2018 IEEE International Conference on Big Data (Big Data)

FinanViz: Visualizing Emerging Topics in Financial News

Ngan V.T. Nguyen
Computer Science Department
Texas Tech University
Lubbock, TX, USA
ngan.v.t.nguyen@ttu.edu

Vinh T. Nguyen
Computer Science Department
Texas Tech University
Lubbock, TX, USA
vinh.nguyen@ttu.edu

Vung Pham
Computer Science Department
Texas Tech University
Lubbock, TX, USA
vung.pham@ttu.edu

Tommy Dang
Computer Science Department
Texas Tech University
Lubbock, TX, USA
tommy.dang@ttu.edu

***Abstract*—The explosion of social media has paved a way for big data in which entrepreneurs use this data to find out potential customers, market demands, individual behavior, thereby to improve existing products, to create new products according to users' need, or to analyze and evaluate financial risks. The challenges of the heterogeneity and fragmentation of data make it difficult for analysts to fully exploit the benefit of deluge information. Available statistical software lacks customization and address unknown research questions. This paper proposes *FinanViz*, a visual analytics tool for analyzing financial news on social media. The principal aim of *FinanViz* is to observe the dynamic behavior of terms/words over time along with their proximity to other terms/words. The tool provides an intuitive, interactive exploration of the financial topics and what events are emerging in which we would argue that it will give hints for financial marketers in the decision making process.**

***Index Terms*—Financial news, social media, coordinated multiple views, dynamic network, community detection, word clouds.**

I. Introduction

In the era of big data, the fast growth rate of data has transformed the way we consume information. From the early days of the internet, users were only able to view information available on a website; then an inquiry form was provided that allows users to communicate with the system's owners. Later on, the birth of mySpace [1] marked a turning point for social networking, unveiling new kinds of interactions and views (e.g., blogs, groups, friends, music, photos). Nowadays, we have witnessed a plethora of social media applications on the internet (e.g., Twitter, Facebook, Google++, YouTube, LinkedIn) that give users seamless communication channels.

There is no denying that social networking has a lot of benefits for the user, which makes the distance between people shorter, updating the news more and faster. The communication between people on the social network over time has created a huge amount of data unprecedented ever. This is a valuable asset for both policymakers and businesses. Policy advocates use this data source to analyze individual behaviors or group behaviors, thereby making policies and decisions relevant to society.

At the same time, entrepreneurs use this data to find out potential customers, market demands, individual behavior, thereby to improve existing products, to generate new products, or to analyze and evaluate financial risks. In line with this trend, the use of social networking in financial markets has been gaining the attention of researchers for the last 10 years as the number of scientific publications has grown exponentially (as shown in Fig. 1). By the end of 2018, there are about 1700 publications and this number is expected to increase sharply in the coming years (the dashed line in Fig. 1).

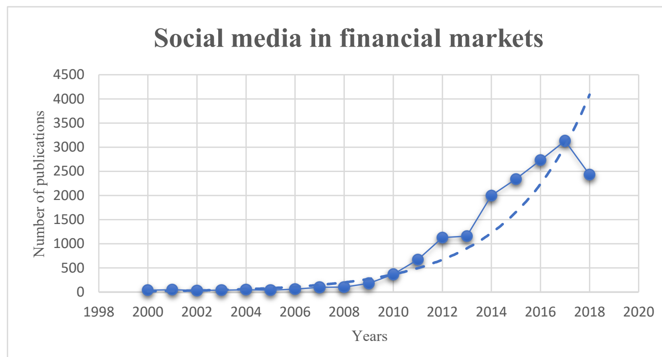

Fig. 1. Number of publications on social media in financial markets from Google Scholar

The explosion of social networking and big data has created a source of inspiration for researchers and a lever for deepening the development of financial markets. In addition to these potentials, there are also many challenges that have not been fully addressed. The first challenge is the heterogeneity of data (such as text, numbers, images, sounds, videos), in which there are many different types of structured data. Moreover, there is no uniform "standard" for these data structures, making it difficult for consolidating information. The second challenge is the fragmentation of the data, meaning that the data comes from different sources with different reliability. We have seen too many "fake news" on the web as a way of distracting competitors so data reliability should be properly considered.

In this context, many researchers from different disciplines have proposed a number of methods to partially address these challenges. For example, the application of mathematical models and statistical probability for searching, filtering, clustering, automatic data collection, and entity mining. The emergence of these models has grown up in the business arena by integrating them into popular software such as R [2], SPSS [3], JMP

978-1-5386-5035-6/18/$31.00 ©2018 IEEE

[4]. However, the drawback of these software packages is that they work on predefined structured data. Moreover, because it is packaged software, users do not have much choice and customization when new tasks and questions are posed. Thus the ability to fully exploit distinct data sources is not solved.

To alleviate this problem, this paper provides *FinanViz*, a visual analytics tool for analyzing financial news on social media. Our main contributions in this paper are:

- We present a new approach for discovering dynamic terms/topics behavior over time, taking into account the consideration of spatial and proximity of terms and topics.
- We demonstrate our proposed approach through a visualization tool called *FinanViz* that provides social analysts an intuitive and interactive exploration of terms and topics through coordinated multiple views.
- We evaluate the usefulness and feasibility of *FinanViz* through a set of social media posts.

The paper is structured as follows: We describe related work in the following section. Then we introduce our *FinanViz* interface and its components. We illustrate the use of *FinanViz* on social media data and present the results of our study in Section IV. Finally, we conclude the paper and discuss future work.

II. Related Work

In the past decade, we have witnessed the importance and effectiveness of network visualization in social media [5], [6]. However, there is still no best practice for visualizing the temporal dynamics of a network. Visualizing topic evolution has emerged as an interesting research area, as social media has become increasingly accessible to billions of users [7], [8]. In the past few years, efforts of visualizing both topic evolutions and relationships can be classified into two strategies with different focuses: (1) network visualizations in which a user can highlight the evolution of entities on demand; (2) topic evolutions in which a user can highlight the relationship between entities at a particular time point [9]. In this section, we briefly review related work including text visualization and financial analysis from social media visualizations.

A. Text Visualization

Visualizing different aspects of textual datasets is a crucial part of visual text analytics. Text visualization has been an umbrella term for techniques that facilitate knowledge discovery and analysis of text data. One of the most popular techniques, the tag cloud, has been widely used as a summarization technique allowing readers to get a glimpse of the content of a document at a glance [10], [11]. Related to tag clouds, *TextArc*[1] is another visual rendering of text on a single page allowing the viewer to quickly get a sense of the content of the text. It places the text words on an ellipse with words central to the text appearing closer to the center of the graph, and like tag clouds font size represents the frequency of the word; however, unlike tag clouds words that co-occur in the same sentence are linked using lines revealing their relationships. Recently Kucher and Kerren [12] published a survey of text visualization techniques that classify these techniques by analytic task, visual task, domain, and data type. Their web site presents 220 different kinds of text visualization techniques published from 1976 to 2015[2]. A number of other recent projects also investigate novel approaches to visualizing text [13]–[19]. A noteworthy work is IBM's AlchemyAPI[3], which also color-codes prominent terms in text according to associated sentiment. The common thread among these works is the goal of facilitating the identification of threads of information and the discovery of emerging topics over time by a human analyst. This is usually achieved by either arranging keywords on a two-dimensional graph in a way that reveals relationships among keywords, or as a time series of rivers that highlight the emergence and disappearance of topics over time.

Theme River [20] employs kernel smoothing of time series, stacking them in a single display. Since stacking each time series depends on the previous stack, it becomes difficult to compare the absolute levels of the series. Interaction can mitigate this issue; for instance, Wattenberg's *Name Voyager* [21] presents an interactive stacked graphs of raw series which allows users to quickly drill-down to a particular time series by a single click. *Stacked graph* [22] is also a popular tool in visualizing and exploring online conversations about large-scale events [8]. *TextFlow* [23] analyzes various evolution patterns that emerge from multiple topics. The *TextFlow* visualization focuses on visually analyzing merging and splitting relationships between evolving topics. Xu et al. [24] employed stacked graphs to display the time-varying competitiveness of topics on social media with a storyline style visualization.

In general, these text visualization techniques focus on topic evolution, but not their correlations [25]–[27]. *EvoRiver* [9] uses the same composite visual design, but separate the negative and positive threads. When there is a turn on competition power (from negative to positive or vice versa), the topic will switch the thread as well. Users can select a time point and see the relationships between different topics by arcs.

B. Financial analysis from social media

It can be seen from Fig. 1 that there are approximately 17000 publications related to this topic. We do not survey all of them, instead, we filter out some studies similar to our approach in terms of data collection, and extraction.

Alanyali et al. [28] conducted a study to understand the correlations between decisions taken in financial markets and developments in the financial news. Data for analysis was extracted from Financial Times in a period of six years (2/2007 - 12/2012). The results of their study showed a positive correlation between mentioned company's posts on the Financial Times and the daily transaction volume on the stock. The calculated results indicated an intrinsically interlinked on movements between financial markets and financial news.

[1]http://www.textarc.org/

[2]http://textvis.lnu.se/

[3]http://www.alchemyapi.com/products/demo/alchemylanguage

Similarly, Guidi [26] studied how social media information can reveal the trend in financial markets by conducting two analysis. The first one was to find if there is any correlation between stock indicators and the moods derived from the social network (i.e., Twitter) about Apple Inc. The results showed a negative correlation between mood indicators, tweet volume and prices and a positive correlation between trading volume and tweet volume. He found that volume was more relevant to explain the relationship above rather than using the sentiment. The second analysis involved tracking the performance of financial indexes (e.g., S&P 500 and FTSE 100) based on 20 financial terms extracted from social media during a period of 4 years (2013 - 2016). The results of this analysis revealed a negative correlation between search word volume with closing price and return and a positive relationship with volatility and trading volume. The main issues with this study were the lack of data, leading to many statistically not significant values.

Challet et al. [29] proposed a study to predict financial index returns based on data from Google Trend. They claimed that the achieved data accommodated sufficient information for prediction with respect to careful choosing keywords for the use of an industry-grade backtest system.

Mao et al. [30] tried to identify indicators which can be used for financial markets. Their data were collected from various sources such as news media (e.g., Wall Street Journal, Bloomberg, Forbes.com, Reuters Business& Finance), Search Engine Data (i.e., Google Insights for Search), and Social Media (e.g., Twitter). They found that the conventional surveys of smart investors are the slow indicators of financial markets but financial search queries on Google Insight Search can be complimentary used. The interesting insights from social media data are that the tweets posted few days before financial index returns are statistically significant predictors.

In line with the approach of mining Twitter social media, Ruiz et al. [31] also tried to figure out the indicators that can reveal a correlation between micro-blogging activities with stock-market events. Trade volume is found to be more statistically significant than that of the stock price.

These studies above all posed interesting research questions by looking data from multiple perspectives and ultimately gave insights of given information. Statistical methods are strongly used to find the correlation between extracted features and the dependent variable. We address the same research questions but with a different approach through a lense of visualization method that allows content to be self-explained.

III. The *FinanViz* architecture

A. The *FinanViz* approach

FinanViz[4] is developed using JavaScript. The primary goal of *FinanViz* is to create an interactive visual analytic tool that presents users a high level view of trending topics/terms along with their information and relations. The tool enables users to investigate and explore the financial topics and terms over a period of time such as 1) the emerging terms in each week for a given topic, 2) the related terms in accordance with the trending terms within a topic and 3) the related terms in accordance with the trending terms across topics. To meet this goal, this paper proposes several features that are implemented in *FinanViz*:

- **Overview Display (F1)**. Display overview of terms, topics over time.
- **Details-On-Demand (F2)**. Present details on demand, including sources of information and its content.
- **Relationship on views (F3)**. Show the relationship among trending terms within and across topics
- **Sudden change (F4)**. Show the sudden change of terms over time.
- **Filter (F5)**. Search or filter out desired financial news

B. Data collection and pre-processing

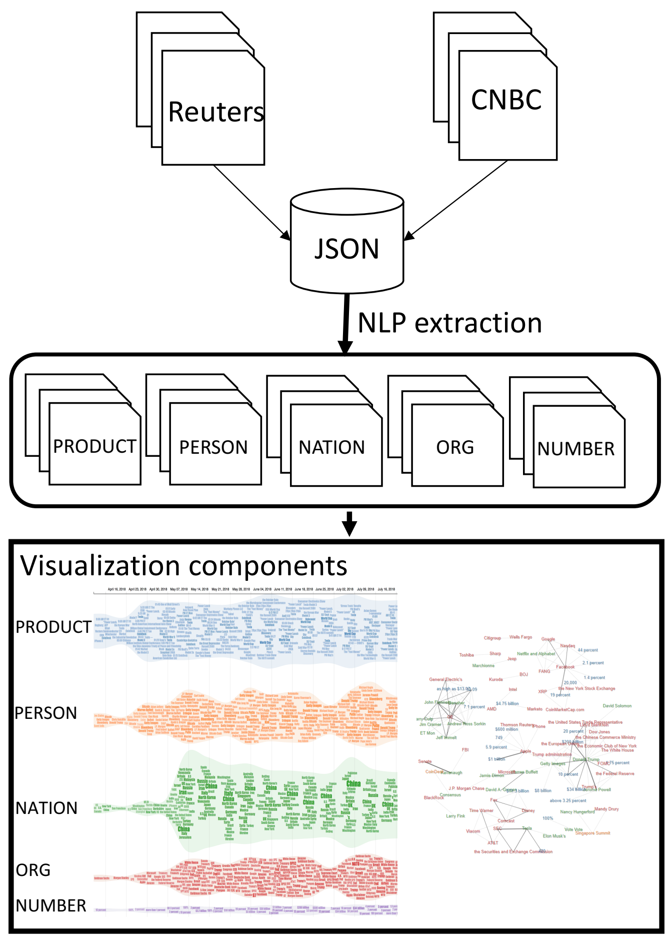

Fig. 2. The *FinanViz* architecture

The dataset for this study was retrieved from two main sources: Reuters[5] and CNBC finance[6]. These two sources are consolidated into a single JOSN format. The final dataset contains approximately 27000 articles. Each article has 6 features

[4]For a short demo of *FinanViz* visit https://vixlab305.github.io/FinanViz/

[5]https://www.reuters.com/

[6]https://www.cnbc.com/finance/

including: *id, title, url, imageurl, posted date and body*. Terms and topics are extracted by using Natural Language Processing library (i.e., spaCY[7]). The topics include: Product (objects, vehicles, foods, etc. (not services.)), Person (people, including fictional), Nation (countries, cities, states), Organization (companies, agencies, institutions, etc) and Number (percentage, including "%", monetary values, including unit). The extracted topics and their terms will be fed into the *FinanViz* tool. The flow of data visualization process is depicted in Fig. 2.

C. *The FinanViz visualization components*

FinanViz consists of three main components (the visualization feature **F1**) as depicted in Fig. 3 where 1) Box A contains *Utility* component, 2) Box B displays *Trending Terms* component, and 3) Box C shows the *Dynamic network* component.

Box A - The *Utility* component allows users to search keyword or terms of interest, layout of the viz tool in Box B can be reorganized in terms of time (daily or weekly by checking/unchecking the "Weekly' option) or uniformly (when users want to see each topic in a separated view or a single view by checking/unchecking the "Stack' option) .

Box B - The *Trending Terms* component encourages users exploring interesting terms for each topic (product, person, nation, organization and number) over time.

Box C - The *Dynamic network* component helps users get an overview of consolidated terms over all topics [27]. In this component, we provide two types of view for the use of interest: graph view, and timeline view. For both views, only terms that suddenly increase mentioned are extracted (the visualization feature **F4**). The graph view consists of nodes and links where each node represents a term, and each link indicates the co-occurrence of two terms in the same article. The more frequency two terms appear together, the thicker link. In the timeline view, each term is represented by a line, connection between terms is denoted by an arc.

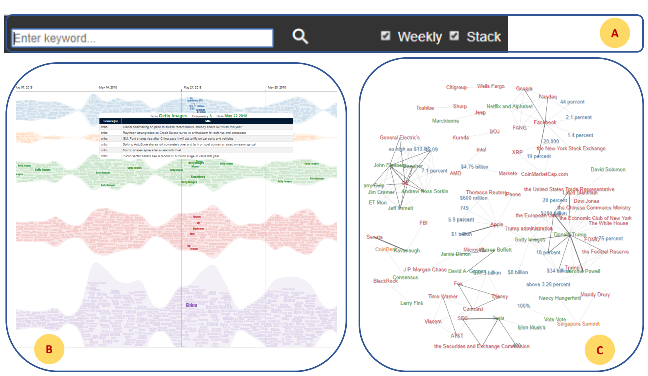

Fig. 3. Three main components of the *FinanViz*: A) Utility component, B) Topics stream component, C) Dynamic network component

[7]https://spacy.io

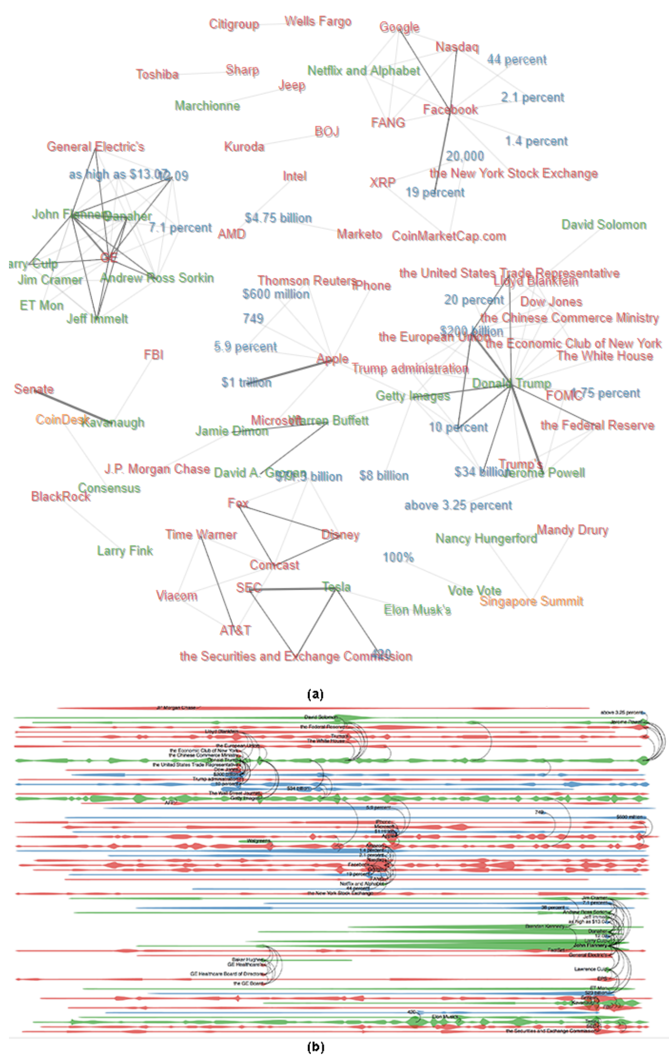

Fig. 4. Dynamic network of the *FinanViz*

D. *User interactions*

The *FinanViz* supports three types of interactions on components as follows:

Mouse over:

When users mouse over a trending term, a tooltip will be popped up to display all articles (the visualization feature **F2**) that contain the current terms. More interestingly, related emerging terms are also highlighted within the term's topic and across other topics. Fig. 5 illustrates an example when users mouse over a term "Treasury", related terms in other topic Fig. 5-A within the same period as well as related terms in the same topic Fig. 5-B are highlighted.

Mouse click: When users click on a given terms, a list of related terms (colored by topics) on the same day will be displayed below the selected term. Unlike mouse over, mouse click event allows user to visually explore a thumbnail of image corresponding to articles related to current term (as depicted in Fig. 6).

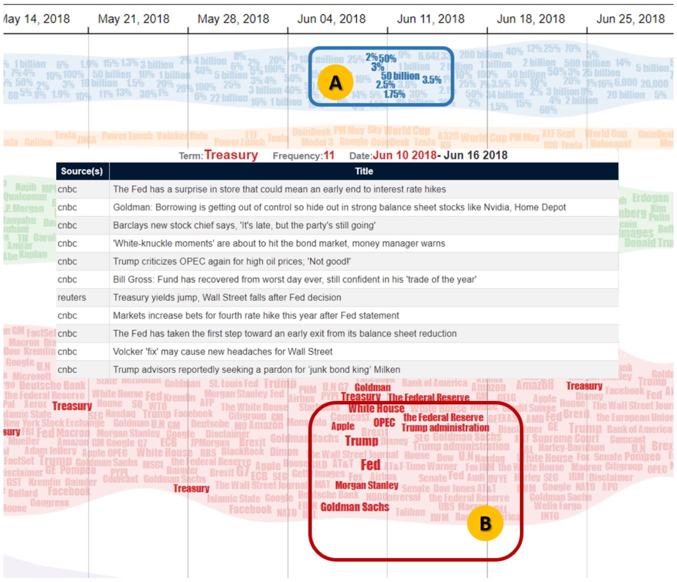

Fig. 5. Mouse over a trending terms highlights its relations to terms within and across topics

Search box: To narrow down the results shown on the overview, or to explore contents of interest, *FinanViz* provides a search box (the visualization feature **F5**), on the top left of the screen, that allows user to input desired information. The five components update the results according to the search word interactively (Fig. 7 - A).

IV. Results

To evaluate the usefulness and effectiveness of the *FinanViz*, we conducted a use case with three volunteer users. The purpose of this study was to (1) assess whether the user could find interesting patterns of emerging trending terms in financial field, and (2) receive feedback from user's experience to improve *FinanViz*. Before using the visualization tool, all users were introduced the basic functionality of *FinanViz* along with some explanations of the how terms and networks are structured.

Findings: Overall, users recognize a common pattern of a trending term "*Brexit*" due to the number of repeatedly occurrences over months. By investigating into detail of this term, they found an interesting influence on other categories. For example, from May 14 to Sep 17, 2018 "*Brexit*" barely impacts "Nation", "Person", "Product" or "Number", however, this behavior changes suddenly in the other half of the timeline, especially in the "Number" topic, leaving the question of the relationship between "*Brexit*" and financial issues in the last quarter of 2018. In addition to "*Brexit*", "*China*" is another popular term that can be easily detected. Nevertheless, it has a long-term impact on all other topics, showing its strong influence not only on financial markets but also on other domains.

On the overview, normally, changes in the "Number" topic may not vary considerably along the timeline. However, when users take a closer look at the common term "Facebook", the "Number" category witness an unusual move as it has a vast increase from Jul 23, 2018 to Jul 30, 2018. The term "more than ... percent" appear multiple times show there may be some transformation in resources. The changes range across all topics, especially "Organization" with many Tech companies' name highlighted, such as "Facebook" or "Google", along with financial organizations like "Nasdaq" market or "the New York Stock Exchange". This whole pattern may leave an impression of some major financial fluctuation relating huge tech companies.

When exploring on the Dynamic network component, users can detect some typical center nodes such as "Donald Trump", "Apple", "Facebook", and "Flannery". Since this component only extracts sudden change terms over months, interesting relationship between a given term and *"financial terms"* is revealed in a more intuitive way. For example, the term "Facebook" has several connections with percentage such as {'44 percent", "19 percent", "1.4 percent", "2.1 percent"} while "Apple" and "Donald Trump' has direct connections to monetary such as {"$600 million","$1 trillion"} and {"$200 billion"," $34 billion"} respectively.

User's feedback: After finishing the experiment, we gathered user's feedback to improve *FinanViz*, including adjusting the overlapping terms when users performed a mouse click, reorganizing the layout (or topology) such such financial concerned will be showed first, and including more information of the tooltip in the *Trending Terms Component* (detailed on total articles, date range and frequency)

V. Conclusion and Future Work

In this paper, we introduced *FinanViz*, a graphical tool that allows entrepreneurs, especially financial experts to gain a better understanding about emerging trends from financial news based on the co-occurrence of related terms within a topic or across multiple topics over time, and the relationship sudden change of trending terms spanning multiple years. The usefulness of the application is evaluated through a case study where users were able to find some typical interesting patterns. The most challenging task in this study is the interpretation of data when there is no financial experts involved in the study. In future work, we try to mitigate the current issue by collaborating with experts in financial department to improve the *FinanViz* and extract, interpret terms in a more meaningful manner.

References

[1] J. Snyder, D. Carpenter, and G. J. Slauson, "Myspace. com–a social networking site and social contract theory," *Director*, p. 07, 2006.

[2] M. Maechler, P. Rousseeuw, A. Struyf, M. Hubert, and K. Hornik, *cluster: Cluster Analysis Basics and Extensions*, 2013, r package version 1.14.4 — For new features, see the 'Changelog' file (in the package source).

[3] S. B. Green and N. J. Salkind, *Using SPSS for Windows and Macintosh, Books a la Carte*. Pearson, 2016.

[4] J. Goupy and L. Creighton, *Introduction to design of experiments with JMP examples*. SAS Publishing, 2007.

[5] N. Elmqvist, T.-N. Do, H. Goodell, N. Henry, and J. Fekete, "Zame: Interactive large-scale graph visualization," in *Visualization Symposium, 2008. PacificVIS '08. IEEE Pacific*, March 2008, pp. 215–222.

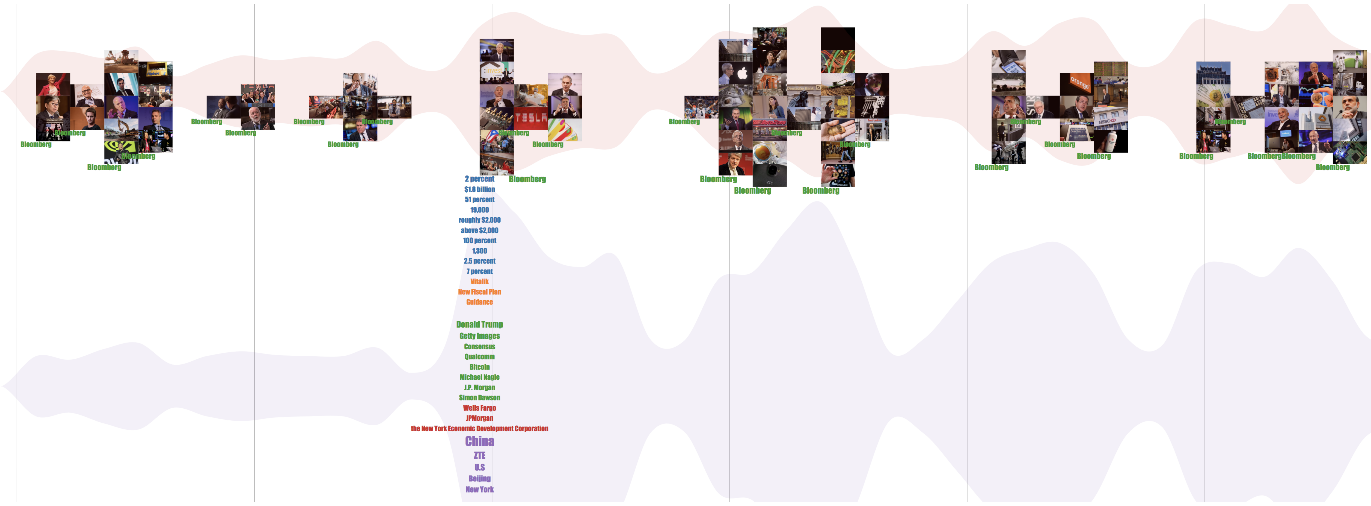

Fig. 6. Mouse click allows users to glance over a thumbnail of article's image over a period of time.

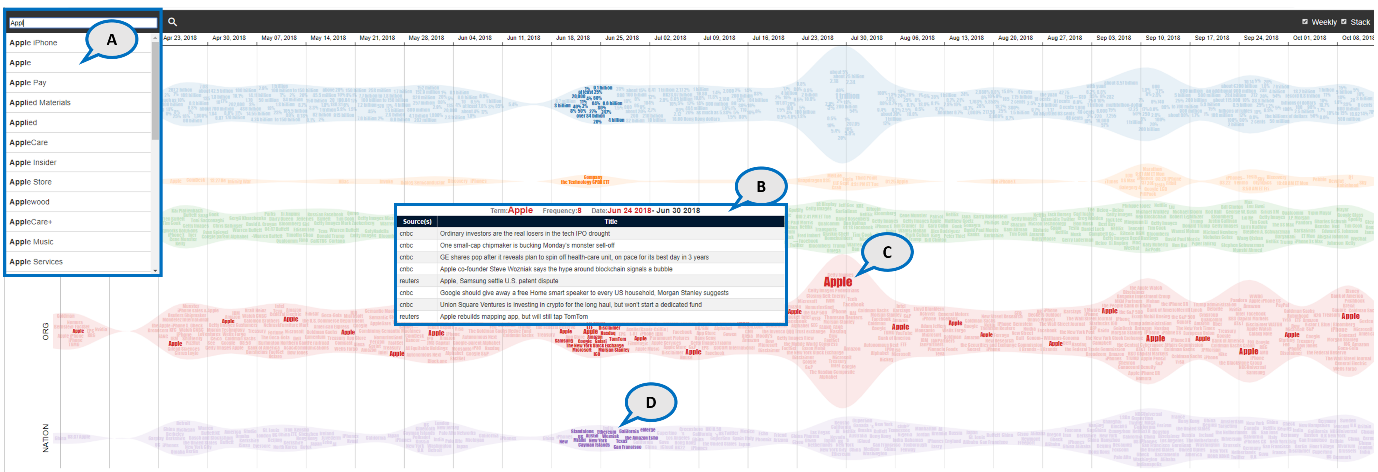

Fig. 7. A) Search box with "auto complete" widget, B) Detail of current term with mouse over, C) Appearance of the terms over time, D) Other terms related to current search word at the same time .

[6] N. Henry, J.-D. Fekete, and M. J. McGuffin, "Nodetrix: A hybrid visualization of social networks," *IEEE Transactions on Visualization and Computer Graphics*, vol. 13, no. 6, pp. 1302–1309, Nov. 2007. [Online]. Available: http://dx.doi.org/10.1109/TVCG.2007.70582

[7] M. Wattenberg, "Baby names, visualization, and social data analysis," in *Proceedings of the 2005 IEEE Symposium on Information Visualization*. Washington, DC, USA: IEEE Computer Society, 2005, pp. 1–7.

[8] M. Dork, D. Gruen, C. Williamson, and S. Carpendale, "A visual backchannel for large-scale events," *IEEE Transactions on Visualization and Computer Graphics*, vol. 16, no. 6, pp. 1129–1138, Nov. 2010. [Online]. Available: http://dx.doi.org/10.1109/TVCG.2010.129

[9] G. Sun, Y. Wu, S. Liu, T. Peng, J. J. H. Zhu, and R. Liang, "Evoriver: Visual analysis of topic coopetition on social media," *IEEE Trans. Vis. Comput. Graph.*, vol. 20, no. 12, pp. 1753–1762, 2014. [Online]. Available: http://dx.doi.org/10.1109/TVCG.2014.2346919

[10] F. Heimerl, S. Lohmann, S. Lange, and T. Ertl, "Word cloud explorer: Text analytics based on word clouds," in *System Sciences (HICSS), 2014 47th Hawaii International Conference on*, Jan 2014, pp. 1833–1842.

[11] B. Y.-L. Kuo, T. Hentrich, B. M. . Good, and M. D. Wilkinson, "Tag clouds for summarizing web search results," in *Proceedings of the 16th International Conference on World Wide Web*, ser. WWW '07. New York, NY, USA: ACM, 2007, pp. 1203–1204. [Online]. Available: http://doi.acm.org/10.1145/1242572.1242766

[12] K. Kucher and A. Kerren, "Text visualization techniques: Taxonomy, visual survey, and community insights," in *Visualization Symposium (PacificVis), 2015 IEEE Pacific*, April 2015, pp. 117–121.

[13] W. Cui, S. Liu, Z. Wu, and H. Wei, "How hierarchical topics evolve in large text corpora," *Visualization and Computer Graphics, IEEE Transactions on*, vol. 20, no. 12, pp. 2281–2290, Dec 2014.

[14] D. Eler, F. Paulovich, M. de Oliveira, and R. Minghim, "Topic-based coordination for visual analysis of evolving document collections," in *Information Visualisation, 2009 13th International Conference*, July 2009, pp. 149–155.

[15] A. G. Forbes, B. Alper, T. Höllerer, and G. Legrady, "Interactive folksonomic analytics with the Tag River visualization," in *Proceedings of the IEEE Workshop on Interactive Visual Text Analytics (TextVis)*, Providence, Rhode Island, October 2011.

[16] S. Havre, E. Hetzler, P. Whitney, and L. Nowell, "ThemeRiver: Visualizing thematic changes in large document collections," *Visualization and Computer Graphics, IEEE Transactions on*, vol. 8, no. 1, pp. 9–20, Jan 2002.

[17] A. Kochtchi, T. von Landesberger, and C. Biemann, "Networks of names: Visual exploration and semi-automatic tagging of social networks from newspaper articles," *Computer Graphics Forum*, vol. 33, no. 3, pp. 211–220, 2014. [Online]. Available: http://dx.doi.org/10.1111/cgf.12377

[18] M. Krstajic, F. Mansmann, A. Stoffel, M. Atkinson, and D. Keim, "Processing online news streams for large-scale semantic analysis," in *Data Engineering Workshops (ICDEW), 2010 IEEE 26th International*

Conference on, March 2010, pp. 215–220.
[19] S. Liu, X. Wang, J. Chen, J. Zhu, and B. Guo, "Topicpanorama: a full picture of relevant topics," in *Proceedings of IEEE Symposium on Visual Analytics Science and Technology (VAST)*, 2014, pp. 183–192.
[20] S. Havre, B. Hetzler, and L. Nowell, "ThemeRiver: Visualizing theme changes over time," in *Proceedings of the 2000 IEEE Symposium on Information Visualization*. Washington, DC, USA: IEEE Computer Society, 2000, pp. 115–123.
[21] M. Wattenberg, "Baby names, visualization, and social data analysis," in *Proceedings of the 2005 IEEE Symposium on Information Visualization*. Washington, DC, USA: IEEE Computer Society, 2005, pp. 1–7.
[22] L. Byron and M. Wattenberg, "Stacked graphs – geometry & aesthetics," *IEEE Transactions on Visualization and Computer Graphics*, vol. 14, no. 6, pp. 1245–1252, Nov. 2008. [Online]. Available: http://dx.doi.org/10.1109/TVCG.2008.166
[23] W. Cui, S. Liu, L. Tan, C. Shi, Y. Song, Z. Gao, H. Qu, and X. Tong, "Textflow: Towards better understanding of evolving topics in text," *IEEE Transactions on Visualization and Computer Graphics*, vol. 17, no. 12, pp. 2412–2421, Dec. 2011. [Online]. Available: http://dx.doi.org/10.1109/TVCG.2011.239
[24] P. Xu, Y. Wu, E. Wei, T.-Q. Peng, S. Liu, J. Zhu, and H. Qu, "Visual analysis of topic competition on social media," *Visualization and Computer Graphics, IEEE Transactions on*, vol. 19, no. 12, pp. 2012–2021, Dec 2013.
[25] T. N. Dang, N. Pendar, and A. G. Forbes, "TimeArcs: Visualizing Fluctuations in Dynamic Networks," *Computer Graphics Forum*, 2016.
[26] M. Guidi, "Extracting information from social media to track financial markets," B.S. thesis, Università Ca'Foscari Venezia, 2017.
[27] T. Dang and V. T. Nguyen, "ComModeler: Topic Modeling Using Community Detection," in *EuroVis Workshop on Visual Analytics (EuroVA)*, C. Tominski and T. von Landesberger, Eds. The Eurographics Association, 2018.
[28] M. Alanyali, H. S. Moat, and T. Preis, "Quantifying the relationship between financial news and the stock market," *Scientific reports*, vol. 3, p. 3578, 2013.
[29] D. Challet and A. B. H. Ayed, "Predicting financial markets with google trends and not so random keywords," *arXiv preprint arXiv:1307.4643*, 2013.
[30] H. Mao, S. Counts, and J. Bollen, "Predicting financial markets: Comparing survey, news, twitter and search engine data," *arXiv preprint arXiv:1112.1051*, 2011.
[31] E. J. Ruiz, V. Hristidis, C. Castillo, A. Gionis, and A. Jaimes, "Correlating financial time series with micro-blogging activity," in *Proceedings of the fifth ACM international conference on Web search and data mining*. ACM, 2012, pp. 513–522.