International Journal of Engineering & Technology, 7 (4.38) (2018) 939-943

International Journal of Engineering & Technology

Website: www.sciencepubco.com/index.php/IJET

Research paper



Review of Data Visualization for Social Media Postings

Nur Atiqah Sia Abdullah, Hamizah Binti Anuar

Faculty of Computer and Mathematical Sciences, Unversiti Teknologi MARA,
40450 Shah Alam, Selangor Darul Ehsan, Malaysia
**Corresponding author E-mail: atiqah@tmsk.uitm.edu.my,*

Abstract

Facebook and Twitter are the most popular social media platforms among netizen. People are now more aggressive to express their opinions, perceptions, and emotions through social media platforms. These massive data provide great value for the data analyst to understand patterns and emotions related to a certain issue. Mining the data needs techniques and time, therefore data visualization becomes trending in representing these types of information. This paper aims to review data visualization studies that involved data from social media postings. Past literature used node-link diagram, node-link tree, directed graph, line graph, heatmap, and stream graph to represent the data collected from the social media platforms. An analysis by comparing the social media data types, representation, and data visualization techniques is carried out based on the previous studies. This paper critically discussed the comparison and provides a suggestion for the suitability of data visualization based on the type of social media data in hand.

Keywords: *Data Visualization; Social Media; Perception; Emotion; Data Representation*.

1. Introduction

Advanced technologies and computers are gaining more computational power lead to high volumes of informative data, which can be collected at a high velocity [1]. IBM survey which involved more than 1,100 businesses and IT professionals showed that less than 26% of the respondents had experience on analyzing extremely unstructured data, where only 35% analyzed on streaming data [2]. In addition, from Intel IT manager survey, 33% of surveyed companies are working with very large amounts of data (500 TB and more); and 84% of IT managers analyzed unstructured data.

Social media services, such as Twitter or Facebook, have experienced huge growth rates [3] and there are good platforms for collecting and sharing information on users' interests. These platforms provide information such as what users are interested in, their shares such as following friends, like the comments or adding friends [4]. On top of that, netizens use these social network services daily in society, businesses and government aspects [5]. A person's emotions, such as anger, joy, and grief, often significantly impact his or her behavior and performance in the real world [6][7], especially when they are posting and sharing information in social media. Hence, understanding a person's emotions and the patterns of these emotions provide great value for both individuals and businesses [8].

Most of the previous studies have their own visual analytics systems [9][10][11][12][13] to visualize the emotions through social media. Their visualization techniques are varied based on the type of data and domain. For example, node-link diagram is used to represent barrier of works and emotions [9]; relationship between members and groups in Faculty of Art are characterized by using node-link diagram and directed graph [10]; line graph is used in representing Twitter datasets and network views [11]; Heatmap is used to display the normal and depressed users with the several of behaviour identity features [12]; topic frequency such as current and new topics can be illustrated by using a stream graph [13]. Therefore, different visualization techniques are used for representing different types of data.

This paper aims to compare these studies and suggest the data visualization technique based on the type of social media data in hand. This paper continues with Section II, which consists of related works to provide a holistic overview of data visualization, social media, and past research on visualization in social media posting. Section III discusses the previous studies in term of social media data types, data representation, and data visualization techniques before concludes the findings in Section IV.

2. Related Work

2.1. Data Visualization

A visualization is a form of visual communication, which is the process of transforming a non-visual data into artifacts that are accessible to the human mind [14]. The examples of visualization tasks for large data [15] include visualization of data streams; visual data mining; visual search and recommendations; big data storytelling using visualization; scalable methods for parallel imaging; modern hardware and architecture for data visualization and analysis; human-computer interface and visualization of large data; visualization of large data applications including cyber intelligence and counter-intelligence, business intelligence, electronic commerce analysis of scientific information, education, visual noise, large image perception, information loss, high performance requirements and high rate of image change.

This field of information visualization has been rapidly achieved presently. Therefore, many of decision makers implement various visualization techniques and applications to make their decisions in all kinds of subject domains such as digital library management; personal information assistants, historical data management; information hierarchy visualization, and many more [16]. The deci-

Copyright © 2018 Authors. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

sion makers believe that the solution of visualization can provide higher intelligence density in order for them to make a better decision with less effort despite the complexity and high occurrence in certain issues [16].

2.2. Social Media

Social media is one of the main sources generating large, unstructured, and redundant big data [9]. Social networking is a subcategory of social media where it is a platform connecting people to each other and shares information, opinions, and experiences. The increasing of mobile technologies has made people even easier to use social media to share their opinions, announcements, information, and behavior. Usually, they share information on their feeling and experiences [17]. Besides, it is also an important platform for government to update the information and interact with citizens [9].

There are millions of active users of many social networking sites (SNS) who publish content on a regular basis. For example, Twitter claims a daily tweet-publishing rate of 250 million [18]. SNS continues to transform to the public discourse and set trends and agendas in most facets of society [18][19]. In addition, there is an open forum which gives user opportunity to broadcast their thoughts, opinions, judgments, and experiences [19].

As social media contains a huge of data, it is difficult to analyze the behavior of the individuals related to the information and data they posted. Therefore, some researchers have studied moods and emotion by analyzing Twitter. For example, the temporal patterns of happiness by observing the data set of tweets posted over the duration of 33 months [20], and the correlations between expressions on Twitter across the United State and a range of emotional, geographic, demographic, and health characteristics [21].

2.3. Visualization in Social Media Posting

Over the past decades, some researchers have developed sophisticated techniques for analyzing and visualizing network data, mainly social network data [10]. This section reviews past researches on visualization techniques. Community emotion visualization [9] used Java Applet to produce a network graph to represent the interactions of online communities. This visualization enables an analyst to explore and understand the quantitative and qualitative features of the underlying data.

The main characteristic of social network visualization lies in to capture the interaction between members by measuring the interaction frequency. One of the challenges is to visualize effectively a large-scale annotated data poses in terms of the presentation of informative layout between people and the annotations result from their posts [9].

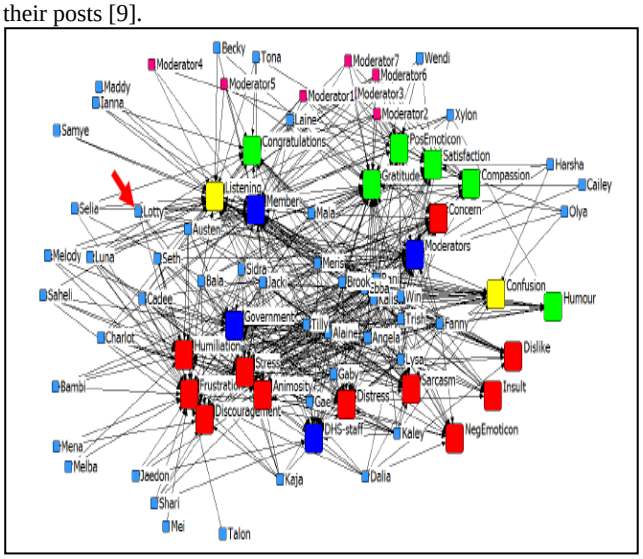


Figure 1. Community emotion visualization

Based on Fig. 1 [9], it visualized the expression of emotions by the community members with different colors (green, yellow, and red) to represent three types of emotions (positive, neutral, and negative emotions). The dark blue nodes (in big squares) represent the government staffs, pink squares for the moderators, and light blue squares for the community members. From the visualization, the members and moderators are in positive emotions, whereas government staffs are in negative emotions. It shows that the members of this community have negative expression because there are many red color nodes towards government staffs. However, they posed their emotions positively about the moderators and other community members.

Besides, the types of work barriers that shared by the community members are visualized using different size of nodes [9]. The larger nodes describe the specific barriers of work, for example, Transport, Flexibility, Job Market, Stereotypes, and Cost of Education. The small nodes represent community members. Linked nodes indicate the community members share same work barriers as in Fig. 2.

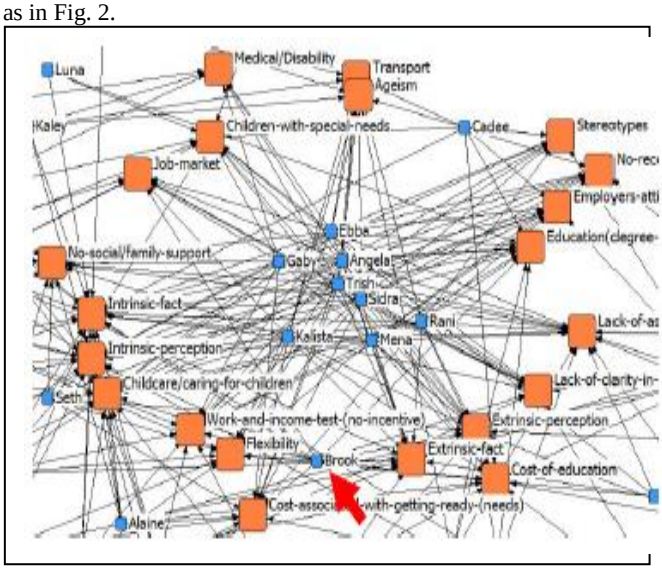


Figure 2. Community barriers to work layout

Social Network Analysis (SNA) usually examines social interaction in computer-mediated environments; identify unique social roles; important individuals; social structures; and dissemination patterns [10]. NodeXL SNA tool is used to produce a bimodal graph [10], which has two different nodes to represent community members and forums they have posted on social media as shown in Fig. 3.

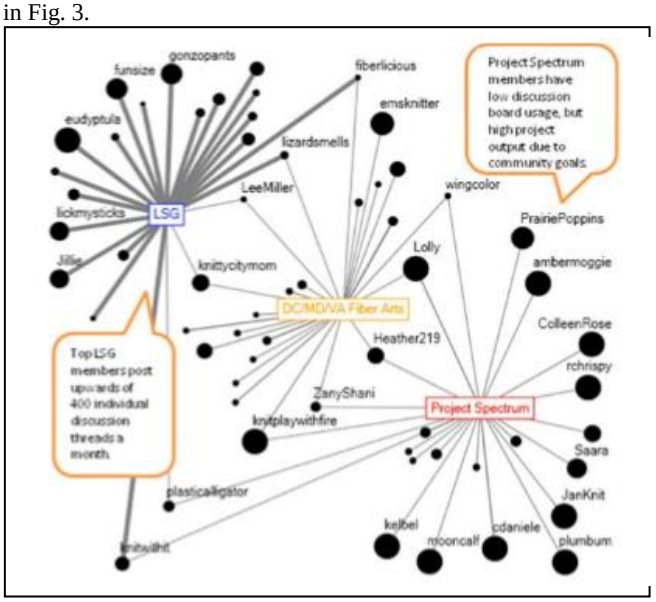


Figure 3. Bimodal graph

The data is differentiated by using visual properties of nodes, such as color and size, to represent important social roles, such as a network administrator or important community members, and corresponding SNA [10]. They visualized the data using node-link tree technique, where the circle shape represents the community members, and the rectangle represents the subgroup. The node size characterized the number of craft projects completed whereas the edge width indicates the number of posts to the subgroup forum. To indicate the strength of a tie between entities and to represent different types of connections, the visual properties of edges such as line thickness or through different colors were used [10]. A directed graph (as in Fig. 4) can be used to represent community member discussions, which illustrates the link between the fashion designers [10]. The size of nodes indicates the degree of node centrality and edge width specifies the tie strength.

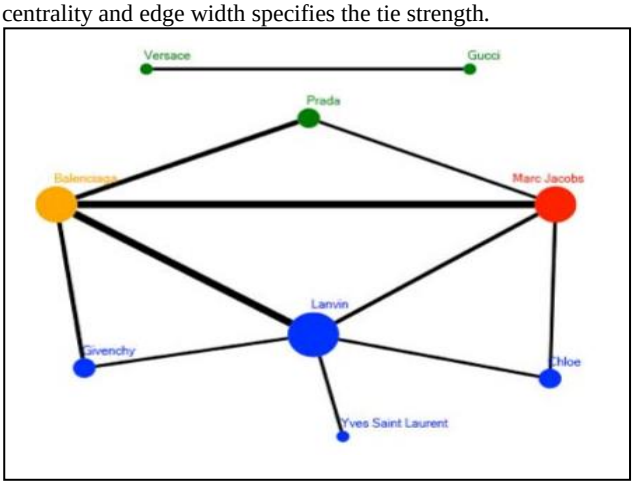


Figure 4. Directed graph

Social networks usually allow multiple views or relations that share maybe all or just a part of the same user set. Generally, these views involve graphs with a variety of edges where each type has a different frequency or weight distribution [11].

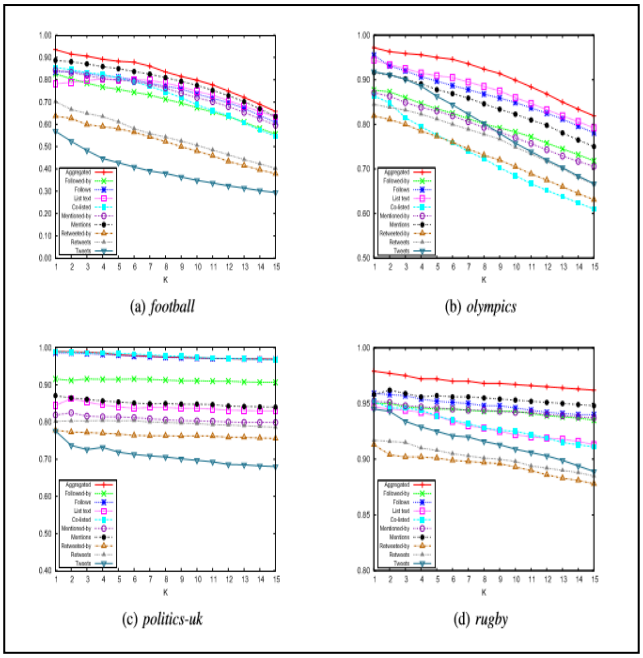


Figure 5. Unified graph representations

Based on the case of analyzed tweets in Fig. 5 [11], they characterized the users by the accounts whom they follow (or who follow them), the users whom they retweet (or who retweet them), and curated lists to which they have been assigned, and so on. An integrated multiple data view is used to provide unified graph representation, which recollects the most informative connections from the original views [11]. These line graphs used four Twitter datasets, which are football, Olympics, politics-UK, and rugby. The network views are based on ten types of Twitter's features, which aggregated, list text, co-listed, tweets, follows, followed-by, mentions, mentioned by, retweets, retweeted-by. The data is differentiated by using different colors and shapes. They plotted the graph based on the application of Singular Value Decomposition (SVD) rank aggregation.

The increasing trend of online interaction through various platforms of social media such as Facebook and Twitter can have an extreme influence on various mental and behavioral health outcomes such as depression, stress, cyberbullying and violence [12]. A heatmap is constructed to distinguish behavioral features of depressed users from non-depressed users (see to Fig. 6).

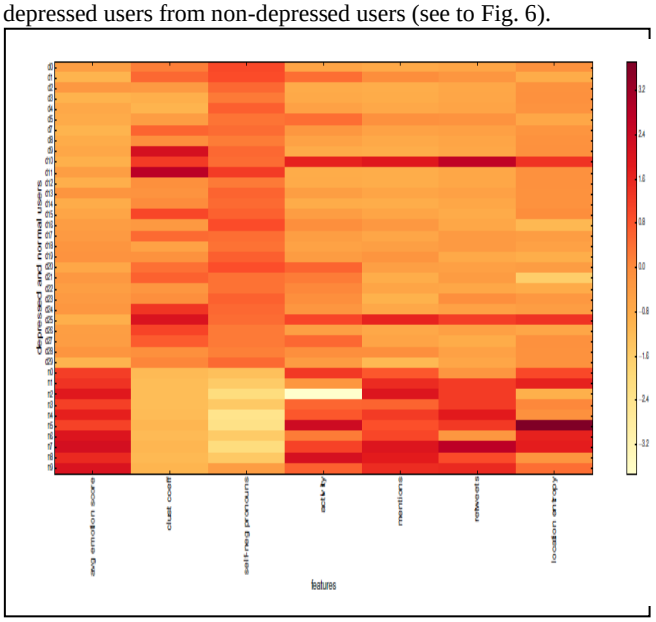


Figure 6. Heatmap visualization

Y-axis shows a sample of users who are depressed and normal classes. On the x-axis, it shows the contextual, linguistic and structural, and network engagement based on features representation to their behavioral identity on social media such as average textual emotion expressed, user clustering coefficient with respect to their one-hop network, linguistic style (the proportion of self and negativity related pronoun usage in their text), average user activity (number of posts), average number of mentions and retweets received by the user and the regional entropy of their ego-network [12]. All these features are indicated by the different type of color bars. The color range indicates normal or depressed user where the higher or lower features based on their behavioral identity on social media.

At present, Twitter is one of the most popular microblogging communities used by people to give comments and impressions during political speeches, sports competitions, natural disasters, and other large events [13]. When more people have exchanged the comments, impressions, and suggestions during events, digital backchannel becomes an interesting communication medium [13]. Thus, a Visual Backchannel [13] is an evolving, interactive, and multi-faceted interface that combine three visualizations with a list of backchannel posts via linking, brushing and filtering [13].

Stream graph is built to visualize a visual backchannel (see Fig. 7). They used the different type of data, which are time and topic. The data is denoted by different color, shape, and label in the stream graph. The x-axis of the graph shows the time dimension from past until present. The blue color represents the newest topic while the green color shows the oldest topic. The y-axis signifies the relative frequency of topics corresponding to the topic at a particular time. The stream graph displays 30 most popular topics and ignores the less popular topics in order to reduce drawing complexity and ensure interactive responsiveness.

In addition, the topic is labeled by the most used word instance of the particular stem. A transition like a zoom is used to allow the viewers to set their time period to represent a wide spectrum of events. A viewer also can click on the corresponding button above the Topic Streams, which is updated through animated transitions when there is a transmitting data for the new time window of the topic visualization to the browser [13].



Figure 7. Stack graph of topic stream

3. Comparison and Discussion

This section compares the past studies on visualization of social media postings. It discusses based on classification category, representation, size of nodes, and finally suggestion on the selection of data visualization techniques. The past studies show that there can be many classifications of issues based on the collected data. For example, the data are classified based on work barriers, groups in the workspace, network views, behavior identity features, and discussion topics. These classifications are different based on the purpose of the studies. Therefore, the classification of data has to be pre-defined before collecting the data. These classifications are set as the theme for the analysis of data. Then the data is categorized according to the pre-defined classification.

For the representation, most of the researchers [9][11][12]][13] used colors to indicate types of data. The use of colors is depending on the number of data used. For example, three colors (green, red, and yellow) are used to represent positive, negative and neutral emotion [9]; while for the barriers of works [9] use one color only, which is orange. The same color is also can be used if the category can be differentiated by the size of nodes, as in [11]. A range of colors is normally applied in heatmap and stream graph [11][13]. Different color spectrum is normally used to differentiate the behavioral features [12] and time-series topics [13]. Node-link tree uses different sizes of nodes to represent the amount of data rather than using colors [10].

Size is normally used to differentiate the amount of data. If the amount of data is huge, the bigger size of shape is used. Instead of using colors, shapes are the most relevant to show the patterns of the data as used in [9][10]11][13] to show their data and domain. There are different types of shapes such as a rectangle, circle, star, triangle, cross shape, hexagon, and etc. When dealing with different categories, more than one shape is used to differentiate the types of categories as in [11]. When the number of the category is small, only one shape is used. The data visualizations also utilize textual labels to express data, as in [9][10][13].

The selection of data visualization techniques is normally depending on the types of social media data. Node-link diagram uses color and shapes to differentiate emotions and barriers of work [9]. Node-link tree uses two types of nodes to represent community members and forums [10]. A line graph is simple but enough to show ten network views that represented by different colors and shapes for four types of tweets datasets [11]. Heatmap applies color range to characterize the behavioral related data [12]. Different colors range represents the normal and depressed user based on the behavior identity features [12]. Stream graph uses colors, shapes, and labels to show the topic frequency in Twitter based on the time dimension. Table 1 shows the comparison of data visualization in social media postings. From this comparison, a network diagram is the most common category of data visualization used to show the relationships between nodes in social media postings.

Table 1. Summary of Data Visualization in Social Media Postings

Research	Type of Social Media Data	Representation	Data Visualization Techniques
[9]	- Emotion (Positive, Negative, Neutral) - Barriers of Work (Job market, Processes & policies, Cultural, Life situation, Logistics, Financial, Intrinsic, Extrinsic)	Color Label Shape	Node-link diagram
[10]	- Community members (75 members) - Groups of forums (LSG, Fiber Arts, Project Spectrum)	Shape Label Node Size	Node-link tree Directed Graph
[11]	- Twitter datasets (football, Olympics, politics-UK, rugby) - Network views (aggregated, list text, co-listed, tweets, follows, followed-by, mentions, mentioned-by, retweets, re-tweeted-by)	Shape Color	Line Graph
[12]	- Users (Normal & Depressed) - Behavior identity features (emotion, clustering coefficient, self-negative pronouns, user activity, mentions, retweets and location entropy)	Color	Heatmap
[13]	- Time dimension (15 intervals – from past to currently) - Topic frequency (new topics, movie titles, artist name, current topics)	Color Shape Label	Stream graph

4. Conclusion

Data visualization is a solution that saves people time to get the information and an easy technique to come out with the decision through analyzing the visual. Social media is a podium that consists of huge data. Based on the data in social media, it can be visualized by using a variety of data visualization techniques to explore the information that people want to look into. However, it depends on the data they want to study. This paper has reviewed different types of visualization techniques in representing social media data. Node link-diagram, node-link tree, directed graph, line graph, heatmap and stream graph are among the visualization techniques used to represent the social media data. To make the decision on what type of data visualization technique is suitable for the collected social media data, one has to process the social media data according to the pre-defined categories and classify the data. To decide on the representation of data, one can consider different colors and sizes of nodes to differentiate the category of data. If the data is huge, organize it into a different level of representations, and use transitions such as zoom, click, hover over to provide extra information. Despite that, different node sizes can be used to differentiate the amount of data in visualization. Bigger nodes will indicate a larger amount of data. Lastly, a network diagram is the most common category of data visualization used to show the relationships between nodes in social media postings.

Acknowledgements

The authors express their appreciation to Faculty of Computer and Mathematical Sciences and Universiti Teknologi MARA, Malaysia for sponsoring this paper.

References

[1] C. Leung, V. V. Kononov, A. G. M. Pazdor and F. Jiang. PyramidViz: Visual Analytics and Big Data Visualization for Frequent Patterns. Proceedings of the 2nd IEEE International Conference on Big Data Intelligence and Computing, (2016) August 8-12; Auckland, New Zealand

[2] Intel Corporation, Big Data Visualization: Turning Big Data Into Big Insights, White Paper, March (2013)

[3] W. Li, C. Eickhoff and A. P. Vries. Interactive summarization of social media. Proceedings of the 5th Information Interaction in Context Symposium, (2014) August 26-30; Regensburg, Germany

[4] F. X. Chao, J. Xu, and C. W. Lin. Mining user interests from social media by fusing textual and visual features. Proceedings of Asia-Pacific Signal and Information Processing Association Annual Summit and Conference, (2016) December 13-16; Jeju, South Korea

[5] P. Wang, M. Gao, and R. O. Sinnott. Enhanced Gender Identification through Social Media Analytics on the Cloud. Proceedings of IEEE 2nd International Conference on Data Science and Systems, (2016) December 12-14; Sydney, Australia

[6] R. J. Davidsonand S. Begley, The Emotional Life of Your Brain: How Its Unique Patterns Affect the Way You Think, Feel, and Live -- and How You Can Change Them, Hodder and Stougton Ltd, New York (2012)

[7] G. Loewenstein and J. S. Lerner, The Role of Affect in Decision Making, in Handbook of Affective Scienc, Edited R. Davidson, H. Goldsmith and K. Scherer, Oxford University Press, Oxford (2003), pp. 619-642.

[8] J. Zhao, L. Gou, F. Wang and M. Zhou. PEARL: An interactive visual analytic tool for understanding personal emotion style derived from social media. Proceedings of IEEE Conference on Visual Analytics Science and Technology, (2014) October 25-31; Paris, France

[9] S. K. Bista, S. Nepal and C. Paris. Multifaceted Visualisation of Annotated Social Media Data. Proceedings of IEEE International Congress on Big Data, (2014) 27 June- 2 July; Anchorage, USA

[10] D. L. Hansen, D. Rotman, E. Bonsignore, N. Milic-Frayling, E. M. Rodrigues, M. Smith and B. Shneiderman. Do You Know the Way to SNA?: A Process Model for Analyzing and Visualizing Social Media Network Data. Proceedings of International Conference on Social Informatics, (2012) December 14-16; Alexandria, USA

[11] D. Greene and P. Cunningham. Producing a unified graph representation from multiple social network views. Proceedings of the 5th Annual ACM Web Science Conference, (2013) May 2-4; Paris, France

[12] N. Vedula and S. Parthasarathy. Emotional and Linguistic Cues of Depression from Social Media. Proceedings of International Conference on Digital Health, (2017) July2-9; London, United Kingdom

[13] M. Dork, D. Gruen, C. Williamson and S. Carpendale, A Visual Backchannel for Large-Scale Events, IEEE Transactions on Visualization and Computer Graphics. 16, 6 (2010)

[14] S. Dubel, M. Rohlig, H. Schumann and M. Trapp. 2D and 3D presentation of spatial data: A systematic review. Proceedings of IEEE VIS International Workshop on 3DVis, (2010) November 9; Paris, France

[15] T. Soklakova, A. Ziarmand and S. Osadchyieva. Big data visualization in smart cyber university. Proceedings of the IEEE East-West Design & Test Symposium, (2016) October 14-17; Yerevan, Armenia

[16] X. Bai, D. White and D. Sundaram. Visual intelligence density. Proceedings of the 10th International Conference NZ Chapter of the ACMs Special Interest Group on Human-Computer Interaction, (2009) July 6-7; Auckland, New Zealand

[17] C. Paris, H. Christensen, P. Batterham, and B. O'Dea. Exploring Emotions in Social Media. Proceedings of IEEE Conference on Collaboration and Internet Computing, (2015) October 27-30; Hangzhou, China

[18] S. Kumar, F. Morstatter, G. Marshall, H. Liu and U. Nambiar. Navigating information facets on twitter. Proceedings of the 18th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, (2012) August 12-16; Beijing, China

[19] C. A.Steed, M. Drouhard, J. Beaver, J. Pyle and P. L. Bogen. Matisse: A visual analytics system for exploring emotion trends in social media text streams. Proceedings of IEEE International Conference on Big Data, (2015) October 29-November 1; Santa Clara, USA

[20] P. S. Dodds, K. D. Harris, I. M. Kloumann, C. A. Bliss and C. M. Danforth, Temporal Patterns of Happiness and Information in a Global Social Network: Hedonometrics and Twitter. . J. PLOS ONE. 6, 12 (2011)

[21] L. Mitchell, M. R. Frank, K. D. Harris, P. S. Dodds and C. M. Danforth, The Geography of Happiness: Connecting Twitter Sentiment and Expression, Demographics, and Objective Characteristics of Place. J. PLOS ONE. 8, 5 (2013)