Segmentation of Social Network Users in Turkey

Mahmut Ali Özkuran, Gulgun Kayakutlu

HAL Id: hal-01950013

https://inria.hal.science/hal-01950013

Submitted on 10 Dec 2018

Segmentation of Social Network Users in Turkey

Mahmut Ali Özkuran[1], Gülgün Kayakutlu[2]

[1] Istanbul Technical University, Istanbul, Turkey
[2] Istanbul Technical University Industrial Engineering Department, Istanbul, Turkey

Abstract. In the digital world that we are living on companies needs as much as knowledge as they can collect to survive. Knowledge creation process requires as much as data to create more dimension on the knowledge. As the knowledge created with data in different dimensions it become more useful for the company. At this point, Knowledge Management helps creation of additional information sources to be used for knowledge accumulation.

In today's digital world social networks are fruitful data sources for nearly all of industries. Companies that utilizes this data source easily add another dimension to their information base and creates more revealing knowledge to their Knowledge Base. With automated processes, up-to-date information can be added to the Knowledge Base of the companies with well-known, easy to apply methods with little effort.

In this paper, we have conducted a research to reveal segments of the Turkish Twitter users using Self Organizing Maps method. Results show that, using segmentation we can create an important knowledge source about focus of interest, which could be used as a tool for analyzing the market penetration of the advertisements.

Keywords: segmentation, social networks, knowledge management, self-organizing maps.

1 Introduction

We are living in the knowledge era. Everything around us from smartphones to "Internet of Things" devices creates uncountable amount of data. According to IDC, size of the digital universe is doubling every two-year. [1]. Without organization of that much data, it is not possible to create information and knowledge that will help to create better processes for the companies. Knowledge Management offers different tools to convert this data for better utilization for the company. Those knowledge creation tools make Knowledge Management as one of the most important processes for the modern businesses.

Smart companies convert every step of their processes to the data, and using this data they create information and knowledge about those processes. Created knowledge and information helps with the optimization of every step of the company's commercial

effort from production to marketing. Especially optimization of the marketing opportunities is limitless if data of the customers can be converted into knowledge.

Usage of direct customer information collecting methods like Marketing Research (Qualitative or Quantitative) and Customer Relation Management have presents different dimensions of the customer knowledge. Adding new dimensions to this knowledge base would be help with the better understanding of the customers' needs and requirements.

As we are under the rain of data today, we have many different possibilities for adding this new dimension. As it offers relatively cheap data that varies over all segments of the society, social networks are very suitable tool for adding this new dimension.

From advertising to customer relation management, social networks are offering direct communication chance between customers and companies. Smart companies are utilizing this direct communication chance with their customers and make their products and processes better using those feedbacks. These optimizations made in the process and products also helps with the increases of their market share. In addition to the direct communication opportunity as an ocean of information, social networks give hints about many different properties of their users. Using these properties, also helps with selecting right audience for the targeted advertisement and opportunity of more concentrated connection with potential or existing customers of the products.

As we mentioned above social networks are widely used by any type of consumer, independent of revenue or belief. This variety in users makes social networks one of the important sources to discover potential markets. As there are many to learn from the social networks we should start with classification of the users of those networks and evaluate the potential marketing opportunities.

Some example research already made on social networks as an example of their productivity on creation of information. A sample work done by García-Palomares et al. shows some possible knowledge that can be learned from the social networks [2]. Researchers are classified cities using the GIS tags of photos shared on the social networks. According to the research every European metropolis have different tourist attraction. As some of the cities like Paris and London have dispersion in their touristic photo hotspots, other cities like Athens and Rotterdam has concentration of photos on some spots.

Today there are thousands of different sized social networks that servicing internet users. Twitter as one of the biggest social networks, (320 Million active users [3]) its data creates many possible areas for researchers to work on. From usage density of the certain words or tags to tweet count by time on any special event, their data is available to researchers. We have used this data source to find followers of the Turkish newspapers and other well-known Twitter accounts to classify Turkish twitter users by using segmentation.

The article organized that the next section is reserved for the literature review and the third section for the method. The application and results will be analyzed in the fourth section and conclusion will be given in the last section.

2 Literature Review

Many aspects of our lives have been digitized in the information era. This digitized data opened many avenues for scientific research. Especially public domain data that created by service users let researchers to created many new methods to work with.

2.1 Social Networks in Research

The Social network data presents limitless possibilities for the scientists to conduct a research on different fields from sociology to market research. For example, social networks are well known for their information exchange speed. This speed in information propagation inspires companies to create social networking tools that let their employees to share their knowledge. Behringer and Sassenberg made a cross-sectional study on the effects of different features on using social media as knowledge exchange [4]. But resistance against those new tools among the employees also affected the social media knowledge sharing applications.

Social networks are not only good for knowledge sharing but creating the knowledge. Companies are using social networks as data source for their strategic plans. Nguyen et al. conducted a research uses Chinese online sector data to examine the relationships between the effect of knowledge acquisition from social media, proactive and reactive market orientation, strategic capability over social media and brand innovation strategy. [5].

Usage of social networks are not only important for big companies but also important for small and middle sized enterprises[6] and individuals. Social networks are not only diffuse information between the users but in their interests too. A research made by D'Agostino et al. showed that parts that constitutes a social network is effected by its neighbors and trends [7].

Reciprocity is an extension of human behavior into the social networks. According to social exchange theory social behavior of a person is the result of an exchange process. Maximizing the benefit with minimized cost is the purpose of the social exchange. At this point we can say that online social networks are perfect for maximizing the profit with less effort. Reciprocity has important part of social exchanges as humans are tends to keep score, increase in the number of the reciprocity messages from the user, increases the number of reciprocity messages from his/her audience [8]. Usage of social networks is also related to other aspects human behavior like extroversion, conscientiousness, agreeableness, openness and emotional stability [9][10][11].

As behavior of individual social network users important to discover the dynamics of social networks they are still made with small samples compared to the times of Big Data. Today data of huge networks are used to find out the habits of users[12]. Not only the habits of users, also relations between users are important source of information. People uses micro blogging sites for following persons with similar interests or sources of information that interested in. This relation data between the users can be used for knowledge creation[13].

3 Methodology

In the fast times that we are living on, every information that created has an expiration date. To overcome information has propagation speed that never heard of before, we need to get use of information as fast possible to create difference from our competitors. With this background given we can say that it is important to create information from real time data with well-known data processing methodologies and automated processes.

3.1 Segmentation

Increase of the digital technologies, number of channels between customers and companies increased unprecedentedly. This increase in number of channels enforces companies to produce different marketing campaigns and products to reach out that many different markets. At this point segmentation of the customers (or even potential customers) become important for today's companies. Segmenting markets gives companies opportunity to create different marketing campaigns and products for different segments of the market.

A research made by Hamka et al. in 2014 used segmentation to differentiate mobile services users [14]. Classical segmentation of mobile device users is generally based on location of the calls or the length of the calls, but in this research, scientists gone one step further and measured the behaviors of users using a software. Researchers found that mobile service users can be segmented by usage of the network and the usage of the content services.

3.2 Self-Organizing Maps

Applications of social networking especially in the problems of coordination and cooperation make social network analysis embedded into many research fields. In any analysis, easiest representation of social networks is representing them as vertices (elements) and edges (relations). Prior to 1970's presentation of social network are made by sociograms [15] and modern graph theory [16]. From the 1970's, as data and processing power of the computers increased analysis and research of social networks has become hot topic.

Among all those advances in social networks research, Finnish professor Teuvo Kohonen introduced the Self-organizing maps in 1980s [17]. Self-organizing map is an artificial neural network that converts some input into a map (usually two dimensional) using unsupervised learning. Those maps are consisting of points called neurons. As name implies working model of neurons is inspired by the human brain data handling method.

At start weights neurons are initialized with random values. Following the initialization, competitive learning starts with supplying of test data consists of vectors to the network. With each new training vector supplied to the network, distance to all weight vectors is calculated and the most similar neuron to this input is called best matching unit. The best matching unit and neighbors of this neuron come closer to the

given vector. As supplied training vector count increases this adjustment amount toward given input decreases. Function that updating the neurons is given at formula **(1)**.

$$W_v(s+1) = W_v(s) + \alpha(s)(\theta(u,v,s)(T - W_v(s)) \tag{1}$$

Where **s** iteration number, **T** training vector, **W** nodes weight vector, **u** index of best matching neuron, **v** index of the neuron to be calculated, **θ(u,v,s)** distance function from best matching neuron to neuron **v** and **α(s)** as training function makes decreases the effect of training vector as supplied data count increases.

As it is a powerful tool for converting multidimensional data in to two dimensional maps self-organizing maps have wide usage area in different topics. Kohonen et al. mentioned about those different usage areas in their paper dated back to 1996. According to this paper, self-organizing maps can be used for feature detection in signal processing, fault detection in process and system analysis, in pattern recognition or even in robotics for navigation [18].

An example research by Bhandarkar et al. gives shows how to use self-organizing maps on image segmentation [19]. Researchers are used an improved version of the original self-organizing map, the hierarchical self-organizing map. This hierarchical self-organizing map is makes application of vector quantization on images possible, which leads to segmentation of the images.

Self-organizing map is also used in many researches related to market segmentation. A research by Bloom dated back to 2004 made market segmentation for the tourists using self-organizing map and backpropagation neural networks [20]. Bloom visualized the segmentation of the tourists visiting South Africa.

A recent research prepared from Iranian ADSL subscriber data also shows implementation of self-organizing map onto market segmentation [21]. The researchers are integrated Fuzzy Delphi method and self-organizing maps to visualize segments of the market.

4 Application and Results

We have made our research on Twitter data. Using Twitter API, we have downloaded IDs the followers of 74 selected Twitter accounts. Twitter accounts of selected Turkish Newspapers, Universities, Art Museums, Online Learning Platforms and Actors as initial research data. We have downloaded those IDs using C# and processed this data using Python and applied self-organizing map algorithms using R. SOM function from Kohonen package of R used for creating self-organizing maps.

Initially we have created main data consists of followers of 22 twitter account that belongs to 18 newspapers, 2 online newspapers and 2 television channels. Furthermore, follower data of those newspaper accounts mixed up with selected accounts related to

Education, Art and Cinema. Combined data used for creating different Self-Organizing Maps.

Figure 1. Training Progress
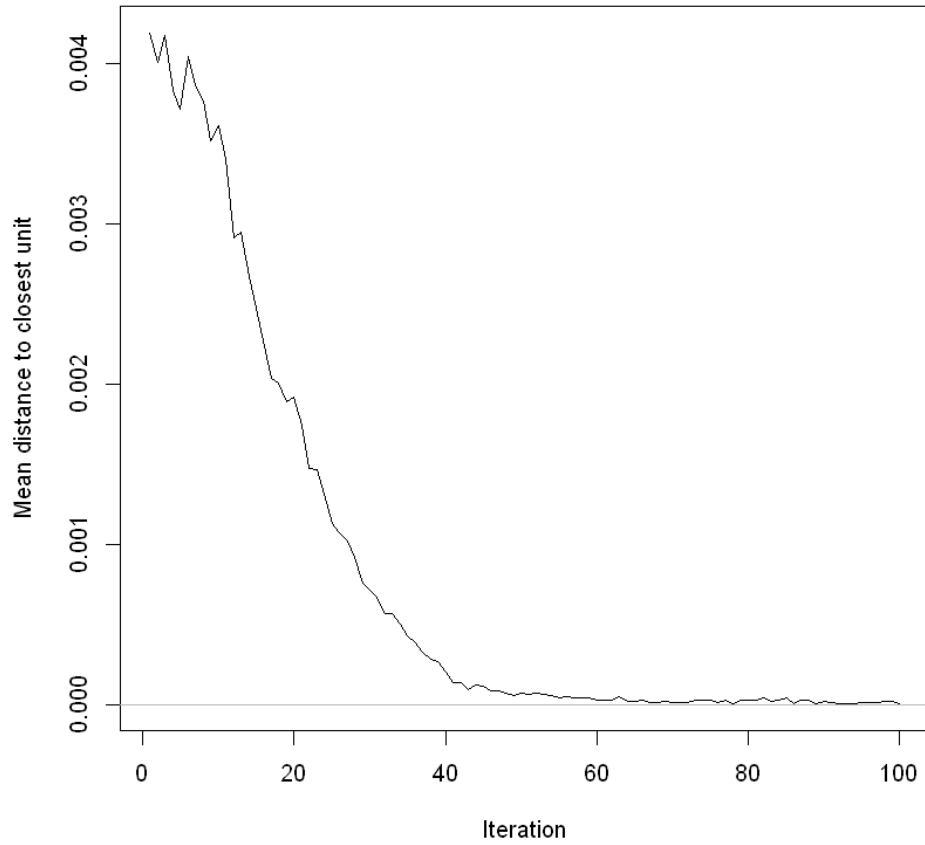
To simulate creation of different datasets for different information creation process we have combined different Twitter accounts to create 5 different combinations of data.

- Dataset 1 consists of the followers of top 5 bestselling Turkish newspapers Twitter accounts.
- Dataset 2 consists of the followers of the all Turkish newspapers Twitter accounts.
- Dataset 3 consists of the followers of the all Turkish newspapers and art related Twitter accounts

- Dataset 4 consists of the followers of the all Turkish newspapers and Cinema related Twitter accounts
- Dataset 5 consists of the followers of Cumhuriyet, Hurriyet, Radikal newspapers and well known universities Twitter accounts.

Application of self-organizing maps on to data has created datasets in size of 30 by 30. For each dataset, Self-Organizing Map algorithm is fits after 60 iterations as seen in Figure 1.

Figure 2 General overview of Dataset 1
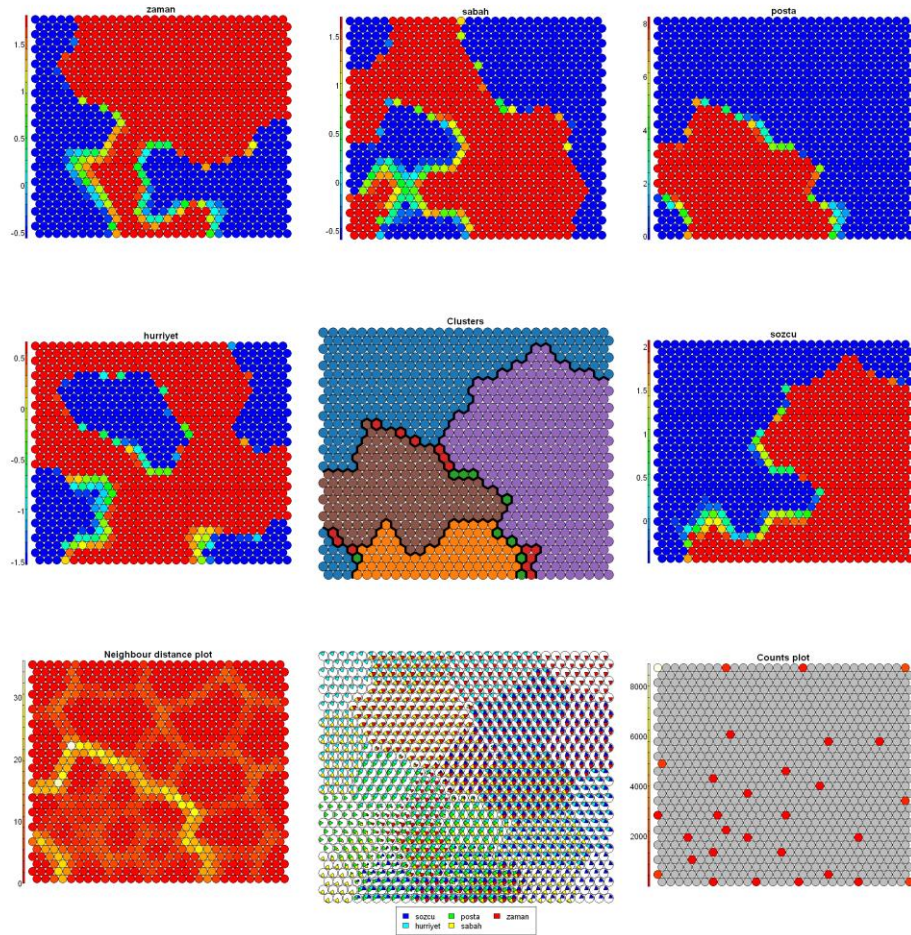
For visualizing results of the self-organizing map method data, we have created heat maps, neighboring distance maps, counts plot and cluster maps using those datasets.

General overview of visual results of research on Dataset 1 is shown in Figure 2.

Application of self-organizing map algorithm on to Dataset 1 undercover 4 big segments which are visualized in Figure 3.

**Figure 3** Clusters created by application of SOM on Dataset 1
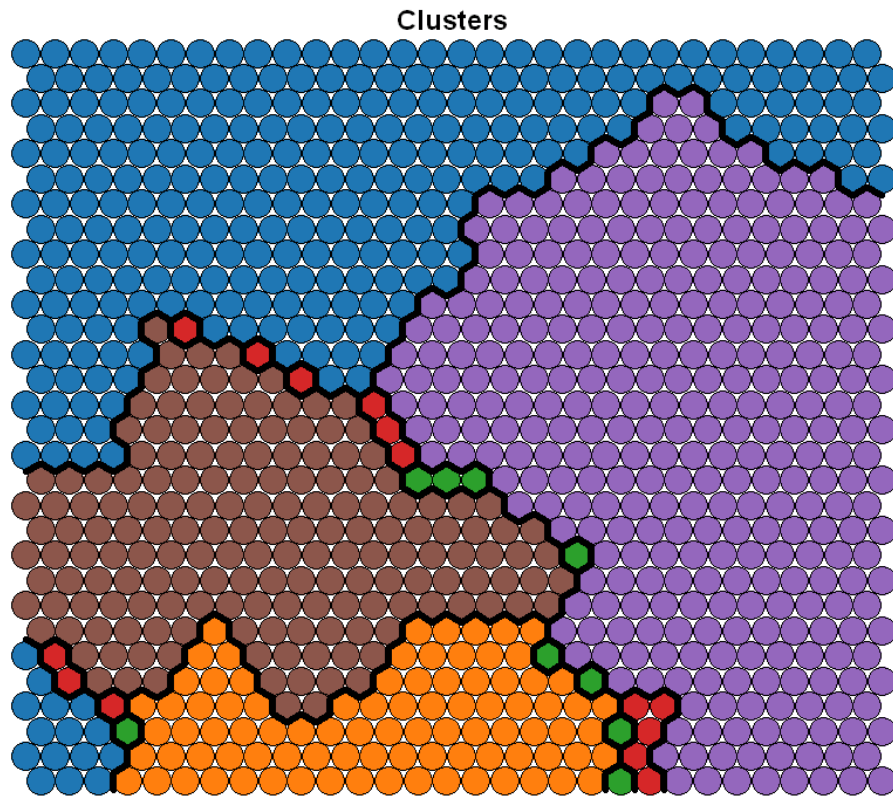
Followers of twitter accounts of mainstream newspapers like Hürriyet, Milliyet and Habertürk is scattered all around the map without creating big chunks of segments **(Figure 4)**.

Followers of twitter accounts of newspapers Akşam, Sabah, Vatan, Yeni Şafak, Star, Zaman and Habertürk are overlapping while Sabah and Zaman have broader follower base. **(Figure 5)**

Figure 4 Dispersion of Mainstream Paper Followers
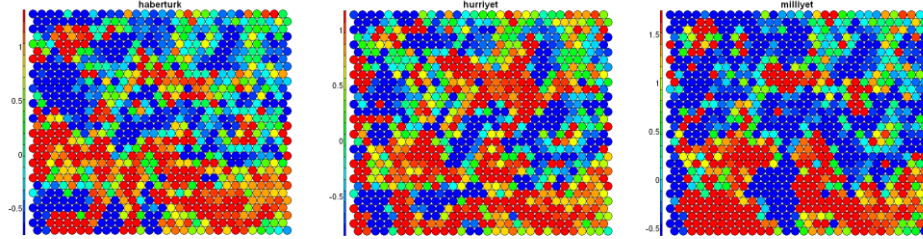
Figure 5 Overlapping Follower Bases of similar media companies - Akşam, Sabah, Vatan, Yeni Şafak, Star, Zaman and Habertürk.
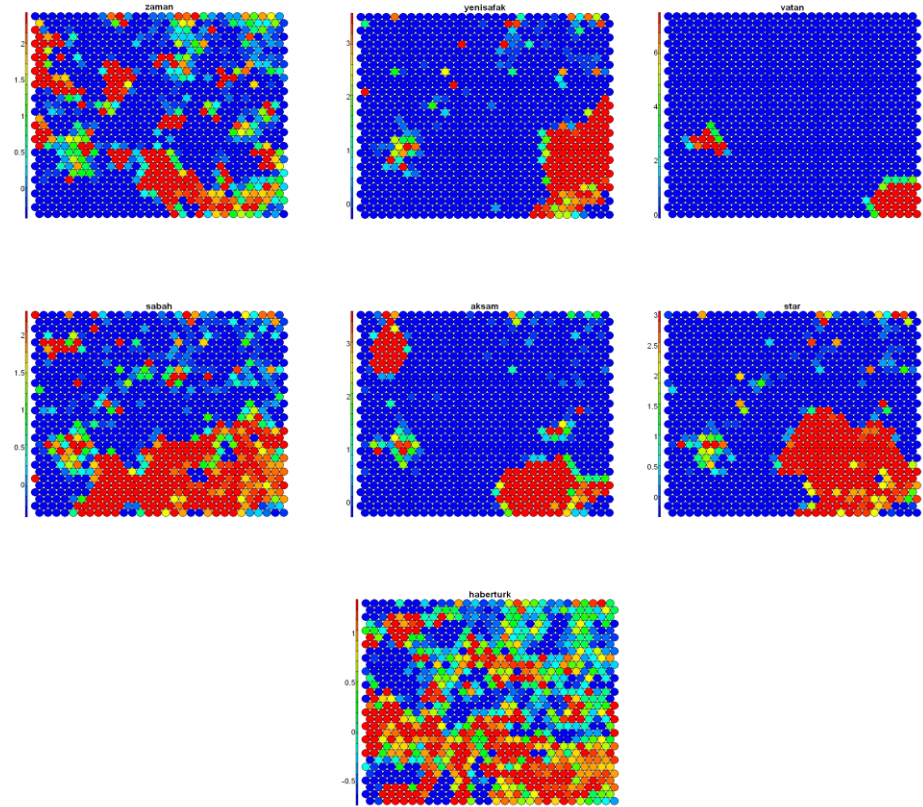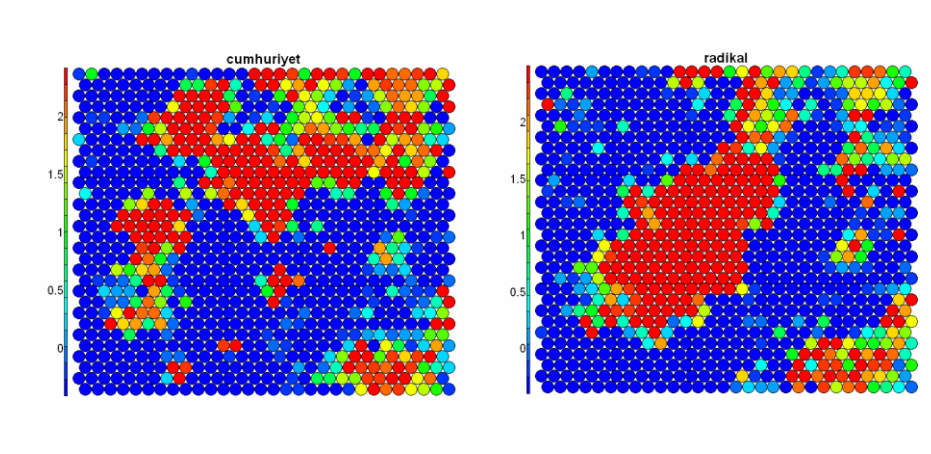
Followers of twitter accounts of newspapers Aydınlık, Cumhuriyet, Diken, Sözcü, Halk TV, Sol and Ulusal Kanal are overlapping while Cumhuriyet and Sözcü have broader follower base **(Figure 6)**.

Figure 6 Overlapping Follower Bases of similar media companies - Aydınlık, Cumhuriyet, Diken, Sözcü, Halk TV, Sol and Ulusal Kanal.
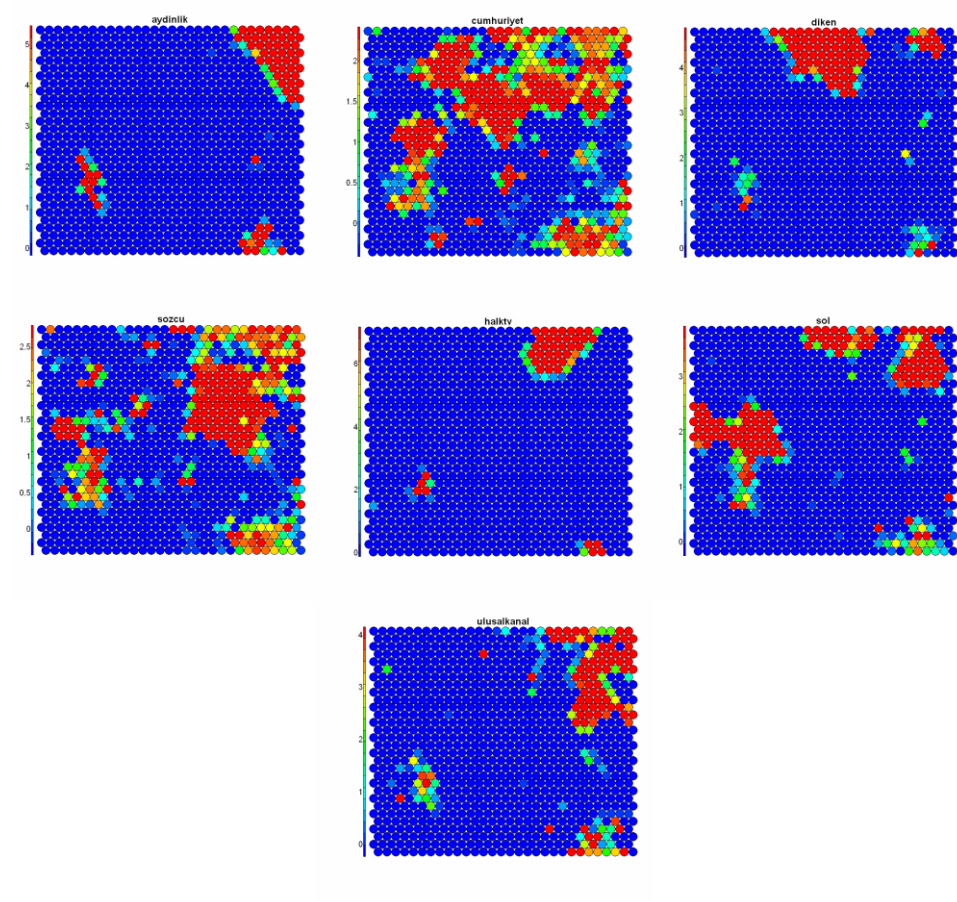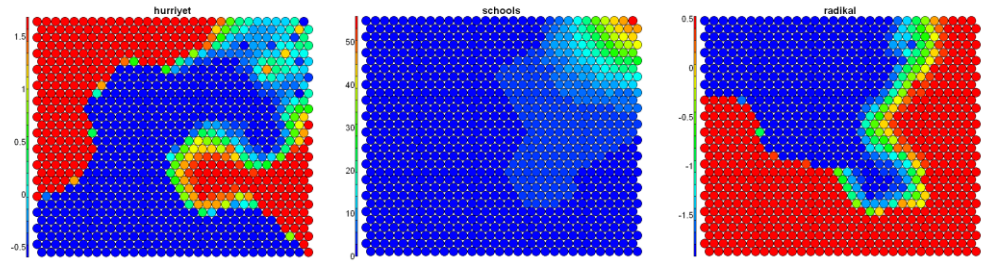

Radikal and Cumhuriyet twitter accounts have similar follower base while Radikal follower more packed up **(Figure 7)**.

Figure 7 Followers of twitter accounts of Cumhuriyet and Radikal

Followers of twitter accounts of Universities are overlapping the followers of Radikal. **(Figure 8).**

Figure 8 Followers of Radikal, Hurriyet and University twitter accounts

Followers of twitter accounts of Cinema related accounts are overlapping the followers of Radikal.

Results shows followers of the twitter account belong to Radikal newspaper are have a remarkable interest in universities and cinema. Consequently, Radikal newspaper seems have a good follower base for adverts of educational institutions and films.

For more solid forecasting for successful advertisement campaign we may use the numerical results of the self-organizing map method.

5 Conclusion

Social Networks are abundant data sources for researchers. Especially in the market research domain conducting researches using this data source can go beyond the classical scope of the market research. Researchers can identify the requirements of the certain segments of market using the knowledge created using information that distilled from social network data.

There are many algorithms and research methods are available to find the segments of any data. In this research, we have selected the self-organizing map method to find the segments of the social network users and visualize them. Application of self-organizing maps on social network data is productive in the means of knowledge creation. Using this data, we have created knowledge about general habits of followers of different Twitter accounts. For example, Turkish followers of the Twitter accounts of the universities are overlapping with followers of the Radikal Newspaper.

This research method also allows automated data creation, which may lead to real-time information retrieval. As we analyze the process of this research, it is possible to get real-time follower data from Twitter API to our data warehouse. Using data in this data warehouse we may run of segmentation scripts continuously and we may get up-to-date information about the segmentation of Twitter users with given inputs.

Several methods can be compared for virtual marketing; or analysis can be enriched by industry based specific subjects. Pointing the target audience is very important for both getting the right result and more effective marketing campaign. Using ids of the users already present in this data we may create tailor made offers for each and every customer.

For future research, we may improve those results with different research techniques like text mining and surveying for creating even more segments of Twitter users. Creation of time series that shows differences in segmentation could be also a good research opportunity. Also, our research resulted interesting group of individuals around the maps, researching those individuals with surveys may reveal focus segments.

References

[1]   V. Turner, J. F. Gantz, D. Reinsel, and S. Minton, "The Digital Universe of Opportunities: Rich Data and Increasing Value of the Internet of Things," *IDC White Paper*, 2014. [Online]. Available: https://www.emc.com/leadership/digital-universe/2014iview/index.htm?cmp=micro-.

[2]   J. C. García-Palomares, J. Gutiérrez, and C. Mínguez, "Identification of tourist hot spots based on social networks: A comparative analysis of European metropolises using photo-sharing services and GIS," *Appl. Geogr.*, vol. 63, pp. 408–417, 2015.

[3]   Twitter, "Twitter | About," 2016. [Online]. Available: https://about.twitter.com/company. [Accessed: 03-Mar-2016].

[4]   N. Behringer and K. Sassenberg, "Introducing social media for knowledge management: Determinants of employees' intentions to adopt new tools," *Comput. Human Behav.*, vol. 48, pp. 290–296, 2015.

[5]   B. Nguyen, X. Yu, T. C. Melewar, and J. Chen, "Brand innovation and social media: Knowledge acquisition from social media, market orientation, and the moderating role of social media strategic capability," *Ind. Mark. Manag.*, vol. 51, pp. 11–25, 2015.

[6]   S. Fernandes, A. Belo, and G. Castela, "Social network enterprise behaviors and patterns in SMEs: Lessons from a Portuguese local community centered around the tourism industry," *Technol. Soc.*, vol. 44, pp. 15–22, 2016.

[7]   G. D'Agostino, F. D'Antonio, A. De Nicola, and S. Tucci, "Interests diffusion in social networks," *Phys. A Stat. Mech. its Appl.*, vol. 436, pp. 443–461, 2015.

[8]   J. Surma, "Social exchange in online social networks. the reciprocity phenomenon on Facebook," *Comput. Commun.*, vol. 73, pp. 342–346, 2016.

[9]   C. Ross, E. S. Orr, M. Sisic, J. M. Arseneault, M. G. Simmering, and R. R. Orr, "Personality and motivations associated with Facebook use," *Comput. Human Behav.*, vol. 25, no. 2, pp. 578–586, 2009.

[10]   Y. Amichai-Hamburger and G. Vinitzky, "Social network use and personality," *Comput. Human Behav.*, vol. 26, no. 6, pp. 1289–1295, 2010.

[11]   K. Moore and J. C. McElroy, "The influence of personality on Facebook usage, wall postings, and regret," *Comput. Human Behav.*, vol. 28, no. 1, pp. 267–274, 2012.

[12]   Z. Wang, L. Tu, Z. Guo, L. T. Yang, and B. Huang, "Analysis of user behaviors by mining large network data sets," *Futur. Gener. Comput. Syst.*, vol. 37, pp. 429–437, 2014.

[13]   B. Liang, Y. Liu, M. Zhang, S. Ma, L. Ru, and K. Zhang, "Searching for people to follow in social networks," *Expert Syst. Appl.*, vol. 41, no. 16, pp. 7455–7465, 2014.

[14]   F. Hamka, H. Bouwman, M. De Reuver, and M. Kroesen, "Mobile customer segmentation based on smartphone measurement," *Telemat. Informatics*, vol. 31, no. 2, pp. 220–227, 2014.

[15]   J. L. Moreno, *Who Shall Survive? A New Approach to the Problem of Human Interrelations*, vol. 58. 1934.

[16]   F. Harary, R. Z. Norman, and D. Cartwright, *Structural models: an introduction to the theory of directed graphs*. Wiley, 1965.

[17]   T. Kohonen, "Self-organized formation of topologically correct feature maps," *Biol. Cybern.*, vol. 43, no. 1, pp. 59–69, 1982.

[18]   T. Kohonen, E. Oja, O. Simula, A. Visa, and J. Kangas, "Engineering applications of the self-organizing map," *Proc. IEEE*, vol. 84, no. 10, pp. 1358–1383, 1996.

[19]   S. M. Bhandarkar, J. Koh, and M. Suk, "Multiscale image segmentation using a hierarchical self-organizing map," *Neurocomputing*, vol. 14, no. 3, pp. 241–272, 1997.

[20]   J. Z. Bloom, "Tourist market segmentation with linear and non-linear techniques," *Tour. Manag.*, vol. 25, no. 6, pp. 723–733, 2004.

[21]   P. Hanafizadeh and M. Mirzazadeh, "Visualizing market segmentation using self-organizing maps and Fuzzy Delphi method - ADSL market of a telecommunication company," *Expert Syst. Appl.*, vol. 38, no. 1, pp. 198–205, 2011.