SocialOcean: Visual Analysis and Characterization of Social Media Bubbles

1st Alexandra Diehl
University of Konstanz
Konstanz, Germany
alexandra.diehl@uni.kn

2nd Michael Hundt
University of Konstanz
Konstanz, Germany
michael.hundt@uni.kn

3rd Johannes Häußler
University of Konstanz
Konstanz, Germany
Johannes.3.Haeussler@uni.kn

4th Daniel Seebacher
University of Konstanz
Konstanz, Germany
daniel.seebacher@uni.kn

5th Siming Chen
University of Bonn
Bonn, Germany
Siming.Chen@iais.fraunhofer.de

6th Nida Cilasun
University of Konstanz
Konstanz, Germany
nida.cilasun@uni.kn

7th Daniel Keim
University of Konstanz
Konstanz, Germany
keim@uni.kn

8th Tobias Shreck
Graz University of Technology
Graz, Austria
tobias.schreck@cgv.tugraz.at

***Abstract*—Social media allows citizens, corporations, and authorities to create, post, and exchange information. The study of its dynamics will enable analysts to understand user activities and social group characteristics such as connectedness, geospatial distribution, and temporal behavior. In this context, social media bubbles can be defined as social groups that exhibit certain biases in social media. These biases strongly depend on the dimensions selected in the analysis, for example, topic affinity, credibility, sentiment, and geographic distribution. In this paper, we present *SocialOcean*, a visual analytics system that allows for the investigation of social media bubbles. There exists a large body of research in social sciences which identifies important dimensions of social media bubbles (SMBs). While such dimensions have been studied separately, and also some of them in combination, it is still an open question which dimensions play the most important role in defining SMBs. Since the concept of SMBs is fairly recent, there are many unknowns regarding their characterization. We investigate the thematic and spatio-temporal characteristics of SMBs and present a visual analytics system to address questions such as: What are the most important dimensions that characterize SMBs? and How SMBs embody in the presence of specific events that resonate with them? We illustrate our approach using three different real scenarios related to the single event of Boston Marathon Bombing, and political news about Global Warming. We perform an expert evaluation, analyze the experts' feedback, and present the lessons learned.**

***Index Terms*—Geospatial Visual Analytics, Echo chambers, social media**

I. Introduction

Social media and news media allow citizens, corporations and authorities to create, post, and exchange content [30]. Recently, the term Social Bubbles or Social Media Bubbles, as described in this paper, captured the attention of the society in an unprecedented way. Recently, awareness has risen regarding the influence of social media in shaping our lives, decisions, and democracy. Recent investigations look into how social media has influenced, e.g., public voting like in the UK Brexit referendum, and in Donald Trump's presidential election [25]. The magazine Wired posted: *The social bubbles that Facebook and Google have designed for us are shaping the reality of your America* [56]. The Guardian, Forbes, and many other news media are alerting of these phenomena. In this context, we use the term Social Media Bubbles (SMBs) as *social networks connected by social media that exhibit a tendency or bias towards a specific topic, event, or matter*.

Personalization or personalized experience in search engines and social media consists of the use of previous users' activities to customize what the users see or the results of a search. The outcome are filter bubbles where the users see only a limited subset of results depending on their background and previous activities, and also where different users see different results. The rise of *personalization* on search engines and consequently, filter bubbles, are expected to reinforce phenomena such as SMBs, echo chambers [24], and spreading of rumors [47]. SMBs are social groups that could exhibit a certain bias towards a topic, influencer, or interest, but their opinions are not necessarily polarized, uninformed, or reinforced. One of the characteristics of SMBs is the presence of discussions and diversity in opinions and sentiment towards the bias that tight them together, contrary to the concept of echo chambers. The phenomena are increasing in interest because, during the last years, the social networks *Facebook* and *Twitter* communicated new changes in their search and ranking algorithms to create a more personalized experience for the users, based on their interests, family and friends, and previous positive sentiments ("likes" in Facebook). There is evidence that those changes could have increased filter bubbles [22, 5, 17] and therefore SMBs.

The detection of SMBs represents a complex and ill-defined problem because social interactions depend on a diversity of factors such as where and when they take place, demographics (gender, religion, education, occupation), affinities, abilities, beliefs, and previous experiences [36]. In our work, we characterize SMBs using intrinsic dimensions from social media and the concept of *homophily*, as proposed in the social sciences [36]. We limit our analysis to explicit connections among users. Confirmation bias refers to the seek or partial interpretation of evidence based on previous or existing beliefs, expectations, or hypotheses in hand [38]. We use the term "bias" in this

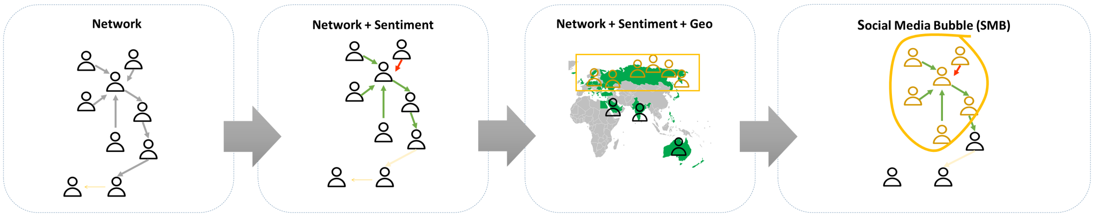


Fig. 1. Social media bubbles (SMBs) can be identified and characterized based on different dimensions. Baseline dimensions include the connectedness (e.g., depending on the platform expressed as follows, likes, retweets etc.) and message interest or shared sentiment towards a topic. In addition, we also consider geographic location and user attributes like gender, opinion etc. as supporting dimensions to perform the SMB analysis.

work as an inclination in favor or against a composed set of dimensions. The analysis of the confirmation bias is important for our work because it is what ties the social groups in social media and therefore transforms them in an SMB. We analyze the qualitative and quantitative aspects of these biases using our visual analytics tool.

We use an exploratory data analysis approach to investigate different aspects of SMBs. Figure 1 shows our simplified tasks workflow. The study of SMB presents several challenges that are difficult to tackle by using automated algorithms alone. Among them, vagueness or uncertainty, geo-located inaccuracy and heterogeneity, thematic inaccuracy, credibility, and trustworthiness of the users that set the information on the social media. We propose a visual analytics approach that combines state-of-the-art machine learning and automated algorithms with interactive visual analysis. Our approach serves as a starting point to understand how people connect and collectively act as social groups using social media. We want to investigate the following research questions:

- R1. How does space, time, and thematic attributes of social media shapes SMBs?
- R2. How is their internal structure regarding the distribution of topics, sentiments, and connectivity in-between users?
- R3. Are SMBs tight to events or do they exist independently of those events? and What characteristics do they show?

Our tool, *SocialOcean*, is a visualization system based on a coordinated multiple views (CMV) architecture [44] that allows the experts to explore diverse aspects of the data and homophily dimensions. It contains two main visual components: a mention graph that shows the connection between different users, and a map view that shows the geographical location distribution. Machine learning topics and sentiments classification algorithms help analyzing details and context data. Additional views, such as histogram views and details-on-demand allow for the analysis of trending topics, sentiment in the messages, and user metadata. Figure 2 shows the visual components of our system. Our main target users are data scientists, political scientists, and journalists, who regularly disentangle the complexity of the social phenomena.

The main contribution of our work is the integration of interactive visual analysis with automated methods for the thematic and spatio-temporal analysis of SMBs. To the best of our knowledge, this is the first work that presents all these aspects in one visual analytics system with the focus on the characterization of SMBs. We evaluated our visual analytics approach thoroughly with domain experts from political sciences and journalism. All of them found our system useful to characterize social media bubbles, and supportive to perform further analysis such as situational awareness.

II. Related Work

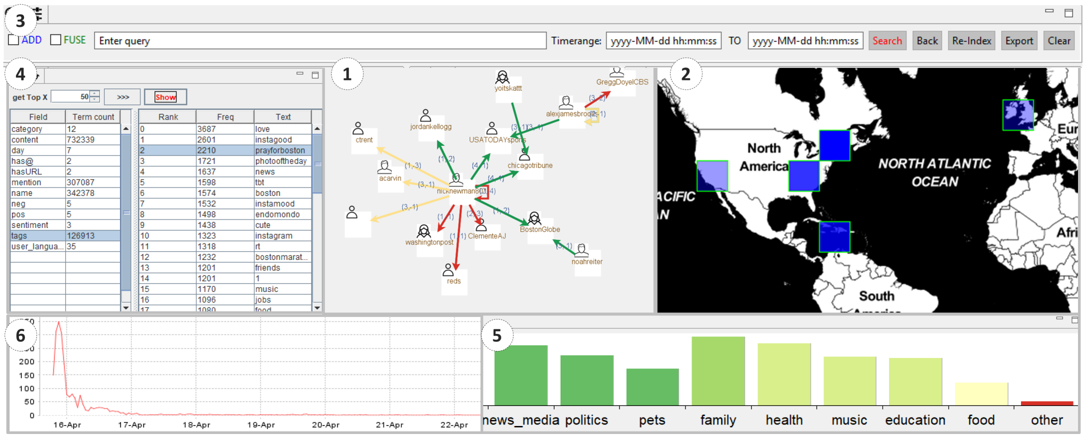
We first discuss related work concerning general social media analysis, community detection and networks, geo- and semantic-visualization and social media bubble analysis in social media.

A. Social Media Analysis

Social media analysis has received much research attention in the visual analytics domain. On the one hand, because the data from social media allows researchers to conclude current topics and the behavior of users, and on the other hand, because the large and heterogeneous data volumes present a challenge for researchers and promote the development of new analysis methods and visualizations. Chen et al. [13] presented a recent state-of-the-art report about social media visual analytics, where they proposed a taxonomy of work in the social media domain and provided an overview about analysis and visualization techniques. The presented approaches include work about the detection and evaluation of topics in social media [18, 40], geographic analysis of social media data [34, 2] and network analysis [26]. Text information is important in social media analytics, related techniques can be found in [4, 7].

B. Community Detection in Networks

Community detection in networks is a multidisciplinary topic with application in biology, physics, economics, social and political sciences and many more [51, 52]. However, the problem of community detection is ill-defined, making it a diverse topic with no clear-cut guidelines. A good overview and a critical analysis of the problem of community detection is presented by Fortunato and Hric [23]. Prominent examples of community detection algorithms are *Newmans community identification algorithm* [37], which, for example, was used by Heer and Boyd in their system *Vizster* to visualize social networks [27] or the *Louvain algorithm* by Blondel et. al. [6]. Similarly, Wade et al. identify communities in blog networks [53], and OpinionRings analyzed the opinion networks with mining and visualization techniques [20]. D-Map visualized ego-centered user groups and profiles with the map metaphor [14]. However,



□ ADD □ FUSE Enter query Timerange: yyyy-MM-dd hh:mm:ss TO yyyy-MM-dd hh:mm:ss

get Top X 50

Field	Term count
category	12
content	732339
day	7
has@	2
hasURL	2
mention	307087
name	342378
neg	5
pos	5
sentiment	3
tags	126913
user_langua...	35

Rank	Freq	Text
0	3687	love
1	2601	instagood
2	2210	prayforboston
3	1721	photooftheday
4	1637	news
5	1598	tbt
6	1574	boston
7	1532	instamood
8	1498	endomondo
9	1438	cute
10	1323	instagram
11	1318	rt
12	1232	bostonmarat...
13	1201	friends
14	1201	1
15	1170	music
16	1096	jobs
17	1080	food

Fig. 2. *SocialOcean*: Visual Analysis and Characterization of Social Media Bubbles (SMBs). The *Mention Graph* (1) highlights social networks where people mention and/or follow each other. There are two Mentions Graphs: Overview and Detailed. (1) the figure shows the Detailed Mention Graph. The edges of the graph embed the sentiment associated to the graph structure. The map view (2) provides information about the distribution of the social networks that can characterize SMBs. The full-text search (3) and the Interactive Filter Space (4) are the starting point for the analysis of a topic or content of interest. The histogram (5) with the topic distribution shows the average sentiment using a divergent colormap ranging from positive to negative (6). The timeline provides temporal information to the analyst about the SMB in the analysis. In this example, the graph shows a SMB with a diverse range of sentiments and a central user. This diversity could be evidence of some discussion or divided opinions in-between the bubble.

these algorithms cannot be used directly to identify social media bubbles, which reflects the community in multiple dimensions. Therefore, appropriate methods must be used which take into account the characteristics of social media bubbles.

C. Geo- and Semantic- Visual Analysis in Social Media

Geo-tagged social media data reflects the spatial and temporal distribution of the data. It can be used to identify event distribution [12], people's movement [15] and geo-spatial information diffusion [9]. Scatterblogs combine the event detection and classification in investigating the geo-tagged social media, to enable situation awareness [50]. Together with sentiment and geo information, Li et al. proposed VisTravel to understand the users' traveling patterns with sentiment [33]. According to Chen et al.'s survey [13], keywords, topic and Sentiment analysis are three perspectives in semantic analysis. Xu et al. visualized topic competition with river-like metaphors [55]. With the similar river metaphor, FluxFlow conducted sentiment analysis and visualization in anomaly analysis [58]. Dou et al. integrated other data sources such as demographic information to identify the grouping patterns in social media users [19]. To address on sentiment, MultiConVis visualized the particular users comments connecting to global topics [28]. There are related works addressing on public sentiment [8, 54]. SentiView addressed on the sentiment divergence between user communities over time in Twitter [54]. SocialHelix visualized the sentiment divergence between user communities over time with a DNA-like visual metaphor [8]. However, these works did not consider combining networks, sentiment analysis, geographic distribution visualization, temporal and thematic analysis to detect SMBs. Our focus is to identify and analyze the Social Media Bubbles with multiple perspectives.

D. Analysis of Social Media Bubbles

A well-known phenomenon that is often associated with social networks are the so-called echo chambers, which is used synonymously for various phenomena that occur in social networks. However, the term predates the modern social networks and was used as early as 1990 [46] and generally refers to "an environment in which a person encounters only beliefs or opinions that coincide with their own, so that their existing views are reinforced and alternative ideas are not considered" [16]. Echo chambers often occur in social networks, such as Facebook, Twitter, etc. and can be reinforced by the "filter bubble" effect, which describes how technology can bias the exposure of certain groups to new information. A prominent example are social networks, which are increasingly personalizing content with the help of recommender systems and machine learning [22], creating social media bubbles in which individuals are largely exposed to conforming opinions. Social Media Bubbles can tie together people from different opinions or beliefs about a certain topic, but still connected by certain influential users. To better differentiate between general echo chambers and these new phenomena of social networks that present a tendency in social media, the term "social media bubble' was coined by Nikolov et al. [39] in 2015. To identify social media bubbles, one must not only consider the connectivity of persons in social networks, but also many other sociodemographic, behavioral, and intrapersonal characteristics. An overview of these characteristics is given by McPherson et al. [36] and includes, but is not limited to: age, gender, education, social class, network position, geography, sentiment, and behavior. McPherson et al. argue, that people only have significant contact with people who are homogeneous regarding these characteristics and that interacting with people that share the same affinities, reinforces their existing views.

Many social media platforms, such as Twitter, Facebook, etc., provide access to the aforementioned features, such as geolocation or sociodemographic characteristics, such as age, gender, or education. Additional information, such as behavioral or sentiment characteristics can be extracted using text analysis methods. For instance, what a person is talking about using topic modeling [1], how a person is talking about something using sentiment analysis [49], or how credible a person is [11].

The identification and analysis of social media bubbles is a timely and important topic. Additionally, Chen et al. [13] pointed out in their recent state-of-the-art report on social media visual analytics, that this topic is currently not highly-researched. As pointed out in section II-D, one must consider many sociodemographic, behavioral, and intra-personal characteristics to identify social media bubbles, which can also be dependent on topic, community, and timeframe. Chen et al. identified only three works, that addressed the necessary data to identify social media bubbles, namely WeiboEvents [42], Whisper [10] and the time-varying visual analysis of microblog sentiment by Zhang et al. [57]. However, none of these applications are directly designed for the analysis of social media bubbles. Thus, in order to enable experts to manage this complex task and to fill the gap in the current research, we propose a novel visual analysis-based approach for the characterization, identification, and examination of social media bubbles.

III. SocialOcean Approach for Visual Analysis of Social Media Bubbles

The detection and investigation of social media bubbles and echo chambers demands a deep analysis task. We provide the users with several analytical views following the Coordinated Multiple Views (CMV) architecture proposed by Roberts et al. [44]. Each view of the system contributes towards the definition and description of the social bubbles. Figure 2 shows a snapshot of the tool. The Visual Analytics workflow is divided into four different steps: (1) data preprocessing (cleaning, and filtering), (2) natural language processing methods for the topics classification and sentiment analysis, (3) visualization of temporal, topical, geospatial and network properties, and (4) interaction. Next, we describe these steps in more detail and evaluate them afterwards.

A. Data Preprocessing

We focus on Twitter media communication, one of the most popular micro-blogging services worldwide. We collected around 10% of the worldwide Twitter stream data for two case studies: (1) We review 10 days of data, surrounding the Boston Marathon Bombing, from April 14 2013 until the April 24, 2013, and (2) a global warming controversy which was initialized by a statement of President Donald Trump. We collected data from November 10 until November 25, 2016. We set up a preprocessing workflow in KNIME [32] for our Twitter data. It incorporates URL removal, abnormal content analysis, Tweet source analysis, language filtering, and credibility scoring. We also analyzed other dimensions of homophily presented in McPherson et al. [36]. For the geocoding, we use different levels of granularity. For tweets with geolocation metadata we used the provided coordinates, otherwise, we used the user's location.

B. Topic Modeling and Sentiment Analysis

Due to the informal and specialized language used in tweets and their short message format, tools and, techniques which work well for classical text media, such as news, perform quite poorly when applied to tweets [43]. To tackle this problem and to enable a topic modeling for tweets, we employ a hierarchical feature subset selection algorithm, as proposed by Fiaidhi et al. [21]. We used the *LingPipe* [1] library, to classify our tweets into one of the 12 categories *music, news media, family, health, pets, education, marketing, recreation-sports, politics, food, computers-technology, other,* using the default n-gram size of 5 and selecting the topic with the highest probability. Our tool can be used to generate training data for future improved topic detection.

To determine the sentiment of the processed tweets, we employed the SentiStrength library of Thelwall et al. [49]. It estimates the strength of positive and negative sentiment in short texts, even for informal language. The algorithm is based on a series of lookup tables and is also able to deal with some domain-specific language, including emojis, booster words, negations, and even irony terms. It approaches human-level accuracy in most tested cases [48], thus making it an appropriate choice for the sentiment analysis of tweets.

C. Visual Design of SocialOcean

The visual design of our tool integrates graph visualization with the sentiment analysis embedded in the graph structure and combine it with an overview of the geographic, temporal, and thematic attributes, as shown in Figure 2. The main visual components are the Mention Graphs Overview, the Detailed Mention Graph (see Figure 2.1 showing the Detailed Mention Graph), and the map view (Figure 2.2). Furthermore, it comprises a full-text search (Figure 2.3), the interactive search space (Figure 2.4), a topic histogram (Figure 2.5), a timeline (Figure 2.6), and a settings panel (Figure 2.7). These components allow the user to set up a particular context consisting of hashtags, topics, content, sentiments, and other addressed dimensions of homophily.

We designed a tasks workflow for exploratory data analysis following the visual analytics mantra [31]. The principal tasks can be reduced to:

- **T1**. Selection of the starting topics of interest.
- **T2**. Overview and steering of the results.
- **T3**. Analysis of the internal structure of the SMBs.
- **T4**. Shrinking or expanding the current SMBs for further analysis.

The first step (T1) in the analysis consists of the selection of interesting hashtags and topics that could reveal and characterize SMBs. Once, those input parameters are selected, the system displays the results accordingly, using four different main overviews: a mention graph overview, a map overview, a timeline, and a topics histogram. We chose these four overview visualization to cover what we considered the most prominent aspects of identifying SMBs: time, geolocation, connectivity, and topics. The second step (T2) comprises an explanatory data analysis of the four overviews above. The user can zoom and filter, and add and fuse information, to steer each one of the data views and ”shape” the SMBs s/he is interested in. In the third step (T3), the user analyzes details of the selected SMB. For example, the distribution of sentiments along the SMB, gender distribution, the content details of the tweets, the connectivity topology, key users such as hubs and gatekeepers, etc. We used the normalized betweenness- and degree centrality, as described in [**Himelboim2017**], to quantify the number of connections that can lead to potentially influential users. In this last step (T4), the user can continue refining and shaping the identified SMBs or adding new topics and hashtags to expand the SMB boundaries. This tasks could lead to new findings wrt to the interrelation between SMBs.

1) Mention Graphs: SocialOcean provides two different levels of detail for the analysis of the graph structure. One general mention graph with undirected edges and one detailed-graph with directed edges (see Figure 3).

Mention Graph Overview. This visual component shows the Mention Graph Overview distribution. We use a uniform layout from the JUNG graph library[29] to help the users to identify salient subgraphs. The user can select a particular graph according to the betweenness or degree centrality measures, the density of nodes, or particularity of the structure. Using this view, the user can highlight interesting edges, nodes, and select whole graph structures. We incorporated normalized betweenness and degree centrality measures which range between 0 and 1 and two radio buttons so that the analyst could switch between them. The user can select particular graph structures or find outstanding users according to the betweenness or degree centrality measures that we have incorporated. Degree centrality quantifies the number of connections and reflects potentially influential users. Betweenness centrality describes how often a node lies along the shortest path between two other nodes. Nodes with high betweenness centrality values are bridging otherwise separate parts and are therefore interesting for our SMB analysis. Selecting key persons or whole groups of interest could reveal more details in other parts and create the detailed graph. Selected edges are highlighted in yellow, nodes are highlighted as well and labeled by the user’s Twitter screen-name. Dense graphs are colored red. The graph density is calculated using the density formula for simple undirected graphs:

$$D_{ensity} = \frac{2|E|}{|V|(|V|-1)}$$

Detailed Mention Graph. The Detailed Mention Graph maps every mention in a tweet to a directed edge, see Figure 3. The edge color represents the sentiment by default. Every edge is labeled according to its positive and negative SentiStrength [45] values.

2) Map: The map view components visualize a layer of tweets using two visual encodings: (a) a heatmap, and (b) glyphs. Similar to MacEachren et al. [35] we also differentiated between tweet and user locations. Locations extracted from the tweet metadata or content are colored in blue, while given user locations are colored in red, as shown in Figure 4.

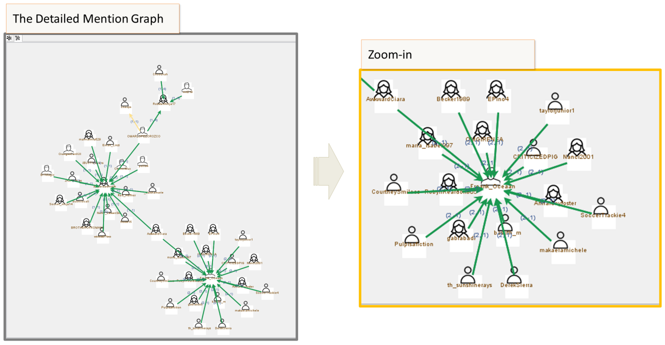


Fig. 3. After selection from the general graph, a detailed graph is created. The colored directed edges show the direction, the sentiment, and the SentiStrength value information. Icons for the nodes present the user’s gender.

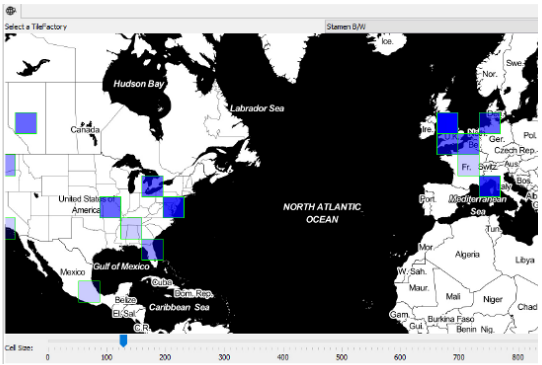


Fig. 4. Our map visualization enables viewing user and tweet locations at once, on a country level or in a variable grid-based heatmap, by merging colors of overlapping regions.

D. Details-On-Demand

SocialOcean provides ancillary view components to display details-on-demand for the SMB and individual tweets and users. For example, a tweet or a user will open a pop-up window with the corresponding detail and metadata.

Topic Histogram. This visual component shows the topic distribution of the 12 news categories. Additionally, it shows the average sentiment of tweets within a category using a divergent colormap from positive 🟩🟩🟩🟨🟨🟧🟧🟥🟥 to negative. The main goal is to provide the analyst with an overview of the topic diversity and to reveal emotional states associated with those topics.

Timeline. This component aggregates the amount of tweets by hour and thus, it visualizes information about the temporal topic spread. Peaks reveal topic bursts. The analyst can filter for a given time range by dragging the mouse over the desired span, see Figure 5. The precision of the time adapts to the visualized time span.

Settings and System Feedback In a settings part the analyst can switch the color scheme of the tweets from sentiment and SentiStrength to categories. Further, the user can decide if

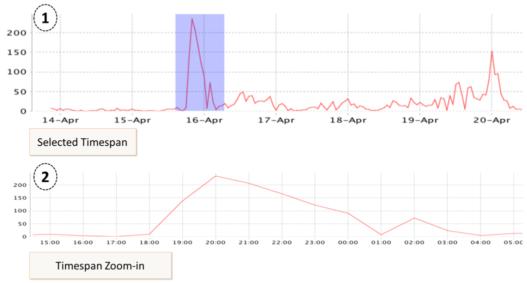


Fig. 5. The timeline shows a temporal peak at April 16, 2013. (1) Shows time range selection by the user, and (2) is a zoom-in of the timeline for the selected time span. The precision of the time changes to hours, if the selection is smaller than a day.

he wants the tweets- and/or user-locations to be visualized in the map, either in a heatmap or in colored country shapes. A console view prints and save the logs of the current system states and interaction feedback allowing for provenance of the system.

E. Implementation

The system is implemented using a client-server architecture. The back-end consists of a KNIME Workflow that pre-processes the data and stores it in a PostGreSQL database with PostGis extension[41]. The data is indexed using Lucene [3]. We chose the Lucene database for text queries in the application because it is a powerful platform for text querying. We used LingPipe for topic classification, SentiStrength for sentiment analysis, and Apache Tika Language Detection for language detection. The client is implemented in Java as an Eclipse e4 RCP project for plugin development.

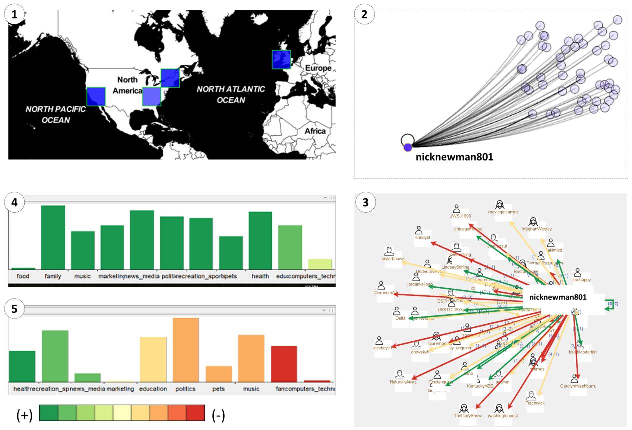


Fig. 6. Findings of the first scenario (BMB). (1) Geographic distribution of the SMBs. (2) Selected SMB at the beginning of the analysis. (3) Extended SMB with more tweets of the user:*nicknewman801*. (4) and (5) topic distribution of the initial and extended SMB, respectively.

IV. Use Cases

The goal of the present use cases is to illustrate the capabilities of our approach. We use the proposed tasks workflow to structure the scenarios and showcase the use of the proposed VA system, *SocialOcean*.

A. Boston Marathon Bombing (BMB)

The first scenario takes place in the time frame of the Boston Marathon Bombing from April 14 to April 24, 2013. We selected this time frame because there is vast research done around the BMB. Our hyphothesis is that events can promote the expression of SMBs or even create SMBs around them. In this scenario, the analyst wants to analyze SMBs whose users are interested in providing help or first aid to the people affected by BMB.

T1. S/he looks at the Interactive Filter Space and finds the hashtag #prayforboston at rank 2 in the ranking list to start the analysis.The user explores more connections adding the hashtag #bostonmarathon, with rank 12 in the ranking list. The selected hashtags reveal the major event of the BMB happening at that time frame. **T2**. The user is interested in the analysis of SMBs which exhibits a positive sentiment . The user uses the sentiment to identify SMBs in favor or with positive bias to help. **T3**. S/he selects the content:fundraising, content:donation, and content:help. **T4**. The user filters tweets that expose the geolocation to analyze the geographic distribution world wide, and have a qualitative estimation of the amount of geolocated SMB. The resulting map view shows how the SMBs are spreading geographically mainly in USA, Canada, and Europe (see Figure 6.1). **Further Analysis** The user selects an SMB from the Mention Graph Overview and visualizes it in detail. The distribution of the sentiment across the SMB structure is predominately positive. All the views are linked and show the results filtered for the selected SMB.

The analyst looks for more information about the user with more mentions. The results are displayed in Figure 6. A new SMB appears where the user *nicknewman801* appears central. Further analysis could be done to divide or isolate bi-polar SMBs for this specific user: mostly positive and mostly negative, filtering by sentiment.

B. Global Warming

The second scenario takes place in 2016, from November 10 to November 24, 2016, when US president Donald Trump tweeted: ”The concept of global warming was created by and for the Chinese in order to make U.S. manufacturing non-competitive”. In this scenario, the analyst is interested in exploring SMBs related to another influential user, US President Donald Trump, and trending hashtags and content such as China Hoax and Conspiracy Theory. Figure 7 shows the results of the case study. **T1**. To start the investigation, the user selects the hashtags: #Trump, #China, #hoax, #Globalwarming, and #climatechange. The user selects the tweets classified as positive. **T2**. S/he analyzes the geographic distribution of the SMB, selecting only the connections that are geolocated. **T3**. S/he picks a subgraph from the Mentions Graph Overview, identified as potential SMB (see Figure 7.2). The SMB is significantly small and is distributed mostly between USA and Europe (see Figure 7.1).

T4. Afterwards, the user filters the content by *conspiracy theory* to analyze changes on the selected SMB. The SMB is expanded, the connectivity between the SMB increased significantly (see Figure 7.3). The topic histogram shows a slightly more diverse distribution of tweets per category and sentiments (see Figure 7.4).

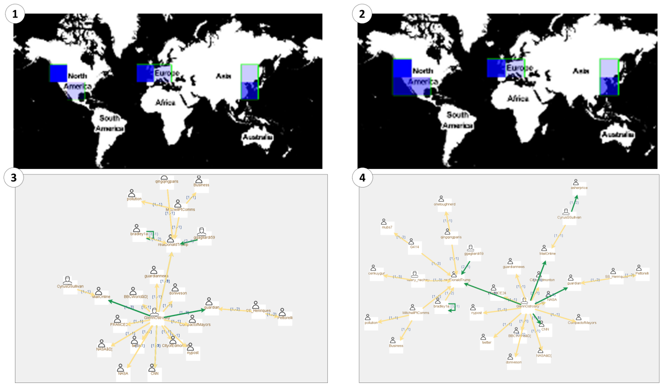


Fig. 7. Results of the second scenario. (1) Geographic distribution and topic histogram of the SMB initially studied. (2) Detailed Mention Graph of the SMB. (3) Geographic distribution and topic histogram after adding tweets with content *conspiracy theory*. (4) The same SMB extended after the addition of the new tweets.

V. Evaluation

We conducted a preliminary expert evaluation that consisted in a semi-structured pre- and post-interview and a paired analytics session with a total of five experts. The interviews were recorded, encoded, and made available as appendix material to this paper. Two experts were females, two were males, and one of them chose to not disclose this information. Their ages range from 24 to 40 years and all of them have international background. They were four political scientists ranging between four to nine years of experience in the field, and a journalist with more than 15 years of experience.

A. Semi-Structured Pre-Interview

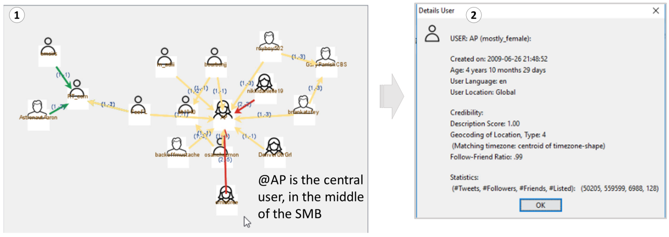
During the semi-structured pre-interview we gathered some important information about their background and familiarity with the terms: SMB, Echo chambers, Filter Bubbles, and Personalization. All our participants had some previous experience using social media such as Facebook, Twitter, Instagram, LinkedIn, and WhatsApp. They checked the news regularly everyday, in different formats, two of the participants read printed newspapers, but most of them read Internet news. They were all interested in domestic, national, and international news. Four participants had previous experience with the term SMBs and Echochambers. Some of them confused both terms or considered synonyms. Most of them except for the last one, were not familiar with the definitions Filter Bubbles and Personalization. Only three participants had a general knowledge about bias but were not familiar with the term *confirmation bias*.

B. Paired Analytics

The objective of the paired analytics session was to let the participant hypothesize and characterize possible SMBs in Twitter, using *SocialOcean*. The findings and feedback of the participants were gathered using a *Think-Aloud* approach. Each participant could select freely which data set to use and the starting strategy for the exploratory analysis.

The first participant (P1) proposed the following scenarios and research questions. (1) What are the tweets rated as positive in the context of the Boston Marathon Bombing (BMB)? (2) What are the SMB related to a specific user? and (3) What are the SMBs that expose a predominant positive sentiment on the mention graph?

She chose a positive group and explored how it was organized. She looked at the people in the middle of the groups that could be influential users. The participant was interested in seeing if phenomena, for example, *positive attracts positive* and *negative attracts negative* could shape the bubbles. She searched for an influencer, in this case, US ex-president Obama to start the analysis. The system returned the SMBs where the US ex-president Obama was mentioned. She picked a group and analyzed the sentiment across that group. The participant stated: *It is interesting to see how different emotions change in-between the group*. The results are shown in Figure 8. The participant provided us with suggestions for improvement, such as attaching pictures to the tweets and including view components that could show the evolution of the SMBs over time, and additional quantitative information.



@AP is the central user, in the middle of the SMB

USER: AP (mostly_female):

Created on: 2009-06-26 21:48:52
Age: 4 years 10 months 29 days
User Language: en
User Location: Global

Credibility:
Description Score: 1.00
Geocoding of Location, Type: 4
(Matching timezone: centroid of timezone-shape)
Follow-Friend Ratio: .99

Statistics:
(#Tweets, #Followers, #Friends, #Listed): (50205, 559599, 6988, 128)

Fig. 8. P1 searched for an influencer US President Obama. (1) The participant identified a group and analyzed the sentiment across the group. S/he found interesting how the sentiment spread along the graph. (2) A detailed view on the central user reveals that it is a journalist or broadcasting channel. In this case, the user or broadcasting channel is the interest that connects the SMB.

The second participant (P2) was interested in the topics: corruption, conflict, conspiracy, Trump, and climate change. S/he mostly used the query search and looked at the climate change data sets. S/he explored SMBs for #Climatechange and identified a bubble where Trump was in the middle, but the network has two influential users, one Trump from USA, the other Glenn Ostrosky from Europe. The bubble was distributed along USA and Europe. Afterwards, S/he added the hashtag: Corruption. S/he could find the same SMB, extended with more connections, again from USA and Europe. The same influencer in the middle. The participant found interesting that the tool indicates with color the sentiment in the detailed graph. As a conclusion of this part of the interview, P2 said that there are some members of the SMB that could have contrary beliefs, but they want to be in the bubble to follow what this group is talking about. Figure 9 shows some of the findings of P2.

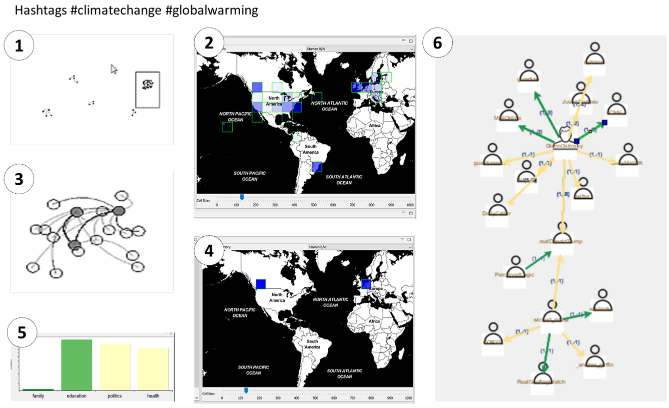
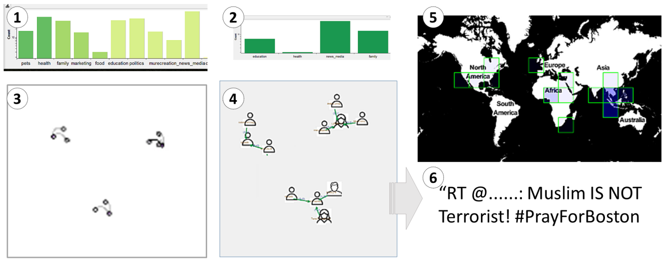


Fig. 9. The figure shows the characterized SMB that was found by the second participant (1) and (3).The map above shows the tweets distribution before selecting the SMB (2). The map below shows how the SMB is distributed on the map, specifically, between Canada and Europe (4). For the sentiment distribution it could be an SMB of Trump's supporters (6). A further analysis on topics classification (5) could reveal more details about the affinities or shared interests.

The third participant (P3) selected the specific case of the BMB. P3 wanted to investigate SMBs interested in muslims and Islam. The participant used the full-text query search to select hashtags and content related to these two topics. P3 wanted to identify groups that approve or disapprove muslims. She found mainly positive sentiments or neutral dominated the social network structures of the SMB, for this analysis. The results of the experiment are shown on Figure 10.



“RT @......: Muslim IS NOT Terrorist! #PrayForBoston

Fig. 10. The figure shows the results of the hashtags muslims and content Islam. Several small SMBs were identified by the participant. (1) shows the topic distribution of tweets. (2) shows the selection of a specific SMB, and the positive tendency, even more pronounced. (3) and (4) shows the mention graph overview and detailed view of SMBs and (5) the map view shows the geographic distribution of SMBs. The SMBs seems to be distributed mostly in US, Europe, and Asia. A positive trend can be observed along the topic histogram and in the Detailed Mention Graph.

The fourth participant (P4) wanted to characterize an SMB in the context of the BMB. Initially, he started his analysis looking at trending hashtags like police, traffic, radios, related to the event or that could appear as a result of the event. As part of the *Think-aloud* approach, the participant could define two types of SMB he wanted to investigate: (1) People that are sorry for what happened and want to communicate their sentiment and transmit it to families of victims, (2) other kind of SMBs that could be related to an anger response to terrorism.

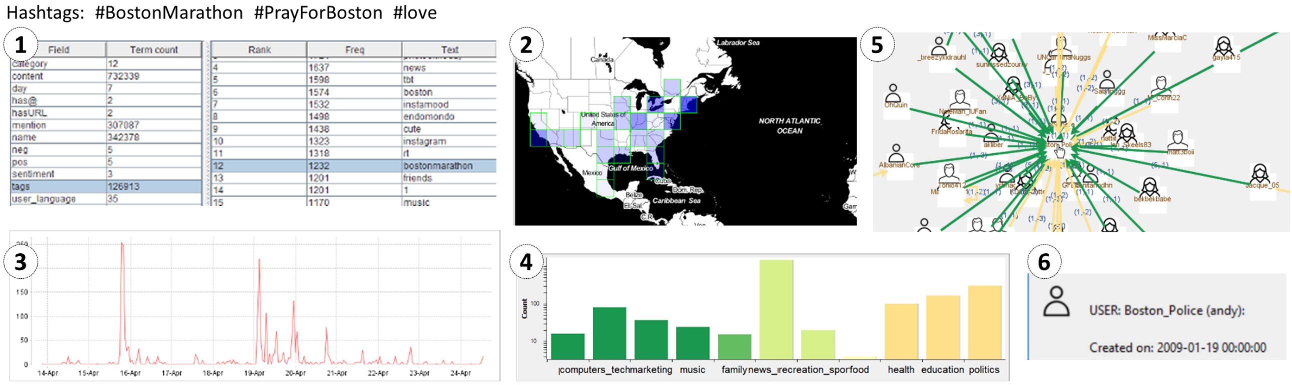
For the first type of SMBs, the participant investigated the hashtags and content: #PrayForBoston, #BostonMarathon, #love. Afterwards, he filtered the SMBs by positive sentiment. He found it interesting that there were bubbles that have no connection at all to the event. P4 would have expected that Twitter is more focused on famous journalists or influencers and that they would be the main players on the SMBs. Figure 11 shows some of the findings from the interview. Our last participant (P5) selected the datasets of the tweeted event about President Trump and Global Warming. P5 used the full-text query search to investigate the hashtags #Trump and #BernieSanders and the content China. He was interested in visualizing only the geolocated SMBs. P5 found interesting a particular SMB (see Figure 12.1) that contains two main groups connected by two influential users, President Trump and another politician, Paul Mitchell. The participant found interesting that Trump is mostly mentioned with neutral sentiments. He concluded that the selected SMB should consist of followers of Trump (see Figure 12).

C. *Semi-Structured Post-Interview*

We performed a post-interview session to collect information about the usability of the tool and general feedback. Among the most useful dimensions they found: connectivity in-between the SMB and its graph structure, geographic distribution, influential users, sentiment analysis, and hashtags. They also mentioned with lesser importance dimensions such as language, gender, and temporal peaks. From this results, we learned that tasks workflow was useful guidance for the participants to explore the system. Most of the designed components showed to be useful. In particular, the topic categories were less attractive than the hashtags, and the full-text search query was the preferred one for most of the participants to start the analysis.

VI. Discussion and Future Work

Our evaluation demonstrated the effectiveness of *SocialOcean* in characterizing SMBs. The results are positive and promising because all the participants were able to identify several SMBs in different contexts and with different analytical goals. The research questions defined at the beginning of the paper were assessed during the design process and evaluation. Regarding R1, we observed that the participants used all the dimensions and components provided by the system, though they weighted their utility differently. We hypothesize that their background has influenced their interaction with the tool, exploration, and preferences in terms of the topics and datasets. The bias is an intrinsic aspect of the phenomena in the study. We considered the bias as part of the context of the exploratory analysis. These observations also reaffirmed our claims of SMB as a fuzzy and ill-defined phenomenon where the inclusion of the user is essential, and therefore a visual analytics approach is particularly promising and required. To address R2, we encoded the sentiment in several views: graph, map, and topic histogram. Most of the participants emphasized how useful it was for them to see different emotions or sentiments distributed or structured in-between the graph. Some of them related the sentiment



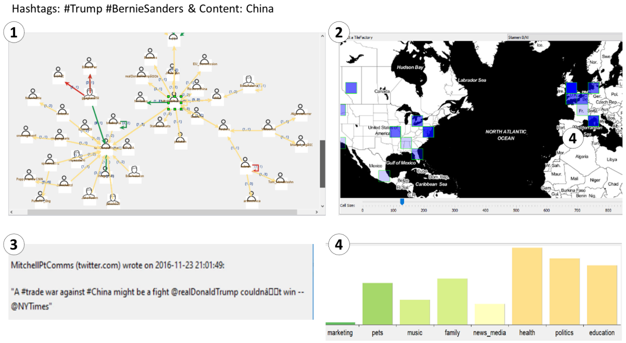
Field	Term count
category	12
content	732339
day	7
has@	2
hasURL	2
mention	307087
name	342378
neg	5
pos	5
sentiment	3
tags	126913
user_language	35

Rank	Freq	Text
4	1637	news
5	1598	tbt
6	1574	boston
7	1532	instamood
8	1498	endomondo
9	1438	cute
10	1323	instagram
11	1318	rt
12	1232	bostonmarathon
13	1201	friends
14	1201	1
15	1170	music

USER: Boston_Police (andy):

Created on: 2009-01-19 00:00:00

Fig. 11. Findings of P4. (1) Ranking list. (2) Map distribution of the SMB. (3) Timeline showing peaks around the time of the bomb. (4) Topic distribution. (5) Central user: Boston_Police. (6) Details of the central user.



MitchellPtComms (twitter.com) wrote on 2016-11-23 21:01:49:

"A #trade war against #China might be a fight @realDonaldTrump couldnâ□□t win -- @NYTimes"

Fig. 12. Findings of P5. (1) The participant identified a bi-polar SMB. The SMB contains two people that he could call leaders. He could possibly identify a reciprocal relationship. (2) The map shows that it is a phenomena covering mostly North America, Europe and Asia. The southern hemisphere does not contain tweets. (3) The timeline shows a temporal peak in coincidence with Trump sayings on the November 16, 2016. (4) The topic histogram shows that the most important trending topics were health, politics, and education.

structure to the *behavior* of the group and the individuals in the group. In the case of R3, there were some evidence of SMBs linked to the proposed events during the time frame of the study, but this findings require further investigation along different time frames. Also, polarized opinions were addressed by a combination of graph connectivity and sentiment analysis, easily distinguishable in the graph structures. These views can be combined with the map and timeline to extend the analysis of polarization in other dimensions of SMBs.

One of the limitations of our approach is that for certain cases, machine learning algorithms perform poorly, misleading the analysis. This happened particularly with the topic classification and corresponding histogram component. This could be caused by the noised characteristics of Twitter data. We argue that the mismatches and back and forward could confuse the analyst and affect the trust-ability on the topic histogram. These mismatches could be the reason why the participants found it the less attractive component. The same reasoning could justify the divided results w.r.t. the sentiment analysis. In future work, we will incorporate new algorithms for topic classification and sentiment analysis. Future work will also include other uncertainty measures, such as content quality, thematic accuracy, and multi-resolution geospatial accuracy, among others. The domain experts that evaluated our tool asked us to add more quantitative views and functionalities to compare SMBs. We will foster collaborations with some of the domain experts that showed interested in our work.

VII. Conclusions

In this paper, we study the problem of characterizing the phenomena of Social Media Bubbles. We provided a typical tasks workflow for the exploratory analysis of this kind of phenomena in social media analytics. We presented *SocialOcean*, an interactive visual analytics system that combines social network visual analytics, sentiment analysis, and map visualization as a backbone for the characterization of SMBs. Our paper is the first work to the best of our knowledge that combines a graph with the sentiment analysis embedded in the structure, the geographic distribution, topic modeling, and other social media measures to characterize SMBs. We evaluated with domain experts that showed that the tool could be efficiently used to detect, refine and describe in detail SMBs of interest. All the participants were able to find SMBs and emphasized the potential of the tool.

Acknowledgment

The research presented in this paper has been supported by the German Research Foundation DFG within Priority Research Program 1894 (DFG SPP VGI). The authors would like to thank Lukas Kirchner for facilitating us the Twitter data.

References

[1] Alias-I. *LingPipe*. http://alias-i.com/lingpipe/index.html.

[2] Natalia Andrienko et al. “Visual analytics for understanding spatial situations from episodic movement data”. In: *KI-Künstliche Intelligenz* 26.3 (2012), pp. 241–251.

[3] *Apache Lucene Index*. https://lucene.apache.org.

[4] Daniel Archambault et al. "ThemeCrowds: Multiresolution Summaries of Twitter Usage". In: *Proceedings of the 3rd International Workshop on Search and Mining User-generated Contents*. SMUC '11. Glasgow, Scotland, UK: ACM, 2011, pp. 77–84. ISBN: 978-1-4503-0949-3. DOI: 10.1145/2065023.2065041. URL: http://doi.acm.org/10.1145/2065023.2065041.

[5] Eytan Bakshy, Solomon Messing, and Lada A Adamic. "Exposure to ideologically diverse news and opinion on Facebook". In: *Science* 348.6239 (2015), pp. 1130–1132. DOI: 10.1126/science.aaa1160.

[6] Vincent D Blondel et al. "Fast unfolding of communities in large networks". In: *Journal of statistical mechanics: theory and experiment* 2008.10 (2008), P10008.

[7] Anthony Brew et al. "Deriving Insights from National Happiness Indices". In: *Proceedings of the 2011 IEEE 11th International Conference on Data Mining Workshops*. ICDMW '11. Washington, DC, USA: IEEE Computer Society, 2011, pp. 53–60. ISBN: 978-0-7695-4409-0. DOI: 10.1109/ICDMW.2011.61. URL: https://doi.org/10.1109/ICDMW.2011.61.

[8] Nan Cao et al. "SocialHelix: visual analysis of sentiment divergence in social media". In: *Journal of Visualization* 18.2 (2014), pp. 221–235.

[9] Nan Cao et al. "Whisper: Tracing the Spatiotemporal Process of Information Diffusion in Real Time". In: *IEEE Trans. Vis. Comput. Graph.* 18.12 (2012), pp. 2649–2658. DOI: 10.1109/TVCG.2012.291. URL: http://doi.ieeecomputersociety.org/10.1109/TVCG.2012.291.

[10] Nan Cao et al. "Whisper: Tracing the spatiotemporal process of information diffusion in real time". In: *IEEE Transactions on Visualization and Computer Graphics* 18.12 (2012), pp. 2649–2658.

[11] Carlos Castillo, Marcelo Mendoza, and Barbara Poblete. "Information credibility on twitter". In: *Proceedings of the 20th international conference on World wide web*. ACM. 2011, pp. 675–684.

[12] Junghoon Chae et al. "Spatiotemporal social media analytics for abnormal event detection and examination using seasonal-trend decomposition". In: *2012 IEEE Conference on Visual Analytics Science and Technology, VAST 2012, Seattle, WA, USA, October 14-19, 2012*. 2012, pp. 143–152. DOI: 10.1109/VAST.2012.6400557. URL: http://dx.doi.org/10.1109/VAST.2012.6400557.

[13] Siming Chen, Lijing Lin, and Xiaoru Yuan. "Social media visual analytics". In: *Computer Graphics Forum*. Vol. 36. 3. Wiley Online Library. 2017, pp. 563–587.

[14] Siming Chen et al. "D-Map: Visual Analysis of Egocentric Information Diffusion Patterns in Social Media". In: *Proc. of IEEE Visual Analytics Science and Technology*. 2016, pp. 41–50.

[15] S. Chen et al. "Interactive Visual Discovering of Movement Patterns from Sparsely Sampled Geo-tagged Social Media Data". In: *Visualization and Computer Graphics, IEEE Transactions on* PP.99 (2015), pp. 1–1. ISSN: 1077-2626. DOI: 10.1109/TVCG.2015.2467619.

[16] *Definition of echo chamber in English by Oxford Dictionaries*.
https://en.oxforddictionaries.com/definition/echo_chamber, journal=Oxford Dictionaries.

[17] Dominic DiFranzo and Kristine Gloria-Garcia. "Filter bubbles and fake news". In: *XRDS: Crossroads, The ACM Magazine for Students* 23.3 (2017), pp. 32–35.

[18] Wenwen Dou et al. "Leadline: Interactive visual analysis of text data through event identification and exploration". In: *Visual Analytics Science and Technology (VAST), 2012 IEEE Conference on*. IEEE. 2012, pp. 93–102.

[19] W. Dou et al. "DemographicVis: Analyzing demographic information based on user generated content". In: *Visual Analytics Science and Technology (VAST), 2015 IEEE Conference on*. Oct. 2015, pp. 57–64. DOI: 10.1109/VAST.2015.7347631.

[20] Xiaolin Du et al. "OpinionRings". In: *Decis. Support Syst.* 75.C (July 2015), pp. 11–24. ISSN: 0167-9236. DOI: 10.1016/j.dss.2015.04.007. URL: https://doi.org/10.1016/j.dss.2015.04.007.

[21] Jinan Fiaidhi et al. "Developing a hierarchical multi-label classifier for Twitter trending topics". In: *International Journal of u-and e-Service, Science and Technology* 6.3 (2013), pp. 1–12.

[22] Seth Flaxman, Sharad Goel, and Justin M Rao. "Filter bubbles, echo chambers, and online news consumption". In: *Public Opinion Quarterly* 80.S1 (2016), pp. 298–320.

[23] Santo Fortunato and Darko Hric. "Community detection in networks: A user guide". In: *Physics Reports* 659 (2016), pp. 1–44.

[24] Eric Gilbert, Tony Bergstrom, and Karrie Karahalios. "Blogs are echo chambers: Blogs are echo chambers". In: *System Sciences, 2009. HICSS'09. 42nd Hawaii International Conference on*. IEEE. 2009, pp. 1–10.

[25] W. Hall, R. Tinati, and W. Jennings. "From Brexit to Trump: Social Media's Role in Democracy". In: *Computer* 51.1 (Jan. 2018), pp. 18–27. ISSN: 0018-9162. DOI: 10.1109/MC.2018.1151005.

[26] Jeffrey Heer and Danah Boyd. "Vizster: Visualizing online social networks". In: *Information Visualization, 2005. INFOVIS 2005. IEEE Symposium on*. IEEE. 2005, pp. 32–39.

[27] Jeffrey Heer and Danah Boyd. "Vizster: Visualizing online social networks". In: *Information Visualization, 2005. INFOVIS 2005. IEEE Symposium on*. IEEE. 2005, pp. 32–39.

[28] Enamul Hoque and Giuseppe Carenini. "MultiConVis: A Visual Text Analytics System for Exploring a Collection of Online Conversations". In: *Proceedings of the 21st International Conference on Intelligent User Interfaces*. IUI '16. Sonoma, California, USA: ACM, 2016, pp. 96–107. ISBN: 978-1-4503-4137-0. DOI: 10.1145/2856767.2856782. URL: http://doi.acm.org/10.1145/2856767.2856782.

[29] *Jung*. http://jrtom.github.io/jung/.

[30] Andreas M. Kaplan and Michael Haenlein. "Users of the world, unite! The challenges and opportunities of Social Media". In: *Business Horizons* 53.1 (2010), pp. 59–68.

[31] Daniel Keim et al. "Visual analytics: Definition, process, and challenges". In: *Information visualization*. Springer, 2008, pp. 154–175.

[32] *KNIME*. https://www.knime.org.

[33] Qiusheng Li et al. "VisTravel: Visualizing Tourism Network Opinion from the User Generated Content". In: *J. Vis.* 19.3 (Aug. 2016), pp. 489–502. ISSN: 1343-8875. DOI: 10.1007/s12650-015-0330-x. URL: http://dx.doi.org/10.1007/s12650-015-0330-x.

[34] Alan M MacEachren et al. "Geo-twitter analytics: Applications in crisis management". In: *25th International Cartographic Conference*. 2011, pp. 3–8.

[35] Alan M. MacEachren et al. "SensePlace2: GeoTwitter analytics support for situational awareness". In: *Visual Analytics Science and Technology (VAST), 2011 IEEE Conference on*. 2011, pp. 181–190. DOI: 10.1109/VAST.2011.6102456.

[36] Miller McPherson, Lynn Smith-Lovin, and James M Cook. "Birds of a feather: Homophily in social networks". In: *Annual review of sociology* 27.1 (2001), pp. 415–444.

[37] Mark EJ Newman. "Fast algorithm for detecting community structure in networks". In: *Physical review E* 69.6 (2004), p. 066133.

[38] Raymond S Nickerson. "Confirmation bias: A ubiquitous phenomenon in many guises." In: *Review of general psychology* 2.2 (1998), p. 175.

[39] Dimitar Nikolov et al. "Measuring Online Social Bubbles". In: *CoRR* abs/1502.07162 (2015). arXiv: 1502.07162. URL: http://arxiv.org/abs/1502.07162.

[40] Junbiao Pang et al. "Unsupervised web topic detection using a ranked clustering-like pattern across similarity cascades". In: *IEEE Transactions on Multimedia* 17.6 (2015), pp. 843–853.

[41] *PostGIS*. http://postgis.net.

[42] Donghao Ren et al. "Weiboevents: A crowd sourcing weibo visual analytic system". In: *Visualization Symposium (PacificVis), 2014 IEEE Pacific*. IEEE. 2014, pp. 330–334.

[43] Alan Ritter, Sam Clark, Oren Etzioni, et al. "Named entity recognition in tweets: an experimental study". In: *Proceedings of the conference on empirical methods in natural language processing*. Association for Computational Linguistics. 2011, pp. 1524–1534.

[44] Jonathan C Roberts. "State of the art: Coordinated & multiple views in exploratory visualization". In: *Coordinated and Multiple Views in Exploratory Visualization, 2007. CMV'07. Fifth International Conference on*. IEEE. 2007, pp. 61–71.

[45] *SentiStrength*. http://sentistrength.wlv.ac.uk.

[46] DAVID SHAW. *Where Was Skepticism in Media? : Pack journalism and hysteria marked early coverage of the McMartin case. Few journalists stopped to question the believability of the prosecution's charges.* http://articles.latimes.com/1990-01-19/news/mn-226_1_media-coverage, journal=Los Angeles Times, publisher=Los Angeles Times. Jan. 1990.

[47] Kate Starbird et al. "Rumors, false flags, and digital vigilantes: Misinformation on twitter after the 2013 boston marathon bombing". In: *iConference 2014 Proceedings* (2014).

[48] Mike Thelwall. "Heart and soul: Sentiment strength detection in the social web with sentistrength, 2013". In: *Cyberemotions: Collective emotions in cyberspace* (2013).

[49] Mike Thelwall et al. "Sentiment strength detection in short informal text". In: *Journal of the Association for Information Science and Technology* 61.12 (2010), pp. 2544–2558.

[50] Dennis Thom et al. "Spatiotemporal anomaly detection through visual analysis of geolocated Twitter messages". In: *Proc. IEEE PacificVis*. 2012, pp. 41–48.

[51] Corinna Vehlow, Fabian Beck, and Daniel Weiskopf. "The State of the Art in Visualizing Group Structures in Graphs". In: *Eurographics Conference on Visualization (EuroVis) - STARs*. Ed. by R. Borgo, F. Ganovelli, and I. Viola. The Eurographics Association, 2015. DOI: 10.2312/eurovisstar.20151110.

[52] Corinna Vehlow et al. "Visualizing the evolution of communities in dynamic graphs". In: *Computer Graphics Forum*. Vol. 34. 1. Wiley Online Library. 2015, pp. 277–288.

[53] K. Wade et al. "Identifying Representative Textual Sources in Blog Networks". In: *Proc. 5th International AAAI Conference on Weblogs and Social Media (ICWSM'11)*. 2011.

[54] C. Wang et al. "SentiView: Sentiment Analysis and Visualization for Internet Popular Topics". In: *IEEE Transactions on Human-Machine Systems* 43.6 (Nov. 2013), pp. 620–630. ISSN: 2168-2291. DOI: 10.1109/THMS.2013.2285047.

[55] Panpan Xu et al. "Visual Analysis of Topic Competition on Social Media". In: *IEEE Trans. Vis. Comput. Graph.* 19.12 (2013), pp. 2012–2021. DOI: 10.1109/TVCG.2013.221. URL: http://dx.doi.org/10.1109/TVCG.2013.221.

[56] *YOUR FILTER BUBBLE IS DESTROYING DEMOCRACY*. https://www.wired.com/2016/11/filter-bubble-destroying-democracy/.

[57] Chenghai Zhang, Yuhua Liu, and Changbo Wang. "Time-space varying visual analysis of micro-blog sentiment". In: *Proceedings of the 6th International Symposium on Visual Information Communication and Interaction*. ACM. 2013, pp. 64–71.

[58] Jian Zhao et al. "FluxFlow: Visual Analysis of Anomalous Information Spreading on Social Media". In: *IEEE Trans. Vis. Comput. Graph.* 20.12 (2014), pp. 1773–1782. DOI: 10.1109/TVCG.2014.2346922.