Chaos

Geographical fragmentation of the global network of Twitter communications

Cite as: Chaos **30**, 073133 (2020); https://doi.org/10.1063/1.5143256
Submitted: 24 December 2019 . Accepted: 24 June 2020 . Published Online: 21 July 2020

Leila Hedayatifar, Alfredo J. Morales, and Yaneer Bar-Yam

COLLECTIONS

F This paper was selected as Featured

Chaos ARTICLE scitation.org/journal/cha

Geographical fragmentation of the global network of Twitter communications

Cite as: Chaos 30, 073133 (2020); doi: 10.1063/1.5143256
Submitted: 24 December 2019 · Accepted: 24 June 2020 ·
Published Online: 21 July 2020

View Online Export Citation CrossMark

Leila Hedayatifar,[1,a)] Alfredo J. Morales,[1,2] and Yaneer Bar-Yam[1]

AFFILIATIONS

[1] New England Complex Systems Institute, 277 Broadway, Cambridge, Massachusetts 02139, USA
[2] MIT Media Lab, 75 Amherst St. Cambridge, Massachusetts 02139, USA

[a)] **Author to whom correspondence should be addressed:** leila@necsi.edu

ABSTRACT

Understanding the geography of society represents a challenge for social and economic sciences. The recent availability of data from social media enables the observation of societies at a global scale. In this paper, we study the geographical structure of the Twitter communication network at the global scale. We find a complex structure where self-organized patches with clear cultural, historical, and administrative boundaries are manifested and first-world economies centralize information flows. These patches unveil world regions that are socially closer to each other with direct implications for processes of collective learning and identity creation.

The invention of the Internet rapidly enabled unprecedented global connectivity and communication independent of physical boundaries. People, however, do not seem to be communicating with each other homogeneously. Mapping the organization of social systems is crucial for our understanding of society as well as for effective policies involving ethnic groups and economic strata. Here, we study global patterns of human communication using Twitter data and characterize the geographic structure of emergent social networks. In this network, edges between geographical locations 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 communicate between two locations. The betweenness centrality shows that the most central nodes in the network are cities, with seven of the top ten located in Europe and the USA. The distribution of distances between users shows three primary peaks, at 5 km, 300–500 km, and 10 000 km, manifesting social interactions at intercity, international, and intercontinental scales. Using community detection analysis, we show that the Twitter communication network has a multiscale community structure. The dominant geographic area of a community may contain a single country or several that are grouped by either cultural or linguistic similarities. There are also links to other locations, perhaps due to travelers or individuals with specific social or cultural ties, or interests that connect them to the dominant domain. We demonstrated that economic relations, common languages, common borders, and colony/colonizer history have statistically significant effects on the associations of countries in the Twitter communication network. More specifically, we found that countries that had a common colonizer have a decreased probability of interaction, suggesting that the hierarchical interactions with the colonizer inhibit lateral interactions among the colonies. We discuss multiple specific interesting features of community structures that may reflect the historical and current state of international relationships.

I. INTRODUCTION

The introduction of the Internet rapidly enabled unprecedented global connectivity. In about two decades, communication became instantaneous, affordable, and independent of physical barriers. People, however, do not seem to be communicating with each other homogeneously. Online communication networks appear to be polarized and echo-chambers are manifesting for political[1–5] and non-political[6–9] reasons. Understanding the emergent structure of information flows on the Internet and its implications for the complexity of society is a challenge for scientists and technology developers. Mapping the organization of social systems is crucial for effective policy making involving ethnic groups[10] as well as economic growth.[11]

Modular structures are a typical form of self-organization in social networks.[12] From groups of friends, co-workers and families, social networks show clusters of highly connected individuals that are weakly linked to other clusters.[13] The aggregation of these clusters can span from cities up to national regions.[7] Geographical patches of social systems have been seen through human mobility[14–19] and communication[20–26] data. Natural, cultural, and administrative borders play an influential role in such structures.[27] Given the difficulties of obtaining census or survey data about human activity on a global scale, studies are often focused on national and sub-national scale.

The recent availability of social media data enables the observation of social activities around the world, including those associated with social fragmentation. By interacting with each other on social media, people reveal part of their preferences and provide information about the behavior of the social system at large.[28] While individual behaviors are highly heterogeneous and varied,[29] aspects of such behaviors are collective and shared among the members of the social group. Collective behaviors are linked to the set of interdependencies that lead to emergent properties in complex systems.[30] The aggregation of collective behaviors enables the observation of the structure and functioning of social systems at multiple scales.[31,32]

Considering the multiscale structure of the social space is crucial for designing policies and interventions that effectively respond to the challenges presented at each scale.[33,34] The multiscale structure of social ties determines outcomes of social challenges that include organizational performance,[35] the spread of news and information,[36] political polarization,[4,37] the ethnic composition of countries,[20] the synchronization of global regions,[38] and international trade.[11] The fragmentation of social networks influences the channels where information flows[13] and consequently the behaviors that arise in social groups,[37,39] including political or physical conflict.[40–42]

The way information flows among people is fundamental for understanding the behavior of social systems.[13] People learn by imitation during processes of collective learning,[43] and therefore, the information they receive influences behaviors and beliefs.[44–48] Social phenomena such as innovation,[49] business and cultural activities,[50] or crime[51] are directly influenced by the structure of social ties and the properties of the social space. Evidence of such influence is found in the geographic distribution of economic activities.[11] The location of industries, the structure of international trade, and the growth of economic complexity is deeply linked to the spatial structure of social networks where knowledge spreads with clearly defined geographic patterns.

Here we study global patterns of human communication using Twitter data and measure the geographical structure of emergent social networks. We analyze over 50×10^6 geolocated tweets posted in December 2013 from all around the globe. We show that the Twitter communication network has a multiscale community structure whose modules are hierarchically organized and span from parts of a city up to whole nations and intercontinental regions. We show that the structure of this network is influenced by geographic proximity as well as by cultural, historic and economic factors.

II. MATERIALS AND METHODS

A. Data

We use geo-located Twitter data to generate geographical networks based on where people communicate with one another using the "mentions" mechanism. We consider over 50×10^6 geo-located tweets posted in December 2013 from all around the globe. The tweets were collected using the Twitter Streaming Application Programming Interface (API), which provides over 90% of publicly available geo-located tweets in real time.[52] Geo-located tweets contain precise latitude and longitude coordinates at the moment of their creation. Twitter activity has been previously analyzed to understand patterns of global synchronization,[38,53] international human mobility,[54,55] languages and ideologies,[56,57] emergency situations,[58] and the fragmentation of social systems due to politics,[4,59] interests,[60] happiness,[61] and geography.[7] Other sources of data such as Facebook have also shown typical properties of social networks such as heterogeneous degree distributions, and short distances between users and modular structures with geographic features.[62]

We analyze the way people communicate with each other by looking at the exchange of tweets using "mentions." This interaction mechanism consists of explicitly including (mentioning) someone's @username in the text of a tweet. A user whose @username is mentioned is sent an alert by the system including the text of the tweet. By mentioning each other, users are intending to send direct information to other users on the online platform. The analysis does not differentiate positive or negative messages but rather intended information transfer. Our dataset includes messages that contain at least one mention to another user and also retransmitted messages (or retweets) that either contain user mentions or use the "RT @username" notation to indicate that the tweet is not original. Other interaction mechanisms such as "likes" are not captured in our sample dataset.

Twitter users are known to be potentially biased toward younger and more urban populations.[63–65] However, the unique ability of the public Twitter platform to represent large-scale dynamics across a large proportion of the world makes its study worthwhile. Despite possible limitations and biases in Twitter samples,[63,64] researchers have used Twitter data on mobility,[66] communication,[26,58] and immigration patterns[65] to study the underlying structure of countries and the existence of geo-located communities. In a previous study,[7] we showed that human mobility and communication networks extracted from Twitter data show consistent geo-located communities, which are highly similar to each other across multiples scales of observation. The patterns observed on Twitter data are very similar to those extracted from other sources of data, including mobile phones[17,21] and postal activity.[15]

Twitter usage has been reported to vary across different countries.[67] The penetration rate of Twitter increases with gross domestic product (GDP) and the platform is banned in many countries. While these facts limit the scope of our study, the communities we observe show countries that are more tightly connected with each other within the online platform and are consistent with previous studies of human mobility using the same type of data.[67]

B. Networks

We analyze global communication patterns by building geographical networks. Nodes represent a lattice of 1° latitude $\times 1^\circ$ longitude cells overlaid on a map of the world. Each cell is approximately 100 km wide. There are about 8000 cells comprising inhabited areas across the globe. Network edges reflect communication on Twitter via the mentions mechanism. In this 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 communicate between locations i and j. There are about 730 000 edges in the network. The network aggregates the heterogeneities of human activities in a large-scale representation of social collective behaviors.[31] We also performed and compared results to studies on a finer resolution network of nodes for a lattice of 0.1° latitude $\times 0.1^\circ$ longitude.

C. Community analysis

We apply the Louvain method[68] to the communication networks built from Twitter data in order to analyze their community structure. We consider undirected networks to create the communities. To turn the network into an undirected one, we sum the weights of the directed edges. No distinction is made for the direction of action. The Louvain method initially considers each node as a single community and maximizes the metric modularity (M) by merging nodes and communities until finding an optimal partition. Modularity is a scalar value, $-1 < M < 1$, that quantifies the distance of the detected communities in comparison with a random layout. Positive and high values of M (ideally above 0.3) show optimal partitions of the network.[69–71]

We study communities at multiple scales by applying a generalized version of the modularity optimization algorithm,[72] which controls for the coarseness of the communities with a resolution parameter γ. The conventional modularity equation uses $\gamma = 1$. If $\gamma < 1$, larger communities are prevalent. If $\gamma > 1$, smaller communities appear. The method has multiple maxima, and we choose partitions that are robust to multiple runs of the algorithm.

The concept of network fragmentation is commonly used to describe processes of network dismantling.[73,74] Here, we use it to represent the modular structure of society due to the presence and absence of communication links. This is consistent with the terminology from previous community detection methods such as the Girvan–Newman method.[75]

III. RESULTS

A snapshot of the Twitter global communication network is shown in Fig. 1 (top panel). Edges have been colored in yellow and represent user interactions between two locations (see Sec. II). Large areas such as Europe, the USA, Brazil, or South East Asia are densely connected inside and weakly connected to other areas, which shows the geographic structure of strong and weak ties at the global scale. While strong ties keep groups cohesive, weak ties integrate groups at the large scale and are responsible for the spread of information system wide.[13] The active "mention" interactions shown in Fig. 1 occur through a much smaller set of social contacts[76] compared to the more common passive *friends* or *followers* relationships.[62]

The global weak ties have a characteristic structure in the Twitter communication network (the top panel in Fig. 1). They do not appear to be spread homogeneously across the globe. Instead, very few of them connect south-to-south regions, while Europe and the USA appear to be at least one end of most of these edges. Thus, Europe and the USA centralize the global flow of information, and the historical colonial hierarchy is still present in the online world (see Sec. S1 and Fig. S1 in the supplementary material for more details).

In order to quantify the structure of information flows, we measured the betweenness centrality of the Twitter communication network (the bottom panel in Fig. 1). Betweenness centrality measures the number of shortest paths that go through every node and consequently highlights the locations that serve as connectors between many other places. The distribution of betweenness centrality spans multiple orders of magnitude, indicating that while most nodes in this network are peripheral, a few of them are central. We found that most central nodes in the network are cities and that seven out of the top ten are located in Europe and the USA (see the bottom panel in Fig. 1).

Unsurprisingly, cities are hubs in the structure of the Twitter communication network, concentrating a large density of users and connections. Most of the interactions on Twitter are local and embedded within sub-urban areas. Figure 2 shows the distribution of distances between every pair of interacting users. The distribution shows several peaks. The most prominent peak occurs at around 5 km, indicating that most interactions take place within the same city. A second one appears at around 300–500 km, and a third peak is located at around 10 000 km. This manifests the existence of social interactions at intercity, international, and intercontinental scales. In a previous work, we showed that although cities are highly connected at the large scale, their internal structure has an additional layer of complexity with fragmentation due to the socio-economic segregation of homes, with direct implications for the emergent social behaviors.[77]

We analyzed the community structure of the Twitter communication network by applying modularity optimization.[68] We present the results by identifying each community with a distinct color in the top panel of Fig. 3 and showing sub-communities corresponding to each of these communities independent from the other ones in the bottom panel. These communities are detected using the conventional modularity optimization method ($\gamma = 1$ in the generalized method). Each community and their sub-communities are shown separately in the panels of Figs. 4 and 5. In total, we found 16 significant communities. These geographic social modules have global spread, including an area of dominance along with widespread individuals in other regions—perhaps due to travelers or individuals with specific social or cultural ties or interests. The dominant geographic area of a community may contain a single country or several, which are grouped by either cultural or linguistic similarities. The boundaries imply specific cross-cultural, cross-national, and cross-linguistic associations rather than being dominated by any one of these alone. In the Americas, three large areas are manifest. One area corresponds to countries that speak English in the North [Fig. 4(a)], another area includes the Latin American countries in Central and South America [Fig. 4(b)], and a third one includes Brazil [Fig. 4(c)]. In Europe, we find multiple communities. One includes the UK and Ireland and extends all the way to Australia and New Zealand [Fig. 4(d)]. One includes France and extends to the Maghreb (in North-West Africa) and some Sub-Saharan countries [Fig. 4(e)]. Another one includes Portugal and

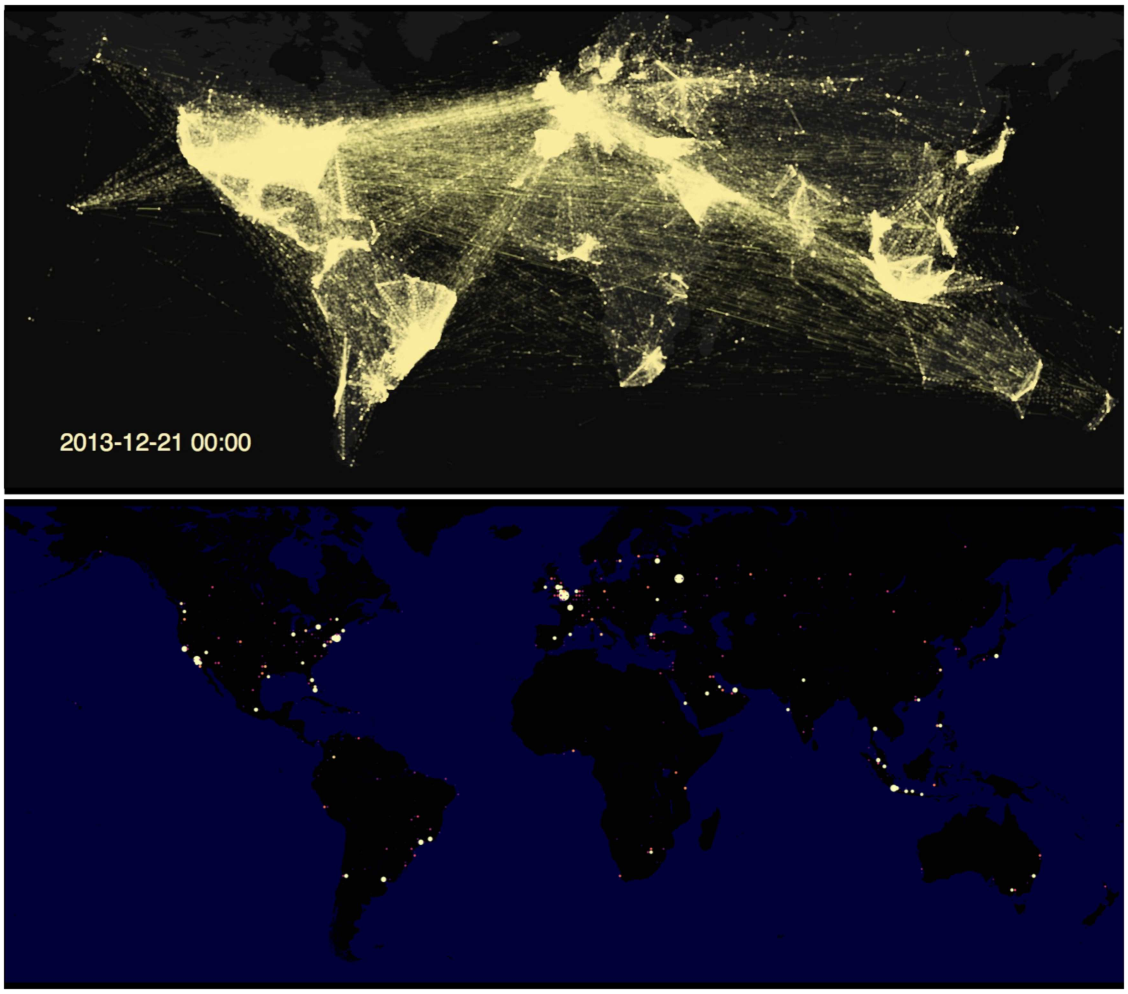

FIG. 1. Top panel: Visualization of the network associated with Twitter global communication. Edges (yellow) connect the location of the tweet emitter to the location of the tweet receiver according to the mentions mechanism (see methods). A video showing the network dynamics can be found at https://www.youtube.com/watch?v=MUTFGt2tYbE. Bottom panel: Betweenness centrality of the Twitter communication network. Dots represent the most central nodes. Some cities may appear displaced from their actual locations because of the $1^\circ \times 1^\circ$ cells.

Spain [Fig. 4(f)], while Italy [Fig. 4(g)] and the Baltic countries [Fig. 4(h)], respectively, have their own community.

Another area includes Germany, Central Europe, and Turkey [Fig. 5(a)]. Eastern Europe and Russia comprise another zone [Fig. 5(b)]. Sub-Saharan Africa appears as a single large community [Fig. 5(c)]. India and the Middle East region appear together [Fig. 5(d)], and the same occurs with Japan, Korea, China, and parts of South East Asia [Fig. 5(e)]. Finally, Indonesia [Fig. 5(f)], the Philippines [Fig. 5(g)], and Malaysia [Fig. 5(h)] have their own distinct communities.

In order to further characterize these large-scale communities, we applied the same algorithm to each community and found their internal sub-community structure (the bottom panel of Fig. 3). At this finer scale, countries separate in Latin America and Sub-Saharan Africa. Distinct sub-national regions appear in the USA (which is separate from Canada), Brazil, and other European countries such as Spain, the UK, and France. In Spain, the areas of the north, center, and south are distinguished from one another, while Portugal remains as a single cohesive community. India is split from the Middle East but not from Pakistan or Bangladesh. The Middle East shows national borders, and Turkey is internally fragmented, distinguishing Istanbul from Anatolia and its subareas. Japan, Korea, and China are distinguished from each other. Russia shows internal modularity along the Trans-Siberian Railway. Australia and New Zealand continue to be tightly connected.

We tested the significance and stability of these communities by running the modularity optimization algorithm multiple times over the geographic network of Twitter communications and

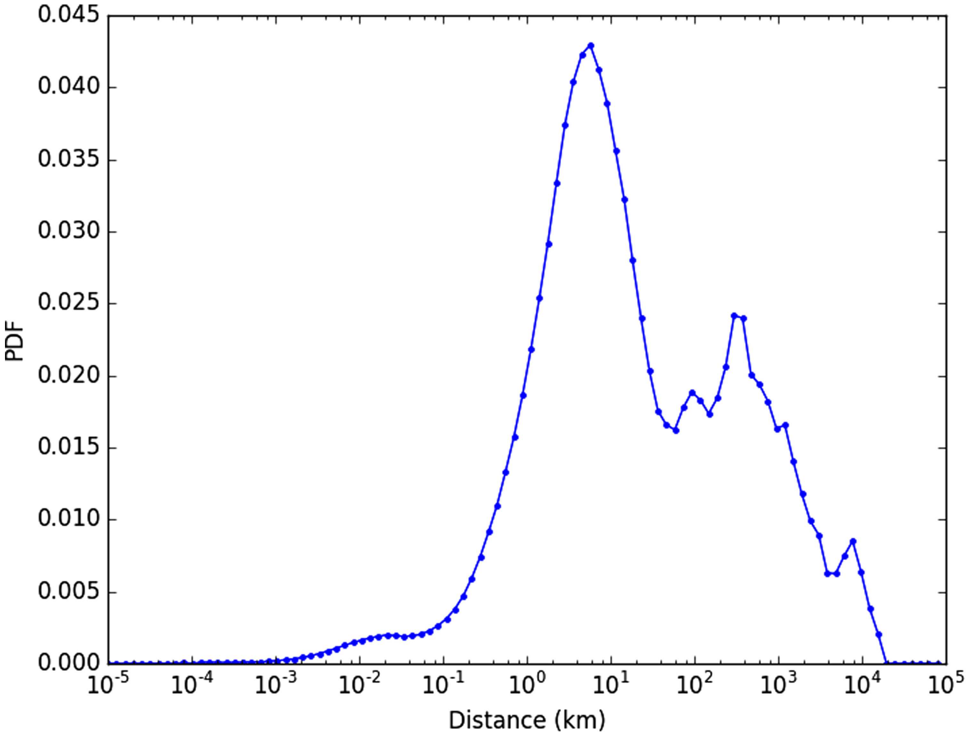

FIG. 2. Probability density function of the distance between users that communicate with each other using mentions on Twitter. Note the logarithmic distance scale.

counting the number of communities obtained in each realization. In over 85% of the cases, the number of communities is invariant, and in the remaining cases, the algorithm detects only one additional community (see Sec. S2 in the supplementary material). The geographic structure of the network is consistent across multiple realizations and, therefore, shows that the partition reported is robust. In Fig. 6(a), we present the map of locations colored by their stability across realization, i.e., remaining in the same community or changing communities. Most areas remain invariant across realizations (green). Red locations indicate areas whose communities change. The areas that differ across realizations are not randomly distributed across the globe but rather are localized around Eastern Canada and Central Europe. A few coupled central European countries, including Germany as well as part of Switzerland, Austria, Hungary, Romania, Croatia, Serbia, and Bulgaria, are alternatively linked to France or the UK. These results show that the structure of community overlaps may be informative about the current state of international relationships.

We further validated the results of the community detection algorithm by filtering edges by their weights. Edges are filtered by discarding those whose weights are below a given threshold. We analyze the community structure after removing edges below 2, 10, and 20 interactions, including the top 60%, 20%, and 10% of edges, respectively. The communities that emerge from these backbone networks are very similar to the ones obtained from the original network in most cases. In Figs. 6(b)–6(d), we compare the community structure of the filtered networks with the original one by constructing a matrix that counts the number of overlapping nodes of communities arising from the networks. Rows have been normalized by the size of each community in the original network. Most communities from the original network are almost identical to the filtered ones and, therefore, show a high overlap (the red diagonal in the matrices). Others are similar but not identical. A few communities diverge and are merged into other communities from the original network (green and light blue). The modular structure is remarkably consistent. Differences become substantial only after removing 60% of the links. Maps with the modular structure of the filtered networks are shown in Fig. S4 of the supplementary material.

In order to analyze the effects of the grid size, we compared the original network with a new one with a higher resolution. In the higher resolution network, nodes represent locations from a lattice of 0.1° latitude $\times 0.1^\circ$ longitude cells, compared to the original 1° resolution. The community structure of this network can be found in Fig. S5 of the supplementary material. The higher resolution network has a similar structure to the original one. Communities are well defined geographically, and national borders matter. In most cases, the communities remain invariant (as in Africa, Brazil, and most of Europe). In other cases, the communities are split into smaller ones (as in North America and Latin America). In Fig. 6(e), we compare the community structure of the higher resolution network with respect to the original one by creating an analogous matrix to the ones shown in Figs. 6(b)–6(d). Red elements show that most communities remain invariant despite the much larger number of edges and a finer scale lattice resolution. A few communities split, and there are instances where small fragments merge into other communities from the original network (green and light blue).

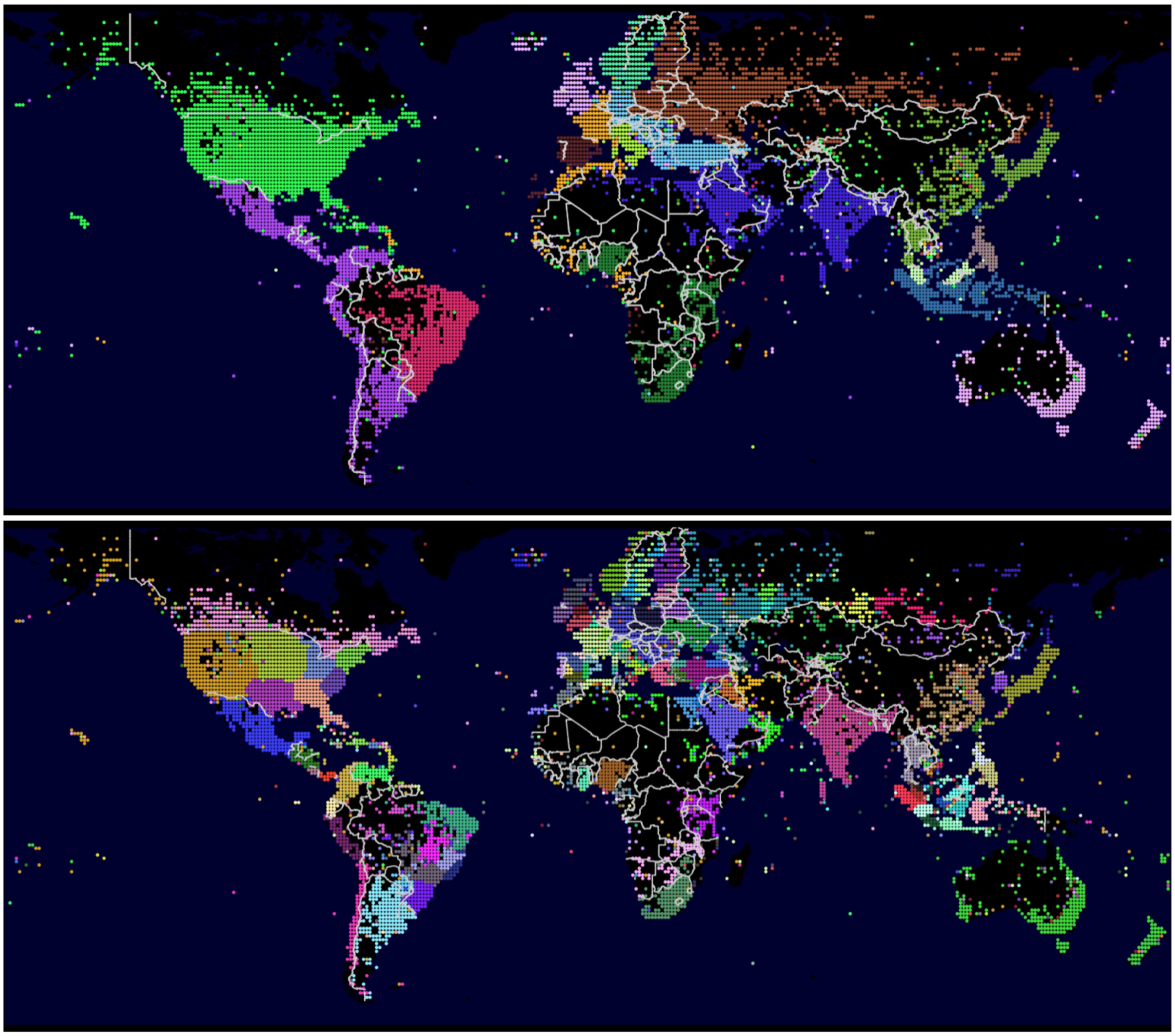

FIG. 3. Fragmentation patterns of the network associated with Twitter global communication. The top panel shows the first level of modularity detection. The bottom panel shows the internal modular structure of the first level communities. Colors indicate distinct communities.

The reason behind the hierarchical structure of communities and sub-communities is given in the multiscale nature of the Twitter communication network. In Fig. 7, we show a multiscale analysis of the community structure by applying a generalized version of the Louvain method that controls for the coarseness of communities with a resolution parameter γ (see Sec. II). In this case, the coarsest community includes the whole globe, which is increasingly decomposed into smaller patches as we increase the resolution parameter, showing national and sub-national borders. Brazil and South East Asia are the first two areas to differentiate from the rest ($\gamma = 0.1$), followed by Russia, Eastern Europe, Middle East, China, and Japan ($\gamma = 0.2$) and later on by the Americas ($\gamma = 0.4$) and Europe ($\gamma = 0.8$). Higher values of $\gamma > 1$ show national borders in Latin America and internal decomposition of Africa and sub-national fragmentation of the USA. Further details about the values of modularity at the different values of the resolution parameter are shown in Sec. S3 of the supplementary material.

The hierarchical stochastic block model is another community detection algorithm that can show the partitions at multiple scales.[78] It identifies blocks based on interior connections among nodes and joins blocks together at the larger scales by considering the edge-degree between the blocks. Specifically, the model builds

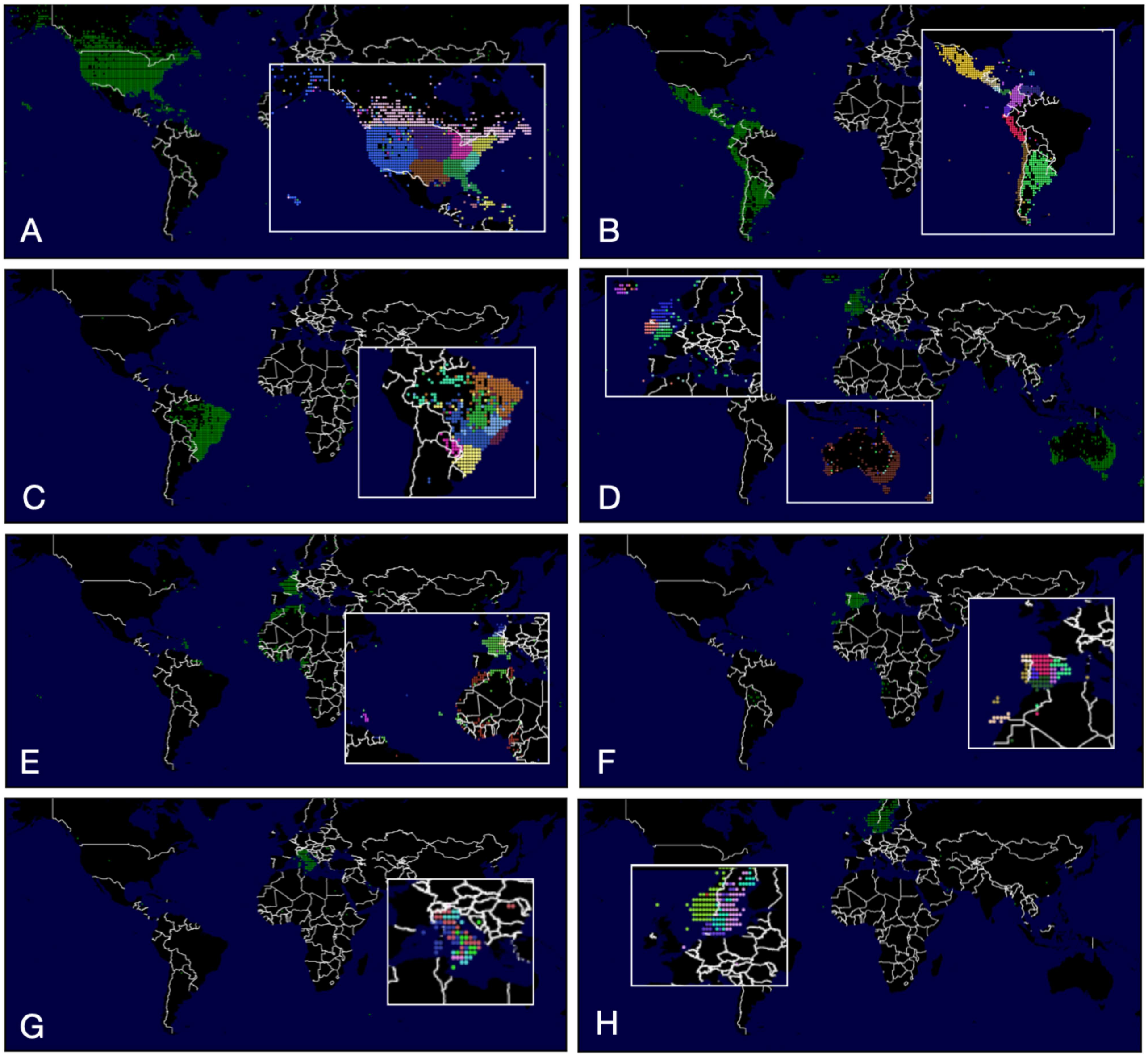

FIG. 4. First set of communities and their sub-communities of the network associated with Twitter global communication shown in separate panels. Panel (a) shows the communities dominated by the USA. Panel (b) shows the communities dominated by Latin American countries. Panel (c) shows the communities related to Brazil. Panel (d) shows the communities dominated by the UK. Panel (e) shows the communities that include France and the Maghreb. Panel (f) shows the communities of the Iberian peninsula. Panel (g) shows the Italian communities. Panel (h) shows the Baltic countries' communities.

the blocks of nodes $\{b_i\}$ in network G, where $b_i \in [1, B]$ is the block membership of node i, by maximizing posterior likelihood $p(G|\{b_i\})$ and minimizing the ensemble entropy. The maximum number of blocks that can be detected according to this model is limited to $B_{\max} = O(N/log(N))$, where N is the number of nodes in the network. Therefore, this model cannot be used to find relatively small partitions. An advantage of this method is that despite the generalized modularity optimization method, it does not need a resolution parameter to detect the communities. Figure 8 shows the hierarchical partitions at four levels, from smaller blocks in level 0 to detecting the whole network as a single community in level 3. Level 0 has 132 partitions, level 1 has 29, and level 2 has 10. At level 0, while some blocks refer to single countries, such as Australia, Canada, Peru, and Russia, some countries including the USA are divided into

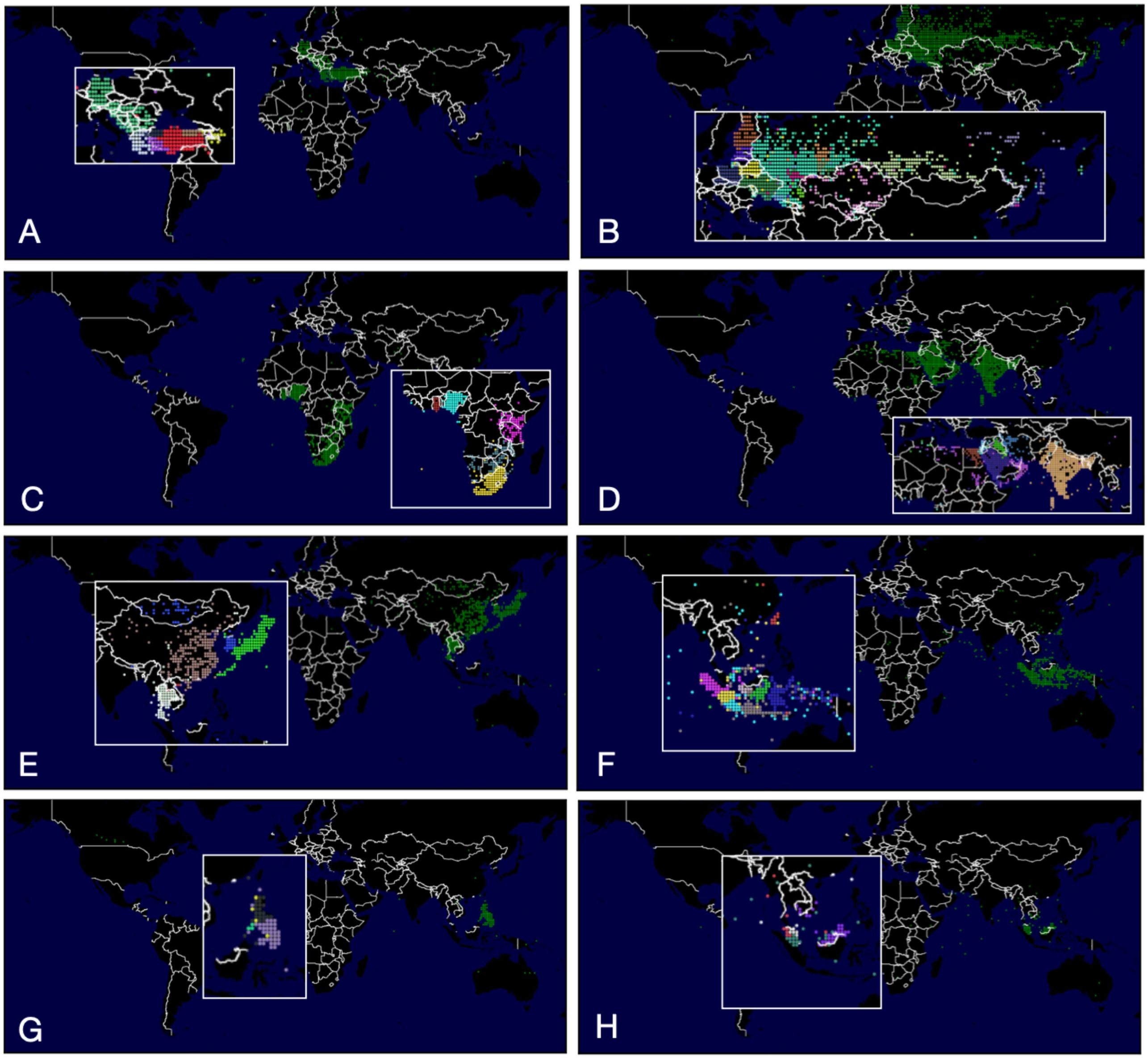

FIG. 5. Second set of communities and their sub-communities of the network associated with Twitter global communication shown in separate panels. Panel (a) shows the communities dominated by Germany, Central Europe, and Turkey. Panel (b) shows the communities dominated by Russia and Eastern Europe. Panel (c) shows the communities related to Africa. Panel (d) shows the communities dominated by the Middle East, North Africa, and India. Panel (e) shows the communities that include China, Japan, and part of South East Asia. Panel (f) shows the communities dominated by Indonesia. Panel (g) shows the communities dominated by Philippines. Panel (h) shows the communities dominated by Malaysia.

several separate blocks, and others combine neighboring countries, for example, countries in the south of Africa. At level 1, many of the blocks in level 0 join together. In this level, north of Africa and west of the Middle East appear as a single community, Russia along with Eastern European countries appear as a single community. While west of the USA is still separated from the east of the USA, it is joined to Canada. At level 2, the whole of Africa and southern Asia except Iran are connected. However, Europe is still separated into several blocks. Eastern Asia and Australia appear as a single community.

In order to better understand the patterns we observe, we created a new network where nodes are countries and edges represent the aggregation of edges between each pair of them. We analyze the weights of these edges according to external information regarding trade, cultural, and geographic proximity. We first create a trade

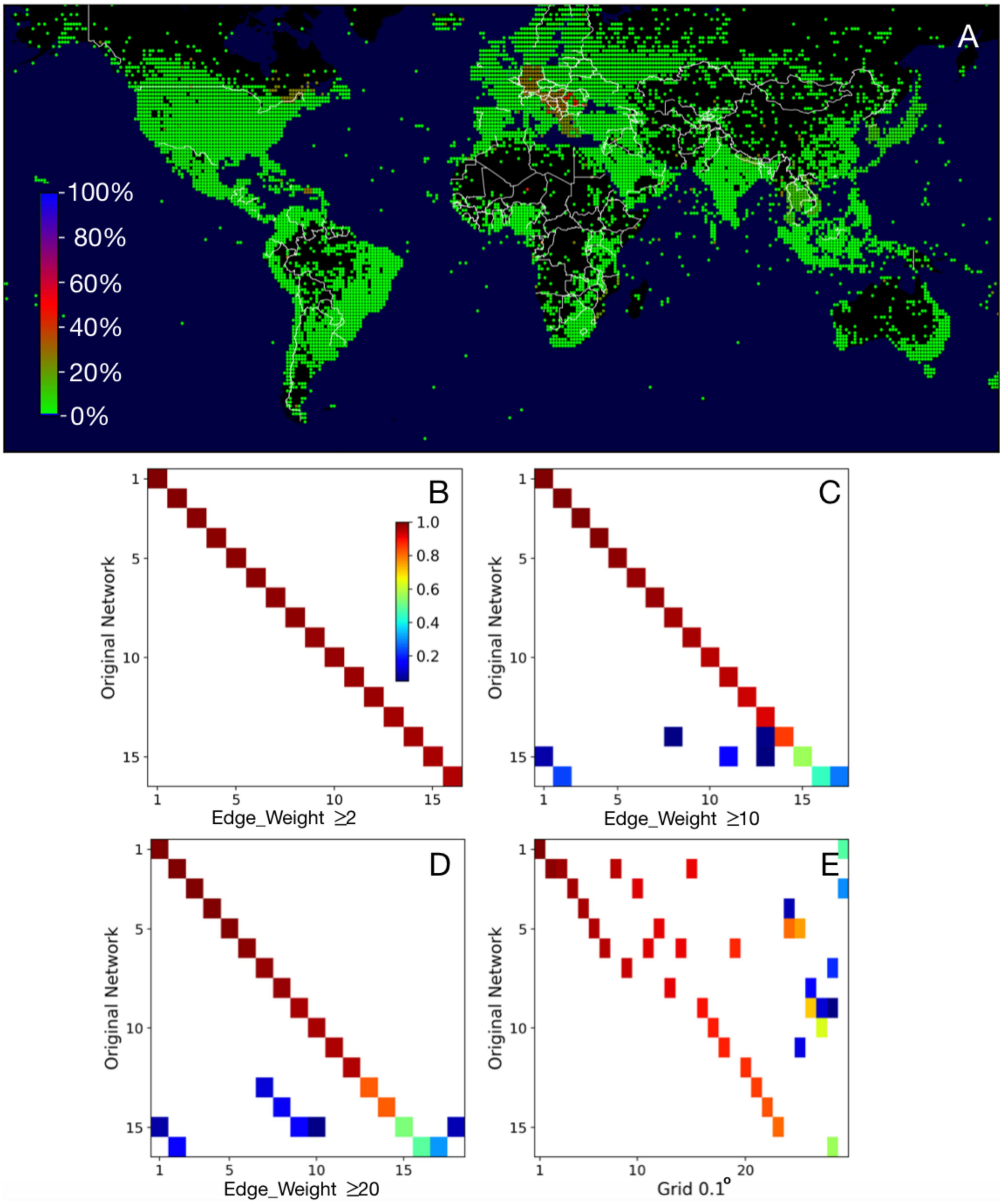

FIG. 6. Validation of the robustness of the community structure of the network associated with Twitter global communication. Panel (a) shows the areas where communities do not overlap (color bar) among multiple algorithmic realizations. Panels (b)–(d) show the similarity of communities in the network after edge removal for multiple edge weight thresholds. Panel (e) shows the similarity of communities at two different grid resolutions. In each case, the vertical axis labels communities of the original network, and the horizontal axis labels the communities after modification (see x-axes labels). In panels (b)–(e), 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 in panel (b)], with no overlap indicated in white. Communities are ordered by decreasing overlap between the networks compared in each matrix.

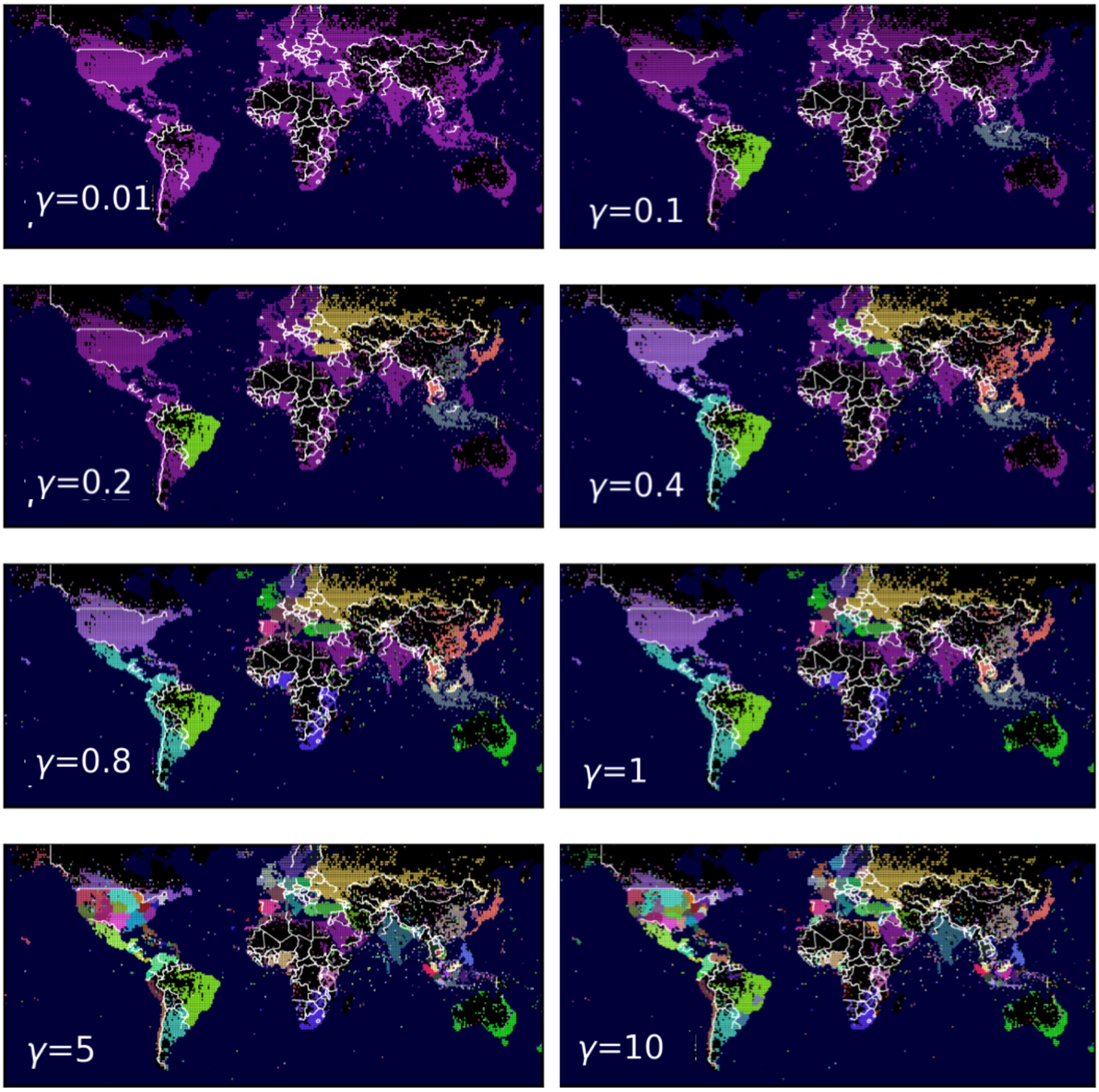

FIG. 7. Multi-scale community detection of the Twitter global communication network. Colors indicate geographical patches detected for values of the resolution parameter γ from 0.01 to 20. When a community is divided into multiple sub-communities at higher resolutions, the sub-community that is most connected to the original community in the former resolution retains the color of the original community, and the other sub-communities are colored differently. The modularity for all of the panels is above 0.3 (see Sec. II C for more details).

network using the data available from the Observatory of Economic Complexity[11] where nodes are countries and edges represent the total amount of traded products in USD during 2013 (the same year as our Twitter dataset). We compared the edge weights between the Twitter communication and trade networks after normalizing them by the degree of the source node such that we compare interaction probabilities conditional on the origin. A significant correlation of $r = 0.43$ ($p < 0.001$) has been found between the edge

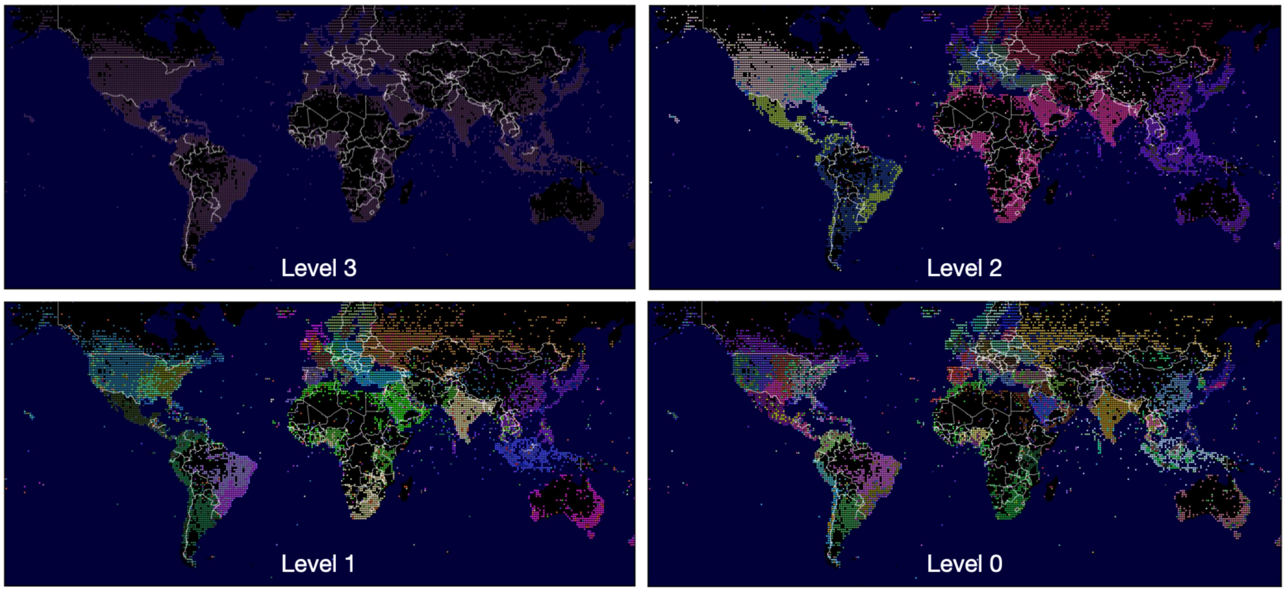

FIG. 8. Hierarchical partitions of the Twitter global communication network obtained from the nested stochastic block model[78] at four levels of resolution.

weights of both networks. Different normalizations yield similar correlations.

Moreover, we analyzed the effects of language, common borders, and colonial history between each pair of countries on the Twitter communication network using the GeoDist dataset from the French institute for research into international economics, CEPII.[79] The dataset contains a set of binary variables indicating if pairs of countries have common languages, share borders, were formerly colonized by each other, or share a common colonizer. We split country pairs in two sets for each variable according to whether they meet such criteria or not. Prior to splitting country pairs, we label each pair by the decile they belong to according to the edge weight that connects them in the Twitter communication network. If the CEPII variables have no association to the distribution of edge weights, the decile labels from the corresponding set of country pairs should follow a uniform distribution. However, if the CEPII variables do have an association, the distribution of decile labels should be different from uniform and reveals the extent of such associations. We found statistically significantly higher deciles for countries who have common languages, contiguous borders, and who were colony and colonizer, implying a strong association of these variables to interactions on the online platform ($p < 0.001$). We also found, perhaps more surprisingly, that countries who had a common colonizer have a *d*ecreased preference of interactions, suggesting that the hierarchical interactions with the colonizer inhibit lateral interactions. The cumulative distribution of deciles per variable (color) and set (solid for true and dotted for false) is shown in Fig. 9. For all variables, the sets that are false are close to the black diagonal and hence follow the uniform distribution. We apply the t-test to show that the averages differ from both distributions for each variable (see Table I). In turn, the sets that are true significantly differ from the uniform distribution ($p < 0.001$).

Finally, it is possible to extend the analysis and focus on specific regions, countries, or urban areas. In Fig. 10, we present a few examples: The Island of Java in Indonesia [Figs. 10(a) and 10(b)], the Iberian Peninsula [Fig. 10(c)], and the city of Istanbul in Turkey [Fig. 10(d)]. We use lattice locations of 1 km wide in the first two cases and 500 m wide in the third case. As in the global scale, social systems show a clear geographical fragmentation in both regional and urban scales. In the case of Java, the island is split into multiple areas, containing large cities and the roads that interconnect them. Jakarta, the main city, appears to be split into East, West, and South regions. A multiscale decomposition of Java communities is shown in Fig. S7 of the supplementary material, analogously to Fig. 7. Similarly in Spain, Madrid and its surroundings represent one community (center of the map), while Catalonia (upper right) and Galicia (upper left) have separate communities. By contrast, Portugal appears as a single cohesive community. Perhaps more surprisingly, the structure of individual cities, as in Istanbul, is also fragmented. Natural boundaries, such as the Bosporus, as well as

TABLE I. Statistical tests (t-test and p-value) of geometrical distribution correlations according to variables from the CEPII database.

	Statistical tests	
Parameters	t-test	p-value
Common spoken language	22.40	3.88×10^{-111}
Common official language	12.25	1.56×10^{-34}
Contiguous	15.80	3.04×10^{-56}
Colony	33.46	1.47×10^{-245}
Common colonizer	50.37	0.00

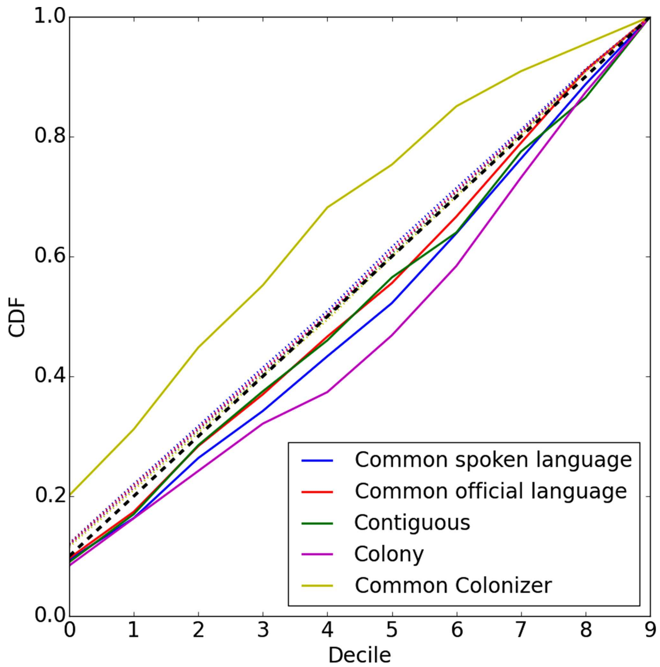

FIG. 9. Statistical tests of edge weight distributions according to variables from the CEPII database. The figure compares the expected cumulative distribution of edges according to their decile from the original distribution. If edges are independently sampled, the cumulative curve would resemble the uniform distribution (black dashed diagonal). Variables represented in the legend are binary and distinct edges. Samples from pairs of countries that meet the variables in the legend are represented in solid lines. Samples that do not meet the variables in the legend are represented in dotted lines.

socioeconomic distinctions separate the network into neighborhood level communities.

IV. CONCLUSION

In nature, organisms differentiate their parts not just through behavioral differences, but also by the structure of information, matter, and energy flows among them. These structural distinctions enable the emergence of functionalities at progressively larger scales. We have shown that the global society may present a similar structure, where parts of the population are clustered and information is selectively shared.

We found that the global network of Twitter communications shows that administrative borders determine the structure of strong and weak social ties, which are reflected in the emergence of clearly defined geographic fragmentation at multiple scales. Such fragmentation implies a highly heterogeneous flow of information across the globe, with direct implications for how people learn from one another and consequently behave. Patches represent areas with higher inner connections than connections to the rest of the areas. While strong connections keep clusters cohesive, weak ones integrate clusters at the larger scales and are responsible for the spread of information across the clusters. At large scales, areas such as Europe, the USA, Brazil, and South East Asia appear as separate patches, which split into distinct sub-national regions at smaller scales. According to the Twitter mentions network, Europe and the USA centralize the global flow of information. Our results show that the historical colonial hierarchy is still present in the online world, and there is a significant correlation between country relationships in Twitter communication and trade networks. Language and common borders are the other two important parameters in the formation of fragmentation patterns.

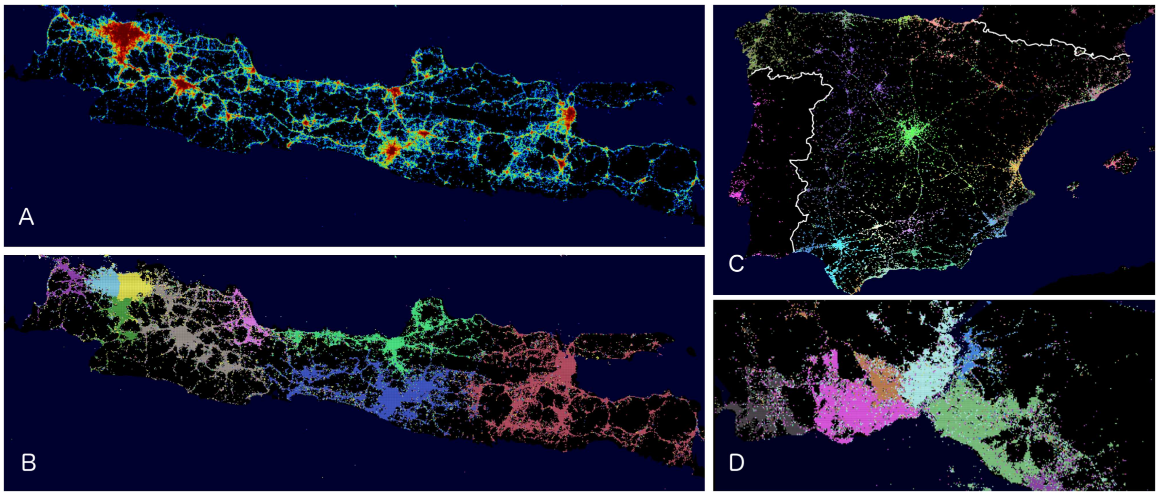

FIG. 10. Twitter communication network structure at higher resolutions in specific regions. Panel (a) shows the degree distribution of the mentions network for the Island of Java in Indonesia. Panel (b) shows the corresponding community structure. Panel (c) shows the community structure in the Iberian Peninsula. Panel (d) shows the community structure in Istanbul, Turkey. The resolution of the lattice in panels (a)–(c) is 1 km wide. The resolution of the lattice in panel (d) is 500 m wide.

The distribution of distances shows that most of the interactions take place within the same city. Statistical tests represent over 85% stability in the detected fragmentation patterns, and even filtering weak ties does not change the communities' borders in most cases. While reducing the lattice cells increases the network resolution, it does not change the community structure.

SUPPLEMENTARY MATERIAL

See the supplementary material for the stability and significance of communities, the grid size, communities at multiple scales, the CEPII database, Refs. 1–5, and Figs. S1–S8.

AUTHORS' CONTRIBUTIONS

All authors contributed equally to the data collection, conceptualization and development of experiments, and in the manuscript write-up.

ACKNOWLEDGMENTS

We thank Matthew Hardcastle for proofreading the manuscript. The authors declare that they have no competing interests.

DATA AVAILABILITY

The data that support the findings of this study are openly available in gml format at https://necsi.edu/world-mentions-network-data, Ref. 80.

REFERENCES

[1] D. Arlt, A. Rauchfleisch, and M. S. Schäfer, "Between fragmentation and dialogue. Twitter communities and political debate about the Swiss "nuclear withdrawal initiative"," Environ. Commun. **13**(4), 440–456 (2019).
[2] A. J. Morales, J. C. Losada, and R. M. Benito, "Users structure and behavior on an online social network during a political protest," Physica A **391**(21), 5244–5253 (2012).
[3] D. Xue, S. Hirche, and M. Cao, "Opinion behavior analysis in social networks under the influence of coopetitive media," IEEE Trans. Netw. Sci. Eng. 1 (2019).
[4] A. Morales, J. Borondo, J. C. Losada, and R. M. Benito, "Measuring political polarization: Twitter shows the two sides of Venezuela," Chaos **25**, 033114 (2015).
[5] M. Nelimarkka, S.-M. Laaksonen, and B. Semaan, "Social media is polarized, social media is polarized: Towards a new design agenda for mitigating polarization," in *Proceedings of the 2018 Designing Interactive Systems Conference* (ACM, 2018), pp. 957–970.
[6] T. Bolsen and M. A. Shapiro, "The US news media, polarization on climate change, and pathways to effective communication," Environ. Commun. **12**(2), 149–163 (2018).
[7] L. Hedayatifar, R. A. Rigg, Y. Bar-Yam, and A. J. Morales, "US social fragmentation at multiple scales," J. R. Soc. Interface **16**(159), 20190509 (2019).
[8] J. Swart, C. Peters, and M. Broersma, "Shedding light on the dark social: The connective role of news and journalism in social media communities," New Media Soc. **20**(11), 4329–4345 (2018).
[9] M. Bailey, R. Cao, T. Kuchler, J. Stroebel, and A. Wong, "Social connectedness: Measurement, determinants, and effects," J. Econ. Perspect. **32**(3), 259–280 (2018).
[10] M. Lim, R. Metzler, and Y. Bar-Yam, "Global pattern formation and ethnic/cultural violence," Science **317**(5844), 1540–1544 (2007).
[11] R. Hausmann and C. Hidalgo, *The Atlas of Economic Complexity: Mapping Paths to Prosperity*, Vol. 1 (The MIT Press, 2014).
[12] M. E. J. Newman, "Communities, modules and large-scale structure in networks," Nat. Phys. **8**(1), 25 (2012).
[13] M. S. Granovetter, "The strength of weak ties," Am. J. Sociol. **78**(6), 1360–1380 (1973).
[14] C. Thiemann, F. Theis, D. Grady, R. Brune, and D. Brockmann, "The structure of borders in a small world," PLoS One **5**, e15422 (2010).
[15] F. Calabrese, D. Dahlem, A. Gerber, D. Paul, X. Chen, J. Rowland, C. Rath, and C. Ratti, "The connected States of America: Quantifying social radii of influence," in *Proceedings of IEEE International Conference on Social Computing (SocialCom)* (IEEE, 2011).
[16] T. Menezes and C. Roth, "Natural scales in geographical patterns," Sci. Rep. **7**, 45823 (2017).
[17] 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), 6 (2014).
[18] F. Xiang, L. Tu, B. Huang, and X. Yin, "Region partition using user mobility patterns based on topic model," in *16th International Conference on Computational Science and Engineering* (IEEE, 2014).
[19] G. D. Nelson and A. Rae, "An economic geography of the United States: From commutes to megaregions," PLoS One **11**(11), e0166083 (2016).
[20] A. J. Morales, W. Creixell, J. Borondo, J. C. Losada, and R. Benito, "Characterizing ethnic interactions from human communication patterns in Ivory Coast," Netw. Heterog. Media **10**(1), 87 (2015).
[21] R. Lambiotte, V. D. Blondel, C. De Kerchove, E. Huens, C. Prieur, Z. Smoreda, and P. Van Dooren, "Geographical dispersal of mobile communication networks," Physica A **387**, 5317–5325 (2008).
[22] 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).
[23] S. Sobolevsky, M. Szell, R. Campari, T. Couronné, Z. Smoreda, and C. Ratti, "Delineating geographical regions with networks of human interactions in an extensive set of countries," PLoS One **8**, e81707 (2013).
[24] V. Blondel, G. Krings, and I. Thomas, "Regions and borders of mobile telephony in Belgium and in the Brussels metropolitan zone," Brussels Stud. **42** (2010).
[25] 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).
[26] Z. 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).
[27] M. Saeedian, T. Jamali, S. V. Farahani, G. R. Jafari, and M. Ausloos, "How visas shape and make visible the geopolitical architecture of the planet," arXiv:1601.06314 (2016).
[28] D. Lazer, A. Pentland, L. Adamic, S. Aral, A.-L. Barabasi, 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 (2009).
[29] Y.-A. de Montjoye, C. A. Hidalgo, M. Verleysen, and V. D. Blondel, "Unique in the crowd: The privacy bounds of human mobility," Sci. Rep. **3**, 1376 (2013).
[30] Y. Bar-Yam, *Making Things Work: Solving Complex Problems in a Complex World* (Knowledge Industry, 2004).
[31] Y. Bar-Yam, "From big data to important information," Complexity **21**(S2), 73–98 (2016).
[32] A. Adam, J.-C. Delvenne, and I. Thomas, "Detecting communities with the multi-scale Louvain method: Robustness test on the metropolitan area of Brussels," J. Geogr. Syst. **20**(4), 363–386 (2018).
[33] Y. Bar-Yam, "Multiscale variety in complex systems," Complexity **9**(4), 37–45 (2004).
[34] M. Ostovari and D. Yu, "Impact of care provider network characteristics on patient outcomes: Usage of social network analysis and a multi-scale community detection," PLoS One **14**(9), e0222016 (2019).
[35] Y.-A. de Montjoye, A. Stopczynski, E. Shmueli, A. Pentland, and S. Lehmann, "The strength of the strongest ties in collaborative problem solving," Sci. Rep. **4**, 5277 (2014).
[36] A. Morales, J. Borondo, J. C. Losada, and R. M. Benito, "Efficiency of human activity on information spreading on Twitter," Soc. Netw. **39**, 1–11 (2014).
[37] E. Bakshy, S. Messing, and L. A. Adamic, "Exposure to ideologically diverse news and opinion on Facebook," Science **348**(6239), 1130–1132 (2015).

[38]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), 20 (2017).
[39]T. Schelling, "Dynamic models of segregation," J. Math. Sociol. **1**, 143–186 (1971).
[40]A. Varshney, *Ethnic Conflict and Civic Life: Hindus and Muslims in India* (Yale University Press, 2008).
[41]D. L. Horowitz, *Ethnic Groups in Conflict* (University of California Press, Berkeley, 2011).
[42]M. J. Esman, *Ethnic Conflict in the Western World* (Cornell University Press, London, 1979).
[43]R. Axelrod, "The dissemination of culture: A model with local convergence and global polarization," J. Confl. Resolut. **41**(2), 203–226 (1997).
[44]C. R. Laub, "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).
[45]B. R. da Cunha, J. C. González-Avella, and S. Gonçalves, "Fast fragmentation of networks using module-based attacks," PLoS One **10**, e0142824 (2015).
[46]S. Wandelt, X. Sun, D. Feng, M. Zanin, and S. Havlin, "A comparative analysis of approaches to network-dismantling," Sci. Rep. **8**, 13513 (2018).
[47]L. M. Shekhtman, M. M. Danziger, and S. Havlin, "Recent advances on failure and recovery in networks of networks," Chaos Solitons Fractals **90**, 28–36 (2016).
[48]J. M. Kleinberg, "Navigation in a small world," Nature **406**, 845 (2000).
[49]N. Friedkin, "A test of the structural features of Granovetter's 'strength of weak ties' theory," Soc. Netw. **2**, 411–422 (1980).
[50]R. L. Breiger and P. E. Pattison, "The joint role structure of two communities' elites," Sociol. Methods Res. **7**(2), 213–226 (1978).
[51]B. R. da Cunha and S. Gonçalves, "Topology, robustness, and structural controllability of the Brazilian Federal Police criminal intelligence network," Appl. Netw. Sci. **3**, 36 (2018).
[52]F. Morstatter, J. Pfeffer, H. Liu, and K. M. Carley, "Is the sample good enough? Comparing data from Twitter's streaming API with Twitter's firehose," in *Proceedings of the 7th International AAAI Conference on Weblogs and Social Media* (The AAAI Press, Palo Alto, CA, 2013).
[53]K. Leetaru, S. Wang, G. Cao, A. Padmanabhan, and E. Shook, "Mapping the global Twitter heartbeat: The geography of Twitter," First Monday **18**(5) (2013).
[54]M. Lenormand, B. Gonçalves, A. Tugores, and J. J. Ramasco, "Human diffusion and city influence," J. R. Soc. Interface **12**(109), 20150473 (2015).
[55]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).
[56]D. Mocanu, A. Baronchelli, N. Perra, B. Gonçalves, Q. Zhang, and A. Vespignani, "The Twitter of Babel: Mapping world languages through microblogging platforms," PLoS One **8**(4), e61981 (2013).
[57]J. Atienza-Barthelemy, S. Martin-Gutierrez, J. C. Losada, and R. M. Benito, "Relationship between ideology and language in the Catalan independence context," Sci. Rep. **9**(1), 17148 (2019).
[58]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," Int. J. Geogr. Inf. Sci. **29**(2), 258–279 (2015).
[59]C. A. Bail, L. P. Argyle, T. W. Brown, J. P. Bumpus, H. Chen, M. B. F. Hunzaker, J. Lee, M. Mann, F. Merhout, and A. Volfovsky, "Exposure to opposing views on social media can increase political polarization," Proc. Natl. Acad. Sci. U.S.A. **115**(37), 9216–9221 (2018).
[60]A. Herdagdelen, 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).
[61]M. R. Frank, L. Mitchell, P. S. Dodds, and C. M. Danforth, "Happiness and the patterns of life: A study of geolocated tweets," Sci. Rep. **3**(1), 2625 (2013).
[62]J. Ugander, B. Karrer, L. Backstrom, and C. Marlow, "The anatomy of the Facebook social graph," arXiv:1111.4503 (2011).
[63]Pew Research Center, "Social media use in 2018," Pew Research Center, March 2018.
[64]M. Duggan and D. Page, "Mobile messaging and social media 2015," Internet Science and Technology, Pew Research Center, 2015.
[65]F. Lamanna, M. Lenormand, M. H. Salas-Olmedo, G. Romanillos, B. Gonçalves, and J. J. Ramasco, "Immigrant community integration in world cities," PLoS One **13**(3), e0191612 (2018).
[66]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," Int. J. Geogr. Inf. Sci. **31**(7), 1293–1313 (2017).
[67]B. Hawelka, I. Sitko, E. Beinat, S. Sobolevsky, P. Kazakopoulos, and C. Ratti, "Geo-located Twitter as proxy for global mobility patterns," Cartogr. Geogr. Inf. Sci. **41**(3), 260–271 (2014).
[68]V. D. Blondel, J.-L. Guillaume, R. Lambiotte, and E. Lefebvre, "Fast unfolding of communities in large networks," J. Stat. Mech.: Theory Exp. **2008**(10), P10008.
[69]M. E. J. Newman, "Modularity and community structure in networks," Proc. Natl. Acad. Sci. U.S.A. **103**(23), 8577–8582 (2006).
[70]A.-L. Barabási *et al.*, *Network Science* (Cambridge University Press, 2016).
[71]A. Clauset, M. E. J. Newman, and C. Moore, "Finding community structure in very large networks," Phys. Rev. E **70**(6), 066111 (2004).
[72]J. Reichardt and S. Bornholdt, "Statistical mechanics of community detection," Phys. Rev. E **74**, 016110 (2006).
[73]A. Braunstein, L. Dall'Asta, G. Semerjian, and L. Zdeborová, "Network dismantling," Proc. Natl. Acad. Sci. U.S.A. **113**(44), 12368–12373 (2016).
[74]X.-L. Ren, N. Gleinig, D. Helbing, and N. Antulov-Fantulin, "Generalized network dismantling," Proc. Natl. Acad. Sci. U.S.A. **116**, 6554–6559 (2018).
[75]M. Girvan and M. E. J. Newman, "Community structure in social and biological networks," Proc. Natl. Acad. Sci. U.S.A. **99**(12), 7821–7826 (2002).
[76]B. Huberman, D. Romero, and F. Wu, "Social networks that matter: Twitter under the microscope," First Monday **14**(1), 9 (2008), arXiv:0812.1045.
[77]A. Morales, X. Dong, Y. Bar-Yam, and A. S. Pentland, "Segregation and polarization in urban areas," R. Soc. Open Sci. **6**(10), 190573 (2019).
[78]T. P. Peixoto, "Hierarchical block structures and high-resolution model selection in large networks," Phys. Rev. X **4**(1), 011047 (2014).
[79]T. Mayer and S. Zignago, "Notes on CEPII distances measures: The GeoDist database," CEPII Working Paper, 2011.
[80]World Mentions Network Data, see https://necsi.edu/world-mentions-network-data.