

VisTwit: Talking Together about News Visualization with Twitter

Deok Gun Park*

Niklas Elmqvist†

Lorraine Kisselburgh‡

Purdue University

ABSTRACT

Interesting online news visualizations are often discussed in Twitter, a conversation medium that is separated from the original visualization. We present VisTwit, a method that feeds back realtime Twitter discussions about a news visualization to the visualization itself, thereby promoting more involved discussions. Integrating VisTwit in a web-based visualization is easy: news editors simply define anchoring points in the visualization where Twitter messages containing associated keywords will be shown. The design space of VisTwit includes selection criteria of the ranking methods to filter messages, design of sparklines to show information scent, and methods for facilitating structured conversation between readers.

Index Terms: H.5.3 [Information Interfaces and Presentation]: Group and Organization Interfaces—Web-based interaction; H.3.5 [Information Interfaces and Presentation]: Online Information Services—Web-based services

1 INTRODUCTION

Web-based news visualizations, which provide a visualization of news content, are gaining in popularity. Media outlets, such as the *New York Times* and *The Guardian*, use web-based visualization as a mean to increase understanding of the data related with news, thereby providing fresh insights to the readers. However, these visualizations are delivered as finished content produced by the journalist, and lack the capacity to reflect crowd-based responses and comments about the work. Even though the readers can comment and link the visualization from social network services, such as Twitter or Facebook, their messages are fragmented to their social networks. This fragmentation limits interaction with other readers.

In this paper, we propose VisTwit, a Twitter-based web service for overlaying real-time Twitter messages onto web-based visualizations to encourage discussion around the data. News visualizations typically have many graphical entities such as circles, rectangles, and lines well as text labels and annotations. To integrate VisTwit with such a visualization, the editors simply create one or several *anchors* and link them to these graphical entities. Anchors also have associated keywords and ranking rules that are used to partition matched Twitter messages between them. Then, using the Twitter Search API, VisTwit (1) aggregates tweets that mention the visualization URL, (2) partitions the tweets between anchors based on the textual content and the anchor keywords, and finally (3) ranks them according to the salience and uniqueness of the messages. When a reader of the news visualization interacts with these entities, matched and aggregated Twitter messages are overlaid on top of the original visualization to show related Twitter messages from other viewers about these subjects.

*e-mail: park473@purdue.edu

†e-mail: elm@purdue.edu

‡e-mail: lorraine@purdue.edu

For example, let us assume we are integrating VisTwit with a heatmap visualization about government budgets. Each rectangular region has a text label about the budget category. In this case, we will create an anchor for each category, linking it to the respective rectangular shape on the visualization and using the name of the budget category as a keyword. When users hover the mouse over each region on the news visualization, VisTwit will overlay matched Twitter messages about each budget item. In this way, VisTwit can be thought as a feedback loop as shown in Figure 1. This integration of commentary with the news visualization will contribute to the experience of readers in the following way:

- An enhanced sense of belonging in the readership community;
- Overview about the acceptance of a news visualization; and
- In-depth discussion at the level of individual keywords, which facilitates structured conversation about a news visualization.

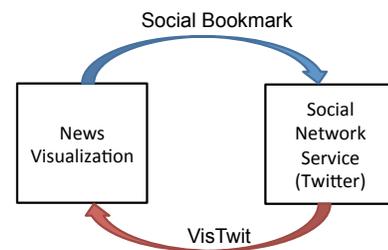


Figure 1: Concept diagram for the VisTwit system. The visualization can be linked to the social network service using social bookmarking techniques. The proposed VisTwit system creates a feedback loop of the messages back into the visualization so that these messages can be used as context-aware annotations of the visualization.

This paper presents the following contributions: (i) the recognition of Twitter messages as a source for automatic annotations for news visualizations; (ii) an anchor-keywords scheme to enable structured conversation on a visualization; and (iii) the design of information scent using sparklines to convey the status of the conversation.

2 RELATED WORK

A number of web-based visualization systems to aid collaborative sensemaking process have been developed [7, 11, 12], most of which have sophisticated comment or annotation functions. However, according to a followup study of Many Eyes conducted by Danis et al, the real discussions on a visualization often happened outside the site [2]. News visualizations are usually accompanied by a social network bookmark, where users link the visualization to their social network and continue discussions in those communities. This results in fragmented and distributed discussion isolated from the visualization itself. This limitation is suboptimal because feedback on a contribution increases additional contributions [1]. Therefore, fragmenting the discussion community decreases the opportunity for comments to generate additional contributions. VisTwit tries to solve this problem by developing a feedback loop from the social network service back to the visualization.

How to present the collective insights from readers is an important research problem. The most common form of generating feedback loops is with list-type comments, which are used in Many Eyes [11]. However, for highly graphical content such as visualizations, annotations superimposed on the visualization may be a better approach [7]. For example, Contextifier [8] tries to annotate a stock line graph from news article sources. In VisTwit, the visualization format can be any information visualization format, and the information source is Twitter messages. Previous work have also use Twitter as an information sources [3]. Visual Backchannel [5] uses them to track large-scale events. Vox Civitas [4] helps journalist find information sources, and introduced the concept of message utility based on uniqueness. In VisTwit, we are extending the audience of this data from journalists to general news readers.

3 THE VISWIT SYSTEM

VisTwit is a Twitter aggregator system that collects, partitions, and summarizes Twitter messages about a visualization and then feeds back this data to the visualization itself as an annotation layer superimposed on the visualization. The basic idea is to aggregate Twitter messages that contain the visualization URL and visualize their content on the visualization, thereby closing the feedback loop.

We will explain VisTwit features and components using an existing visualization named *512 paths to the White House*¹, which was published in the New York Times on November 2, 2012 by Mike Bostock and Shan Carter. Figure 2 shows the visualization, where users can explore permutations of election scenarios based on the results of nine states. It was very popular, resulting in over 10,000 tweets containing a link to the visualization.

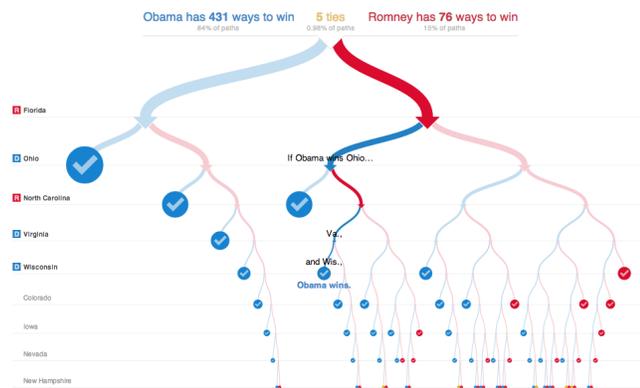


Figure 2: A screenshot of the “512 paths to the White House” news visualization from the New York Times website.

3.1 Aggregating and Partitioning

The Twitter messages that VisTwit overlays on a visualization can be thought of in two dimensions: (a) general discussion, and (b) specific details. General discussion about a visualization is straightforward to collect by simply searching for the URL using the Twitter Search API. Representations of such general discussion may include the total number of tweets, the most retweeted messages, or the most favorited messages. Also, a timeline of tweets mentioning the visualization can show the current stage of the conversation, i.e., whether it is gaining popularity or losing momentum.

¹<http://www.nytimes.com/interactive/2012/11/02/us/politics/paths-to-the-white-house.html>

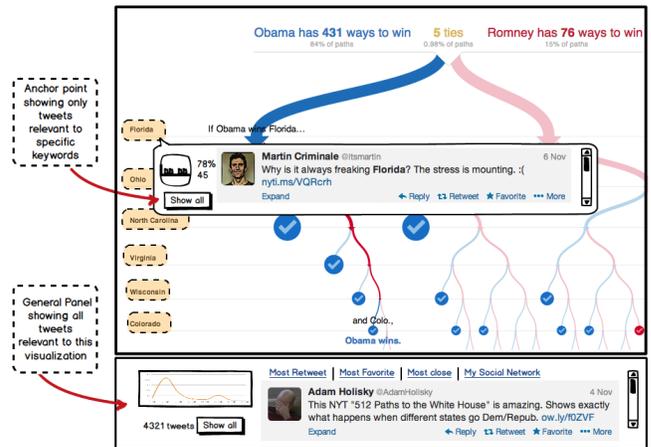


Figure 3: Example of layout showing the general trends and specific details. At the bottom of the visualization, general trends are shown with the line chart showing the volume trend of tweets. Representative messages are shown based on a ranking function. Messages cycle in 3-5 seconds, and the user can adjust the selection criteria, such as “Close to me”, “My social network,” or “Most retweeted.” The specific messages are shown when the mouse hovers over the anchor point. The related messages filtered with predefined keywords are shown. When the message is highly interesting, the user can react by *replying* or *retweeting*, which will promote in-depth discussion.

However, general messages may miss specific details in individual tweets. For example, when a viewer is interacting with the state of Florida on the visualization, tweets about Florida are likely interesting to the viewer. To remedy this, we propose the concept of an *anchor* as a way to partition tweets down to the level of individual entities on the visualization. An anchor consists of (a) a graphical entity on the visualization, and (b) a set of associated keywords. The editors of the news publication create anchors by selecting graphical entities and adding associated keywords. For example, in the 512 paths visualization, the entities may be the text labels representing the states, and the associated keywords can be the name of the states. In this way, readers who hover the mouse over the anchor entity will see a popup showing interesting tweet messages relevant to that entity. Furthermore, based on information foraging theory [10], we draw a sparkline showing the trend associated with the anchors to provide *information scent* to the users.

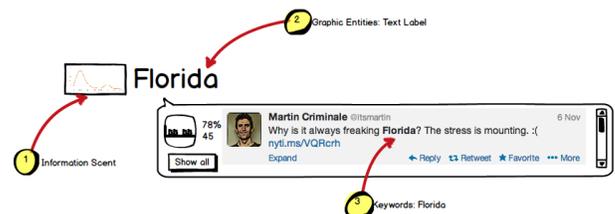


Figure 4: Example of an anchor: The text label *Florida* is associated with the keywords *florida*, *fla*, *fl*. A sparkline representing information scent shows the relative occurrence of this keywords from all relevant tweets. When the user hovers over the Florida label, the tweets mentioning Florida and the visualization will be overlaid as a popup.

3.2 Ranking Tweets

There are approximately 10,000 tweets containing a link to the example visualization. Even with anchors to partition these tweets into specific categories, there is still a large volume of data. For example, there are more than one hundred tweets that each contain the word *Florida* or *Ohio*. This shows the need for a ranking mechanism to sort tweets according to additional criteria:

- **Social Distance:** Message that originate within or near the reader's social network should be exposed more extensively.
- **Influence:** Users with many followers have more impact.
- **Retweets:** Important messages tend to be retweeted.
- **Favorited:** Favorited messages should have more weight.
- **Geographic Distance:** Depending on context, messages from close geographic distances may have more meaning.
- **Recency:** Recent messages are often most relevant to readers.

However, these kind of selection criteria may not be optimal for situations where comments have the form "I like this," which tend to be common in social media services. Incorporating a degree of uniqueness [4] in the ranking mechanism may thus be desirable.

3.3 Design of Information Scent

VisTwit will only show specific Twitter messages when the user hovers the mouse over a graphical element associated with specific anchor. However, gaining an overview of these anchor points and identifying what elements are receiving interests from other users is difficult. According to information foraging theory [10], the concept of *information scent* is one way to provide hints about the existence of information. However, news visualizations are usually designed to convey large amounts of information effectively. Adding new graphical components using VisTwit can interfere with the original visualization, which is not desirable. To alleviate this, VisTwit should provide information scent using a minimal visual footprint.

We suggest using a sparkline, which has a small visual complexity, to communicate the number of tweets in both absolute terms, as well as a relative percentage of that anchor's content. Depending on the design, new Twitter messages being aggregated containing that keyword can be made to flash subtly. In this case, the data entities that are receiving more Twitter messages will draw more interest.

4 DISCUSSION

The major benefit of VisTwit can be explained using the concept of feedback theory [6]. Positive feedback happens when the output of the system promotes the output function. For example, if there are no comments, users will be less motivated to contribute themselves, which is a vicious cycle. However, when there are many interesting comments for the users, the user would be motivated to contribute more in forms of answering question or arguing against earlier comments. It will again increase the possibility of the interesting comments. This creates a virtuous cycle. The role of VisTwit here is to increase the number of impression of other people's response, thus facilitating discussion, which results in quality comments.

Highlighting specific keywords for specific visual elements in the visualization can promote discussions about the specific. For example, another recent visualization, named "Constellations of Directors and Their Stars"², shows the relationship between directors and their frequent actors. However, of 408 tweets containing a link to it, only two of the tweets contained the reference to the specific entities, such as names of directors or actors, even though there are

²<http://www.nytimes.com/newsgraphics/2013/09/07/director-star-chart/>

many such entities in the visualization. The contents of tweets were mostly "I like this" or "check out this interesting thing" kind of messages. This shows the limitation of current conversation mechanisms for visualization. We speculate that adding anchors to specific entities of the visualization may better facilitate the in-depth discussion over visual entities.

Such in-depth discussion enables collective sensemaking, where people can use the comments and discussions of others to interpret and make sense of novel events and information [9]. In other words, the ultimate goal of VisTwit is to transform static and "read-only" visualizations into collective sensemaking platforms.

5 CONCLUSION AND FUTURE WORK

This paper presents the VisTwit system, where Twitter messages about news visualization can be fed back into the visualization. The anchor point, where graphic entities in the visualization are associated with predefined keywords, show relevant comment about that area of the visualization, thus creating context-aware annotation, which promotes in-depth discussion about the data.

REFERENCES

- [1] C. Cheshire. Selective incentives and generalized information exchange. *Social Psychology Quarterly*, 70(1):82–100, 2007.
- [2] C. M. Danis, F. B. Viegas, M. Wattenberg, and J. Kriss. Your place or mine?: visualization as a community component. In *Proceedings of the ACM Conference on Human Factors in Computing Systems*, pages 275–284, 2008.
- [3] N. Diakopoulos, M. De Choudhury, and M. Naaman. Finding and assessing social media information sources in the context of journalism. In *Proceedings of the ACM Conference on Human Factors in Computing Systems*, pages 2451–2460, 2012.
- [4] N. Diakopoulos, M. Naaman, and F. Kivran-Swaine. Diamonds in the rough: Social media visual analytics for journalistic inquiry. In *IEEE Symposium on Visual Analytics Science and Technology*, pages 115–122, 2010.
- [5] M. Dork, D. Gruen, C. Williamson, and S. Carpendale. A visual backchannel for large-scale events. *IEEE Transactions on Visualization and Computer Graphics*, 16(6):1129–1138, 2010.
- [6] D. Fleder and K. Hosanagar. Blockbuster culture's next rise or fall: The impact of recommender systems on sales diversity. *Management Science*, 55(5):697–712, May 2009.
- [7] J. Heer, F. B. Viégas, and M. Wattenberg. Voyagers and voyeurs: supporting asynchronous collaborative information visualization. In *Proceedings of the ACM Conference on Human Factors in Computing Systems*, pages 1029–1038, 2007.
- [8] J. Hullman, N. Diakopoulos, and E. Adar. Contextifier: automatic generation of annotated stock visualizations. In *Proceedings of the ACM Conference on Human Factors in Computing Systems*, pages 2707–2716, 2013.
- [9] L. Nelson, C. Held, P. Pirolli, L. Hong, D. Schiano, and E. H. Chi. With a little help from my friends: examining the impact of social annotations in sensemaking tasks. In *Proceedings of the ACM Conference on Human Factors in Computing Systems*, pages 1795–1798, 2009.
- [10] P. Pirolli and S. Card. Information foraging. *Psychological review*, 106(4):643, 1999.
- [11] F. B. Viégas, M. Wattenberg, F. van Ham, J. Kriss, and M. McKeon. ManyEyes: a site for visualization at internet scale. *IEEE Transactions on Visualization and Computer Graphics*, 13(6):1121–1128, 2007.
- [12] M. Wattenberg and J. Kriss. Designing for social data analysis. *IEEE Transactions on Visualization and Computer Graphics*, 12(4):549–557, 2006.