Potential-driven random walks on interconnected systems

Barbara Benigni [*]
Department of Information Engineering and Computer Science, University of Trento, Via Sommarive, 9, 38123 Povo, Trento, Italy and CoMuNe Lab, Fondazione Bruno Kessler, Via Sommarive 18, 38123 Povo, Trento, Italy

Riccardo Gallotti and Manlio De Domenico [†]
CoMuNe Lab, Fondazione Bruno Kessler, Via Sommarive 18, 38123 Povo, Trento, Italy

(Received 12 March 2021; accepted 12 July 2021; published 16 August 2021)

Interconnected systems have to route information to function properly: At the lowest scale neural cells exchange electrochemical signals to communicate, while at larger scales animals and humans move between distinct spatial patches and machines exchange information via the Internet through communication protocols. Nontrivial patterns emerge from the analysis of information flows, which are not captured either by broadcasting, such as in random walks, or by geodesic routing, such as shortest paths. In fact, alternative models between those extreme protocols are still eluding us. Here we propose a class of stochastic processes, based on biased random walks, where agents are driven by a physical potential pervading the underlying network topology. By considering a generalized Coulomb dependence on the distance on destination(s), we show that it is possible to interpolate between random walk and geodesic routing in a simple and effective way. We demonstrate that it is not possible to find a one-size-fit-all solution to efficient navigation and that network heterogeneity or modularity has measurable effects. We illustrate how our framework can describe the movements of animals and humans, capturing with a stylized model some measurable features of the latter. From a methodological perspective, our potential-driven random walks open the doors to a broad spectrum of analytical tools, ranging from random-walk centralities to geometry induced by potential-driven network processes.

DOI: 10.1103/PhysRevE.104.024120

I. INTRODUCTION

Communication, transport, and mobility are just a few examples of real domains involving the movement of particular entities—information, commodities, animals, or human beings—from one place A to another place B. *How* point B is reached from A varies according to the routing strategy adopted by the agents. Notably, routing strategies on networks are mostly based on shortest path or random walk protocols. On the one hand, shortest path protocols assume global knowledge of the network, driving agents to follow the routes minimizing a cost function among the available ones. However, this approach is not always feasible because we lack either global knowledge about the system or the computational resources to exploit it, such as for searches in the World Wide Web. On the other hand, random walk protocols rely only on local knowledge and, consequently, take a longer time to reach a destination, therefore providing a misrepresentation of real-world navigation processes where, for example, it is rare for an agent to randomly proceed on her journey.

The limits of this dichotomy is manifest if we consider human navigation, where route choice strategies are, in the first approximation, driven by the objective of minimizing time and travel costs [1]. However, route choice can be at same time also influenced by other external factors, such as traffic restrictions, unexpected queuing times due to special events [2], the application of congestion pricing schemes [3], the relative pleasantness of routes [4], or other preferences such as the desire of minimizing the number of turns [5] which may strongly variate the cost function between individuals. On the other hand, while being the most efficient way of proceeding, even when considering possible detours due to traffic congestion or other user initiatives, following the shortest path requires a complete knowledge of network topology [6]. Often such knowledge is not available and only partial information on network structure can be considered (i.e., the degree of neighbors). The problem of routing information within a network without global knowledge about the system is related to a variety of applications from neuroanatomy and social sciences to communication and infrastructure engineering. For instance, it has been shown that the small-world topology of some systems, characterized by the presence of long-range links, favors the finding of paths which allow for the efficient delivery of information towards the destination [7] (see also [8] for a review), as well as for the efficient navigation of an interconnected system, the latter being favored by the presence of a latent metric space [9]. Recently, the trade-off between information routing through the shortest path, network entropy, and stability has been pointed out [10]. Even when complete information is available, human navigation is limited by the cognitive resource that can be assigned to the task [11], which has as consequences the need for

[*]bbenigni@fbk.eu
[†]mdedomenico@fbk.eu

heuristics to simplify mental representation of the space [12] and the over-reliance on habitual routes [13]. The combination of all these uncontrollable and individual factors yields to the ensembles of trajectories empirically observed at a high level of randomness. Regardless of the urban layout, a significant portion of trajectories within a city does not follow the shortest path between two specific origin and destination points [14], but rather prefers some other eligible paths contained in the ellipse generated from these origin and destination points, coinciding with the two foci of such an ellipse [15]. As a consequence, even if it is known that the human routing is based on cost minimization, for some scenarios of analysis random walks have also been proposed as suitable alternative to develop routing strategies on networks [6,16,17].

It is thus clear that real human trajectories over transportation networks fall between these two opposite paradigms, shortest path routing and random walks. The same can be said in other notable cases such as the flow of information across the brain [18] or animal movements, where the "base" model is that of random walks and diffusion, but observed paths often display a high level of correlation at the microscopical level [19] and are ultimately driven by optimized strategies reached through evolution [20]. There is therefore a need for a continuous spectrum of dynamics between shortest path and random walk routing that can describe paths balancing efficiency and randomness integrating at the same time local and global information [18]. Moreover, the need for overcoming deterministic routing strategy and embracing a degree of randomness is motivated by a realm of ordinary and extraordinary situations where uncertainty eventually proves to be useful. Avoiding congestion and/or avoiding predictability of a routing strategy for security purposes are a few common examples of such an eventuality. This problem is known in the literature as the randomized shortest path problem (RSP) [21]. As illustrated above, in the field of human mobility, the ensembles of trajectories empirically observed exhibit emergent behaviors that overcame the shortest paths, demonstrating that humans do not make optimal decisions, but suboptimal ones. In the human brain, a phenomenon known as neuroplasticity enables brain neural networks to change, reorganize, and grow leading to significant implications for healthy development, learning, memory, adaptation to changing environment, and recovery from brain damage [22]. This means, for example, that brain routing strategies are flexible and not fixed to one optimal path [23]. Due to this flexibility, the pattern of cortical activation can change over time allowing for new skills and abilities to be learnt [23]. The behavior of people moving in a city and the one of neural connections evolving in the brain due to neuroplasticity could be lively examples of the effect due to a particular field influencing such a behavior. In the animal kingdom, a phenomenon known as stigmergy drives animals to follow some preferential paths according to the trace left in the environment by other individuals of the same species. Stigmergy is an example of indirect communication where individuals communicate with each other by modifying the surrounding habitat. In this case we could say that a field emerges from the interaction between animals and their environment. An extraordinary example of routing strategies evolving in space based on the availability of resources is given by the brainless slime mold, whose efficiency to form networks is comparable to those of real-world infrastructure networks [24]. This is brilliantly proved in a study where, arranging food in a scattered pattern such as that of Japanese cities around Tokyo makes the slime mold build networks of nutrients which strikingly resemble the layout of the Japanese rail system [24]. Here we could compare the food resources to physical particles giving rise to a field and driving the slime mold towards specific targets. Finally, a system can benefit from continuous exploration, mostly when considering nonstationary conditions: For instance, one can think about the animal foraging in a changing environment, where agents routing towards the optimal path would miss the opportunity to encounter new patches to feed on. Exploitation of what is known and exploration of what is unknown is an ineluctable trade-off in space, society, and even the human mind [25]. Here we take a step further in modeling this trade-off by overcoming pure deterministic routing strategies and allowing for a continuous exploration of the system, embedding our model with information about preexisting fields or emerging fields.

To work in this direction, we build upon the most recent literature about random walks on networks. Two main approaches of *biased* random walk have been proposed that encapsulate any available information on particular network features. The first approach is the *degree-biased random walk* [27–30], where the transition probability from one node to its neighbors is biased according to the degree of its neighbors. Depending on the sign of the bias parameter, the walker will explore the network by visiting the hubs or by passing through poorly connected nodes. The second approach is the *maximal-entropy random walk* [31,32], where the step transitions probabilities are defined in such a way that the walker disperses maximally in the network. Also in this second case, the transition probabilities from one node to another one are biased according to some topological features (e.g., degree), or to any relevant property for the diffusion dynamics (e.g., level of congestion), of its nearest nodes. Recently, a *memory-based random walk* has been proposed to extend the local information up to the next-nearest neighbors [33]. Despite the efforts made to include more and more information about nodes' features up to the second-order nearest neighbors, little is known about routing strategies when the known information is scattered over the network or when it affects only some particular nodes.

In the last decade, in parallel to the family of *biased* random walks, a new framework aimed to interpolate between he shortest path and random walk has emerged, the randomized shortest paths (RSPs) [21,34,35]. The RSPs exploit a thermodynamic formalism by considering the temperature distribution over paths. By adjusting such a temperature parameter the walker mediates between a minimal travel cost and the maximal exploration of the network. A close proposal, similar in intent—i.e., interpolating between shortest path and random walk—yet different in the implementation has been provided to investigate the routing of neural signals [18]. This last proposal suggests a possible model for network communication merging local and global information about topology and generating an alternative kind of biased random walk. For each node, the bias is encoded by a degree of knowledge of the underlying network topology [18]. Specifically, the transition

probabilities from one node to its neighbors depend on two factors: The length of the edge connecting the node with its neighbors (local information) and the degree of knowledge regarding the distance of such neighbors from the target node (global information). Nevertheless, the shortest path between all pairs of nodes is always required to build the transition matrix of such a random walk [18]. In fact, in all cases except for the trivial unbiased random walk, the shortest path between all pairs of node contributes in defining the degree of (global) knowledge of the network, properly modulated by a bias parameter.

In this work, we define a walk, grounded on physics, which tends to minimize distances (like a shortest path, a global feature) while keeping some flexibility in random exploration (like a random walk, a local feature). In particular, we propose a process, the potential-driven walk, that effectively interpolates between shortest path and random walk routing protocols due to the presence of a potential field defined on the top of the network and acting at each node. By combining knowledge of local features and (partial) global information about network topology, we propose a type of biased random walk where the bias is generated by the potential, which in turn can take different functional forms and can be expressed in different ways depending on background assumptions. The potential-driven random walk does not require one to control the randomness of exploration by fixing the entropy spread—or other global variables—on top of the network or to define a temperature parameter regulating the free wandering in the system (as in the case of RSPs), which is a computational advantage besides being a conceptual shift. In our framework, we do not fix or constrain any variables *a priori* (e.g., entropy constraint); instead we establish where to put the potential node. The dynamical process we proposed minimizes the distances while allowing for some flexibility in the exploration of the system, by relying on simple and well-known network measures, properly biased to continuously interpolate between the random walk and shortest path. Thereby, our walker acts as a physical particle which moves according to stochastic rules but potential-driven. Notably, our methodology allows for embedding the transition matrix with information about preexisting fields or emerging fields. Remarkably, our framework is able to reproduce some salient features regarding animal and human movements.

II. POTENTIAL-DRIVEN RANDOM WALK

Let us consider a finite connected graph of nodes $i = 1, \ldots, N$ whose connectivity is defined by the adjacency matrix $\mathbf{A}$. The element $a_{i,j}$ of $\mathbf{A}$ is equal to 1 if a link exists between node i and node j and 0 otherwise. Now let us arbitrarily choose a drifting node ℓ and imagine the following dynamical process: A free wandering walker moving from node i to node j is subjected to a potential responsible for drifting the walk towards (or away from) the drifting node ℓ of the system. The master equation of the potential-driven random walk reads

$$p_j(t+1 \mid \ell) = \sum_{i=1}^{N} T_{ij}(\ell) p_i(t \mid \ell), \tag{1}$$

where $p_j(t+1 \mid \ell)$ is the probability of being in j at time $t+1$ given the presence of ℓ and $T_{ij}(\ell)$ is the transition probability of the potential-driven random walk from node i to j. Specifically, we proposed a biased transition matrix whose elements are defined by

$$T_{ij}(\ell) = \frac{c_j(\ell) a_{ij}}{\sum_{n=1}^{N} c_n(\ell) a_{in}}. \tag{2}$$

It is to be noticed that the denominator serves as a normalization factor and n is another index indicating the network nodes. At this point we have to define what is the bias factor $c_j(\ell)$. As stated before, the bias generated by the potential can take a different functional form. To define the bias of the potential-driven random walk, we consider the combination of two fundamental variables: (1) the topological distance from the potential $d(\ell)$ and (2) the node's degree k. Here we assume that the potential has a gravitational-like form, and it is centered in a node ℓ so that a walker in node i will randomly select a neighbor j while (1) being attracted (repulsed) by ℓ inversely (directly) proportional to their relative topological distance and (2) being biased by the presence of network hubs in its neighborhood. Under these assumptions the bias factor $c_j(\ell)$ reads

$$c_j(\ell) = d_{j\ell}^{\gamma} k_j^{\beta}, \tag{3}$$

where $d_{j\ell}$ is the topological distance between j and the potential ℓ and k_j is the degree of node j, and γ and β are the bias parameters. It is to be noticed that, when there is no dependence on any other node in the network but only of the neighbor at most, i.e., when $c_j(\ell) = c_j$, we would have the classical biased random walk. When, instead $c_j = \text{const}$, it is easy to show that the classical nonbiased random walk is restored. In Fig. 1 we give an illustrative example of the so-defined potential-driven random walk. For ease of reading, only the topological distance from potential is considered in the bias factor [i.e., $c_j(\ell) = d_{j\ell}^{\gamma}$], which actually represents the innovative aspect of this work. By tuning the bias parameters γ and β we define the dependence of the process on the potential and on node degree. When $\gamma < 0$ the walker is drifted towards the node ℓ, while $\gamma > 0$ implies a repulsive potential from ℓ, and $\gamma = 0$ is the classical biased random walk. Hence, in the case of attractive potential, the target node of a potential-driven random walk coincides with ℓ. When $\beta < 0$ the motion is biased towards poorly connected nodes, while if $\beta > 0$ the walker will favor hub nodes. In the case $\beta = 0$ (and $\gamma = 0$) the classical random walk is restored. To sum up: γ defines where the walker is potential-driven, towards or away from ℓ, and β establishes how the walker is driven, i.e., towards hubs or poorly connected nodes. By considering the stochastic process of the potential-driven random walk in terms of its Markov chain representation we can characterize the walk through the steady-state distribution $\mathbf{p}^*$ and the mean first passage time matrix **MFPT**. In fact, if the Markov chain is ergodic, i.e., every state of the chain can be reached by any other state (the graph is connected), then the steady-state distribution $\mathbf{p}^*$ exists and is unique and

$$\lim_{t\to\infty} T^t(\ell)\mathbf{p}(0) = \mathbf{p}^*(\ell) \quad \forall \mathbf{p}(0), \tag{4}$$

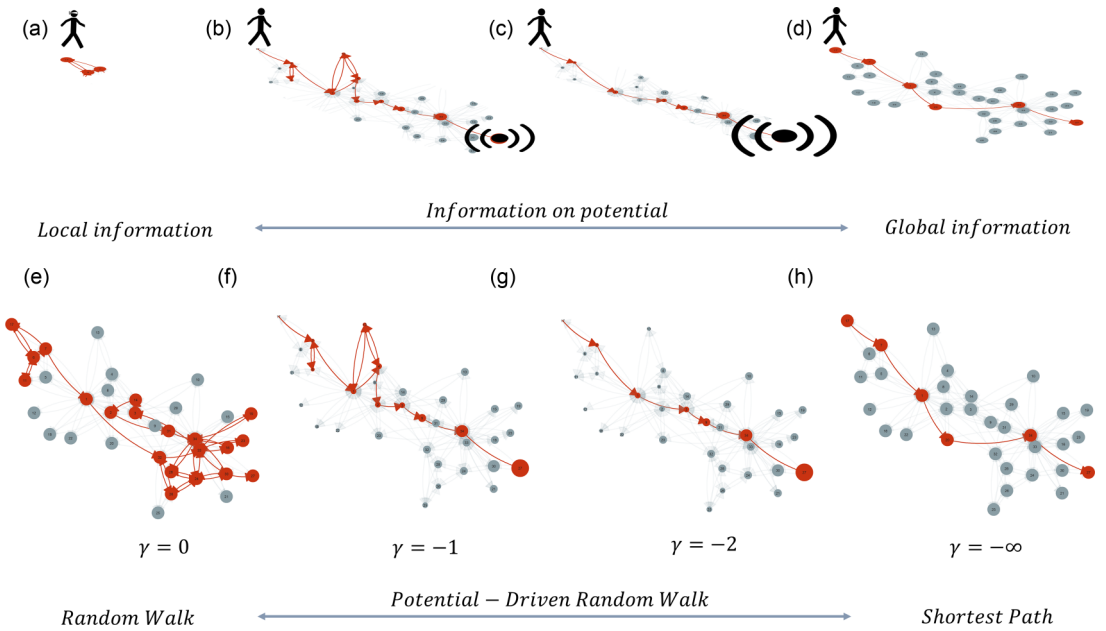

FIG. 1. *Interpolating between random walk and shortest path: The potential-driven random walk.* Illustrative example of potential-driven random walk with different values of bias parameter γ on the Zachary's karate club network [26]. The group of networks on the top half of the figure [panels (a)–(d)] encode the same concept of the group on the bottom [panels (e)–(h)] but in terms of available amount of information. Specifically, the random walker in (a) knows ("sees") nothing about the network topology, while the walkers in (b) and (c) are driven by the potential in ℓ (stylized by a black dot with waves) and they thus know ("feel") something about network topology. Finally, the walker in (d) knows ("measures") everything about network topology and can therefore take the shortest path. Red nodes and red links encode the path followed by the potential-driven random walker varying γ. When $\gamma = 0$ in (a) and (e) the classic random walk is restored. When $\gamma = -1$ in (b) and (f) and $\gamma = -2$ in (c) and (g) we are in the case of an *attractive* potential-driven random walk. When $\gamma = -\infty$ in (d) and (h) we are in the case of the shortest path. Note that the size of the nodes in the case of a potential-driven random walk, i.e., in (b), (c), (f), and (g), is inversely proportional to the distance from ℓ, which corresponds to the target node (node id = 27).

where $\mathbf{p}(0)$ is the distribution of the initial state. Following a procedure similar to the one presented in [31], below we derive the analytical expression for the steady-state distribution $\mathbf{p}^*(\ell)$ of the potential-driven random walk.

Let us consider the probability to go from node i to node j in t time steps $P_{i\to j}(t)$. If the network is undirected, then $a_{ij} = a_{ji} \forall i, j$, and so the relation between the probability of going from i to j in t time steps and the probability of going from j to i in the same time can be defined as [31]

$$b_i c_i P_{i\to j}(t) = b_j c_j P_{j\to i}(t), \tag{5}$$

where $b_i = \sum_j a_{i,j} c_j(\ell)$. For the steady-state distribution $\mathbf{p}^*$, the the same relationship applies so that $b_i c_i p_j^* = b_j c_j p_i^*$, and, through the detailed balance condition, it can be shown that

$$p_i^*(\ell) = \frac{b_i(\ell)c_i(\ell)}{\sum_v b_v(\ell)c_v(\ell)} = \frac{\sum_j a_{i,j} d_{j,\ell}^{\gamma} k_j^{\beta} * d_{i,\ell}^{\gamma} k_i^{\beta}}{\sum_v \sum_j a_{v,j} d_{j,\ell}^{\gamma} k_j^{\beta} * d_{v,\ell}^{\gamma} k_v^{\beta}}. \tag{6}$$

Note that, in the case γ and β are equal to 0, the classic random walk is restored:

$$p_i^* = \frac{\sum_j a_{i,j}}{\sum_v \sum_j a_{v,j}} = \frac{k_i}{2m} \tag{7}$$

with m corresponding to the number of edges in the network. At this point, we probe the mean first passage time matrix **MFPT** of the potential-driven random walk. The **MFPT** matrix defines the average number of steps required to a walker starting in node i to reach a specific node j. To obtain the **MFPT** matrix we followed the matrix solution proposed by [36]. Particularly, we make use of the fundamental matrix $\mathbf{Z}$:

$$\mathbf{Z}(\ell) = (\mathbf{I} - \mathbf{T}(\ell) + \mathbf{W}(\ell))^{-1}, \tag{8}$$

where, in our case, $\mathbf{I}$ is the identity matrix, $\mathbf{T}$ is the transition probability matrix of the potential-driven random walk, and $\mathbf{W}$ is a matrix having all rows equal to $\mathbf{p}^*(\ell)$. For an ergodic Markov chain, the entries $m_{ij}(\ell)$ of the mean first passage time matrix **MFPT** can be obtained from the fundamental matrix $\mathbf{Z}(\ell)$ as [36]

$$m_{ij}(\ell) = \frac{z_{jj}(\ell) - z_{ij}(\ell)}{p_j^*(\ell)}. \tag{9}$$

Finally, we can compute the estimated variance numerically, while its analytical formulation lies outside the purpose of this current work. For the reader interested in proceeding with this formulation we refer to [37].

III. ANALYSIS OF SYNTHETIC NETWORKS

We now have defined the key descriptors for the dynamical process of the potential-driven random walk, i.e., the steady-state distribution and the mean first passage time matrix. In this section, we put these indicators into play on different network topologies. In particular, we show how the stationary distribution varies according to the presence of attractive or repulsive potential(s) in the case of one single potential (monopole), two potentials (bipole), and more than two

potentials (multipole) on top of ordered topology networks, i.e., on lattices. Specifically, in the case of a monopole, a single node is considered as a source of potential—which can be either attractive or repulsive—while in the case of a multipole a scenario with multiple sources of potential (nodes)—attractive, repulsive, or both—is considered. We prove as well that our approach is effective for computing the mean first passage time matrix by comparing analytical results with simulated results on disordered network topologies, i.e., on random networks, scale-free and small-world networks, random geometric graphs, (hierarchical) stochastic block models, and Lancichinetti-Fortunato-Radicchi (LFR) networks. Finally, we provide a measure that defines to what extent the potential-driven random walk deviates from the shortest path in different topologies according to the value of bias parameter γ and β. We call this measure the straightness index on networks, reminiscent of the homonymous index defined by Batschelet in 1981 [38] to define the tortuosity of animals' paths in the physical space.

A. Lattices

Let us suppose the following scenarios. Imagine one is in a city and the agent has to go, for instance, to the mall. Which route would she (most likely) follow? What if there is a more convenient mall around or multiple ones? One can also imagine the scenario where, in addition to a potential destination (e.g., a mall), there are also restricted traffic zones that one should avoid [39]. How does the knowledge of all this information shape one's routing strategy? In this section, we illustrate how the potential-driven random walk provides a suitable way to model such scenarios in a physically grounded and elegant way, giving plausible answers to these questions. Before illustrating the modeling of the aforementioned scenarios, we generalize the bias factor and the steady-state distribution to the case of a multipole, as in following Eq. (10) and Eq. (11) respectively:

$$c_j(\ell_1,\ldots,\ell_L) = k_j^{\beta}\prod_{\ell=\ell_1}^{\ell_L} d_{j\ell}^{\gamma}, \tag{10}$$

where ℓ are the variables indicating the potential nodes within the system

$$p_i^*(\ell_1,\ldots,\ell_L) = \frac{\sum_j a_{i,j}k_j^{\beta}\prod_{\ell=\ell_1}^{\ell_L} d_{j\ell}^{\gamma} k_i^{\beta}\prod_{\ell=\ell_1}^{\ell_L} d_{i\ell}^{\gamma}}{\sum_v\sum_j a_{v,j}k_j^{\beta}\prod_{\ell=\ell_1}^{\ell_L} d_{j\ell}^{\gamma} k_v^{\beta}\prod_{\ell=\ell_1}^{\ell_L} d_{v\ell}^{\gamma}}. \tag{11}$$

It is to be noticed that the generalization to the multipole requires the product of the distance from the nearest neighbors (j) to all potential nodes ($\ell_1,\ldots,\ell_L$).

Let us consider a 9×9 regular lattice (81 nodes) and first pick a drift node ℓ on top of this lattices [as in the first lattice of Fig. 2(a)]. Now imagine a walker, subjected to the presence of such a drift node ℓ, i.e., a potential-driven random walker. According to the nature of this potential—repulsive or attractive—the dynamical process described by the master equation (1) will drive the walker respectively far from the node ℓ or towards node ℓ. This is well described by the value of the steady-state (SS) distribution, represented by the node's size, in the first lattice of Fig. 2(b)—repulsive potential—and in the first lattice of Fig. 2(c)—attractive potential. The SS distribution has been computed by means of Eq. (6) considering a linear dependence from the distance from ℓ (i.e., $\gamma=\pm 1$). Dealing with lattices, we neglected the bias on the degree (i.e., we put $\beta=0$) in computing the steady state. Curiously, in the case of repulsive potential [first lattice of Fig. 2(b)] the highest probability of the steady state is not in the farthest node from ℓ (id node = 81), which is a node on the border. Instead, the highest value of the S distribution corresponds to the farthest node from ℓ of maximum degree (id node = 71). In fact, the SS distribution of a node i depends on its distance from ℓ but also on the distance of all its neighbors from ℓ [see Eq. (6)]. Therefore, being in a node i having neighbor nodes far from ℓ contributes to increase the probability of being in that node i. The same is true in the case of a repulsive bipole and multipole [second and third lattices of Fig. 2(b)]. Specifically, in the case of a bipole the highest probability of the steady state divides the four nodes of highest degree, which are simultaneously the farthest nodes from the two repulsive potentials (id nodes = 8, 18, 64, 74). Similarly, in the case of a multipole, the nodes where the walker is most likely to be when time goes to infinity are the most distant from the four poles and the ones of maximum degree simultaneously. The situation is completely reversed in the case of an attractive potential. As can be seen in the first lattice of Fig. 2(c), the drift node ℓ generates a basin of attraction that drives the walker toward the neighbors of ℓ and ℓ itself. In particular, the probability of being in ℓ when time goes to infinity is 0.5 (we are in the case of linear dependence from the distance from ℓ, i.e., $\gamma=-1$). The rest of the SS distribution equally divides between the four nodes nearest to ℓ. A similar situation appears in the case of an attractive bipole and multipole [second and third lattices of Fig. 2(c)]. Specifically, in the case of a bipole two basins of attraction arise around the two potentials , but the steady-state probability divides unequally around them. In fact, the nearest neighbors of one pole which are also the nearest to the other pole (the "internal nodes") have a greater occupation probability than the external ones. The same unbalanced configuration arises in the case of a multipole, where the "internal" nodes have a greater occupation probability than the "external" ones. It is to be noticed that the pattern around each pole is the same [see the second and third lattices of Fig. 2(c)].

We now reveal what happens in the case of a mixed multipole. Let us pick, for example, two attractive potentials and two repulsive potentials on our lattice [Fig. 2(d)]. Let us assume also that the two flavors of the potential—repulsive and attractive—act with different intensity on the lattice, for example, establishing a linear dependence on the distance from attractive potentials and a quadratic dependence on the one from repulsive potentials. The arising pattern of the steady state is equal—in terms of distribution—and, at the same time, opposite—in terms of direction—to the one resulting from the case of an attractive multipole [second lattice in Fig. 2(c)]. In fact, also in this case the two attractive potentials share half of the occupation probability (i.e., 0.25 each). The second half of the stationary probability is divided among the nearest neighbors of these two attractive potentials, but this time the "external" nodes (id nodes = 12, 20, 62, 70) have a greater probability than the "internal" nodes (id nodes = 22, 30, 52, 60). The reason why this happens is due to the presence of

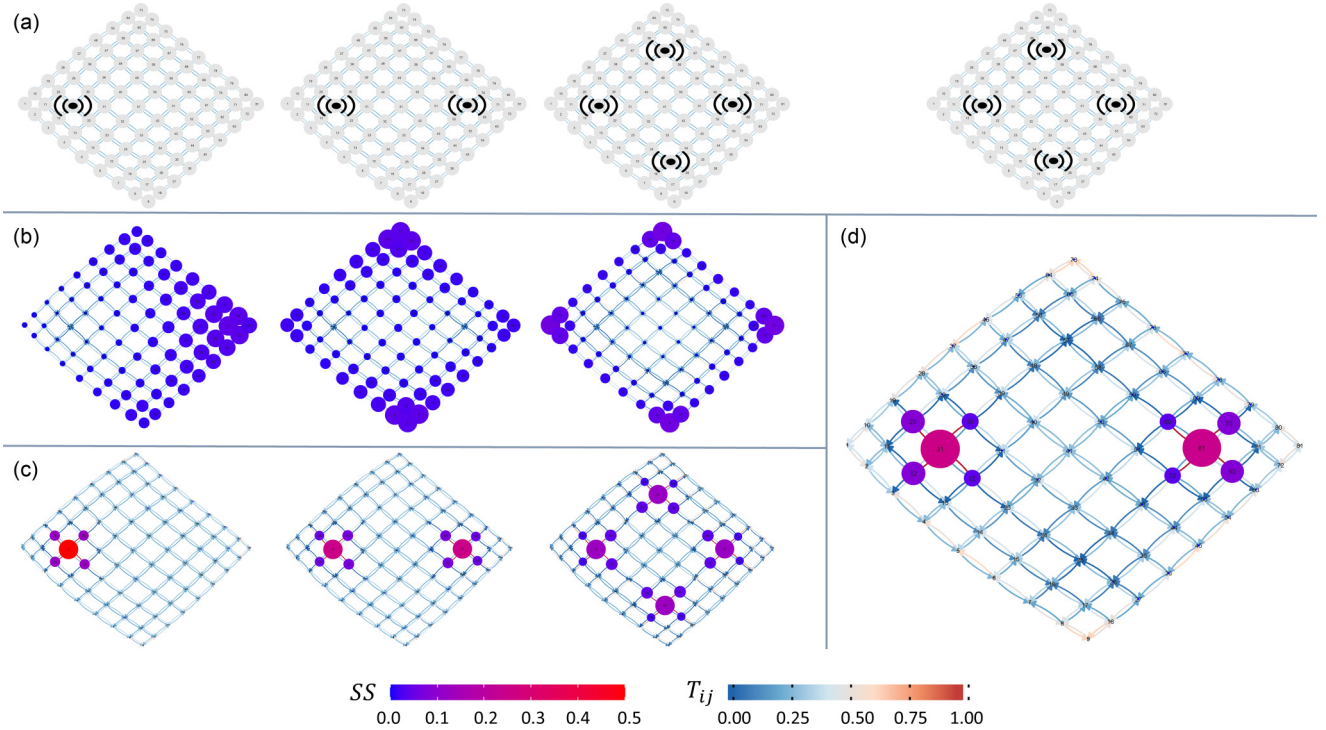

FIG. 2. *Probing the potential-driven random walk on lattices.* In (a) we indicates the position(s) of the potential(s) (i.e., where the potentials are) on top of the lattices, represented as a black dot with waves. From left to right we have the case of a monopole, bipole, and multipole (shown twice), to test distinct scenarios of interest. In (b) we show the values of the steady-state (SS) distribution for each node in the case of repulsive potential(s), specifically for $\gamma = 1$, i.e., what happens if the potentials were repulsive. In (c) we show the values of the SS distribution for each node in the case of attractive potential(s), specifically for $\gamma = -1$, i.e., what happens if the potentials were attractive. In (d) we report a mixed case where potentials are both attractive—extreme left and extreme right—and repulsive—up and down—i.e., what happens if the potentials were mixed. In (d) attractive potentials have $\gamma = -1$, and repulsive potentials $\gamma = 2$. The color of nodes encodes the value of its steady state, and the color bar applies to all lattices. As well, the size of the node is proportional to the value of the steady state, but it is rescaled for each lattice; in this way one can better appreciate the direction towards which the walker is driven. The bigger the node, the bigger the probability to find the walker there at the steady state. Edges' color encode the probability to drive along that link (T_{ij}).

repulsive potentials concomitantly with the attractive ones, which drives the walker towards (one of) the two basins of attraction generated by the attractive poles while keeping the walk as far as possible from the repulsive poles within the basin of attraction, or, more precisely, while preferring nodes far from repulsive potentials among the nearest neighbors of the attractive potentials.

B. Disordered topologies

Usually real systems are far from being adequately modeled by lattices. Brain networks, for example, exhibit small-world properties [40] such as the Internet [41] or social interactions [42]. Other examples are protein-protein interactions which besides being small-world are revealed to be scale-free [43] such as the world airline network [44,45]. Furthermore, some key features of real systems, such as modular structure or spatial embedding, are properly reproduced when considering more complex topologies, such as stochastic block models and random geometric graphs, respectively. For this reason, in the current section we overcome regular lattices to consider more disordered topologies. Specifically, we now investigate the dynamical process of potential-driven random walks considering eight different models, each one encoding one specific feature, covering a broad spectrum of the topological characteristics emerging from real-world networks. These models are Barabási-Albert (BA) [46], Erdős-Rényi (ER) [47], Hierarchical Stochastic Block Model [48] with four dense groups per level (HSBM4), Lancichinetti-Fortunato-Radicchi (LFR) [49], Stochastic Block Model [50] with four dense groups (SBM4), Random Geometric Graphs (RGG) [51], Scale-Free [46] with scaling exponent -2 (SF2), and Watts-Strogatz (WS) [42]. All networks consist of 256 nodes, and their parameters are chosen to obtain an average degree $\langle k \rangle = 12$ or, equivalently, to have about 3000 links, on average.

To prove that our approach is effective for computing the mean first passage time (**MFPT**) matrix we run 200 simulations of the potential driven random walk for each topology, and we compute as well the theoretical values of the **MFPT** matrix by means of Eq. (9). It is to be noticed that we can compute the **MFPT** matrices only for the values of $\gamma \leqslant 0$ (and for all values of β). In fact, for values of $\gamma > 0$ the Markov chain would have a transient state coincident with ℓ (the node where we put the potential). This means that when starting in ℓ there is a nonzero probability to never return in ℓ. For this reason, the Markov chain would not be ergodic, consequently making Eq. (9) nonapplicable. This is the reason why we compute the MFPT only for $\gamma \leqslant 0$ (and $\forall \beta$), or in other words, in the case of attractive potential, i.e., when ℓ

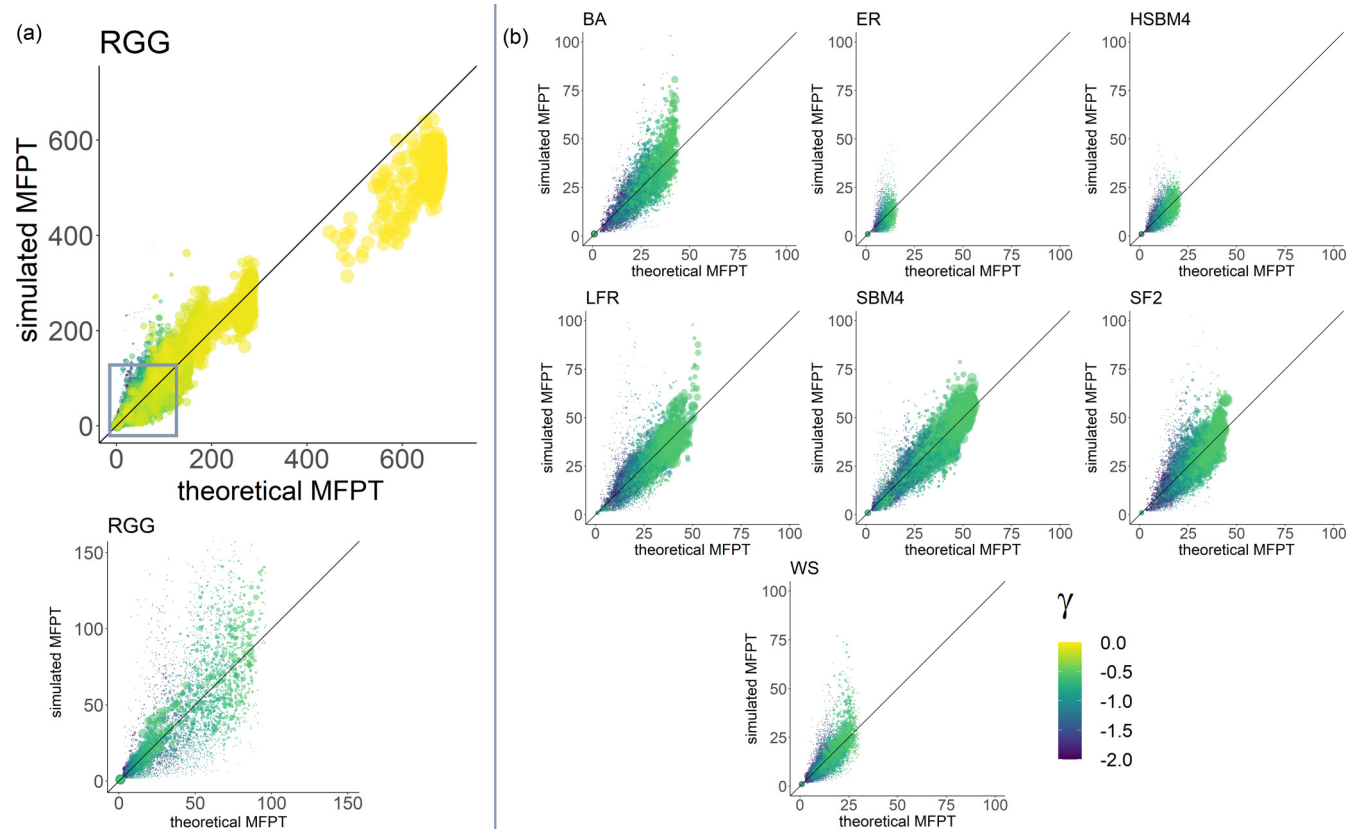

FIG. 3. *Theoretical versus simulated mean first passage time from all possible source nodes to the target node on different network topologies.* In the top of (a) we show the values of the simulated mean first passage time versus the theoretical ones for all values of $\gamma \leqslant 0$ and for $\beta = -2$ considering a random geometric graph (RGG) of 256 nodes. Underneath, we display a zoom of the same graph (in cobalt squares) to better appreciate the values of the mean first passage time for potential-driven random walk (i.e., when $\gamma < 0$). In (b) we directly report the zooms for all the considered network topologies of 256 nodes: Barabási-Albert (BA), Erdős-Rényi (ER), Hierarchical Stochastic Block Model (HSBM4), Lancichinetti-Fortunato-Radicchi (LFR), Stochastic Block Model (SBM4), Scale-Free (SF2), and Watts-Strogatz (WS). The color of the points encodes the values of γ, while the size of the node is proportional to the frequency of that sampled values in the simulations.

coincides with the arrival node. In Fig. 3 we report, as an example, the values of the simulated mean first passage time versus the theoretical ones, for all $\gamma \leqslant 0$ and for $\beta = -2$. In particular we plot the value of the mean first passage time from all the possible source nodes to the target node on the eight different network topologies. By design, the target node is the one where we put the (attractive) potential. In each panel of Fig. 3 the size of the points is proportional to the frequency of the sampled values in the simulations. As can be noticed, most of the values lie on the bisector: These values are also the most frequent in the simulations (bigger points), showing that the theoretical results are in agreement with the simulations. At this point, we investigate how the theoretical values of the MFPT change for different values of bias parameters, with the results shown in Fig. 4. As can be seen in Fig. 4, the trend of the theoretical MFPT is qualitatively similar for all the analyzed topologies, with the values of the MFPT increasing as γ increases, until the peak in $\gamma = 0$, i.e., when the random walk is restored. Interestingly, for some topologies there are critical points in γ, i.e., points where increasing β does not lead to decreased values of the MFPT. This regime shift is visible in the following topologies: LFR, RGG, and SBM4 and partially visible also in BA and SF2.

Let us now randomly pick a simulated values of the **MFPT** matrix from a source node s to a target t. We investigate how good we are in returning the corresponding values of γ and β, i.e., in guessing to what extent the process is potential-driven. The rationale behind the proposed speculation on matching the right value of γ and β is motivated by the intention to recast a stochastic routing strategy to a potential-driven random walk. More specifically, we investigate if and when we are able to infer the values of γ and β given only the information on (1) estimated mean first passage time and (2) the topology of the network. Tracing back the value of γ and β would allow for a more informed definition of a dynamical diffusion process, providing information about the direction towards which and the way a walker reaches a specific destination simply by measuring the time spent on a given topology. This proposal has potential broad applications, e.g., in the context of human and animal mobility to determine if a walker was forced to avoid a given restricted area and/or to establish if the path strategy was preferring hubs or poorly connected nodes, given the values of the mean first passage time (MFPT) from source node to target node and the network topology. To this end, we compare a given estimated value of the MFPT to the whole spectrum of the theoretical MFPT values resulting from all

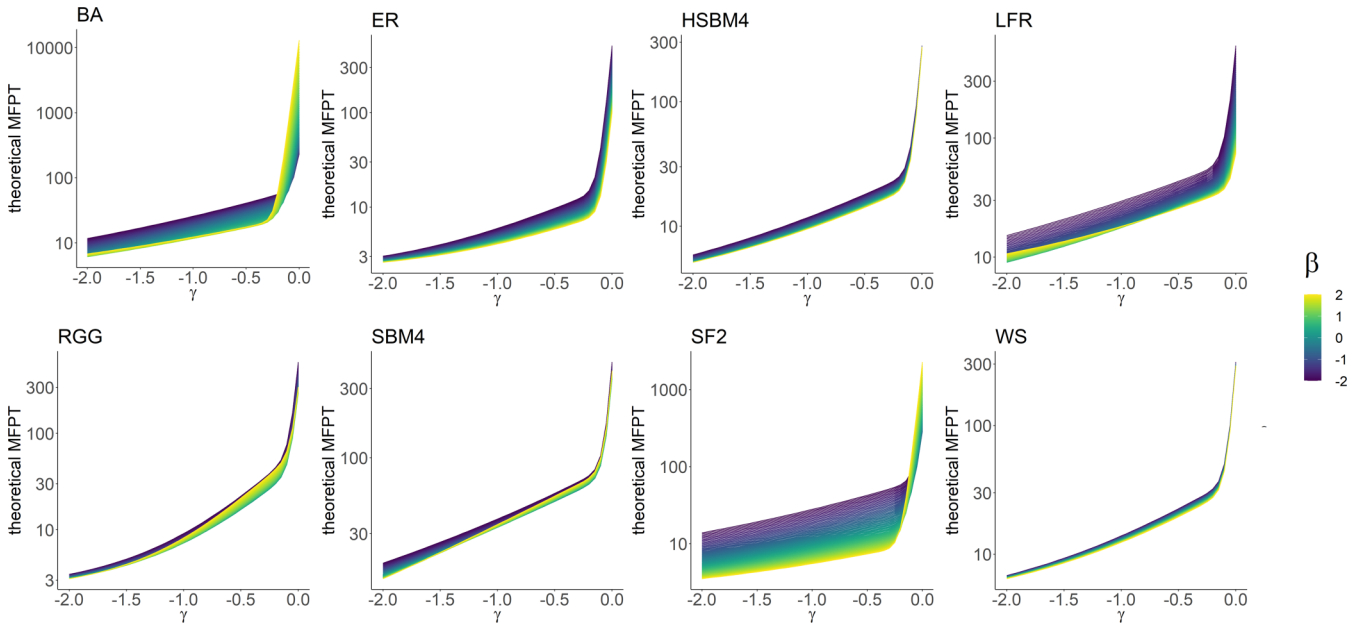

FIG. 4. *Theoretical mean first passage time for all values of* γ *and* β, for all the considered network topologies: Barabási-Albert (BA), Erdős-Rényi (ER), Hierarchical Stochastic Block Model (HSBM4), Lancichinetti-Fortunato-Radicchi (LFR), Random Geometric Graph (RGG), Stochastic Block Model (SBM4), Scale-Free (SF2), and Watts-Strogatz (WS). The color of the line encodes the value of β.

values of γ, in the range $[-2, 0]$, and β in the range $[-2, 2]$. We compute the absolute value of the difference between the estimated value of the MFPT and each theoretical value. The theoretical MFPT giving the minimum of such an absolute value determines the inferred values of γ and β. On average, for all the topologies we match the right value of γ 70% of the time, which is 47% more than what we match for β. This makes us speculate that the process is more sensitive to the bias on the distance from the potential (γ) than to the bias on the degree (β). Specifically, we are more likely to guess the values of γ in the case of stochastic block models and hierarchical stochastic block models, followed by Erdős-Rényi and a random geometric graph. Instead, for what concern scale-free networks, we match the values of γ only 55% of the time. However, the theoretical value of the MFPT closest to the simulated one—in terms of absolute value—always results to being within one standard deviation.

C. Straightness index on networks

In 1981, E. Batschelet, a mathematician devoted to the study of animal path orientation, defined a novel indicator to measure the tortuosity of such animal paths: the *straightness index*. In this vein, we provide a similar indicator for the purpose of measuring to what extent the potential-driven random walk deviates from the shortest path (or how close it gets to the random walk) in different topologies, according to the value of bias parameters γ and β. We call this measure *straightness index* (SI) on networks and, given a source node s and a target node t, we define it as follows:

$$\mathrm{SI} = \frac{\mathrm{SPL}}{L(\ell, \gamma, \beta)}, \tag{12}$$

where SPL is the shortest path length between s and t, while $L(\ell, \gamma, \beta)$ is the length of the potential-driven random walk from s to t, i.e., the path length of a potential-driven random walk in terms of traversed links to reach target node t from source node s, given a potential node ℓ and the bias parameters γ and β. It is to be noticed that a similar formulation was used to develop the straightness centrality for spatial networks [52]. The straightness index on networks is bounded between 0 and 1. Specifically, when $\mathrm{SI} = 0$ we are in the case of a random walk, while $\mathrm{SI} = 1$ corresponds to the shortest path. For $0 < \mathrm{SI} < 1$ we are in the case of a potential-driven random walk. Figure 5 shows how the value of SI changes on different network topologies, varying γ and β. In particular, in Fig. 5(a) we report the values of SI in the eight different topologies considering γ and β as the axes of the heatmaps. The color of the tile encodes the values of SI: the closer the value is to dark blue, the closer the walker behavior is to that of a random walker. Instead, a color tile close to yellow indicates a behavior that tends to the shortest path. By design, when $\gamma < 0$ the potential node ℓ coincides with the target node, while when $\gamma = 0$, the position of ℓ does not affect the process since we are in the case of random walker. As can be seen from these heatmap, the values of SI closest to the shortest path are reached in the Erdős-Rényi network topology, in a random geometric graph, and in the scale-free topology. To better appreciate how the pattern of SI evolves by varying γ we refer to Fig. 5(b). For each topology it is evident that the lower the value of γ, the higher the value of SI, i.e., the closer the walk to the shortest path. This is because the walker is potential driven towards the target node, coincident with the attractive pole ℓ. As γ approaches zero, the value of SI tends to zero as well, denoting a random walk behavior. At $\gamma = 0$, the walker is no longer affected by the potential, and the value of SI dramatically goes to zero with an apparently jump discontinuity in all network topologies. The process at this point is a random walk. Compared to γ, the value of β is less critical to determine how the process approaches a

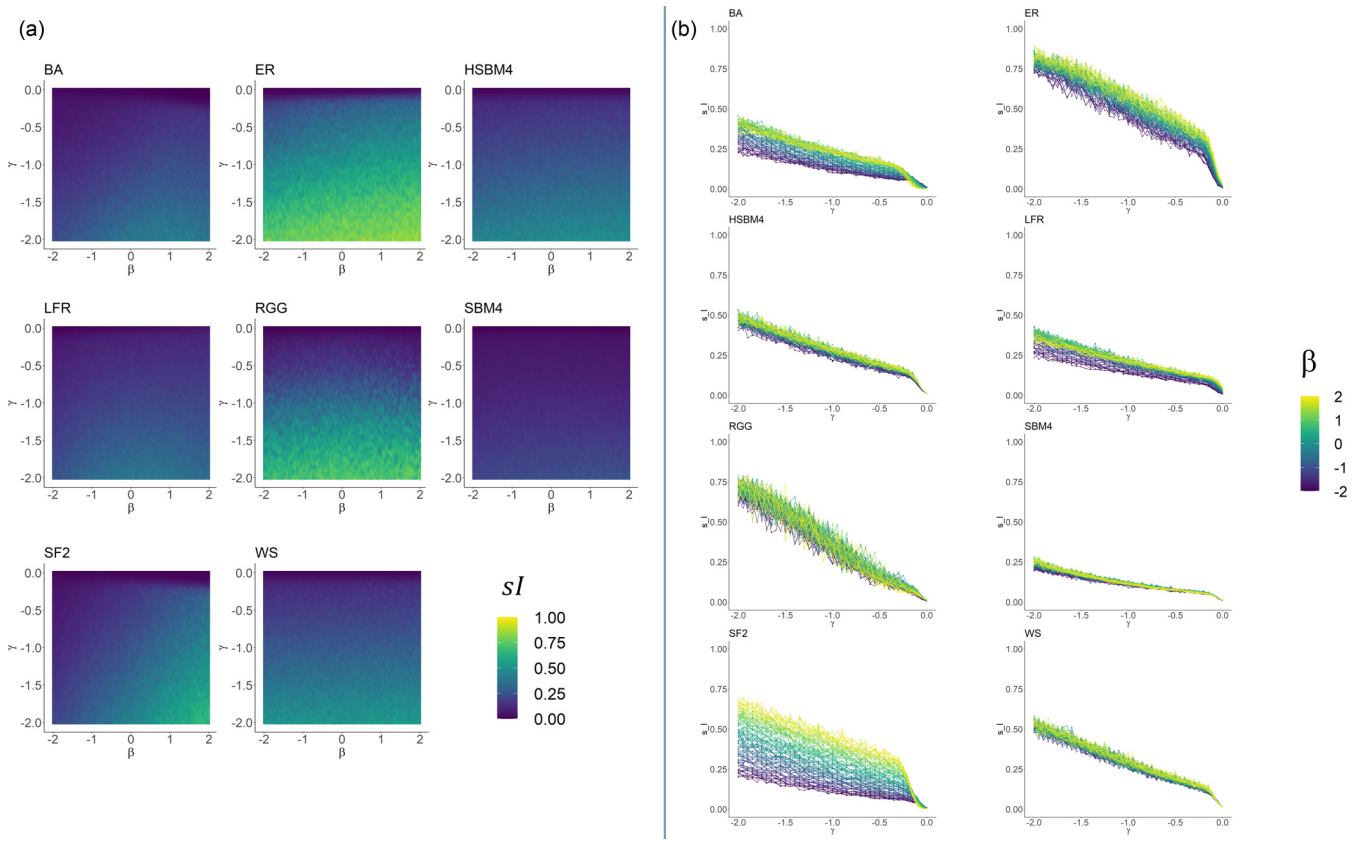

FIG. 5. *Straightness index on networks.* In (a) we report the values of SI for all the considered network topologies of 256 nodes: Barabási-Albert (BA), Erdős-Rényi (ER), Hierarchical Stochastic Block Model (HSBM4) Lancichinetti-Fortunato-Radicchi (LFR), Random Geometric Graph (RGG), Stochastic Block Model (SBM4), Scale-Free (SF2), and Watts-Strogatz (WS). The values of bias parameter γ and β are the axes of the heatmap, while tile color encodes the values of SI. In (b) we express the values of SI as a function of γ, and the color line encodes the value of β.

shortest path behavior rather than a random walk, albeit in different ways for each topology. Quantitatively, the BA, SF2, and LFR topologies seem to be more sensitive to the variation of this bias parameter β, while qualitatively all topologies exhibit similar behavior. For example, β is crucial in the scale-free networks and, to some extent also for BA topologies, to achieve high values of SI, i.e., to be close to the shortest path. This means that, unsurprisingly, in these topologies a walker favoring hubs would reach the target node faster. A slighter dependence from β in determining the values of SI, concurrently with γ, is also shown in HSBM4, LFR, and ER topologies. Conversely, in these case there is not the transition at $\gamma = 0$ as in scale-free networks. Finally, RGG, SBM4, and WS seem not to be affected by the bias on the degree, β, in reaching the target node, as much as they are by γ.

To sum up, a potential-driven random walk with negative low values of γ can resemble a shortest path when considering ER, RGG, and SF2 network topologies. Also for the other topologies (HSBM4, SBM4, LFR, BA, and WS) the most efficient path—in terms of path length—can be reached with lower values of γ, but in this case the path length is, at most, twice the shortest path. Whether lower values of γ than those considered in this study would ensure more efficient route should be investigated in further development of this work. For what concerns the degree bias β, it mostly affects the SF2 topologies and, in a smaller portion, also the BA topologies, while the potential-driven random walk on the other networks seems not to be influenced by this bias in reaching the target node. In other words, for all the analyzed networks it is evident that γ is the leading bias parameter in determining the values of SI. In fact, considering the analyzed range of γ, the potential-driven random walk spans from very short paths, sometimes even very close to the shortest path (as in ER, RGG, and SF2), up to completely different and longer paths, regardless of the bias on the degree (β) in all network topologies. This demonstrates that the potential-driven random walk is able to effectively interpolate between shortest path and random walk.

IV. EMPIRICAL 2D TRAJECTORIES

As discussed in the introduction, the movements of both humans and animals do not strictly follow either the shortest path or the random walk paradigm, but their routing falls somewhere in between. In this section, we show that applying our potential-driven random walks over a very stylized network allows us to successfully reproduce the very broad range of characteristics that are observed in empirical human trajectories.

In most cases, the empirical data collected about human and animal trajectories are embedded into a bidimensional space. Note that the analysis presented in this section comes

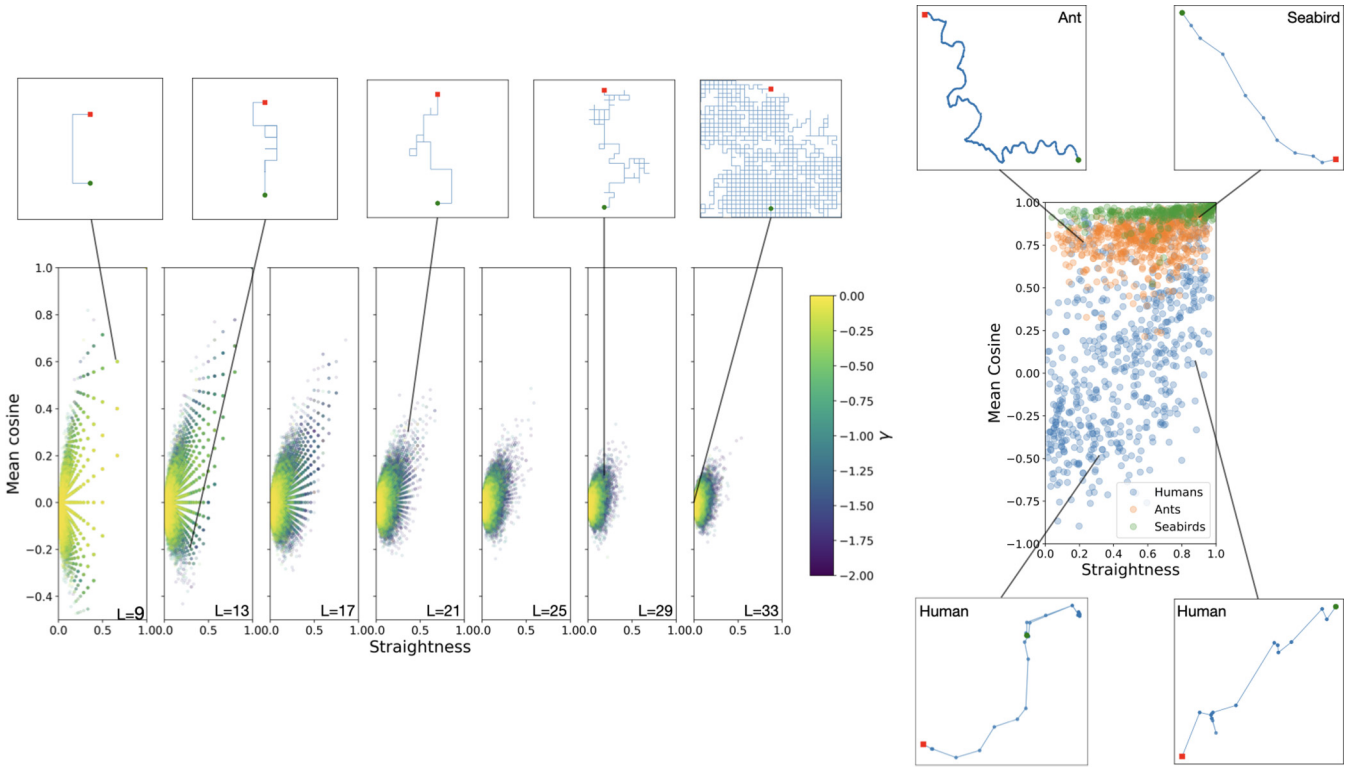

FIG. 6. *Straightness vs autocorrelation diagrams for walks in a 2D space.* Left: Potential-driven walks over a square lattice. We display the values of SI and MC for a range of values of lattice sizes L (different subplots) and γ (different colors). The trajectories above the plot illustrate examples of the trajectories generated, the black line identifying the associated values of SI and MC. For each (L, γ) we generated 1000 trajectories starting from the origin point (green circles in the trajectory examples) and arriving to the destination point (red squares) where the potential is also placed. In these examples, the shortest path is along a straight vertical line. In the examples, we clearly see how the trajectory becomes more and more straight as γ approaches -2 (blue; the upper limit of -2 is chosen here because it is rare to observe larger potential in physical potential fields in nature) and is totally random for $\gamma = 0$ (yellow), which corresponds to a pure random walk. Right: Empirical trajectories of humans, ants, and seabirds. We display the values of SI and MC for three types of empirical trajectories: 600 human trajectories (blue), captured by anonymized GPS data generated by mobile phone applications; 600 ant trajectories (orange), captured optically in laboratory conditions around an artificial nest; and 300 seabird trajectories (green), captured by GPS trackers. In these cases, trajectories are again on a plane, but not limited to follow the lattice topology. Above and below the plot we illustrate two examples of human trajectories and one example each for ants and seabirds (a Wandering Albatross in particular), again with the black line identifying the associated values of SI and MC. The ant trajectory is highly auto-correlated although not straight, while the albatross example is at the same time straight and autocorrelated. The two human examples allow us to show instead a trajectory with negative MC (probably accumulated in the small-scale movements on the top right of the diagram) and straight but not strongly correlated trajectories.

without any biological or ecological context. In fact, the empirical trajectories here are taken into account for the sole purpose of verifying the model reliability in reproducing the diffusion component of both animal and human mobility, whatever the causes or the contexts. To create walks over a network reproducing a similar condition, we place our synthetic walkers over networks defined as square lattices having side L of varying sizes (so that the total number of nodes is $N = L \times L$). For humans, a lattice can be seen as a null model for a street network over which the movements are constrained. On these lattices, similarly to what we have for other networks in the attractive case $\gamma < 0$, we also define a potential node which coincides with the target node. The initial position of the walker and the target node are indicated as green circles and red squares, respectively, in the examples on the top left of Fig. 6. All networks considered for this analysis being square lattices, all nodes with the exception of the borders have the same same degree $k = 4$; therefore the effect of a varying β would be here only an effect of attraction or repulsion at the boundaries. We instead focus on the effect of varying γ in the attractive range $-2 \leqslant \gamma \leqslant 0$ and the lattice size $L \in [9, 13, 17, 21, 25, 29, 33]$.

This approach clearly interprets animal and human trajectories as goal-oriented trajectories, at the end of which a attractive site is located. Our model allows a description where the movements of a living organisms are informed and directed toward a goal, but this information is incomplete or noisy. The stronger the potential, the more information leading towards the goal is readily available. For humans we would expect this goal-oriented behavior to be strong in recurrent mobility, as it is known that the tendency of returning to previously visited locations [53,54] is a clear driver of this behavior. For animals, it will surely depend upon the species, but several examples of attraction-driven movement are known [55], and the cause of attraction can be because either the path is towards one's nest, the lair is known, or information about

the attractiveness of particular areas is available through the organisms' senses.

To characterize both the synthetic trajectories generated over the lattices and the empirical trajectories we use the straightness index SI discussed above as indicator of tortuosity and the mean cosine MC between the trajectory steps as an indicator of autocorrelation defined as the average along the trajectory of the cosine of the turning angle θ between two subsequent steps [19].

As we can see on the left side of Fig. 6, where we represent the distribution of SI and MC in our synthetic trajectories over lattices, for $\gamma = 0$, the paths generated are pure random walks. These random walks are characterized by vanishing values of both SI and MC as the trajectories wrap around themselves. For decreasing values of γ, the trajectories become progressively straighter, as they align more and more to the shortest path, which, given the origin and destination set on our lattices, is simply a straight line. The values of MC characterize the tendency for a trajectory over the lattice to locally deviate from the shortest path or backtracking with steps inverting the direction with respect to the previously taken one. Differing from what often happens for most animal or human trajectories [15,55,56], the potential-driven random walks we used here allow for backtracking, which is possible, although progressively more unlikely, for stronger attractive potentials and decreasing values of γ. Another factor influencing the likelihood of backtracking is here the size L of the lattice, as the larger is the number of steps made, the more likely the event of a deviation from the shortest path even for stronger potentials. For this reason, the synthetic trajectories with the larger values of SI and MC are found for small lattices and strong attractive potentials. This can be seen in our coarser lattices $L = (9, 13, 17)$, where we observe very broad distribution of values of SI and MC spacing (note that in our plots both are limited to values of about 0.8 due to the discrete nature of the trajectories).

On the right side of Fig. 6 we represent the distribution of SI and MC of three types of empirical trajectories: (1) privacy-enhanced GPS trajectories describing the movements of anonymized users of mobile applications who opted in to location-based services through a GDPR compliant framework, randomly extracted from a database provided us by Cuebiq (a location intelligence and measurement platform that collects GPS trajectories from mobile app users who have opted in to provide access to their aggregated location data anonymously) covering the Trentino province, an area of approximately 6000 km^2 in northern Italy (blue); (2) optically tracked trajectories describing the movements of *Temnothorax albipennis* ants while exploring a large arena outside of an artificial nest [57] (orange); and (3) GPS trajectories describing the movements of three types of seabirds (Wandering Albatross, Laysan Albatross, and Streaked Shearwater) [58] over the ocean surface (green). Note that in all three cases considered, the trajectories have been segmented with the Infostop library [59], and only the movements between two stopping points have been analyzed. The distribution of points in the SI-MC diagram for humans is very broad, and most trajectories are very far from both a pure random walk and a straight line. Animals' trajectories present instead a high level of autocorrelation, probably due to the fact that the movements of these animals over the ocean for seabird and in the arena created for the ants is relatively unconstrained. This is different for humans whose movements are constrained over the topology of the street network.

Comparing the diagrams of the synthetic trajectories in the left panel and empirical ones in the right one, we can appreciate how our very stylized model, which includes a single potential source over a square lattice, allows us to already partially reproduce the very broad range of trajectories characteristic of human mobility. A clear improvement in this sense is observed when we used a coarser description of the movement space (small L), which on one hand possibly better describes the limited options dictated by the street network, and on the other hand limits the incurring of backtracking, which is not expected for humans. The trajectories of ants and birds appear instead to be largely more autocorrelated to what we were able to describe with the few ingredients we introduced here.

The exploration in Fig. 6 is, however, limited by the parametric span of $\gamma \in [-2, 0]$. For this reason, in Fig. 7 we expand the simulations, again over a 33×33 square lattice, to $\gamma \in [-12.5, 0]$ with the goal of covering a broader range in the straightness-cosine plane by extending towards the shortest-path area (1,1) our observations. Furthermore, to show the utility of our model as compared to competing models of animal movements describing directed paths, we also simulated 1000 times correlated random walks [19] and Lévy walks [60].

As can be appreciated in Fig. 7(a), similarly to the $L = 33$ example in Fig. 6 (left), we use a relatively large square lattice to generate trajectories embedded into a 2D metric. The origin and destination nodes lie at distance 2 from the sides of the square. The shortest path is in this case a straight path across 29 edges, dividing the square in two [Fig. 7(b)]. In Fig. 7(c), we first can appreciate how the shortest path (purple circle) is characterized by SI = 1 and MC = 1. Random walks (RWs, orange dots) are instead distributed around SI = 0 and MC = 0. We further compare with (1) correlated random walks (red dots) generated via the Poisson process a probability $p > 0$ of having a step of the walk on the same direction of the preceding (otherwise, the step is random as in RWs), thus generating movements constituted by sequences of steps aligned towards the same direction, forming straight displacements of length ℓ, where ℓ follows an exponential probability distribution. (2) Lévy walks (green dots), where the path is again formed by sequences of steps towards the same direction, but where the length of the sequences of aligned steps is instead distributed as a power law $p(\ell) \propto \ell^{-\alpha}$. Comparing the different model, we see how the potential-driven random walk successfully bridges between random walks (0,0) and shortest paths (1,1) as the potential becomes stronger when the value of $-\gamma$ increases from 0 to 12.5. (3) Our potential-driven random walks (blue dots), with $k = 0$ and γ ranging here between 0 and -12.5. The same data are presented in Figs. 7(d)–7(f) to illustrate the dependence over the model's free parameter. The correlated random walk is close to a normal random walk for small p, and then has increasing values of MC as p grows. Since we have origin and destination nodes aligned over the network, for large values of p we observe a walk coinciding with the shortest path. The Lévy walks behave similarly to the

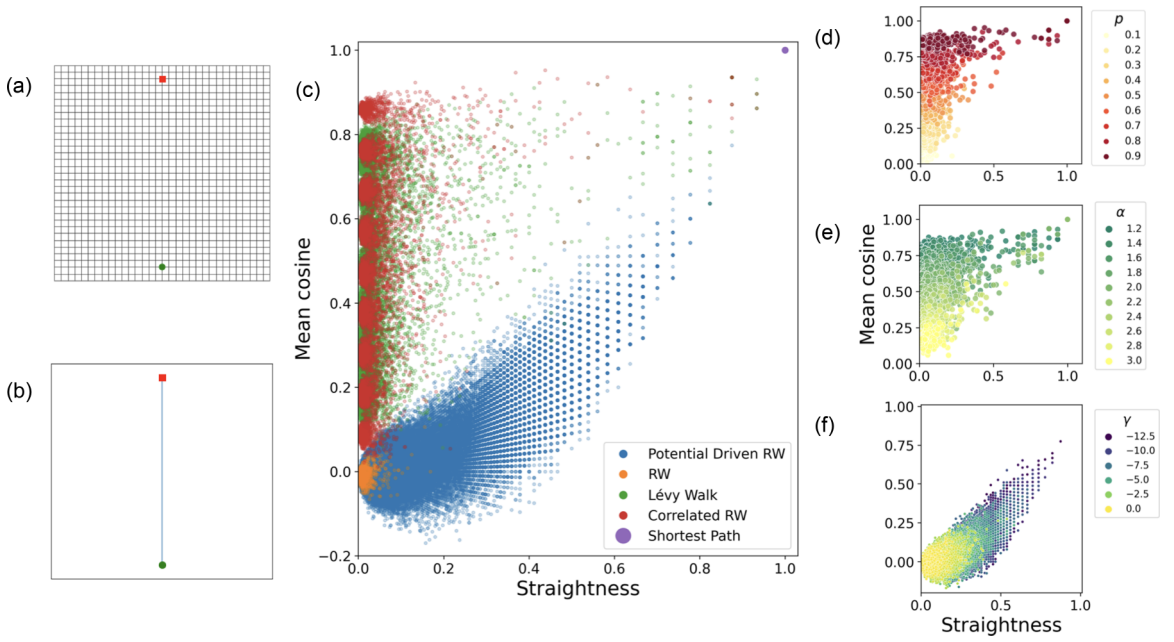

FIG. 7. *Comparison of different walk models on a* 33 × 33 *lattice.* (a) The square lattice over which the trajectories are run from the origin point (green circle) to the destination (red square). (b) The shortest path, in this case a straight line. (c) The SI-MC diagram describing the range of characteristic values for trajectories generated with the different models. (d) Same values as (c) for correlated random walks, where is highlighted the dependency over p. (e) Same values as (c) for Lévy walks, where is highlighted the dependency over α. (f) Same values as (c) for a potential-driven random walk, where is highlighted the dependency over γ.

correlated random walk, with α mostly driving the value of MC and with even more likelihood of having some paths close to the shortest path for large α given the opportunity of making a long jump towards the destination. In the potential-driven random walk, as the potential becomes stronger (larger values of $-\gamma$) we have increasing values of both SI (as the path becomes more similar to the shortest path) and MC $= 1$ (as the path become more straightly directed towards the destination).

These stylized examples of course can be largely improved in order to better describe specific types of behaviors. We offer here three possible directions. First, the network topology surely can be refined from a simple lattice in order to describe the real physical constraint faced by the moving individual. Second, introducing a penalty to backtracking would allow one to create artificial trajectories closer to the trajectories observed pushing towards shapes normally described as composite correlated random walks. Third, introducing multiple potentials activated at specific times, such as, for instance, after the arrival at an intermediary destination, would permit us to account for more complex behaviors like round trips or the chaining of multiple way points in navigating a complex environment.

V. DISCUSSION

In this work, we have introduced a stochastic process, the potential-driven random walk, which effectively interpolates between shortest path and random walk protocols on a network by taking into account at the same time a preference for nodes of high or low degree and the effect of attraction towards the destination node. This dynamical process aims at minimizing the distance (as the shortest path) by considering only partial information about the network, i.e., the position of a drifting node ℓ, and by maintaining a certain flexibility in the exploration of the network (as random walk). We characterized the process described by the potential-driven random walk by means of its steady-state distribution and mean first passage time matrix. We investigated the patterns of these indicators on synthetic networks considering both ordered topologies (i.e., lattices) and disordered topologies (eight different network models), accounting for Barabási-Albert (BA), Erdős-Rényi (ER), Hierarchical Stochastic Block Model (HSBM4), Lancichinetti-Fortunato-Radicchi (LFR), Stochastic Block Model (SBM4), Random Geometric Graphs (RGG), Scale-Free (SF2), and Watts-Strogatz (WS) topologies, providing evidence of agreement between simulations and theoretical expectations. As well, we investigated how the theoretical values of the MFPT change varying the values of bias parameters, pointing out a regime shift in the MFPT—for the LFR, RGG, SBM4, BA, and SF2 topologies—corresponding with some critical points in γ. By generating a large number of paths over networks characterized by a disordered topology, we also showed how the process can be inverted as we were able to successfully infer most of the time (70% overall) the values of the parameter γ only on the basis of the observed the mean first passage time. From a computational point of view, the shortest path scales as $N\log(N)$ and the potential-driven random walk as N^2. Inspired by ecological studies related to the movement of animals in physical space, we proposed a metric, the *straightness index* on networks, quantifying to what extent the potential-driven random walk is close to the shortest path (or to the random walk). Specifically, by tuning the bias parameter related to the distance from the potential, γ, we can define efficient

paths—in terms of path length—which in SF2, ER, and RGG topologies are revealed to be very close to the shortest path when $\gamma = -2$ while stronger potential are necessary in large lattices. For what concerns the bias parameter on the degree, i.e., β, it mainly affects SF2 networks and, in smaller portion, BA topologies. In general, it is evident that γ is the leading bias parameter in determining the values of the *straightness index* on networks, i.e., in directing the process towards a shortest path or a random walk, in all the analyzed topologies.

The straightness index, together with a measure of trajectory autocorrelation on a 2D space, allowed us to also compare empirical trajectories of human and animals with walks generated on a square lattice with a single attractive potential at the destination node. Even within the strong limitations inherent with the very stylized model presented here, we were able to show how our model is able to replicate the broad range of behaviors typical of real human trajectories better than alternative models such as correlated random walks or Lévy walks.

Our results provide a fundamental starting point to build upon the understanding of the movements of agents and information over a network in terms grounded on physics potentials. On the one hand, a more accurate description of real trajectories will be possible by building upon our model, which can be possibly extended to include memory effects and already can account for more complex time-varying potentials that can combine positive and negative charges to describe at the same time the attractiveness of certain areas and the avoidance of others. On the other hand, our framework has a broad spectrum of methodological applications: For instance, one can use potential-driven random walks to analytically define network indicators able to identify the centrality of nodes in a continuous between the family of random walk [61] and geodesic [62] centralities, such as a betweenness and closeness centrality, in both classical and multilayer systems [63]. Such indicators would allow for a set of applications in scenarios such as urban mobility [16] or network neuroscience [18], where human and information flows are known to lie in the gray area between random and geodesic paths. We expect, in particular, that alternative perspectives over these networks will be driven by the fundamental question of what topological conditions make possible to radically change the distribution of the random walk centrality when even small potentials are introduced. Finally, the framework is expected to open doors for the analysis of functional clusters emerging from collective phenomena, where diffusion geometry induced by random walk dynamics [64] can be extended to span from diffusion to geodesic distance, allowing the analysis of the complex interplay between structure and dynamics from alternative perspectives.

ACKNOWLEDGMENTS

The authors acknowledge interesting discussions with the members of the Complex Multilayer Networks Lab and with E. Estrada at the Mediterranean School of Complex Networks 2019. We thank Filippo Privitera and Brennan Lake at Cuebiq for providing us free access to anonymized and privacy-enhanced trajectory data through their Data for Good program.

[1] M. Wardman, Public transport values of time, Transport Policy **11**, 363 (2004).

[2] A. Bassolas, R. Gallotti, F. Lamanna, M. Lenormand, and J. J. Ramasco, Scaling in the recovery of urban transportation systems from massive events, Sci. Rep. **10**, 1 (2020).

[3] A. Solé-Ribalta, S. Gómez, and A. Arenas, Decongestion of urban areas with hotspot pricing, Netw. Econ. **18**, 33 (2018).

[4] D. Quercia, R. Schifanella, and L. M. Aiello, The shortest path to happiness: Recommending beautiful, quiet, and happy routes in the city, in *Proceedings of the 25th ACM Conference on Hypertext and Social Media* (Association for Computing Machinery, New York, 2014), pp. 116–125.

[5] E. Manley and T. Cheng, Exploring the role of spatial cognition in predicting urban traffic flow through agent-based modelling, Transport. Res. A **109**, 14 (2018).

[6] N. Masuda, M. A. Porter, and R. Lambiotte, Random walks and diffusion on networks, Phys. Rep. **716**, 1 (2017).

[7] J. M. Kleinberg, Navigation in a small world, Nature (London) **406**, 845 (2000).

[8] J. Kleinberg, Complex networks and decentralized search algorithms, in *Proceedings of the International Congress of Mathematicians (ICM)* (European Mathematical Society, Madrid, 2006), Vol. 3, pp. 1019–1044.

[9] M. Boguna, D. Krioukov, and K. C. Claffy, Navigability of complex networks, Nat. Phys. **5**, 74 (2009).

[10] E. Estrada, Informational cost and networks navigability, Appl. Math. Comput. **397**, 125914 (2021).

[11] R. Gallotti, M. A. Porter, and M. Barthelemy, Lost in transportation: Information measures and cognitive limits in multilayer navigation, Sci. Adv. **2**, e1500445 (2016).

[12] E. Manley, S. W. Orr, and T. Cheng, A heuristic model of bounded route choice in urban areas, Transport. Res. C **56**, 195 (2015).

[13] S. Larcom, F. Rauch, and T. Willems, The benefits of forced experimentation: Striking evidence from the London underground network, Q. J. Econ. **132**, 2019 (2017).

[14] S. Zhu and D. Levinson, Do people use the shortest path? An empirical test of Wardrop's first principle, PloS ONE **10**, e0134322 (2015).

[15] A. Lima, R. Stanojevic, D. Papagiannaki, P. Rodriguez, and M. C. González, Understanding individual routing behaviour, J. R. Soc., Interface **13**, 20160021 (2016).

[16] M. Akbarzadeh and E. Estrada, Communicability geometry captures traffic flows in cities, Nature Human Behav. **2**, 645 (2018).

[17] A. Ghavasieh and M. De Domenico, Enhancing transport properties in interconnected systems without altering their structure, Phys. Rev. Research **2**, 013155 (2020).

[18] A. Avena-Koenigsberger, X. Yan, A. Kolchinsky, M. van den Heuvel, P. Hagmann, and O. Sporns, A spectrum of routing

strategies for brain networks, PLoS Comput. Biol. **15**, e1006833 (2019).
[19] E. A. Codling, M. J. Plank, and S. Benhamou, Random walk models in biology, J. R. Soc., Interface **5**, 813 (2008).
[20] G. M. Viswanathan, S. V. Buldyrev, S. Havlin, M. Da Luz, E. Raposo, and H. E. Stanley, Optimizing the success of random searches, Nature (London) **401**, 911 (1999).
[21] M. Saerens, Y. Achbany, F. Fouss, and L. Yen, Randomized shortest-path problems: Two related models, Neural Comput. **21**, 2363 (2009).
[22] A. Pascual-Leone, C. Freitas, L. Oberman, J. C. Horvath, M. Halko, M. Eldaief, S. Bashir, M. Vernet, M. Shafi, B. Westover *et al.*, Characterizing brain cortical plasticity and network dynamics across the age-span in health and disease with TMS-EEG and TMS-FMRI, Brain Topog. **24**, 302 (2011).
[23] A. Pascual-Leone, A. Amedi, F. Fregni, and L. B. Merabet, The plastic human brain cortex, Annu. Rev. Neurosci. **28**, 377 (2005).
[24] A. Tero, S. Takagi, T. Saigusa, K. Ito, D. P. Bebber, M. D. Fricker, K. Yumiki, R. Kobayashi, and T. Nakagaki, Rules for biologically inspired adaptive network design, Science **327**, 439 (2010).
[25] T. T. Hills, P. M. Todd, D. Lazer, A. D. Redish, I. D. Couzin, C. S. R. Group *et al.*, Exploration versus exploitation in space, mind, and society, Trends Cogn. Sci. **19**, 46 (2015).
[26] W. W. Zachary, An information flow model for conflict and fission in small groups, J. Anthropol. Res. **33**, 452 (1977).
[27] Z. Eisler and J. Kertész, Random walks on complex networks with inhomogeneous impact, Phys. Rev. E **71**, 057104 (2005).
[28] S. Lee, S.-H. Yook, and Y. Kim, Centrality measure of complex networks using biased random walks, Eur. Phys. J. B **68**, 277 (2009).
[29] A. Fronczak and P. Fronczak, Biased random walks in complex networks: The role of local navigation rules, Phys. Rev. E **80**, 016107 (2009).
[30] M. Bonaventura, V. Nicosia, and V. Latora, Characteristic times of biased random walks on complex networks, Phys. Rev. E **89**, 012803 (2014).
[31] J. Gómez-Gardenes and V. Latora, Entropy rate of diffusion processes on complex networks, Phys. Rev. E **78**, 065102 (2008).
[32] R. Sinatra, J. Gómez-Gardenes, R. Lambiotte, V. Nicosia, and V. Latora, Maximal-entropy random walks in complex networks with limited information, Phys. Rev. E **83**, 030103 (2011).
[33] L. Basnarkov, M. Mirchev, and L. Kocarev, Random walk with memory on complex networks, Phys. Rev. E **102**, 042315 (2020).
[34] F. Bavaud and G. Guex, Interpolating between random walks and shortest paths: A path functional approach, in *4th International Conference on Social Informatics* (Springer, Switzerland, 2012), pp. 68–81.
[35] I. Kivimäki, B. Van Moorter, M. Panzacchi, J. Saramäki, and M. Saerens, Maximum likelihood estimation for randomized shortest paths with trajectory data, J. Complex Netw. **8**, cnaa024 (2020).
[36] C. Grinstead and L. J. Snell, *Introduction to Probability* (American Mathematical Society, Providence, 2006).
[37] J. J. Hunter, Variances of first passage times in a Markov chain with applications to mixing times, Linear Algebra Appl. **429**, 1135 (2008).
[38] E. Batschelet, *Circular Statistics in Biology* (Academic Press, New York, 1981), p. 388.
[39] M. De Domenico, A. Lima, M. C. González, and A. Arenas, Personalized routing for multitudes in smart cities, EPJ Data Sci. **4**, 1 (2015).
[40] D. S. Bassett and E. Bullmore, Small-world brain networks, Neuroscientist **12**, 512 (2006).
[41] R. Albert, H. Jeong, and A.-L. Barabási, Diameter of the World-Wide Web, Nature (London) **401**, 130 (1999).
[42] D. J. Watts and S. H. Strogatz, Collective dynamics of 'small-world' networks, Nature (London) **393**, 440 (1998).
[43] S. Wuchty, Scale-free behavior in protein domain networks, Mol. Biol. Evol. **18**, 1694 (2001).
[44] R. Guimera and L. A. N. Amaral, Modeling the world-wide airport network, Eur. Phys. J. B **38**, 381 (2004).
[45] T. Verma, N. A. M. Araújo, and H. J. Herrmann, Revealing the structure of the world airline network, Sci. Rep. **4**, 1 (2014).
[46] R. Albert and A.-L. Barabási, Statistical mechanics of complex networks, Rev. Mod. Phys. **74**, 47 (2002).
[47] P. Erdős and A. Rényi, On random graphs. I, Publ. Math. Debrecen **6**, 290 (1959).
[48] T. P. Peixoto, Hierarchical Block Structures and High-Resolution Model Selection in Large Networks, Phys. Rev. X **4**, 011047 (2014).
[49] A. Lancichinetti, S. Fortunato, and F. Radicchi, Benchmark graphs for testing community detection algorithms, Phys. Rev. E **78**, 046110 (2008).
[50] P. W. Holland, K. V. Laskey, and S. Leinhardt, Stochastic blockmodels: First steps, Social Netw. **5**, 109 (1983).
[51] J. Dall and M. Christensen, Random geometric graphs, Phys. Rev. E **66**, 016121 (2002).
[52] P. Crucitti, V. Latora, and S. Porta, Centrality measures in spatial networks of urban streets, Phys. Rev. E **73**, 036125 (2006).
[53] C. Song, T. Koren, P. Wang, and A.-L. Barabási, Modelling the scaling properties of human mobility, Nat. Phys. **6**, 818 (2010).
[54] R. Gallotti, A. Bazzani, and S. Rambaldi, Towards a statistical physics of human mobility, Int. J. Mod. Phys. C **23**, 1250061 (2012).
[55] P. Turchin, *Quantitative Analysis of Movement: Measuring and Modeling Population Redistribution in Animals and Plants* (Sinauer Associates, Chicago, 1998).
[56] F. Bartumeus, J. Catalan, G. Viswanathan, E. Raposo, and M. Da Luz, The influence of turning angles on the success of non-oriented animal searches, J. Theor. Biol. **252**, 43 (2008).
[57] E. R. Hunt, R. J. Baddeley, A. Worley, A. B. Sendova-Franks, and N. R. Franks, Ants determine their next move at rest: Motor planning and causality in complex systems, R. Soc. Open Sci. **3**, 150534 (2016).
[58] Y. Yonehara, Y. Goto, K. Yoda, Y. Watanuki, L. C. Young, H. Weimerskirch, C.-A. Bost, and K. Sato, Flight paths of seabirds soaring over the ocean surface enable measurement of fine-scale wind speed and direction, Proc. Natl. Acad. Sci. USA **113**, 9039 (2016).
[59] U. Aslak and L. Alessandretti, Infostop: Scalable stop-location detection in multi-user mobility data, arXiv:2003.14370 (2020).

[60] V. Zaburdaev, S. Denisov, and J. Klafter, Lévy walks, Rev. Mod. Phys. **87**, 483 (2015).

[61] M. J. Newman, A measure of betweenness centrality based on random walks, Social Netw. **27**, 39 (2005).

[62] U. Brandes, A faster algorithm for betweenness centrality, J. Math. Sociol. **25**, 163 (2001).

[63] A. Solé-Ribalta, M. De Domenico, S. Gómez, and A. Arenas, Random walk centrality in interconnected multilayer networks, Physica D **323**, 73 (2016).

[64] M. De Domenico, Diffusion Geometry Unravels the Emergence of Functional Clusters in Collective Phenomena, Phys. Rev. Lett. **118**, 168301 (2017).