Towards the Concept of Spatial Network Motifs

José Ferreira[1], Alberto Barbosa[1,2], and Pedro Ribeiro[1,2]

[1] Departamento de Ciência de Computadores, Faculdade de Ciências
Universidade do Porto, R. Campo Alegre s/n, 4169–007 Porto, Portugal
[2] CRACS & INESC-TEC, Portugal,
jcff@fc.up.pt, alberto.barbosa@fc.up.pt, pribeiro@dcc.fc.up.pt

Abstract. Many complex systems exist in the physical world and therefore can be modeled by networks in which their nodes and edges are embedded in space. However, classical network motifs only use purely topological information and disregard other features. In this paper we introduce a novel and general subgraph abstraction that incorporates spatial information, therefore enriching its characterization power. Moreover, we describe and implement a method to compute and count our spatial subgraphs in any given network. We also provide initial experimental results by using our methodology to produce spatial fingerprints of real road networks, showcasing its discrimination power and how it captures more than just simple topology.

Keywords: spatial networks, subgraphs, network motifs

1 Introduction

Complex networks are a very powerful abstraction of real-world systems that allow us to analyze the underlying interactions [11]. Many of these systems have a correspondence to the physical world, such as transportation networks (e.g. road, train or subway), power grids or brain networks. Their components are therefore embedded in space and topology alone does not capture all the relevant information [1]. Being able to understand and analyze these spatial networks is therefore a crucial task with multidisciplinary applicability [2,3].

Subgraphs can be seen as the building blocks of networks and they are the core of rich characterization concepts such as network motifs [9] or graphlet degree distributions [13]. Despite extensions to incorporate dimensions such as weight [4], time [12], color [15] or multiple layers [16], to the best of our knowledge there is no general and widespread subgraph abstraction that incorporates the spatial dimension. We should note that for specific domains there has been some related work, such as in football, where passing networks between different regions of the playing field have been created [10], but these remain specialized and restricted to their own field of study.

In this paper we try precisely to aim towards a general concept of spatial motifs able to characterize networks from any domain. Our first contribution (Section 2) is a novel subgraph abstraction that incorporates spatial information in a way that is general enough to incorporate several spatial dimensions (e.g. 2D or 3D) and granularities (e.g. large macroscale vs small microscale regions). The key idea is to automatically create a spatial partition of the subgraph bounding and to color the nodes according to the region they are on. Our second contribution (Section 3) is an initial methodology and fully functional framework to detect and count these spatial motifs, based on enumerating subgraph occurrences and then computing their spatial and topological type. Our third and last contribution (Section 4) is a proof of concept experimental section, in which we analyze several real word road networks, showing that unlike purely topological motifs, we can distinguish between grid and non grid-like layouts.

2 A Novel Concept of Spatial Motifs

There are several possible ways for expressing spatial properties. For instance, the distance between nodes can be used as edge weight [7], but this would not take into account the relative position of the nodes. Another option would be to use angles between nodes, hence losing the distance information. Our approach relies on first creating a *bounding box* around the found subgraph using the nodes spatial location, and then partitioning this box into regular-sized regions, thus taking into account both the relative position and the distance between nodes.

For the sake of simplicity and given the space constraints, we will mainly focus on a 2D example divided into 2×2 quadrants, but as explained later, our approach is general and extends naturally to higher dimensions. The creation of the bounding box is straight-forward: we require a set of coordinates for each node on the input, and for each found subgraph, we calculate the maximum and the minimum of both the x and the y values, which gives us the limits of our box. Then, we simply calculate the relative position of each node when referring to the center of the bounding box, assigning a quadrant to the node on the form of a color. An example can be seen in Figure 1, where the original spatial network is given above, in the blue nodes, and all its three node spatial subgraph occurrences are given below.

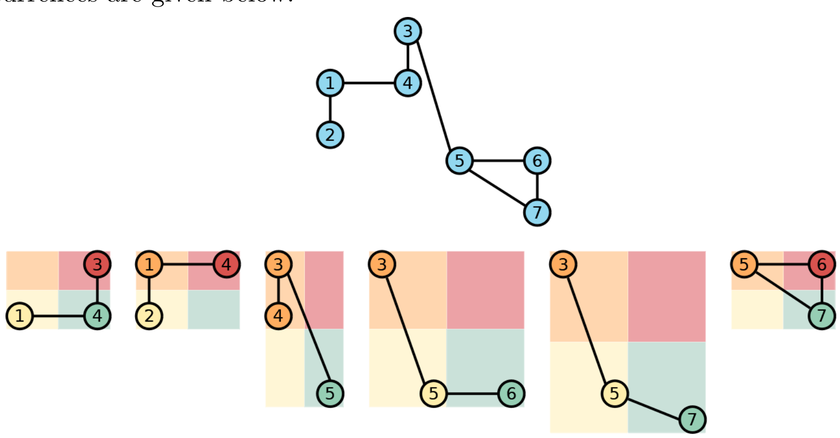


Fig. 1: Example of subgraphs with spatial coloring by 2D quadrants

Taking into account the spatial dimension of the nodes on the previous figure, we can enumerate five different subgraph types, being the fourth ($\{3, 5, 6\}$) and the fifth occurrence ($\{3, 5, 7\}$) of the same type: they both have three nodes in the same quadrants (one orange, one yellow and one green) and the same connections (one orange-yellow edge and another yellow-green edge).

By contrast, if only purely topological properties were used, there would exist only two subgraph types, as depicted in Figure 2, with the first five occurrences being a chain of three nodes and the last one (($\{5, 6, 7\}$) being a triangle.

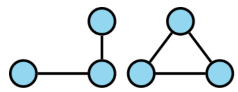


Fig. 2: Chain (type A) and Triangle (type B) topological subgraphs

The above example already illustrates how much richer our spatial representation is, but we would like to emphasize how general our conceptual approach is. From a scale point of view, it naturally extends to higher numbers of nodes (just consider more nodes in each subgraph). From a topological point of view, it is also able to organically integrate features such as direction (just consider that when distinguishing between different isomorphic types). From a granularity point of view, we can also consider any regular division. Here we exemplified with 2×2 quadrants, but we could use any $n \times n$ partition, depending on what we want to measure (and moreover we could even use on the same analysis subgraph occurrences at different n sizes to create a richer set of features). Finally, our approach also naturally extends to higher dimensions (for instance, in a 3D space one could use $2 \times 2 \times 2$ octants as the equivalent of 2D quadrants).

3 Finding and Counting Spatial Motifs

In this section we explain our methodology for finding and counting the occurrences of spatial motifs as defined in the previous section. The motivation for counting will become clearer on Section 4, but essentially by computing subgraph frequencies we are able to obtain numerical features characterizing the underlying network. Counting subgraphs is therefore a core network analysis primitive. A fully detailed survey on how to count purely topological motifs can be seen in [14], including approximate and parallel approaches.

Our proposed initial approach has two steps: (i) we first enumerate all subgraph occurrences of a given size k, obtaining sets of k connected nodes; (ii) for each occurrence we identify its spatial type by producing a canonical labeling that is unique to each colored isomorphic type. A fully functional implementation is available at github [3].

[3] we will make available a github link to the source code if the paper is accepted

3.1 Enumerating Subgraph Occurrences

In order to enumerate the occurrences of subgraphs with k nodes, we opted to use ESU [17], a general purpose subgraph enumeration algorithm capable of finding each occurrence only once, avoiding symmetries. In short, this is done by starting from a single node and expanding from there, using only vertices that have an index (label at the original graph) greater than that of the original node and that can be neighbors of a newly added vertex but not of any other one previously added. In Figure 3 we illustrate this process with a small example for $k = 3$ and an original network of six nodes. Inside each tree node box we indicate two node sets: first the current subgraph being enumerated ($V_{subgraph}$) and secondly the set of nodes which can expand it ($V_{extension}$).

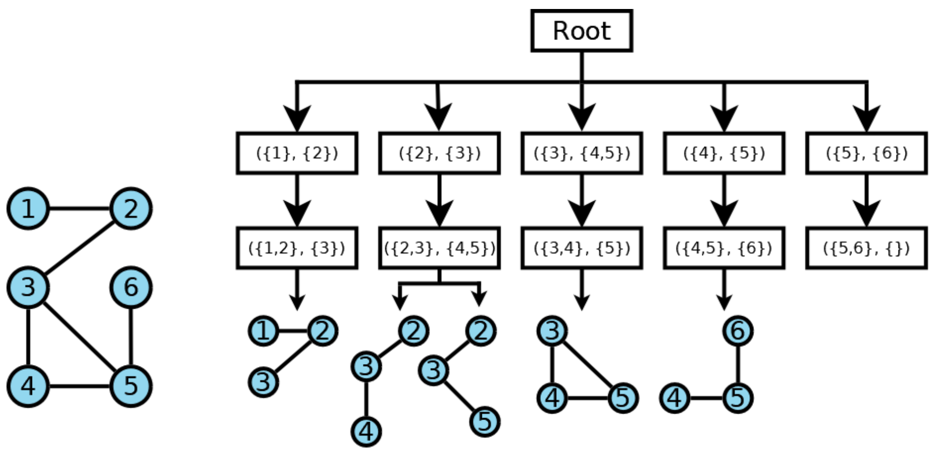


Fig. 3: Example of an ESU search tree for k-subgraph enumeration with $k = 3$

The root of the search tree is a starting point to evaluate the subgraph. It's children, on the second level, correspond essentially to one branch per node, with the extension sets being their immediate neighbors with a larger index than the node itself. For instance, the second branch contains $V_{subgraph} = \{2\}$ and $V_{extension} = \{3\}$ (3 is a neighbor of 2 and 1 is not considered since $1 < 2$ and the subgraph with $\{1, 2\}$ would be already considered in the first branch). This process continues in the following tree levels: we add 3 to $V_{subgraph}$ since it has two neighbors that meet the requirements, those are added to $V_{extension}$, resulting in $V_{subgraph} = \{2, 3\}$ and $V_{extension} = \{4, 5\}$. Now we have two possible branches, $V_{subgraph} = 2, 3, 4$ and $V_{subgraph} = 2, 3, 5$. In both these cases we have $|V_{subgraph}| = 3$ and we have reach the desired node set size.

After doing this to every single node we end up with the subgraphs of size 3 represented on the leafs of the tree. The required conditions for a node to be added to $V_{extension}$ make sure that no subgraph is found twice.

3.2 Subgraph Types and Canonical Labeling

After having the node sets that correspond to each subgraph occurrence, we still need to discover the spatial type of each one, so that we can increment its frequency. For instance, as we could observe in Figure 1 that the fourth and fifth subgraphs belong to the same type.

In our approach, we first determine the bounding box of each occurrence by computing the minimum and maximum values of each spatial dimension. We then partition the box into the desired number of regions and we "color" the nodes according to the region in which each one falls, effectively obtaining what could be considered a colored subgraph [15]. Afterwards, we compute a canonical labeling such that two subgraph occurrences will have the same labeling if and only if they correspond to the same (colored) isomorphic type.

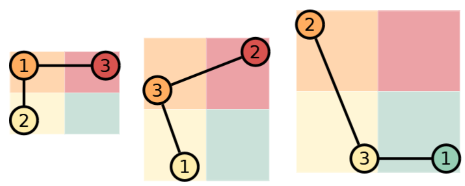


Fig. 4: The first two subgraphs are of the same type and should have the same canonical labeling; the third one should have a different labeling

Figure 4 illustrates the need for a canonical labeling that takes node colors into account. From a purely topological point of view, all three subgraphs are chains and therefore indistinguishable. However, when incorporating spatial information, this is not the case. We want the first and second subgraphs to have the same label, as they have the same colored topological properties: one node in each quadrant except the fourth one (represented by the colors orange, red and yellow), and two connections (an orange-red edge and orange-yellow edge). The third subgraph has different spatial properties that correspond to different colored nodes and edges.

In general, even without colors, computing canonical labelings is a very hard computational task, closely related to the graph isomorphism problem [5]. We therefore resorted to `nauty` [8], a third-party and very efficient set of procedures to determine the automorphism group of a vertex-colored graph. Since `nauty` has built-in support for colored nodes, a call to the default method with the required arguments and the quadrant as color is enough to give us the canonical label. To achieve a labeling using colors, `nauty` requires the colors to be given in some order, and the edge labels will be returned in the order the colors were provided, that is, first the edges with the first color, then the ones with the second color, and so on.

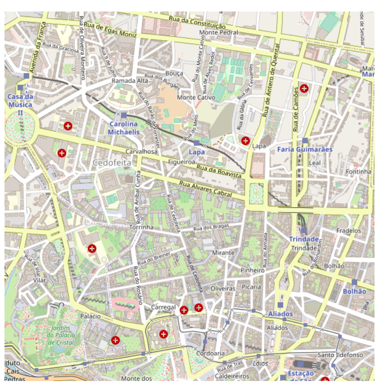
4 Experimental results

In order to test our implementation and to showcase the applicability of our proposed subgraph abstraction, we now provide a proof of concept. We use real world road networks, which can be considered one of the quintessential spatial network examples to which everyone can relate to. We selected two cities with "grid-like" street layouts (Espinho, Portugal and Detroit, USA) and two with "non grid-like" layouts (Porto, Portugal and Oxford, UK), aiming to distinguish between these two layout groups using our proposed approach.

Our original source of raw street data was OpenStreetMap (OSM) [6], which is a collaborative project that aims to provide a free editable geographic database of the entire world. In Figure 5, we show an image of the approximate area used for each of the cities mentioned above, taken from OSM, which clearly indicates the nature of the road layouts that are being analyzed.

(a) Espinho

(b) (Northwest) Detroit



(c) Porto

(d) Oxford

Fig. 5: Layout of the four cities used as input

From the raw data we create a network in which the nodes are true road intersections and edges represent roads between them (this implied the creation of an automated script that given a geographical bounding box will extract all OSM features from it, which are further processed and simplified to create the desired intersection network). In this network, we fix the subgraph size to $k = 3$ and count the number of occurrences of each spatial subgraph. Figure 6 exemplifies this process for a small portion of a map, illustrating the extracted network and all its occurring subgraph occurrences, some of them belonging to the same spatial type. For instance, $\{1, 2, 3\}$ and $\{3, 4, 5\}$ are of the same type, and the same can be said for $\{1, 2, 7\}$ and $\{3, 4, 6\}$.

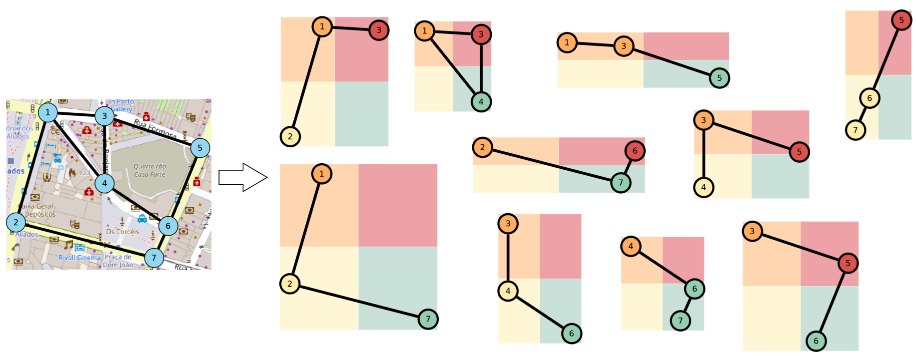


Fig. 6: The network corresponding to a map and its subgraph enumeration

To better understand and visualize the differences in subgraph occurrences, we opted to further divide spatial types into *classes* (families of subgraphs), that corresponds to the four 90 degrees rotations of the same simple type. Figure 7 illustrates this concept and one class of subgraphs.

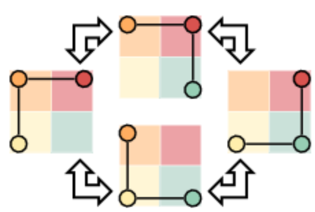
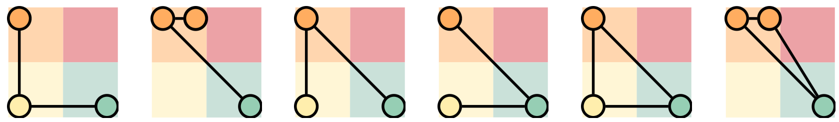


Fig. 7: Subgraph class 1 and its four spatial subgraph types, corresponding to 90 degree rotations

Figure 8 illustrates representatives of the six most frequent classes of subgraphs that we found in the studied road networks (the frequency of the other possible classes is residual and their very low relative frequency does not impact the conclusions of our analysis).

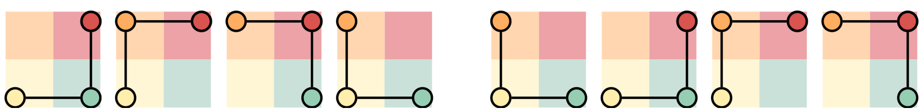


(a) Class 1 (b) Class 2 (c) Class 3 (d) Class 4 (e) Class 5 (f) Class 6

Fig. 8: Representatives of the most frequent classes of subgraphs considered

4.1 Results for "grid-like" street layouts

Figure 9 represents the top-4 (in order, from left to right) of the subgraphs with most occurrences in Espinho and Detroit. The results are as expected and capture the grid-like nature of the layout. Even if the exact order of simple types is not exactly the same, this top-4 corresponds to class 1 subgraphs, whose representation resembles a right angle, that is, where each node is in a different quadrant and the subgraph is a chain that connects nodes in consecutive quadrants.



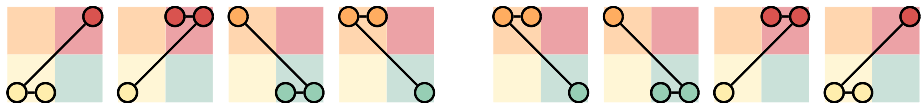
(a) Espinho (b) Detroit

Fig. 9: Top-4 of subgraphs with most occurrences in the two grid-like cities. Note that all the subgraphs are of class 1 (as defined in Figure 8)

The difference in frequency from the 5th to the 4th most common subgraph is noticeable, particularly in the case of Espinho. In both cases, subgraph types from the 5th to the 8th positions are of class 2. Tables 1 and 2 give more detail on the results, showing the relative frequency (percentage of total occurrences) of the 10 most common subgraph types and their associated class. In total, there were 22 different subgraph types found for Espinho and 32 for Detroit.

4.2 Results for "non grid-like" street layouts

On the other hand, if the city does not have a well defined grid layout, we can observe that the most frequent subgraphs are very different, as can be seen in Figure 10. In fact, the top-4 for these two cities does not have any subgraph type in common with the grid-like cities. Here, the topological chain type of subgraph is still the most common, which means that if only the topological information of the subgraphs was used, conclusions with this level of detail would not be possible, but in this case instead of having each node in a different quadrant, two nodes share a quadrant, there is a connection between them, and one of them connects to a node in the opposite quadrant, with the entire top-4 of most frequent subgraphs being of the same class.



(a) Porto (b) Oxford

Fig. 10: Top-4 of subgraphs with most occurrences in the two non grid-like cities. Note that all the subgraphs are of class 2 (as defined in Figure 8)

As on the previous section, we give detailed results of the top-10 most common subgraph types in Tables 3 and 4. Again there is a noticeable increase in frequency from the 5th to the 4th most common subgraph, and the same class appears from rank 5 to rank 8. In total, there were 28 different subgraphs found for both Porto and Oxford.

Rank	Subgraph Type	Relative Frequency
1	1	0.178258
2	1	0.167738
3	1	0.161894
4	1	0.160140
5	2	0.089421
6	2	0.080070
7	2	0.046756
8	2	0.044418
9	4	0.010520
10	3	0.009351
Top-10 total	—	0.948567

Table 1: Spatial subgraph frequencies in Espinho

Rank	Subgraph Type	Relative Frequency
1	1	0.139369
2	1	0.138028
3	1	0.135329
4	1	0.134110
5	2	0.108859
6	2	0.097870
7	2	0.086585
8	2	0.083102
9	3	0.019452
10	3	0.019208
Top-10 total	—	0.961913

Table 2: Spatial subgraph frequencies in Detroit

Rank	Subgraph Type	Relative Frequency
1	2	0.133904
2	2	0.131815
3	2	0.118655
4	2	0.118446
5	1	0.081888
6	1	0.080426
7	1	0.079590
8	1	0.075621
9	3	0.023605
10	3	0.022143
Top-10 total	—	0.866095

Table 3: Spatial subgraph frequencies in Porto

Rank	Subgraph Type	Relative Frequency
1	2	0.149768
2	2	0.148675
3	2	0.133916
4	2	0.121618
5	1	0.074064
6	1	0.071331
7	1	0.070511
8	1	0.066684
9	4	0.024050
10	4	0.022683
Top-10 total	—	0.883302

Table 4: Spatial Subgraph frequencies in Oxford

4.3 Comparison between cities

In Figure 11 we can observe a bar chart of the relative frequency of each subgraph class per city, with the usage of their spatial properties.

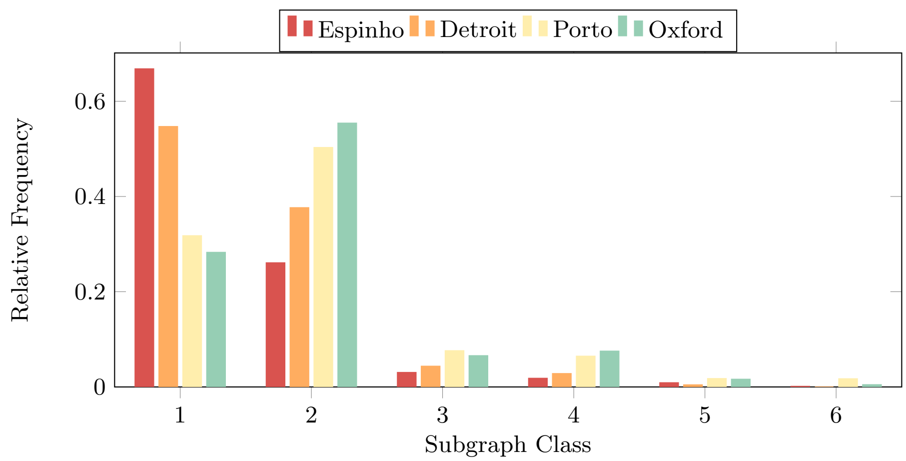


Fig. 11: Spatial subgraph fingerprint of each of the studied cities

If we remove the spatial component from the subgraphs, we are left with only two types of subgraphs: chains and triangles. This means that spatial classes 1 to 4 will be of the single topological type A (chain), and classes 5 and 6 will of the the topological type B (triangle). Using the same data as the previous plot but ignoring the spatial properties results in the bar chart depicted in Figure 12.

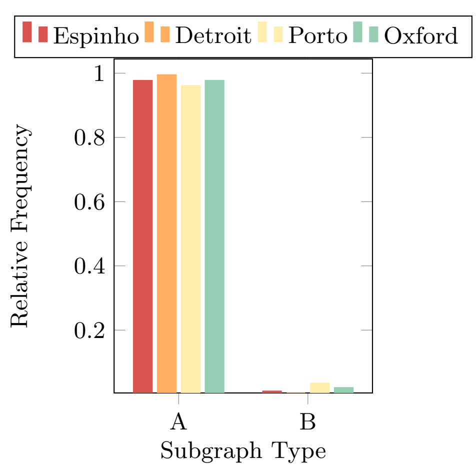


Fig. 12: Purely topological subgraph fingerprint of each of the studied studies

We can observe from Figure 11 that the distribution of the subgraph classes clearly shows a predominance of class 1 in the two grid-like cities, whilst class 2 is more common in the two cities without this layout, which allows us to easily distinguish them, using only this distribution. Conversely, using the data from Figure 12 it is not possible to make that distinction, as all cities show a clear dominance of type A subgraphs with around the same difference in frequency when compared to type B, which is really uncommon in street networks. It is also interesting to note that the cities without a grid layout have a bigger frequency of other types of subgraphs other than classes 1 and 2, even if though those types are still by far the most relevant.

As a final note, we would like to remark that using a normal laptop the subgraph counting and labeling phase takes less than a minute to compute even in the largest considered network (Detroit, with 16 029 nodes and 24 773 edges). A potential drawback of our proposed strategy is that increasing the size k of the subgraph will inevitably lead to an exponential growth of the number of subgraph occurrences and hence on the execution time. However, in this paper we were mainly concerned with proving that the concept could be useful and there are still many improvements that can be made regarding efficiency.

5 Conclusions and Future Work

In this paper we present a set of contributions aiming to incorporate spatial properties into subgraph analysis. We first offer a novel abstraction that relies on a bounding box and regular spatial partitions to attribute node colors that describe the relative position of nodes within the subgraph. We then describe an implementation of a framework capable of discovering and counting these spatial subgraphs, based on enumerating occurrences and then discovering their type using a specialized canonical labeling mechanism. Finally, we provide a proof of concept experiment using real life data in which we show that our approach is able to go beyond classical topological motifs, capturing enough information to distinguish between different road network layouts.

We believe these are promising results that could lead into new insight on the characterization and comparison of network with spatial information. Our end goal is to be able to provide a universal spatial concept of network motifs that can be generally applicable to networks of any domain.

In order to further extend our work, we intend to study the incorporation of higher dimensional data, such as 3D brain networks, and we want to make an extensive evaluation of the role of the granularity in the information gained, by carefully analysing what happens when we use different amounts and sizes of spatial partitions. Furthermore, we want to study how changing the point of reference would impact the results (e.g. what happens to the patterns when we make arbitrary rotations?) and we intend to explore different symmetries and subgraph families that could provide classes that are invariant to spatial transformations (e.g. mirror symmetry). We also want to understand how we can assess statistical significance of the subgraph frequencies, by studying what could be appropriate spatial null models.

Finally, we want to improve the efficiency, not only by improving the exact counting computation, but also by trading accuracy for speed (e.g. using sampling) or using parallelism (e.g. using several threads in multicore machines).

Acknowledgements: This work is partially financed financed by National Funds through the Portuguese funding agency, FCT - Fundação para a Ciência e a Tecnologia within project LA/P/0063/2020 and grant SFRH/BD/136525/2018. We would also like to thank the reviewers for the insightful comments.

References

1. Barthélemy, M.: Spatial networks. Physics reports 499(1-3), 1–101 (2011)
2. Barthelemy, M.: Morphogenesis of spatial networks. Springer (2018)
3. Boguñá, M., Krioukov, D., Almagro, P., Serrano, M.Á.: Small worlds and clustering in spatial networks. Physical Review Research 2(2), 023040 (2020)
4. Choobdar, S., Ribeiro, P., Silva, F.: Motif mining in weighted networks. In: 2012 IEEE 12th International Conference on Data Mining Workshops. pp. 210–217. IEEE (2012)
5. Grohe, M., Schweitzer, P.: The graph isomorphism problem. Communications of the ACM 63(11), 128–134 (2020)
6. Haklay, M., Weber, P.: Openstreetmap: User-generated street maps. IEEE Pervasive computing 7(4), 12–18 (2008)
7. Hasan, S., Schneider, C.M., Ukkusuri, S.V., González, M.C.: Spatiotemporal patterns of urban human mobility. Journal Statistical Physics 151(1), 304–318 (2013)
8. McKay, B.D., Piperno, A.: Practical graph isomorphism, ii. Journal of symbolic computation 60, 94–112 (2014)
9. Milo, R., Shen-Orr, S., Itzkovitz, S., Kashtan, N., Chklovskii, D., Alon, U.: Network motifs: simple building blocks of complex networks. Science 298(5594), 824–827 (2002)
10. Narizuka, T., Yamamoto, K., Yamazaki, Y.: Statistical properties of position-dependent ball-passing networks in football games. Physica A: Statistical Mechanics and its Applications 412, 157–168 (2014)
11. Newman, M.E.: The structure and function of complex networks. SIAM review 45(2), 167–256 (2003)
12. Paranjape, A., Benson, A.R., Leskovec, J.: Motifs in temporal networks. In: Proceedings of the tenth ACM international conference on web search and data mining. pp. 601–610 (2017)
13. Pržulj, N.: Biological network comparison using graphlet degree distribution. Bioinformatics 23(2), e177–e183 (2007)
14. Ribeiro, P., Paredes, P., Silva, M.E., Aparicio, D., Silva, F.: A survey on subgraph counting: concepts, algorithms, and applications to network motifs and graphlets. ACM Computing Surveys (CSUR) 54(2), 1–36 (2021)
15. Ribeiro, P., Silva, F.: Discovering colored network motifs. In: Complex networks V, pp. 107–118. Springer (2014)
16. Sallmen, S., Nurmi, T., Kivelä, M.: Graphlets in multilayer networks. Journal of Complex Networks 10(2), cnac005 (2022)
17. Wernicke, S.: Efficient detection of network motifs. IEEE/ACM transactions on computational biology and bioinformatics 3(4), 347–359 (2006)