SPATIAL ANALYSIS ABOUT USERS COLLABORATION ON GEO-SOCIAL NETWORKS IN A BRAZILIAN CITY

Aline Morais

[1]Universidade Federal de Campina Grande, Brasil

ABSTRACT

Geo-Social Networks (GSNs) are collaborative systems that has the geolocated information as main component. The geolocation resource integrates virtual and real worlds, allowing the comprehension about these two scenarios at same time. Based on that, this work define a process of spatial analysis of shared information on a GSN. The present work proposes the usage of six spatial features as feedback about collaborative behaviour on city. The spatial analysis aims understand if users' collaboration change among city census sectors. Understanding how users deal with GSNs in an area, will help about collaborative patterns per urban region. As result, this work detected spatial patterns among users in the GSN Foursquare of a Brazilian city. These patterns indicates that users' collaboration receive influences of extrinsic and intrinsic features of GSN and the comprehension about their users is a complex task.

KEYWORDS

Geo-Social Networks, Spatial Analysis, Collaborative Behaviour

1. INTRODUCTION

The Geo-Social Networks (GSNs) are social networks that share geolocated information, based on location of visited places by their users. The analysis of interactions in GSN can show how users collaborate in a place. Many investigations about GSNs has explored spatial analysis in order to map preferences and patterns of displacement of users. Other type of investigation about GSNs is about spatial analysis focused on a delimitated area, as a city [2]. This last line of investigation is few explored because of complexity to mining geolocated data in a specific region.

Based on that, the present work has as research problem the absence of investigation about collaborative behaviour of GSN users in Brazilian cities. The acceptation of GSNs in Brazil increase considerably in last years. At last, this work has as main aim the definition of spatial metrics able to measure the collaborative behaviours in GSNs. The data used to calculate these metrics is tips shared on GSN Foursquare. It is possible because tips have geographic coordinates integrated with their content.

The metrics observed in the present work are: a) coverage, b) proximity, c) spreading, d) activity, e) diversity and f) actuality. The coverage measures the percentage distribution in a region. The proximity evaluates the distance among tips shared consecutively on time and in a same region. Spreading represents the distance among tips in a same area, without temporal reference. Activity is the number tips in an area. Diversity corresponds to a number of different tips' authors in a same area. Lastly, the actuality represents the lifetime of tips in a region. At last, this work observed these metrics in subareas of city, called to census sectors.

This paper proposes methods to analyze spatially the GSN interaction over a city. First, the Section 2 describes the related works about spatial analysis in social networks. At next, the present work will explain about methodology adopted to describe these metrics over a Brazilian city. This section will also describe the first results about the metrics. At last, the section 4 will summarize about main findings of this work.

2. RELATED WORKS

GSN are social networks that qualify location as the core object of its structure [3]. There are investigations with GSNs that adopt the association between urban features and GSN features in order to understand human relationships and interactions [4] [5]. Related to the association between urban features and GSN features, it has a few works that observed it [6] [7]. Many of these works aim to extract Points of Interested (POIs) of its users in specific areas, mainly for recommendation techniques. However, the present paper found few works that focus on gap between digital worlds and socioeconomic factors about localities [19].

There are many metrics on GSN able to have relation with socioeconomic features on a city. Related to that, the present work shall highlight features used by it. The first metric is the diversity and it represents how users may vary based on collaboration [9] and origin [10], for example. Other metric used broadly is the activity. Normally this metric measures the users' participation on collaborative systems [8]. Related to coverage metric, some investigations adopt this metric in order to understand better about the percentage of some variables in a context generic. In collaborative systems, the coverage is useful to measure percentage of participation to associate with other occurrences outside of systems [11].

Some previous works highlight the importance to consider metrics with distance during analysis of features on collaborative systems [12]. Because of that, the present work will use two distance metrics on GSN: spreading and proximity. Spreading is a metric that measures the space among points in an area [13]. The proximity measures data as the number or frequency of contacts between individuals, the size of share-holding relations between enterprises, the number of members in common between organizations, and so on [14]. The last metric is the actuality and it represents the novelty degree shared in collaborative systems. Normally, this metric is based on lifetime features [15].

3. METHODS AND RESULTS

This work developed a longitudinal study, which aimed to identify peculiarities about the collaborative participation of GSN users over regions of a Brazilian city. To understand it, the present paper chose Campina Grande city to be part of analysis. It is a medium Brazilian city with around 626,937 people and it is in Northeast of Brazil [20]. Moreover, this work observed areas of Campina Grande, called census sectors. They are political-administrative divisions that represent territorial units, created and registered by *Instituto Brasileiro de Geografia e Estatística (IBGE – Statistics and Geography Brazilian Institute)*. The Campina Grande city has 439 census sectors, only in the urban area. At last, the chosen GSN for this work is the Foursquare. It is one of the most popular GSNs on Brazil in 2014. Furthermore, this system present 40 million users over the world.

As mentioned previously, the spatial metrics are focused on tips shared by GSN Foursquare. Tips represent annotations created by users in order to comment publicly about visited places in GSN. The present work analysed tips shared in census sectors of Campina Grande during March, April and May of 2014. The present work define the spatial metrics based on tips features, at following:

- • Diversity (user per sector): mean value about number of users in a census sector;
- • Activity (tips per sector): number of tips in a census sector;
- • Actuality (days): mean of lifetime of tips in a census sector;
- • Coverage (%): distribution percentage about tips in a census sector;
- • Spreading (meters): distance among users in a census sector;
- • Proximity (m): spatial distance among tips of different places in a census sector.

This work observes spatial patterns with assistance of ArcGIS 10.1 software, mainly to plot GSN metrics in census sectors and to calculate the metrics. With this software, the present work executes three geographic operations able to distinguish the presence of spatial patterns. The spatial techniques used in investigation are:

- • Spatial autocorrelation (Moran's I): It measures and analyzes the degree of dependency among tips in a census sector. This method shows measures how much close objects are in comparison with other close objects [16].
- • Hot Spot Analysis (Getis Ord Gi*): Hot Spot Analysis is a detection method of spatial clusters, which identifies spatial concentrations, statistically significant, of the high and of low values associated with a set of geographic features[17];
- • Cluster and Outlier Analysis (Anselin Local Moran's I): Given a set of weighted features, identifies where high or low values cluster spatially, and features with values that are very different from surrounding feature values [18].

3.1. Diversity

The diversity of tips reflects the popularity of places in an urban region. This work has as suspicion that sectors with high number of users per place, also present: a) Popular places; b) Places with updated tips; c) Places with tips' collaborators close among them. The Figure 1(a) details about diversity is in Campina Grande. The darker the sector, the greater the diversity. The darkest areas are located mainly in center of Campina Grande. It is possible realize the variation about diversity indexes over census sectors of the Brazilian city.

Firstly, it investigates if census sectors with same diversity index are near spatially on the city. To solve that the present work executed spatial auto-correlation test of Moran's I, as explained previously. If autocorrelation exists in a map, then this violates the fact that observations are independent from one another. The Moran's I index is a value between -1 (disperse) and 1 (grouped). According with the result in Figure 1(b), it is not possible say the number of different users per census sector (diversity) presents a dispersed or grouped behaviour over Campina Grande city, due to p-value higher than 0.05.
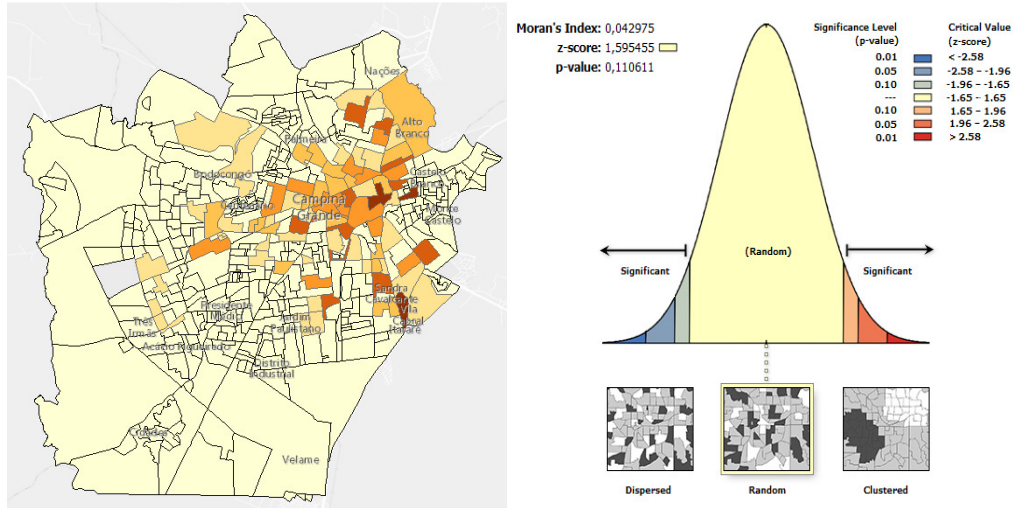
Figure 1. (a) Diversity in Campina Grande. (b) Result of spatial autocorrelation about Diversity.

The next step was the local analysis about diversity on census sectors, after detect the inexistence of a global pattern among census sector. Thus, this investigation applied Anselin Moran's I method as an indicator of spatial association. The results is in Figure 2 (a).

By the map, the majority of census sectors present not significant index compared with other diversity values in census sectors. The next procedure is the usage of a method able to identify groups, according to diversity on census sectors. The chosen method was hotspot analysis based on Getis-Ord Gi index. The hotspot analysis calculates the statistics about Gets-Ord Gi index and GiZ-score informs about groups with high and low values spatially. Observing the map, the central area presents a group of census sectors with the biggest quantity of different users in the same area (red sectors in Figure 2b). In minority, it saw a few sectors with diversity lower than average (blue sectors in Figure 2b). The yellow sectors represent diversity values without statistical significance.
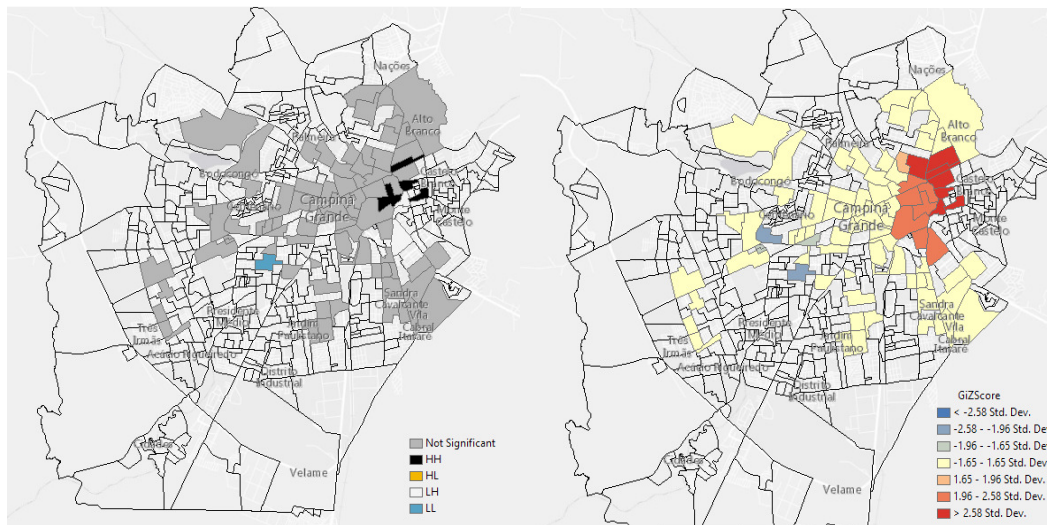

Figure 2. (a) Result of group mapping with Anselin Moran's I about Diversity. (b) Result of Hot Spot analysis about diversity.

If a census sector has a high diversity and does not present popular places, probably the census sector present new places to visit. Related to this, it is an urban area with many novelties shared on GSN. Thus, novelty gives seasonality to services and features of places. At last, this work will investigate this possibility through actuality investigation, in next sections.

3.2. Activity

The activity represents the number of tips shared per square kilometer in each census sector of Campina Grande city. The initial suspicious is census sectors, which present places with high activity, also have places with recent tips on it. The description about activity on census sectors is in Figure 3 (a).

The activity is higher in areas located on the city center. However, the major part of census sectors did not present any activity, as it shows the clearest areas of Figure 3(a). As in diversity perspective, this paper investigated about spatial patterns related to activity index over census sectors. Again, it did a spatial autocorrelation of Moran's I, in Figure 3(b).
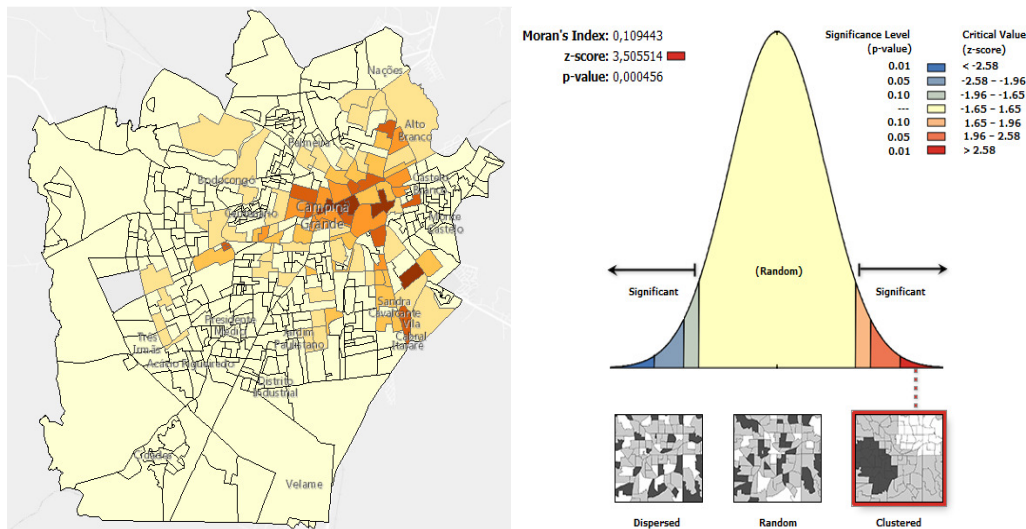

Figure 3. (a) Activity over census sectors of Campina Grande. (b) Result of spatial autocorrelation about Activity.

The Moran's index (I = 0.10) shows a spatial pattern slightly aggregated based on activity index over census sectors. It presents significant results due to the p-value less than 0.05 and the z-score higher than 0.96. These results mean that the activity about creation of geolocated content over Campina Grande are dependent to other factors.

After detect the presence of global spatial patterns about the number of tips per square kilometer, it is necessary evaluate on a local scale. Then, it applied Anselin Moran's I (Figure 4a) and Hot Spot analysis (Figure 4b).

According to Anselin Moran's I results, the activity behaves differently over census sectors. For example, there are some areas with high contribution of tips per squared kilometer, rounded by sectors with high activity (black sectors in Figure 4a). According to Figure 4(b), there are two main groups based on activity: sectors with activity higher than average (sectors in red colour) and sectors with activity lower than average (sectors in blue colour). In this last group, there is only one sector.
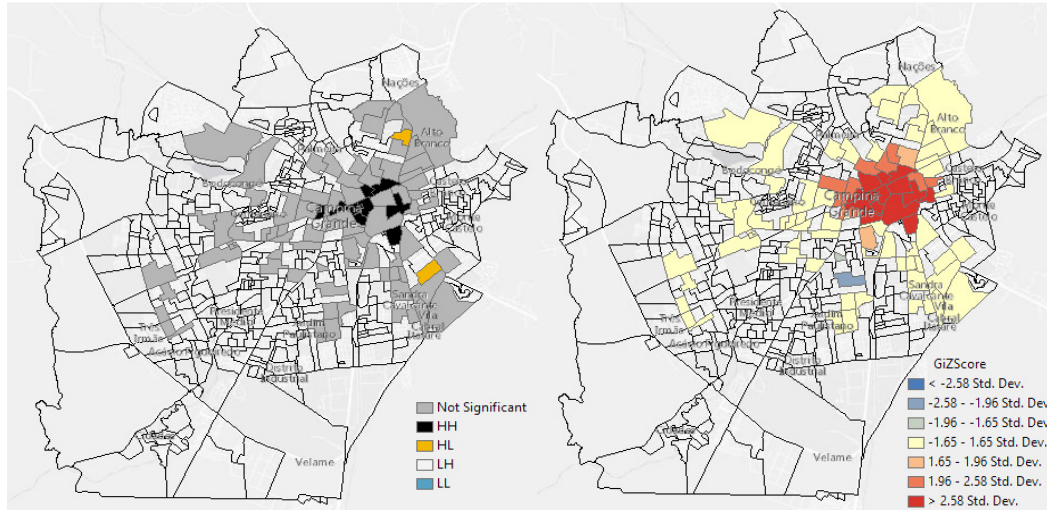
Figure 4. (a) Result of group mapping with Anselin Moran's I about Activity. (b) Result of Hot Spot analysis about activity.

Therefore, some sectors with high activity in Hot Spot Analysis are not significant in Anselin Moran's I. It happens because these sectors are spatially isolated. In this sense, the center of Campina Grande has the highest quantity of tips per square kilometer with some areas of high activity isolated among other census sector with high activity.

3.3. Coverage

The coverage has as aim to measure the distribution of tips on census sector. The suspicious is sectors with high coverage, also present recent tips. According to Figure 5 (a) the coverage in Campina Grande is heterogeneous over census sectors. As happened with two previous metrics, there is a presence of high indexes in the center of Campina Grande city. Moreover, the majority of sectors present coverage near to 0% between March and May of 2014.
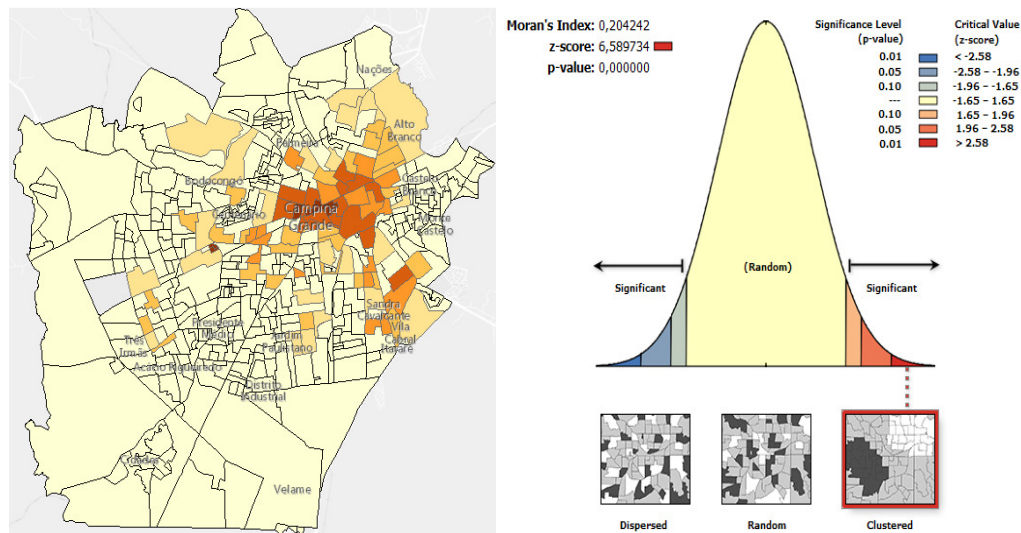

Figure 5. (a) Coverage over census sectors of Campina Grande. (b) Result of spatial autocorrelation about Coverage.

6

Again, this work applied the spatial autocorrelation in order to understand the existence or not of patterns on Campina Grande city. According to results of Figure 5 (b), it is present a slightly clustered behaviour related to coverage metric over census sectors. Related to initial results about presence of patterns, the present investigation applied in sequence the Anselin Moran's I (Figure 6a) and Getis-Ord Gi* (Figure 6b) methods in order to obtain details about these patterns.
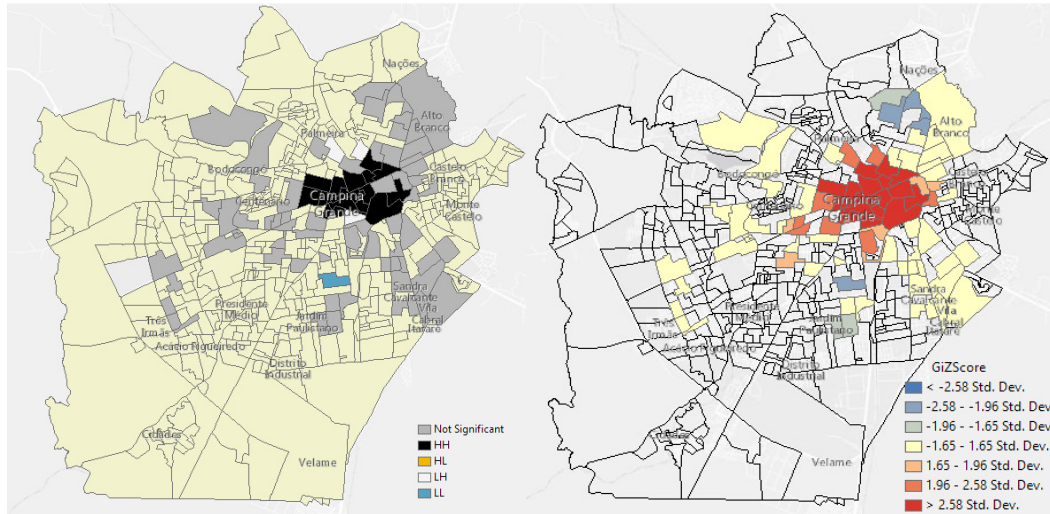

Figure 6. (a) Result of group mapping with Anselin Moran's I about Coverage. (b) Result of Hot Spot analysis about Coverage.

The results about Anselin Moran's I method presented, until now, the best quantity of sectors with high coverage (HH Legend in Figure 6a). All of these sectors is concentrated in the center of Campina Grande city. One part of Liberdade district has low coverage (LL Legend on Figure 6a). Moreover, part of Palmeira, Conceição and São José district present a low coverage rounded by sectors with high coverage (LH Legend on Figure 6a).

According to results of Getis-Ord Gi* method, it is possible to distinguish two main groups. The former belongs to sectors with a coverage higher than average (red sectors in Figure 6b). The latter group is about sectors with coverage lower than average (blue sectors in Figure 6b). The blue sectors are part of districts belonging to Liberdade, Louzeiro, Jardim Continental, among others.

3.4. Actuality

The actuality measures how new are the shared tips per sector. The first impression is census sectors with high actuality also have an area with popular places, collaborative users and a high coverage. The panorama about actuality metric in Campina Grande city is showed in Figure 7(a).
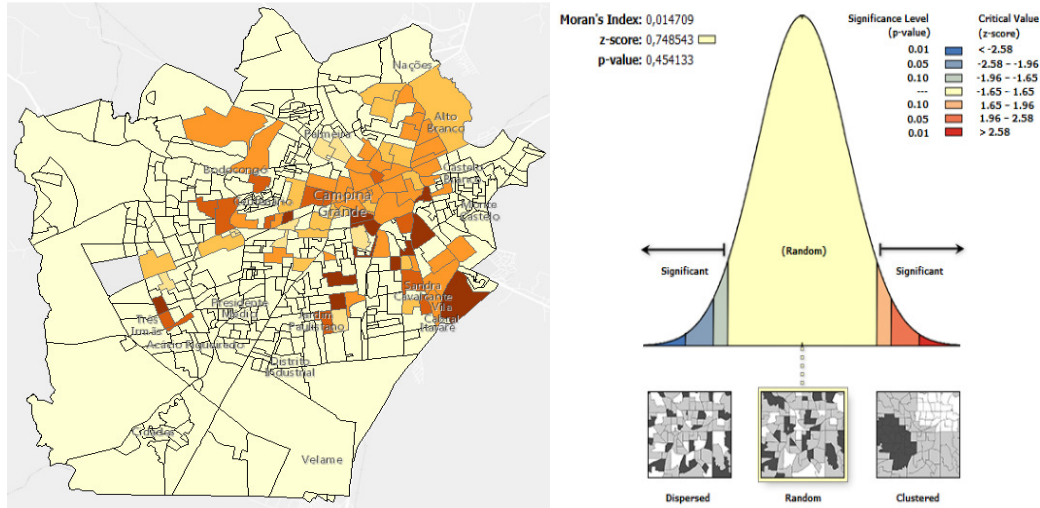
Figure 7. (a) Actuality over census sectors of Campina Grande. (b) Result of spatial autocorrelation about Actuality.

According to Figure 7(a), it is possible realize a centralized behaviour at the first moment. In this case, clear sectors mean sectors with the best index of actuality. In this work, the actuality is a measure between 69 and 159 days of the lifetime. It happens because has a relation with collecting date and mining date.

Even as previous features, the present work did a spatial autocorrelation to get more results about spatial patterns. According to results of Figure 7b, it cannot say anything about actuality in Campina Grande sectors because the result do not have statistical significance (p-value > 0.05). However, the local spatial analysis is able to understand the actuality over census sectors. Based on that, it executed Anselin Moran's I method. As explained previously, this method helps to identify outliers and spatial groups, according to a specific feature. The results are in Figure 8a.
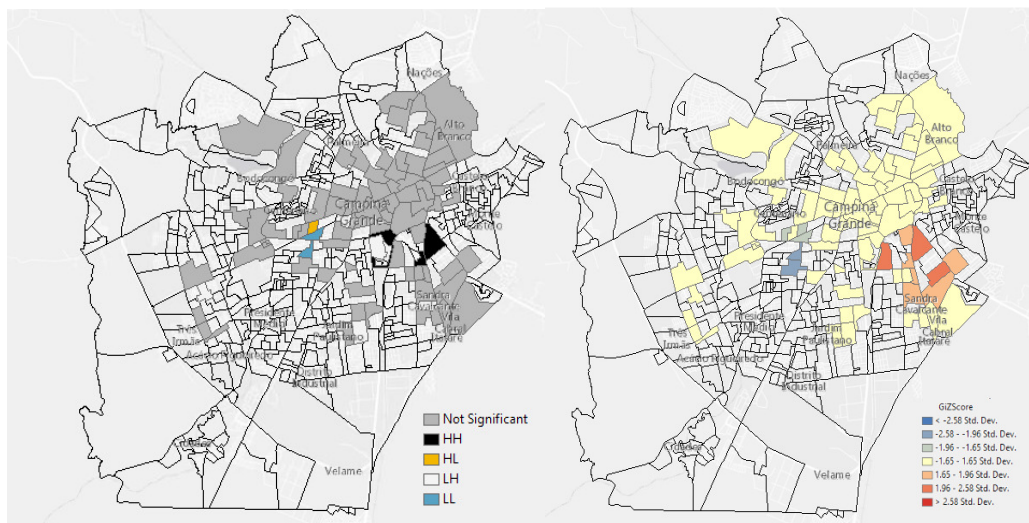

Figure 8. (a) Result of group mapping with Anselin Moran's I about Actuality. (b) Result of Hot Spot analysis about Actuality.

8

The results of Figure 8(a) showed that there is not big cluster related to actuality in Campina Grande city. It has three sectors (referee to Catolé and Estação Velha districts) that present high values of actuality (HH Legend in Figure 8a). In other words, these sectors present places with a little novelty. Perhaps, the tips on these places did not update. The sector belongs to Centenário district present a high value related to lifetime, but it is rounded by others sectors with younger tips (HL Legend in Figure 8a). At last, the Santa Rosa and Centenário districts presented the youngest tips compared to all sectors (LL Legend in Figure 8a).

It found out that census sectors of a same district present opposite behaviour related to actuality metric. It happens in Centenário district. The north part of this district presented ancient tips and the south area presented younger tips. It happens because the north area is a residential area while the south area has banks and restaurants that attract people flow. The next step is mapping significative groups based on tips actuality. To do this, it ran Getis-Ord Gi* method (Figure 8b). According to results of Figure 8(b), it detects the absence of concentrated areas related to actuality metric in Campina Grande city. It detected some sectors belonging to Catolé, José Pinheiro and Sandra Cavalcante districts, which presented values higher than average of tips. Otherwise, the part of Santa Rosa district presented sectors with tips more recent.

3.5. Spreading

The spreading reflects how consecutives tips are near overtime. Census sectors with low spreading probably will present popular places, places with high participation, places with new tips and low coverage. The general configuration about spreading is showed in Figure 9(a).
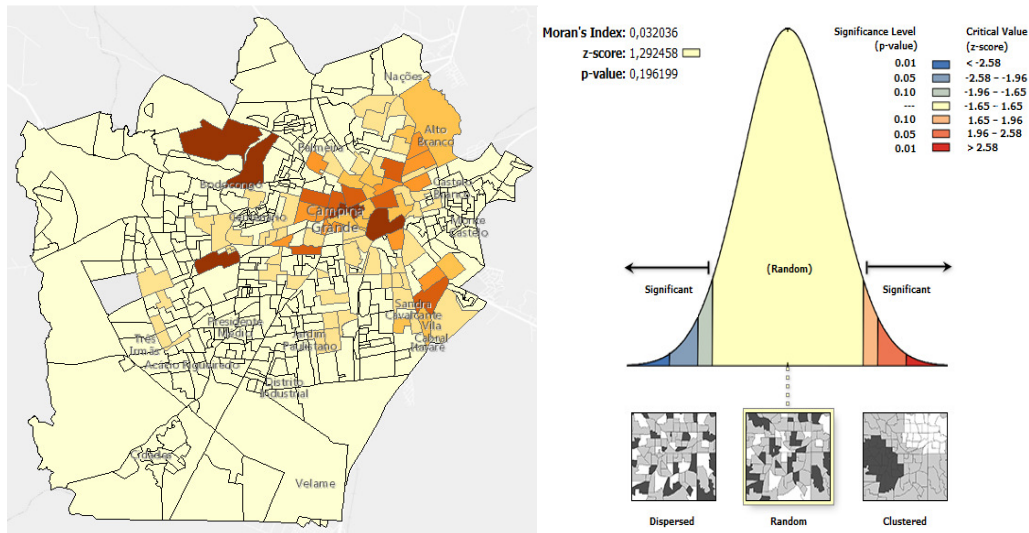

Figure 9. (a) Spreading over census sectors of Campina Grande. (b) Result of spatial autocorrelation about Spreading.

The sectors with the highest spreading are belonging to districts of Bela Vista (3.3 meters), Catolé (8.2 meters), Centenário (19.4 meters) and Center (19.6 meters). Otherwise, there are some sectors with spreading near to 0 meters because they have tips created in the same or near places. It happens in districts of Liberdade, São José, Palmeira, and Conceição, for example. The sectors with tips shared more distant among themselves are at Bairro Universitário, Centro and Dinamérica districts.

According to results in Figure 9(b), it is not possible conclude anything about Moran's I spatial autocorrelation because the results are not statistical significance (p-value > 0.05 and z-score < 1.96). Although it did not detect a global pattern, it is possible mapping cluster in a local scale by Anselin Moran's I method, as seen in Figure 10 (a).
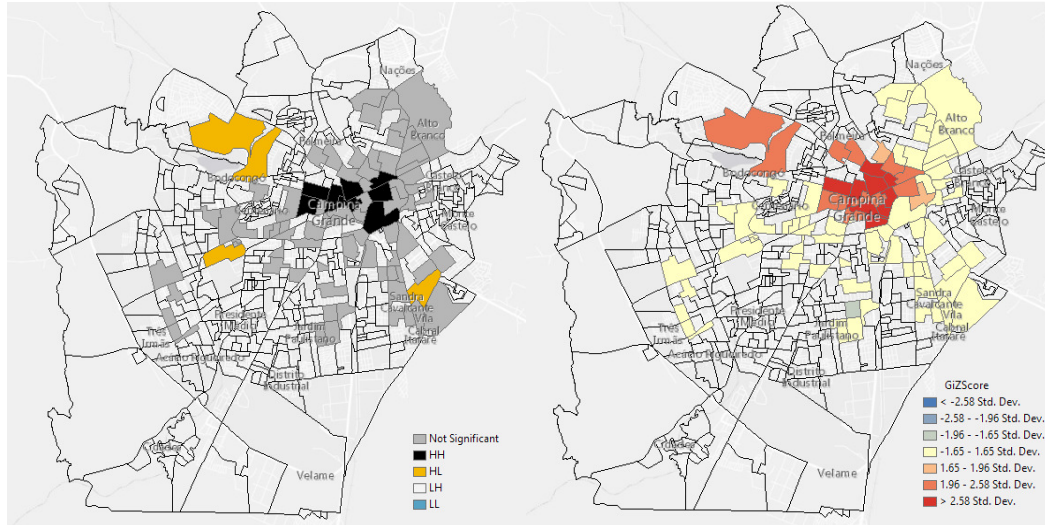

Figure 10. (a) Result of group mapping with Anselin Moran's I about Spreading. (b) Result of Hot Spot analysis about Spreading.

Based on results of Figure 10 (a), it is possible to observe two main groups. The former is sectors with a long distance among tips (legend black in Figure 10 (a). The latter is three sectors with a long distance among their tips and rounded by sectors with opposite behaviour. The sectors correspond to part of Bairro Universitário, Dinamérica and Sandra Cavalcanti districts.

At next, it will see this pattern based on Getis-Ord Gi* method. Based the results of Figure 10 (b), it detected the presence of clusters on central area of city and in Bairro Universitário district. They present high level of spreading in sectors. At last, a part of Jardim Paulistano district has tips near among their tips

3.6. Proximity

The proximity is a spatial metric that aims understand distance among tips in different places. Its formula is a ratio between the mean of distance and quantity of places in each sector. The panorama is in Figure 11 (a).

Based on results of Figure 11 (a), it can see a high distance among places mainly in the district of Bairro Universitário. It happens because this district presents few constructions, especially two universities. The next stage is the execution of spatial autocorrelation in order to evaluate the patterns about proximity metric (Figure 11b). The result of p-value about Moran's I method did not reveal statistical significance on results.
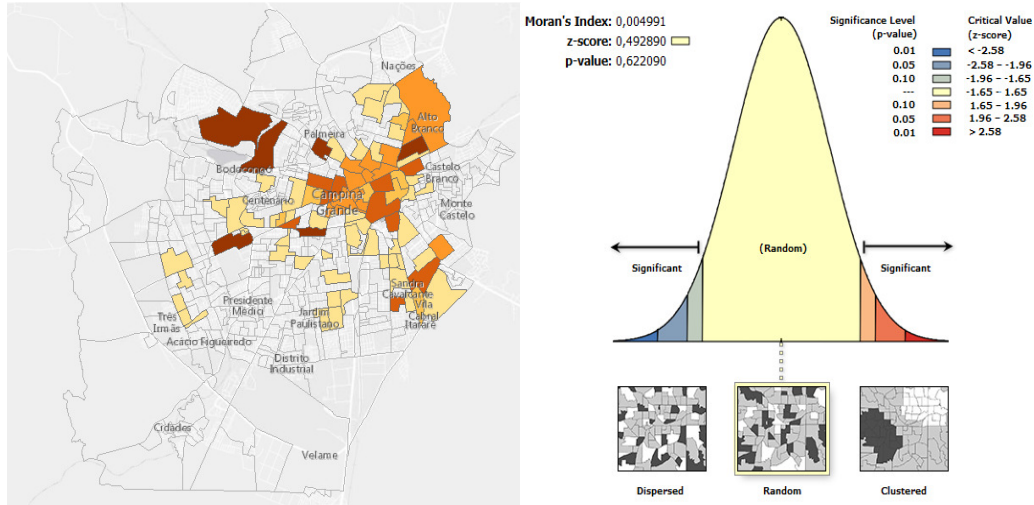
Figure 11. (a) Proximity over census sectors of Campina Grande. (b) Result of spatial autocorrelation about Proximity.

The other way to find patterns about proximity is applying Anselin Moran's I method. Based on the results of Figure 12(a), it highlights only four sectors. A part of the district of Prata and Centro present high distance among tips in different places. At last, a part of Quarenta and Dinamérica districts have a high value of proximity, rounded by sectors with small values of proximity among places.
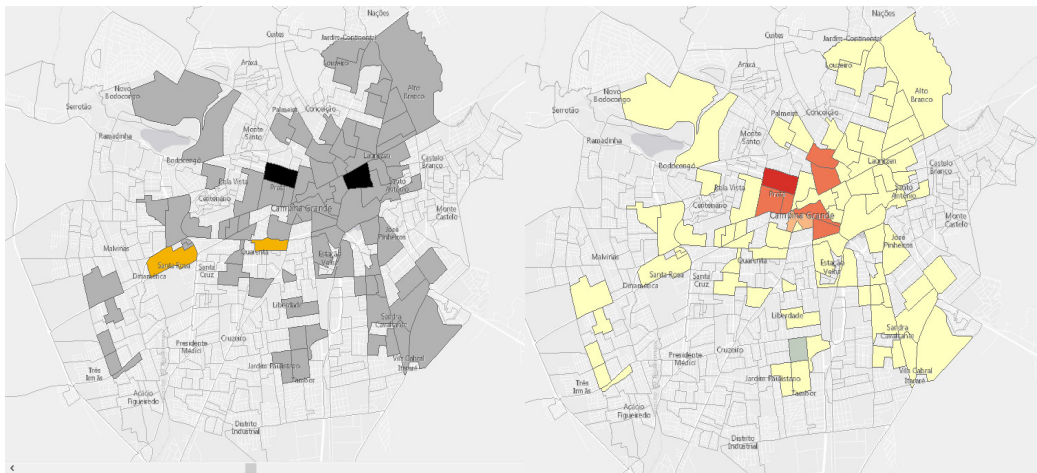

Figure 12. (a) Result of group mapping with Anselin Moran's I about Proximity. (b) Result of Hot Spot analysis about Proximity.

At next, the present paper executed Getis-Ord Gi* method to mapping better these sectors. Based on results of Figure 12(b), it detected a cluster with proximity value lower than the average of sectors. This cluster is represented by parts of Prata, Centro e São José districts.

4. CONCLUSIONS

Related to diversity, the maps showed that sectors, which present high quantity of users per sector, also have users so near among them, a high coverage of tips and a high number of tips per square kilometer. Census sectors with high activity also are sectors with a high number of tips per square kilometer and a high number of tips per place. Moreover, sectors with high activity also present high coverage and distance among consecutive tips.

Related to coverage, this work found out sectors of Campina Grande, which present high coverage also present a high activity, generally concentrated in the city center. Based on the age of tips per sector (actuality), the present paper found out that census sectors with novel tips are areas with a low number of tips per place (activity).

Comparing maps of spreading and proximity, it is possible realize some sectors with low spreading because their tips belonged to the same place. For example, some sectors in the city center were much tagged in popular places. Otherwise, a part of Dinamérica district showed tips shared in different places but near among them. Then, it shows how important consider spreading and proximity during the spatial analysis process.

Another interesting fact is during comparisons between spreading and proximity in an area of Campina Grande, called Açude Velho. The spreading index in this area is high and it happens due to two possible reasons. The first is because consecutive tips are done in opposite sides of the area. The second possible reason is due to physical impediment. Observing proximity index, it confirms the second possibility as true. It happens because this work realized that this spreading index is homogeneous among census sectors on that region.

Finally, sectors with high proximity, also present a high number of tips per square kilometer and a high coverage. For next phase of this research, this work will compare these results with other cities.

Thus, the results about the present investigation reinforce the importance for understanding urban context. As the future works, it will map which urban factors are able to affect these different behaviors about GSN collaboration over the city. The second planning is understand how this comprehension may be useful for other GSN users, during the decision making process. Probably, it will be through recommendation techniques based on the kind of place that a GSN user want visit.

ACKNOWLEDGEMENTS

REFERENCES

[1] Bao, J., Zheng, Y., Wilkie, D., & Mokbel, M. (2015). Recommendations in location-based social networks: a survey. GeoInformatica, 19(3), 525-565.
[2] Jiang, B., & Miao, Y. (2015). The evolution of natural cities from the perspective of location-based social media. The Professional Geographer, 67(2), 295-306.
[3] Symeonidis, Panagiotis, Dimitrios Ntempos, and Yannis Manolopoulos. Recommender Systems for Location-based Social Networks. Springer, 2014.
[4] Davis Jr, C. A., Pappa, G. L., de Oliveira, D. R. R., & de L Arcanjo, F. (2011). Inferring the location of twitter messages based on user relationships. Transactions in GIS, 15(6), 735-751.

[5] Silva, A. D. S. E. (2008). Alien revolt (2005-2007): A case study of the first location-based mobile game in Brazil. Technology and Society Magazine, IEEE, 27(1), 18-28.

[6] Pinquart, M., & Sörensen, S. (2000). Influences of socioeconomic status, social network, and competence on subjective well-being in later life: a meta-analysis. Psychology and aging, 15(2), 187.

[7] Gao, H., Tang, J., Hu, X., & Liu, H. (2015, February). Content-aware point of interest recommendation on location-based social networks. In Proceedings of the 29th AAAI Conference on Artificial Intelligence.

[8] Viren, P. P., Vogt, C. A., Kline, C., Rummel, A. M., & Tsao, J. (2015). Social network participation and coverage by tourism industry sector. Journal of Destination Marketing & Management.

[9] Weng, L., & Menczer, F. (2015). Topicality and Impact in Social Media: Diverse Messages, Focused Messengers. PloS one, 10(2), e0118410.

[10] Morais, A., & Andrade, N. (2014, January). The Relevance of Annotations Shared by Tourists and Residents on a Geo-Social Network During a Large-Scale Touristic Event: The Case of São João. In COOP 2014-Proceedings of the 11th International Conference on the Design of Cooperative Systems, 27-30 May 2014, Nice (France) (pp. 393-408). Springer International Publishing.

[11] Ellison, N. B. (2007). Social network sites: Definition, history, and scholarship. Journal of Computer Mediated Communication, 13(1), 210-230.

[12] Bjørn, P., Esbensen, M., Jensen, R. E., & Matthiesen, S. (2014). Does distance still matter? Revisiting the CSCW fundamentals on distributed collaboration. ACM Transactions on Computer-Human Interaction, 21(5), 1-27.

[13] Wong, L. H., Pattison, P., & Robins, G. (2006). A spatial model for social networks. Physica A: Statistical Mechanics and its Applications, 360(1), 99-120.

[14] Scott, J., & Carrington, P. J. (2011). The SAGE handbook of social network analysis. SAGE publications.

[15] Wilson, C., Boe, B., Sala, A., Puttaswamy, K. P., & Zhao, B. Y. (2009, April). User interactions in social networks and their implications. In Proceedings of the 4th ACM European conference on Computer systems (pp. 205-218). Acm.

[16] Carrington, Peter J., John Scott, and Stanley Wasserman, eds. Models and methods in social network analysis. Vol. 28. Cambridge university press, 2005.

[17] Grubesic, T. H. (2006). On the application of fuzzy clustering for crime hot spot detection. Journal of Quantitative Criminology, 22(1), 77-105.

[18] Shaker, R., Craciun, A., & Gradinaru, I. (2010). Relating land cover and urban patterns to aquatic ecological integrity: a spatial analysis. Geographica Technica, 1, 76-90.

[19] Sales, A., Alves, L., Araújo, M., Menezes, A., Morais, A., & Andrade, N. (2013, October). O uso de uma rede geossocial nas cidades brasileiras e sua relação com fatores socioeconômicos. In Proceedings of the 12th Brazilian Symposium on Human Factors in Computing Systems (pp. 142-147). Brazilian Computer Society.

[20] IBGE- Instituto Brasileiro de Geografia e Estatística. Available at: http://cidades.ibge.gov.br/painel/painel.php?codmun=250400. Acessed in June 2015.

AUTHOR

Aline Morais

She is PhD student at Universidade Federal de Campina Grande (UFCG). She was a professor at Instituto Federal de Educação, Ciência e Tecnologia da Paraíba (IFPB) in courses of Electrical Engineering, GIS Technology, Computer Networking Technology, Systems in Telecommunications Technology and Systems for the Internet Technology. She is technologist in Computer Networking, by IFPB, and is Msc. in Informatic at Universidade Federal da Paraíba (UFPB). Nowadays, she is in Laboratório de Sistemas Distribuídos (LSD), in UFCG. Recently, she has collaborated with L3S Reseach Center (Hannover, Germany) for working in PhD sandwich.