Eur. Phys. J. B (2021) 94:192
https://doi.org/10.1140/epjb/s10051-021-00202-y

THE EUROPEAN PHYSICAL JOURNAL B

Regular Article - Statistical and Nonlinear Physics

Spatial distribution of graffiti types: a complex network approach

Eric K. Tokuda[1,a], Henrique F. de Arruda[1], Cesar H. Comin[2], Roberto M. Cesar Jr.[1], Claudio T. Silva[3], and Luciano da F. Costa[1]

[1] University of São Paulo, São Paulo, Brazil
[2] Federal University of São Carlos, São Carlos, Brazil
[3] New York University, New York, USA

Received 23 July 2021 / Accepted 9 September 2021 / Published online 27 September 2021

Abstract. Urban art constitutes an important issue in urbanism. Previous studies on the spatial distribution of graffiti rarely consider visual categories and how the city topology can impact graffiti production. In this work, after assigning graffiti occurrences to three categories, we analyzed their spatial distribution while searching for possible biases. Concepts from complex networks were adopted. First, communities defined by the connectivity profiles of the city network were obtained and the prevalence of each type of graffiti over these regions was analyzed. Next, to study the relationship between the density of graffiti occurrences with the visibility of specific city localities, a measurement (*accessibility*) based on the dynamics of the network was related to the distribution of the occurrences of graffiti of each type. Our case study considered the city of São Paulo, Brazil. The results showed good agreement of the detected communities with the main city areas and no biases between the relative density of each type of graffiti. Relatively high correlations were obtained between the density of graffiti and the accessibility when calculated inside each of the identified communities, suggesting that graffiti tends to appear more frequently in certain regions of the city with high accessibility.

1 Introduction

Modern cities represent the perfect stage for different forms of urban expression. Paintings on public walls are commonly known as *graffiti*, being part of the landscape of most large cities. These *urban canvases* may be composed not just by drawings, but also by combinations of symbols and writings. There is a wide range of ways in which these graffiti can be made, regarding the material, the size, the colour, and the respective content. This visual signature may be associated with the socio-economical context of the location [5]. Despite their great diversity, for simplicity's sake here we call all these forms *graffiti*.

It can be observed that graffiti is primarily created in urban regions. From a mathematical point of view, cities can be seen as an irregular, complex graph that evolves in time [2]. A diverse range of properties can be obtained from the network structure, including its average degree, clustering coefficient, and centrality measures [11]. Interestingly, these topological features can be related to several dynamical processes taking place in the network, such as mobility [14]. Previous works have analyzed the distribution of graffiti and its correlation with socio-economical indices [1,8,10,16]. These works considered just particular cases of graffiti, and more systematic studies considering the spatial distribution of different categories of graffiti are still necessary. Another subject deserving further attention consists in trying to better understand graffiti in the light of the city network topology and dynamics. In particular, having assigned categories to graffiti, it would become possible to study their spatial distribution and relationships with topological features of cities such as communities. Among other results, observed heterogeneities could provide indication about the context and motivation of graffiti. In this work, we aim at approaching these problems in a more systematic way.

Street-level images are initially collected from the location of interest. Categories are then defined based on a visual exploration of the data. The collected images are classified in terms of the observed categories of graffiti. The distribution of the location of each type is calculated and its deviation to the overall distribution is estimated. To investigate these differences, we extract the street network and analyze the respective structural and dynamics properties. From the topology, we obtain the communities and analyze the difference in the participation of each type of graffiti across the communities. We additionally perform random walks in the network to estimate the *accessibility* of the nodes. The results suggest differences among the spatial distribu-

[a] e-mail: tokudaek@gmail.com (corresponding author)

tions of each type of graffiti. The results of the dynamics show that there are moderate positive correlations among the density of each type and the accessibility. Another compelling result concerns the association between accessibility and the density of occurrences of graffiti. Despite showing moderate values of Pearson correlation coefficient when the entire city was considered, much higher values of correlation were observed when the correlation was measured within some communities. In particular, considerable values have been obtained in the northern region of the city of São Paulo.

The paper is organized as follows: the method and the underlying theory are initially presented, then the experiments are described followed by the respective conclusions.

2 Materials and methods

Instead of manually inspecting the regions in the city, the proposed approach considers street-level images collected from Google Maps [7]. These images are obtained from sensor-equipped cars as they move along the streets, with similar weather and lighting of the scene within a given region. The images are accompanied by metadata, which includes the geo-location of the scene. In the presented method, a geographical region of interest is initially defined (e.g. corresponding to a specific city). Given that a very large number of images is usually found in [7], it often becomes a challenge to process all of them. In this work, we randomly sample, with uniform distribution, a percentage of these images. As the images available in [7] do not contain metadata regarding the presence of graffiti, it is necessary to perform some pre-processing to separate the images that are to be considered.

When the overall distribution of graffiti is not uniform, presenting variations of density at distinct positions, two main hypotheses may be considered to explain such heterogeneity distributions: (i) though originally distributed in a uniform manner, the graffiti was sampled in some heterogeneous manner; and (ii) the graffiti is, in fact, not uniformly distributed. A third possibility would correspond to a combination of these two hypotheses, namely a heterogeneous graffiti distribution being sampled in a non-uniform manner. For simplicity's sake, hypothesis (i) was adopted in the present work. Such an approach is associated with a choice of a minimum density. Therefore, it becomes necessary to consider some indication of relevance that allows the less densely sampled portions of the data not to be taken into account. This criterion is described in more detail in *Experiments*.

Differently from previous related works, the graffiti present in the image is categorized according to the visual form: simple scribbles (type A), painted writings (type B), and canvases or textures (type C). This division is made after a preliminary visual exploration of the graffiti occurrences, taking into account the expected time to make them, while taking care to achieve well-separated categories. Please refer to Fig. 1 for an illustration of each of these three types. The collected sample is manually labeled whether containing or not each of the three types described, resulting in $\sum_{m=0}^{3} \binom{3}{m} = 8$ possible combinations of labels for an image.

2.1 Definition of types and annotation

To compare the spatial distribution of each of the above types of graffiti, a non-parametric density estimation with smoothing kernel can be calculated on the location of the occurrences. A kernel density estimation [4] with Gaussian kernel is therefore computed for each of the graffiti types. To compare how the distributions *differ* one another, the Kullback–Leibler divergence (KL) [9] is calculated. The KL divergence is a non-symmetric measure that originated from information theory, to estimate the information lost between the expected (Q) and the observed signal (P). It is calculated as $D_{\mathrm{KL}}(p(x) \parallel q(x)) = \int_{-\infty}^{\infty} p(x) \log\left(\frac{p(x)}{q(x)}\right) \mathrm{d}x$. According to the equation, lower values corresponds to *similar* distributions.

2.2 Network structure

Real networks [3] are characterized by presenting heterogeneous connectivity patterns. One classical problem when handling graphs is *community detection*, which aims at identifying modules or subgraphs having vertices that are more intensely interconnected internally to each module than across modules. Many works have tackled this problem, including the more traditional partition-based, hierarchical-based, and spectral-based methods [6]. Instead of handling expensive computations on the adjacency matrices, Rosvall et al. [15] proposed a community detection methodology based on random walks and information diffusion. The method minimizes the information required to describe the process of information diffusion across the graph. For instance, in a two-level description, communities in the graph may be uniquely identified, while the name of the vertices may be recycled across these structures. This idea is analogous to repeating street names across different cities. Huffman coding is used for the vertex names and the method aims at finding the partitioning of the graph which minimizes the description length of an infinite walk. In the case of no well-defined clusters, transitions between the potential clusters will be frequent and there will be little benefit in using a two-level description. The optimization of the possible partitioning schemes is computed using greedy search and simulated annealing. The described approach is applied to the identification of clusters in the city network.

2.3 Network dynamics

The structure of a network plays an important role in the dynamics of the processes that can take place on

(a) Type A **(b)** Type B **(c)** Type C

Fig. 1 Categories of graffiti considered in this work (from a personal collection of pictures)

it (e.g. [3]). *Accessibility* [17] is an interesting concept in this regard, as it provides information about both the structure and dynamics of complex networks. This measurement quantifies the influence of a given node on a set of vertices in its neighborhood, with respect to a specific dynamics (e.g. neuronal dynamics, epidemics, random walks). The random walk represents one of the central paradigms in dynamical systems, partly due to its simple formulation. The dynamics of random walks have been extensively studied (e.g. [12]) in the literature.

In this work, besides the traditional random walk, we consider a variation in which the walker is not allowed to repeat any vertex or edge because it is not expected that the graffiti makers keep going back and forth in the same streets. The agent starts from a source node and is allowed to move in a walk with no repeating vertices or edges, which results in a path with length h (please refer Fig. 2). The transition probabilities ($p^h_{i,j}$) between vertices can be obtained by considering the number of times the vertex j was reached from vertex i, $c^h_{i,j}$ scaled by the number of realizations (number of paths). The outward accessibility of each vertex can then be calculated with Eq. (1).

$$A^{out}_h(i) = exp\left[-\sum_{j=1}^{n} p^h_{i,j} log(p^h_{i,j})\right]. \tag{1}$$

This measure reaches maximum value when it can reach the vertices in level h with the same probability. Consider the network depicted in Fig. 2. In this case, the accessibility of the vertex i for $h = 2$ has maximum value, $A^{out}_{h=2}(i) = exp\left[-(\frac{1}{2}log(\frac{1}{2}) + \frac{1}{2}log(\frac{1}{2})\right] = 2$, because it can reach the vertices in level 2 with the same probability. On the other hand, this same measure achieves a lower value for vertex k due to the different probabilities of reaching the vertices 2 levels apart from k.

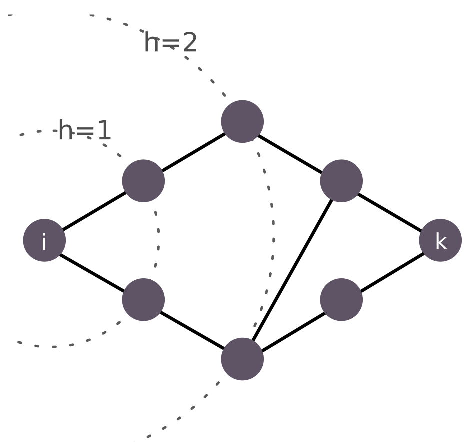


Fig. 2 Reachable vertices in the accessibility measurement. Concentric vertices represent nodes reachable in the same number of steps, $h = 1, 2$, departing from node i. This measure achieves a higher value for vertex i than for vertex k, assuming $h = 2$

3 Experiments

We considered the city of São Paulo, Brazil, as a case study. We initially obtained from [13] the street network, which is directed and composed of 112,014 nodes and 287,592 edges.

The images were initially obtained and manually annotated in terms of the three types of graffiti or their absence. A total of 3,154 images were identified as containing at least one type of graffiti. The numbers of occurrences of graffiti for types A, B, and C, were 2086, 476, and 592 respectively. In relative terms, 19.7% of the images contain at least an occurrence of graffiti, with an approximate 66%–15%–19% mixture of the types A, B, and C, respectively. That means that in the analyzed region, type A appears in two of three affected images picked at random, on average. The dominance of this particular type on the ratios may be associated with that type being more easily produced when compared to the other types (please refer to Fig. 1). The affected images may present more than one type of graffiti as can be seen in Fig. 4. Despite 88% of the images having

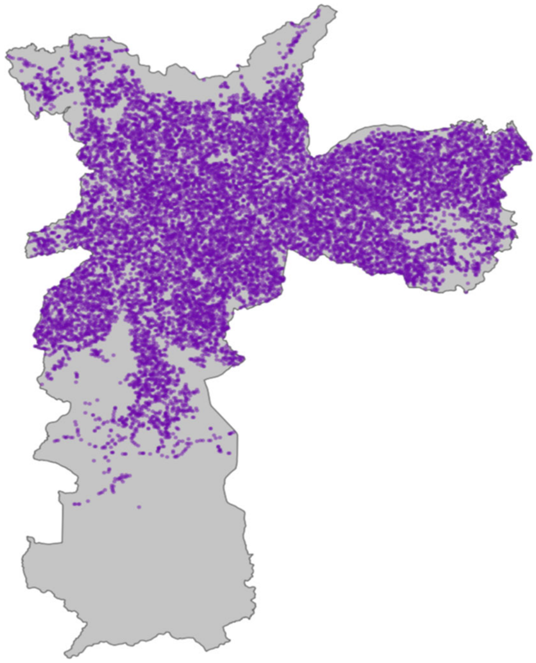


Fig. 3 Spatial distribution of the sample taken into account in the analysis. The shaded region corresponds to the city of São Paulo and each dot (16,000) represents an image considered in our analysis

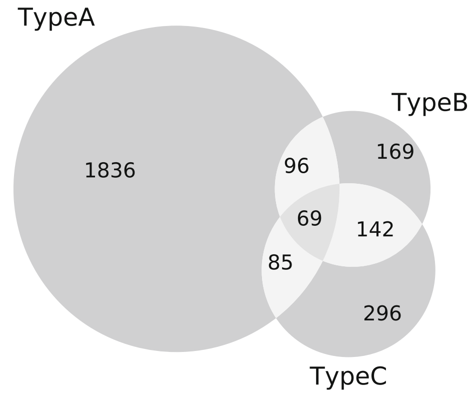


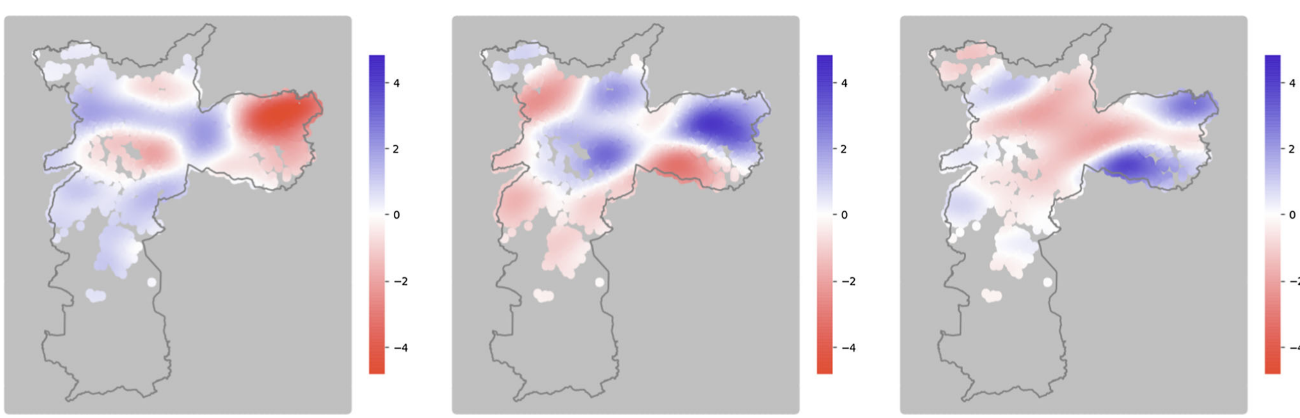
Fig. 4 Number of images containing each type of graffiti. Different types may be present in a single image and are represented by the intersection regions

occurrences of type A only, just 35% of type B occur in isolation.

A kernel density estimation with a Gaussian kernel was used to estimate the spatial distribution of the occurrences per type. Given our interest in getting the difference between these patterns, in Fig. 5 we depict the difference of the distribution of each type with respect to the mean distribution. Different patterns are observable for each type. It can be noticed that the northeast region of the map presents a low concentration of type A (red region in (a)) and a high concentration of type B (blue region in (b)). Also, two focuses of high concentration of type C can be observed in the eastern region (c). To discriminate between zero-valued points and sparse regions, regions with densities below one occurrence of graffiti per 500 m (0.31 mi) were filtered out from the analysis and in Fig. 5, they are shown in grey. The KL divergence between each distribution and the mean distribution was calculated as a means to quantify the discriminability of these patterns and 0.0137, 0.0141, and 0.009 were obtained for the distributions of type A, B, and C, respectively. It shows that the spatial distribution of type B differs the most from the average case.

3.1 Analyzing the network topology

The application of the Infomap community detection algorithm [15] resulted in the six network communities (C1–C6) shown in Fig. 6. The figure shows the obtained communities, with urban features annotated, including the meaning of the borders obtained and particular features of each community. As can be seen, the regions of separation mainly correspond to arterial roads and rivers that cross the city, which are natural separators of the topology of the city.



(a) Type A **(b)** Type B **(c)** Type C

Fig. 5 Graffiti occurrences distribution compared to the mean. The heatmaps show the difference between the distribution of each type and the mean distribution. The region in grey corresponds to sparsely sampled locations (below one occurrence of graffiti per 500 m)

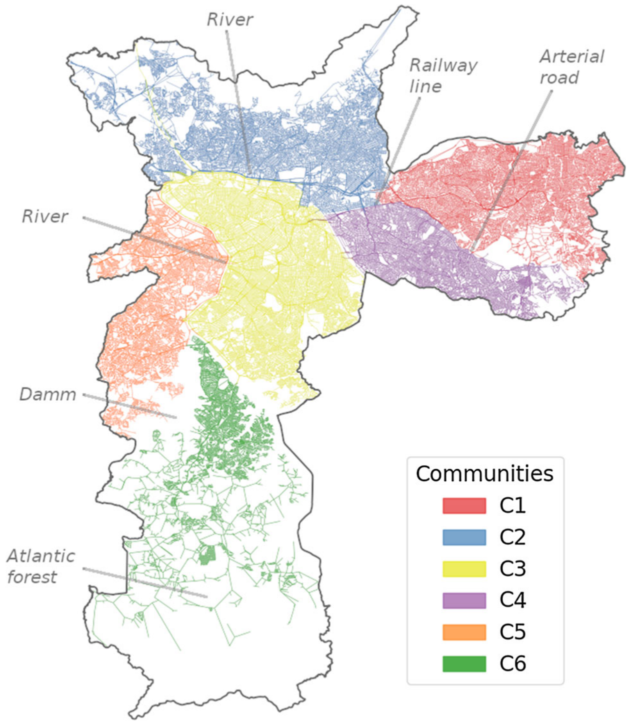


Fig. 6 Network communities obtained using Infomap [15]. The annotations show the consistency of the results obtained with the populated portions of the environment

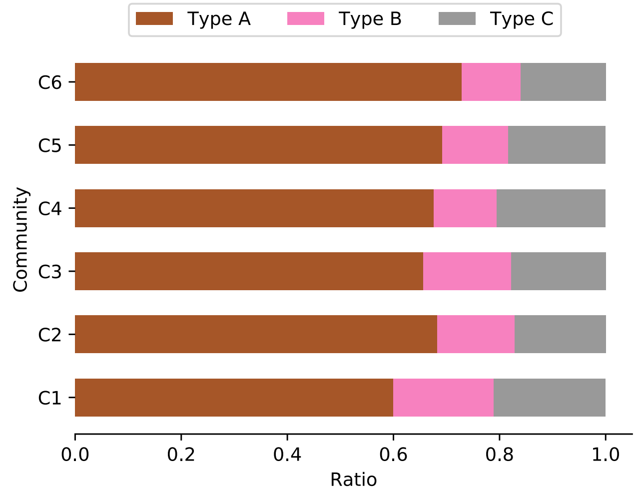


Fig. 7 Ratio of types of graffiti occurrences for the obtained communities. Each bar corresponds to a community and the partitions along each bar are scaled by the ratios of each type over the number of occurrences inside that community

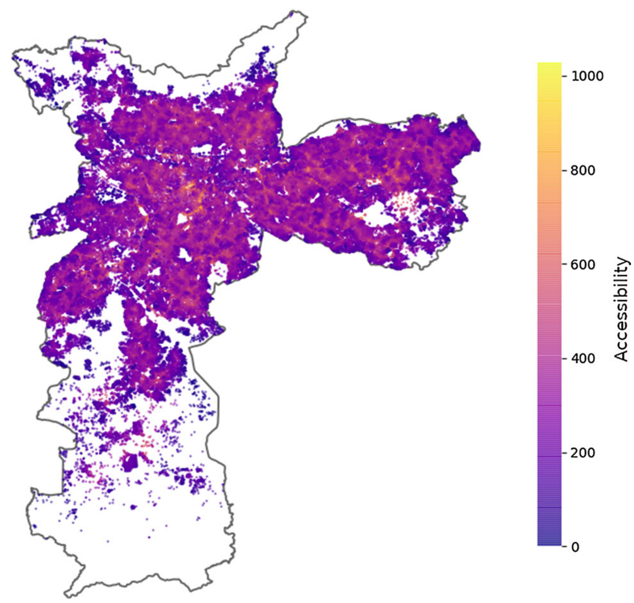


Fig. 8 Accessibility of the nodes of the street network of the city of São Paulo. As expected for this measurement, higher values are observed in the inner regions of the network, while smaller values tend to correspond to the network border. The network links were omitted in order to not clutter the image

To analyze the predominance of one type over another in each community, the relative occurrence of the graffiti types per community was calculated and is depicted in Fig. 7. This chart shows a pattern common to all communities, with most of the occurrences of type A, followed by type C and B. Despite the deviations found, the results indicate no significant variations in the proportions of each type of graffiti associated with the communities.

3.2 Analyzing the relationship between the topology and graffiti density

A 22-step random-walk ($h = 22$) was considered for the computation of the entire-network accessibility and the results can be seen in Fig. 8. The figure shows the vertices of the network colored by the accessibility measure. It can be noticed that lower values tend to appear at the border, while the peak values occur inside the identified regions. That means that, in general, the border regions can reach fewer regions than the central regions for the adopted topological scale of 22 steps.

To analyze the relationship of this vertex-based measurement and the spatial distribution of the graffiti occurrences, the graffiti count of each vertex was extracted from the previously obtained density. Vertices in sparse regions (grey regions in Fig. 5) were not considered. Fig. 9 shows the graffiti count and the accessibility measure for each vertex of the network, considering the whole city. The Pearson correlation coefficients between each type and the accessibility values were computed and 0.35, 0.36, and 0.29 of correlation were obtained with respect to each of the three types of graffiti. These results indicate small positive correlation between the accessibility value and the graffiti occurrences.

Figure 10a shows the correlations between the average accessibility and the graffiti density obtained con-

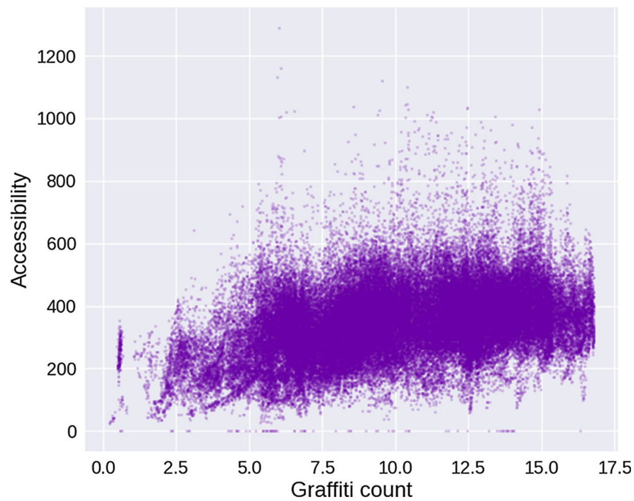


Fig. 9 Relationship between the network accessibility and the overall count of graffiti occurrences. Each point represents a vertex in the network and the correspondent graffiti occurrence count and the accessibility of the vertex in the network

sidering several values of accessibility steps for traditional random walks. The results respective to self-avoiding random walks are depicted in Fig. 10b. Therefore, the results in Fig. 10 provide a more systematic view of the possible overall relationship between the accessibility of the city network and the respective graffiti density with respect to varying random walk lengths and for two important types of random walks, traditional and self-avoiding.

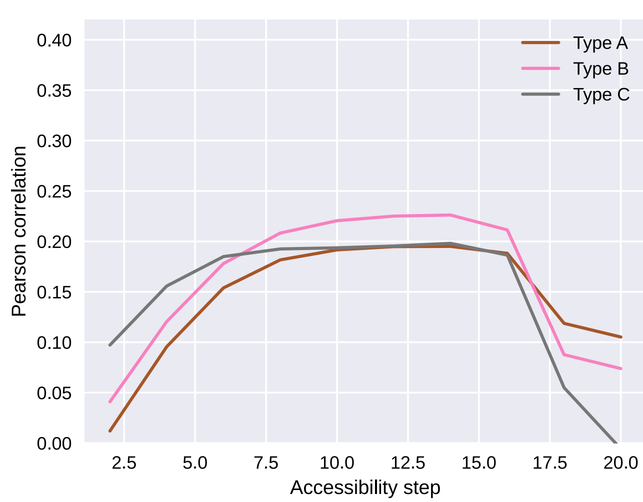
Interestingly, a peak of correlation takes place in both situations at similar positions. More specifically, in the case of traditional random walks, the peaks are observed for accessibility steps ranging from approximately 7.5–17.5, while the peaks are found between approximately 15 and 25 in the case of self-avoiding random walks. These results have interesting implications. First, they show that the type of random walks does not influence the correlation in a particularly strong manner, indicating that similar correlation values would be obtained in any of these cases. Then, we also have that the existence of peaks in well-defined intervals of accessibility extension would indicate that the facilitation of existence of graffiti at network places with higher accessibility depends critically on the extension of the walks as measured in blocks, suggesting that the graffiti production tends to take place at the observed displacements or spatial scales.

In the light of these correlations, we hypothesize that the accessibility is associated with the location of graffiti in two antagonistic ways: more accessible regions are more transited and thus gets more visibility for the produced content; and at the same time, this accessible ways are more monitored, which inhibit the production of graffiti in these regions.

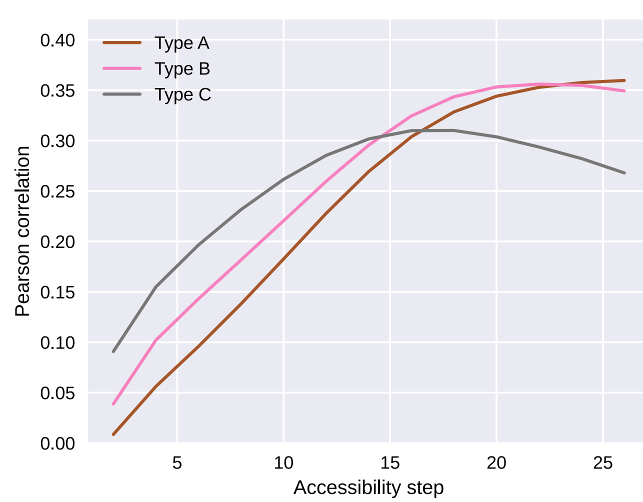
3.3 The correlation between the accessibility and graffiti density observed within each community

To investigate the above hypothesis further, we considered the correlations between the accessibility and the graffiti density within each of the six identified communities. We focus on accessibility length of 22 blocks considering self-avoiding walks, as this parameter configuration led to the highest correlations. Table 1 presents the so-obtained correlations for each of the three types of graffiti.

Some particularly high correlation values were obtained. In particular, correlations of 0.51, 0.56 and 0.49 were observed within community C2. Relatively high correlation values were found also within community C1. These two communities, C2 and C1, correspond roughly to north region of the city. Smaller cor-



(a) Traditional random-walk



(b) Self-avoiding random walk

Fig. 10 Correlation between the graffiti count and the accessibility versus the accessibility step. Calculation considering **(a)** a traditional and **(b)** a self-avoiding random walks

Table 1 Correlation between the accessibility values and each type of graffiti per community, considering the accessibility with a step size of h = 22. The obtained values range from almost no correlation (0.05), to moderate correlation (0.56)

Community	Type1	Type2	Type3
C1	0.41	0.39	0.34
C2	0.51	0.56	0.49
C3	0.13	0.29	0.21
C4	0.35	0.35	0.22
C5	0.27	0.36	0.29
C6	0.12	0.05	0.06

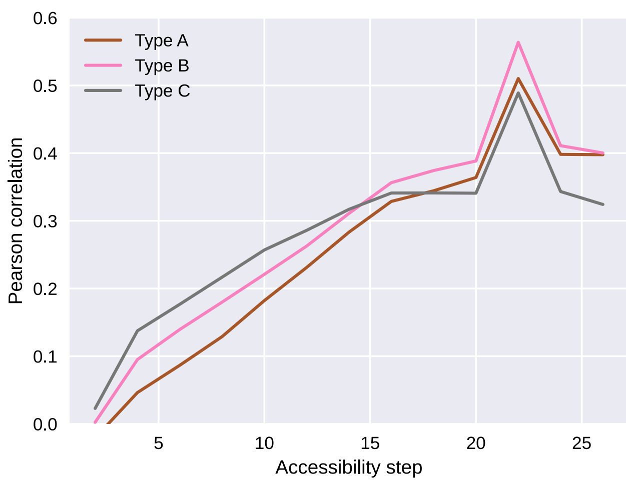


Fig. 11 Correlation of occurrences of graffiti and accessibility inside community C2. Variation with the accessibility step size (h)

relations were obtained for the other regions, such as the values 0.12, 0.05, 0.06 observed for community C6 which is related to the south region of São Paulo.

These results are particularly interesting because they suggest that the relationship between the accessibility of the city locations and the density of graffiti depends strongly on the urban communities or regions. In the case of community C6, we have that the respective urban region is more sparse, with a more heterogeneous topology, and also has substantially less graffiti. On the other hand, the two communities leading to higher correlations correspond to the regions with the two largest overall number of graffiti, being also characterized by a more uniform topology.

4 Concluding remarks

Graffiti is a common feature of many cities worldwide. Some previous works have already analyzed graffiti spatial distribution, but they typically do not consider the different visual forms of graffiti. Besides, they do not take into account city spatial and topological organization in the analysis, which would allow the respective topology to be tentatively associated with dynamical aspects such as mobility.

In this work, a network science-based method for analyzing the different types of graffiti, which are then studied by considering the modularity and accessibility measurements, is reported. A case study was performed considering the city of São Paulo, Brazil, and three graffiti types. The results showed that there are differences among the spatial distribution of these three types. The moderate positive correlation values between the accessibility on the network and the graffiti types indicates spatial distributions biases of the graffiti production in the city of São Paulo.

Possible future steps in the reported study could include the consideration of other cities, as well as the subdivision of the graffiti types in further subcategories.

Acknowledgements The authors would like to thank FAPESP (#2019/01077-3, #2015/22308-2, #2019/16223-5, #2018/09125-4 and #2018/10489-0), CAPES and CNPq (Grant no. 307085/2018-0); NSF awards CNS-1229185, CCF-1533564, CNS-1544753, CNS-1730396, CNS-1828576; the Moore-Sloan Data Science Environment at NYU, DARPA D3M and C2SMART. Any opinions, findings, conclusions and recommendations expressed in this material are those of the authors.

Author contributions

EKT was involved in all steps of this work. HFA contributed to the experiments and writing. CHC contributed to the analysis and writing. RMCJ and CTS contributed to the writing. LFC contributed to the analysis and writing.

Data availability statement This manuscript has no associated data or the data will not be deposited. [Authors' comment: While the data used in this data is originally public, the images were collected using the Google Maps platform, which Terms of Service forbid any kind of redistribution of the data.]

References

1. A. Alonso, Urban graffiti on the city landscape, in *Western Geography Graduate Conference* (1998)
2. M. Barthélemy, Spatial networks. Phys. Rep. **499**(1–3), 1–101 (2011)
3. S. Boccaletti, V. Latora, Y. Moreno, M. Chavez, D.-U. Hwang, Complex networks: structure and dynamics. Phys. Rep. **424**(4–5), 175–308 (2006)
4. R.A. Davis, K.-S. Lii, D.N. Politis, Remarks on some nonparametric estimates of a density function, in *Selected Works of Murray Rosenblatt*, pp. 95–100. Springer (2011)
5. J. Ferrell, *Crimes of Style: Urban Graffiti and the Politics of Criminality* (Garland, New York, 1993)

6. S. Fortunato, Community detection in graphs. Phys. Rep. **486**(3–5), 75–174 (2010)
7. Google. Google Maps (2005), https://www.google.com/maps. Accessed Apr 2021
8. B. Haworth, E. Bruce, K. Iveson, Spatio-temporal analysis of graffiti occurrence in an inner-city urban environment. Appl. Geogr. **38**, 53–63 (2013)
9. S. Kullback, *Information Theory and Statistics* (Courier Corporation, Chelmsford, 1997)
10. V. Megler, D. Banis, H. Chang, Spatial analysis of graffiti in San Francisco. Appl. Geogr. **54**, 63–73 (2014)
11. M. Newman, *Networks* (Oxford University Press, Oxford, 2018)
12. M. Newman, A.-L. Barabási, D.J. Watts, *The Structure and Dynamics of Networks* (Princeton University Press, Princeton, 2006)
13. OpenStreetMap Foundation. OpenStreetMap (2017), https://www.openstreetmap.org. Accessed Apr 2021
14. H. Priemus, P. Nijkamp, D. Banister, Mobility and spatial dynamics: an uneasy relationship. J. Transport Geogr. **9**(3), 167–171 (2001)
15. M. Rosvall, C.T. Bergstrom, Maps of random walks on complex networks reveal community structure. Proc. Natl. Acad. Sci. **105**(4), 1118–1123 (2008)
16. E.K. Tokuda, R.M. Cesar Jr., C.T. Silva, Quantifying the presence of graffiti in urban environments, in *International Conference on Big Data and Smart Computing*, IEEE, pp. 1–4, Japan (2019)
17. B.A.N. Travençolo, LdaF . Costa, Accessibility in complex networks. Phys. Lett. A **373**(1), 89–95 (2008)