

INVESTIGATION OF SPATIAL DATA WITH OPEN SOURCE SOCIAL NETWORK ANALYSIS AND GEOGRAPHIC INFORMATION SYSTEMS APPLICATIONS

L. Sabah ^{a,*}, M. Şimşek ^b

^a Department of IT, Düzce University, 81620 Düzce, Turkey – leventsabah@duzce.edu.tr

^b Department of Computer Engineering, Faculty of Engineering, Düzce University, 81620, Düzce, Turkey – mehmetsimsek@duzce.edu.tr

KEY WORDS: Social Network Analysis, Gephi, QGIS, Open Source Geographical Information Systems, Spatial Data

ABSTRACT:

Social networks are the real social experience of individuals in the online environment. In this environment, people use symbolic gestures and mimics, sharing thoughts and content. Social network analysis is the visualization of complex and large quantities of data to ensure that the overall picture appears. It is the understanding, development, quantitative and qualitative analysis of the relations in the social networks of Graph theory. Social networks are expressed in the form of nodes and edges. Nodes are people / organizations, and edges are relationships between nodes. Relations are directional, non-directional, weighted, and weightless. The purpose of this study is to examine the effects of social networks on the evaluation of person data with spatial coordinates. For this, the cluster size and the effect on the geographical area of the circle where the placements of the individual are influenced by the frequently used placeholder feature in the social networks have been studied.

1. INTRODUCTION

A point of interest (POI) is a spatial location that is useful, attractive specific point location such as Tourist attractions, restaurants, cafés, bars, cinemas, hospitals etc. This POI locations providing individuals to more opportunities to experience modern life. (Gao et al., 2015). With usage of mobile devices individuals' check-in with this POIs to explore places or decide "where to go" according to their personal interests. And so on influencing their relationships more easily to their decisions too (Gao et al., 2015). Increasing of mobile device usage and easy internet access, location-based social services become prevalent (Cheng et al., 2012). Online location-based social networks (LBSNs) have attracted millions of users to share their social friendship, experience and tips of Point of- interest (POI) via check-ins (Cheng et al., 2012). In this study, we used poi locations to examine individuals' connections to other individuals.

2. MATERIALS AND METHODS

2.1 Materials

In this study, SNAP (Stanford Large Network Dataset Collection) datasets of Stanford University (Leskovec et al., 2014) were used to perform social network analysis on spatial data. The Brightkite dataset is used for location based online social networks from within datasets. Brightkite is the first site that brings social media placemarks. 2009 data are used because it is given for 12 months in the Brightkite dataset. Europe has been chosen as the study area where the data density is high. To obtain data in Europe, QGIS (QGIS, 2009), an open source geographic information system, was used. The latitude and longitude coordinate values in the dataset were created as point layer with QGIS. Then the intersection of the European

continent with the vector shp layer was made. After this process, the dataset data stored in the European borders were recorded in csv format. In the dataset, spatial intersection is performed through the QGIS and the area defined by the European border lines, and the point node data is filtered according to the spatial data of the individuals. Thus, it has been tried to determine the intensity of the people that they affect in local settlements. The edge data is also filtered according to the id values of the users in the filtered data. In the edge data, there are source and target user id values. Source id is the id value of the individual doing check-in. The target value is the id values of the other related individuals.

2.2 Methods

In this study, the levels of individual check-ins influenced by other individuals in social networks were tried to be analyzed using spatial data. In this context, the SNAP Brightkite data set is used. Check-ins in the data set are defined as node data. And their relations are defined by edge data. Node values are individuals' user id, check-in time, latitude, longitude and location id values. There are many software for social network analysis. GEPHI (Bastian et al., 2009), is an open source alternative software. By importing nodes and edges in the Brightkite dataset, it is possible to divide the available data into groups according to modularity values. In social network analysis, there are many measurements that help to detect communities. Modularity is one measure of the structure of networks or graphs. It was designed to measure the strength of division of a network into modules and used for community structure detections in networks. (V. K. Sihag et al, 2014). Networks with high modularity have dense connections between the nodes within modules but sparse connections between nodes in different modules.

* Corresponding author

15 groups with the highest modularity value were colored. The id, date, latitude, longitude, and location id values of the placeholder are used in the node import process. In the edge import process, the source node and target node are selected as type undirected. Weight values are left as 1. After importing the data, the data in the Gephi interface is displayed on the laboratory screen. The modularity value to be used for clustering of the data is obtained by pressing the modularity button on the Statistics screen. There are various algorithms for clustering data on the home screen. In this study, Force Atlas 2 (Jacomy et al., 2014) method which supports up to 1.000.000 nodes in $O(n \log n)$ complexity is used for network analysis.

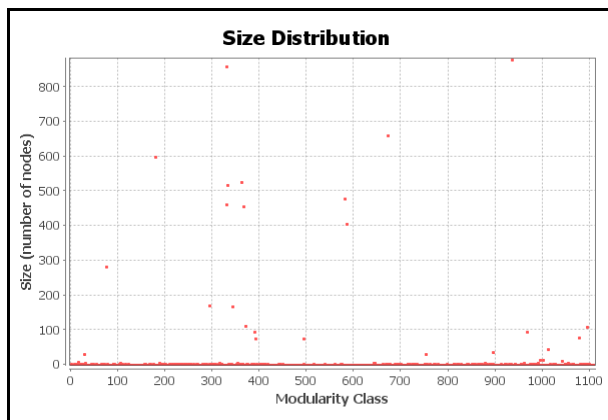


Figure 1. Modularity class values of the analyzed social network

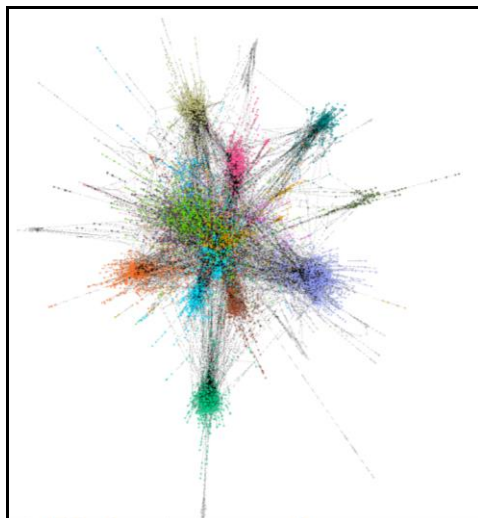


Figure 2. Show on social network with ForceAtlas2 layout

The coloring of nodes in Figure 2 is made according to modularity values. These values were found by running the modularity command in the Gephi Statistics panel for community detection (Figure 3). After, the color palette was created for community, which has the highest modularity value according to the modularity values found (Figure 4). After the color palette is created, the first 15 communities with the highest modularity value such as Figure 5 are colored according to the determined color palette.

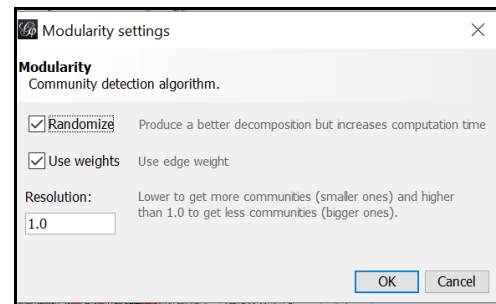


Figure 3. Community Detection in GEPHI

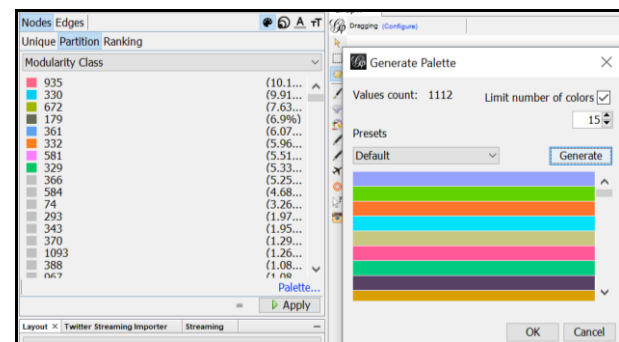


Figure 4. Generating color palette for top 15 modularity valued communities

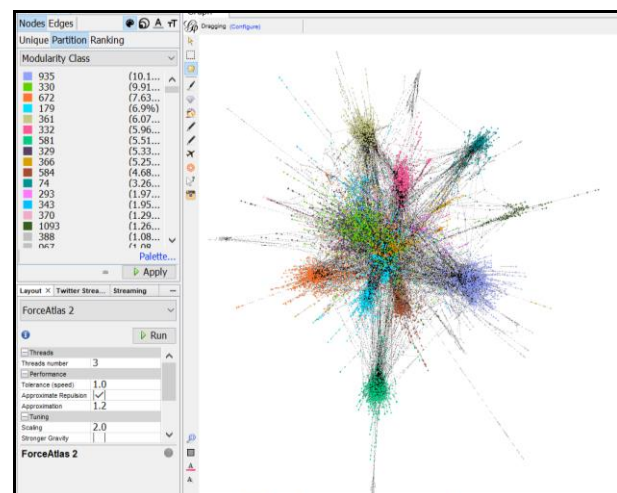


Figure 5. Coloring the communities with modularity values

As a result of the analysis, 1112 communities were formed (Figure 1). In Figure 1, the graph shows the node counts corresponding to the modularity classes. The overall social network analysis was conducted with 15 groups with the highest modularity value (Figure 2). According to this, placements made by individuals are generally influenced by friends. It is envisaged that each node will be effective on an area of no more than 50 km in order to ensure that the location reported in a geo-targeted social network is easily accessible to other individuals. The heat map in Figure 6 is obtained. It has been found that the obtained heat map is enough to propagate by affecting the other individuals.

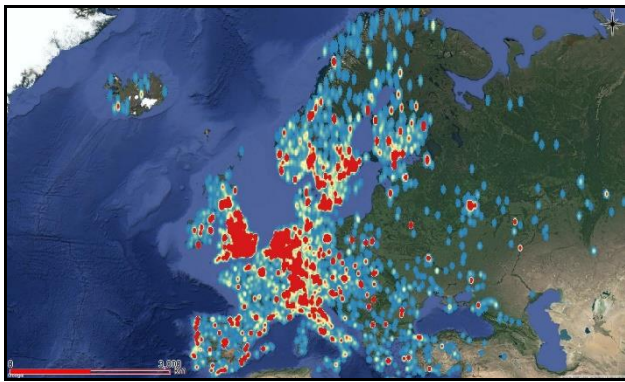


Figure 6. Heat map of 50 km radius nodes

3. RESULTS AND RECOMMENDATIONS

3.1 Results

The use of social media by individuals is increasing day by day. The main reason for this is the increase in the number of individuals who are online. The use of check-ins in social media basically means that individuals want to get a quick decision on where to go and get a decision not to go there. After running modularity command 1112 communities found. Top 15 of communities with highest modularity value presents %77.07 nodes. By examining the sample dataset, according to this result it has been seen that individuals generally influence their friends' decisions. But check-in distance is also an important criterion. Individuals check-in locations As the check-in locations of the individuals intensify, the intensity of the heat map is shown in red.

3.2 Recommendations

The prominence of spatial data is increasing day by day as individuals become mobile. With social network analysis, it is important to make large and complex data as meaningful data in today's fast-lived world. With the choice of open source software in analysis processes, it will provide widespread use in meaningful data production.

Firstly, the analysis will be made of the people who influence the groups obtained with modularity depending on the time and secondly, depending on the time between the groups, an analysis of the inter-group displacement of the individuals will be made are our future work.

REFERENCES

- Gao H., Tang J., Hu X., Liu H., 2015. Content-Aware Point of Interest Recommendation on Location-Based Social Networks. In: Twenty-Ninth AAAI Conference on Artificial Intelligence
- Cheng C., Yang H., King I., Lyu M. R., 2012. Fused Matrix Factorization with Geographical and Social Influence in Location-Based Social Networks. In: Twenty-Sixth AAAI Conference on Artificial Intelligence
- Leskovec J., Krevl A., 2014. SNAP Datasets : Stanford Large Network Dataset Collection, <http://snap.stanford.edu/data> (03 Aug. 2017)
- QGIS Development Team, 2009. QGIS Geographic Information System. Open Source Geospatial Foundation, <http://qgis.osgeo.org> (03 Aug. 2017)

Bastian M., Heymann S., Jacomy M., 2009. Gephi: An Open Source Software for Exploring and Manipulating Networks. In : *International AAAI Conference on Weblogs and Social Media*

V. K. Sihag et al., "Detecting community structure based on traffic at node in networks," in 2014 IEEE Students' Conference on Electrical, Electronics and Computer Science, 2014, pp. 1–9. DOI: 10.1109/SCEECs.2014.6804502

Jacomy M., Venturini T., Heymann S., Bastian M., 2014. ForceAtlas2, a Continuous Graph Layout Algorithm for Handy Network Visualization Designed for the Gephi Software. PLoS ONE 9(6): e98679. <https://doi.org/10.1371/journal.pone.0098679>