A greedy-navigator approach to navigable city plans

Sang Hoon Lee[1, *] and Petter Holme[1, 2, 3]

[1]*IceLab, Department of Physics, Umeå University, 901 87 Umeå, Sweden*

[2]*Department of Energy Science, Sungkyunkwan University, Suwon 440–746, Korea*

[3]*Department of Sociology, Stockholm University, 106 91 Stockholm, Sweden*

Abstract

We use a set of four theoretical navigability indices for street maps to investigate the shape of the resulting street networks, if they are grown by optimizing these indices. The indices compare the performance of simulated navigators (having a partial information about the surroundings, like humans in many real situations) to the performance of optimally navigating individuals. We show that our simple greedy shortcut construction strategy generates the emerging structures that are different from real road network, but not inconceivable. The resulting city plans, for all navigation indices, share common qualitative properties such as the tendency for triangular blocks to appear, while the more quantitative features, such as degree distributions and clustering, are characteristically different depending on the type of metrics and routing strategies. We show that it is the type of metrics used which determines the overall shapes characterized by structural heterogeneity, but the routing schemes contribute to more subtle details of locality, which is more emphasized in case of unrestricted connections when the edge crossing is allowed.

*Electronic address: sanghoon.lee@physics.umu.se

I. INTRODUCTION

Making cities easy to navigate without maps or electronic devices is a desirable objective (although certainly not the only one) for urban planning. Based on models of how humans find their way in a partially unknown environment, one can evaluate street maps and choose where to put new streets to optimize them for better navigability. The setup for such a simulation is to assume every pair of start and finish points for a navigator with partial information and measure the ratio between the shortest path length and the actual path length. In the previous work of ours, an index defined by the ratio is within the range $(0, 1]$ with 0 representing a worst and 1 a perfect navigability of *given* spatial graph layouts [1]. In this work, we instead focus on *designing* optimal transport systems for such concept of navigability under limited resources, which is an important engineering problem and, mathematically, a typical example of the constraint optimization problem [2] where it is well-known that finding the exact optimum of such systems is hard. The problem gets even more complicated if the measure or object function to be optimized is not exactly given. Constructing transport networks on which real navigators move precisely correspond to this situation, due to the fact that the navigability considering real navigators' behavior is not given as a simple mathematical description [1, 3–7].

Our idea of modeling real navigators' behavior is essentially exploiting the spatial information on a local level, and thus it can be modeled as a simple greedy routing strategy [1, 5, 6]. In this respect, our main idea on this work is to use the performance of a simple greedy routing scheme called greedy spatial navigation (GSN) based on the directional information ("taking the road with the direction closest to the direction to the target" introduced in Ref. [1]), along with the real shortest path navigation (SPN) for comparison, to optimize the interconnected structures for a given distribution of vertices on two-dimensional (2D) space under limited resources, i.e., the total length of edges [8–12]. Instead of solving the mathematically cumbersome nondeterministic polynomial time (NP)-complete problem of finding the exactly optimal configurations, we focus on the realistic approach of constructing *shortcuts* greedily from a spanning tree structure as the skeleton [13].

Based on the simulation results from a simple shortcut construction scheme starting with randomly distributed vertices to be connected, we first find that the type of metric (hopping vs. Euclidean distance) is crucial to determine the final structures. In addition, the type of

navigability (shortest vs. GSN path length) also plays an important role, which is reflected by the different degree of inefficiency caused by greedy navigators between those cases. The difference between the two navigability measures is quite prominent if we allow the crossing among edges, which leads to the unrealistic edge condensation for the shortest path length but only the fat-tailed degree distribution for the GSN path length.

II.   SHORTCUT CONSTRUCTION

The aim of the model is to construct navigability-friendly structures under the constraint of resources given by the total length of the edges. This fully deterministic model only depends on the initial configuration of vertices without any free parameter involved. First, assume that we have a set of vertices in a 2D space with their coordinates given, without any prescribed edges. To guarantee the connectivity and minimum initial resource, the minimum spanning tree (MST) T, which is the connected subtree minimizing the total length of edges, is generated with Kruskal's algorithm [14]. This T is the initial state of the evolving graph G so that $G(t = 0) = T$. We define the navigability as the path length averaged over all the pairs of vertices as sources and targets. The navigability is first classified as the one using GSN pathways and the other using the real shortest path (global optimum) in terms of the graph topology and the Euclidean path length. Furthermore, we here distinguish the metrics based on the hopping distance (the number of hops needed to reach the target) and the Euclidean distance (the sum of Euclidean distance along the path). The former is appropriate for the situation when the time spent on the pathway is relatively insensitive to the length of each segment of pathways, or the waiting time for each junction (vertex in this case) is significant, e.g., the airline network or express way network with significant delays in the junctions due to the traffic light. On the other hand, the latter is more suitable for the case when the junction does not play a significant role, e.g., the road network without severe traffic and traffic light. Distinction between GSN and SPN representing whether navigators only use the directional information or have the ability to fully access the real shortest path, combined with those two different metrics, yields four different navigation strategies in total.

At each time step t, the "shortcut" among the vertex pair (not already connected by an edge in $G(t)$), which maximizes the performance of one of the four navigation strategies: Greedy Spatial Navigation with Hopping distance (GSNH) [1], Greedy Spatial Navigation
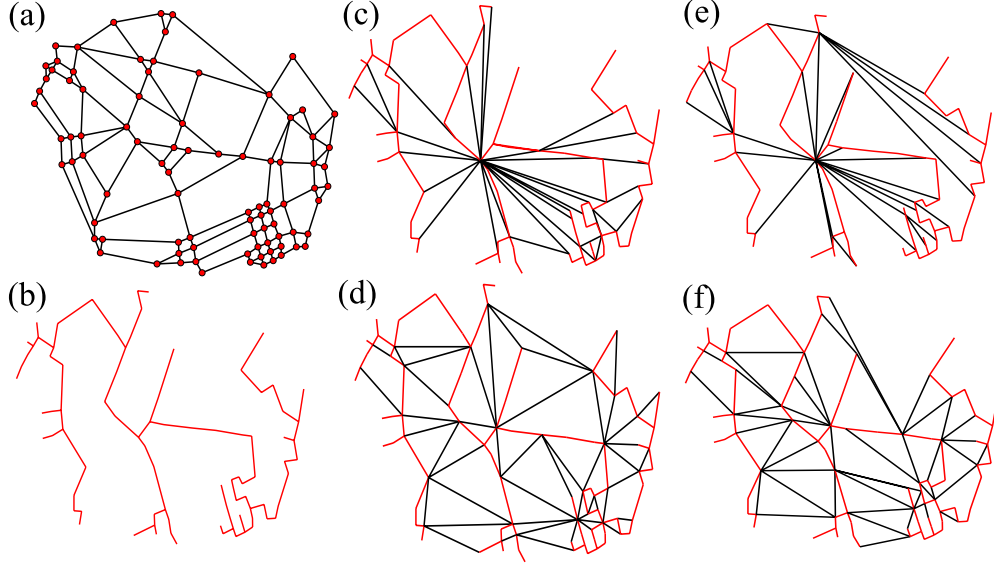
FIG. 1: Original Boston road structure (a), MST (b), GSNH (c), GSNE (d), SPNH (e), and SPNE (f). Red edges are the ones in MST (b), and black edges correspond to shortcut edges added during the evolving process, for (c)-(f). Note that the edges of MST (b) in this case are chosen among the original edges (a).

with Euclidean distance (GSNE), Shortest Path Navigation with Hopping distance (SPNH), and Shortest Path Navigation with Euclidean distance (SPNE), is selected. The shortcut is connected by a new edge, unless the candidate edge crosses one of the existing edges of $G(t)$ in the 2D space [15]. This shortcut construction process is repeated as long as the total length $l(t)$ of edges of $G(t)$ does not exceed a length constraint l_{max}. Figure 1 illustrates the shortcut construction model based on the set of vertices in the major thoroughfares of Boston road structure [1, 16]. The constraint l_{max}, in this case, is given by the total length of the original edges in Fig. 1(a), and the performance is summarized in Table I. From Fig. 1, we observe that the type of metric [hopping distance for (c) and (e) vs. Euclidean distance for (d) and (f)] greatly affects the final structure of the network.

TABLE I: Total edge lengths and routing performances for the Boston road example depicted in Fig. 1. The total edge lengths and the GSN/SP distances are shown in an arbitrary unit.

network	length	routing performance
original [Fig. 1(a)]	1.038×10^5	6.820 (GSN steps)
		4.619×10^3 (GSN distance)
		5.716 (SP steps)
		4.259×10^3 (SP distance)
MST [Fig. 1(b)]	3.818×10^4	25.78 (GSN steps)
		1.413×10^4 (GSN distance)
		17.02 (SP steps)
		8.173×10^3 (SP distance)
GSNH [Fig. 1(c)]	1.036×10^5	5.266 (GSN steps)
GSNE [Fig. 1(d)]	1.029×10^5	4.228×10^3 (GSN distance)
SPNH [Fig. 1(e)]	1.037×10^5	3.950 (SP steps)
SPNE [Fig. 1(f)]	1.029×10^5	4.103×10^3 (SP distance)

III. RESULTS

A. Topological properties of emerged network structures

For more systematic approach, we first generate N number of vertices on the square with the unit length. Then, as described in Sect. II, the MST is constructed and serves as a starting point of the shortcut construction. Examples of constructed networks from a vertex configuration in the space are illustrated in Fig. 2. All the results reported here are from graphs with $N = 10^2$ and averaged at least 20 independent graph generations. At the first glance, the difference between GSN and SP [(a) vs. (c) and (b) vs. (d)] may not be notable enough, but the fraction of Braess edges (defined as the edges whose removal enhances the greedy navigability) [1] causing the inefficiency, shown in Table II, is much smaller for GSN, from its construction purpose toward the optimized structure for the GSN pathways. The comparative time series of performance and the total edge length shown in Figs. 3–6(a) indicate that the GSNH tends to connect edges with shorter distance first.
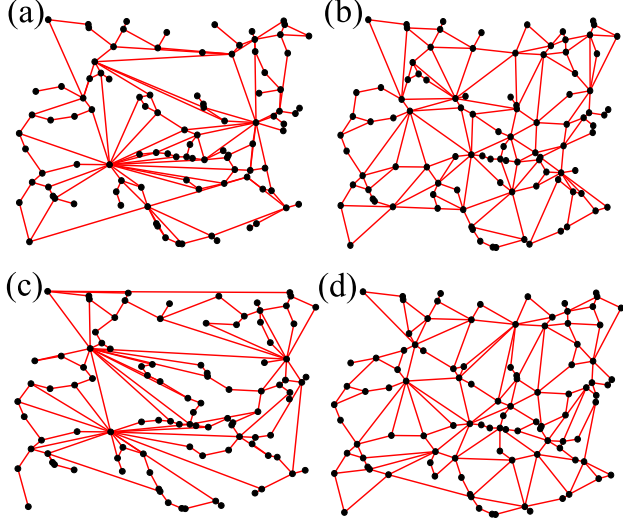
FIG. 2: Example structures of model networks for GSNH (a), GSNE (b), SPNH (c), and SPNE (d), starting from the same randomly distributed vertices with $N = 10^2$ and $l_{max} = 20$.

TABLE II: Fraction of Braess edges P_B with negative edge essentiality e values [1] for emerged networks with different greedy shortcut addition schemes, for the model network with $N = 10^2$ and $l_{max} = 20$. The e values defined with the hopping (Euclidean) distance are used for GSNH and SPNH (GSNE and SPNE), respectively.

	GSNH	GSNE	SPNH	SPNE
P_B	12.12%	1.249%	21.70%	5.215%

The most noticeable feature distinguishing between using the hopping distance and the Euclidean distance is the heterogeneity in the number of connections attached to each node, or degree distribution shown in Figs. 3–6(b). Hubs, or the nodes with emergence of hubs for hopping-distance-based scheme, as shown in much fatter tail from Figs. 3(b) and 4(b) than Figs. 5(b) and 6(b). Looking more closely, it is also shown that even some large hubs located relatively far from the centroid, showing the central governance (SPN) vs. decentralized local governance (GSN) hub locality [Figs. 3–6(c)]. The different importance of geometric distance for routing schemes is reflected in the average connection probability for vertex pairs with a certain Euclidean distance, shown in Figs. 3–6(d), where the connection probability is exponentially decreased for GSNE and SPNE, while the relatively long-range connections are observed for GSNH and SPNH. The average clustering coefficient $\langle C(k) \rangle$ [17] as a function
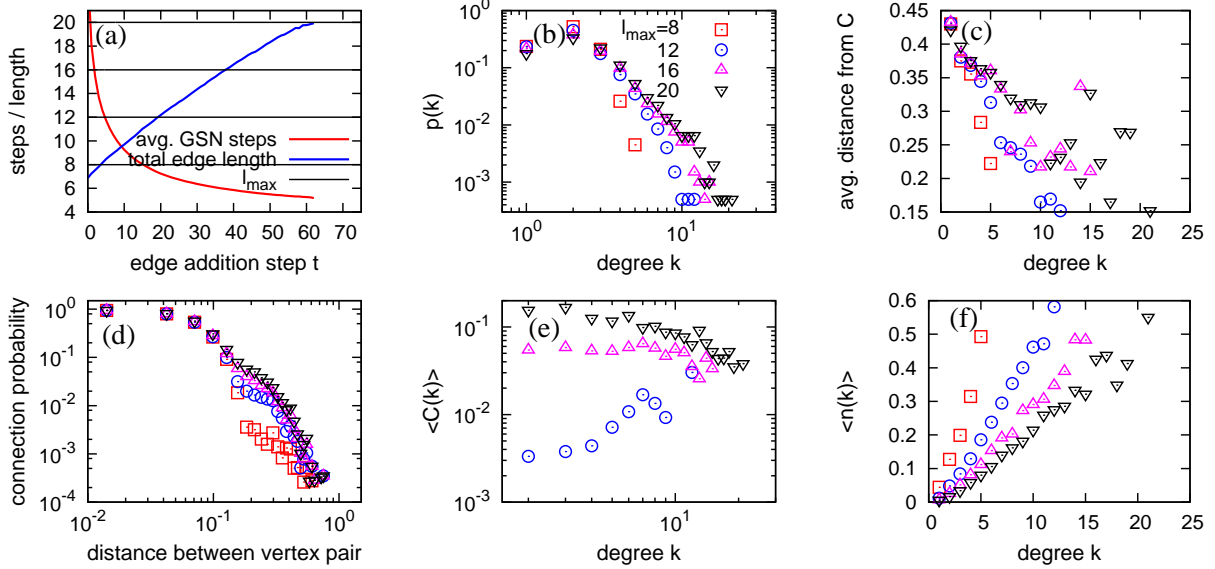
FIG. 3: The topological properties of emerged network structures for several different cutoff values of $l_{\max}$, in case of GSNH, where $N = 10^2$. Time series of performance and total edge length is shown (a), where the black horizontal lines correspond to $l_{\max}$ values used in (b)–(f). The others are degree distribution (b), position centrality of vertices as functions of degree (c), connection probability of vertex pairs as functions of Euclidean distance (d), average clustering coefficient for vertices with given degree (e), and average vertex navigator centrality for vertices with given degree (f), depending on the change of $l_{\max}$ values.

of degrees seem to be increased with the degree [Figs. 3–6(e)], in contrast to topological networks without geometric embedding. In addition, naturally, there is a strong correlation between the vertex navigator centrality $\langle n(k) \rangle$ [1] and degrees [Figs. 3–6(f)].

B. Geometric properties: triangular block, angle and area distributions

One can observe lots of characteristic triangular blocks for the optimized structures for all the cases, as shown in Figs. 1 and 2. Those triangular blocks are quantitatively counted in comparison to the random counterparts, using the clustering coefficient of a whole network [18]. Table III shows the results. Based on the ratio $C_\triangle / C_r$, obviously the Euclidean-distance based strategy produces more triangular blocks, but for both metrics, it is also notable that GSN induces more triangles than SPN. Therefore, along with the degree distributions, the triangular block statistics also support the conclusion that GSN encapsulates
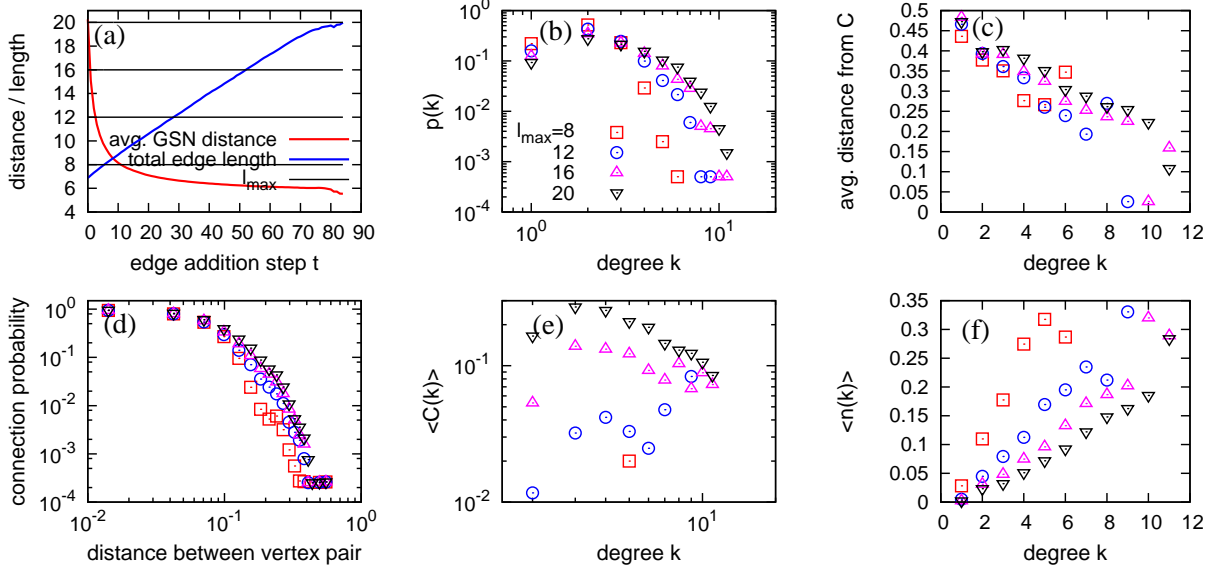
FIG. 4: The topological properties of emerged network structures for several different cutoff values of l_{max}, as in Fig. 3, in case of GSNE.
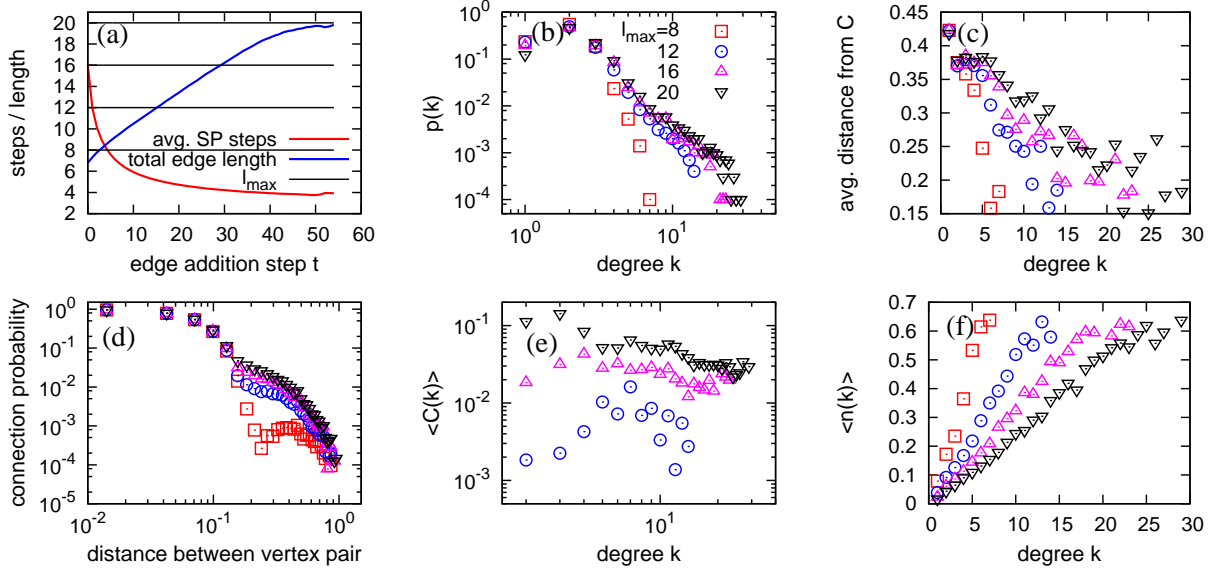

FIG. 5: The topological properties of emerged network structures for several different cutoff values of l_{max}, as in Fig. 3, in case of SPNH.

the refined local structures. In reality, of course, due to other factors such as packing the buildings with square cross-sections into blocks, the square blocks rather than triangular blocks prevail. However, it is important to note that such a simple objective function based on navigability can generate the local block structures. Those triangular structures are in-
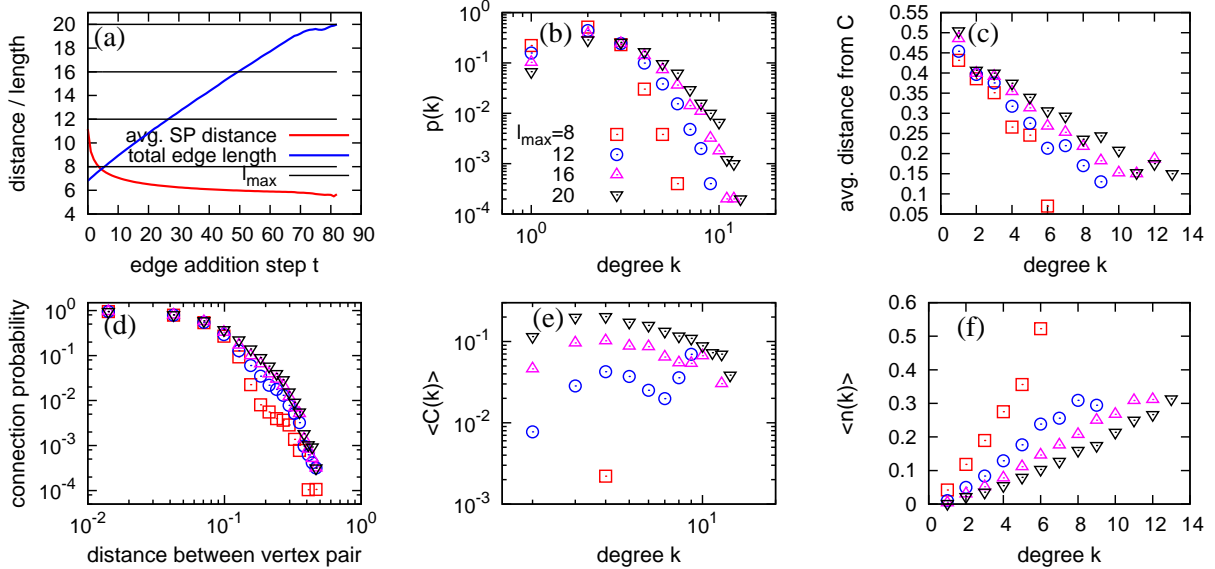
FIG. 6: The topological properties of emerged network structures for several different cutoff values of $l_{\max}$, as in Fig. 3, in case of SPNE.

TABLE III: The clustering coefficient based on the number of triangles ($C_{\triangle}$), compared to the random counterpart ($C_r = 2M/N^2$, where N and M are the numbers of vertices and edges, respectively).

method	$C_{\triangle}$	C_r	$C_{\triangle}/C_r$
GSNH	1.04×10^{-1}	3.20×10^{-2}	3.26
GSNE	1.89×10^{-1}	3.66×10^{-2}	5.15
SPNH	6.29×10^{-2}	2.98×10^{-2}	2.11
SPNE	1.56×10^{-1}	3.44×10^{-2}	4.53

deed observed in reality, such as Roman roads in Britain [19], the sheer scale of which is not imposed with such restricting factors at the time of the construction. Such a difference is shown to be even more prominent when we remove the no-crossing rule, which will be discussed in Sect. III C.

Another geometric aspect of the optimized networks is seen by observing the distributions of area enclosed by edges and the angles by adjacent edges for vertices, i.e., the angles between adjacent roads at the intersections. We observe that both the enclosed area distribution and angle distribution are notably distinguishable depending on the metrics used as
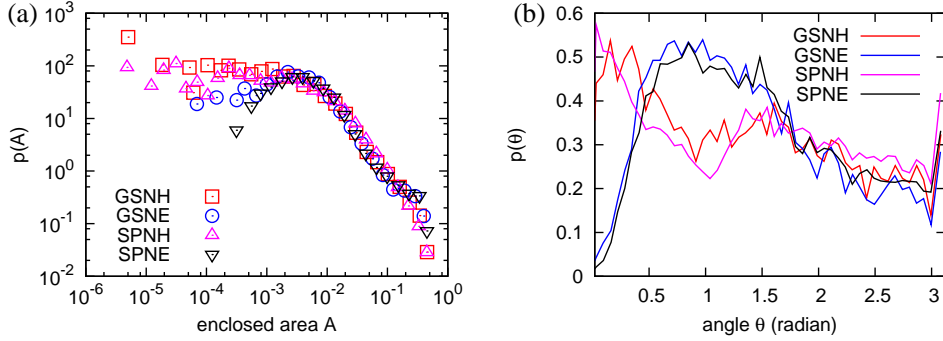
FIG. 7: The distributions of area enclosed by edges (a) and the angles for the intersections (vertices) (b), in the optimized structure for each navigability ($N = 10^2$ and $l_{\max} = 20$).

shown in Fig. 7. For hopping-distance based strategies, more heterogeneous enclosed area distributions compared to the Euclidean-distance based strategies are observed (Fig. 7(a)). In addition, very sharp angles (Fig. 7(b)) are abundant due to the existence of hubs (see Fig. 2 as well for examples), and the tendency is slightly more significant in SPNH than GSNH. On the other hand, the angles are distributed around $\simeq 60°$, the characteristic angle of the regular triangle for Euclidean-distance based GSNE and SPNE. Therefore, the enclosed area and angle distributions also indicate the emergence of relatively more regular triangular structures with uniform enclosed area for GSNE and SPNE.

C. Remarks on the no-crossing rule for edges

So far, we have not allowed the crossing between edges in the construction process since we consider such a crossing as effectively generating a new junction or vertex. What if there is no such rule? If we allow edge-crossing, star-graph-like structures are naturally emerged for SPNH as shown in Fig. 8(c), which leads to the "condensation" in terms of degree, as the bimodal distribution for $l_{\max} = 12, 16$, and 20 shown in Fig. 9(b) [20]. Interestingly, this severe condensation is not observed in case of GSNH, as shown in Figs. 8(a) and 9(a) (power-law-like fat tailed distribution but no condensation). The no-crossing rule, therefore, effectively prevents such a condensation for SPNH, while no such explicit rule is necessary for GSNH, because the consideration of greedy navigators itself naturally avoid such condensed situations and allows the *local* hubs.
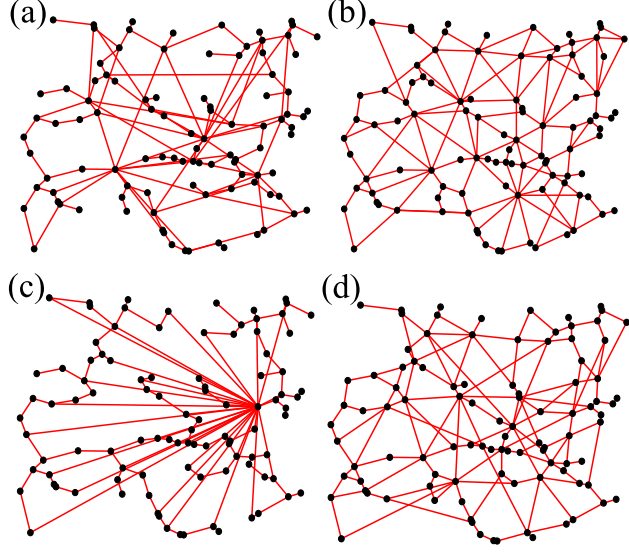
FIG. 8: Example structures of model networks for GSNH (a), GSNE (b), SPNH (c), and SPNE (d), if edge-crossing is allowed, starting from the same randomly distributed 10^2 vertices ($l_{max} = 20$) as in Fig. 2.
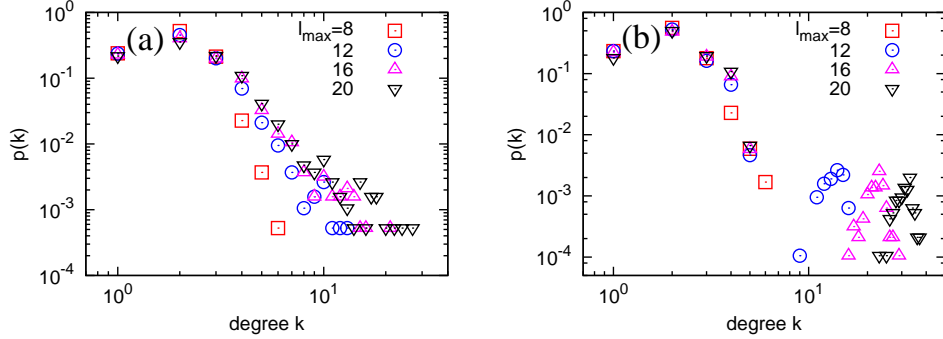

FIG. 9: The degree distributions of emerged network structures for GSNH (a) and SPNH (b), for several different cutoff values of l_{max}, where $N = 10^2$.

IV. SUMMARY AND DISCUSSIONS

With a simple greedy optimization under the limited sum of length, we have investigated the various structural properties of optimized networks for different types of navigability. Due to the complexity of finding the exact optimal configurations of transportation systems, in practice, heuristic methods are usually adopted for real systems. Therefore, we believe that even though our scheme is simple and idealistic, it has some degree of implications in real transportation systems. The most essential findings of ours are the emergence of hubs in

11

case of the consideration of the hopping distance, inherently geometric aspect of optimized structures reflecting the types of metrics used, and the fact that the user-based routing scheme (GSN) effectively induces the local hubs even without the explicit no-crossing rule. The emergence of triangular blocks from the simple optimization procedure is also notable— would city plans be optimized for pedestrian navigability rather than traffic planning and building design, then we should probably see more triangular blocks.

More elaborated or realistic approaches, for instance, using the different types of backbones, stochastic edge additions (similar to the simulated annealing process); different objective functions such as the combination of navigability and length [8] or among different metrics, etc. are good candidates for the future work. The reality check using human subjects, of course, would be necessary for the application to the real systems.

Acknowledgments

[1] S. H. Lee and P. Holme, Phys. Rev. Lett. **108**, 128701 (2012).

[2] A.K. Hartmann and M. Weigt, *Phase Transitions in Combinatorial Optimization Problems* (Wiley-VCH, 2001)

[3] T. Wolbers and M. Hegarty, Trends in Cognitive Science **14**, 138 (2010)

[4] R. Thomas and S. Donikian, *Spatial Cognition V, LNAI* 4387, pp. 421-438 (2007).

[5] J. M. Kleinberg, Nature **406**, 845 (2000).

[6] M. Boguñá, D. Krioukov, and K. C. Claffy, Nat. Phys. **5**, 74 (2008); M. Boguñá and D. Krioukov, Phys. Rev. Lett. **102**, 058701 (2009)

[7] M. Moussaïd, D. Helbing, and G. Theraulaz, Proc. Natl. Acad. Sci. USA **108**, 6884 (2011)

[8] M. T. Gastner and M. E. J. Newman, Phys. Rev. E **74**, 016117 (2006)

[9] G. Li, S. D. S. Reis, A. A. Moreira, S. Havlin, H. E. Stanley, and J. S. Andrade, Jr., Phys. Rev. Lett. **104**, 018701 (2010)

[10] M. Brede, Phys. Rev. E **81**, 025202(R) (2010)

[11] Y. Hu, Y. Wang, D. Li, S. Havlin, and Z. Di, Phys. Rev. Lett. **106**, 108701 (2011)

[12] W. Liu, A. Zeng, and Y. Zhou, e-print arXiv:1112.0241

[13] K.-I. Goh, G. Salvi, B. Kahng, and D. Kim, Phys. Rev. Lett. **96**, 018701 (2006)

[14] J. B. Kruskal, Proc. Am. Math. Soc. **7**, 48 (1956)

[15] This rule is applied since we consider the road-like structure, where each vertex corresponds to the junctions of roads. Therefore, a new crossing point generated by a new edge can be considered as a new vertex, which effectively changes the system size, so we intentionally avoid this situation.

[16] H. Youn, M. T. Gastner, and H. Jeong, Phys. Rev. Lett. **101**, 128701 (2008)

[17] D. J. Watts and S. H. Strogatz, Nature **393**, 409 (1998)

[18] Note that there are two slightly different kinds of clustering coefficients, i.e., the one averaged over the individual vertices' clustering coefficient which is used for $C(k)$ in this work, and the other taking the ratio of numbers of triangles to the numbers of vertex triads used for $C_\triangle$.

[19] I. D. Margary, *Roman Roads in Britain* (John Baker, 1973) The image is available online at `http://en.wikipedia.org/wiki/File:Roman.Britain.roads.jpg`.

[20] G. Bianconi and A.-L. Barabási, Phys. Rev. Lett. **86**, 5632 (2001)