“Visual” #Thanksgiving

Yan Ma Yam14@pitt.edu
Jiayu Liu Jil146@pitt.edu
Yiran Li Yil129@pitt.edu
Shujun Zhang Shz53@pitt.edu

ABSTRACT
As a capturer of 2014 American Thanksgiving on Twitter, our project aims at providing a platform to allow people share this festival feeling, discovering what others are talking about on Twitter, the people living in different states may have various concerns about this Thanksgiving festival.

So with the efforts of reviewing previous related researches, we enriched our project with detailed technical support, and diverse approaches to visualize social networks. And with the “almost comprehensive” project design, we also encountered problems. However, we are not the first ones. By borrowing the former explorers’ solutions, we found out how to use Twitter API to access the instant Twitter data, use php to clear up the messy data, and make right choices to employ various layouts and interactions to reveal the Thanksgiving topic in the United States.

As a result, we are proud to show this visualization with unique topic on Thanksgiving to users, and gained their approval and also feedbacks. These will promote future improvements on this visualization system definitely.

Categories and Subject Descriptors
H.5 [**Information Interfaces and Presentation**]: Miscellaneous;

D.1.7 [**Software**]: Visual Programming

General Terms
Documentation, Design, Human Factors.

Keywords
Social Visualization, Information Visualization, Twitter, Geographical Map, Node-Link, Force-Layout.

1. INTRODUCTION
With the rapid development of social network, the distance between people is shortened. Nowadays, the social network like Facebook and Twitter construct almost all the social communication among human beings who use the Internet. On this kind of network, new friends are made; best wishes and cares are sent among family members and friends.

Presentation’Group-jil146, December 10, 2014, Pittsburgh, PA, the United States.

As students in Information Sciences School in University of Pittsburgh, who are studying how to make visualization design, the idea of capturing 2014 Thanksgiving on Twitter, the second year living and studying in America, come to our minds. Through our project, the users can witness the warm wishes and concerns via browsing Tweets posted during the Thanksgiving period (November 21st to December 6th). For an instance, mothers and wives would post some Turkey and dessert recipe in order to make delicious family-gather-together meals; shoppers would focus on Black Friday and Cyber Monday to purchase on-sale stuff, and also buy some Christmas gifts in advance; etc. So the motivation of this Visualizing Thanksgiving on Twitter is to capture every possible feeling in Thanksgiving in the United States, so that the ones who missed Thanksgiving dinner can take this chance go over this American festival.

As for the uniqueness of this visualization project, we find that developers rarely notice this topic on Thanksgiving. In other words, our project will enrich the domain of Twitter visualization and add a little humanity into it. So as we put this initial designing idea forward to our professor, she encouraged us to go forward, and also gave us some meaningful suggestions on how to construct the whole visualization.

However, while having this ambitious designing conception, we also encounter several problems on designing overall visual layout, crawling data from Twitter, employing php to clear up and organize huge amount of Tweets, etc. Fortunately, there are other people encountered these kinds of issues before, so that we can borrow their approaches to solve our own problems.

2. RELATED WORK
As for related work, although our group had a rough picture about what we were moving forward, the literature review does provide a supplement of details. The articles help us to find the best ways to collect, analyze, and present data.

Heer and Boyd’s *Vizster: Visualizing online social networks* [4] gives our group a framework that guides us how to perform the whole project. On the comprehensive level, as the goal of both our group and the authors is to visualize the social network, especially to emphasize the characteristics of the community structure, this article reminds us of the importance of the node-link as a visualizing technique. In practice, node-link is being used to visualize e-mail communication [3], early online social networks [1], co-authorship in publication domain [6], etc. In the other word, the layout of link-node can reveal the relationship between components. On the practical level, this article provides us how to present collected data from the sections of layout, basic interaction, exploration, navigation, and search. Especially, the structure and organization of this paper benefits us how to develop this report, which section should contain what kind of information so that readers can better understand and use this. And the research on end-users’ usage really reminds us of the importance

of developing user-friendly system: exploring the system from users' perspective rather than developers'.

Besides the above main article, two other ones supply some basis of visualizing Twitters. França, Sayama, McSwiggen, Daneshvar, and Bar-Yam [2] employed static visualizing methods to show daily, weekly, and sleep-waking patterns of more than 6 million Tweets in New York City and surrounding areas. Apart from that, Marcus, Bernstein, Badar, Karger, Madden, and Miller [5] designed a Tweets crawler and a timeline-based visualization, which are called TweeQL and TwitInfo respectively. The former can extract and facilitate the analysis of Tweets data; the latter one is a dashboard that combines 6 kinds of methods to visualize Tweets dynamically.

In a conclusion, all of these three papers do supplementary function to construct our approach to visualize Twitter. Although the first fundamental paper does not mention visualizing Twitter, the authors teach us that using node-link layout to visualize the connections within social network would be a good choice. And the other two projects show us different perspectives or approaches of our project.

So, as beginners of Twitter visualizers, these researchers really help our team to construct the Twitter visualization project from the aspect of framework, crawling data techniques, variety options of visualizing methods.

3. VISUALIZATION DESIGN

This section reveals the rationale of designing the overall visualization. With the understanding of reasons why we chose each visualizing method, the latter followers can pick up the useful or effective methods, and discard ineffective ones. Apart from following implementers, the designing rationale can also help users know the background knowledge of each layout component so that they can better make use of the visualization.

3.1 Design process

Technically speaking, the processing of designing and developing this Thanksgiving on Twitter visualization project can be separated into the following steps.

Initial designing idea. With the very foremost idea of tracing Americans' Thanksgiving topic on Twitter, we used Microsoft Word to generate the initial blueprint as shown in **Figure 1**. The pull-down menu enables users to select dates. As one date is selected, the map will change its depth of color according to amount of Twitters. The detail about state's name and Twitter's number will be in a small window with mouse touching the state. And when one state of America is clicked, there will be a new curve graph to show Twitter's hourly amount in the United States.

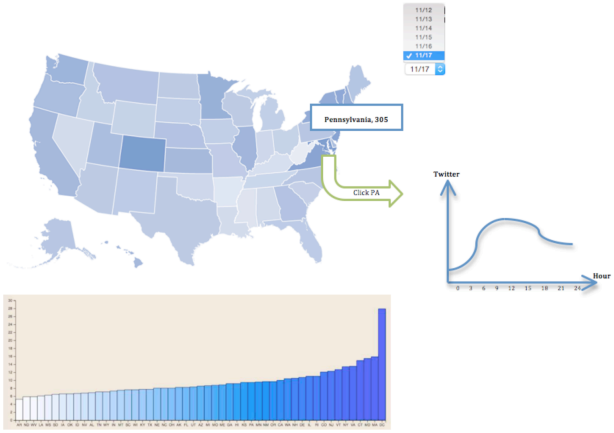

Figure 1. Initial designing idea

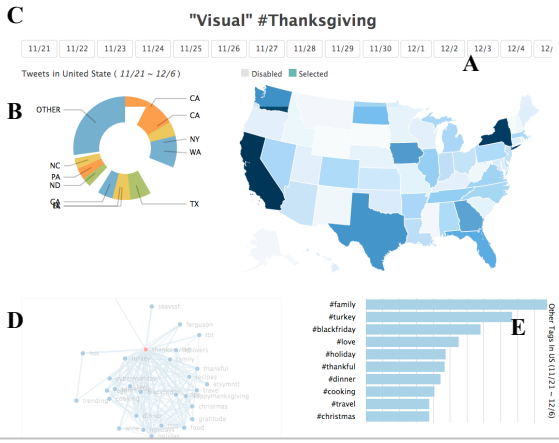

Figure 2. Final layout of the design

A is an American map with each state selectable. B is a pie chart to compare the amounts of Tweets. C is a list of dates with selectable buttons. D is a node-link that shows popular hash tags. E is a bar chart to present the most popular hash tags. This screenshot is the initial situation of the visualization. In the following **Interaction** sub-section, there will be more detailed explanation of how each layout interacts with others.

Crawling data via Twitter APIs. With efforts to search online websites about the experience of former Twitter visualization explorers, we began to use Twitter API to collect Tweets with "#thanksgiving" as their hash tag in their Tweets' content. And as the fluency of crawling, we set crawler to collect according Tweets every half hour from November 21st through December 6th. In this way, we can capture as many targeted Tweets as possible in order to construct the whole situation of Thanksgiving. As a result, we got around 48,000 Tweets as shown in **Figure 3**

Using php to clear up data. As we found that some Tweeters' location were not the location names rather like "all around the world, the place in your heart." So we decided to employ php to filter Tweets, and to collect items with American states' abbreviation. In the end, we filtered and got 29,399 Tweets from the initial 48,000.

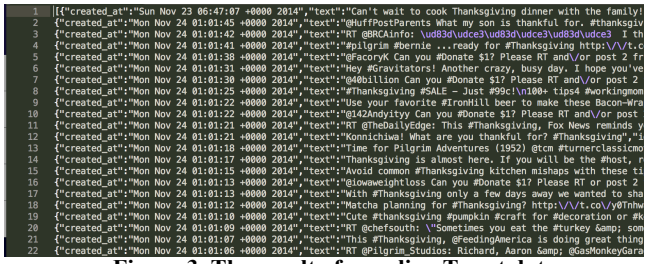

Figure 3. The result of crawling Tweet data

Improving layout. As we found that the layouts in the initial designing were not interconnected with others, we built five components with interactions as shown in **Figure 2**: Geographical American map (A), Pie Chart (B), Calendar Buttons (C), Link-Node (D), and Bar Chart (E). And their relationships will be explained in the following sub-section **3.4 Interactions**.

3.2 Layout

Geographical map (A). The reason why we selected to use geographical map is that: we want to compare the Tweets' amounts throughout the whole America country. The color depth variation could deliver the differences of the Tweets' amounts in each state in a specific period at the first glance. Geographical map could highlight the relationship between the Tweets' amounts and the location distribution, and express this relationship in a clearer way. And if the states can be clicked, there can be more interactions with other components.

Pie Chart (B). The difference of the Tweets' amounts in different states is essential to our visualization task. We take advantage of a pie chart to label this difference in a numerical way. By comparing the sizes variations in the pie chart, the changes of Tweets' amounts are obvious. And combining with the state's name label, the comparison work about different states' Tweets amounts can be completed

Calendar Buttons (C). The Calendar Buttons are involved to highlight the changes in date. The comparison works not only can be done through the whole country, but also could be implemented across distinctive day.

Node-Link (D). As node-link having the capacity to reveal the relationships between different nodes by drawing links based on some numerical correlations. We design a node-link graph based on force-directed layout, which set each hash tag as an individual node, to explore

Bar Chart (E). Bar chart is another carrier to complete the comparison between different states' Tweets' amounts. The reason why we choose this format is that the bar chart is the most fluent used layout, which means this layout is also intuitive to most people.

3.3 Visual Encodings

Technically speaking, most layouts bear the most intuitive data. Geographical Map (A) has the Tweets' amounts of each American state; Pie Chart (B) not only compares Tweets' amounts among different states, but also presents the percentage that particular state's Tweets' amount/other states' total Tweets; Calendar Button (C) is most intuitive for users to select the date; Node-Link enables end-users to discover the relationship between the 30 most popular hash tags, which is also complementary to the Bar Chart (E); and although the Bar Chart has only 10 most popular hash tags' popularity, it is most widely used format to compare. Also, we hope that Bar Chart performs as a welcomer to initial users to explore the relationships within the overall visualization by themselves.

3.4 Interactions

Almost all the layouts can interact with each other.

Scenario 1: the overall situation in the nation on one specific date. So the most possible solution is to select the targeted date in the Calendar Button (C), which will lead other layouts to present the according data. The result is shown in **Figure 4**.

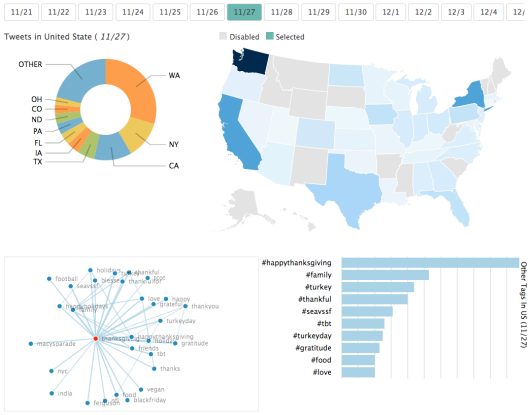

Figure 4. The overall situation on Thanksgiving Day (11/27)

Scenario 2: discover one specific date and one hash tag. The solution would be: click on the targeted date in the Calendar Button (C) and then click on the targeted hash tag in Bar Chart (E) or Node-Link (D). Usually, it is more convenient to find a targeted hash tag in the Bar Chart (E). And the relationship between the most popular 30 hash tags will be in the Node-Link layout (D). The result is shown in **Figure 5**. Also, the geographical map reveals which states' people talked most.

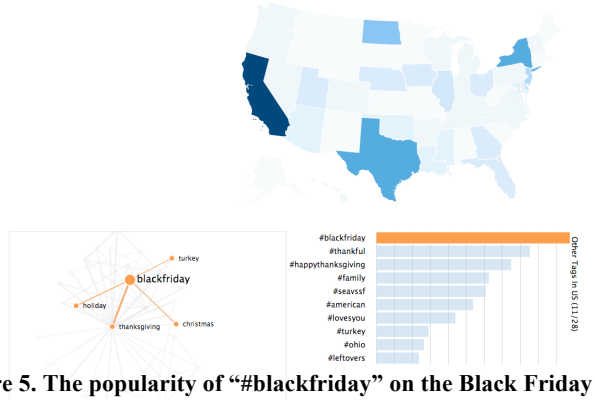

Figure 5. The popularity of "#blackfriday" on the Black Friday (11/28)

Scenario 3: the overall situation in one specific state on one particular date. The most intuitive option is to select on that targeted state, as shown in **Figure 6**.

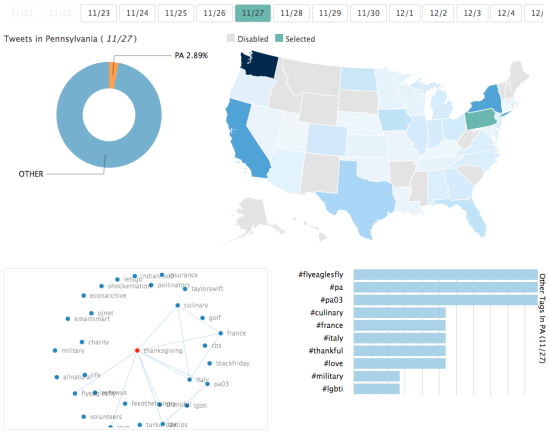

Figure 6. The overall situation with the state of PA selected

Scenario 4: if the user interests at a particular topic. The solution would be selecting on the targeted hash tag in the Bar Chart (E) or in the Node-Link (D).

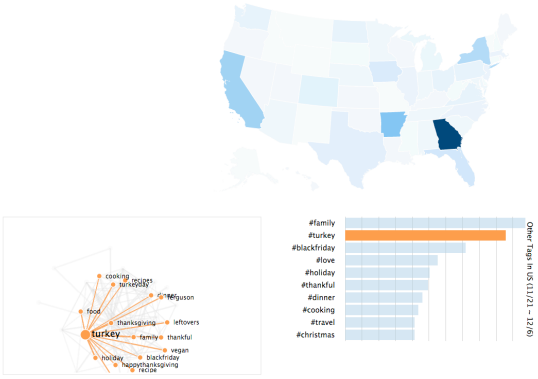

Figure 7. Focus on "#turkey"

4. IMPLICATION NOTES

In this "Visual" #thanksgiving visualization project, we mainly use d3, php, and nodeJS.

Specifically, we first used **php** to capture targeted data from Twitter every half hour in our targeted duration. And after getting such a messy dataset, we used **nodeJS** to filter effective data items, by deleting those items with other location information such as "the place in your heart", etc, and organizing data into a format that **d3** can use. With ideal dataset, we used **d3** to construct the end-user interface to present the data. Specifically, we added layouts one by one, by referring the examples in d3js.org (https://github.com/mbostock/d3/wiki/Gallery); refining them, like linking one layout with other layouts, adjusting some arguments to make some small changes, and adding some other elements to make them better looking.

5. USAGE OBSERVATION

After accomplishing our visualization layout designing, we survived three friends of our group members.

All of these four subjects are aged around 24 (2 females and 2 males), born in China. And they can speak normal-to-fluent English. The differences between them are their majors: two girls are studying in GSPIA (Graduate School of Public and International Affairs, University of Pittsburgh), one male is a student from Carnegie Mellon University, interested in computer sciences and can develop websites, and the last male subject is Information Sciences student from our school. So all of subjects know how to operate computer, search the Internet to find information. Through our observation, they know what the visualization is before the experiment.

At first, we told our subjects the main idea of this Twitter visualization is to reveal what people are concerning about around the period of Thanksgiving in the United States. After they getting the idea of what the visualization is and that we want to collect users reviews from them, subjects began to explore the layouts. And our group members would sit besides them, in order to answer their questions, observe their exploring movements, and collect every possible feedback from them.

Through close observations, we find that the exploration can be separated into the following steps: 1) knowing what the purposes of the visualization have, 2) randomly pressing buttons and trying to find the connections and interactions between components, 3) bearing some doubts, 4) if there are someone answering questions or having some friends explore the visualization together with them, there will be positive emotions to inspire them explore deeper, 5) having some conclusions about this visualization, 6) by comparing with other similar systems that they used before, subjects would like to tell their usage experience and give improvement feedbacks.

Besides with the above exploring pattern, our group members also find a turning point of exploration. Although all subjects felt interested and excited about our visualization topic at first, it seemed to be difficult for beginners to continue exploring if their questions not get answered. And as they know how the interaction functions, they will turn to be very happy. As a positive result, they would like to tell us how it functions, telling us that they found new functions, or new way to explore this visualization system, and definitely will recommend this system to their other friends. However, as a negative reaction, the subjects would quit the exploration once they ran into dead ends: especially, they could not figure out the connection of the interaction.

As the main part of users' feedbacks, we feel that we still need to take many aspects into consideration. "Why there is no search column?" "Your project group can add a initial welcome page, with animation to instruct users how to use this visualization." "More complementary descriptive languages or pictures should be added in order to better connect these components!" Thanks to feedbacks like these, we know that this project needs future optimization. But, so far, with the limited time and developing skills, we did our best.

6. DISCUSSION

As a project to visualize Thanksgiving on Twitter, it has many advantages. Firstly, it is a small program so that it needs little time to run. As for initial users of our visualization, it is essential. Because the slow reaction would definitely hinder the processing of users explore the new system. And the enthusiasm will be curbed somehow. Secondly, although the project is small, there are a lot of components to explore. Through the observation of user experience, we find that it takes at least three minutes at

average to learn this visualization system. And subjects would find many interaction and connections, which attracts deeper exploration. Thirdly, as we focusing on Thanksgiving, people who missed spending time with their family or travelled away from the United States can use our accomplish to capture the already past time. Through using the visualization system, they can find out what people are talking about (hash tagging) Thanksgiving all around the United States.

Although we are very proud to have this project finish, there are also many places need improved in the future. 1) As the initial page, we think that adding a welcome instrument on how to use this system would make this system user-friendlier. 2) More components can be added into this system, such as the emotion analysis, the picture of original Twitter, etc.

7. CONCLUSION

As a summarization of what we have done, we are inclined to write this paper to introduce our initial motivation of doing this project, review of related work on visualizing social network, description of the developing rationale and visualization system architecture, and also a small survey on user experience.

Through searching the Internet and reviewing previous papers on social networks, we find that our topic, Thanksgiving on Twitter, is very unique. And this try will definitely enrich the domain of Twitter visualization.

In the processing, we had the help from Dr. Lin and precious researchers to construct and refine the whole visualization layout, solve technical problems like how to collect data from Twitter, improving the interaction by linking all components, etc.

And after finishing the designing and development, we also conducted a survey to observe how people use and think about our Thanksgiving visualization. More importantly, their feedbacks inspired us, and let us know how to improve it.

ACKNOWLEDGEMENT

Our thanks to Dr. Yu-Ru Lin, who is the lecturer of INFSCI 2415 / LIS 2690: Information Visualization. This course is so fundamental and essential that gives us the basics of visualization designing, especially the on-hand activities to operate visualizing tool - d3.js and basic concepts of visualization layout, e.g. forced-layout, link-node. And thanks to Dr. Lin's encourage and suggestions, the whole visualization design of this Visualizing Thanksgiving Twitter project get improved. Specifically, the layout turns to be more user-friendly and connective between components.

After all, besides all the visualizing techniques we learnt, we also want to thank the opportunity to meet each other and form this wonderful group. In the process of doing projects and group activities, we try to respect and inspire others, thus fortunately turning to be a cooperative group.

REFERENCES

[1] Adamic, L., Buyukkokten, O., & Adar, E. A social network caught in the web. *First Monday, 8*(6), 2003.

[2] França, U., Sayama, H., McSwiggen, C., Daneshvar, R., & Bar-Yam, Y. 2014. Visualizing the "Heartbeat" of a City with Tweets. *arXiv preprint arXiv:1411.0722*. November 2014.

[3] Fisher, D., & Dourish, P. Social and temporal structures in everyday collaboration. *Proceedings of the SIGCHI conference on Human factors in computing systems*. ACM. April 2004, 551-558.

[4] Heer, J., & Boyd, D. Vizster: Visualizing online social networks. In Information Visualization, 2005. INFOVIS 2005. IEEE Symposium. IEEE. October 2005, 32-39.

[5] Marcus, A., Bernstein, M. S., Badar, O., Karger, D. R., Madden, S., & Miller, R. C. Processing and visualizing the data in tweets. *ACM SIGMOD Record, 40*(4), 2012, 21-27.

[6] Newmann, M. E. J. Co-authorship networks and patterns of scientific collaboration, *PANS*, 101: 5200-5205, 2004.