A Qualitative Examination of using Information Visualization on Discussions on Social Media Networks

SUBMITTED IN PARTIAL FULFILLMENT FOR THE DEGREE OF MASTER OF SCIENCE

R. (ROBERT) SPIER
10762884

MASTER INFORMATION STUDIES
INFORMATION SYSTEMS

FACULTY OF SCIENCE
UNIVERSITY OF AMSTERDAM

AUGUST 15TH, 2020

1st Examiner
dr. F.M. (Frank) Nack
Faculty of Science
University of Amsterdam

2nd Examiner
prof. dr. P.T. (Paul) Groth
Faculty of Science
University of Amsterdam

A Qualitative Examination of Information Visualization in Discussions on Social Media Networks

R. (Robert) Spier (10762884)
University of Amsterdam
Amsterdam, The Netherlands
robert.spier@student.uva.com

ABSTRACT

Information Visualization is being used by academics, businesses and creatives to display large amounts of data in an aesthetic, meaningful way to quickly learn from large amounts of data. Within the field of social media, the application of visualization is not new, but is not being used in all aspects. In this study, the usefulness and possibilities of using visualization on discussions is being researched. To do so, people that use social media in a professional context were asked to share their knowledge of social media in order to create a prototype, which was then validated using the same interviewees. In the end, a framework on which future studies can be conducted has been created.

KEYWORDS

Information Visualization, Visual Analytics, Datavisualization, Social Media, Network Analysis

1 INTRODUCTION

Public conversations can be traced back to the earliest form of online *fora*. These platforms, such as Usenet[1] and 4chan[2], allowed users to post content, identifying as whoever they wanted, while also allowing them to stay anonymous all together. Over time, these fora have evolved into a standard of communication serving millions of users. Nowadays, social media is a vital part of the internet. Services like Facebook[3], Twitter[4], Reddit[5]. connect billions of people on a daily basis.[21] On these sites, the functionalities of posting content and speaking to one another are tightly intertwined.

The most popular social media in the Netherlands as of February 2020 is currently Facebook, with an average market share of around 56% over the past year. It is followed by Pinterest[6] (2̃8%), Twitter (1̃0%) and Instagram[7] (7̃%)[21]. In 2008, People were already using social media as a way to read/respond to notes/messages, read comments/posts on their profile and/or timeline or browse other peoples pages[22]. In 2018, it was also confirmed that a significant amount of social media users were using the platform as a mean of finding news and information[1, 10] In fact, circulating links on social media such as Facebook has become a source of news for a significant amount of people[11] This is valid for all ages, but younger people tend to obtain their news more often online than offline. In comparison, a smaller number of users cites Twitter as their source for information, but this has been contributed due to the lower nominal users of this platform[11].

All of these online conversations can be analyzed using specific tools, such as Google Analytics[8] or Facebook Creator Studio[9]. These metrics can be very valuable for a plethora of companies. Marketing agency's use said metrics to measure effectiveness of online advertising campaigns. Media companies can measure how successful their brand is online compared to similar companies and event organisations can make an informed decision on how their event will sell due to online response. In addition, social scientists use these metrics to analyze behaviour of people in an online context.

Analyzing this data is often a tedious and difficult process. In order for analysts to gain significant insights into peoples online conversations, several layers of data have to be extracted. At first, absolute data, such as the number of likes, reactions or shares of a message on a companies social media page can be used to see how a message, and possibly a brand behind it, performs. Second, underlying or subliminal messages from users, the application of sarcasm in messages and context of these messages which can deceive or divert researchers from the actual meaning. Third, the usage of non-verbal communication, where emojis can be converted to sentiment. Images and videos are harder to classify, because these also can provide a more in depth meaning or sentiment behind the basic image or video.

Because the total amount of data that can be harvested on the internet is ever-increasing, methods have been developed over time to help humans comprehend huge amounts of information. One of these methods is visual analytics, an interdisciplinary term which stands for a broad range of visual tools to help humans understand data, and extract information and possibly knowledge from it.[12] Closely related to this methodology are information design and information visualization, which are also ways of transforming raw data into information and knowledge. [19]

This goal of this paper is to provide businesses and individuals with a framework and guidelines when applying information visualization to discussions on social media. By combining the current literature with qualitative research, applicable standards and requirements can be constructed upon which future researchers or businesses can build their implementation. To achieve this goal, a main research question, accompanied by subquestions is posed, these questions provide a scope to define the boundaries of this paper.

The research question is posed as follows:

[1]https://www.usenet.com/
[2]https://4chan.org/
[3]https://facebook.com/
[4]https://twitter.com
[5]https://reddit.com
[6]https://www.pinterest.com/
[7]https://www.instagram.com
[8]https://analytics.google.com/analytics/web/
[9]https://business.facebook.com/creatorstudio

"How can Information Visualization be used to help businesses and researchers extract knowledge from discussions on social media posts?"

In order to define the requirements and conditions in which this question can be answered, a set of sub questions have been posed:

(1) What are the current technical possibilities and limitations in the field of information visualization?
(2) What do researchers and businesses wish to learn from a visualization of discussions on their social media channels?
(3) How can we make sure that these visualizations provide the correct information to these businesses and researchers?

The structure of this paper is as follows: First, a literature study was conducted. This resulted in *a)* a review of current work in information visualization, network analysis and social media and *b* an overview of the current standards and methodologies within information visualization. After this, the visualization requirements for the researchers and businesses will be laid out using formative interviews, which is followed by the prototype section containing the system description and techincal specifications. Next up is the validation section, in which the results from summative assesments will be analyzed to see how researchers and businesses react. The discussion of these tests will be in the second-last section, followed by a conclusion in which potential caveats, opportunities and future options will be discussed.

2 RELATED WORK

In order to gain a better understanding of the current possibilities and limitations of all the vectors regarding this subject, a literature study regarding visualization, social media, sentiment analysis and topic modelling has been conducted to see which elements could provide support and guidelines when creating the framework and prototype.

2.1 Information Visualization

Manovich described Information Visualization as "the mapping between discrete data and a [sic] visual representation"[14, p. 17]. Manovich also divided the subject into four domains which are vital parts of data visualization on their own ranking.

- Graphics
- Information Design
- Information Visualization
- Visual Analytics

These domains can be roughly divided into different amounts of data that are being visualized. For small amounts of data, the domain of visualization is graphics. Datasources up until a hundred can be classified as information design and for units up to a thousand, actual information visualization happens when the data is comprised until a thousand unique data points. Finally, visual analytics stands for the visualization of datasets containing more than thousand items. Because discussions on social media can rank up a lot of responses, including a lot of metadata, this paper will focus on the visualization of multivariate data, balancing between the concepts of Information Visualization and Visual Analytics, as defined by Manovich[14].

As a baseline concept, all visualizations should follow the so-called Gestalt laws of pattern perception[23]. Within Information Visualization, these laws can be defined as:

- Proximity, the (negative) space between elements
- Similarity, the visual resemblance of elements
- Continuity, consistent usage of elements
- Symmetry, or asymmetry to discover inconsistencies
- Closure, by avoiding seemingly random endings.
- Relative size, in comparison to other elements
- Figure and ground, using elements to form other elements

All of these pattern perception-based laws trigger a human response when encountered. This response can be utilized to extract knowledge or indicate significance within the dataset. By applying these laws in visualization, humans will interact more intuitively and have a better natural understanding of any visualization. This paper also builds on the work by Keim et al. in the field of Visual Analytics. In 2006, Keim et al. provided a standard which any visualization should follow. *"Analyse First - Show the Important - Zoom, Filter and Analyse Further - Details on Demand"*[12, p 16]. In 2008, Keim et al. concluded that there is no specific and single right (technological) way of visualizing a dataset, they posed three questions on which a visualization could be judged:

- relevance of specific information
- adequacy of data processing methods and validity of results
- acceptability of the presentation of results for a given task

Keim et al. stated that focusing on specific methods is not the best practice in this field, and rather than using a singular method of visualization, there should be an interdisciplinary solution which proposes a set of visualization options, rather than specific graph types.

Ward, Grinstein, and Keim (2010) provided an extensive set of current visualization methods and data types.[24] Their work allows us to make informed choices regarding the creation of specific visualization methods, manners of testing and validation.

2.2 Social media analytics

Social media analytics *"is concerned with developing and evaluating informatics tools and frameworks to collect, monitor, analyze, summarize, and visualize social media data … to facilit[ate] conversations and interactions … to extract useful patterns and intelligence… "*[7, p 76]. Since social media are an integral part of the internet, analyzing the usage of it can yield valuable information. Examples of social analytics would be politicians that wish to understand what voters think of their latest TV performance or businesses that wish to understand better how they can sell their products to paying customers[26]. For social scientists, researchers and media analysts, this information is also very valuable.

Within Social media analytics, Visual analytics is a branch of analysing data by using a visual representation. This type of analytics is usually applied to combat what is called: "the information overload problem"[13]. Since there are millions of messages posted on social media per day, the amount of raw data that can be used to analyze can be overwhelming. Visual analytics aim to remove that overload by creating visualizations that are easily comprehensible and understandable by humans, in order for them to act upon the findings from them[26].

2.3 Current social network visualization tools

Fisher et al. proposed *Narratives*.[9] A visualization which places trends and changes of topics in discussions on social media networks on a temporal based graph to follow the changing of narrative events while developing. They have applied several techniques to visualize how events relate to one another, such as their *four forms of correlation* which categorizes events based on different parameters. As well as their feature to measure the interest certain topics gained from external sources. Its core view shows several clusters of messages on a temporal axis to see how a topic changes over time.

Wu, Bartram & Shaw proposed *Plexus*, a real-time way of monitoring and analyzing public emotions on Twitter.[25] They have established a system based on three modules: processing data, emotion analysis and visualization. It aims to solve the problem of information overload by creating a prototype that measures peoples engagement. They provide emphasis on the usage of symbols and colors to distinguish different emotions and interaction between them.

Chen et al. proposed D-Map (Diffusion Map), a so-called *Visual Analysis of Ego-centric Information Diffusion Patterns in Social Media*.[6] They presented a dashboard which contains multiple graphs, with a large focus on a hexagon-based map that visualized the landscape of the researched data source. It allows users to analyse information diffusion and propagation via a map metaphor. Users can find key players, communities and important information diffusion paths via a customizable interface. They proposed revised versions of this visualization, additional work was done to create D-Map+[5], E-Map[4] and R-Map.[3] In E-Map, the evolution of events was researched while R-Map explored the same dashboard in the context of reposting content.

2.4 Topic modelling

Topic modelling describes the categorisation and classification of topics in a corpus. It is usually done by applying a statistical model to discover the abstract keywords from a body of text. Topic modelling can be used to structure and categorise large sets of unstructured data. There are several existing algorithms available that enable one to accurately identify topics

Fang et al.[8] tested three existing models on small bodies of text extracted from Twiter: Latent Dirichlet allocation, Twitter Latent Dirichlet allocation and Pachinko allocation. They found that all of these models can yield resourceful topics based on small bodies of text. Especially when their so-called “T-PMI” metric is applied, pointwise measured intelligence using background crawled data to supplement Twitter-related data, such as hashtags and abbreviations.

Further research into LDA for Twitter was conducted by Ostrowski, which focused on the usefulness of the model in regards to commercial goals.[16] They concluded that using LDA could support social media analysis in a complementary way, without the need to supervise the classification.

2.5 Sentiment analysis

Sentiment analysis is being used to computationally identify positive, negative and neutral emotions in a corpus. It can be used to identify how opinions are placed in large datasets. By classification of certain keywords or natural language processing (NLP), unstructured data can be transformed into sectioned elements which can provide researchers with insight in user behaviour and mindset.

Pak & Paroubek (2010) used Twitter as a corpus to mine opinions and identify sentiment among tweets.[17] They created a workflow of downloading tweets using the public Twitter API and build a classifier to analyze the sentiment of these tweets. After this, they optimized their classifier by testing different parameters for the given dataset.

Saif, He & Alani (2012) proposed an algorithm that could accurately classify tweets as either positive or negative.[20] Their hypothesis stated that they could accurately classify tweets using machine learning techniques. They found that best results were achieved when interpolating the generative model of words given semantic concepts into a unigram language model.

A hands-on approach to sentiment analysis was proposed by Nielsen as ANEW, an improved version of an earlier proposed word list named AFINN.[15] Both ANEW and AFINN contain a selection of words with a positive or negative association. By running these words on a corpus, a general sentiment can be extracted. Nielsen proved that running the ANEW word list could yield reliable results in terms of sentiment analysis for social media, specifically on Twitter.

3 REQUIREMENTS

In order to extract the needs of people that actively analyse social media, interviews were conducted with a diverse group of (social) researchers, information specialists and marketing executives. These individuals were initially selected using convenience sampling. However, due to the availability of the required individual types, purposive sampling was applied. In total, 5 people were interviewed.

All interviews were conducted via Zoom.[10] Zoom allows interviews without requiring a physical presence between the interview participants. The main advantage this gives us is that the interviewee is using the application in their own trusted and familiar environment, which yields more accurate responses. To ensure preservation, all interviews were recorded with permission given by all interviewees.

[10]https://zoom.us/

3.1 Interview methodology

The five interviewees were selected from diverse working areas. Two of them are social scientists working in an academic environment, where one researched noise pollution reporting in cities (#1), and one researches terrorism networks and propaganda online (#2). Two other interviewees are full time marketing consultants in the entertainment industry and rely heavily on social media feedback for their work(#3 & #5). The final interviewee is an information scientist working in a consultancy role for ICT companies (#4). This divergent background was chosen to ensure that answers from these persons came from different wishes and d. This helped to ensure a diverse, but saturated response.

Interviews were conducted following an open question model, which was based on the work of Robson[18]. Several starting points were introduced in order to start the conversation, but further questions were all following up on the initial answer. The interviewees were asked to describe how they currently work with (large) social media data, and if they utilize or act upon the information they retrieve from this data. After this, the users were asked how they would filter or order the data, which elements did they find interesting or valuable and how do they use or amend these in order to gain insights in the dataset? Following up on this, interviewees were asked about potential stakeholders to satisfy, and goals they wished to achieve using a dataset. When these questions were answered, the interviewees were asked to describe an ideal visualization for a discussion on social media and were asked to describe how they would interact with it. Finally, they were asked if they would visualize, and if so: how, media such as URLs, images and videos.

3.2 Results

The recordings of the interview were transcribed after conducting them. When all the interviews were completed, the transcripts were transferred to ATLAS.ti 8[11] for analyzing. In ATLAS.ti, codes were created based on expected literature, such as grouping and clustering, which corresponds to the Gestalt principles of proximity and continuity. After the interviews were coded, the obvious and underlying wishes and requests from the interviews became apparent. The interviews yielded three common themes shared among all the participants:

(1) Context
(2) Customization
(3) Clustering

In addition to these themes, several features were queried among the interviewees:

(1) Implementation of sentiment analysis
(2) The usage of topic modelling in tweets
(3) Media visualisation, images and videos

[11] https://atlasti.com/product/mac-os-edition/

3.2.1 Context

With context, the importance of the individual message in a discussion as compared to all the messages is being implied. As well as an individual message when taken from the entire discussion as a whole. Interviewee 1# said: *"A visualization is always a continuation. So you start with summarizing general information, how many people have responded in total, how many likes and retweets does [sic] a message have?. This is something you really want to visualize."*

The context was deemed important by all interviewees, although different underlying motives were discovered. Interviewee 4# noted how they found it tough to believe if a program could identify how certain sentiment should be interpreted: *"I think it's very hard to say in some sort of graph if you're saying "discussions". Because that is quite multi-interpretable, would that be mad or not mad, or neutral, or agreeing or disagreeing or "cool" or not "cool". I would not use a graph for that. I would want to see the top reactions to judge myself. Because that's a concrete example I'm interested in."* In addition, bias was coined as a reason to distrust artificial sentiment analysis, because word association and context interpretation is subjective, as opposed to fully objective.

3.2.2 Customization

The ability to highlight or hide information that is deemed unnecessary or important was a major factor for interviewee 1#, stating: *"I am very much into interactive dashboards because you can easily see insights, as well as discern them. I find this very important".* A core motivation for this was to be able to tweak the visualization to find interesting elements. Interviewees 4# and 5# mentioned similar points, but their motivation stemmed from the ability to filter on their target audience, and view data only interesting for them to sell tickets on. Interviewee 2# mentioned the possibility of building several networks before answering any questions in the first place, mentioning different network graphs based on several deviating parameters. This can be interpreted as the desire to build more than a single graph, or the ability to change one existing graph by removing or adding additional data layers.

3.2.3 Clustering

The grouping of nodes, people, messages and sentiment was often named by all interviewees, with similar interests and goals among all of them. Interviewees 4# and 5#, for instance, were concerned whether a negative message, resulting in negative replies arose from users that were actively targeted in their marketing campaign.

Interviewee 4#: *"Let's state it globally, for your brand it would be interesting to see how certain groups of people think about it. Imagine, the most extreme and dedicated ultra-marathon runners say "these Nike shoes are shit". How much does that mean to Nike? Because Nike just wants to sell shoes that are being worn by all the kids."*

Interviewee 5# stated that the performance of their online posts could be skewed by users deviating from the initial subject posed. A single user could state a controversial opinion, which attracts a lot of replies that do not necessarily contribute in a commercially viable way. Therefore deforming the absolute statistics of a post, which claims over a hundred people replied, yet few were actually commenting on the initial post itself.

Furthermore, interviewees 1#, 2# and 3# all spoke about the ability to cluster or group entities as if it was a standardized function, with a core implementation. Their comments all discussed types of clustering rather than the functionality itself, whereas Interviewee 2# gave clustering based on country as a prime example, and Interviewee 1# and 3# spoke about the feature in a more global matter, mentioning variables as sentiment, topic and rich media such as images.

4 PROTOTYPE

To see how context, customization and clustering can be applied in information visualization, an interactive prototype was designed and build. The prototype had to be accessible for users without the interviewer present, so an MVP using HTML, CSS and JavaScript was chosen. An additional benefit to building an MVP would be for future researchers and businesses to build on the experiences and work provided in this paper.

4.1 System description

The prototype is built as a web application. It consists of an HTML document, accompanied with a CSS stylesheet to add styling, and JavaScript for interactive elements. This method of testing was chosen so rapid prototyping could be possible, while versions could easily be hosted online to enable user testing. In order to easily implement visualizations, the usage of D3 or *data-driven documents* was required.[2] The application uses the latest version to date, v5[12] to visualize the aforementioned dataset.

4.2 Dataset

In order to populate the visualization with an actual discussion, a dataset containing a social media post with replies was reqruired. To extract the dataset, a script was written in JavaScript, utilizing NodeJS[13] was used to communicate with the Twitter API. Tweets were manually selected beforehand and their ID's were then scraped using the Twit library[14]. When completed, full tweet objects were stored in a JSON format as a local file.

The dataset to test the prototype which contains 79 unique tweet objects. Tweet objects contain a single tweet including all available metadata it has, such as the user who posted it, all available hashtags used, unique identifier, date and time, and so on. The dataset was based around a single tweet posted by The Times[15]. It was chosen for a number of reasons: the language was English, which is the optimal language for topic modelling and was a language spoken by all interviewees. It was posted by an international media outlet, which would expose the tweet to a large and diverse audience and it received a lot of replies, which were the core part of the visualization. After the initial tweet ID was saved, all the replies on the tweet and all the replies on those tweets were recursively scraped using the API.

[12] https://github.com/d3/d3/releases/tag/v5.16.0
[13] https://nodejs.org/en/
[14] https://www.npmjs.com/package/twit
[15] https://twitter.com/thetimes/status/1287739259886215170

4.3 Data sanitation

In order to use the dataset to work in a prototype, several rounds of sanitation and amendments were applied. First, the dataset was split into two arrays. One containing all the tweet objects called the "nodes" and another one containing all the connections between the nodes called the "edges". Every single tweet object was checked for the textitin_reply_to_status_id_str value to see if which tweet objects were direct replies to other tweets in the dataset. If a match was found, that set of tweets was stored as a link in the edges array.

An implementation of topic modelling was done using the LDA package[16]. A JavaScript implementation similar to the method used by Fang et al.[8] All tweet objects were expanded with a topics property containing a list of topics discovered by using the LDA algorithm or left empty if no clear topics were identified. Alongside was an implementation of sentiment analysis following the work of Nielsen using their AFINN list.[15] The package Sentiment[17] was used which was built around this list of words to identify positive and negative word associations in order to calculate the sentiment of every individual tweet.

4.4 User interface

The user interface of the application can be found in figure 1. It consists of two elements, the left side is the D3 implemented adaptation of a classical network graph, the right side consists of a sideboard showing deeper information about single nodes when they are clicked. The network graph was chosen because it allows us to visualize context by linking messages together, allow the user to customize the visualization, and provide clustering of the nodes. At the top left 1, three buttons enable users to switch between different node type views. Toggle sentiment changes the nodes into a circle with a color range between red and gray when the sentiment is negative, and gray and green when the sentiment is positive as illustrated in figure 2. Toggle topics change the nodes into one or two words when the LDA algorithm was able to extract topics from the message or display the message "no topics found" when it was unable to do so. This is shown in figure 3. The last button, toggle users change the nodes in an image which is the avatar used by the poster of that message on Twitter.

The visualization on the left side is a *network graph*. It's intended use is to originate from a single node in the centre (2), but it allows the creator of the dataset to enter multiple starting points. The centred node in the middle contains a black border to give the user the signal that it is a starting point in the visualization. From the centred starting node, several so-called branches expand where people directly replied to the first tweet. Every line or edge as they are called within the graph displays a relation between two messages, as one is a reply or quote of another. When several messages are being posted in reply to a single message, clustering automatically occurs around the popular node (3). The concept also visualizes discussions as *chains* of messages which constantly reply to each other (4). The concept allows users to create a *tree diagram*, using the functionalities of a network graph. The prototype diverts from the original concept of a tree, because it allows the user to create multiple starting points. Another key

[16] https://www.npmjs.com/package/lda
[17] https://www.npmjs.com/package/sentiment

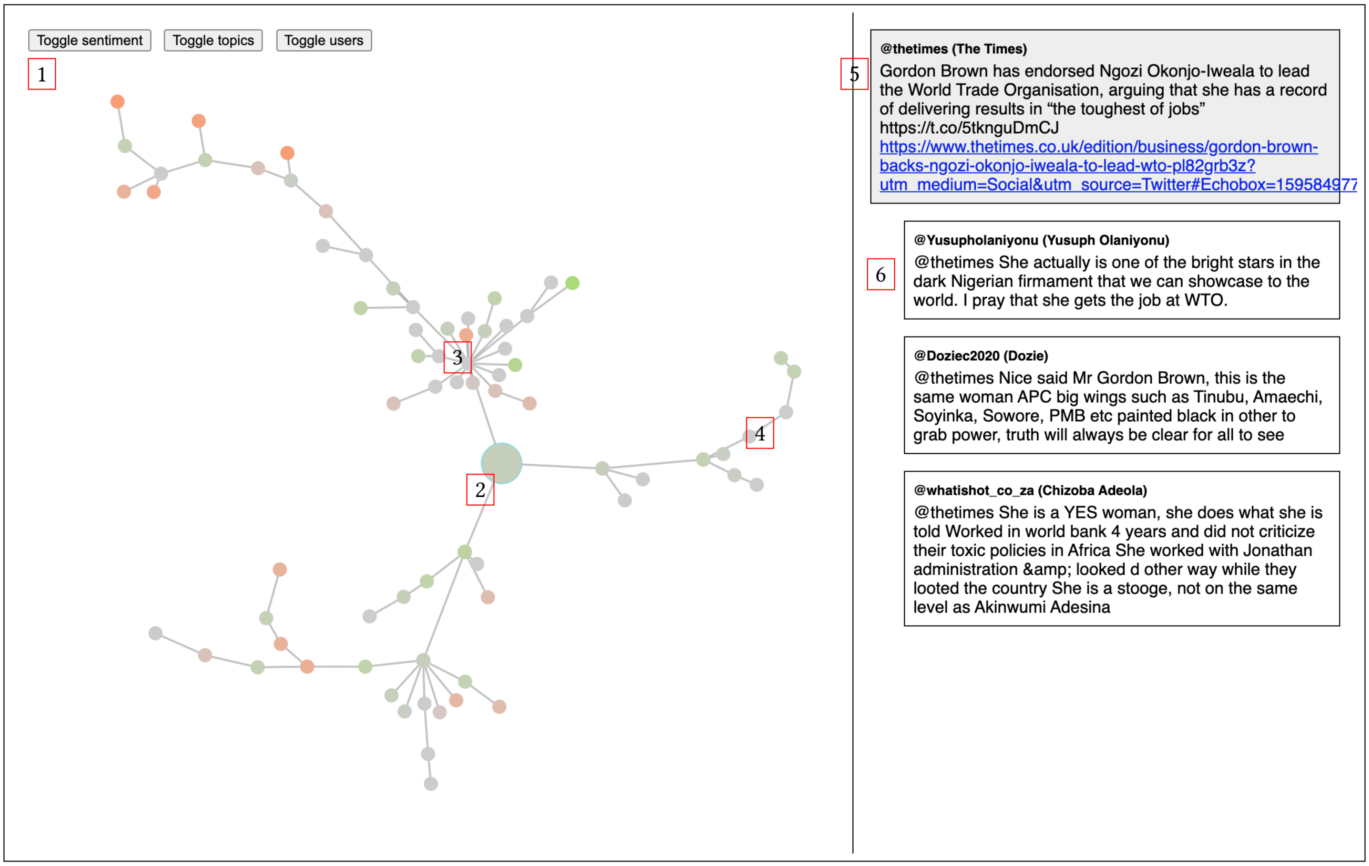

Figure 1: Interface of the web application used for user testing

visualization method was incorporated when the topic view is enabled. The network graph converts all nodes into words and creates a *word cloud* without hierarchy in terms of weighted words, but rather uses the edges to connect subjects as they change through the messages in the branches. (see figure 3). Users are able to interact with the visualization itself, as they can drag individual nodes to positions of their own choice. Because all nodes in the graph are connected, this allows them to reshape the entire visualization to their own preferences. They're also able to zoom in and out on the visualization, and drag their current view to see a concise part of the graph.

The right side of the prototype reserves space for the user to perform deeper analysis in single nodes and their direct parent and replies. Whenever a user clicks on a particular node, the message behind it including the metadata of the tweet is shown in the right column. The clicked message has a subtle gray background to distinguish it from the other tweets (5). If there is a parent message available, it will be shown above the clicked tweet, and the clicked tweet will show a small margin on the left to indicate it is a child. In addition to a potential parent tweet, all the available replies on the tweet are indented on the left side as well to indicate they are in a similar, but later order than the original message (6). Whenever a node is clicked, a small blue border appears around the node to give the user a visual cue which node is currently being viewed on the right.

All code used can be found on GitHub[18], as well as a live demonstration which is being hosted online as well.[19]

5 VALIDATION

To validate the prototype, user tests were conducted with the same five interviewees as originally were used to discover the initial requirements. Interviewee 5# was unable to be interviewed but was replaced by a colleague of them in an identical role to minimize discrepancies between their answers and requirements. As was done with the initial interviews as described in chapter 4, the user tests were also conducted via Zoom or Skype[20]. All users were asked to share their screen with the interviewer so they would be able to exactly see which actions the interviewees were performing when being tested.

5.1 Interview methodology

Having determined the requirements and goals of the interviewees during the first round of interviews, a set of three questions were

[18]https://github.com/roberrrt-s/Context-Visualizer
[19]http://robertspier.nl/uva/thesis
[20]https://www.skype.com/

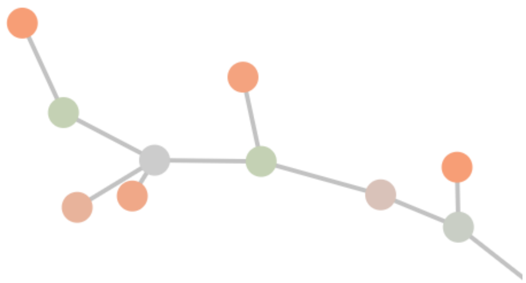

Figure 2: Individual nodes in the sentiment view, showing red nodes for negative and green nodes for positive comments

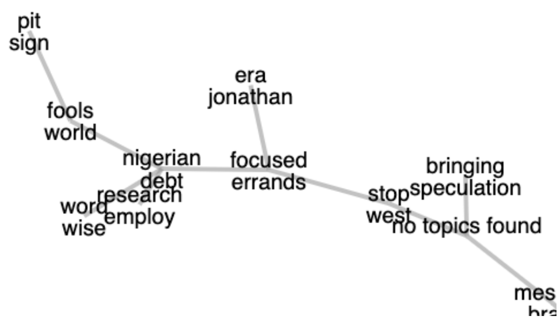

Figure 3: Individual nodes in the topics view, showing two identified topics for every comment

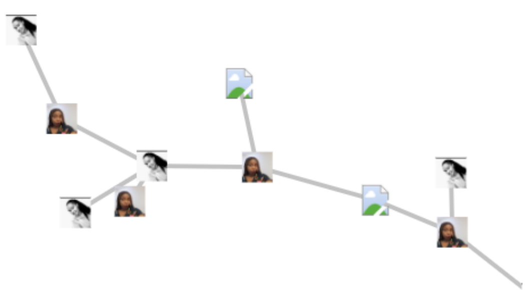

Figure 4: Individual nodes in the users view, showing the Twitter avatar of the poster of the comment

asked to verify if users would intuitively understand how the prototype would function. After these questions, interviewees were asked to speak out loud while completing three scenarios. Interviewees shared their screen so any interaction could be monitored and they were asked to communicate any non-verbal thoughts, such as confusion, joy or satisfaction.

5.1.1 Questions

The following questions were asked before the user was asked to interact with the application:

(1) *What do you think the three buttons at the top left corner of your screen do?*
(2) *What do you think the original message was in this discussion?*
(3) *How do you think people responded in this discussion? Were they negative, or positive, did they divert from the subject?*

Question one was asked for two reasons: first to see if people would understand that the button would change a view, and show a different perspective on the visualization. Secondly, to discover how they expected the visualization to change. Because the users had no prior interaction with the prototype, their initial pattern of thought could be discovered. Question two was asked with the purpose of identifying the centre, or start of the visualization. As the prototype renders the visualization starting from the middle, it was assumed users would choose a node in the middle. To make this more obvious, however, a small black border was added to the starting node to signal the users this node was the starting point of the discussion. Finally, question three triggered users to look deeper into the visualization itself. The goal here was to see if users could identify nodes as messages by users, and assume messages to be connected with lines or edges.

5.1.2 Scenarios

The users were asked to complete the following tasks:

(1) *There are multiple "sub" discussions happening, you want to find out what they are about, what would you do?*
(2) *You want to identify who is the most negative actor in this discussion, how would you do that?*
(3) *You want to find out how every branch talks about your message, how would you do that?*

The tasks were not labelled with right or wrong, and the interviewees received no feedback during the completion of the tasks. Whenever an interviewee stated that they completed the task, a question was asked to elaborate on why they believed why their task was complete, or to explain why they were unable to finish the task because they were missing information.

5.2 Results

This section outlines the findings of the interview based on the answers and non-verbal interaction of the user with the prototype. Users, in general, were neutral *I like this concept, I'm not sure if this bubble-way of visualizing would be the right way for analysis in organizational context* (#3) to positive regarding the usage of visualization when analyzing a discussion on social media. Interviewee 5# stated: *"I would have definitely used something like this if I wanted to measure how my posts [on social media] were received"*. People were able to identify the distinct elements of the network visualization as a discussion when they were informed the dataset was a Twitter conversation.

Question 1.

Interviewee 1# started by associating sentiment to colours immediately, stating *"Red is probably negative, and green is probably positive".* The interviewee assumed that the toggling implied changes of views, stating topics would probably change the nodes to words, and users would show posters and individual messages. Interviewee 3# and 4# replied with similar answers, whereas interviewee 2# assumed it would zoom in on particular elements of a node, as opposed to showing different elements. Interviewee 5# did not suspect any deformation or changing of nodes but rather assumed it would filter the existing nodes.

Question 2.

All interviewees except for number 5# were able to immediately identify the correct node as the initial message in the discussion. Interviewee 5# reacted confused and requested more information to answer the question. When reformulated as "Can you pinpoint the starting point in this visualization", the interviewee was able to answer correctly. Methods of identification happened because of the centrality of the node: *I think the source of the discussion [is from] the middle?* (1#, 3#, 4#) and the outgoing links that expanded from it (2# and 5#).

Question 3.

The first reaction, in general, was of confusion. Interviewees were hovering their mouse over nodes to see if it would give feedback and immediately stated they were unable to draw any conclusions at all purely from the visualization itself: *"I don't know what I'm seeing yet?"* (1#). After 10 to 20 seconds, all interviewees came up with assumptions regarding the shape of the visualization and the formation of the branches: *"Quite varied, but more people agree on certain points. One chain is concentrated, so this person said something so this is a hot topic, and this guy isn't relevant."* (4#). Because the initial view displayed sentiment, interviewees tried to identify the colours of the nodes as well: *I think people are positive, the colours are probably from the stoplight, grey is 0, I think this bubble [with the most reactions].* (3#)

Scenario 1.

Interviewee 5#:

> *"By clicking on nodes, it started with the biggest node, then there were branches, one is going to the right, and that gained two reactions, and another one that had reactions and so on. To follow such a discussion, you could click on nodes to find out what they are about. So first look at the post, and then recursively follow the nodes, the top left one had a lot of reactions, on a post, and all these surrounding are based on that. A lot of engagement."*

All interviewees used the centred node they previously identified as the beginning of the discussion as the starting point of their analysis. Their first instincts were to toggle all the buttons to view their effect (1# & 2#) or to click on the individual nodes (3#, 4# and 5#). Interviewees started playing around with the three toggles when they interacted with them, comparing different views with each other to see if any information arises. After switching around, the interviewees preferred the initial sentiment view above the topics- and user view. When identifying what the discussion was turning in to, interviewee 2# tried to use the topic view. All other interviewees clicked on individual nodes to investigate tweets. When they understood the padding and light background as visual cues for the hierarchy, they reacted positively to this way of investigating single messages. *"I like this simple way of visualising, using padding and indentation. This is (though) raw data, perhaps this is the core data?"* (3#).

A distinction in the analysis was however observed between social scientists and marketing consultants. Social scientists were aimed at identifying the people in the specific clusters based on their behaviour, or in the words of interviewee 1#: *"Three finds: 1 rants, 1 is positive and 1 goes full conspiracy".* Marketing consultants were looking specifically for the subject of a discussion. All interviewees were interested in the nodes with high sentiment values, positive and negative. This was also the case for nodes which had a lot of connected edges.

Scenario 2.

This question was interpreted in three different ways, first: the most negative actor is the person in the visualization that has the most (light) red dots. Second, the most negative actor would be the person that attracts the most negative comments. Third, the negative actor has the lowest sentimental score of all messages in a single message. Interviewee 1# identified the user with the most negative messages in a discussion by continuously toggling sentiment and user images. The interviewee skipped their so-called "boring" chain before because there were no negative messages in that branch. Interviewees 4# & 5# used the same methodology to conclude their answer. Interviewee 2# tried to identify the influencer responsible for the most negative messages. The interviewee did so by trying to visually identify the single node that had the most connecting (light) red nodes. *"I found one person who is responsible for all the negativity, so she's an influencer. She's the negative actor".* Interviewee 3# identified the most distinct red absolute colour in the visualization and assigned the poster the label of the most negative actor. The interviewee stated that they would have liked to use topics to find negative topics, and called them "a potential" but "very unclear in this visualization" before settling on the user which made the single most negative comment.

Scenario 3.

This scenario yielded the most divergent answers. Interviewee 1# went back to the centred node and used the three branches and their direct descendent nodes to identify how people were starting conversations about the subject. The interviewee judged the so-called "conspiracy cluster" and the "boring discussion cluster" to be of no obvious value, and focused on the polarized discussion in the final cluster. Interviewee 2# started with a more holistic approach, first defining the sentiment of each cluster, before categorizing them and adding tags. If a cluster would be in their field of interest, the interviewe would further analyze it. Interviewee 3# created a word cloud using the topic view but was easily confused about the widely broad range of words. The interviewee stated they would prefer a combination of sentiment analysis with topic modelling using some sort of hovering principle. Their preferred method to investigate this was the usage of topics over sentiment. Interviewee 4# used sentiment as guidance to analyze the different clusters but was soon surprised by incorrect contextual sentiment. The usage of sarcasm in certain messages gave the nodes a positive rating but was actually meant to be negative. Finally, interviewee 5# went through the entire visualization using the topic view, and read every individual tweet before giving a detailed description of the activity of every cluster.

6 DISCUSSION

Since the nature of this study is to provide social media analysts with a framework from which can be worked on in future projects, several limitations can be listed which have impacted the result of the study. However, the lack of definite answers and quantitative validation can be used to further continue the current prototype and requirements for a more specific setting.

6.1 Limitations

A major limitation of this paper is the relatively small sample size in respect to the testing population. In order to discover more interests and demands, or to specify the current requirements, a broader selection of people should be interviewed. Next to this, the validation and discovery phase revolved around user interviews. There was no quantitative validation done to verify whether or not the proposed solutions and finds were also fitting for all possible users.

In addition, the exploratory nature of the study diverted the attention and focus of the interviewees. While this paper focused on the visualization of discussions, interviewees were inclined to answer the questions with solutions or requests that were deemed out of scope for this investigation. Correcting this pattern of thoughts was considered disruptive and possibly steering the interviewees, so only marginal adjustments in terms of asking questions were made to compensate for this.

A tough issue that arose during questions related to identifying what users discussed was the choice of the dataset. Because interviewees had no prior knowledge of the dataset, users were easily confusion about the interpretation of certain words or sentences. This showed two problems with the current experiment. First, interviewees had different perceptions of the semantic value of words. Second, it made testing the prototype with a dataset that was irrelevant for the user less reliable. Interviewees may have been able to identify certain elements, such as sarcasm, more easily if they would have chosen their own dataset.

Finally, the chosen system of using a network visualization as a tree, combined with a high level of customization caused confusion for all interviewees at at least one point during the validation. Therefor, the threshold of understanding and easily using such a visualization could be higher than anticipated, which could void the benefits that a visualization could offer in the first place.

6.2 Opportunities

Despite the lack of thoroughly validated results, a lot of options to continue this research were based on them. As all interviewees expressed fun and enjoyment, as well as direct benefits to their area of expertise. Optimization of the current prototype would be a preferred next step, while also increasing the amount of interviewees. Because the prototype is built in JavaScript using open source frameworks, the prototype can be developed and tested further, exploring options such as the merging of views, integration with more social media, topic modelling for clusters and live interpretation of discussions.

It should be noted however, that sentiment analysis can be incorrectly interpreted when the researcher using the visualization is unfamiliar with the dataset and the additional usage of language, such as mass usage of sarcasm. For topic modelling, it is suggested to provide ad hoc topic modelling of clusters as opposed to individual tweets, as the corpus on these was usually insufficient to generate usable topics.

6.3 Future work

In terms of continued research, three points of direct improvement can be tackled:

(1) Applied topic modelling for clusters and branches
(2) Optimize sentiment analysis per dataset
(3) Improve the visualization of context and relationships
(4) Flexibility in terms of the dataset

Since the word cloud caused some confusion for the interviewees, as the corpus was very small in some messages, experimentation with the modelling of entire clusters and branches, customizable by the user of the visualization can yield interesting results. In addition, since interviewees had issues with identifying sarcasm due to the sentiment analysis using a score based on words instead of sentences, it is advised to experiment with different, or possibly multiple options of sentiment analysis which should be optimized for the dataset of choice by the interviewee. The context in which messages were send and their relationships were hard to identify, customization may be a solution here, as hiding certain branches or applying different weights on properties such as likes, favorites or replies can help users identify context more easily and should be researched. Finally, customization of the dataset, or even datasets generated or curated by the interviewees could yield more accurate results when identifying the context of messages.

7 CONCLUSION

In this paper, the visualization of discussions on social media networks was explored. It aimed to identify which types and aspects of Information Visualization could be extracted, applied and/or combined in order to help academic researchers and businesses. The methodology of discovery yielded a baseline of requirements from which a prototype was developed to test these requirements. Based on the findings of the interface validation, it can be concluded that an interactive tree visualization, using a network graph concept, extended with a word cloud was supportive and constructive for the interviewees to conduct analysis.

Visualization of discussions can be used to discover influencers in discussions, see how a discussions topic changes over time, depending on the different branches of people replying and display polarization as a discussion progresses. It allows researchers and businesses to skip a step of running data analysis on a dataset to view absolute numbers, and instead emphasizes how context within a discussion can be discovered, in order to get a better image of how people talk about something, or discuss a topic with one another.

However, the most prominent find was the perception of the interviewees, which were unanimously positive about using visualization as a way to analyze and conceptualize interaction on social media. Even though most of them heavily relied on raw data from messages, the barrier of reading them in the context of the visualization seemed to decrease significantly. It also proved that sentiment analysis and topic modelling are valuable tools when analyzing a dataset, even when used in combination with information visualization.

On a technical level, the application of the D3 framework and its high level of customization allows researchers and businesses to optimize a visualization for their specific goals and requirements. Usage of the D3 force-directed graph was chosen and drew positive replies from all interviewees and can be recommended as a way of visualizing a discussion. Customization of the visualization also allows researchers and businesses to pin-point their specific interests. For social researchers, this usually involved group dynamics and user behaviour, where marketing consultants preferred to discover which subjects were touched inside the discussions. This paper provides a framework and broad defaults which can be used to further pursue the analysis of discussions on social media using visualization.

Clustering of nodes within the visualization led to the discovery of influencing factors within a discussion, whether a single message or person was responsible for a lot of engagement within the discussion helped the interviewees with measuring which actors are of interest. It also helped interviewees discover whether or not the replies were directly related to the subject when used in combination with topic modelling, which provides a more truthful representation of a discussion as opposed to raw numbers.

In all, the study found interesting results which should be pursued by analysts. Furthermore, future researchers are encouraged to build upon the defaults established in this paper to see how more specific or divergent discussions can be optimally visualized.

REFERENCES

[1] Annika Bergström and Maria Jervelycke Belfrage. News in Social Media: Incidental consumption and the role of opinion leaders. *Digital Journalism*, 6(5):583–598, 2018.

[2] M. Bostock, V. Ogievetsky, and J. Heer. D³ Data-Driven Documents. *IEEE Transactions on Visualization and Computer Graphics*, 17(12):2301–2309, dec 2011.

[3] Shuai Chen, Sihang Li, Siming Chen, and Xiaoru Yuan. R-Map: A Map Metaphor for Visualizing Information Reposting Process in Social Media. *IEEE Transactions on Visualization and Computer Graphics*, 26(1):1–1, 2019.

[4] Siming Chen, Shuai Chen, Lijing Lin, Xiaoru Yuan, Jie Liang, and Xiaolong Zhang. E-Map: A Visual Analytics Approach for Exploring Significant Event Evolutions in Social Media. *2017 IEEE Conference on Visual Analytics Science and Technology, VAST 2017 - Proceedings*, pages 36–47, 2018.

[5] Siming Chen, Shuai Chen, Zhenhuang Wang, Jie Liang, Yadong Wu, and Xiaoru Yuan. D-MAP+: Interactive visual analysis and exploration of ego-centric and event-centric information diffusion patterns in social media. *ACM Transactions on Intelligent Systems and Technology*, 10(1):1–26, 2018.

[6] Siming Chen, Shuai Chen, Zhenhuang Wang, Jie Liang, Xiaoru Yuan, Nan Cao, and Yadong Wu. D-Map: Visual analysis of ego-centric information diffusion patterns in social media. *2016 IEEE Conference on Visual Analytics Science and Technology, VAST 2016 - Proceedings*, pages 41–50, 2017.

[7] Weiguo Fan and Michael D. Gordon. The power of social media analytics. *Communications of the ACM*, 57(6):74–81, 2014.

[8] Anjie Fang, Craig Macdonald, Iadh Ounis, and Philip Habel. Topics in Tweets: A User Study of Topic Coherence Metrics for Twitter Data. In *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, volume 9626, pages 492–504. 2016.

[9] Danyel Fisher, Aaron Hoff, George Robertson, and Matthew Hurst. Narratives: A visualization to track narrative events as they develop. *VAST'08 - IEEE Symposium on Visual Analytics Science and Technology, Proceedings*, pages 115–122, 2008.

[10] Alfred Hermida and David Domingo. The active recipient: Participatory journalism through the lens of the Dewey-Lippmann debate. *International Symposium on Online Journalism*, pages 0–21, 2011.

[11] Alfred Hermida, Fred Fletcher, Darryl Korell, and Donna Logan. SHARE, LIKE, RECOMMEND: Decoding the social media news consumer. *Journalism Studies*, 13(5-6):815–824, 2012.

[12] D.A. Keim, Florian Mansmann, J. Schneidewind, and Hartmut Ziegler. Challenges in Visual Data Analysis. In *Tenth International Conference on Information Visualisation (IV'06)*, number Iv 2006, pages 9–16. IEEE, 2006.

[13] Daniel Keim, Gennady Andrienko, Jean-daniel Fekete, G Carsten, and Guy Melan. Visual Analytics : Definition , Process and Challenges To cite this version : Visual Analytics : Definition , Process , and Challenges. 2008.

[14] Lev Manovich. What is visualisation? *Visual Studies*, 26(1):36–49, 2011.

[15] Finn Årup Nielsen. A new ANEW: Evaluation of a word list for sentiment analysis in microblogs. *CEUR Workshop Proceedings*, 718:93–98, 2011.

[16] David Alfred Ostrowski. Using latent dirichlet allocation for topic modelling in twitter. *Proceedings of the 2015 IEEE 9th International Conference on Semantic Computing, IEEE ICSC 2015*, pages 493–497, 2015.

[17] Alexander Pak and Patrick Paroubek. Twitter as a corpus for sentiment analysis and opinion mining. *Proceedings of the 7th International Conference on Language Resources and Evaluation, LREC 2010*, pages 1320–1326, 2010.

[18] Colin Robson. *Real world research : a resource for users of social research methods in applied settings.* John Wiley & Sons Ltd, Chichester, 2016.

[19] Jennifer Rowley. The wisdom hierarchy: Representations of the DIKW hierarchy. *Journal of Information Science*, 33(2):163–180, 2007.

[20] Hassan Saif, Yulan He, and Harith Alani. Semantic Sentiment Analysis of Twitter. pages 508–524, 2012.

[21] Statcounter. Social Media Stats Netherlands, 2020.

[22] Kaveri Subrahmanyam, Stephanie M. Reich, Natalia Waechter, and Guadalupe Espinoza. Online and offline social networks: Use of social networking sites by emerging adults. *Journal of Applied Developmental Psychology*, 29(6):420–433, 2008.

[23] Dejan Todorovic. Gestalt principles. *Scholarpedia*, 3(12):5345, 2008.

[24] Matthew O Ward, Georges Grinstein, and Daniel Keim. *Interactive data visualization: foundations, techniques, and applications.* CRC Press, 2010.

[25] Xiaodong Wu, Lyn Bartram, and Chris Shaw. Plexus: An Interactive Visualization Tool for Analyzing Public Emotions from Twitter Data. 2017.

[26] Daniel Zeng, Hsinchun Chen, Robert Lusch, and Shu Hsing Li. Social media analytics and intelligence. *IEEE Intelligent Systems*, 25(6):13–16, 2010.