2012 ASE/IEEE International Conference on Social Computing and 2012 ASE/IEEE International Conference on Privacy, Security, Risk and Trust

City Flow: Prototype Exploration for Visualizing Urban Traffic Conversations

Jiayu Wu, Zhiyong Fu, Zhiyuan Liu*, Jiajia Pan, Huiling Long, Xu Lin, Haoqing He*, Xinxiong Chen*, Jiayu Tang*
Academy of Arts and Design, Tsinghua University
Beijing, China
*Natural Language Processing Group, Tsinghua University
Beijing, China
e-mail: {wujy08, fuzhiyong}@tsinghua.edu.cn, {liuliudong, kaluojiajia , hehq.09, amiucxx, tjy430}@gmail.com, willinglong@163.com, linxu8891@yahoo.cn

Abstract— **The paper presents City Flow, an urban traffic visualization prototype based on real post streams captured from Sina Weibo, the most popular social networking site in China. With the increasingly pervasive use of online social networks in China, these new channels attract more users than conventional social media. People use them to seek comments and points exchanges about urgent event and popular news. However, the high noise level precludes conscious or objective standpoints. Especially regarding urban traffic, the challenge is to synthesize some general characteristics of cities rather than to follow a particular event.**

City Flow captures a mass of data from the conversations about traffic issues through Sina Weibo and generates visualizations of general patterns in addition to including detailed insights. This paper describes the interface design, interaction modes, and data capturing methods. It also presents initial feedback from a usability study and observational findings in terms of the cities' characteristics. Our design considerations present both an overall sentiment and detailed perspectives. We define a hierarchy of categories including City Sentiment, Related Topics, and Time-based Posts Stream based on factors such as topic variables and time-span dynamics. This organizational strategy is intended to guide viewers away from the noise and towards a closer examination of facts.

Online social networks add a new dimension of offering views at an individual level rather than authority level when considering social issues. But this contributes to a new phenomenon of people paying excessive attention to their personal views. Our intention is to explore ways to guide people from a micro view towards a macro view in looking into online conversations. Through the prototype of City Flow we experiment and explore the methods of designing for both collective sentiment and individual attitudes on traffic conversations. The main goal is to inspire people to reflect on the general representations and to form opinions by considering individuals' perspectives.

Keywords-*social networks; information visualization; social media; design; traffic conversation visualization*

I. INTRODUCTION

Since online social networks were developed to offer people channels to talk about public topics or make personal statements, it contributes to a phenomenon of people paying much more attention to their own feelings than others'. This does not consider the situation of society as a whole. Similar to Twitter, Chinese online social networks attract people for daily chatter, conversations, sharing information, and reporting news [1]. Online social networking posts can be grouped into two major types: presentation of one's feelings and opinions on social topics or on friends' posts, and reporting trivialities of one's own daily life. Both types of posts focus on micro level of seeing the world. Our design concept starts from this problem. City Flow aims to synthesize the real streamed data related to urban traffic theme on online social networks in order to provide a macro view of social issues.

Built on the real post streams about urban traffic in Sina Weibo[1], the most popular social networking site in China, City Flow presents an overall sentiment, major topics, and detailed contents in each city. One reason we chose urban traffic is because it is an everyday life issue that is related to a lot of relevant topics. Another reason is that traffic situation is one of the most important living quality indices. It is a severe issue for most Chinese cities, so common characteristics across cities can provide a general basis for the visualization. We design a live visualization system, based on real data of traffic conversations, which allows user interactions. According to the design goal, the system is intended to transform people's views from the micro level, which has been seen as the typical observing pattern for online social networks, to the macro level, which we propose is a more meaningful way to view social issues.

[1] Sina Weibo is a Chinese social networking site, combining the typical features of Twitter and facebook, allows people to seek information and to connect with friends. With the user population rising over 250 million, it becomes the most popular social network in China.

2012 ASE/IEEE International Conference on Social Computing, Amsterdam, The Netherlands, September 3-5, 2012.

978-0-7695-4848-7/12 $26.00 © 2012 IEEE
DOI 10.1109/SocialCom-PASSAT.2012.40

We introduce three main views in City Flow: City Sentiment, an overview of the prominent sentiment of 38 major cities according to the data that can be captured on Sina Weibo (Figure 1); Related Topics, a topic summary for each city which includes the ten most popular topics and closely related topics (Figure 2); Time-based Posts Stream, a timeline presentation of posts related to a specific topic in a city (see Figure 3). By managing the views in this order, City Flow leads viewers to consider urban traffic issues step by step from a general point of view to a personal point of view in order to make informed conscious judgments.

The work of building City Flow falls into two parts. Interface and interaction design determine what elements should be shown and how to present them. Data mining methods enable the visualizations. In this paper we present the project from a design perspective. As designers, we particularly emphasize presentation of interface element arrangement and interaction mode. We also include insights on leveraging data mining to serve appropriate visualizations.

The rest of the paper is organized as follows. Section 2 provides an overview of prior related work. Section 3 describes design of the visualization system's interface and interactions. Section 4 illustrates data mining methods as they relate to visualization implementation. Section 5 presents initial feedback from a usability study. Section 6 analyzes observational findings in terms of Chinese city contexts. Section 7 discusses design principles we reflected on during the work. And section 8 concludes the paper and envisions the future work.

II. RELATED WORK

Much of the research in reporting urban traffic condition focuses on tracing physical movements to obtain real situations by using pervasive systems such as cell phone networks or GPS [10-12]. Visualizing mental expressions on urban traffic through online social networks is a relatively novel domain for information visualization. Our work is an experiment of collecting ongoing individual online posts about urban traffic to analyze the mental activities that reflect physical conditions. By building the prototype to visualize the traffic conversations we explore ways to make use of social media intelligence to refine mental expressions on traffic related issues. Instead of complementing conventional social media by presenting different views, we want to encourage viewers to engage not only active participations but also reflective observations. City Flow addresses the challenge of balancing accurate analysis and semantic interpretation when designing visualizations for traffic themes. There is some research in information visualization domain that has influenced our project on data analysis and visualization presentations.

A. Visualizing Social Topics

Much of the previous research in visualizing social topics focuses on analyzing large-scale events [4,5] or presenting news stories (Newsmap[2]). For large-scale event, the frequency of keywords appearing, the density of participation, and image gathering is important to be presented as in [4]. By providing these visual elements viewers are able to grasp a general view of what has been happening with knowing what people have said and seen regarding the ongoing event. From the journalism perspective, the part of the news that caught the most sensational movements would be valuable to assist reporting of a specific event [5]. Treemap visualizations such as Newsmap provide a way to present the connectivity between a title appearing on the visualization interface and the amount of articles related to the title. In City Flow prototype, we do not present any specific event or story but rather provide a method to mentally extract information from collected events related to urban traffic. Therefore, we choose to mainly present general views such as collective sentiments, hot topics, and the dynamic of mental activities based on time changes. We employ a treemap in the main interface of City Flow because it provides a good device for connecting sentiment (in colors) and the amount of traffic conversations (in shapes).

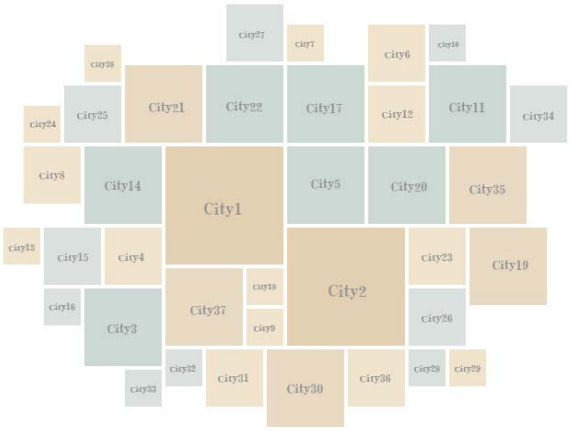

Figure 1. City Sentiment

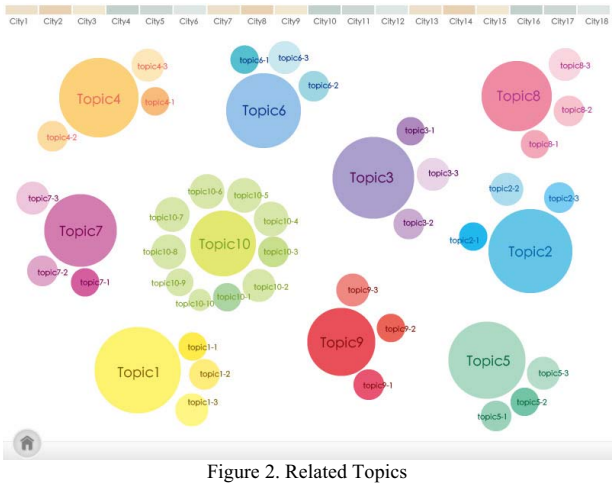

Figure 2. Related Topics

[2] http://newsmap.jp/ (Retrieved 2012-5-1)

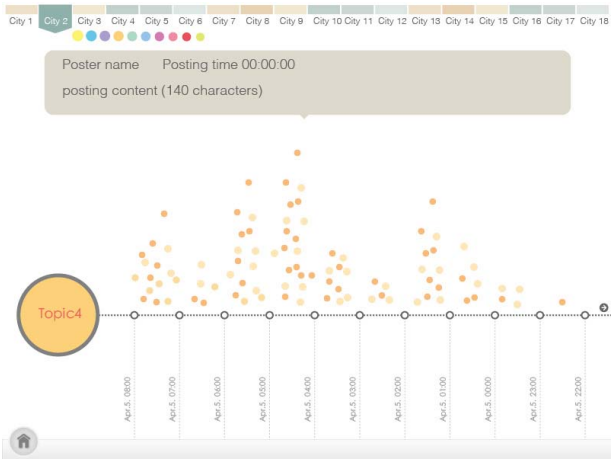

Figure 3. Time-based topic

B. Using Online Social Networks for Conversational Analysis

A variety of works explore designs for visualizing conversations through online social networks. Showing the facts is the major motivation for analyzing the online conversations such as [4,5]. There is another angel from which to look at the conversational data, which is in an illustrative or hedonistic view. We Feel Fine [6] and PeopleGarden [3] are the good examples. We Feel Fine explores designing interesting visual elements to support a comparative view so as to introduce a big picture of sensational movements. PeopleGarden creates a semantic expression to present an overview of participants' patterns for over a period of conversational activities. Both of the two works stress the importance of abstract visual elements in carrying out the conversational analysis. Our work in building City Flow prototype focuses on what semantic expression can bring the data analysis to viewers on the one hand and in what extent the facts can be reflected on the other hand.

C. Presenting Aggregate Changes

Plenty of works have presented aggregate changes from minutes to years [9]. Visualize Yahoo! Mail[3] continually presents email deliveries for each moment and the constantly changing aggregate email keywords for the last 30 minutes. The minute changing visualization makes people understand what is happening at current moment. The aggregate keywords provide a linguistic explanation to connect the temporal meaning. Other works such as Themail [9] presents conversational histories according to months and years of keywords. It stresses keywords evolution which regarding the long-term conversational activities. City Flow is designed to present the sentiment of the now and the posting dynamics of the recent past. So we employ a continually changing treemap (see Figure 1) and a 24-hour's timeline in the visualization prototype (see Figure 3).

City Flow is a relatively novel attempt to design for mental expressions reflected on urban traffic conditions. The issue of urban traffic has long been seen as the responsibility of government or urban transit organizations. What purpose citizens' casual comments can serve is not clear. We believe the rapid development of online social networks in China and the large amount of attention on traffic experiences can contribute reflective insights on living quality. We hope to extend our work on social topic visualization by exploring sentiment analysis and content refinements based on general social expressions instead of specific events.

III. INTERFACE DESIGN AND INTERACTION MODE

In this section, we mainly focus on the design aspect of the visualization prototype through discussing the three views and the interaction mode with considering the use of the data source. In order to explain the interface and interaction design, the visual implementations in terms of the benefits and constraints of the data source from Sina Weibo must be articulated. First, we will contextualize what data we choose to use by considering the data capturing permissions. The characteristics of Sina Weibo as a public communication channel for Chinese users will be explained as compare to Twitter and facebook for western users. Next, we discuss design decisions by demonstrating the three views of the visualization prototype.

A. Sina Weibo as Data Source

Sina Weibo is the most popular social networking site in China. It combines the characteristics of Twitter and facebook by integrating status updates and long-term relationship maintenance. It can be seen as a news resource provider to keep people updated with what is happening. It is also a social connection tool used to keep in touch with close friends. Some features come from Twitter type's online social networks. The 140 Chinese-character-length posting limit inspires users to keep updating and recycling news continually. Account mechanisms attract popular celebrities and public institutions to create forums for public issues. Other features resemble facebook type's social networks such as chat boxes and private conversation buttons that facilitate users to maintain close relationships.

In terms of the data that Sina Weibo authorized to capture, available data includes user ID information and the content of each post, but not the IP address of the post. User ID information includes registered gender, location, age, nickname, school name, company name, and self tags. Without the posting IP address the real traffic situation cannot be located accurately. So it can't reflect the live road information. We can collect information on both patterns of the contacts and contents of the conversations. We do not use the patterns of the contacts in our first online prototype since we concentrate on development of general sentiments and overview of the relative topics. We use the contents of the conversations in order to show viewers a micro view on observing the data flow. Since all the views of City Flow are based on city location, we make use of posters' registered locations to refine the sentiment and to classify the topics. In the Time-based Posts Stream view some of the poster's ID information such as the nickname and the time of posting

[3] http://visualize.yahoo.com/mail/ (Retrieved 2012-5-1)

appears with the posting content in the visualization prototype (see Figure 4).

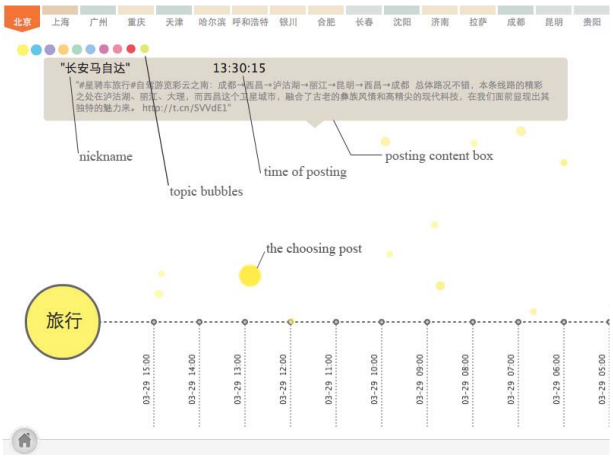

Figure 4. Posting content box with the poster's nickname and the time of posting

B. Interface Design

1) City Sentiment

In order to present viewers with a sense of general traffic patterns cross China, we introduce City Sentiment (see Figure 1) as the first view of the visualization presentation. When viewers start using the online prototype this view immediately shows up. We choose 38 major cities in this view in order to cover every province and municipality. City Sentiment presents two visual elements that are considered to be keys to viewers: the traffic sentiment of each city, and the overview of the sum of posts in each city. We employ treemap in this view to combine color with the sentiment and size of square with the sum of posts. We define colors in two ranges of hues to illustrate positive and negative sentiments. The positive sentiment is designated in warm hues and the negative in cool hues. Each range of hue is classified into four levels. The squares are defined in four sizes according to four levels of hues. Therefore, as can be seen in the figure, each square represents a city with one of the four sizes representing the amount of posts and one level of the hues representing good or bad mood. The more posts the city has the darker in color and the bigger in size of the square. Positions of squares are arranged geographically in order to display an abstract Chinese map in viewers' mind.

2) Related Topics

Related Topics is the second view of City Flow. This view presents ten most popular topics in a given city represented by bubbles (see Figure 2). In order to indicate the level of popularity of each topic, we define the size of the bubbles. The more popular the topic the bigger the bubble is. The colors are assigned to ten topics with the given size of bubbles. To manage topics, we assume the connectivity between hot topics and other closely related topics might offer clues for interpretation. In addition to present the ten major topics, other related topics are displayed in smaller bubbles by the side of the bigger topic bubbles. By this arrangement, we hope to lead viewers enter a colorful keywords view to notice the complexity of opinions after grasping the general sentiment in the first view (see Figure 6).

3) Time Based Posts Stream

According to the design goal of guiding viewers towards a closer examination of facts, we introduce Time-based Posts Stream as the last view (see Figure 3). The main area in this view is occupied by a dot-stream indicates the dynamics of conversations about a given topic over 24 hours. Each dot represents a Weibo post, viewers can scroll the timeline to the left and right to check the dynamic pattern of the posts. On the top of the posts stream there is an empty space left for present the post's content with user ID and the exact posting time (see Figure 4). By looking at this view viewers can observe the city traffic through detailed perspectives.

Through the arrangement of the three views, we hope to guide the viewer to consider social issues as a sophisticated museum visitor might regard a painting: by standing back first to look at the whole picture sometimes with squinted eyes to catch a general expression, then stepping closer to see the details of brushstrokes and color mixtures.

Figure 5. Picture symbolization of a city with the number of Weibo posts on traffic issues

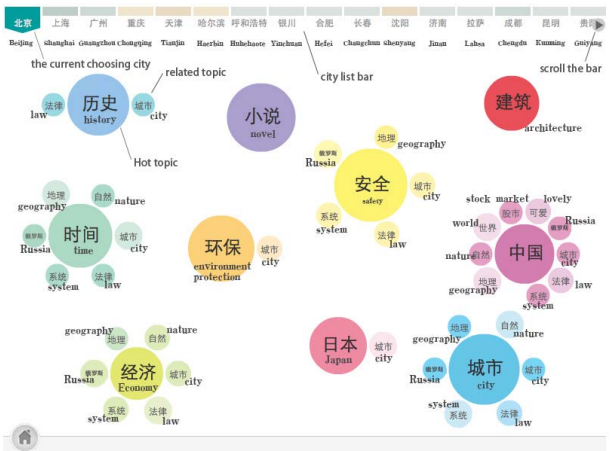

Figure 6. Related Topics view with interaction mode

C. Interaction Mode

In terms of simplicity of the three interfaces of City Flow, we keep the interaction mode simple and easy to use. The prototype can be operated simply by clicking and hovering. Animations are employed in the view switches. The front-end prototype is written in ActionScript 3.0.

In the City Sentiment view, viewers can identify the whole situation by knowing the amount of posts in each city square. When the cursor hovers over one city square it becomes a picture symbolizing that city with the number of posts showing on top of the picture (see Figure 5). Click one city square all of the squares fly to the top of the view and line up in small rectangles with city names below. The Related Topics view appears. In this view, the currently selected city is indicated by color and shape in the city list bar at the top of the screen. Put the cursor over the far right or left ends of the bar a scroll button appears letting to change the city (see Figure 6). In this way, viewers can switch to another city's Related Topic view. Move the cursor onto one of the topic bubbles, the bubble shakes indicating the topic has been chosen. Click the bubble the Time-based Posts Stream view appears. When the cursor hovers over one posting dot the post's content shows with the poster's user ID and the time of posting in the posting content box (see Figure 4). In this view the viewer can switch to another city's Related Topics by clicking the city list bar at the top of the screen. Switching to another topic's Time-based Posts Stream is also allowed by clicking the small colorful bubbles above the posting content box (see Figure 4). The timeline can be slid to browse forward and backward within 24 hours by hovering cursor over the far right or left ends. The interval on the timeline is 1 hour. When viewers enter into the timeline view, they see the most recent time in the first sight. When they move the timeline to the left, previous times appear. The first view can be accessed from each view by clicking the homepage button on the bottom of the screen.

The interaction mode is set up to facilitate switching between macro and micro views. In this way, we hope to turn viewers from a passive participant to an engaged observer when considering social topics.

IV. DATA ANALYSIS

The data analysis process includes data mining, sentiment analysis, and data connection with the visualization prototype. First, posts about traffic are filtered from all kinds of topics on Sina Weibo. Second, sentiment analysis is carried out on these traffic conversations in order to classify positive and negative posts. Then, keywords are generated within each city. All of these organized data is contained in an XML file to connect the SWF visualization frontier. The data is captured and analyzed during the last 24 hours. The online prototype is updated every 10 minutes representing the combined data in the last 24 hours.

A. Data Mining

Data is captured through Sina Weibo over 24 hours in the current version of the prototype. Posts that include "traffic", "transit", and "road condition" are collected as the initial data. Then a filter is used to distinguish posts that refer to real traffic conditions from advertisements and common phrases. The first step is to filter out common phrases like "the Bank of Communications" or "the University of Communications". The word "communication" in Chinese is equal to "transit" in some circumstances. So these phrases contain words related to traffic but not the topics we need for the visualization. The second step is to cut off advertising posts with a user ID containing words like "real estate" or "group purchase". By doing this, the posts that contain comments made by citizens about traffic are picked up for further analysis.

B. Sentiment Analysis

After fetching the traffic conversations from the Sina Weibo posts, sentiment extraction defines the emotional aspect of each post. A list that composed of positive and negative words is constructed. Each word in this list has a value for measuring emotion. The post is split into single words ready for emotional evaluation by the values of the word in the emotion dictionary. In this way, every post is assigned an emotion value and the sentiment of a city is delivered.

Emotion dictionary construction is essential for measuring post's sentiment. Existing emotion dictionaries are relatively too small for traffic themes. For example, "lovely" is a positive word in general situation, but it seems to be appropriately considered as neutral word in traffic conversations. Therefore, it is important to improve the emotion dictionary in terms of traffic issues. The method employed in this project for improving the emotion dictionary is to measure the similarity and common information appears in a word regarding the existing words in the emotion dictionary and the traffic keywords. The data used in the experiment for improving the emotion dictionary is the total posts during a week on Sina Weibo. By using this method, the emotion value of the existing emotion words in terms of traffic themes can be defined.

To measure each post's emotion, we split the post into words and pick up the emotion words, evaluate the words by the new emotion dictionary, and then aggregate the emotion values. Finally the sentiment of a post is delivered.

C. Data Connection Between Analysis and Visualization

All of the posts are catalogued by city with user ID labeled on each one. Top-ten keywords are picked up from the posts of each city during the last 24 hours using term frequency and inverse document frequency (TF-IDF) measure. For each keyword, ten relevant words at most are extracted according to their proximities to the keyword. This part is shown in the Related Topics view in bigger bubbles representing keywords and smaller bubbles representing relevant words to the given keyword. For connecting the raw data with the visualization frontier, an XML file is generated containing the emotions of each city, the relevant posts of each keyword, and the total number of posts, which will be read by the SWF visualization prototype.

V. INITIAL FEEDBACK FROM POTENTIAL USERS

We have launched the first version of the online prototype. In order to improve the City Flow system, we did a usability study in the lab. The insights from potential users contribute to designing visualizations both for urban traffic and for a wider domain of other social topics. To summarize the initial feedback, some factors are important for developing the prototype, such as the features the users are interested, the information they expect to see, and the usage scenarios they suggest.

Generally, the users are very interested in a city's sentiment, especially the comparison among cities. The city's topics and the extended relevant topics are considered to be a window for learning about a city. The users all show strong interest in the time-line based dynamic fluctuations. They state that the Time-based Posts Stream is good for observing the pattern of peak density at different times on one topic. By switching between cities, comparison of peak times for one subject can also be found in the timeline presentations.

A. Features Users Are Interested in

During the lab usability study, when the users were asked what the most attractive thing in each view is, they provided some insights.

1) Size of squares in the City Sentiment view.

Users state that the critical element of the first view is how many people are talking about traffic issues in each city. They would like to click the most popular one to find out what people are talking about. The city they live in or their hometown is the second place users like to look at and usually compare with the previous city.

2) Association between keywords and a city.

Users are interested in the connectivity between keywords and a specific city. We found that users spent quite much time trying to figure out why a keyword belongs to a city by reading the posts' contents.

3) Distinctions of keywords between cities.

Users are interested in the horizontal comparisons of topics people are focusing on in each city.

4) Timeline pattern.

Users are interested in the overview of the timeline pattern and the detailed posting time. They are most interested in posting-dots distributions since it contextualizes the message that allows them to grasp the meaning in a glance.

B. Information Users Expect

When we asked the user, what would you like to see in terms of social topic visualization, they mentioned several viewpoints.

1) Comparisons of conversational contents and sentiments according to one topic across cities are considered interesting.

2) Regarding to show the post content, the original post or the most popular comment is more attractive than the post with less people commented on.

3) Personal view should be added in the visualization prototype. The user's filter is an important personal view to select the visual items based on one's interest.

4）Images that help users to understand the situation of an emergent event are important for visualizing urban traffic. Pictures taken by witnesses at the scene of an event and post to Sina Weibo are the facts showing the causes and results. They are valuable for tracing the essence of the topic. However, for non-accident issues, pictures seem too monotonous for the user. For example, the pictures of a road and crowded vehicles do not seem very attractive.

C. Potential Use Scenario

When talking about the expected scenario of using this applet, some users claim that waiting at traffic lights in a car or on the public transport are the expected scenarios to use the tool. Due to the traffic-related theme, we assume the potential City Flow use scenario in transit. Therefore, mobile devices like cell phones and tablets would be the medium rather than a computer. Furthermore, the goal of the City Flow applet is to provide a channel for browsing traffic conversations and sentiments rather than examining real traffic information. This suggests that entertaining and gaming aspects could bring added value to the tool. By providing portable device versions, the applet can be accessible in fragmented intervals.

VI. OBSERVATIONAL FINDINGS

Since the online prototype of City Flow started capturing data, we conducted a continual observation for several weeks on the patterns of traffic conversational distributions. Here we discuss the findings that might shed light on future improvement of the visualization system. These findings provide us, as designers, some social contexts and practical insights to reflect on. Some Chinese urban lifestyles are revealed in the visualization prototype, which can be used as the inspirations for visualizing social media intelligence.

A. Observing the Cities

We summarize Chinese cities' traffic conditions in terms of the data captured by the City Flow prototype and compare the cities' contexts. There are some stable patterns we found during the continuous observation on the fluctuated data on the online prototype. Beijing, Shanghai, Guangzhou, Shenzhen, and Hangzhou are the five cities that have the most Weibo posts about traffic. These cities are all located in the east part of China, which belong to the economically developed area. Lhasa, Huhehaote, Xining, Yinchuan, and Haikou are the five cities that have the least Weibo posts about urban transit. Most of these cities are located in the west part of China. The reason for having less conversations on traffic issues is either due to the inconvenience of transit and media coverage such as in Lhasa, or owing to the relatively low density of population with simplicity of life such as in Haikou on Hainan Island in the far south part of China.

We discovered that there is close connectivity between cities. In one city's keyword view there always appears another city's name. For example, in Beijing's Related Topics Shanghai shows in one of the topic bubbles. This phenomenon substantiates insights about comparisons between cities by potential users. Mutual focus usually happens between two cities that share similar economic situations or other features for instance, proximity or ethnic relations.

B. *Observing the Topics*

There is a stable pattern for the Related Topics view. The topics that people discuss the most according to urban transit are “city”, “time”, “environmental protection”, “safety”, and “history”. No distinctive divergence has been found on topics across cities except accidental event happens. When some accident happens the Related Topics usually shows “news”, “journalist” or the keyword that represented the accident as one of the ten top keywords. For example, in the observation record of Mar. 29, 2012, in the Related Topics view of Beijing, it appears Chengdu and Children (see figure 7). As we mentioned that the comparison of cities usually happens between two cities that have similar features, Beijing and Chengdu seems an exception since Chengdu is a small southeast city with relaxing life in contrast with Beijing a big city with speedy lifestyle. When click Chengdu and Children separately we realize that it is an emergent event of children trafficking happened in Chengdu and talked widely around China. Because of human trafficking needs to move people from one place to another usually in a long distance, it is relevant to traffic here.

C. *Observing the Timeline*

There are two ways to watch the Time-based Posts Stream view. One is to watch the density of the posting dots by sliding the timeline. In this way, people get to discover the emergent event by noticing the sudden increase of the dots’ density (see figure 8). The other way is to switch the views between cities for observing the distinctions in terms of same topic. Through this way, we found the subtle differences of lifestyles among cities. For example, from the starting point and ending point of the dense posting dots during a day the general life pattern such as working time starts from getting on the transportation tool, lunch break, and the time getting back home can be unveiled. Here we introduce some discrimination. The time starts from taking transportation tool (since many people use mobile devices to access Sine Weibo on the trip to office, we assume the dense posting dots in the morning are sent on the transportation tool) in Beijing is 6 a.m., Shanghai is 7 a.m., Chengdu is 8 a.m. This maybe means Beijing has the heavier traffic stress than the other cities while Chengdu has the most relaxing living pace. The lunch break is generally between 12a.m. to 2 p.m., but Guangzhou starts the afternoon work earlier than other main cities (in the list of the top cities have the most Weibo posts about traffic) at 1 p.m.

VII. DISCUSSING DESIGN DISCIPLINES

Card et al. defines information visualization: is to amplify cognition by using computers to provide interactive visual representations [2]. For this project, Amplify cognition, means to embed insights in the visual representations for sense making rather than yield visual elements for decision making. We choose traffic issue，which usually been visualized to support decision making applications, as the experimental object in hope of exploring ways to display the general inclinations on traffic related issues by using social media intelligence. Through the prototype observation and the usability study, we also draw the reflections the potential users

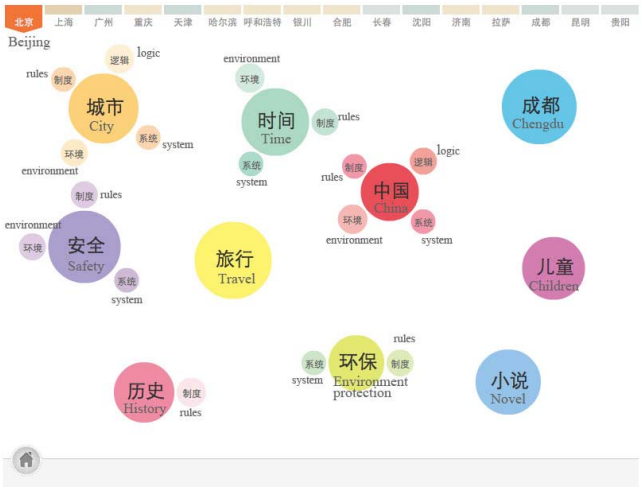

Figure 7. Screen shot of Beijing’s related topics on Mar.29. 2012 showing the accidental event on that day led by abnormal keywords “Chengdu” and “Children”

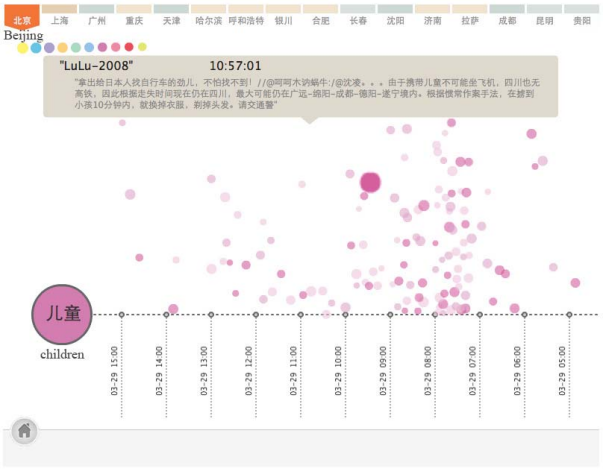

Figure 8. The sudden increase of the posting dots density for a child trafficking event, screen shot on Mar.29.2012

provided and the rules the traffic conversational visualization unearthed. With this attempt we wish to study on a critical issue for social media intelligence visualization. That is, in what extent the collective attitudes on social topics will bring viewers to think and reflect on their surroundings. To explore ways to set up general evaluations as a visual focus in the traffic topic visualization is our design challenge. In this session, we discuss about the design principles that used to visualize social topic especially traffic issues in terms of the lessons we learned and the problems to be emphasized.

A. The balance between overall inclinations and viewer’s focus

The biggest challenge for visualizing general impressions for social topics is to not only express overall inclinations but also keep the viewers' focus on details. Much of the research in predicting the overall inclinations based on a large amount of information mainly focus on synthesizing and analyzing the preference features through collecting customer's information from different resources in terms of a specific subject. The aim of this kind of research is to gather feedback during product development [7]. The target viewer is professional analyst seeking for long-term trend to determine the prospect of the product and to envision the market. However, City Flow is not designed for expert or for any working purpose, but for the novice who's just interested in knowing the patterns of attitudes and the trends of sentiments. In this sense, the visualization system needs to offer focus to satisfy the general viewers. But the general pattern of a social issue such as some relevant topics seems not close enough to make viewers to stick on. They hope to see something personally related to them. The user's relationship to the data shouldn't be too far because that is not good for understanding the whole plot. It can't be too close either because that can make the user get lost in the details. An effective solution may be providing the self-filter options in order to create personal view for viewers.

B. Connection Between Keywords and Posting Contents

Keyword is an important element for social topic visualization. Keywords are usually extracted from the conversations. So that the keyword could come from the relevant areas to the given topic, which turns out not quite overtly connected with the given topic. For example, during the observation we found words like "history" and "architecture" are not directly related to traffic topic. Therefore, to make viewer be conscious of the connection between keyword and the conversational contents is design challenge for social topic visualizations. For keyword generates a macro perspective while conversational contents a micro angle. When browsing the social topic visualization, viewers are very likely to get lost while switching between these two views. In the usability study, some users mentioned that they don't want to see every Weibo post, only the original post and the most popular comment is interesting to them. Since the potential users are interested in the connection between the keywords and the city, to display the keywords and the original posts with the most popular comments can provide a clue to understand why people talking about this topic in this city. In this way, viewers are not only able to access the contents of the conversations but also easy to learn how the keywords come from. Without the disturbance of the large amount of less important posts viewers won't get lost. We believe, social topic visualization should present an individual point of view in showing the conversational contents, so that organizing and refining the massive data in terms of viewers' value is crucial.

C. The use of Images

As the potential users claim, image is a very important element for presenting the facts and contexts of the accidental event. But the picture without distinctive features will make the viewer fell boring especially the monotonous image with crowded cars and pedestrians. Displaying a lot of unattractive images will decrease the value of people to use the system. Therefore, even image is supposed to be a necessary element in designing traffic topic visualization, filter the pictures regarding the clue of important events can be a feasible way of doing it.

VIII. CONCLUSIONS AND FUTURE WORK

City Flow is an attempt for visualizing sentiment and attitudes on urban traffic. Through synthesizing conversations in which people talk about city transit, City Flow supports analysis of emotions and issues behind the physical traffic situation. In order to nudge viewers towards making objective judgments while considering others' opinions and to have a general evaluation for their living quality, City Flow uses data from online social networks to display both collective inclinations and individual view points. The design challenge is to make these two views equal in the viewer's mind.

Currently, information visualization meets the argument that if it should support scientific information analysis or abstract semantic interpretation. "Amplify cognition" is undoubtedly been seen the crucial aspect in information visualization design. Zachary P. et al. extended the definition of information visualization with claiming casual infovis should be defined in the infovis domain. Casual infovis differs from conventional scientific analytical visualization by emphasizing "depict personally meaningful information" and "support everyday users in non-work situations" [8]. Card's definition about "amplify cognition" is redefined as "depict personally meaningful information" here, together with stressing "everyday users", Zachary's definition is more focused on the user's relationship to the data. To what extent information visualization should guide data interpretation is a critical issue faced by designers. City Flow supports to make every viewer an information analyst of urban traffic themes. Through the process of observing, thinking, and reflecting, viewers of information visualizations can learn how to apply analysis methods while considering social issues objectively. In this way, viewers can be both active participants and reflective observers by using online social networks. We believe, besides bringing forth insights, information visualization should also be an incubator for developing reflections for viewers.

To sum up information visualization based on urban traffic, basically there are two ways to conduct research. One is to display live traffic situation, for example, the light and heavy traffic routes. The other is to take urban traffic as a social issue to analyze and present the clue of a specific event in the cause, plot, and result. City Flow uses social media intelligence to generate overall inclinations of sentiment and hot topics. In doing this, it explores design a sense making and reflecting visualization application for everyday use. The future work includes:

1) Organize traffic related topic's data to deliver effective interpretations.

To avoid focus losing between city and its hot topics, several related topics about urban traffic should be designated

according to the closest connection to traffic theme. We found some topics in current version of the prototype that can be used for evidence. Through all of the cities, the most popular topics are “safety”, “environment protection”, “time”, and “economy”. Evaluating the topics by the criteria of making direct sense to urban traffic, several will be chosen to show in the Related Topics view. In this way, viewers won’t be disturbed by non-direct-related topics. In addition, the posting data should be organized in some way instead of showing all of them.

2) Design for viewer dominated filter mechanism in terms of features of portable devices.

How to transform the computer based information visualization into portable devices would be one subject for the future work. The exploration of reading patterns and interaction modes according to the small screen is a research direction for information visualization. In this sense, the definition of information visualization can be extended from based on “the use of computers” [2] to “computers and mobile devices”. The viewer dominated filter functions would grow simpler and easier to be used during moving positions.

3) Extend the research on traffic topic visualization to other social topics.

City Flow project starts from the traffic topic study in mapping social media landscape. We hope to extend the topic to every aspect of city life in the future. By analyzing how people talk about social issues through online social networks, City Flow intends to discover how we can use these data to do what for citizens. The research goal in this project is to collect information from the public then present it in a reflective way to return the information back to the public.

ACKNOWLEDGMENT

We would like to acknowledge Microsoft Research Asia as the project founder to support the research, especially to Chen Zhao for her coordination. Special thanks also give to Sara Jacobsen for detailed language correcting and critical revise opinions. Thank you also to Joanna Saad-Sulonen for your valuable advices and professor Jikun Liu for providing review feedback.

REFERENCES

[1] Akshay J., Xiaodan S., Tim F., and Belle T. 2007. Why We Twitter: Understanding Microblogging Usage and Communities. In the Proceedings of the 9th WebKDD and 1st SNA-KDD 2007(August 12, 2007, San Jose, California , USA). ACM, New York, NY, USA.

[2] Card S. K., Mackinlay J., and Shneiderman B. Readings in Information Visualization, Using Vision to Think, chapter 1, pages 1–34. Morgan Kaufmann, 1999.

[3] Donath J. 2002. A Semantic Approach to Visualizing Online Conversations. Communications of the ACM. 45,4 (April, 2002).

[4] Do ̈rk M., Gruen D., Williamson C., and Carpendale s. 2010. A Visual Backchannel for Large-Scale Events. IEEE Transactions on Visualization and Computer Graphics. 16,6 (Nov./Dec. 2010).

[5] Diakopoulos N., Naaman M., Kivran-Swaine F. 2010. Diamonds in Rough: Social Media Visual Analytics for Journalistic Inquiry. In proceedings of Visual Analytics Science and Technology (VAST), 2010 IEEE Symposium, (25-26 Oct. 2010), 115-122.

[6] Kamvar S. D., Harris J. 2011. We Feel Fine and Searching the Emotional. In Proceedings of WSDM’11, ACM Press (2011).

[7] Liu B, Hu M, Cheng J. Opinion observer: Analyzing and comparing opinions on the Web, Proceedings of the 14th International Conference on World Wide Web, 2005, 342-352.

[8] Pousman Z., Stasko J. T., and Mateas M. 2007. Casual Information Visualization: Depictions of Data in Everyday Life. Ieee Transactions On Visualization And Computer Graphics, Vol. 13, No. 6 (November/December. 2007).

[9] Viégas F. B., Golder S., Donath J. 2006. Visualizing Email Content: Portraying Relationships from Conversational Histories. In proceedings of CHI 2006, (Apr 22-27,2006).

[10] Yuan J., Zheng Y., Zhang L. H., Xie X., Sun G. Z. 2011. Where to Find My Next Passenger? In Proceedings of UbiComp’11, ACM Press (2011).

[11] Yu Z., Liu Y. C., Yuan J., Xie X. 2011. Urban Computing with Taxicabs. In Proceedings of UbiComp’11, ACM Press (2011).

[12] Zhang D. Q., Li N., Zhou Z. H., Chen C., Sun L., Li S. J. 2011. iBAT: Detecting Anomalous Taxi Trajectories from GPS Traces. In Proceedings of UbiComp’11, ACM Press (2011).