Uncovering space-independent communities in spatial networks

Paul Expert[a,b], Tim S. Evans[b], Vincent D. Blondel[c,d], and Renaud Lambiotte[a,e,1]

[a]Complexity and Networks Group, Imperial College London, London SW7 2AZ, United Kingdom; [b]Blackett Laboratory, Prince Consort Road, Imperial College London, London SW7 2AZ, United Kingdom; [c]Massachusetts Institute of Technology, Laboratory for Information and Decision Systems, 77 Massachusetts Avenue, Cambridge, MA 02139; [d]Institute of Information and Communication Technologies, Electronics and Applied Mathematics, Université catholique de Louvain, Avenue Georges Lemaitre 4, B-1348 Louvain-la-Neuve, Belgium; and [e]Naxys, Facultés Universitaires Notre-Dame de la Paix, B-5000 Namur, Belgium

Edited by Kenneth Wachter, University of California, Berkeley, CA, and approved March 22, 2011 (received for review December 20, 2010)

Many complex systems are organized in the form of a network embedded in space. Important examples include the physical Internet infrastucture, road networks, flight connections, brain functional networks, and social networks. The effect of space on network topology has recently come under the spotlight because of the emergence of pervasive technologies based on geolocalization, which constantly fill databases with people's movements and thus reveal their trajectories and spatial behavior. Extracting patterns and regularities from the resulting massive amount of human mobility data requires the development of appropriate tools for uncovering information in spatially embedded networks. In contrast with most works that tend to apply standard network metrics to any type of network, we argue in this paper for a careful treatment of the constraints imposed by space on network topology. In particular, we focus on the problem of community detection and propose a modularity function adapted to spatial networks. We show that it is possible to factor out the effect of space in order to reveal more clearly hidden structural similarities between the nodes. Methods are tested on a large mobile phone network and computer-generated benchmarks where the effect of space has been incorporated.

complex networks | social systems

Understanding the principles driving the organization of complex networks is crucial for a broad range of fields including information and social sciences, economics, biology, and neuroscience (1). In networks where nodes occupy positions in an Euclidian space, spatial constraints may have a strong effect on their connectivity patterns (2). Edges may either be spatially embedded, such as in roads or railway lines in transportation networks or cables in a power grid, or abstract entities, such as friendship relations in online and offline social networks or functional connectivity in brain networks. In either case, space plays a crucial role by affecting, directly or indirectly, network connectivity and making its architecture radically different from that of random networks (3). A crucial difference stems from the cost associated to long-distance links (4–12), which restricts the existence of hubs (i.e., high-degree nodes), and thus the observation of fat-tailed degree distributions in spatial networks.

From a modeling viewpoint, gravity models (13–15) have long been used to model flows in spatial networks. These models focus on the intensity of interaction between locations i and j separated by a certain physical distance d_{ij}. It has been shown for systems as diverse as the International Trade Market (16), human migration (17), traffic flows (18), or mobile communication between cities (19, 20) that the volume of interaction between distant locations is successfully modeled by

$$T_{ij} = N_i N_j f(d_{ij}), \qquad \textbf{[1]}$$

where N_i measures the importance of location i, e.g., its population, and the deterrence function f describes the influence of space. Eq. **1** emphasizes that the number of interactions between two locations is proportional to the number of possible contacts $N_i N_j$ and that it varies with geographic distance, because of financial or temporal cost. In many socioeconomic systems, f is well fitted by a power law $\sim d_{ij}^{-\alpha}$ reminiscent of Newton's law of gravity, with population playing the role of a mass.

Whereas a broad range of models have been specifically developed for spatial networks (21–25), dedicated tools for uncovering useful information from their topology are poorly developed. When analyzing spatial networks, authors tend to use network metrics where the spatial arrangement of the nodes is ignored, thus disregarding that useful measures for nonspatial networks might yield irrelevant or trivial results for spatial ones. Important examples are the clustering coefficient, because spatial networks are often spatially clustered by nature, and degree distribution, where high-degree nodes are suppressed by long-distance costs. This observation underlines the need for appropriate metrics for the analysis and modeling of networks where spatial constraints play an important role (26–28).

This need is particularly apparent in the context of community detection. The detection of communities (modules or clusters) is a difficult task that is important to many fields, and it has attracted much attention in the last few years (29–35). In a nutshell, modules are defined as subnetworks that are locally dense even though the network as a whole is sparse. Community detection is a central tool of network theory because revealing intermediate scales of network organization provides the means to draw readable maps of the network and to uncover hidden functional relations between nodes (32). In the case of spatial networks, important practical applications include (*i*) the design of efficient national, economical, or administrative borders based on human mobility or economical interactions, instead of historical or ad hoc reasons (36–39); (*ii*) the modeling of historical or prehistorical interactions based on limited archaeological evidence (40, 41); *iii*) the identification of functionally related brain regions and of principles leading to global integration and functional segregation (42, 43).

In practice, the current state-of-the-art for finding modules in spatial networks (44, 45) is to optimize the standard Newman–Girvan modularity, which, as we argue below, overlooks the spatial nature of the system. In most cases, this scheme produces communities which are strongly determined by geographical factors and provide poor information about the underlying forces shaping the network. For instance, social and transportation networks are typically dominated by low-cost short-ranged interac-

Author contributions: P.E., T.S.E., V.D.B., and R.L. designed research; P.E., T.S.E., V.D.B., and R.L. performed research; P.E., T.S.E., V.D.B., and R.L. analyzed data; and P.E., T.S.E., V.D.B., and R.L. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

[1]To whom correspondence should be addressed. E-mail: r.lambiotte@imperial.ac.uk.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1018962108/-/DCSupplemental.

tions leading to modules which are compact in physical space. As a result, modularity optimization is blind to spatial anomalies and fails to uncover modules determined by factors other than mere physical proximity. This point brings us to the central question of our work: In spatial networks, how can one detect patterns that are not due to space? In other words, are observed patterns only due to the effect of spatial distance, because of gravity-like forces, or do other forces come into play? If that is the case, can one go beyond a standard network methodology in order to uncover significant information from spatial networks?

Social Networks and Space

In order to illustrate these concepts and to clarify the goal of this paper, let us elaborate on social networks, where the dichotomy between network and space has been studied for decades. On the one hand, research has attempted to explain the organization of social networks purely in terms of the structural position of the nodes. Structural mechanisms underpinning the existence of social interactions include triadic closure (46) and link reciprocity (47). On the other hand, research has identified ordering principles that explain edge creation in terms of nonstructural attributes, mainly homophily (48) and focus constraint (49). Homophily states that similarity, e.g., in terms of status or interests, fosters connection (48), because similar people tend to select each other, communicate more frequently, and develop stronger social interactions (50). The second ordering principle is focus constraint (49), which refers to the idea that social relations depend on opportunities for social contact. A dominant factor for focus constraint is geographic proximity, which offers opportunities for face-to-face interaction and encounters between individuals (51). Focus constraint thus depends indirectly on distance through its dependence on transportation networks, which themselves typically exhibit a gravity law.

Although homophily and focus constraint are different mechanisms, they are often interrelated, because frequent contacts drive groups toward uniformity, through social influence, and that alike individuals tend to live in the same neighborhoods (52). Moreover, both aspects can be seen as originating from proximity in a high-dimensional social space, which summarizes people's interests and characteristics—i.e., nodes have a tendency to connect with neighboring nodes in social space (53). When uncovering modules of strongly connected nodes in complex networks, one deals with an extremely intricate situation where structural and nonstructural effects, including homophily and focus constraint, are mingled. Modules uncovered by community detection are thus underpinned by an uncontrolled mixture of possibly antagonistic forces, from which few conclusions can be drawn (54). Our aim is the following: When the spatial positions of the nodes are known, as more and more often is the case, is it possible to take out the effect of space in order to identify more clearly homophilious effects and thus hidden structural or cultural similarities.

Modularity and Space

Let us now introduce the notations and formalize the problem of community detection. In the following, we focus on weighted, undirected networks characterized by their adjacency matrix A. By definition, A is symmetric and A_{ij} is the weight of the link between i and j. The strength of node i is defined as $k_i = \sum_j A_{ij}$; $m = \sum_{i,j} A_{ij}/2$ is the total weight in the network. The distance between nodes i and j is denoted by d_{ij}. From now on, by distance, we mean Euclidian distance between nodes when measured on the embedding space, and not network distance, which is the number of edges traversed along the shortest path from one vertex to another. As discussed above, the nature of space and its associated distance may be abstract (i.e., affinity in a social network) or physical (i.e., geographical distance between cities).

The fundamental idea behind most community detection methods is to partition the nodes of the network into modules. Contrary to standard graph partitioning algorithms, the detection of communities is performed without a priori specifying the number of modules nor their size and aims at uncovering in an automated way the mesoscale organization of the network (31). Behind most community detection methods, there is a mathematical definition measuring the quality of a partition. The widely used modularity (55) of a partition $\mathscr{P}$ measures if links are more abundant within communities than would be expected on the basis of chance, namely,

$$Q = (\text{fraction of links within communities}) - (\text{expected fraction of such links}). \qquad [2]$$

In a mathematical expression, modularity reads

$$Q = \frac{1}{2m} \sum_{C \in \mathscr{P}} \sum_{i,j \in C} [A_{ij} - P_{ij}], \qquad [3]$$

where $i,j \in C$ is a summation over pairs of nodes i and j belonging to the same community C of $\mathscr{P}$ and therefore counts links between nodes within the same community.

What is meant by chance (i.e., the null hypothesis) is an extra ingredient in the definition (56) and is embodied by the matrix P_{ij}. P_{ij} is the expected weight of a link between nodes i and j over an ensemble of random networks with certain constraints. These constraints correspond to known information about the network organization (i.e., its total number of links and nodes), which has to be taken into account when assessing the relevance of an observed topological feature. In general, if A_{ij} is symmetric, P_{ij} is also chosen to be symmetric and one also imposes that the total weight is conserved* (i.e., $\sum_{ij} A_{ij} = \sum_{ij} P_{ij} = 2m$). Beyond these basic considerations, different null models can be constructed depending on the network under consideration (57–59). The most popular choice, proposed by Newman and Girvan (NG) (55) is

$$P_{ij}^{\mathrm{NG}} = k_i k_j / 2m, \quad \text{then } Q = Q_{\mathrm{NG}}, \qquad [4]$$

where randomized networks preserve the strength of each node. Constraining the node strengths goes along the view that the network is well mixed, in the sense that any node can be connected to any node and that only connectivity matters. In that case, node strength is a good proxy for the probability of a link to arrive on a certain node. Different types of heuristics can be developed in order to approximate the optimal value of the corresponding NG modularity (56, 60–62). These methods have been shown to produce useful and relevant partitions in a broad class of systems (31), even if modularity suffers from limitations such as resolution limit (63) and a possible high degeneracy of its landscape (64, 65).

The NG null model only uses the basic structural information encoded in the adjacency matrix. Therefore, it is appropriate when no additional information on the nodes is available but not when additional constraints are known. In networks where distance strongly affects the probability for two nodes to be connected, a natural choice for the null model is inspired by the aforementioned gravity models

*This constraint can be relaxed in order to change the characteristic size of the network and thus to tune the resolution at which communities are uncovered (66). A discussion can be found in *SI Text*

$$P_{ij}^{\mathrm{Spa}} = N_i N_j f(d_{ij}) \quad [5]$$

where N_i is, as in Eq. **1**, a notion of importance of node i and where the deterrence function

$$f(d) = \frac{\sum_{i,j|d_{ij}=d} A_{ij}}{\sum_{i,j|d_{ij}=d} N_i N_j}, \quad [6]$$

is the weighted average of the probability $A_{ij}/(N_i N_j)$ for a link to exist at distance d. It is thus directly measured from the data[†] and not fitted by a determined functional dependence, as is often the case (15). By construction, the total weight of the network is conserved as required. Depending on the system under scrutiny, N_i may be the number of inhabitants in a city or the degree of a node when it corresponds to a single person in a social network. It is worth mentioning that in the latter case and if the embedding in space does not play a role—i.e., where $f(d)$ is flat—the standard NG model is exactly recovered (*SI Text*).

From now on, let us denote by Q_{Spa} the version of modularity (**3**) whose null model P_{ij}^{Spa} is given by Eq. **5**. Q_{Spa} incorporates nonstructural information about the nodes (i.e., their position in physical space). By definition, Q_{Spa} favors communities made of nodes i and j such that $A_{ij} - P_{ij}^{\mathrm{Spa}}$ is large—i.e., pairs of nodes which are more connected than expected for that distance. Compared to Q_{NG}, Q_{Spa} tends to give larger contributions to distant nodes and its optimization is expected to uncover modules driven by nonspatial factors.

Numerical Validation

Belgian Mobile Phone Data. To compare the partitions obtained by optimizing Q_{NG} and Q_{Spa}, let us first focus on a Belgian mobile phone network made of 571 communes (the 19 communes forming Brussels are merged into one) and of the symmetrized number of calls $\{A_{ij}\}_{i,j=1}^{571}$ between them during a time period of 6 mo (see ref. 38 for a more detailed description of the data). This network is aggregated from the anonymized customer–customer communication network of a large mobile phone provider by using the billing commune associated to each customer. The number of customers in each commune i is given by N_i. This network provides an ideal test for our method because of the importance of nonspatial factors driving mobile phone communication, namely, the existence of two linguistic communities in Belgium:[‡] a Flemish community and a French community mainly concentrated in the north and the south of the country, respectively. As reported in ref. 38, when the weights between communes are given by the average duration of communication between people, a standard NG modularity optimization recovers a bipartition that closely follows the linguistic border.

Both versions of modularity are optimized using the spectral method described in ref. 62. Visualization of the results are shown in Fig. 1. The NG modularity uncovers 18 spatially compact modules, similar to those observed in other spatially extended networks and mainly determined by short-range interactions between communes. Although boundaries of this partition coincide with the linguistic separation of the country (38), the unaware would not discover the existence of two linguistic communities only from Fig. 1. The spatial modularity uncovers a strikingly different type of structure: an almost perfect bipartition of the country where the two largest communities account for about 75% of all communes (see *SI Text* for more details) and nicely reproduce the linguistic separation of the country. Moreover, Brussels is assigned to the French community, in agreement with the fact that approximately 80% of its population is French speaking and despite the fact that it is spatially located in Flanders. The remaining smaller communities (not bigger than 10 communes each) originate from the constraints imposed by a hard partitioning, which is blind to overlapping communities and might thus misclassify Flemish communes strongly interacting with Brussels and communes that have mixed language populations. A similar bipartition is found by considering only the signs of the dominant eigenvector of the modularity matrix (see *SI Text*).

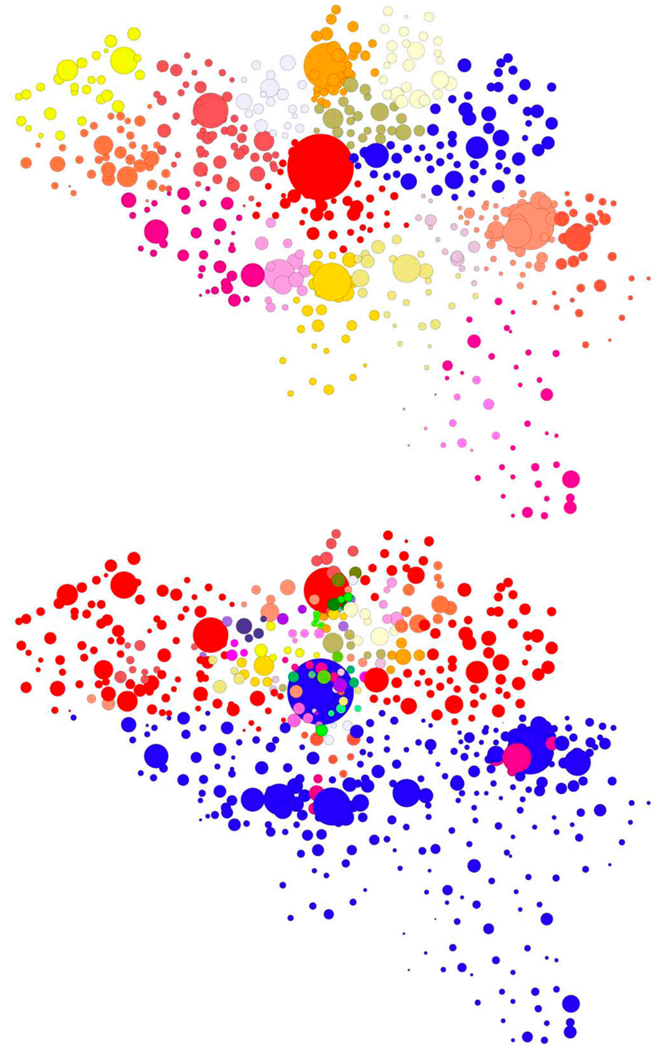

Fig. 1. Decomposition of a Belgian mobile phone network into communities (see main text). Each node represents a commune and its size is proportional to its number of clients N_i. (*Upper*) Partition into 18 communities found by optimizing NG modularity. (*Lower*) Partition into 31 communities found by optimizing Spa modularity.

Statistical Tests. The values for the optimal modularities can be found in Table 1. It is important to stress that a direct comparison of Q_{NG} and Q_{Spa} is meaningless because modularity is a way to compare different partitions of the same graph and so its absolute value is inconsequential. Moreover, the value of modularity is expected to be lower when its null model is closer to the real

[†]In practice, when analyzing empirical data, the distance between two cities is binned such as to smoothen $f(d)$. The dependence of our results on bin size is explored in *SI Text*.

[‡]There also exists a German-speaking community made of 0.73% of the national population

APPLIED MATHEMATICS

SOCIAL SCIENCES

Table 1. Statistical comparison of the modularity measured on the original dataset with the modularity values measured on the randomized data

		Q^{norm}_{obs}	$\langle Q^{norm}_{rand} \rangle$	z score
Weights	spatial	0.0881	0.0049	803
	NG	0.7961	0.7059	55
Positions	spatial	0.0881	0.2383	−90

structure of the data, as it is the case for Q_{Spa}. In order to assess the significance of the uncovered partitions, one needs instead to resort to statistical tests by comparing modularity with that of an ensemble of random networks (60).

Two types of random networks are constructed: (*i*) networks where weights are randomized. Starting from the empirical $f(d)$, we generated weights between two communes i and j according to a binomial of mean $\rho N_i N_j f(d_{ij})$. In the following, we chose $\rho = 1$, thus conserving (up to some fluctuations) the total number of calls in the system and the spatial dependence between nodes. Let us keep in mind that ρ allows to tune the importance of finite-size fluctuations and that $A_{ij}/\rho = N_i N_j f(d_{ij})$ in the limit $\rho \to \infty$; and (*ii*) networks where the geographical position of the nodes is randomized while leaving the weights unchanged. This second ensemble of random networks is radically different from the first one because it keeps the topology of the network unaffected and only randomizes node attributes. Because NG does not make use of geographical information, it is unaffected by this reshuffling. By construction, the effect on Spa is to make space less important by changing the function $f(d)$, thus leading to an expression closer to NG (see *SI Text*). For each type of randomization, we produce $N = 100$ networks and optimize their modularities Q_{NG} and Q_{Spa}.

The significance of the partitions found in the original data is first evaluated by comparing their modularity with that of the randomized data through a z score (60), defined as

$$z = \frac{Q - \langle Q \rangle_{random}}{\sigma}, \quad [7]$$

where σ is the standard deviation across the 100 realizations. Results are summarized in Table 1 and clearly show that the original data are significantly more modular than networks where the weights are randomized. The z score is an order of magnitude larger for the spatial modularity. For the spatial randomization, in contrast, the z score is negative, which reflects the fact that useful information is lost by randomizing node positions and that the resulting randomized null model is further away from reality than the original.

As a next step, we focus on the variability across the uncovered partitions. This step is done by using normalized variation of information (VI) (67), which is a measure of the distance between partitions. VI is equal to zero only when two partitions are identical and is between 0 and 1 otherwise. Results are summarized in Table 2 where we observe that partitions obtained from NG and Spa are genuinely different. In the case of weight randomization, the important point is that VI between partitions uncovered in random networks is much smaller for NG (0.09) than for Spa (0.58), thus indicating that very similar partitions are found by NG across random networks (i.e., only due to spatial interactions between communes). Another interesting point is the high similarity between partitions found by NG in the original data and by Spa in the spatially randomized networks, as their VI is found to be equal to 0.16, in agreement with the fact that Spa becomes similar to NG when space is irrelevant.[§] This observation is confirmed by the similar values of VI between the partitions found by NG and Spa in the original data, as shown in Fig. 1 (i.e., 0.38), and between partitions found by Spa in the original data and in the spatially randomized data (0.35 in Table 2).

[§]It is important to stress that the spatial randomization does not entirely remove the effect of space on network connectivity because self-loops (i.e., intracommune links) are preserved.

Table 2. Average VI measured between the partition found on the original dataset and the randomized ones (Orig–Rand) and the average VI among the randomized dataset (Rand–Rand) for both null models and randomization procedures

		Orig–Rand	Rand–Rand
Weights	spatial	0.54 ± 0.02	0.58 ± 0.02
	NG	0.23 ± 0.02	0.09 ± 0.05
Positions	spatial	0.35 ± 0.02	0.07 ± 0.04

Gravity Model Benchmark. To test the validity of our method in a controlled setting, let us now focus on computer-generated benchmarks for spatial, modular networks. The underlying idea is to build spatially embedded random networks where the probability for two nodes to be connected depends on their distance, as observed in real-world examples, and on the community to which they are assigned. We implement benchmarks in the simplest way by throwing 100 nodes at random in a two-dimensional square of dimension 100×100 and by randomly assigning them into two communities of 50 nodes. Contrary to the previous example, where nodes (communes) could have different sizes, we

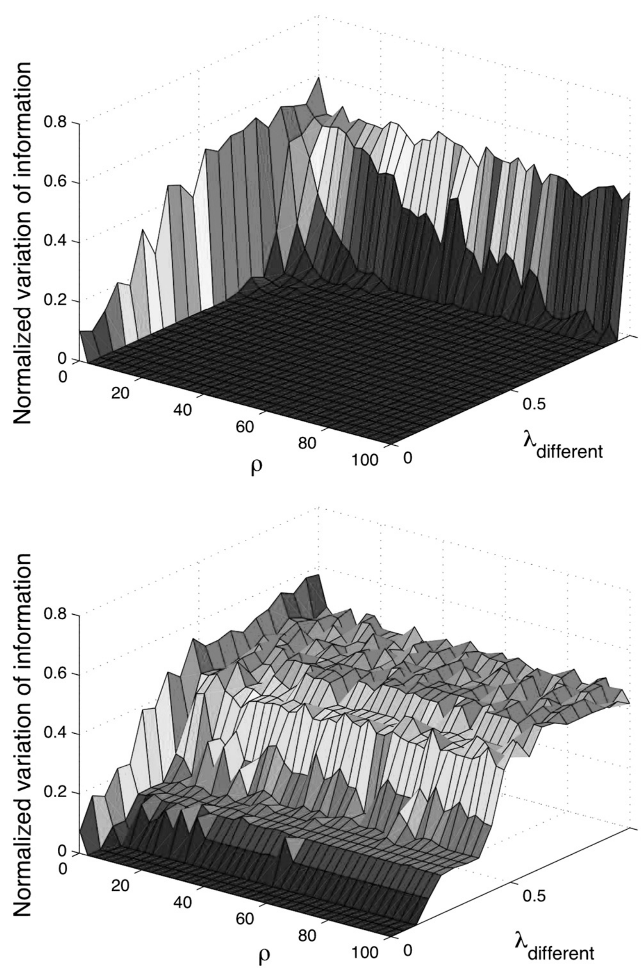

Fig. 2. Variation of information over the ($\lambda_{different}$, ρ) parameter space for Spa (*Upper*) and NG (*Lower*) when tested on the spatial benchmark. Spa is able to recover the correct communities over a wide range of parameters' values, whereas NG fails to find the correct communities almost as soon as the interaction $\lambda_{different}$ is turned on.

assume that all nodes have the same size. The probability that a link exists between nodes i and j has the form

$$p_{ij} = \frac{\lambda(c_i,c_j)}{Zd_{ij}}, \quad [8]$$

where c_i is the community of node i. The function $\lambda(c_i,c_j)$ determines the community linkage. By definition, it is equal to 1 if $c_i = c_j$ and $\lambda_{\text{different}}$ otherwise. When $\lambda_{\text{different}} = 0$, only nodes in the same community are connected, although no distinct communities are present when $\lambda_{\text{different}} = 1$. A normalization constant, Z, ensures that $\sum_{i>j} p_{ij} = 1$. These networks, directly inspired by gravity models, are built by placing $L = \rho N(N-1)/2$ links with probability p_{ij}, where $\rho \geq 0$ determines the density of links in the network. Multiple links are allowed and interpreted as weights. The parameter ρ controls finite-size fluctuations around the expected number of edges Lp_{ij}.

In order to compare the efficiency of Q_{Spa} and Q_{NG}, we generated one realization of the random model for different values of $\lambda_{\text{different}} \in [0,1]$ and $\rho \in [0.01,100]$, and optimized their modularity. As a measure of the quality of the uncovered partitions, we compared them with the known bipartition of the network by using normalized VI. Our simulations show that Q_{Spa} outperforms Q_{NG} and that the improvement becomes larger and larger as the density of links is increased (see Fig. 2). In the limit $\rho \to \infty$, where fluctuations become negligible, our simulations show that Spa perfectly identifies the correct communities for any $\lambda_{\text{different}} < 1$, whereas NG fails even for small values of $\lambda_{\text{different}}$. It is also interesting to note that results presented in Fig. 2 are obtained for single realizations of the random networks (i.e., because when dealing with empirical datasets one does not analyze an ensemble of networks), and yet the precision of Spa is significantly better than that of NG (results smoothed by averaging over several realizations are presented in *SI Text*).

Discussion

Despite the increasing availability of affordable long-distance travel and new communication media, the "death of distance" (68) has been greatly exaggerated (11, 69). Furthermore, the emergence of new technologies entangling physical and virtual worlds has stimulated new research and produced new applications for social and human mobility networks embedded in space (70). This importance of space is not limited to social networks—a broad range of economical and biological networks are also spatially embedded, with strong consequences on their topological organization. The main purpose of this paper has been to find an alternative way to uncover significant patterns in spatial networks. To do so, we have taken advantage of the flexibility of a quantity called modularity defined for community detection. Modularity incorporates a null model, which represents what is expected by chance, namely, the expected probability that two nodes are connected. Unlike the standard null model, we incorporate nonstructural attributes into our null model and use this as a comparison with empirical data. By doing so, we construct null models which portray more closely the network under scrutiny and provide the means to exploit known attributes (e.g., spatial location) in order to uncover unknown ones (e.g., homophilious relations).

We believe that our general framework is suitable for a wide range of networks and that it opens avenues of quantitative exploration of spatially distributed systems. Interesting lines of research include the development of more general null models, for instance by interlacing structural and nonstructural information, and the detection of hierarchies in spatial networks either by tuning the resolution of modularity (66) or by looking for local maxima of the modularity landscape (64) (see *SI Text* for more details). Moreover, our methodology is not limited to situations where distance is measured in physical space as it may be applied whenever one can use node attributes to define a separation between nodes. For instance, in many social networks, age may be a dominant factor, yet by building a null model on the age difference between actors, other types of relationships may be revealed for little extra computational effort. A further advantage is that, by incorporating relevant information, a partitioning approach can be applied even if modules are pervasively overlapping (71, 72).

ACKNOWLEDGMENTS. We would like to thank K. Christensen, M. Gastner, and H. J. Jensen for fruitful discussions, N. Zachariou for proofreading, and P.J. Mucha for sharing his code for modularity optimization. R.L. acknowledges support from the United Kingdom Engineering and Physical Sciences Research Council. This work was conducted within the framework of the European Cooperation in Science and Technology Action MP0801 Physics of Competition and Conflicts.

1. Newman MEJ, Barabási A-L, Watts DJ (2006) *The Structure and Dynamics of Networks* (Princeton Univ Press, Princeton).
2. Barthelemy M (2011) Spatial networks. *Phys Rep* 499:1–101.
3. Gastner MT, Newman MEJ (2006) The spatial structure of networks. *Eur Phys J B* 49:247–252.
4. Kitchin R, Dodge M (2000) Placing cyberspace: Geography, community and identity. *Information Technology, Education and Society* 1:25–46.
5. Amaral LAN, Scala A, Barthelemy M, Stanley HE (2000) Classes of small-world networks,. *Proc Natl Acad Sci USA* 97:11149–11152.
6. Guimerà R, Amaral LAN (2004) Modeling the world-wide airport network. *Eur Phys J B* 38:381–385.
7. Kaiser M, Hilgetag CC (2004) Modelling the development of cortical systems networks. *Neurocomputing* 58:297–302.
8. Barrat A, Barthlemy M, Vespignani A (2005) The effects of spatial constraints in the evolution of weighted complex networks. *J Stat Mech* P05003.
9. Hayashi Y (2006) A review of recent studies of geographical scale-free networks. *IPSJ Journal* 47:776–785.
10. Cummings JN, Kiesler S (2007) Coordination costs and project outcomes in multi-university collaborations. *Research Policy* 36:1620–1634.
11. Liben-Nowell D, Novak J, Kumar R, Raghavan P, Tomkins A (2005) Geographic routing in social networks. *Proc Natl Acad Sci USA* 102:11623–11628.
12. Bassett D, et al. (2009) Cognitive fitness of cost-efficient brain functional networks. *Proc Natl Acad Sci USA* 106:11747–11752.
13. Carrothers V (1956) A historical review of the gravity and potential concepts of human interaction. *J Am Inst of Planners* 22:94–102.
14. Wilson AG (1967) A statistical theory of spatial distribution models. *Transp Res* 1:253–269.
15. Erlander S, Stewart NF (1990) *The Gravity Model in Transportation Analysis: Theory and Extensions* (Brill Academic, Utrecht, The Netherlands).
16. Bhattacharya K, Mukherjee G, Saramäki J, Kaski K, Manna SS (2008) The international trade network: Weighted network analysis and modelling. *J Stat Mech* P02002.
17. Levy M (2010) Scale-free human migration and the geography of social networks. *Physica A* 389:4913–4917.
18. Jung W-S, Wang F, Stanley HE (2008) Gravity model in the Korean highway. *Europhys Lett* 81:48005.
19. Lambiotte R, et al. (2008) Geographical dispersal of mobile communication networks. *Physica A* 387:5317–5325.
20. Krings GM, Calabrese F, Ratti C, Blondel VD (2009) Urban gravity: A model for inter-city telecommunication flows. *J Stat Mech* L07003.
21. Kleinberg J (2000) Navigation in a small world. *Nature* 406:845.
22. Barthélemy M (2003) Crossover from scale-free to spatial networks. *Europhys Lett* 63:915–921.
23. Manna SS, Sen P (2002) Modulated scale-free network in Euclidean space. *Phys Rev E Stat Nonlin Soft Matter Phys* 66:066114.
24. Gastner MT, Newman MEJ (2006) Optimal design of spatial distribution network. *Phys Rev E Stat Nonlin Soft Matter Phys* 74:016117.
25. Wong LH, Pattison P, Robins G (2006) A spatial model for social networks. *Physica A* 360:99–120.
26. Kosmidis K, Havlin S, Bunde A (2008) Structural properties of spatially embedded networks. *Europhys Lett* 82:48005.
27. Scellato S, Mascolo C, Musolesi M, Latora V (2010) Distance matters: Geo-social metrics for online social networks. *Proceedings of the Third Conference on Online Social Networks* (USENIX Assoc., Berkeley, CA), p 8.
28. Colizza V, Barrat A, Barthelemy M, Vespignani A (2006) The role of the airline transportation network in the prediction and predictability of global epidemics. *Proc Natl Acad Sci USA* 103:2015–2020.
29. Girvan M, Newman MEJ (2002) Community structure in social and biological networks. *Proc Natl Acad Sci USA* 99:7821–7826.
30. Danon L, Díaz-Guilera A, Duch J, Arenas A (2005) Comparing community structure identification. *J Stat Mech* P09008.
31. Newman MEJ (2006) Modularity and community structure in networks. *Proc Natl Acad Sci USA* 103:8577–8582.

APPLIED MATHEMATICS

SOCIAL SCIENCES

32. Guimerà R, Amaral LAN (2005) Functional cartography of complex metabolic networks. *Nature* 433:895–900.
33. Gulbahce N, Lehmann S (2008) The art of community detection. *BioEssays* 30:934–938.
34. Porter MA, Onnela J-P, Mucha PJ (2009) Communities in networks. *Not Am Math Soc* 56:1082–1097 1164–1166.
35. Fortunato S (2010) Community detection in graphs. *Phys Rep* 486:75–174.
36. Brockmann D (Feb 2010) Following the money. *Phys World* 31–34.
37. Thiemann C, Theis F, Grady D, Brune R, Brockmann D (2010) The structure of borders in a small world. *PLoS One* 5:e15422.
38. Blondel VD, Krings G, Thomas I (Oct 4, 2010) Regions and borders of mobile telephony in Belgium and in the Brussels metropolitan zone. *Brussels Studies*(42), http://perso.uclouvain.be/gautier.krings/docs/EN_129_BruS42EN.pdf.
39. Ratti C, et al. (2010) Redrawing the map of Great Britain from a network of human interactions. *PLoS One* 5:e14248.
40. Rihll TE, Wilson AG (1987) Spatial interaction and structural models in historical analysis: Some possibilities and an example. *Histoire & Mesure* 2:5–32.
41. Knappett C, Evans TS, Rivers RJ (2008) Modelling maritime interaction in the aegean bronze age. *Antiquity* 82:1009–1024.
42. Bassett D, et al. (2010) Efficient physical embedding of topologically complex information processing networks in brains and computer circuits. *PLoS Comput Biol* 6:e1000748.
43. Meunier D, Lambiotte R, Bullmore ET (2010) Modular and hierarchically modular organization of brain networks. *Front Neurosci* 4:200.
44. Guimerà R, Mossa S, Turtschi A, Amaral LAN (2005) The worldwide air transportation network: Anomalous centrality, community structure, and cities' global roles. *Proc Natl Acad Sci USA* 102:7794–7799.
45. Onnela J-P, Arbesman S, Barabási A-L, Christakis NA (2010) Geographic constraints on social network groups. arXiv:1011.4859.
46. Davis JA (1970) Clustering and hierarchy in interpersonal relations: Testing two graph theoretical models on 742 sociomatrices. *Am Sociol Rev* 35:843–851.
47. Gouldner AW (1960) The norm of reciprocity: A preliminary statement. *Am Sociol Rev* 25:161–178.
48. McPherson JM, Smith-Lovin L, Cook JM (2001) Birds of a feather: Homophily in social networks. *Annu Rev Sociol* 27:415–444.
49. Feld SL (1981) The focused organization of social ties. *AJS* 86:1015–1035.
50. Kossinets G, Watts DJ (2006) Empirical analysis of an evolving social network. *Science* 311:88–90.
51. Crandall DJ, et al. (2010) Inferring social ties from geographic coincidences. *Proc Natl Acad Sci USA* 107:22436–22441.
52. Schelling TC (1971) Dynamic models of segregation. *J Math Sociol* 1:143–186.
53. Watts DJ, Dodds PS, Newman MEJ (2002) Identity and search in social networks. *Science* 296:1302–1305.
54. Evans TS, Lambiotte R, Panzarasa P (2010) Communities and patterns of scientific collaboration. arXiv:1006.1788.
55. Newman MEJ, Girvan M (2004) Finding and evaluating community structure in networks. *Phys Rev E Stat Nonlin Soft Matter Phys* 69:026113.
56. Newman MEJ (2006) Finding community structure in networks using the eigenvectors of matrices. *Phys Rev E Stat Nonlin Soft Matter Phys* 74:036104.
57. Guimerà R, Sales-Pardo M, Amaral LAN (2007) Module identification in bipartite and directed networks. *Phys Rev E Stat Nonlin Soft Matter Phys* 76:036102.
58. Nicosia V, Mangioni G, Carchiolo V, Malgeri M (2009) Extending the definition of modularity to directed graphs with overlapping communities. *J Stat Mech* P03024.
59. Barber MJ (2007) Modularity and community detection in bipartite networks. *Phys Rev E Stat Nonlin Soft Matter Phys* 76:066102.
60. Guimerà R, Sales-Pardo M, Amaral LAN (2004) Modularity from fluctuations in random graphs and complex networks. *Phys Rev E Stat Nonlin Soft Matter Phys* 70:025101(R).
61. Blondel VD, Guillaume J-L, Lambiotte R, Lefebvre E (2008) Fast unfolding of community hierarchies in large networks. *J Stat Mech* P10008.
62. Richardson T, Mucha PJ, Porter MA (2009) Spectral tripartitioning of networks. *Phys Rev E Stat Nonlin Soft Matter Phys* 80:036111.
63. Fortunato S, Barthélemy M (2007) Resolution limit in community detection. *Proc Natl Acad Sci USA* 104:36–41.
64. Sales-Pardo M, Guimerà R, Moreira AA, Amaral LAN (2007) Extracting the hierarchical organization of complex systems. *Proc Natl Acad Sci USA* 104:15224–15229.
65. Good BH, de Montjoye Y-A, Clauset A (2010) The performance of modularity maximization in practical contexts. *Phys Rev E Stat Nonlin Soft Matter Phys* 81:046106.
66. Reichardt J, Bornholdt S (2006) Statistical mechanics of community detection. *Phys Rev E Stat Nonlin Soft Matter Phys* 74:016110.
67. Meila M (2003) Comparing clusterings by the variation of information. *Learning Theory and Kernel Machines*, Lecture Notes in Computer Science (Springer, New York), 2777/2003, pp 173–187.
68. Cairncross F (1997) *The Death of Distance* (Harvard Univ Press, Cambridge, MA).
69. Florida R (October 2005) The world is spiky. *Atlantic Monthly* p 48 The Agenda.
70. Song C, Qu Z, Blumm N, Barabási A-L (2010) Limits of predictability in human mobility. *Science* 327:1018–1021.
71. Ahn Y-Y, Bagrow JP, Lehmann S (2010) Link communities reveal multi-scale complexity in networks. *Nature* 466:761–764.
72. Evans TS, Lambiotte R (2010) Edge partitions and overlapping communities in complex networks. *Eur Phys J B* 77:265–272.