

# Data Visualisation - Assignment 1 Report

## Visualising Big Social Data

Submitted by: Kalyani Prashant Kawale

Student ID: 21237189

### 1 Introduction

Every day we make use of some or other social media channel such as Twitter, Instagram, YouTube, LinkedIn; for work, for leisure, for different reasons. The amount of social media activities carried out in a single minute across the globe is incomprehensible and so is the data generated from these activities. The data thus generated is termed as social data which encompasses the properties of big data, being huge in volume, appearing in variety of forms, rapidly changing with a high velocity, having the potential of being translated into value [1].

The translation of the social data into value can be achieved in many forms. While, most of these forms are commercial leading to monetary benefits, such as gaining insights into the spending habits of a customer and exploiting that information to get more business out of that customer, some forms are for social benefit as well, like using social media data during pandemic to get quick updates on the spread of the disease, understand trends, connect and convey important information. However, just the availability of the social data is not enough to gain value out of it, the data also needs to be processed and analysed before it starts reaping benefits.

The techniques and technologies available for analysing social data are immense and analysis through visualisations is one of them. Due to the inherent network structure that social media has, Social Network Analysis has a particular and very important applicability in analysing social media data. Multiple studies have researched the use of visualisations in social network analysis from various perspectives. Different types of social networks, their visualisations and the value that can be gained from these visualisations has been discussed thoroughly in [6]. [3] reports an interesting experiment, which highlights the importance of visualisation in social network analysis by conducting a study of non-technical students who learn to apply the concepts of social network analysis. The usefulness of data visualisations in social data analysis is, however, not limited to network visualisations but includes a varied range of other techniques which are reported in detail by [7].

### 2 Visual Analytics

Common analyses carried out on the data collected from social media platforms are sentiment analysis for under-

standing the reaction of consumers towards a certain product or service, mining the data to create similarity matrices to be used in recommendation systems, opinion mining for tasks such as detecting hate speech, and so on. Machine learning is a key technology which processes the data and enables data analysis. Visualisations on the other hand have the power of providing a visual means for representing the data or the results of the analysis, which can then be used and understood by general users. The biggest challenge with visualisation of social data is the sheer size of it all. Machine learning techniques such as principal component analysis can reduce the size of this data and visualisations can then be applied to make sense of the data. Moreover, interactive visualisations such as graphs that can be zoomed in, filters that could change the view of the visualisation, tooltips on charts that display detailed information, etc. can aid the social data visualisations. This combination of machine learning and visualisation techniques working hand in hand is described by [7] as visual analytics.

[7] also suggest that the analysis of the entire social data performed solely by autonomous means is quite impractical and the data needs to be explored even before analysis methods are applied by users who can greatly benefit from visualisations. The importance of exploring the data through visualisations before defining the goal of analysis has also been emphasized by [3] based on their study of students who developed graphical representations using social network analysis methods. It was reported by [3] that only after creating an initial graphical visualisation of the data were most of the students able to understand what could be achieved from it.

### 3 Social Networks as Graphs

The social media data comes in many forms, mainly text, images, videos and location [7]. The peculiarity of the social data is that most of it can be structured as a network and the most common way to visualise a network is through a graph. The simplest example of social media data translated into network would be where the users on platform such as Twitter are mapped into nodes and their connections to other accounts are mapped into links or edges in the graph.

The aesthetics involved in creating graphs for social data would typically be the size and shape of each node which may vary depending upon some criteria that depicts the importance of the node, lines with different weights indi-

cating the strength of the connection, and colour used to differentiate between clusters or communities in the network. As such, [3] reported that use of similar aesthetics for encoding the variables was seen among the visualisations created in the study they conducted. Along with exploring the challenges faced by non-technical people in applying social network analysis methods and visualising a graph, [3] also listed the NetViz Nirvana guidelines [2], which can help achieve quality network visualisations. The guidelines [2], include the suggestions that every node should be visible, the degree of every node should be visible enough that it can be counted, the path of every edge should be visible to be followed from source to destination, and clusters in the network and any outliers should be identifiable. The study presented by [3] suggests that using network visualisation can lead to far easier understanding of the network properties and make relationships in the data visible for further analysis.

The NodeXL [4] software program has been a prominent choice for creating the network visualisations and performing analysis in all of the reviewed studies. [6] details different types of networks created using NodeXL; the examples include Polarized networks, In-Group networks, Twitter Brand Map, Bazaar network, Broadcast networks and Support networks. Each one of the stated networks can be visualised and provide an effective tool for either analysis or presentation of social media data for different purposes.

## 4 Beyond Graphs

While the networks created from social media data visualised as graphs provide a proficient means of collecting and representing information, and understanding the relationships among various entities, there are many other visualisations that could provide just as efficient outcomes which have been described in detail with examples by [7]. Recently, different visualisations such as bar charts with 7-day moving average, stacked charts, choropleths, heatmaps were heavily used to follow the trends of Covid-19. Such visualisations can be connected to real time data streams and updated accordingly, this behaviour of dynamically adopting to the change in background knowledge has been highlighted by [8] as well. Some of the more specific examples of social data visualisations examined by [7], includes a temporally adjusted stacked graph of topics used to monitor the ever-changing discussions on online forums, heatmaps drawn on top of satellite images highlighting the places in a city where high percentage of photographs were taken indicating that these were the travel hotspots, flow maps of city indicating popular routes which could be used for traffic congestion analysis, parallel-sets depicting the most important topics extracted from news articles over a fixed period of time.

An approach for visualising unstructured textual data obtained from sources such as tweets, comments, and reviews, called SentenTree has been introduced by [5] which adopts from the word cloud visualisation for displaying the important words in bigger fonts, and extends it by adding interactive features that let the viewer ex-

plore the context of the words in the form of a node-link diagram. The SentenTree visualisation tries to tackle the common issues of no flow, lack of context and no layout involved in other text visualisations [5].

## 5 Conclusion

The most popular and traditional choice of visualising social media data is a network graph which makes sense for most of the social data. Through these network visualisations the process of social network analysis gets simplified which provides valuable insights about the entities and their relationships in the network. This information can be used for targeting and improving products and services based on the information gained through the network analysis. The effectiveness of using visualisation to represent a graph is evident in the fact that a viewer can easily and immediately identify group of similar entities differentiated using different colours, compared to the time it would take for the same viewer to understand the groups by looking at a data table of thousands of entries. The dimensionality is definitely a challenge in visualising the graphs, but it could be made easier by applying machine learning techniques to the data.

The visualisations that can be applied to social data are not limited to network graphs however, and dozens of alternatives exist that may be applied depending on the context and goal of the analysis being carried out. A general theme identified in the social data visualisations was that of the use of interactive features. Data visualisations are instrumental tools that greatly help the social media data analysis process, be it during the pre-processing phase or during the presentation of the results [7].

## References

- [1] Big data. Available at: [https://en.wikipedia.org/wiki/Big\\_data](https://en.wikipedia.org/wiki/Big_data) [Accessed on: 10/04/2022]
- [2] Dunne, Cody & Shneiderman, Ben. (2009). Improving graph drawing readability by incorporating readability metrics: A software tool for network analysts. HCIL Technical Report, College Park, MD: University of Maryland.
- [3] Hansen, Derek & Rotman, Dana & Bonsignore, Elizabeth & Milic-Frayling, Natasa & Mendes Rodrigues, Eduarda & Smith, Marc & Shneiderman, Ben. (2012). Do You Know the Way to SNA?: Process Model for Analyzing and Visualizing Social Media Network Data. 304-313. 10.1109/SocialInformatics.2012.26.
- [4] Hansen, Derek, Ben Shneiderman, & Marc A Smith. (2001). Analyzing Social Media Networks With NodeXL: Insights from a connected world.
- [5] Hu, Mengdie & Wongsuphasawat, Krist & Stasko, John. (2016). Visualizing Social Media Content with SentenTree. IEEE Transactions on

- Visualization and Computer Graphics. 23. 1-1. 10.1109/TVCG.2016.2598590.
- [6] Lieberman, Michael. (2014). Visualizing Big Data: Social Network Analysis. Digital Research Conference, San Antonio, Texas, pp 1–23
- [7] Schreck, Tobias & Keim, Daniel. (2013). Visual Analysis of Social Media Data. Computer. 46. 68-75. 10.1109/MC.2012.430.
- [8] Viegas, Fernanda & Wattenberg, Martin & Ham, Frank & Kriss, Jesse & McKeon, Matt. (2007). ManyEyes: a Site for Visualization at Internet Scale. IEEE transactions on visualization and computer graphics. 13. 1121-8. 10.1109/TVCG.2007.70577.