HackerNets: Visualizing Media Conversations on Internet of Things, Big Data, and Cybersecurity

Hao Van
Department of Computer Science
Texas Tech University
Lubbock, Texas, USA
hao.van@ttu.edu

Huyen N. Nguyen
Department of Computer Science
Texas Tech University
Lubbock, Texas, USA
Huyen.Nguyen@ttu.edu

Rattikorn Hewett
Department of Computer Science
Texas Tech University
Lubbock, Texas, USA
rattikorn.hewett@ttu.edu

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

***Abstract*—The giant network of Internet of Things establishes connections between smart devices and people, with protocols to collect and share data. While the data is expanding at a fast pace in this era of Big Data, there are growing concerns about security and privacy policies. In the current Internet of Things ecosystems, at the intersection of the Internet of Things, Big Data, and Cybersecurity lies the subject that attracts the most attention. In aiding users in getting an adequate understanding, this paper introduces *HackerNets*, an interactive visualization for emerging topics in the crossing of IoT, Big Data, and Cybersecurity over time. To demonstrate the effectiveness and usefulness of *HackerNets*, we apply and evaluate the technique on the dataset from the social media platform.**

***Index Terms*—Internet of Things, Cybersecurity, Big Data, Interactive Visualization, Venn Diagram, Media User Network, Social Media Mining.**

I. Introduction

The Internet of Things (IoT) refers to a giant network of physical objects, connecting a massive number of sensors or devices through the global infrastructure to provide value-added services [1], [2]. Current virtual assistants such as Amazon Alexa [3] or Google Assistant [4], [5] are examples of IoT systems, in which voice commands are utilized for controlling IoT applications. Demands for innovations come with increasing requirements for data and security insurance in order to enhance the trustworthiness of such a system.

The innovation of IoT originates from the data, which is witnessing an explosive growth in volume, variety, and velocity, introducing a new era of Big Data. The more complex the IoT system is, the higher the chance for the data to be distributed in different places, under different policies, and with different objectives [6]. Data collected from smart IoT devices can contain sensitive personal information and can be used for various purposes, from improving customer satisfaction to being sold to third parties for further analysis [1]. With the vast amount the data presents in the IoT sheer scale, the IoT era poses complex challenges and significant demands for security and privacy.

In terms of data sources, social media presents an emerging platform for discussion of these innovations, where ideas can be formed and discussed. This can provide an abundant source of data for collection and analysis [7]. There is a gap between utilizing the data source from social media and capturing the ongoing discussion among the prevalent subjects in technology. To fill this gap, we introduce the technique of *HackerNets* for visualizing the media conversation between these topics and the interests from users that shift over time in forms of networks and relations. We consider the topics regarding three main areas: IoT, Big Data, and Cybersecurity, to showcase the concerns about the current situation, issues, challenges, and solutions. *HackerNets* is built on time-series data, which combines the classic Venn diagram for relation representation incorporated with user network visualization for connections between users. *HackerNets* supports topic evolution visualization for showcasing the trending of a specific topic over time.

This paper is organized as follows. Section II gives an overview of related literature to *HackerNets*. Section III describes design considerations and design choices. Section IV presents the architecture of *HackerNets*. The result and discussion are presented in section V. Section VI discusses the conclusion of *HackerNets* and future work to further improve this technique.

II. Related Work

In this section, we present previous literature related to *HackerNets*, with regard to innovation aspects and concerns from IoT, Big Data and Cybersecurity, as well as data visualization techniques for addressing the target problem.

978-1-7281-0858-2/19/$31.00 ©2019 IEEE

A. Big Data and Cybersecurity in the context of IoT

The importance of addressing IoT, Cybersecurity, and privacy challenges have been discussed toward IoT solutions in previous literature [8]–[11], and brought in widespread attention. As reported by The EU Commission on the IoT, in The Cluster of European Research Projects on the Internet of Things [12], Cybersecurity and privacy are identified as a major IoT research challenge, including models for decentralized authentication and trust; privacy-preserving technology for heterogeneous sets of devices; and the integration of sensor information and agents in discovery services. In addition, according to TRUSTe, consumer privacy concerns could be a significant barrier to IoT growth [13].

The emerging IoT systems generate the data streams from billions of devices, posing another degree of challenges in terms of management, storing, and processing data. An approach to the combined challenges in the collection, use, and management of Big Data at the intersection of security and privacy is described by Sollins [6]. This problem is magnified in the context of the current IoT. The constraints between IoT, Big Data, and Cybersecurity lies in the security capabilities and the IoT device, resulting in the trustworthiness of the data to be sent and received with corresponding privacy policy and enforcement. Extraction of knowledge from Big Data can lead to potential user privacy issues. However, existing technologies and regulations are insufficient to support a privacy-guaranteed data management lifecycle [1], and there is no single perfect data management solution for the cloud to manage big data [14].

B. Social media data analysis

The growth of social media use brings new opportunities in various aspects: Communicating and understanding local communities and people [15], conducting business [16], as well as research in social science, health and medicine [17], [18]. Utilizing social media is especially prevalent in computer science with methods such as mining text data [19], data visualization [20] and recommendation system [21].

In forum-based social media, the original news is typically followed by other reader's opinion, in the forms of comments and votes [21]. This hierarchy structure of content aids readers in following and keep track of the main thread. In this work, we explore the data retrieved from Hacker News[1], a forum-based social media focusing on news in computer science and entrepreneurship. Apart from social influence bias, the online popularity on Hacker News is a relatively stable reflection of intrinsic quality [22].

C. Visualization techniques

In terms of visualization for relation presentations, the Venn diagram is a classic method for logical relations between different sets [23]. Recent variations of Venn diagrams includes the applications on various domains, including gene lists [24], graph theory [25], [26], pharmacological dataset [27] or on interactive features [28]. The network visualization can be seen as a popular method for visualizing the evolution of relationships among objects in a timely manner [29], [30]. In *HackerNets*, we apply the model of the Venn diagram incorporated with the user network, in which users can see the interaction between entities inside a larger overlapping area.

From a set of topics, Bayesian rose trees [31] model is a popular method for topic hierarchy. TopicPanorama [32] employs this model and presents a combination of a radially stacked tree with a density-based graph visualization, supporting topic-graph-based analysis at different degrees of detail. We utilize similar techniques in our *HackerNets*, with clustering inside the network and forming community in distinct areas.

Regarding visualization text in time-series data, several methods are addressing the task of topic evolution. TIARA [33] utilized the combination of the word cloud [34] and stacked graph to demonstrate visual representation from abstract and complex text summarization. Recently, WordStream [35] expanded the technique in terms of emphasizing essential terms temporally, combining the model of the word cloud and stream graph [36] to form a unified stream for each category. Inside the stream, key terms can be highlighted to showcase explicitly topics they belong to, how frequent it appears in the text source. Interactive features are developed to bring in-depth analysis from the WordStream, providing context for such keywords in terms of similar post, location, authors, and timestamp [37]. Our *HackerNets* employs the technique of WordStream to visualize the evolution of three main areas: IoT, Big Data, and Cybersecurity.

III. Design Goals and Design Decisions

A. Design goals

We apply the technique of *HackerNets* on data from social media, to present the conversations and concerns from the perspective of online users. Data from social media regularly have multidimensional structure, which encompasses attributes such as authors, metadata from authors, messages, and their corresponding metadata, features that show the interactions from other authors like comments or ratings. These entities can have many-to-many relationships, forming a complex network of users. Moreover, some posts can be made of non-formal content, making it challenging to pre-process text content. Understanding the underlying structure of this data type plays a vital role in data processing and knowledge extraction. In order to capture the broad overview as well as detail user network from social media, the following design criteria are adopted as goals for our technique:

G1. Present the evolution of topics and network as they shift over time.

G2. Connect the related entities closer together and form communities.

G3. Display detail on demand and highlight the structure of the network.

[1]https://news.ycombinator.com/

B. Design decisions

To achieve the above design goals, we implement the following visual designs:

D1. The time axis is set horizontally. The time axis plays the role of both the timeline and time filtering component for the visualization. This satisfies the design goal **G1**.

D2. User network. The force-layout is used to visualize the network, in which each node represents a user, and the link indicates the interactions between authors. This meets the design goal **G2**.

D3. Venn diagram and selection. Venn diagram can show all possible logical relations between users from three topics: IoT, Big Data, and Cybersecurity. The overlapping area presents conversations that attract more than one topic. We implement a function for selecting a range of entities on the network and highlight them. This meets the design goal **G3**.

D4. WordStream for topic evolution. From the detail text content, extracted keywords are visualized in a timely manner, showing the evolution of the topic over time. This helps to meet the design goal **G3**.

IV. The *HackerNets* architecture

A. The HackerNets approach

HackerNets[2] is developed using JavaScript and the library D3.js [38]. Besides, venn.js and WordStream are employed to implement the Venn diagram and show trending topics over time. *HackerNets* brings a visual interface that shows social media users' concerns on the Internet of Things, Big Data, and Cybersecurity topics with their interactions. *HackerNets* contains a Venn diagram to help users easily envision the relationships between the three main topics when users mention them. The user network is developed and incorporated with the Venn diagram to visualize user characteristics and their interactions with one another. For topic evolution, we utilize the technique of WordStream [35] to express essential keywords from the posts over time. To meet this end, this paper proposes visualization features [39] that are implemented in *HackerNets*:

F1. Overview of presentation. Display overview user network, Venn diagram of the three major categories (IoT, Big Data, and Cybersecurity) and WordStream of main topics within these categories.

F2. Detail-on-demand. Show details of users on request, including user name, number of posts, and number of comments, and ratings from other users.

F3. Filtering. Filter posts and users by time range or by topic category.

F4. Ranking. Rank users on the network by various criteria, such as overall ranking or degree of centrality in the network.

Figure 1 shows a schematic overview of main components in our *HackerNets* prototype.

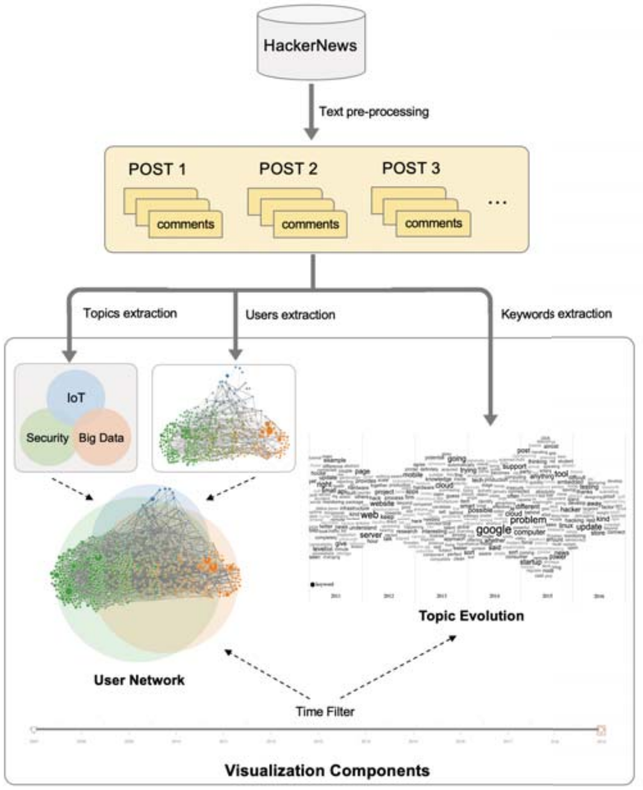

Fig. 1. The schematic overview of *HackerNets* visualization.

B. Data collection and pre-processing

The dataset in this study is collected from Hacker News social news website [40] through their API and Google BigQuery [41]. This website is a social media platform in technology and security, developed by Paul Graham. Users on this platform can post stories, and other users will interact with these stories by voting and leaving comments. On social media platforms, users often use special characters to replace characters commonly used in the text, such as "0" instead of "O" [42]. We have used these special rules to perform searches for posts on each topic. The data collection process is achieved through the following steps:

1) Collecting stories by executing queries with keywords: "internet of things", "iot", "internet 0f things", "i0t", "big data", "bigdata", "security". Google BigQuery is used because they provide faster queries.
2) Collecting comments that belong to each of the above stories. Because of no restricted comment levels of an story, we recursively get the comments from Hacker News API.

The raw dataset contains 22,884 posts and 30,506 comments from the year 2007 to early 2019. For the pre-processing process, we filter out authors who have not had many contributions to the platform while preserve the ten-year period to maintain the general picture. The dataset has the posts and comments in the time range of over ten years, so the authors who had the number of posts less than ten may be inactive. The records in the authors' data are retrieved according to their attributes such as user name, id, topics they are interested

[2]App demo link of *HackerNets*, visit https://bit.ly/35RvMpY

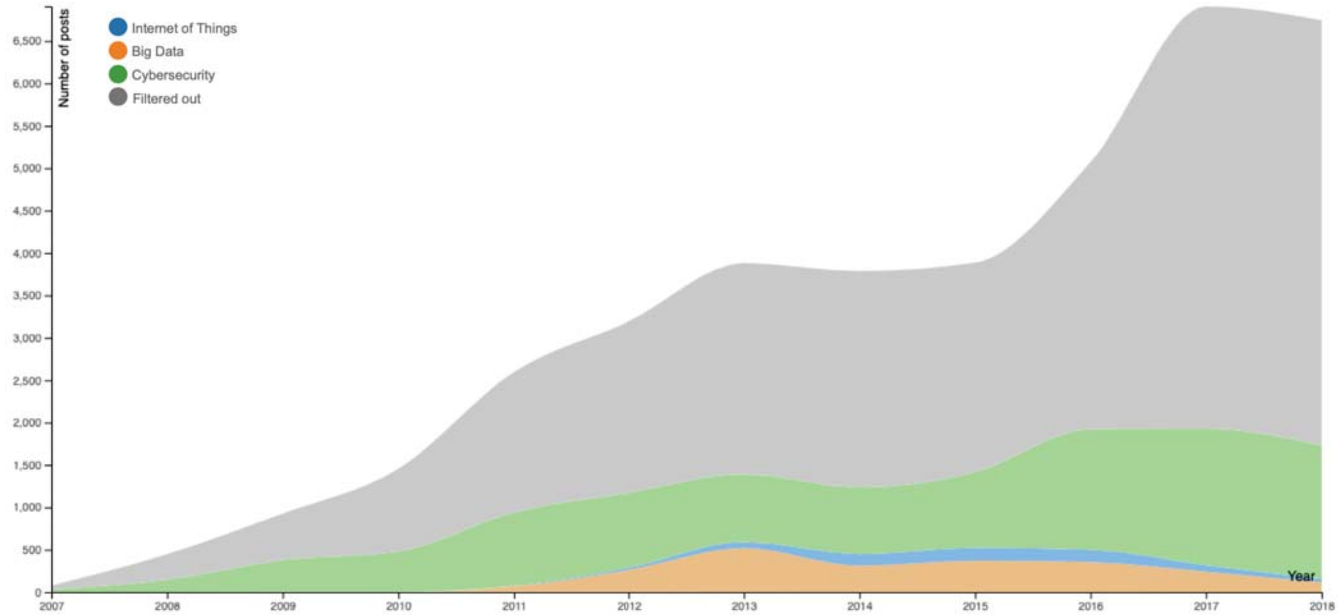

Fig. 2. Number of posts on each topic categories significantly increase over the past 12 years: blue for IoT, orange for Big Data, green for Cybersecurity. and gray for other posts.

in, and the posts, comments they created. Then, we filter out authors who have less than ten posts or comments from 2007 to early 2019, and the dataset now has 7,325 posts and 6439 comments from 649 authors. The number of posts per each topic with corresponding colors: blue for IoT, orange for Big Data, and green for Cybersecurity, along with the number of posts removed during the filtering process in gray, are shown in Fig. 2. After extracting topics and users, the processed data is used for visualizing the network and Venn diagram, showing the relationship between users in terms of the main topics of IoT, Big Data, and Cybersecurity. To go into greater detail of the text content, we employ the technique of WordStream for an in-depth analysis of the topic evolution over time. We generate the data for WordStream from the previously processed dataset with two steps:

1) All comments belong to each post are grouped, and the texts will be processed to get a list of common keywords and a list of users who interacted with this story.
2) All records are grouped by year. Again, we continue to pull out the most common keywords during that time.

C. The HackerNets visualization components

HackerNets consists of three main components as depicted in Fig. 3 where 1) Box A contains the *trending keywords* distributed yearly, 2) Box B has time filters and toggle buttons for interactive features, 3) Box C includes user network, the connection between them within the Venn diagram for three topics: IoT, Big Data and Cybersecurity.

The WordStream component (Box A): This component shows the most popular keywords which are extracted from Hacker News stories and comments. The essential terms, having a high frequency in comparison to other terms, are emphasized by font size chronologically along the timeline. Through this visualization, users can quickly see the most mentioned words in both three topics of IoT, Big Data, Cybersecurity, and their mentioned level each year to catch the global trend. Fig. 4 shows a closeup view of the topic stream for four years from 2012 to 2015.

The control panel (Box B): Users can filter the post by time. Specifically, by dragging the double range slider, users can set the time interval for topic stream and user network. The toggle buttons are for showing or hiding the Venn diagram in the main view.

The network component (Box C): This is the main component in our tool, including two parts: Venn diagram and user network overlaid on top of the Venn diagram. Each node in the network presents a Hacker News author or user. The radius of a circle is computed based on the number of posts, including stories and comments that the author created. Each node contains a pie chart that shows the percentage of the number of posts on the three topic categories: blue for IoT, orange for Big Data, green for Cybersecurity. We use this color encoding consistently throughout the paper. A link between two nodes indicates how often the two authors interact with each other. For example, in this case, at least one of them has commented on the other's article. The thickness of the links shows the number of interactions between the two authors.

D. User interactions

The *HackerNets* supports four types of interactive operations.

Mouse over:

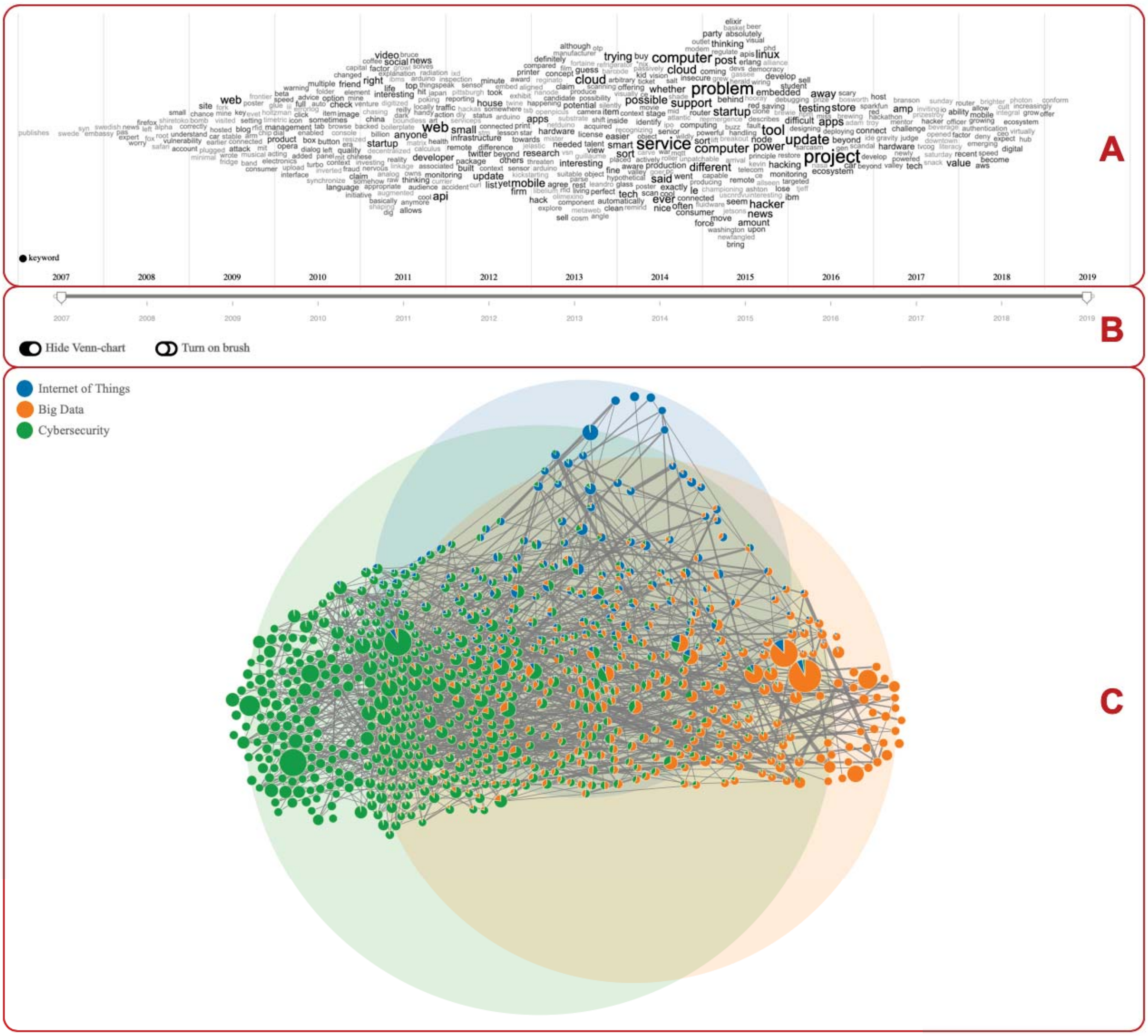

Fig. 3. Visual interface of *HackerNets*: A) Topic evolution of trending keywords extracted over the period of 2007 to 2019, B) Toolbox, including time filter for selecting a period of time for the corresponding presentation, and toggle buttons for Venn diagram visibility and network brushing, C) User network, containing connections between users and Venn diagram for topic covering.

a) Box A. When users mouse over a term, all the similar terms in the whole timeline will be highlighted. This will help users quickly see the popularity of words throughout the years.

b) Box C. When users mouse over a node in the user network, a pop-up panel appears in order to show the details information about the author's name, the number of posts and the number of interactions that the author has.

Mouse click: When users click on a term, the WordStream interface shows the stream of the selected term.

Mouse drag:

a) Box B. Users can drag the filter to select the time period. The WordStream, Venn diagram, and user network will be updated after users release the mouse, as shown in Fig. 5.

Brush: After turning on the brush mode by clicking on the toggle button, users can drag to draw a rectangle on the network (Box C). If the rectangle covers the nodes and the links, they will be highlighted, and the WordStream will be updated to show common keywords from selected users. An example is depicted in Fig. 6.

Toggle Venn diagram: Venn diagram will be hidden by default. When a user clicks the toggle button, the diagram will be displayed, as shown in Fig. 7.

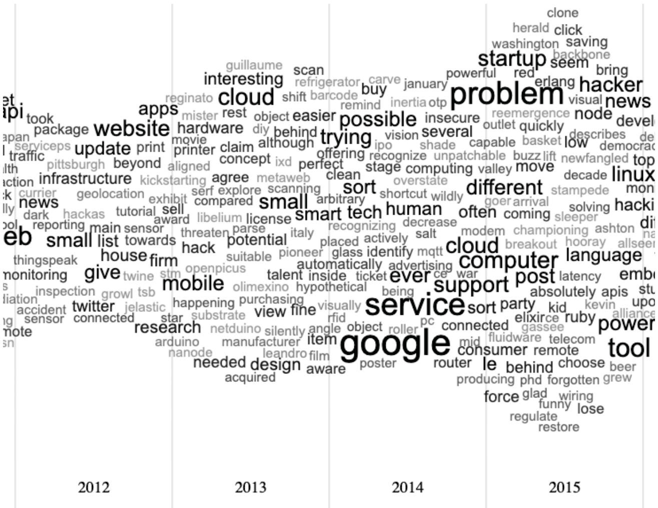

Fig. 4. Trending keywords in 2012, 2013, 2014, 2015. The significant terms are “google”, “service”, and “problem” for these four years.

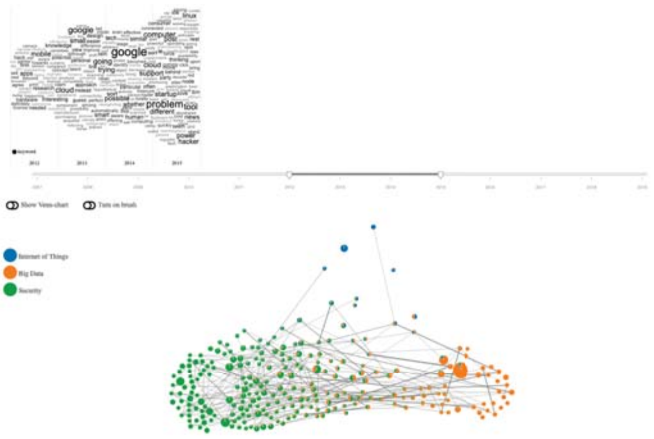

Fig. 5. The charts are updated after selecting new time period, i.e., time filtering features support updates to multiple linked views in the topic evolution visualization, WordStream and the user network.

Fig. 6. The nodes and links are highlighted after using the brush. The WordStream for trending terms in posts and comments is updated according to the selected users.

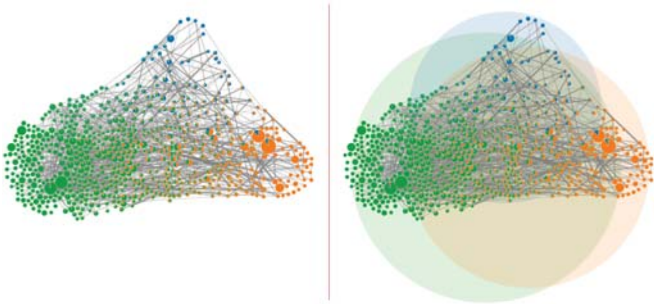

Fig. 7. The Venn diagram before (left panel) and after (right panel) clicking the toggle button. The Venn diagram shows the coverage of each topic in the conversation, where the users and their interests belong to. The intersection of more than one area represents the mixed topics that the conversation refers to.

V. Results and Discussion

We handled a use case with four volunteer users who have experience in using social media and visualization to gather qualitative responses and then evaluated the usefulness and effectiveness of the *HackerNets*. The purpose of this study is to gather users’ feedback to improve *HackerNets* in terms of gaining insights from the visualization from the users’ perspective. At the beginning of the study, users are introduced to *HackerNets*, the data overview, and interactive features of this tool, focusing on media conversations in IoT, Big Data, and Cybersecurity. After that, the users got to familiarize themselves with the application and its supported features. After having experimented with the tool, users are asked a set of questions, and they try to answer them using our *HackerNets* visualization tool.

- (R1) Which topics are more important, and how they change over time?
- (R2) Who are the influential authors and what topics do they care?
- (R3) What are the major events and the reasons behind these events?

Findings:

For the first question (R1), the users can easily find out that in the entire timeline, the Cybersecurity topic has the greatest concerns from authors by using the Venn diagram. Fig. 8 illustrates the expansion of the topic over the years, with accumulated data from the starting point of 2007. The topic of Cybersecurity appeared from the early year (in this case, 2007) and has been the dominant topic throughout the whole period. This helps to confirm the observation from the stream graph in Fig. 2. The IoT-related topic started to emerge in 2008, and the Big Data topic appeared four years later in 2012. However, the Big Data topic grew faster than IoT, demonstrated by the thicker streams in later years since 2013. Interestingly, during user study, one user found out that we can observe the same result by considering the colors and numbers of the nodes in the user network, while the Venn diagram presents an overall picture of the development. The overlapping areas between the topics depict the cross-topics conversations. It is observed that, from Fig. 8, the circles for IoT and Big Data are all located inside the circle of Cybersecurity almost entirely, demonstrating that users who invest in these two topics have major concerns regarding Cybersecurity: Each and every conversation in each of these two topic categories includes Cybersecurity-related terms.

As we slide along the timeline, the intersection area between the main three topics grows larger: more and more users now consider these topics together. The conversations now have a more extensive range of subjects, presenting the connection among IoT, Big Data, and Cybersecurity. The data generated from IoT devices are now collected in the Big Data system. The infrastructure of Big Data provides the facilities for IoT devices to produce faster responses, more comprehensive analysis. IoT and Big Data impact each other in a way that requires a more sophisticated and advanced mechanism for Cybersecurity. On the other hand, Cybersecurity innovations help strengthen the Big Data system and facilitate reliable IoT devices. These three topics have a reciprocal relationship in the era of Industry 4.0.

We divide the time period into non-overlapping intervals, as shown in Fig. 9. Even in this separated context, the areas for overlapping is significantly larger towards the current time. Moreover, in the most recent interval (2016-2018), there is barely any area that is accounted for only one topic, i.e., Cybersecurity and IoT, or Cybersecurity and Big Data would very much likely appear together.

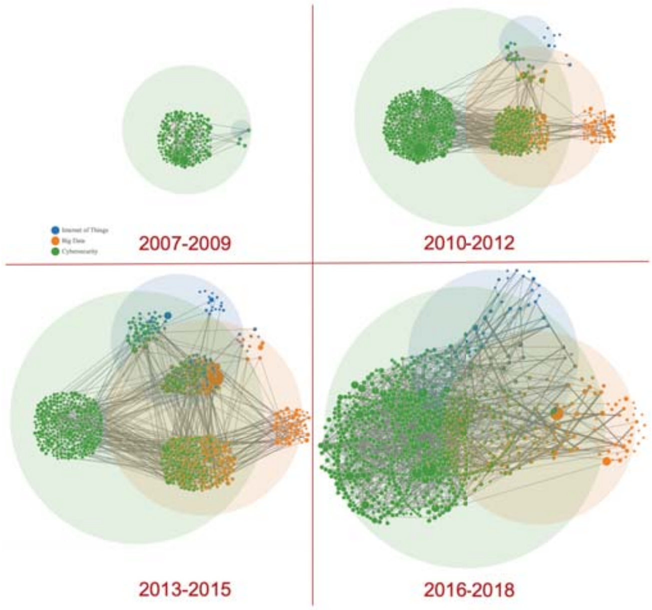

Fig. 9. Topics expansions in every three-year interval: Most authors are concerned about Cybersecurity and a tiny part of the appearance of IoT from 2007-2009; Cybersecurity, IoT and Big Data from 2010-2012, 2013-2015, 2016-2018.

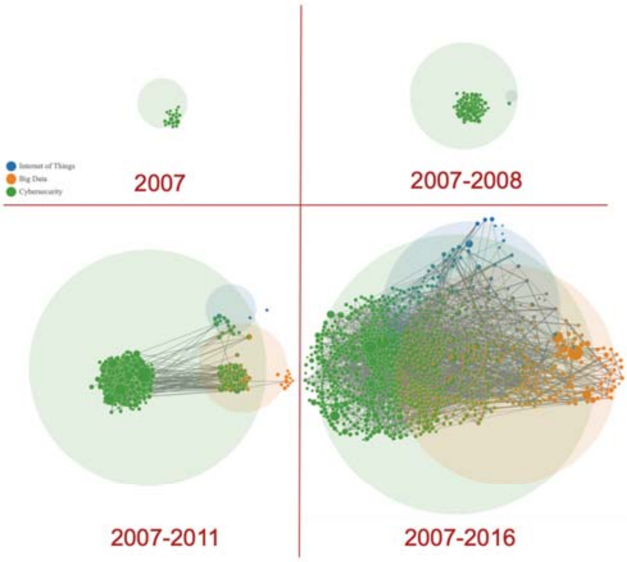

Fig. 8. Topics expansion over the years, from the starting year 2007: Cybersecurity in 2007; Cybersecurity and IoT from 2007-2008; Cybersecurity, IoT and Big Data from 2007-2011; Cybersecurity, IoT and Big Data from 2007-2016.

Regarding the second research question (R2) on detecting the most influential authors on Hacker News, the participants tend to mouse over larger nodes to see the details about the author's name, number of posts, number of interactions, and average ranking. While the size of a node indicates the number of posts that an author creates, it is the connections that present the number of interactions to other accounts, hence their influence towards others. This shows that the use of size may lead to ambiguity for perception.

Fig. 10 presents the detail-on-demand of an author. Most of the users conclude that "tptacek" is the most influential user of the Cybersecurity topic and also the whole network, shown in Fig. 10 A. The author named "BigDataStartups" (in the lower panel of Fig. 10) has a massive number of posts, but there is no connectivity to other authors. One potential reason for this pattern is that this author is an auto-bot, which usually creates spam content. This kind of auto-bot is popular among social media platforms, especially in forums with large impacts, due to large viewers and attention. In fact, according to further findings, BigDataStartups[3] (now known as Datafloq) is an online big data platform where organizations that want to take full advantage of big data and develop a big data strategy can find Big Data startups from around the world. Chances are the posts from this author that are created for advertising and publicity. We utilize brushing features to dive into the detail of content in this case. As depicted in the lower panel of Fig. 10, many technical keywords are shown such as "tool", "hacker", "linux" and "tech". However, these keywords do not give too much information about happened events in the timeline. They are general words that have common meanings about technology, usually, appear in articles or news. This is quite similar to the purpose of this author, as we described. For

[3]https://www.datafloq.com/

this user network, we believe that the betweenness centrality is a more significant indicator for influential people compared to the number of posts.

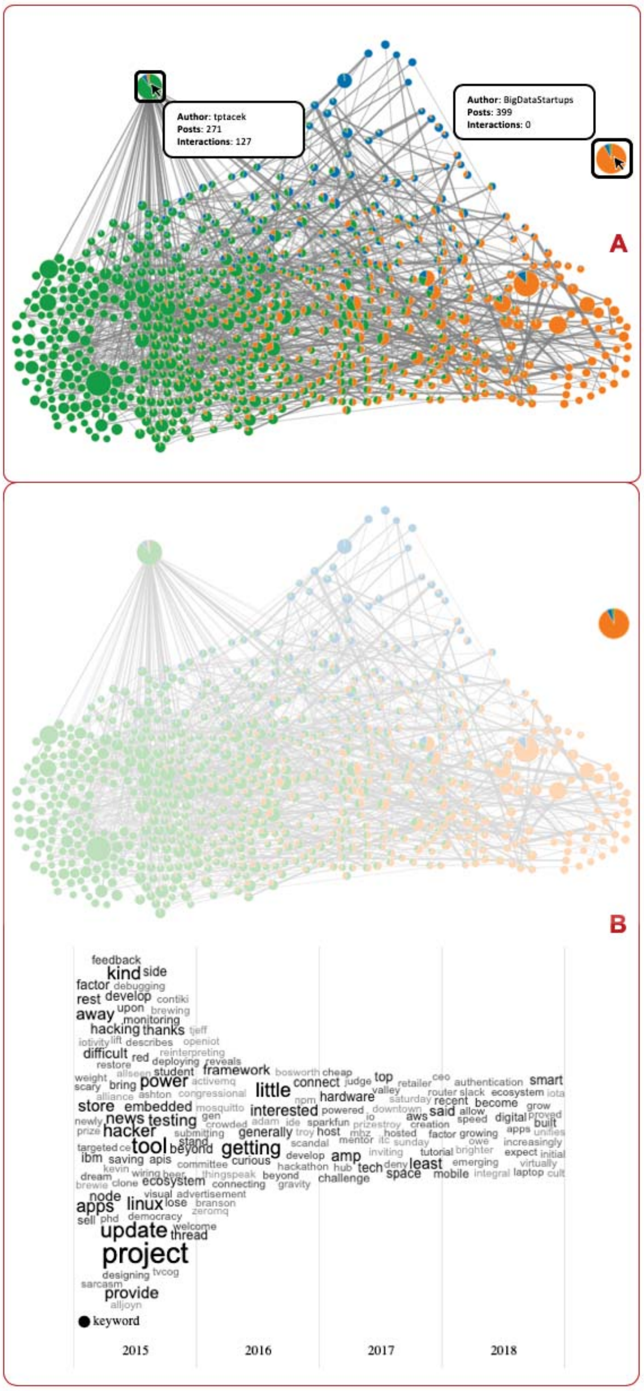

Fig. 10. Detail-on-demand: A) Mouse-overing nodes for detail, including author "tptacek" and author "BigDataStartup". B) Brushing for node "BigDataStartup" and the corresponding topic evolution presenting the content this author created.

For the third research question (R3) regarding the major events, the users can quickly observe some larger keywords such as "Google", "cloud", and "web" as shown in Fig. 11. With the word "Google", the users also spot the word "glass" in the same topic stream. They are mentioned frequently in 2012 and 2013. This is the time that Google prepared and released Google Glass, a smart and hands-free glass, which can connect to many devices and display personal information. Google Glass is a good example of the overlappings between big data and IoT, but there is also a big concern about privacy. Terms "cloud" and "web" are conspicuously the most frequent keywords in all three topic categories. During the period from 2010 to 2011, the topic stream shows that keywords like "twitter", "social", and "news" are being mentioned frequently together. Twitter is a social network platform for microblogging, which was created in 2006 by Jack Dorsey, Noah Glass, Biz Stone, and Evan Williams[4]. Twitter had a breakthrough development by introducing a new interface and adding many features. The number of tweets in a day increases rapidly[5].

Besides the stories and insights gained from the visualization, in terms of improving the systems, the users also suggest the feature of specifying detail posts by a particular author, that could provide the evolution of author concerns over time and easily explain the context of trending words.

VI. Conclusion and Future work

This paper presents an interactive visualization technique for media conversations in topics of IoT, Big Data, and Cybersecurity. *HackerNets* aims to showcase the competition between topics and their detailed user network fluctuation overtime. We demonstrate the applications on the dataset from social media, showing that *HackerNets* could quickly highlight the emerging topics and could assist users in exploring entities' relations. Cybersecurity has been the dominant topic for discussion for over twelve years, while IoT and Big Data are the emerging fields in the current Industry 4.0 and will continue to be the topics of interest.

Future work will focus on expanding the visualization to other related aspects such as privacy policies and user behavior. A user study for a broader scope of users is expected to produce more generalized results. Also, more interactive features should be supported to optimize *HackerNets* layout.

References

[1] C. Perera, R. Ranjan, L. Wang, S. U. Khan, and A. Y. Zomaya, "Big data privacy in the internet of things era," *IT Professional*, vol. 17, no. 3, pp. 32–39, 2015.

[2] L. Atzori, A. Iera, and G. Morabito, "The internet of things: A survey," *Computer networks*, vol. 54, no. 15, pp. 2787–2805, 2010.

[3] Alexa, "Amazon Alexa Official Site: What is Alexa?" https://developer.amazon.com/en-US/alexa, Amazon, Accessed: 2019-10-18.

[4] Alphabet Inc., "Actions on Google — Actions on Google," https://developers.google.com/assistant, Google Develop., Mountain View, CA, USA, Accessed: 2019-10-18.

[5] Google Developers, "Google Assistant—Your Own Personal Google," https://assistant.google.com, Accessed: 2019-10-18.

[6] K. R. Sollins, "Iot big data security and privacy versus innovation," *IEEE Internet of Things Journal*, vol. 6, no. 2, pp. 1628–1635, 2019.

[4] https://www.twitter.com/

[5] https://www.buffer.com/resources/how-twitter-evolved-from-2006-to-2011

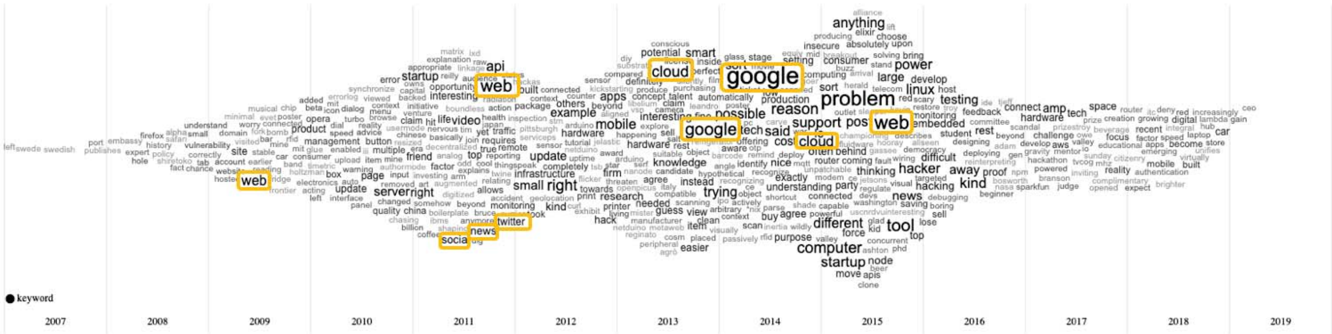

Fig. 11. Topic stream in our *HackerNets* visualization. Frequent keywords are highlighted: "google", "cloud", "api", "social", "news", "twitter", and "web".

[7] A. Parsons, "Using social media to reach consumers: A content analysis of official facebook pages." *Academy of marketing studies Journal*, vol. 17, no. 2, 2013.

[8] C. P. Mayer, "Security and privacy challenges in the internet of things," *Electronic Communications of the EASST*, vol. 17, 2009.

[9] R. Roman, P. Najera, and J. Lopez, "Securing the internet of things," *Computer*, no. 9, pp. 51–58, 2011.

[10] Article 29 Data Protection Working Party, "Opinion 8/2014 on the recent developments on the internet of things," https://www.trustarc.com/blog/2014/05/29/internet-of-things-industry+-brings-data-explosion-but-growth-could-be-impacted+-by-consumer-privacy-concerns, 2014.

[11] H. N. Nguyen, V. T. Nguyen, N. V. Nguyen, V. Pham, and T. Dang, "IoTNegViz: An interactive tool for visualizing negative aspects of IoT," in *2018 IEEE International Conference on Big Data (Big Data)*, Dec 2018, pp. 4565–4568.

[12] H. Sundmaeker, P. Guillemin, P. Friess, and S. Woelfflé, "Cluster of european research projects on the internet of things," *European Commision Vision and challenges for realising the Internet of Things*, 2010.

[13] TRUSTe Research, "Internet of Things Industry Brings Data Explosion, but Growth Could Be Impacted by Consumer Privacy Concerns," 2014.

[14] A. Zaslavsky, C. Perera, and D. Georgakopoulos, "Sensing as a service and big data," *arXiv preprint arXiv:1301.0159*, 2013.

[15] P. Nummi, "Social media data analysis in urban e-planning," in *Smart Cities and Smart Spaces: Concepts, Methodologies, Tools, and Applications*. IGI Global, 2019, pp. 636–651.

[16] A. Macarthy, *500 Social Media Marketing Tips: Essential Advice, Hints and Strategy for Business Facebook, Twitter, Pinterest, Google+, YouTube, Instagram, LinkedIn, and More!* CreateSpace Independent Publishing Platform, 2018.

[17] C. A. Bail, L. P. Argyle, T. W. Brown, J. P. Bumpus, H. Chen, M. F. Hunzaker, J. Lee, M. Mann, F. Merhout, and A. Volfovsky, "Exposure to opposing views on social media can increase political polarization," *Proceedings of the National Academy of Sciences*, vol. 115, no. 37, pp. 9216–9221, 2018.

[18] A. Benton, G. Coppersmith, and M. Dredze, "Ethical research protocols for social media health research," in *Proceedings of the First ACL Workshop on Ethics in Natural Language Processing*. Valencia, Spain: Association for Computational Linguistics, Apr. 2017, pp. 94–102. [Online]. Available: https://www.aclweb.org/anthology/W17-1612

[19] C. C. Aggarwal and C. Zhai, *Mining text data*. Springer Science & Business Media, 2012.

[20] N. V. Nguyen, V. T. Nguyen, V. Pham, and T. Dang, "Finanviz: Visualizing emerging topics in financial news," in *2018 IEEE International Conference on Big Data (Big Data)*. IEEE, 2018, pp. 4698–4704.

[21] Q. Li, J. Wang, Y. P. Chen, and Z. Lin, "User comments for news recommendation in forum-based social media," *Information Sciences*, vol. 180, no. 24, pp. 4929–4939, 2010.

[22] G. Stoddard, "Popularity dynamics and intrinsic quality in reddit and hacker news," in *Ninth International AAAI Conference on Web and Social Media*, 2015.

[23] J. Venn, "I. on the diagrammatic and mechanical representation of propositions and reasonings," *The London, Edinburgh, and Dublin philosophical magazine and journal of science*, vol. 10, no. 59, pp. 1–18, 1880.

[24] M. Pirooznia, V. Nagarajan, and Y. Deng, "Genevenn-a web application for comparing gene lists using venn diagrams," *Bioinformation*, vol. 1, no. 10, p. 420, 2007.

[25] D. S. Johnson and H. O. Pollak, "Hypergraph planarity and the complexity of drawing venn diagrams," *Journal of graph theory*, vol. 11, no. 3, pp. 309–325, 1987.

[26] S. Chow and F. Ruskey, "Drawing area-proportional venn and euler diagrams," in *International Symposium on Graph Drawing*. Springer, 2003, pp. 466–477.

[27] B. Martin, W. Chadwick, T. Yi, S.-S. Park, D. Lu, B. Ni, S. Gadkaree, K. Farhang, K. G. Becker, and S. Maudsley, "Vennture–a novel venn diagram investigational tool for multiple pharmacological dataset analysis," *PloS one*, vol. 7, no. 5, p. e36911, 2012.

[28] P. Bardou, J. Mariette, F. Escudié, C. Djemiel, and C. Klopp, "jvenn: an interactive venn diagram viewer," *BMC bioinformatics*, vol. 15, no. 1, p. 293, 2014.

[29] T. N. Dang, N. Pendar, and A. G. Forbes, "Timearcs: Visualizing fluctuations in dynamic networks," in *Computer Graphics Forum*, vol. 35, no. 3. Wiley Online Library, 2016, pp. 61–69.

[30] J.-w. Ahn, C. Plaisant, and B. Shneiderman, "A task taxonomy for network evolution analysis," *IEEE transactions on visualization and computer graphics*, vol. 20, no. 3, pp. 365–376, 2013.

[31] C. Blundell, Y. W. Teh, and K. A. Heller, "Bayesian rose trees," *arXiv preprint arXiv:1203.3468*, 2012.

[32] S. Liu, X. Wang, J. Chen, J. Zhu, and B. Guo, "Topicpanorama: A full picture of relevant topics," in *2014 IEEE Conference on Visual Analytics Science and Technology (VAST)*. IEEE, 2014, pp. 183–192.

[33] S. Liu, M. X. Zhou, S. Pan, Y. Song, W. Qian, W. Cai, and X. Lian, "TIARA: interactive, topic-based visual text summarization and analysis," *ACM TIST*, vol. 3, no. 2, pp. 25:1–25:28, 2012. [Online]. Available: http://doi.acm.org/10.1145/2089094.2089101

[34] F. B. Viegas, M. Wattenberg, and J. Feinberg, "Participatory visualization with wordle," *IEEE transactions on visualization and computer graphics*, vol. 15, no. 6, pp. 1137–1144, 2009.

[35] T. Dang, H. N. Nguyen, and V. Pham, "WordStream: Interactive Visualization for Topic Evolution," in *EuroVis 2019 - Short Papers*, J. Johansson, F. Sadlo, and G. E. Marai, Eds. The Eurographics Association, 2019.

[36] L. Byron and M. Wattenberg, "Stacked Graphs - Geometry & Aesthetics," *IEEE Transactions on Visualization and Computer Graphics*, vol. 14, no. 6, pp. 1245–1252, 2008. [Online]. Available: http://ieeexplore.ieee.org/document/4658136/

[37] H. N. Nguyen and T. Dang, "EQSA: Earthquake Situational Analytics from Social Media," 2019.

[38] M. Bostock, V. Ogievetsky, and J. Heer, "D3 data-driven documents," *IEEE Trans. Vis. Comput. Graph.*, vol. 17, no. 12, pp. 2301–2309, 2011.

[39] R. Amar, J. Eagan, and J. Stasko, "Low-level components of analytic activity in information visualization," in *Proc. of the IEEE Symposium on Information Visualization*, 2005, pp. 15–24.

[40] YCombinator. (2007) Hacker news. https://news.ycombinator.com/.
[41] G. Inc. (2012) Bigquery. https://cloud.google.com/bigquery/.
[42] V. Pham, V. T. Nguyen, and T. Dang, "IoTViz: Visualizing emerging topics in the internet of things," in *2018 IEEE International Conference on Big Data (Big Data)*. IEEE, 2018, pp. 4569–4576.