Discover Diamonds-in-the-Rough Using Interactive Visual Analytics System: Tweets as a Collective Diary of the Occupy Movement

Wenwen Dou and Derek Xiaoyu Wang and Zhiqiang Ma and William Ribarsky

University of North Carolina at Charlotte

9201 University City Blvd,

Charlotte, NC, 28262

Abstract

The phenomenally wide-adoption of social media has stimulated a new means in organizing and carrying-out modern social movements. Exemplified by the Occupy Movement (OM), rich information, including protest-related events and people's responses to those events, is posted and shared through social media sites such as Twitter. However, it is quite challenging to make sense of such valuable information in a collective manner, as it is often submerged by all the other content on Twitter. In this case study, we demonstrate the combination of computational methods (e.g., topic modeling and event detection) and interactive visual analytics in facilitating users to examine how relevant tweets can reflect a collective view of a social movement. In particular, we focus on discovering and associating key events throughout the OM. Based on the event frequencies, our system helps users to divide the movement into three distinct stages. Information regarding "what" the events were about, "when" and "where" the events occurred, and "who" were involved is extracted from the tweets to describe each stage of the movement. The resulting case studies show that we can indeed construct a collective diary of the social movement by analyzing events extracted from the content of the tweets.

Many discussions have been generated recently on the topic of social media and the effect it may play on the formation and mobilization of social movements. Social movement is a type of group action that involves individuals and organizations who focus on bringing social changes. In this process, social media serves as a medium for the masses to make their voices heard and to initiate and organize social movements. Recently, we start to see the impact of the pervasive use of social media on organizing large-scale social movement. Such usage is illustrated by the Bank Transfer Day (BTD), a Facebook-launched call aimed to move money from big financial institutions to local credit unions. While the BTD lasted for only two weeks (Christian 2011), it produced significant results, including an estimated 1 million consumers moving their accounts to credit unions (CSMonitor 2011). Through assessing whether the self-proclaimed goals have been met, it is relatively straightforward to determine the impact of such bursty and focused protests.

However, evaluating the impact of large-scale social protests with massive participation and a wide range of goals, exemplified by the Occupy Movement (OM), is an extremely challenging task. Even though such movement has subsided significantly in recently months, it is still on-going a year after the proclaimed start date–Sept 17, 2011. Although traditional media report selective excerpts about Occupy protests, they often do not adequately capture the scale, the response, or the opinions of the protestors' and citizens'. Understanding the overall Occupy Movement and the impact of its protest activities can shed light not only on how the movement was organized and evolved, but also on the effect of public policies dealing with the protests. Since the Occupy Movement is known to use social media to advertise, organize, and attract protesters (Wikipedia 2012b), we believe social media data such as tweets can be used as a direct source to depict all major events.

In this paper, we demonstrate the use of our visual analytics system, which combines interactive human-discoveries and a set of computational methods (including topic modeling and event detection), to enable domain users to effectively extract major events throughout the Occupy Movement. An interactive visual interface is designed to present the events in an intuitive manner for making sense of the causes, the actions and the impact of the events. More importantly, through presenting information regarding "who, what, when and where" of the protests, our system facilitates the domain users (e.g., law enforcement officers) to the construction of a collective diary of the movement using tweets.

Data and Computation Architectures

Different from previous work (Marcus et al. 2011; Shamma, Kennedy, and Churchill 2010), in this paper, our system focuses on extracting event-related insights from the content of the tweets. Such analysis is conducted based on the tweets collected from Twitter's GardenHose API from 07-11-2011 to 09-18-2012. To provide more focused analyses on the movement, this tweet collection is further filtered with a general hashtag #occupy*. The resulting dataset captures a wide range of relevant tweets (∼430,000) and hashtags, including ones created to represent ideas and concepts exemplified by #occupydemocracy, #occupyteaparty, #occupybank, etc.; and others that denote specific protest locations such as #OccupyBoston, #OccupySeattle, #OccupyMadrid, and etc.
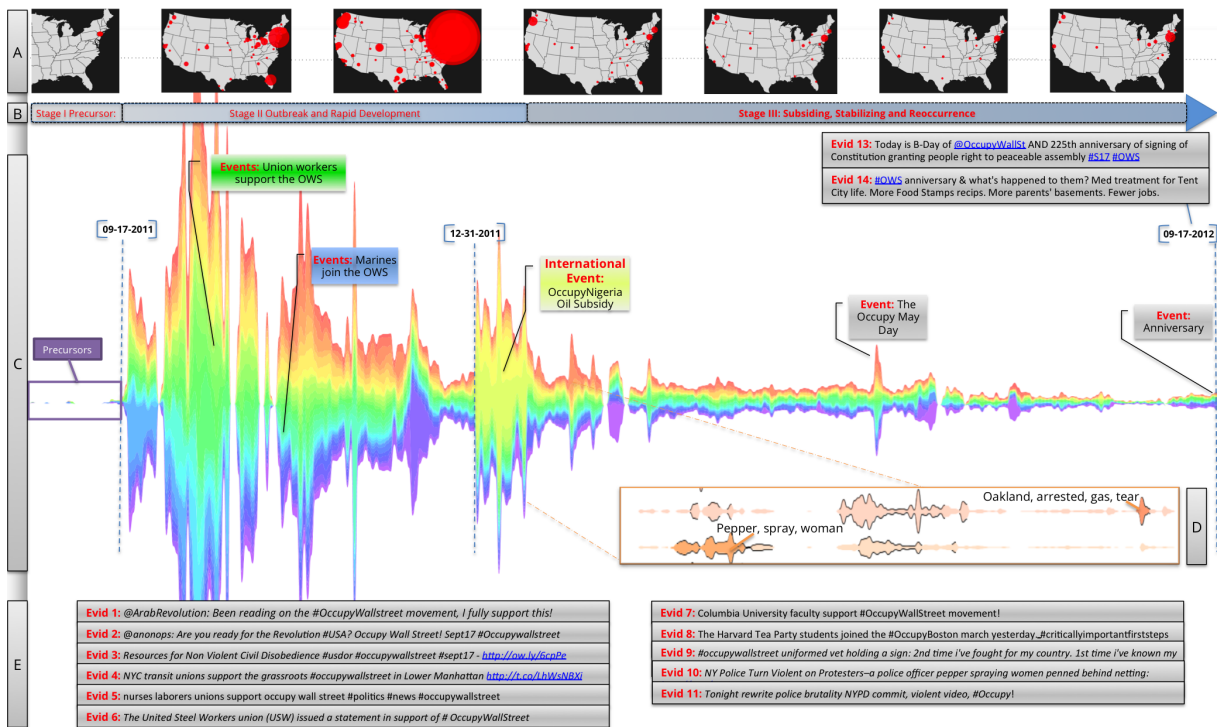
Figure 1: Overview of the OM in our VA system. *A*: Occupy hotspots over time. *B*: Three stages of the OM divided based on the rise and fall of the overall activities. *C*: Visual summary of the Occupy activities. The x-axis represents time; each color-coded ribbon represents a topic extracted from the tweets. Event detection is performed on individual topic to identify bursts as indicators of events. *D*: Sample events labeled with corresponding keywords. *E*: Evidence [Evid.X] mentioned in the paper. Note, here multiple visualizations are placed together due to the page limit. Detail about other views refer to (Dou et al. 2012)

The combination of the automatic data analytics and human-discovery is at the heart of our visual analytics system. Using the computational architecture introduced by Wang et al. in (Wang et al. 2012), we first applied data cleaning-and-aggregation processes to remove the contextual noise in the tweets. Then we performed topic modeling and event detection to extract meaningful insights, such as topical trends and patterns (Dou et al. 2012). Furthermore, we also sifted through features like Hashtags to extract key geospatial and temporal features.

The gist of the analysis environment provided by our system can be thought of as first organizing all tweets based on topic and time, which is visualized in Figure 1 B with each stream representing all tweets that focus on the specific topic. With the X-axis being time, the volume (or the thickness) of the stream denotes the amount of tweets discussing the topic on a given day. Then an event detection method (Dou et al. 2012) was applied to identify bursts within each topic stream as indicators of events. Figure 1 D illustrates the resulting events highlighted by black contours and labeled with a few keywords that best describe the event. In summary, our computational architecture extracts events in the form of topic bursts comprised of tweets discussing the event. The visual interface further presents the extracted events throughout the OM in a temporal flow, with keywords indicating what each event was about.

Human-discovery is involved in this process at any given time. Users can then interactively make sense of the individual events through the visual interface, examining the tweets discussing a specific event to discover information regarding "who, what, when, and where", as well as analyzing people's responses to certain events. Altogether, we provide users a interactive visual analytics system that helps them effectively construct a diary about a social movement.

Case Study: Developmental Overview of the Movement

Based on the major rise and fall of the overall Twitter discussions of the Occupy protests, we as users can divide the movement into three major stages as shown in Figure 1.

Stage I: Precursor and Preparation
Are Occupy protests orchestrated or spontaneous?

In Figure 1, our system confirms the self-proclaimed start date of the OM and clearly illustrates its outbreak (significant boost in volume) on 09-17-11. However, the more interesting and provoking pattern (highlighted in purple) lies in the activities before the official start date. Such pattern is picked up by our computational methods and strongly signals a precursor to the protest. This precursor let us believe that the OM is an orchestrated event rather than a spontaneous gathering. It further provides us insights on how the

movement was organized and the major players involved in this preparation stage of the movement.

Specifically, the earliest tweet in our dataset was posted on 07-21 with content showing strong support of the movement [Evid.1]. With our interactive visualization, one can directly link to a tweet pointed to a blog published by Adbusters (a Canadian-based anti-consumerist organization) on 07-13. This tweet proposed a peaceful occupation of Wall Street to protest both corporate influence on democracy and the increasing inequality in wealth distribution (Adbusters 2011). In addition, our geographical visualization suggested that, at this planning stage of the movement, all geo-entities extracted from the tweets are directly concentrated on New York City (Figure 1 map left), using both GPS location and our Named Entity Recognition techniques. Based on these findings, one can infer that the Adbusters was involved in the early organization of the movement, which is later verified by other sources such as major news media (e.g., LATimes).

Our VA system further helped extract evidence that signaled the diverse contributions from other major organizers that eventually lead to the outbreaks of the movement. For example, the hacker group Anonymous posted tweets to motivate the crowd on Aug 19 [Evid.2]. Meanwhile, other contributing groups such as US Day of Rage (#usdor) advocated a non-violent movement by providing legal info for the movement on 08-27-2011 [Evid.3]. Just by quickly glancing through our visualization of tweets from this stage, one can already identify these three major groups that helped organizing the Occupy Movement, and further determine the orchestrated nature of the movement.

Stage II: Outbreak and Rapid Development

Who helped OM jump on the public bandwagon? Who joined the protest? And for what cause?

The organization of the OM has led to a successful outbreak on the 09-17. As suggested by the significant traffic jump on the #occupy topic in the following two months, one can tell that the movement entered a rapid development stage. As further evidenced in the maps in Figure 1, the rapid, almost instantaneous, geospatial spread was another contributing factor to the successful outbreak. The key to decode such rapid growth and to assess its impact lies in users' depiction of the demographics and the motivations of the people contributing to the movement.

Through interacting with the event results from our computation methods, one can quickly identify a few representative groups that participated and influenced the Occupy Movement. Through examining tweets discussing the events, the cause for their endorsement of the movement can be derived. For example, our visualization clearly suggested (Figure 1) that, on 09-28, the New York Transit Workers Union voted to support the OM, bringing hundreds more protestors on the street [Evid.4]. Furthermore, our Hashtag analysis revealed that this also led to a much larger endorsement from several more labor unions supporting the OM, only within days from prior move [Evid.5,6]. Only a few days later, our visualization (the blue event rectangle) further indicated that burst topics labeled with keywords "marine, country, fought" emerged on 10-02. This captured the facts

that veterans joined the movement to act as a first line of defense between the police and protesters (DailyKosGroup 2011)[Evid.9]. One week later, university faculties and students showed their support to the OM on 10-10. [Evid.7&8], where they banded together to make their voices heard, citing the rising amount of student loan debt and the increasing cost of college (HuffingtonPost 2011).

The massive endorsements from these diverse groups, in such a short time period, suggested that the success of Occupy protest largely relies on the wide spectrum of citizen participation. Some came to protect the protestors, while others wanted to voice their frustrations. Using the analysis and visual interaction, one was able to quickly identify the groups and infer their motivations for endorsing the OM.

Is there violence?

Although the OM was set to hold peaceful demonstrations as advocated in Stage I, a few events with violence-related keywords raised our suspicion. With large numbers of confrontations between the protestors and the police throughout the movement, however, some protests inevitably involved violent acts. Judging by the keywords associated with the events, one could quickly identify a few events that were associated with violence. On Sep 24, an event associated with keywords "pepper, spay, officer" (Figure 1 D) emerged, and the discussion of such violent acts lasted for at least two weeks on Twitter. There were tweets accounting for and condemning the violence that police used in stopping protestors and voiced for justice [Evid.9,10]. A week after that incident, people started to report more violence via Twitter, as shown in [Evid.11]. These kind of events attracted lots of public attention and media coverage (Observer 2011), which pressured NYPD to speak publicly about and investigate the incident. More recently, on Aug 3, 2012, New York City has reported to opt out to defend the officer pepper-spraying two female protesters.

Another event labeled with "oakland, arrested, tear, gas" (Figure 1 D) occurred on Jan 28, 2012 also suggested the involvement of violence. The message *"#OccupyOakland being teargassed smoked bombed & shot at w rubber bullets."* has been retweeted multiple times. Later on during the event, people tweeted *"It appears that an Iraq War Veteran was arrested #occupyoakland"*, and expressed dismay in *"Mass arrests in #Oakland. SO un-American. Sickening to watch."*

Through interactive examination of the activities throughout the movement, one can quickly depict events that occurred at different times but of similar violent nature, ascertaining the involvement of violence in the protests.

Stage III: Subsiding, Stabilizing and Reoccurrence

After the outbreak and rapid rise of the OM, the discussion of domestic (U.S.) occupy activities on Twitter started to subside significantly after the first week of Jan 2012 (shown in Figure 1). Multiple US-based events before the sudden decrease were related to evictions of the Occupy encampment. It is, therefore, reasonable to infer that the sudden drop in tweets was due to such eviction policies. This finding is validated by evidence in a recent article on a brief look back at the Occupy Movement. After a series of evictions started

in New York City on 11-15-2011, Occupy lost its ability to organize without places to gather (ABCNews 2012).

After the evictions, domestic Occupy activities started to stabilize, as the general tweet volume shrunk (Figure 1). One small climax of the Occupy activities returned on May 1, 2012 as the Occupy May Day was organized as an effort to re-energize the movement (ABCNews 2012). More recently, activities during late August and early September were protests during the Republican National Convention and the Democratic National Convention (Charlotte, NC shown as a hotspot in the map to the right). Although the protests were organized weeks before, some were disappointed in the low turnout of the participation. When the Occupy Movement marked it first anniversary on Sep 17, 2012, many were cheerful [Evid.13], while others were not quite happy with the progress so far [Evid.14].

Discussion

To analyze the overall OM, our visual analytics system first facilitated users to divide the movement into three major stages based on the rise and fall of the activities. Within each stage, users can visually select a few representative events that are automatically discovered using the computational methods (Wang et al. 2012; Dou et al. 2012) . Due to the page limit, we focus the discussion mainly on domestic events throughout the OM. But it is worth noting that the movement is widely spread across more than 30 countries (as discovered in our analysis) in addition to the US. As a matter of fact, one significant sub-event (in terms of strength and length) detected by our system was the OccupyNigeria (Figure 1 light green) that was widely participated in Nigerian because of the end of government oil subsidies.

Precision and Recall of our VA System

To validate the accuracy of our automatically extracted events, we compared our results against the only comprehensive timeline of the OM available through Wikipedia (Wikipedia 2012a; 2012b). The Wiki timeline listed 41 major events while we have 60 events extracted from our Occupy related tweets. Two experimenters independently graded the accuracy of extracted events against the Wiki timeline, producing average recall of 79%. In terms of precision, we did not limit the baseline to the Wikipedia Occupy timeline. Instead we used news media to validate the existence of our identified events, which produced a high precision at 95%. Despite the precision and recall results, we acknowledge that there may be other events missed by our detection, should there be a well-compiled timeline of events throughout the OM. There are also cases that multiple co-occurring events appearing to be one big event in the detection results. Therefore, more user interactions need to be supported to allow users to refine the automatically discovered event results. However, we believe our approach provides a novel method to construct post-hoc diary of the OM and lead to comprehensive understandings of the movement. The system not only enables the depiction of meta-information of the "when" and "where", but also provides content information regarding the scale of the event, its

lasting time period, and the responses to the event from protestors and citizens, etc. Through presenting the case study, we hope to showcase the capability of our computational methods and interactive visual interfaces in analyzing and reconstructing social movements in a collective manner.

Conclusion

In this paper, we showcase a novel visual analytics system that facilities the users to construct a collective understanding of the rise and the fall of the social movements. Exemplified in studying the OM on Twitter, our system can effectively help users to identify different stages of the movement. Through studying the events near the shifts between the stages, it can further lead to the discovery of the probable cause for the changes. Last, examining tweets discussing certain events, the system helps users to derive information including who were involved, where were the hotspots, people's opinions, and many others alike.

References

ABCNews. 2012. A look back at the rise and fall of occupy. http://abcn.ws/OYiVo2.

Adbusters. 2011. Occupywallstreet. http://bit.ly/n4va4T.

Christian, K. 2011. Bank transfer day. http://www.facebook.com/Nov.Fifth.

CSMonitor. 2011. Bank transfer day: How much impact did it have? http://bit.ly/vWaOgv.

DailyKosGroup. 2011. Marines join ows movement. http://bit.ly/S3PuQB.

Dou, W.; Wang, X.; Skau, D.; Ribarsky, W.; and Zhou, M. X. 2012. Leadline: Interactive visual analysis of text data through event identification and exploration. In *Visual Analytics Science and Technology (VAST), 2012 IEEE Conference on*.

HuffingtonPost. 2011. College sympathizers of ows walk out of class in support. http://huff.to/pnWOmS.

Marcus, A.; Bernstein, M. S.; Badar, O.; Karger, D. R.; Madden, S.; and Miller, R. C. 2011. Twitinfo: aggregating and visualizing microblogs for event exploration. In *Proceedings of the 2011 annual conference on Human factors in computing systems*, CHI '11, 227–236. New York, NY, USA: ACM.

Observer. 2011. Is ray kelly's nypd spinning out of control? http://bit.ly/S4v6Ev.

Shamma, D.; Kennedy, L.; and Churchill, E. 2010. Conversational shadows: Describing live media events using short messages. *International AAAI Conference on Weblogs and Social Media*.

Wang, X.; Dou, W.; Ma, Z.; Villalobos, J.; Chen, Y.; Kraft, T.; and Ribarsky, W. 2012. I-SI: Scalable Architecture of Analyzing Latent Topical-Level Information From Social Media Data. *Computer Graphics Forum* 31(3):1275–1284.

Wikipedia. 2012a. Occupy movement. http://bit.ly/tMOGfp.

Wikipedia. 2012b. Occupy wall street timeline. http://bit.ly/qRcMEE.