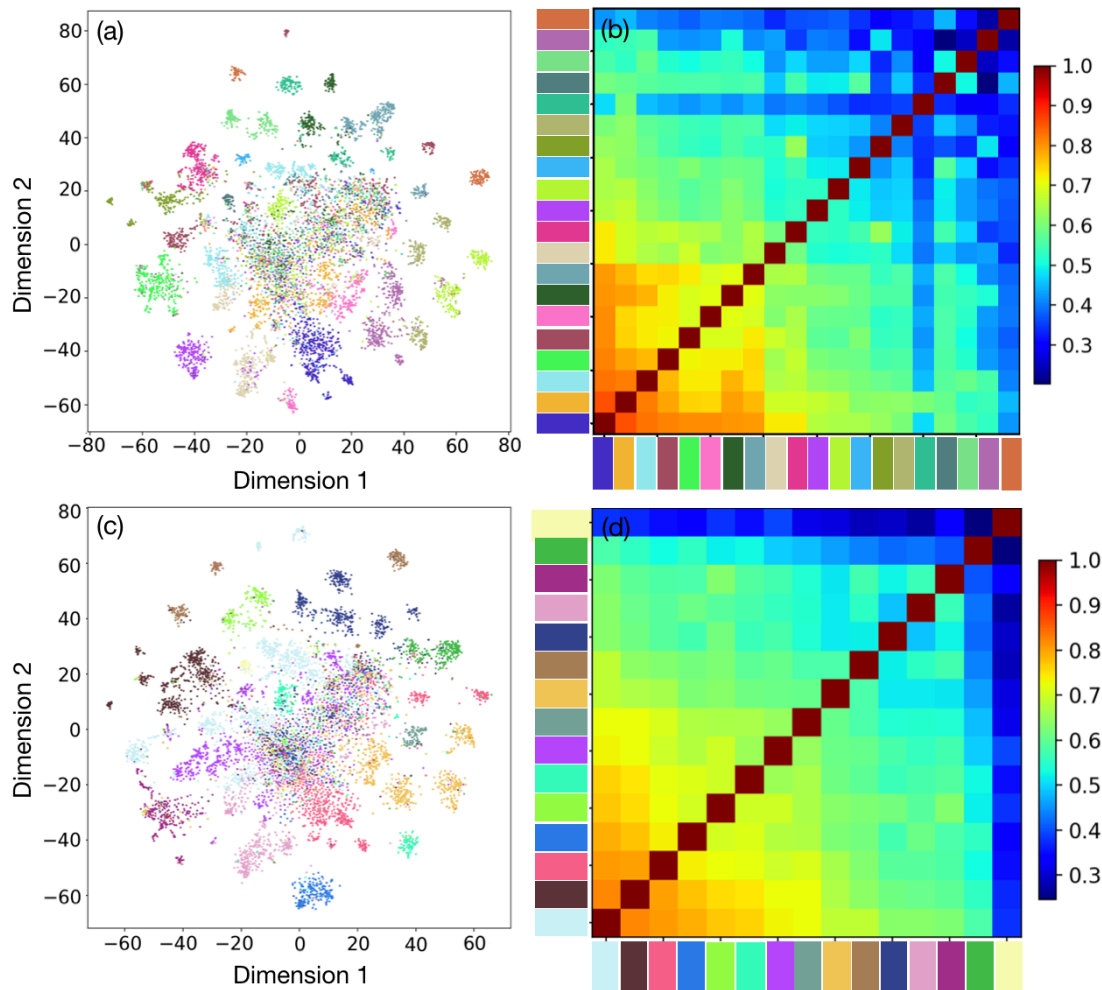
11

Figure 7: Dimensional reduction analysis of hashtag use by location and cosine similarity of communities based on hashtags. Panel (a) shows the results of t-SNE analysis on the first 100 components of PCA analysis of hashtags in locations of the mobility network. Panels (b) shows cosine similarity of hashtag use in the communities of the mobility network at $\gamma = 1$. Panels (c) and (d) show the corresponding t-SNE result and cosine similarity for the communication network. Colors match those of the communities in Figures 1 and 2, panel (b).

connected locations. Links are undirected and weighted to represent the iteration of links over time. Origins are picked randomly (independent from destinations) if their normalized value of $< k_{nn} >^\nu$ exceeds a random threshold. To allow all the locations in the lattice to participate in the dynamics, for the first N time steps, we turn off the origin priority selection and let the system choose origins from a random order of locations, where N represents the number of locations. The probability of selecting destinations is a combination of the preferential attachment mechanism and geographical distance gravity as shown in Equation 1. Thus, locations that are nearer to the origin location and have a higher degree have a higher probability to be chosen. Simulations continue

until reaching a stable state in which communities form and do not change in number.

Figure 8 shows the results of model simulations in terms of the spatial degree distribution (top panels) as well as modular structure (bottom panels) for different values of α (rows) and β (columns) and a fixed value of $\nu = 0.1$. If we do not include the effects of either preferential attachment ($\alpha = 0$) or gravity ($\beta = 0$), the destinations of edges are independently distributed among all nodes and the resulting communities have no spatial pattern. If $\alpha > \beta$, then a few hubs and one or two communities arise without significant geographic effects. Spatial fragmentation arises when the gravity mechanism is stronger than the preferential attachment ($\beta > \alpha$), either without hubs ($\alpha = 0$) or with hubs ($\alpha > 0$). Increasing ν leads to more localized high-activity areas (cities), but this also destroys localized patches, leading to lower values of modularity. For additional results exploring variation of the spatial growth mechanism while keeping α and β constant, see Section S1.7 and Figure S9.

We validated the model results against Twitter data by first testing whether the degree distributions from both sources are drawn from the same distribution and second comparing the modularity values. For each set of parameters, we created 20 model realizations and analyzed their statistical behavior. We applied the Kolmogorov-Smirnov statistical test (K-S) to compare the average degree distribution from the model realizations to that of the mobility network, and similarly for the communication network. Figure 9 (a) shows the values of the test results for different values of α and β (rows and columns of the matrix) and $\nu = 0.1$. Lower K-S values (red) indicate more similarity, and higher K-S values (blue) indicate less similarity. The average modularity values for the simulations in Figure 9 (a) are shown in Figure 9 (b), ranging from 0 (no modular structure) to 1 (high modular structure). We find that $\alpha = 0.9$, $\beta = 1.5$, and $\nu = 0.1$ give a good fit between simulations and observed data. Results for the K-S statistic with variation of all three parameters are shown in Section S1.7 and Figures S10 and S11.

4   Discussion

Understanding the structure and dynamics of groups is an essential aspect of understanding social interactions generally. The functioning of human societies arises not only from the activities of individuals but also from their interaction and integration by means of social ties. We analyzed the structure of social ties in the U.S. using Twitter data and found multi-scale, self-organized fragments that span from urban up to national scales for mobility, communication, and hashtag use. Our results show that the structures emerging from these different types of interaction are highly consistent, revealing that social ties couple the integration and separation of groups in both physical and virtual spaces. Despite potential biases in Twitter samples [36, 35], the similarity of the detected communities between mobility and communication networks shows that the networks reveal the underlying social structure.

We also constructed a model of network growth that is consistent with the statistical property of emergence of the observed patterns from the Twitter data. Our model shows that social fragmentation may result from short-distance interactions, in support of hierarchical models of social network formation [69]. However, this mechanism alone does not explain the emergence of highly connected places such as cities. We model the emergence of cities using preferential attachment and spatial growth mechanisms, which increase heterogeneity in degree distribution but may destroy spatial fragmentation if cities grow large enough. Other generative models can also create fat-tails and power-law behaviors. For example, the emergence of city centers can also be
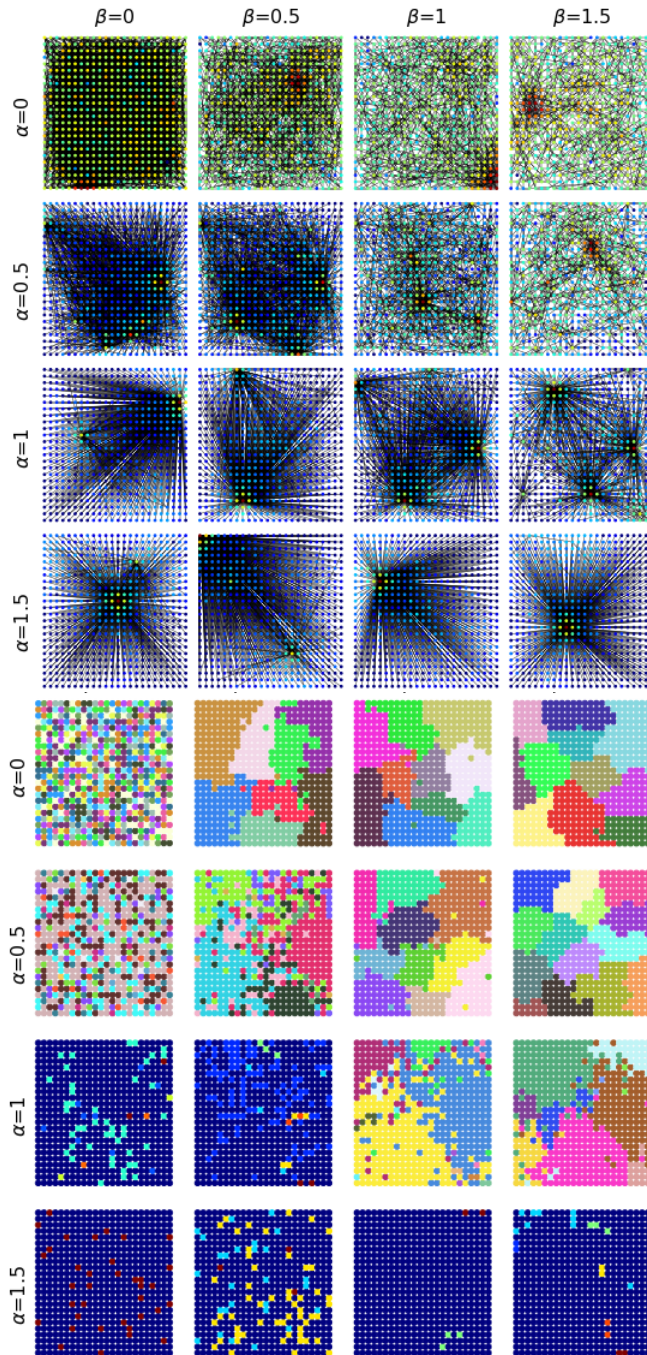
Figure 8: Spatial degree distribution and modular structure for model simulations with different values for parameters α (preferential attachment) and β (geographical distance gravity) and a fixed value of 0.1 for ν (spatial growth). Top panels show the spatial degree distribution (from weakly connected in blue to highly connected in red). Bottom panels show the modules of each graph, with each color identifying a single community.
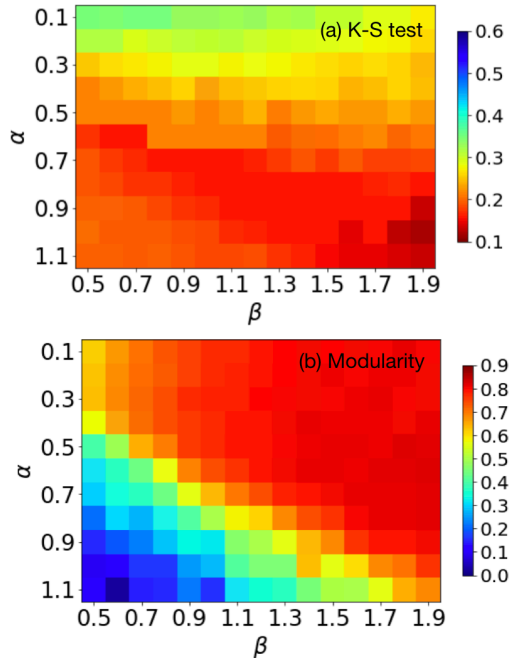
14

Figure 9: Kolmogorov-Smirnov score and modularity for simulations with varying model parameter, α and β, and fixed resolution parameter $\nu = 0.1$. (a) Colors indicate the Kolmogorov-Smirnov (K-S) score, with lower scores (red) indicating similarity between the degree distributions of the model and the mobility network. (b) Colors indicate network modularity. Modularity is highest at around 0.8 (dark red), similar to the actual modularity for the mobility network (0.83).

modeled as processes of optimization of social interactions and information flows or as outcomes of multiplicative growth mechanisms [70].

The gravity model [71, 72] describes how the strength of mobility between two locations is directly related to the population density of those locations and inversely related to the distance between the locations, each of which is a power-law relation. Reported values for the scaling exponents vary in the range $0.5 - 2.0$ depending on the system [2, 10, 34, 13, 73, 42]. Thus, the gravity model predicts that cities with higher population densities attract higher mobility. However, an important limitation is that the gravity model may overestimate mobility from a low density population to a high density population, limiting its applicability over wide geographic areas [14]. Furthermore, the gravity model does not allow for cluster growth or changes in the population of locations. We overcome these limitations by creating a model incorporating geographical distance gravity with preferential attachment and spatial growth.

The formation of groups and their interactions are intimately related to the formation of individual identity through self-identification and adoption of group norms and narratives. Thus, while individual identities are highly complex and unique, there are shared patterns among members of self-associating groups. These common patterns define the group identity, which may involve linguistic, cultural, economic, opinion or interest differences from other groups. To investigate divergence of shared social experience, we analyzed hashtag use by location. Hashtags

are a means of discussing shared experiences and ideas, aspects of group formation. Our analysis demonstrated that many of the communities from the mobility or communication networks have also distinct hashtag use. This suggests that the communities shared experiences also diverge from other communities.

When we further examine the mobility and communication networks at different scales (Figures 4 and 5), we observe that many communities follow state lines, but a few do not, suggesting other forces driving community formation. The large metro area around St. Louis, MO creates a community that spills across the Mississippi River and thus the Missouri (MO)-Illinois (IL) state line (mobility network: $\gamma \approx 0.7 - 20$, communication network: $\gamma \approx 1$). Eastern and western Pennsylvania (PA) splits into two communities, roughly along the Appalachian mountain boundary (mobility network: $\gamma \approx 0.4 - 1$, communication network: $\gamma \approx 0.6 - 1$). California (CA) splits into northern and southern communities (mobility network: $\gamma \approx 0.7 - 1$, communication network: $\gamma \approx 2$), following a known cultural and economic divide [74, 75]. The area of eastern Idaho (ID) combines with Utah (UT) (mobility network: $\gamma \approx 0.3 - 0.7$, communication network: $\gamma \approx 0.7 - 2$), corresponding to the area of historical Mormon settlement [76]. Geographers have proposed that many cultural, political, and religious divisions trace back to the original settlers in each area [77], such that America can be divided into corresponding cultural regions or "nations" [78], which has largely been supported by recent genetic studies of the U.S. population [79]. Our observations also support that the communities we observe reflect geographical, cultural, and economic forces that can supersede administrative boundaries in some locations, although state boundaries remain an important factor in social interactions.

Recent trends seem to be accelerating the forces of community formation and divergence we observe. These forces include economic shifts, political polarization, and the rise of social media. Analyses of work commutes have supported the rise of "megaregions," interconnected labor markets with large cities as hubs, reminiscent of the communities we observe [80]. Migrations from one megaregion to another may be motivated by economics, such as the migration over the last decade from the Northeast toward the mountain West and Southwest, which have offered better job prospects and lower housing prices [81]. In addition to economic movements, an increase in political self-sorting behavior has been observed, with people physically moving nearer to like-minded individuals [82]. The percentage of people who identify as "consistently" liberal or conservative has doubled to over 20% in the past two decades, and these individuals express preferences to live near to, be close friends with, and marry those of the same political persuasion [83]. Social media may be exacerbating this polarization, creating spaces in which users interact with like-minded individuals and ignore opposing opinions [4]. Future work will need to examine how patterns of group formation change and whether cultural, political, or economic factors drive this polarization.

Moving forward, there are at least two strategies for policymakers seeking to address social fragmentation in the U.S. One is to fight social fragmentation by promoting intergroup connection and uniformity in society. The other is to recognize that social fragmentation is present and to incorporate it into policy decisions. This means adopting a policy of localism, which involves tailoring policy approaches to each specific area and fostering participation from local political groups [84]. Our analysis suggests that division into two political groups (e.g., Republican and Democrat) is not sufficient in the U.S. today and that sub-groups may require partial local autonomy to address the multi-scale divisions present in society.

5   Conclusion

In summary, we have used geo-located Twitter data to generate networks of U.S. mobility, communication, and hashtag use and to explore how networks fragment at multiple scales. We also developed a model of network growth that incorporates the properties of geographical distance gravity, preferential attachment, and spatial growth and successfully replicates statistical properties of the social fragmentation patterns observed in the data. Overall, our analysis demonstrates there are many boundaries along which fragmentation of U.S. society may be taking place. Moreover, this fragmentation represents a multi-factorial and dynamic process that is ongoing. It is an important question how social fragmentation at multiple levels will affect the stability and dynamism of U.S. society in the future.

References

[1] A. J. Morales, V. Vavilala, R. M. Benito, and Y. Bar-Yam. Global patterns of synchronization in human communications. *J. R. Soc. Interface*, 14(128), 2017.

[2] G. Krings, F. Calabrese, C. Ratti, and V. Blondel. Urban gravity: a model for inter-city telecommunication flows. *J. Stat. Mech.*, 2009:L07003, Jul 2009.

[3] G. D. Nelson and A. Rae. An economic geography of the United States: From commutes to megaregions. *PLOS ONE*, 11(11):1–23, 2016.

[4] A. Herdağdelen, W. Zuo, A. Gard-Murray, and Y. Bar-Yam. An exploration of social identity: The geography and politics of news-sharing communities in Twitter. *Complexity*, 19(2):10–20, 2013.

[5] E. Bakshy, S. Messing, and L. A. Adamic. Exposure to ideologically diverse news and opinion on Facebook. *Science*, 348(6239):1130–1132, 2015.

[6] P. J. Mucha, T. Richardson, K. Macon, M. A. Porter, and J-P. Onnela. Community structure in time-dependent, multiscale, and multiplex networks. *Science*, 328(5980):876–878, 2010.

[7] H. Mao, X. Shuai, Y-Y. Ahn, and J. Bollen. Mobile communications reveal the regional economy in Cte dIvoire. *NetMob D4D Challenge*, 2013.

[8] M. Fujita, P. Krugman, and A. J. Venables. The spatial economy: cities, regions, and international trade. *The MIT Press*, 2001.

[9] T. C. Schelling. Dynamic models of segregation. *J. Math. Sociol.*, 1(2):143, 1971.

[10] R. Lambiotte, V. D. Blondel, C. de Kerchove, E. Huens, C. Prieur, Z. Smoreda, and P. Dooren. Geographical dispersal of mobile communication networks. *Physica A: Statistical Mechanics and its Applications*, 387:5317–5325, 2008.

[11] C. Herrera-Yage, C. M. Schneider, T. Couronn, Z. Smoreda, R. M. Benito, P. J. Zufiria, and M. C. Gonzlez. The anatomy of urban social networks and its implications in the searchability problem. *Scientific Reports*, 5:10265, 2015.

[12] D. Liben-Nowell, J. Novak, R. Kumar, P. Raghavan, and A. Tomkins. Geographic routing in social networks. *Proceedings of the National Academy of Sciences*, 102(33):11623–11628, 2005.

[13] L. Backstrom, E. Sun, and C. Marlow. Find me if you can: improving geographical prediction with social and spatial proximity. *In Proceedings of WWW*, 10:61–70, 2010. doi: 10.1145/1772690.1772698.

[14] F. Simini, M. C. Gonzlez, A. Maritan, and A.-L. Barabsi. A universal model for mobility and migration patterns. *Nature*, 484:96–100, 2012.

[15] M. Granovetter. The strength of weak ties. *American Journal of Sociology*, 78(6):1360–1380, 1973.

[16] R. L. Coser. *'Complexity of roles as a Seedbed of Individual Autonomy' in 'The Idea of Social Structure: Essays in Honor of Robert Merton'*. Harcourt Brace Jovanovich, 1975.

[17] B. R. da Cunha, J. C. González-Avella, and S. Gonçalves. Fast fragmentation of networks using module-based attacks. *PLOS ONE*, 10:e0142824, 2015.

[18] S. Wandelt, X. Sun, D. Feng, M. Zanin, and S. Havlin. A comparative analysis of approaches to network-dismantling. *Scientific Reports*, 8:13513, 2018.

[19] L. M. Shekhtman, M. M. Danziger, and S. Havlin. Recent advances on failure and recovery in networks of networks. *Chaos, Solitons and Fractals*, 90:28–36, 2016.

[20] J. M. Kleinberg. Navigation in a small world. *Nature*, 406:845, Aug 2000. doi: 10.0.4.14/35022643.

[21] F. Noah. A test of the structural features of Granovetter's 'strength of weak ties' theory. *Social Networks*, 2:411–422, 1980.

[22] R. Breigern and P. Pattison. The joint role structure of two communities' elites. *Sociological Methods and Research*, 7(2):213–226, 1978.

[23] B. R. da Cunha and S. Gonçalves. Topology, robustness, and structural controllability of the brazilian federal police criminal intelligence network. *Applied Network Science*, 3:36, 2018.

[24] I. Pool. *Comment on Mark Granovetter's "The Strength of Weak Ties: A Network Theory Revisited."*. Read at the 1980 meeting of the International Communications Association, Acapulco, Mexico, 1980.

[25] M. Granovetter. The strength of weak ties: A network theory revisited. *Sociological Theory*, 1:201–233, 1983.

[26] A. J. Morales, J. Borondo, J. C. Losada, and R. M. Benito. Measuring political polarization: Twitter shows the two sides of Venezuela. *Chaos*, 25:3, 2015.

[27] M. Lim, R. Metzler, and Y. Bar-Yam. Global pattern formation and ethnic/cultural violence. *Science*, 317:1540–1544, 2007.

[28] A. Varshney. *Ethnic Conflict and Civic Life: Hindus and Muslims in India.* Yale University Press, 2008.

[29] D. L. Horowitz. *Ethnic Groups in Conflict.* Berkeley: University of California Press, 2011.

[30] M. J. Esman. *Ethnic conflict in the western world.* London: Cornell University Press, 1979.

[31] C. Thiemann, F. Theis, D. Grady, R. Brune, and D. Brockmann. The structure of borders in a small world. *PLOS ONE*, 5:e15422, 2010.

[32] D. Lazer, A. Pentland, L. Adamic, S. Aral, A.-L. Barabsi, D. Brewer, N. Christakis, N. Contractor, J. Fowler, M. Gutmann, T. Jebara, G. King, M. Macy, D. Roy, and M. Van Alstyne. Computational social science. *Science*, 323(5915):721–723, 2018.

[33] M. C. González, C. A. Hidalgo, and A.-L. Barabási. Understanding individual human mobility patterns. *Nature*, 453(7196):779–782, 2008.

[34] F. Calabrese, D. Dahlem, A. Gerber, D.D. Paul, X. Chen, J. Rowland, C. Rath, and C. Ratti. The connected States of America: Quantifying social radii of influence. *in Proc. of IEEE International Conference on Social Computing (SocialCom)*, 2011.

[35] Pew Research Center. Social Media Use in 2018, March 2018. URL http://www.pewinternet.org/2018/03/01/social-media-use-in-2018/.

[36] Pew Research Center. Mobile Messaging and Social Media 2015, Aug 2015. URL http://www.pewinternet.org/2015/08/19/the-demographics-of-social-media-users/.

[37] A. Kallus, N. Barankai, J. Szüle, and G. Vattay. Spatial fingerprints of community structure in human interaction network for an extensive set of large-scale regions. *PLOS ONE*, 10(5): e0126713, 2015.

[38] B. Hawelka, I. Sitko, E. Beinat, S. Sobolevsky, P. Kazakopoulos, and C. Ratti. Geo-located Twitter as proxy for global mobility patterns. *Cartography and Geographic Information Science*, 41(3):260–271, 2014.

[39] M. R. Frank, L. Mitchell, P. S. Dodds, and C. M. Danforth. Happiness and the Patterns of Life: A Study of Geolocated Tweets. *Scientific Reports*, 3(1):2625, 2013.

[40] J. I. Blanford, Z. Huang, A. Savelyev, and A. M. MacEachren. Geo-located tweets. enhancing mobility maps and capturing cross-border movement. *PLOS ONE*, 10(6):e0129202, 2015.

[41] M. Bakillah, R.-Y. Li, and S. H. L. Liang. Geo-located community detection in twitter with enhanced fast-greedy optimization of modularity: the case study of typhoon haiyan. *International Journal of Geographical Information Science*, 29(2):258–279, 2015.

[42] J. Yin, A. Soliman, D. Yin, and S. Wang. Depicting urban boundaries from a mobility network of spatial interactions: a case study of great britain with geo-located twitter data. *International Journal of Geographical Information Science*, 31(7):1293–1313, 2017.

[43] T. Menezes and C. Roth. Natural scales in geographical patterns. *Scientific Reports*, 7:45823, 2017.

[44] A. Amini, K. Kung, C. Kang, S. Sobolevsky, and C. Ratti. The impact of social segregation on human mobility in developing and urbanized regions. *EPJ Data Sci.*, 3(6), 2014.

[45] F. Xiang, L. Tu, B. Huang, and X. Yin. Region partition using user mobility patterns based on topic model. *16th International Conference on Computational Science and Engineering*, 2014.

[46] C. Ratti, S. Sobolevsky, F. Calabrese, C. Andris, J. Reades, M. Martino, R. Claxton, and S. H. Strogatz. Redrawing the map of Great Britain from a network of human interactions. *PLOS ONE*, 5(12):e14248, 2010.

[47] S. Sobolevsky, M. Szell, R. Campari, T. Couronn, Z. Smoreda, and C. Ratti. Delineating geographical regions with networks of human interactions in an extensive set of countries. *PLOS ONE*, 8:e81707, 2013.

[48] V. Blondel, G. Krings, and I. Thomas. Regions and borders of mobile telephony in Belgium and in the brussels metropolitan zone. *Brussels Studies*, 42, 2010.

[49] S. Šćepanović, I. Mishkovski, P. Hui, J. K. Nurminen, and A. Ylä-Jääski. Mobile phone call data as a regional socio-economic proxy indicator. *PLOS ONE*, 10:e0124160, 2015.

[50] Y. Bar-Yam. From big data to important information. *Complexity*, 21(S2):7398, 2016.

[51] A. Braunstein, L. Dall'Asta, G. Semerjian, and L. Zdeborová. Network dismantling. *Proceedings of the National Academy of Sciences of the United States of America*, 113(44):12368–12373, Oct 2016. doi: 10.1073/pnas.1605083113.

[52] X.-L. Ren, N. Gleinig, D. Helbing, and N. Antulov-Fantulin. Generalized network dismantling. *CoRR*, abs/1801.01357, 2018. URL http://arxiv.org/abs/1801.01357.

[53] M. Girvan and M. E. J. Newman. Community structure in social and biological networks. *Proceedings of the National Academy of Sciences*, 99(12):7821–7826, Jun 2002. doi: 10.1073/PNAS.122653799.

[54] J. Reichardt and S. Bornholdt. Statistical mechanics of community detection. *Phys. Rev. E*, 74:016110, 2006.

[55] V. D. Blondel, J.-L. Guillaume, R. Lambiotte, and E. Lefebvre. Fast unfolding of communities in large networks. *J. Stat. Mech.*, page 10008, 2008.

[56] M. E. J. Newman. Modularity and community structure in networks. *Proceedings of the National Academy of Sciences*, 103(23):8577–8582, 2006.

[57] A.-L. Barabsi. *Network science*. Cambridge university press, 2016.

[58] R. Baeza-Yates and B. de A. N. Ribeiro. *Modern information retrieval*. New York: ACM Press; Harlow, England: Addison-Wesley,, 2011.

[59] C. Manning, P. Raghavan, and H. Schütze. Introduction to information retrieval. *Natural Language Engineering*, 16(1):100–103, 2010.

[60] M. E. Tipping and C. M. Bishop. Probabilistic principal component analysis. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 61(3):611–622, 1999.

[61] L. van der Maaten and G. Hinton. Visualizing data using t-sne. *Journal of machine learning research*, 9(Nov):2579–2605, 2008.

[62] L. Van Der Maaten. Accelerating t-sne using tree-based algorithms. *The Journal of Machine Learning Research*, 15(1):3221–3245, 2014.

[63] L. Bohlin, D. Edler, A. Lancichinetti, and M. Rosvall. *Community Detection and Visualization of Networks with the Map Equation Framework. In: Ding Y., Rousseau R., Wolfram D. (eds) Measuring Scholarly Impact.* Springer, Cham, 2014.

[64] J. Artiles, J. Gonzalo, and S. Sekine. The Semeval-2007 Weps evaluation: Establishing a benchmark for the web people search task. *Workshop on Semantic Evaluation*, 2007.

[65] L. Hubert and P. Arabie. Comparing partitions. *Journal of Classification*, 2:193–218, 1985.

[66] E. B. Fowlkes and C. L. Mallows. A method for comparing two hierarchical clusterings. *Journal of the American statistical association*, 78(383):553–569, 1983.

[67] M. Zappavigna. Searchable talk: the linguistic functions of hashtags. *Social Semiotics*, 25(3): 274–291, May 2015. doi: 10.1080/10350330.2014.996948.

[68] A.-L. Barabsi and R. Albert. Emergence of scaling in random networks. *Science*, 289:509–512, 1999.

[69] D. J. Watts, P. S. Dodds, and M. E. J. Newman. Identity and search in social networks. *Science*, 296(5571):1302–1305, 2002.

[70] M. Mitzenmacher. A brief history of generative models for power law and lognormal distributions. *Internet Mathematics*, 1(2):226–251, 2004.

[71] G. K. Zipf. The P1 P2/D hypothesis: On the intercity movement of persons. *American Sociological Review*, 11(6):677–686, 1946.

[72] D. A. Krueckeberg and A. L. Silvers. *Urban planning analysis: methods and models*. Wiley New York, 1974.

[73] S. Scellato, A. Noulas, R. Lambiotte, and C. Mascolo. Socio-spatial properties of online location-based social networks. *Fifth International AAAI Conference on Weblogs and Social Media*, 2011.

[74] M. Di Leo and E. Smith. *wo Californias: The Myths And Realities Of A State Divided Against Itself*. Island Press, 1980.

[75] M. Bucholtz, N. Bermudez, V. Fung, L. Edwards, and R. Vargas. Hella nor cal or totally so cal?: The perceptual dialectology of california. *Journal of English Linguistics*, 35(4):325–352, 2007.

[76] D. W. Meinig. The mormon culture region: Strategies and patterns in the geography of the American west, 1847-1964. *Annals of the Association of American Geographers*, 55(2): 191–219, 1965.

[77] W. Zelinsky. *The Cultural Geography of the United States*. Prentice-Hall, 1973.

[78] Colin Woodard. *American Nations: A History of the Eleven Rival Regional Cultures of North America*. New York, NY: Penguin Books, 2011.

[79] E. Han, P. Carbonetto, R. E. Curtis, Y. Wang, J. M. Granka, J. Byrnes, K. Noto, A. R. Kermany, N. M. Myres, M. J. Barber, K. A. Rand, S. Song, T. Roman, E. Battat, E. Elyashiv, H. Guturu, E. L. Hong, K. G. Chahine, and C. A. Ball. Clustering of 770,000 genomes reveals post-colonial population structure of North America. *Nature Communications*, 8:14238, 2017.

[80] G. D. Nelson and R. Alasdair. An economic geography of the united states: From commutes to megaregions. *PLOS ONE*, 11(11):e0166083, 2016.

[81] V. Gray, R. L. Hanson, and T. Kousser, editors. *Politics in the American States: A Comparative Analysis*. SAGE/CQ Press, eleventh edition, 2018.

[82] R. Johnston, D. Manley, and K. Jones. Spatial polarization of presidential voting in the United States, 1992-2012: The big sort revisited. *Annals of the American Association of Geographers*, 106(5):1047–1062, 2016.

[83] Pew Research Center. Political Polarization in the American Public, Jun 2014. URL http://www.people-press.org/2014/06/12/political-polarization-in-the-american-public/.

[84] S. A. Ercan and C. M. Hendriks. The democratic challenges and potential of localism: insights from deliberative democracy. *Policy Studies*, 34(4):422–440, 2013.

[85] M. Sales-Pardo, R. Guimerà, A. A. Moreira, and L. A. Nunes Amaral. Extracting the hierarchical organization of complex systems. *Proceedings of the National Academy of Sciences*, 104(39):15224–15229, 2007.

[86] B. H. Good, Y-A. de Montjoye, and A. Clauset. Performance of modularity maximization in practical contexts. *Phys. Rev. E*, 81:046106, 2010.

[87] M. Newman and M. Girvan. Finding and evaluating community structure in networks. *Physical review. E, Statistical, nonlinear, and soft matter physics*, 69 2 Pt 2:026113, 2004.

[88] M. Rosvall and C. T. Bergstrom. Maps of random walks on complex networks reveal community structure. *Proceedings of the National Academy of Sciences*, 105(4):1118–1123, 2008.

Figure Legends

Figure 1. Structure and fragmentation patterns of the network associated with human mobility. (a) Spatial degree centrality of the mobility network. Colors indicate the amount of people traveling at each location, measured by the logarithm of the degree centrality of each node (scale inset). The mobility network was used to generate communities using modularity optimization, with distinct colors indicating (b) 20 patches that can be visually associated to states or regions and (c) 206 smaller sub-communities within the communities of panel (b) that can be visually associated to urban centers.

Figure 2. Structure and fragmentation patterns of the network associated with human communication. (a) Spatial degree centrality of the communication network. Colors indicate the amount of communication at each location, measured by the logarithm of the degree centrality of each node (scale inset). The communication network was used to generate communities using modularity optimization, with distinct colors indicating (b) 15 patches that can be visually associated to states or regions and (c) 168 smaller sub-communities within the communities of panel (b).

Figure 3. Similarity of communities in the communication and mobility networks. Matrix of the regional communities for the communication network (y-axis, $n = 15$) and mobility network (x-axis, $n = 20$), ordered by decreasing overlap between communities. Cell colors represent the number of nodes overlapping between the two networks in each community, normalized by the size of the communities per row (scale inset), with no overlap indicated in white.

Figure 4. Multi-scale decomposition of the mobility network. Colors indicate geographical patches detected in the mobility network for values of the resolution parameter γ varied from $0.05 - 20$ (upper left to bottom right). M stands for modularity.

Figure 5. Multi-scale decomposition of the communication network. Colors indicate geographical patches detected in the communication network for values of the resolution parameter γ varied from $0.05 - 20$ (upper left to bottom right). M stands for modularity.

Figure 6. Similarity between the mobility and communication networks across multiple scales. Similarity is measured by the average of the Purity, Adjusted Rank, and Fowlkes-Mallows Indexes (color scale shown). Scale is defined by the different values of the resolution parameter γ (horizontal and vertical axes).

Figure 7. t-SNE analysis of hashtag use by location and cosine similarity of communities based on hashtags. Panel (a) shows the results of t-SNE analysis on the first 100 components of PCA analysis of hashtags in locations of the mobility network. Panels (b) shows cosine similarity of hashtag use in the communities of the mobility network at $\gamma = 1$. Panels (c) and (d) show the corresponding t-SNE result and cosine similarity for the communication network. Colors match those of the communities in Figures 1 and 2, panel (b).

Figure 8. Spatial degree distribution and modular structure for model simulations with different parameter values α and β and a fixed value of $\nu = 0.1$ (see text for parameter definitions). Top panels show the spatial degree distribution (from weakly connected in blue to highly connected in red). Bottom panels show the modules of each graph; each color identifies a single community.

Figure 9. Kolmogorov-Smirnov score and modularity for simulations with varying model parameter, α and β, and fixed resolution parameter $\nu = 0.1$. (a) Colors indicate the Kolmogorov-Smirnov (K-S) score, with lower scores (red) indicating similarity between the degree distributions of the model and the mobility network. (b) Colors indicate network modularity. Modularity is highest at around 0.8 (dark red), similar to the actual modularity for the mobility network (0.83).

Supplemental Material

S1.1 Contribution of states to communities in the mobility and communication networks

To further quantify how state borders contribute to communities, we counted the number of nodes for each community in each state and normalized to the size of the largest community. Figure S1 shows the contribution of states to communities for both the mobility and communication networks with resolution $\gamma = 1$. For many of the states, the borders of states match their communities, while other states have administrative borders that deviate from the ways people travel and communicate.

S1.2 Dissimilarity between communities in distance-based networks and real networks

The strength of communications between people has an inverse relationship with their distance from each other. Previous research has shown the relationship is inversely proportional to distance or distance squared [12, 13]. To investigate the effect of distance or distance squared on the formation of communities, we re-weighted the links in the mobility and communication networks based on the distance between any two connected locations.

Figure S2 shows the detected communities in both networks with links that are rearranged to be equal to the inverse of distance $1/d_{ij}$ between locations i and j. Panels (a) and (b) in this figure depict the communities, and panels (c) and (d) quantify the similarity of the communities. Although communities are geo-fragmented (36 communities in the mobility network and 30 communities in the communication network), they are not similar to the communities in the real networks. In fact, in some areas, communities almost overlap with each other, but in most areas, the artificial networks end up with smaller patches that do not match with any real community. Figure S3 shows the results for the artificial networks with links that are created based on $1/d_{ij}^2$ between two connected locations i and j. This dissimilarity between the communities in the distance-based networks and the real networks is evidence of the impact of other parameters beside distance.

S1.3 Stability of communities and overlapping communities

In Figure S4, we quantify the stability of detected communities and identify areas in which communities overlap with each other. In the Louvain method, communities refer to the regions in which nodes are more connected to each other than the rest of the network. However, due to the possible local minima in the Louvain algorithm, some nodes may be detected as part of other communities in the next run [85, 86]. We generated an ensemble of 120 realizations to quantify how much the detected communities are stable across the realizations and in which areas communities overlap. In panels (a) and (b) of Figure S4, the color bar shows how stable a location is in a given community as a percentage of realizations. Panel (a) shows that in many areas, communities are 100% stable, and all the areas have at least 40% stability. In some areas, larger communities split into two communities. For example, Georgia (GA) and Alabama (AL) encompass one community in some realizations, while in others, AL splits off as a separate community. Similarly, Indiana (IN), Kentucky (KY) and Tennessee (TN) form a single community in some realizations, while in others, IN manifests as a separate community. In other cases, areas sometimes exhibit a small overlapping part between two stable communities, such as the small section in eastern New York State (NY)

that sometimes appears as part of the red community in Figure 1. Connecticut (CN) is the state that has the largest instability. In Figure 1, it is part of the New England community (purple community), but in many realizations it appears as part of the New York City and surrounding area community (red community). In the communication network (panel (b)), some overlapping areas are the size of multiple states, demonstrating that sometimes a whole state appears in another community.

In panels (c) and (d) of Figure S4, we count the number of connections outside the community for each location. In panel (c), black spots are the locations that only have inside community connections, which tend to be suburban and rural areas around the cities. Red spots represent locations with more than 100 outside community connections, which are city centers. Roads are also clear in the figure, as people tweet on the road far from the locations where they spend most of their time. In panel (d), the number of black spots decreases while the red areas increase as compared to panel (c). These differences show the larger distances that people communicate with each other versus physical travel. Red spots in these two panels have a higher chance to appear as part of another community, especially ones that are on the border of two communities.

Panels (e) and (f) show the frequency of detected communities over the 120 realizations for both networks. For the mobility network, all realizations generated a range of $19 - 22$ communities, with more than 80% of realizations having 20 communities. For the communication network, the range was $14 - 17$ communities, with more than 80% of realizations having 15 communities. This shows the relative stability of the network, with only a few communities that were likely to split into two.

S1.4 Comparing communities detected by the modularity optimization and Infomap methods

In Figure S5, we show the communities detected by the Infomap method with communities detected by the generalized modulariy optimization method for (a) mobility and (b) communication networks. Compared to the modularity method in which communities represent connected areas that deviate the most from a null model [87], the Infomap method is based on the flow of information within the network, and communities represent the areas a random walker tends to stay in for a long time [63]. The algorithm of community detection with Infomap is very similar to the Louvain method [88, 63]. It starts by considering each node as a single module and joining the neighboring nodes into supermodules in a random sequential order. Nodes move to their neighboring modules to reach the largest decrease in the description length given by a map equation. If the movement of a node to a neighboring module does not reduce its description length, it stays in its own module. This procedure is repeated by a new sequential order in each time step, until movements do not decrease the map equation. Movements in each level start from the formed modules in the previous level. This hierarchical procedure continues until reaching a minimum value for the map equation.

The Infomap method detects 512 communities in the mobility network (Figure S5-(a)) and 459 communities in the communication network (Figure S5-(b)). Most of the communities reflect cities and their suburbs. However, in the communication network, the communities are larger, with some in the size of the states. In panel (c) of Figure S5, we compare the similarity of detected communities for the Infomap and modularity methods. The x-axis shows resolution parameters for the modularity method. As a measure of similarity, the y-axis shows the average value of

the three scores, Purity, Adjusted Rank and Fowlkes-Mallows Indexes. The highest similarity of communities in the mobility network is about 55 percent and occurs for resolution parameters of 20 to 40. Meanwhile, the similarity of communities in the communication network is larger at about 60 percent around resolution 8 but then decreases sharply with increasing resolution parameters.

S1.5 Structural similarities of the mobility and communication networks

We compare structural properties of the mobility and communication networks by means of centrality measures, edge weights and multi-scale structure. The structural properties of both networks are consistent with each other, showing an interplay between how people explore the physical space and communicate on Twitter.

In Figure S6 (a), we show a scatter plot of degree centrality for each location in both networks at $\gamma = 1$, colored by their eigenvector centrality in logarithmic scale. While most locations are poorly connected, a few of them have an extremely high degree, corresponding to densely populated areas in large cities. Locations with a higher degree centrality in both mobility and communication networks also have high eigenvector centrality, which means that these locations are central relative to where information flows.

Next, we compare edge weight and length for both networks in Figure S6 (b). The edge length is estimated as the geographical distance between the locations' centroids. Edges that have high weights also have small lengths, reflecting daily, short distance travels seen in cities and localized communication.

Finally, we compare the multi-scale structure of the network fragmentation. Social fragmentation can be seen at multiple scales using the generalized modularity optimization method. This method seeks partitions at various scales by considering the resolution parameter γ. We compare the two networks at roughly similar partition sizes (mobility and communication, respectively, at γ=0.1 and 0.3, γ=0.2 and 0.6, γ=0.6 and 0.9, or γ=0.7 and 1.0). These pairs of γ values represent partitions in both networks with the highest similarities (see Figure 6 in the main text). In Figure S7, we show the partitions comparison using an alluvial diagram for each set of γ values (panels). The alluvial diagrams map the corresponding number of nodes at each module of the mobility network (left axis) onto each module of the communication network (right axis).

S1.6 Comparing t-SNE analyisis in real dataset with the randomized dataset

To validate the clusters we see in the t-SNE analysis of the real data, we performed t-SNE on a randomized dataset (see Figure S8). In the randomized dataset, we kept the number of instances of each hashtag the same but randomized the location for each hashtag use. We compared the distribution of the distances between locations per community in the t-SNE space for the real dataset to those of the randomized dataset. The patterns we see in the t-SNE of the real dataset were statistically different from those of the randomized dataset for all communities ($p < 0.001$).

S1.7 Additional results from varying the model parameters

In Figure S9, we show the effect of changing the spatial growth term ν in the network model. This term gives preference to locations in which the average degree of nearest neighbors is higher. We set the α and β exponents at 0.5 and 1.5, respectively, which generates geographical fragmentation

patterns even with $\nu = 0$ (left panels). As shown in figure S9 (a), increasing the ν exponent (from left to right panels) concentrates the connections around hotspots, recreating the growth of cities. However, Figure S9 (b) and (c) show that while small values of ν increase the number of communities and their modularity, large values of ν lead to less cohesive borders between communities and a decrease in modularity.

Next, we investigated how changing all three parameters in the model causes deviations between the model and the real data. We used the Kolmogorov-Smirnov (K-S) test, a measure of similarity between two distributions, to determine the similarity of the network degree distributions, which are a measure of network connectivity. Degree distributions for the model were calculated by averaging over 20 realizations for each set of parameters. Figure S10 represents the similarity of degree distributions of locations for the model versus the empirical mobility network. Moving from the top left panel to the bottom right panel, we see that the degree distributions for the model and the mobility network data deviate from each other as the ν exponent increases.

We also determined how changing the model parameters affects the modularity of the network. Figure S11 shows the average modularity from 20 realizations for different values of the three exponents. Overall, increasing the strength of the preferential attachment process (controlled by α) and the spatial growth process (controlled by ν) destroys geographical patches and reduces modularity. Conversely, the human mobility gravity process (controlled by β) is the only exponent that increases these fragmented geographical patches. Note that modularity is 0.83 for the mobility network from the Twitter data.
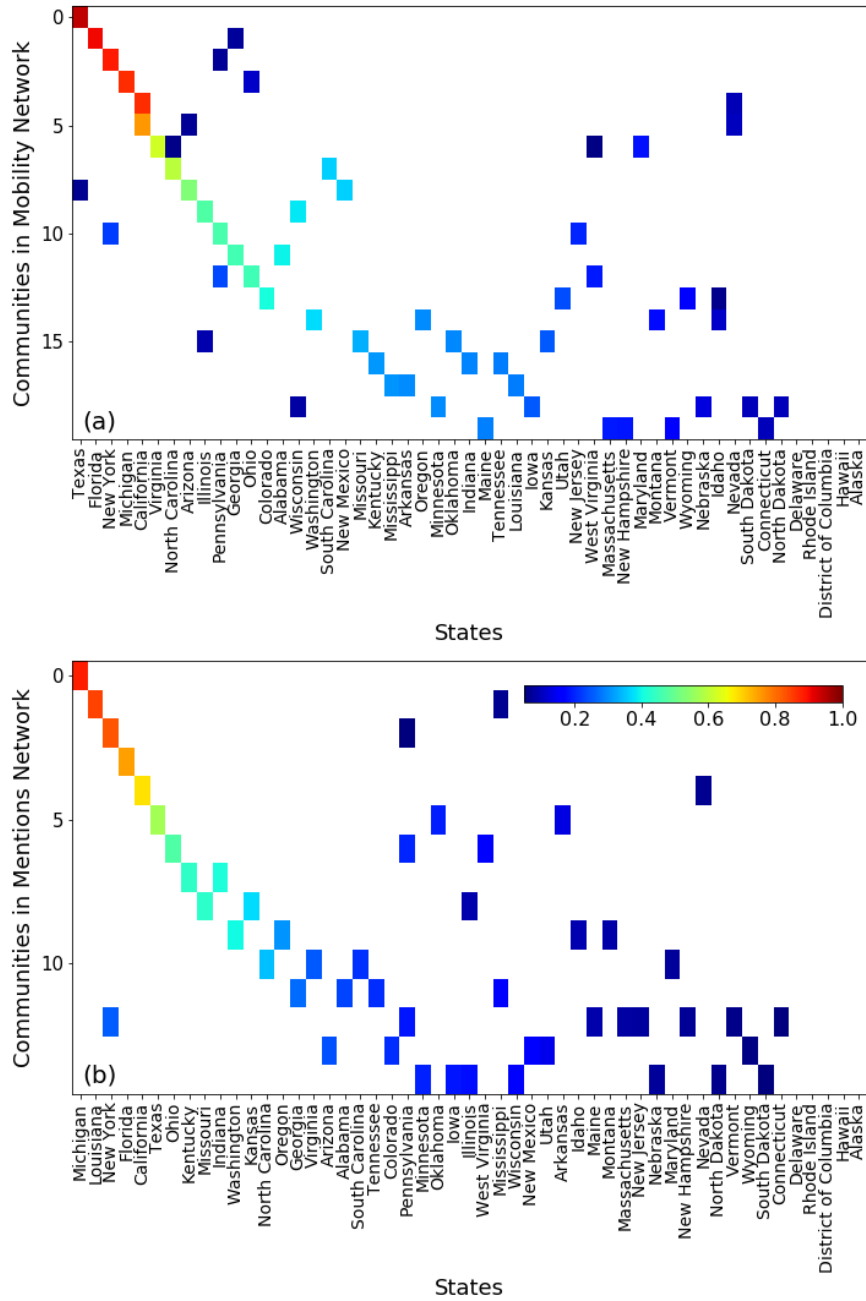
Figure S1: Contribution of states to communities. Graphs depict the number of communities detected (y-axis) in each state (x-axis) for (a) mobility and (b) communication networks generated with $\gamma = 1$. Colors (scale inset, panel (b)) indicate overlap between state and community boundaries, with red indicating greatest overlap.
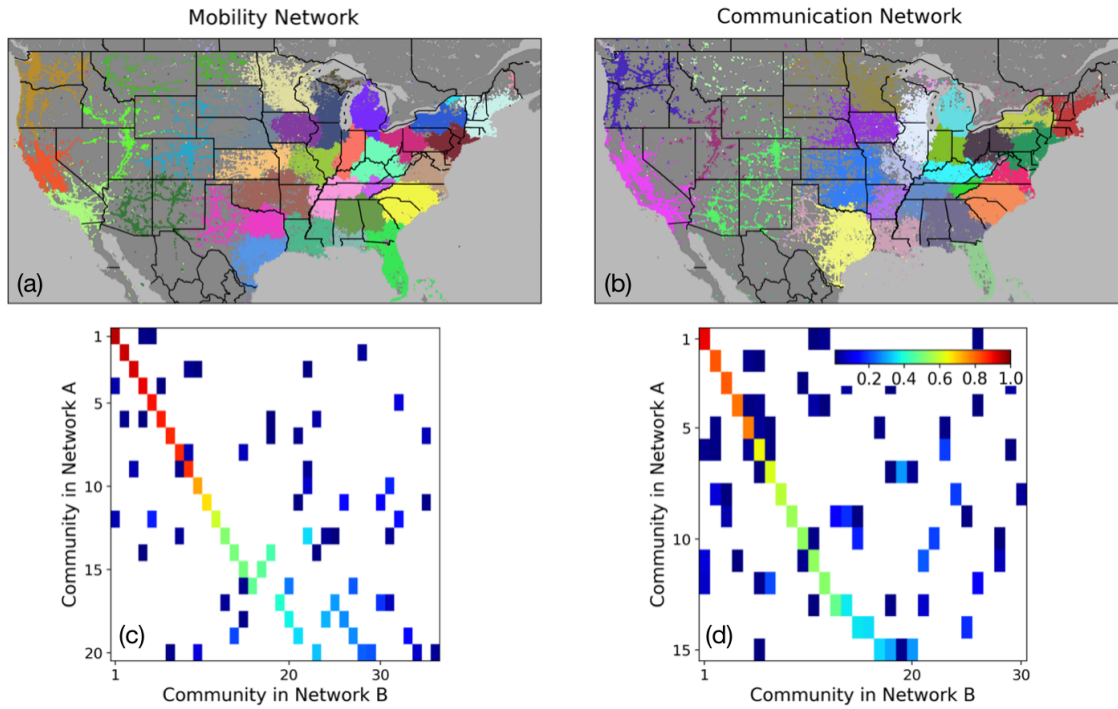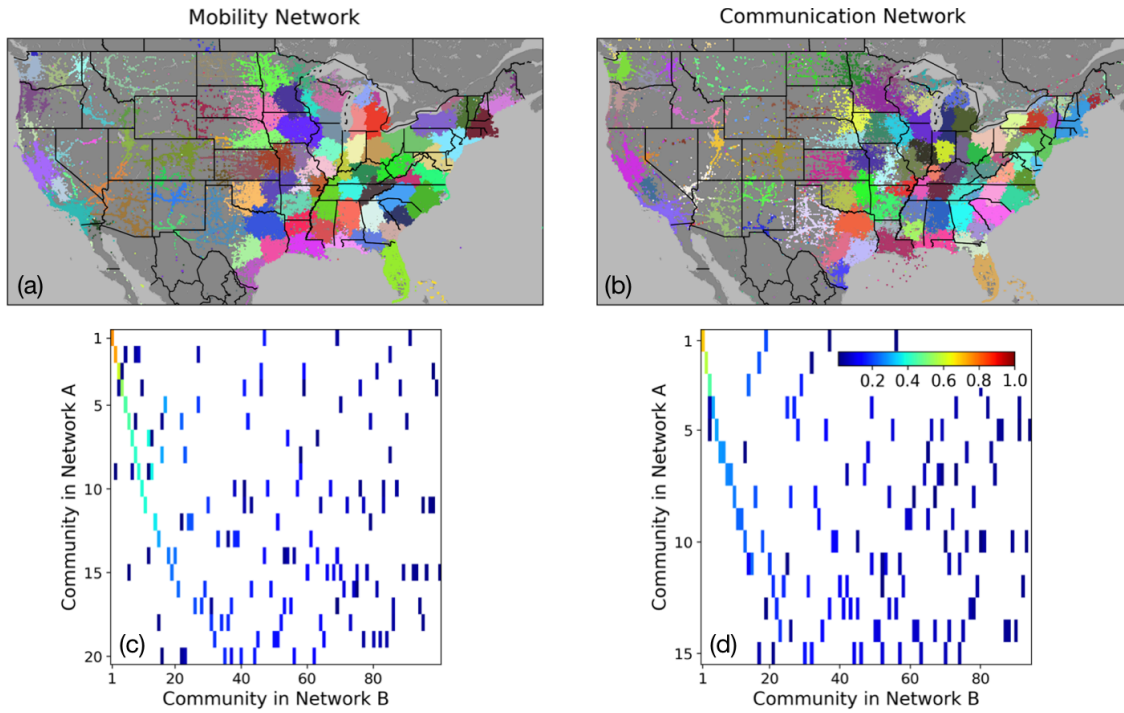
Figure S2: Dissimilarity between communities in real networks (mobility and communication) and artificial ones in which the weight of the links represents the inverse distance between the connected locations i and j, $1/d_{ij}$. Panels (a) and (b) show the detected communities in the artificial mobility and communication networks, respectively. Panels (c) and (d) show the overlap of communities between real and artificial networks of mobility and communication, respectively. The x-axes show the communities in the real networks (labeled "Network B"), and the y-axes show the communities in the artificial networks (labeled "Network A"). Communities are ordered by decreasing overlap. Cell colors represent the number of nodes overlapping between the two networks in each community, normalized by the size of the communities per row (scale inset), with no overlap indicated in white. Although the communities from artificial networks are geo-fragmented, they do not match the communities detected in real networks.

Figure S3: Dissimilarity between communities in real networks (mobility and communication) and artificial ones in which the weight of the links represents the inverse square distance between the connected locations i and j, $1/d_{ij}^2$. Panels (a) and (b) show the detected communities in the artificial mobility and communication networks, respectively. Panels (c) and (d) show the overlap of communities between real and artificial networks of mobility and communication, respectively. The x-axes show the communities in the real networks (labeled "Network B"), and the y-axes show the communities in the artificial networks (labeled "Network A"). Communities are ordered by decreasing level of overlap. Cell colors represent the number of nodes overlapping between the two networks in each community, normalized by the size of the communities per row (scale inset), with no overlap indicated in white. Although the communities from artificial networks are geo-fragmented, they do not match the communities detected in real networks.
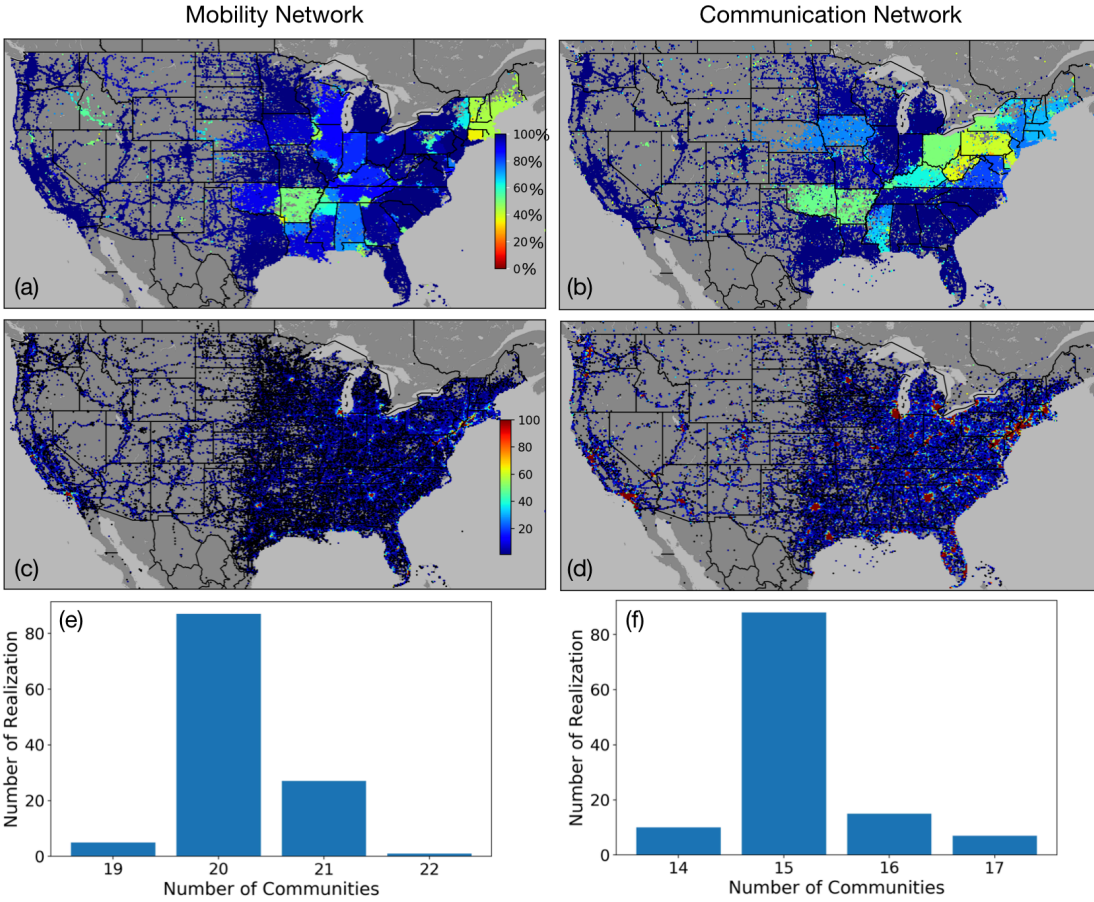
Figure S4: Stability and overlap of detected communities in mobility and communication networks. We created 120 realizations for detection of communities for each network. Panels (a) and (b) show the the stability of locations to their communities. The color bar (inset) represents the percentage at which a location appears in the same community over all realizations; blue areas are the most stable communities. Panels (c) and (d) show how many connections a location makes with locations outside its community in the mobility and communication networks, respectively. The color bar (inset) indicates the number of connections with outside areas, with red areas having more than 100 outside connections and black areas having no outside connections. Panels (e) and (f) show the frequency of detected communities for all realizations; 20 and 15 communities are the most frequent number of detected communities in the mobility and communication networks, respectively.
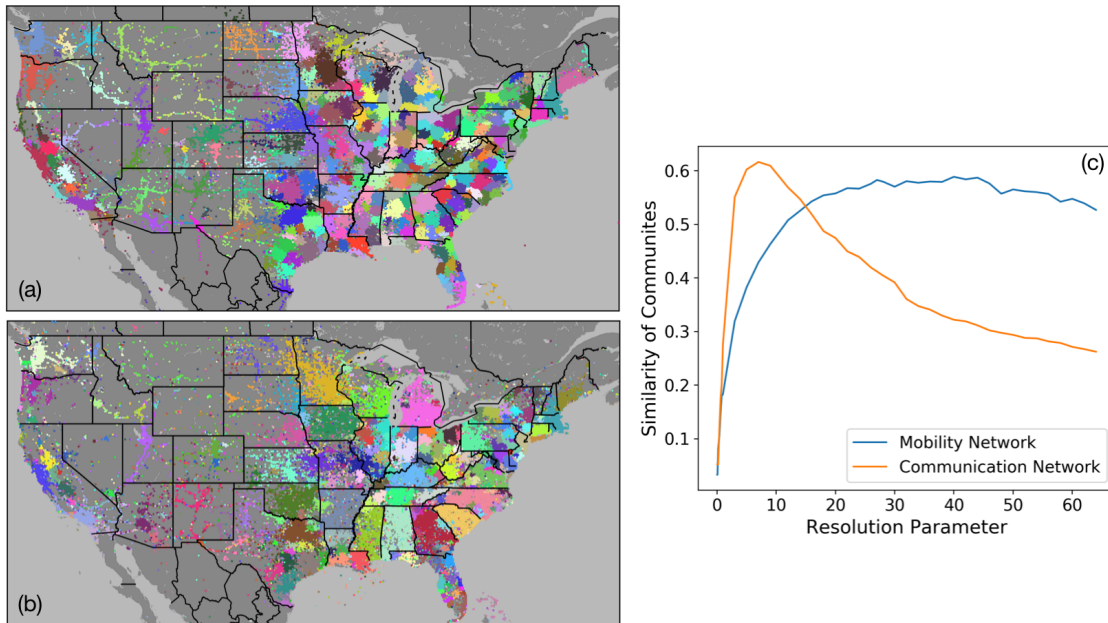
Figure S5: Detected communities in (a) mobility and (b) communication networks using the Infomap method. Applying the Infomap method gives smaller communities that mostly represent the city areas. Some detected communities in the communication network represent states. Panel (c) shows the similarity of detected communities using the Infomap method with the communities detected using the modularity optimization method (y-axis) at different resolution parameters (x-axis). To measure similarity of the communities, we used the average value of the three scores, Purity, Adjusted Rank, and Fowlkes-Mallows Indexes. Detected communities in the mobility network have more that 55 percent intersection over the resolution parameters 20-40. In the communication network, detected communities have the most similarity at resolution 8.
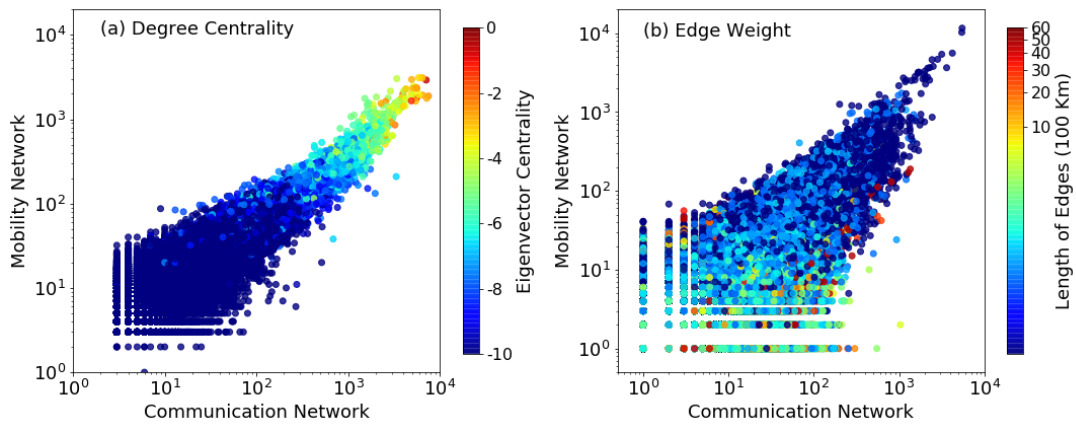
Figure S6: Similarity of network structure between the mobility and communication networks at $\gamma = 1$. Panel (a), scatter plot of degree centrality for each location in the mobility (y-axis) and communication (x-axis) networks, colored by the corresponding eigenvector centrality (scale on right). Panel (b), scatter plot of the edge weights for each location in the mobility (y-axis) and communication (x-axis) networks, colored by the edge length or distance between nodes (scale on right). Axes and color bars are in logarithmic scale.

11
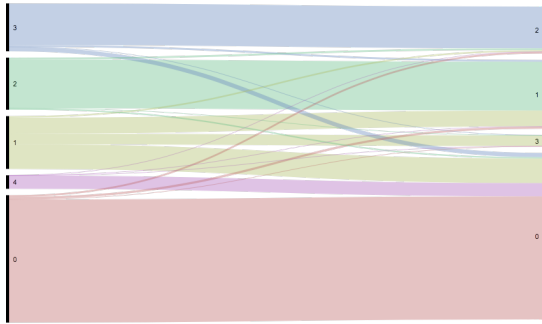
Mobility $\gamma=0.1$ - Communication $\gamma=0.3$
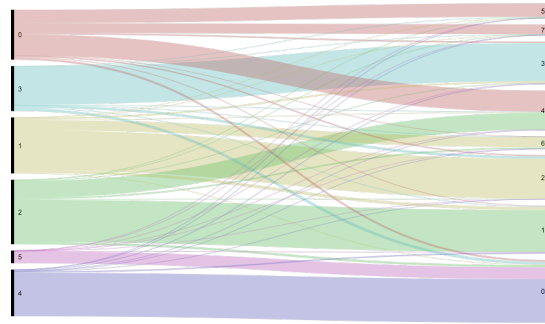
Mobility $\gamma=0.2$ - Communication $\gamma=0.6$
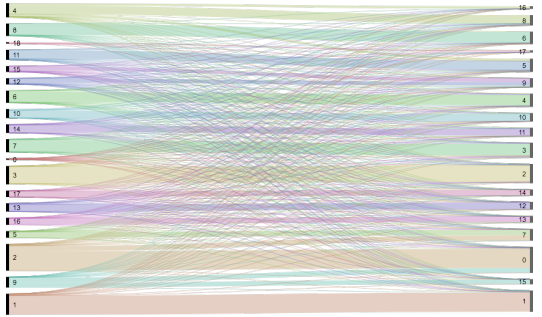
Mobility $\gamma=0.6$ - Communication $\gamma=0.9$
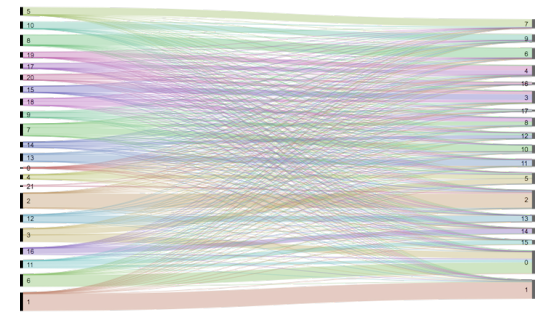
Mobility $\gamma=0.7$ - Communication $\gamma=1.0$

Figure S7: Comparison of the mobility and communication networks at γ values with highest similarity. In each panel, an alluvial diagram maps similarities of detected communities between the mobility (left axis) and the communication (right axis) networks. Values of γ are chosen from the four darkest red cells in Figure 6, which yield the highest similarity between the two networks.
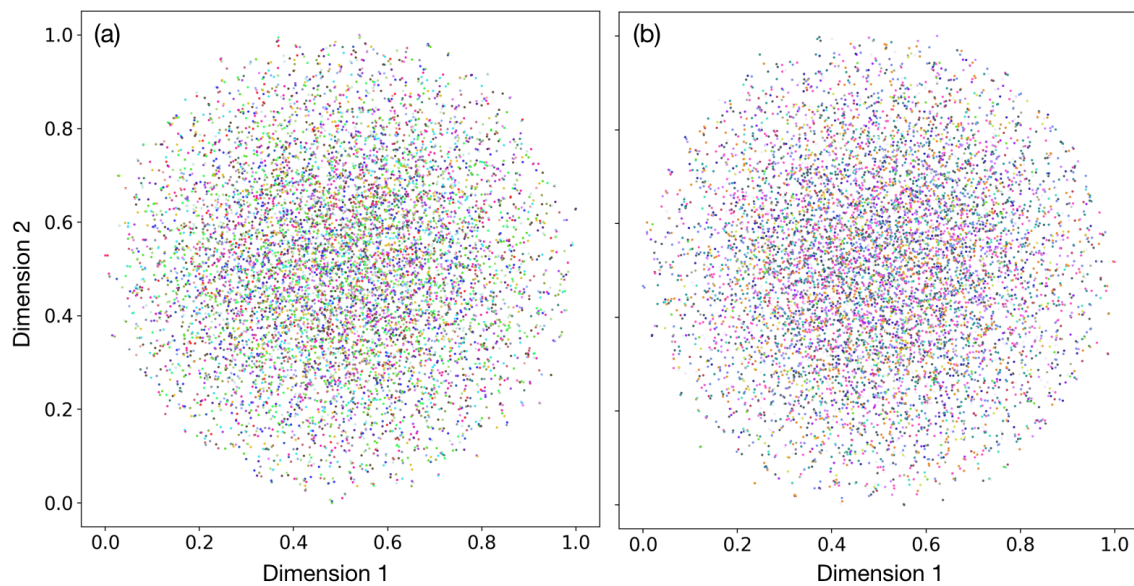
Figure S8: t-SNE analysis of hashtag use for a randomized dataset. Panel (a) shows the results of t-SNE analysis on the first 100 components of PCA analysis of hashtags in locations of the mobility network. Panel (b) shows the corresponding t-SNE analysis for the communication network. Instances of hashtag use have been preserved from the original datasets, but the locations have been randomized. Colors match those of the communities in Figures 1 and 2, panel (b).
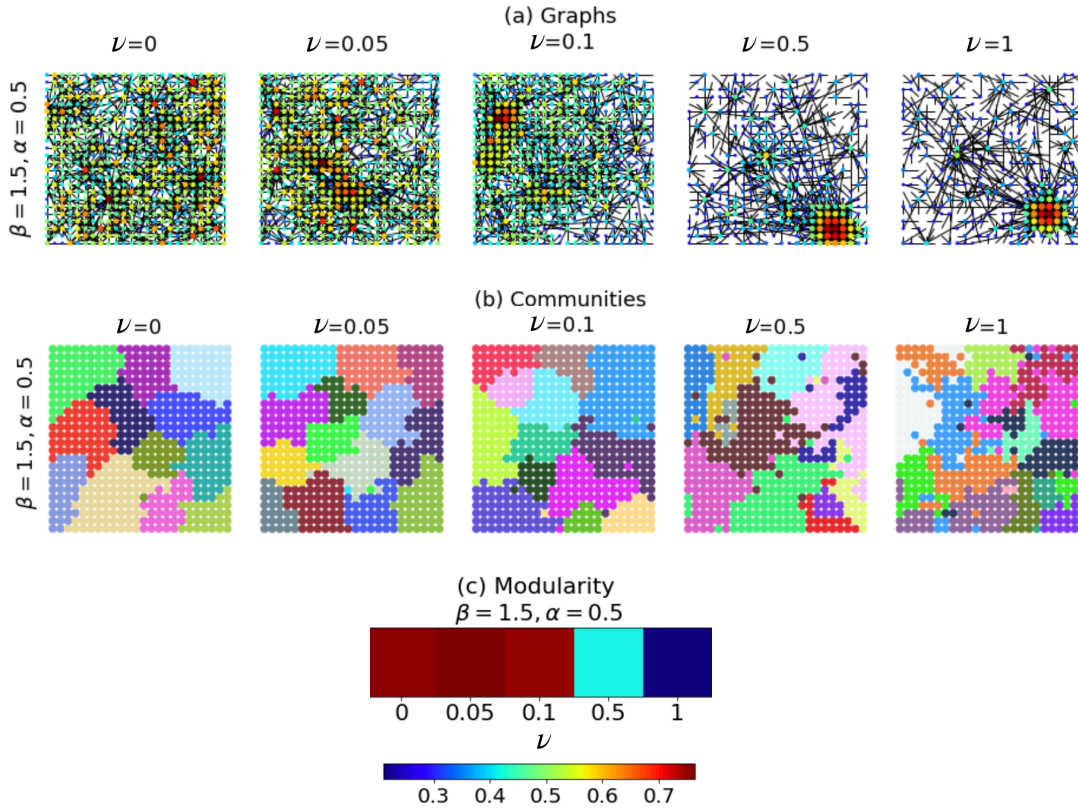
Figure S9: Degree centrality, fragmentation, and modularity for simulations with varied values of the resolution parameter ν (spatial growth). The value of ν is varied from $0, 0.05, 0.1, 0.5, 1$ (left to right), while $\alpha = 0.5$, $\beta = 1.5$, and the size of the lattice is 576 locations. (a) Spatial degree centrality. Nodes are colored by their degree centrality (from blue to red), and edges are plotted in black. (b) Spatial patches, shown in varying colors. (c) Modularity as a function of ν, indicated by color (scale below figure). Overall, increasing ν from 0 to 0.05 increases the connections between hotspots and the modularity of communities, but as ν increases further to 0.5 or 1, the modularity decreases and borders between communities become less clear.
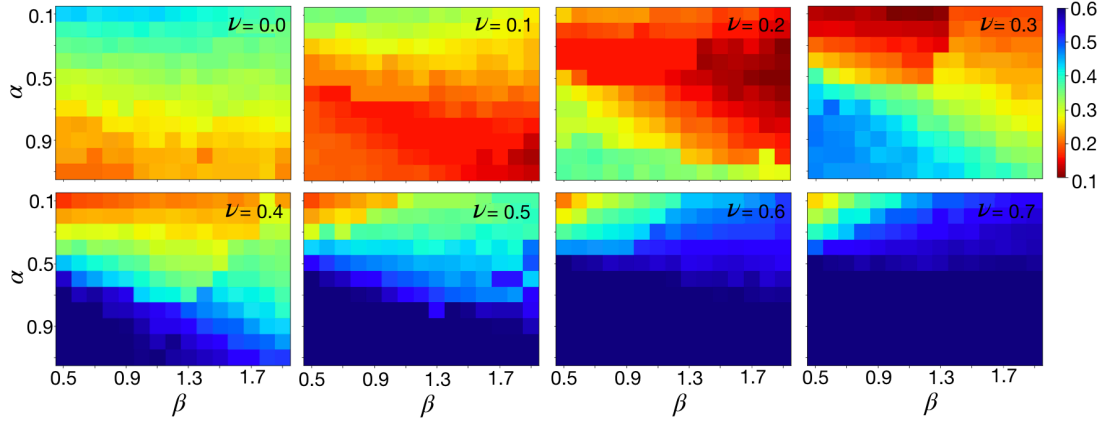
Figure S10: Similarity of degree distributions between simulated and real mobility networks. Matrices show different values of the parameters α (y-axis), β (x-axis) and ν (panels, upper left to lower right). Kolmogorov-Smirnov (K-S) scores are depicted with colors (scale on right), with the lowest K-S values in red, indicating that the degree distributions are similar between simulated and real mobility networks.
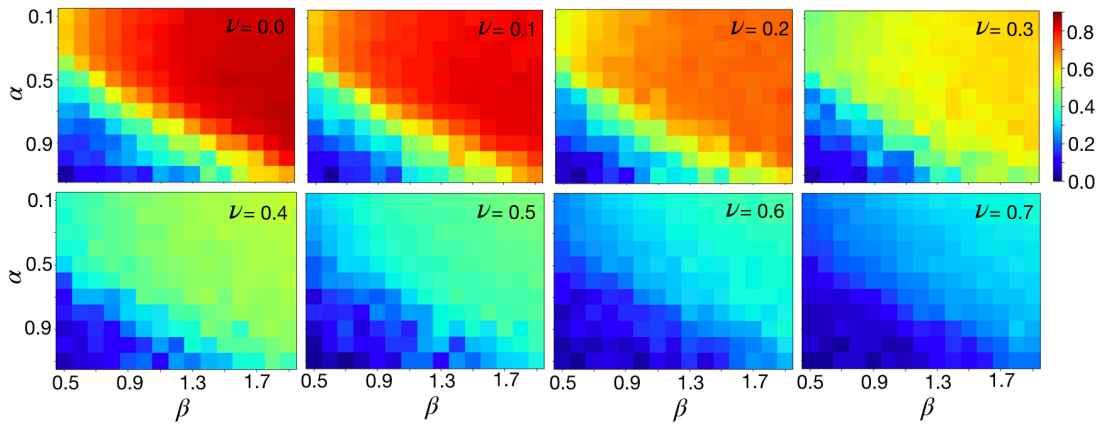

Figure S11: Modularity of detected communities in simulations with varying model parameters. Matrices show different values of the parameters α (y-axis), β (x-axis) and ν (panels, upper left to lower right). Modularity values are depicted with colors (scale on right), with the highest values in red at ~ 0.9. Note that modularity for the mobility network is 0.83.