Int J Software Informatics, Volume 9, Issue 3 (2015), pp. 291–305
International Journal of Software and Informatics, ISSN 1673-7288

E-mail: ijsi@iscas.ac.cn
http://www.ijsi.org
Tel: +86-10-62661040

Visual Analysis of Sentiment and Information Spread on Micro-Blog

Chenghai Zhang[1], Yuhua Liu[1], and Changbo Wang[1,2]

[1](School of Computer Science and Software Engineering, East China Normal University, China)
[2](International Joint Lab of Trustworthy Software, East China Normal University, China)

Abstract In this paper, we combined temporal analysis and spatial analysis together, and proposed the Electron Cloud Model (ECM) which is based on the Schrodinger equation and Niels Bohr atomic theory. The ECM is used to conduct temporal visual analysis of micro-blog sentiments. In the ECM, we made an attempt to mapping a score of sentiment to the electron stability and took neutral sentiments into consideration. We applied kernel density estimation and edge bundling to conduct space-varying visual analysis of sentiment. Kernel density estimation visualized sentiment changes in different levels of detail naturally while edge bundling was used to reduce visual clutter of edge crossing and reveal high-level edge patterns. Finally, we implemented an analysis system, conducted three case studies and made simple comparisons with other visualize methods.
Key words: ECM; sentiment analysis; neutral; micro-blog sentiments; electron cloud model; edge bundling; visualization; kernel density

Zhang CH, Liu YH, Wang CB. Visual analysis of sentiment and information spread on micro-blog. *Int J Software Informatics*, **Vol.9, No.3 (2015): 291–305. http://www.ijsi.org/1673-7288/9/i222.htm**

1 Introduction

Micro-blog, one of the major sources of social information, has such features: easy operation and rapid propagation. Users could post real time messages about any things, such as their opinions on a variety of topics, discuss current issues, complain, and express positive sentiment for products they use in daily life on micro-blog. Highly developed Internet and data mining field enable research on social media data. Governments hope to monitor public opinion through micro-blog thus they could make quick responses to emergencies. Companies are seeking commercial value in these multivariate and heterogeneous data. From micro-blog data, they can get to know what the public are thinking about their company, how they evaluate their products and extent of their new product, etc. Sentiment analysis enables the access to micro-blog users' opinion. Many researchers conducted statistical analysis on micro-blog data and listed the results. To

This paper was supported in part by Natural Science Foundation of China under Grant no. 61532002 and no.61272199, National High-tech R&D Program of China (863 Program) under Grant no. 2015AA016404, the Specialized Research Fund for the Doctoral Program of Higher Education under Grant 20130076110008, and the open funding project of State Key Laboratory of Virtual Reality Technology and Systems, Beihang University under Grant No. BUAA-VR-15KF-14.
Corresponding author: Changbo Wang, Email: cbwangcg@gmail.com
Received 2014-05-31; Revised 2015-05-06; Accepted 2015-10-31.

intuitively represent analytical results, researchers developed a lot of methods, such as timeline (stream river) to get insights of data trend, geography-based method to explore geographical patterns and combinations with other methods. In this paper, we combined features of temporal analysis and spatial analysis, proposed the electron cloud model and finally conducted visual analysis on data.

Researches on micro-blog accumulated and its classification is as follows: content mining, message propagation and users' features and behavior patterns, etc. Retweeting paths and connections between posters are important micro-blog features to visualize. But few researchers concern about the sentiment of each post. On the other hand, most visualization for micro-blogs focuses on temporal or spatial visualization. Each of these methods had its limits to visualize sentiments of micro-blogs overall.

In this paper, we aim to solve these problems and combine both temporal and spatial visualizing methods. For temporal visual analysis of micro-blog sentiments, we proposed an Electron Cloud Model (ECM) based on quantum mechanics. The ECM utilizes electron transition to present changes of sentiment and takes neutral into consideration. A score of micro-blog sentiment is mapped to the state of atomic electron using its stability in Electron Cloud Model. For spatial visualization of micro-blog sentiments, kernel density estimation was used to visualize the influence of micro-blog sentiments among cities based on geographical space. By adjusting the corresponding parameters of kernel density estimation, sentiment changes in different areas can be visualized naturally in different levels of details. In order to layout the structure of edges, we adopted edge bundling based on force-directed model, considering user data of micro-blog retweeets. The use of edge bundling can remedy visual clutter of edge crossing and reveal high-level edge patterns.

The rest of the paper is organized as follows. Section 2 gives an overview of related work on visualization of sentiment for micro-blogs and edge bundling. Sentiment analysis of micro-blogs , the introduction of the Electron Cloud Model (ECM) and space-varying analysis are in Section 3. Section 4 discusses the real-life case studies. Section 5 simply compared similar analyzing methods. Section 6 concludes the paper and points to the future.

2 Related Work

Sentiment analysis has acquired a lot of achievements. There are two main methods to conduct sentiment analysis: annotated word lists and classifiers. But when talking about sentiment analysis of micro-blogs, Ahmed Nagy found it necessary to make some unique modifications. The research conducted by Nielsen proved that different sentiment lexicon and their emotional valence, especially strong obscene words and Internet slang acronyms are very important when doing micro-blog sentiment analysis[2]. Amitava Das addressed sentiment problem from the end user's perspective and proposed a generic opinion 5Ws structure to summarize and track users' sentiment[3]. Jaram Park and Meeyoung Cha et al.[4] used linguistic inquiry and word count to analyze tweeter sentiments and they found that mentions or retweets between tweeter users clearly changes the sentiments of the users, however there is no statistically meaningful level of influence of a social link in the sentiments between two connected users. Cynthia Chew and Gunther

Eysenbach tracked tweet content containing keywords such as swine flu, or H1N1 to study public health and their study illustrated some meaningful results[5]. Sajib Dasgupta and Vincent Ng[3] proposed a novel approach to identify the hidden dimension for text classification which allows a user to easily specify the dimension along which she wants the data points to be clustered and not to inspect only a small number of words. However, previous research on sentiment analysis was mainly about binary classification (positive or negative). Analysis of neutral sentiment was often ignored. Some researchers explored this field. Koppel and Schler found that neutral is very important in sentiment analysis and is not the boundary between negative and positive. It allows more accurate analysis.

To visualize layout of edges in networks, we can spatially bundle edges that traverse similar paths by reducing clutter. A lot of techniques about edge bundling had been proposed, which include hierarchical edge bundling[6], force-directed edge bundling[7], geometry-based edge clustering[8], multi-level agglomerative edge bundling[9], and graph bundling by kernel density estimation. Considering the features of micro-blog retweeting network, our method incorporates edge directions, edge weights, node weights and interactions based on force-directed edge bundling to visualize space-varying sentiments of micro-blogs.

3 Time-Space Varying Analysis

We studied and performed sentiment analysis, proposed the Electron Cloud Model (ECM) to visualize the sentiment, applied kernel density estimation and edge bundling to conduct micro-blog sentiment analysis based on geography network. Then, we combine them together. The ECM starts from a certain radius and forms a ring. Within the ring, we perform visual analysis based on geography network.

3.1 Sentiment analysis and electron cloud model

3.1.1 Sentiment analysis

Each phase was manually annotated with polarity tag: positive, negative and neutral. So a micro-blog post expresses a sentiment. Different sentiments of micro-blog posts can reflect users' attitudes to a given topic. We define a parameter Sentiment Tendency (ST) to measure the sentiment expressed by a micro-blog post.

Firstly, we divide all micro-blog posts into keywords[13] and record the frequency of each individual keyword. We then build an index dictionary for the keywords and their frequencies. So a given micro-blog (MB) is divided into many keywords as shown below. In order to obtain the total score for an entire micro-blog post, we need to sum up the sentiment score of each keyword. Here we use the Wordnet and Hownet sentiment analyzing dictionaries[26], which includes about 5000 positive and 5000 negative keywords. Each keyword has a sentimental score. Positive sentiments are scored within [0, 1] and negative ones within [−1, 0]. For example, we can set "delight" to 0.9, "care" 0.8, "sad" −0.6 and "tragedy" −0.8. Then we calculate the positive sentiment tendency of the micro-blog by using $ST = \Sigma S_i/n$, where n is the number of keywords and S_i is the score of keyword i. If a keyword cannot be found in HowNet dictionary, we set its score to 0.

3.1.2 Electron cloud model

Based on the Schródinger equation, the probability function is deduced. It basically describes a cloud-like region where the electron is likely to be found. The model based on this probability equation can best be described as the electron cloud model (ECM). Here, ECM is a 2D model.

Take hydrogen for example. Given a set of quantum numbers (n, l, m), we could obtain radial distribution, angular distribution, electron density distribution and probability surface in three-dimensional space at this state[12]. So electron cloud would change according to the quantum numbers. Changes in cloud generally include radical changes and angular changes. If we take every dot as an electron, change of electron cloud means that every electron transits from the old state to a new state and rotates at the same time. So our ECM is a dynamic and rotating model with a mass of nodes.

Based on previous researches[14-16], we know that sentiment changes with time. So the change in sentiment is a dynamic process similar to electron transition in electron cloud. Considering the dynamic property of electron cloud, we could take advantage of the ECM to visualize changes in sentiment. That is to say we could utilize a dynamic model to simulate a dynamic process. Following is our mapping rules.

According to Niels Bohr atomic theory[17], an atomic electron has such a property that electrons gain potential energy and become less stable as it moves farther away from the nucleus. The process in which an electron moves farther away from nucleus or closer to the nucleus is named transition. Trying to apply ECM to visualize changes in sentiment, we can map electron's stability to polarity (positive, negative and neutral) of sentiment. The more stable the electron is, more negative the sentiment is. In other words, the sentiment will be more positive if the electron goes further from the nucleus.

Generally, we took neutral as the boundary between positive and negative. However neutral is very important for sentiment analysis. We can't just ignore it. As we know, not every sentence has a sentiment. How would you treat the sentence "The weather is cold"? It certainly does not show whether the speaker likes it or not. Sometimes, comments are highly related to events but without any sentiment. The neutral should not be considered as a state between positive and negative but as a separate part that denotes the lack of sentiment. In Niels Bohr atomic theory, an electron has a ground state. In this state, electrons are in their lowest energy orbits. Not like the Bohr Model, we place the neutral between positive and negative, but not in the innermost orbit. In practice, we expand the neutral so that it could be clearly recognized.

Figure 1 is the sketch map of the ECM. It has three rings: the blue inner ring with an inner radius of r0 and an outer radius r1, the green ring with a width of (r2-r1) close to the blue ring and the outer red ring with an outer radius of r3 close to the green ring. From inner to outer direction, these rings separately represent negative, neutral and positive. The sketch map also has two electron nodes: e1 and e2 and two lines. The Node size indicates the sentiment value. If a node is positive, the bigger the node is, the more positive the sentiment is. Electrons rotate in a clockwise direction and their trajectories form the lines in the map. If an electron moves from external to inner, we utilize a blue (Generally, blue represents negative) line to visualize its trajectory; otherwise we use a red line (Opposite to blue, red represents positive). Time intervals $\Delta t1$ and $\Delta t2$ during which electron e1 transits from A to B, e2 transits from C to D represent rapidity of sentiment changes. The shorter time interval is, the more frequently sentiment changes.

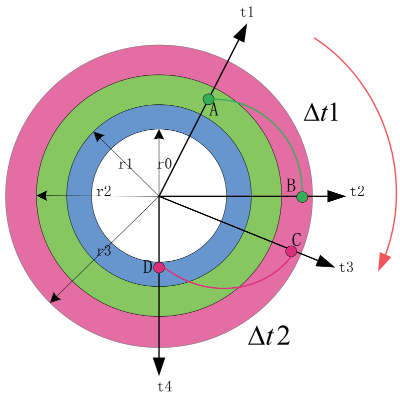


Figure 1. Electron cloud model.

We treat each electron node as a micro-blog user. Through sentiment analysis, we will get a ST score for each micro-blog post. Let the upper limit score of all micro-blog sentiment tendencies be MaxST, the lower limit be MinST, the neutral value ranges from NeutST1 to NeutST2 (NeutST $1 < NeutST2$) and the score for a micro-blog sentiment tendency be the variable st. We linearly map user's sentiment tendency to radial distance. And revised mapping rule is as follows:

$$r = \begin{cases} r_0 + r_1 - \dfrac{r_1 - r_0}{NeutST1 - MinST}(NeutST1 - st) \; (st < NeutST) \\ r_1 + r_2 + \dfrac{r_2 - r_1}{NeutST2 - NeutST1}(st - NeutST1) \; (st = NeutST) \\ r_2 + r_3 + \dfrac{r_3 - r_2}{MaxST - NeutST2}(st - NeutST2) \; (st > NeutST) \end{cases} \tag{1}$$

So st between the blue circle and black circle indicates a negative sentiment. When between the orange circle and the black circle, st represents a positive sentiment. The sentiment will be more positive if the radical distance from st to the nucleus becomes longer.

As user's sentiment changes with the course of time, we linearly map radius angle to time interval between two micro-blog posts. The angle of circumference could represent an appropriate time span like a day or a week, etc. In this paper, we treat it as a day.

3.2 Geography-based visual analysis

Using a map of China as the background can directly show the geographical layouts of network nodes and areal differentiation of sentiment analysis. And we mainly focused on the sentiment trends and the sentiment difference between different cities. In such a case, person-based analysis will be neglected. The map details may clutter the layout of the connections. We therefore use a simplified map with every city represented as one node, whose position is fixed and corresponding to the geographic coordinates of the city. In this section the method of constructing weights of edges and nodes from user data is first introduced. Then we focus on dissemination of information among many Chinese cities and public sentiment.

3.2.1 Constructing geography network

The user data we collected include retweeting connections, users' messages or comments and users' profile information, including gender, address, and membership level. Micro-blog retweeting network is directed and the direction between a pair of nodes is decided by the retweeting connection. We match users with the corresponding coordinates of each city. In the geography network, M and N represent two city nodes. The weight of edge $\overrightarrow{MN}$ denoted as Eweight $(\overrightarrow{MN})$ is defined as how many times the users in city N retweet the micro-blog posts of users in city M in Fig. 2. There are some users of city M retweet the same post from one user in N. And one user in M may retweet posts from different users in N. Anyone in city N retweets a micro-blog post of one user from city M once, the weight of edge $\overrightarrow{MN}$ will be added one. The weight of node M denoted as Vweight (M) is defined as the average sentiment tendency of micro-blog posts (discussed in Sec 3) posted by users in M.

Figure 2. Basic representation of connection.

3.2.2 Bundling of retweeting connections among cities

Considering the features of micro-blog retweeting network based on geography, our method incorporates edge directions, edge weights and interactions based on force-directed edge bundling.

Force-directed edge bundling is a physical mode where each edge is formed by a set of control points. As illustrated in Fig. 3, each control point interacts with adjacent ones via Hooke's law of spring. The spring force denotes as F_s between two adjacent control points p_i and p_j on edge P is computed as

$$F_s(p_i, p_j) = \frac{k_s C}{|p_i - p_j|} \tag{2}$$

where k_s is a global spring constant and C is the number of control points. Besides each control point interacts with control points on other edges via a Coulombic force, which is computed as

$$F_c(p_i, q_j) = \frac{k_c}{|p_i - q_j|^2} \tag{3}$$

where k_c is a global Coulombic constant.

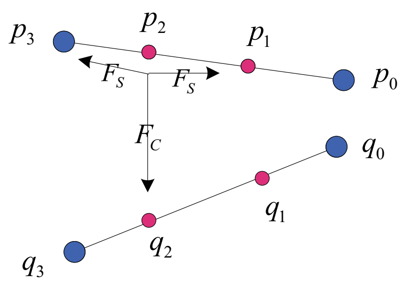


Figure 3. Force-directed edge bundling.

3.2.2.1 Direction compatibility

Prior work[6] introduced four compatibility measures to prevent over-bundling and reduce the computation complexity. Considering the intersection angle, distance, projection overlap and lengths of two edges P and Q, the force between their control points is multiplied by the product of these measures, $C_e(P,Q) \in [0,1]$ to reduce it. But the angle compatibility doesn't involve the direction of edges, which is an important feature about the retweeting connections among cities in the geography network. To separate edges in opposite directions, edges that travel in antiparallel directions are not bundled tighter than those in parallel directions. We now introduce the Direction compatibility $C_d(P,Q) \in [0,1]$ to replace the angle compatibility.

Each edge is considered to be a vector from a source node to a target node. Two edges P and Q go in the same direction if their dot product is positive and their direction compatibility is equal to the angle compatibility computed as $|\cos(\alpha)|$ where α is the intersection angle between P and Q. However, when P and Q are going opposite directions if their dot product is negative, the direction compatibility of them is computed as

$$C_d = k \cdot |\cos(\alpha)| \tag{4}$$

where k is a parameter and $0 \leqslant k \leqslant 1$. If k is smaller, edges travel in antiparallel directions to keep a longer distance.

Color gradient is an effective way for edge direction encoding, so we use a blue (source node) to red (target node) gradient to represent direction in our visualization. Thus we can know the direction of an edge by inspecting the colors of its source node and target node. It is also relatively easy to see if a node is predominantly source or target by comparing the average color.

3.2.2.2 Edge weights

Considering the retweeting connections among cities, an extension incorporating weighted edges is needed. If the retweeting connections between users of two cities are strong, the corresponding edge will exert more influence on the final layout of edges.

The weights of all edges are normalized to [0, 1]. Then the external Coulombic force between two control points of two edges is multiplied by the normalized weight.

More important edges exert more influence over the bundled graph structure. As shown in Fig. 4, the thicker edge R has a large weight while the edge Q has a small weight. Edge R will attract the edge P more. And finally edge R move a short distance to P because one control point p_j of the edge P exerts an attractive force to one control point r_i of the edge P is computed as

$$F_c(r_i, q_j) = \frac{k_c E_{weight}(Q)}{|r_i - q_j|}. \tag{5}$$

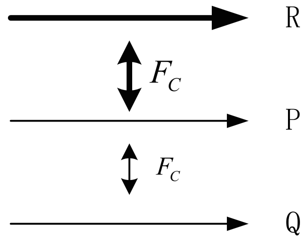


Figure 4. Asymmetric attracting force influenced by the edge weights.

To visualize important edges clearly, the alpha value and thickness of an edge are both in proportion to its weight. However, large edge thickness will make clutters for the endpoints of important edges. So the midpoint of the edge has a maximal thickness. Closer to endpoint, smaller thickness is.

3.2.2.3 Interactive bundling

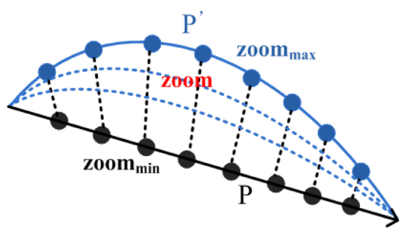


Figure 5. Edge bundling adjusted by zoom operations.

When the whole graph structure is showed, edge bundling is useful to reveal high-level edge patterns. However, if we want to explore local details by zoom and pan operations, edge bundling will hinder us finding many original edges, so we adjust the degree of edge bundling according to the current zoom operations. In the prior work, an iterative refinement scheme is used to calculate the bundling to improve performance. We can record every control point's position in each iteration step. According to the position of the viewpoint window denoted as zoom and its maximal and minimal positions denoted as zoommax and zoommin, the control points' positions in one iteration step can be selected to visual the layout of edges.

But the scheme above performs 141 iteration steps in total and ends with edges that contain 32 subdivision points in the prior work. The storage space recording every control point's positions of all iteration steps is unworthy. So we only record every control point's original position p_i and final position p_i'. A simple linear interpolation can be used to calculate the current position of the control point as

$$p_i(zoom) = p_i - \frac{p_i - p_i'}{zoom_{\min} - zoom_{\max}} \times (zoom_{\min} - zoom) \tag{6}$$

3.2.3 Layout of public sentiment

Our approach to public sentiment layout is inspired by kernel density estimation (KDE). It has the advantage of very naturally visualizing the sentiment changes in different areas. In the following we introduce how to incorporate their method with the sentiment value of micro-blogs. The 2D kernel density estimator for n data points $x_1 \ldots x_n$ is defined as

$$f_{k_H}(x) = \frac{1}{n}\sum_{i=1}^{n} k_H(x - x_i) \tag{7}$$

with kernel function K_H. Each point contributes a weight of one at its position and a decreasing influence to its neighborhood. For simplicity we choose a Triangle Kernel for K with influence matrix H, which is convenient for us to do numerical analysis. A multi-level parameter H is used to control the bandwidth of the kernel function. We define a parameter Sentiment Influence (SI) for every city. If the micro-blogs of users in one city P are retweeted more (Outdegree (P)) and the average sentiment tendency is stronger (Vweight (P)), the sentiment influence of P is larger as

$$SI(P) = out\deg ree(P) \times V_{weight}(P). \tag{8}$$

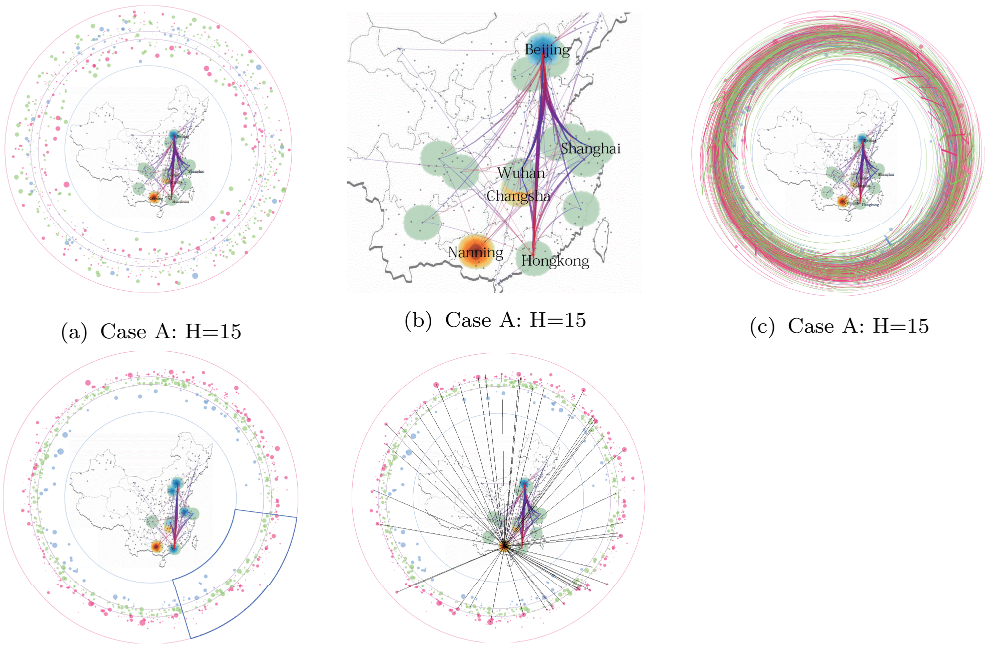
For every city, its sentiment influence will be used in the sentiment density estimator in formula (8). In our visualization, blue and yellow respectively represent positive and negative sentiment density. If the absolute value of sentiment density is larger, the color is darker.

4 Case Studies and Analysis

As we discussed above, the Electron Cloud Model is used to track the micro-blog sentiment changing trend over time while the geography-based network is used to visualize the retweeting relationships between cities. If they are showed in the coordinated views, it is difficult for users to form a mental map and analyze the time-space varying characteristics for the micro-blog sentiments. Therefore we combine them together that the visual analysis based on geography network is placed in the center of the Electron Cloud Model liking the electronic core as shown in Fig. 6(a). In addition, some interactions between the time varying analysis and the space varying analysis can be designed to help users to filter and select some interesting micro-blogs.

A sector-ring liked time window is designed in the ECM to this time-space varying analysis as shown in Fig. 6(d). Users can drag the time window along the ring and set its length. Then the geography-based network will be updated according to the micro-blog retweeting relationships in the current period decided by the time window. When users click a city, then the micro-blogs related to the city will be linked to this city as shown in Fig. 6(e). Through these links, we can better understand the distribution of micro-blogs related to the city over time and know their detailed sentiment tendency. Some alternative interactions could also be designed and added easily. For example, the number of the time windows may be more than one. When a city is clicked, we can highlight the micro-blogs related to the city instead of linking them to the city.

Sina micro-blog which has the largest micro-blog community in China is our experimental data source. We crawled data via Sina micro-blog API by using key words and then interpreted each micro-blog into a sentiment value. Micro-blog posts from about 20,000 users on popular social issues were crawled for our empirical study, which are Contributing Money by Retweeting, Deng's Attempted Assassination of a Government Official, Li Zhuang Case. They are denoted by Cases A, B and C respectively. These cases are all hot topics and caused extensive discussions and social concern. Briefly, Case A is about donating money on micro-blog to help a child who was badly scalded. A real-estate manager promised that anyone retweeted once about this topic, he would donate RMB 1. Case B is on a waitress named Miss Deng Yujiao who stabbed a government official on self-defense after being harassed by the official. The local police took arrested Miss Deng on suspicion of "intentional homicide". Case C is the story of a lawyer committing perjury to harbor criminals. Mr. Li Zhuang, a local lawyer, had defended a dozen of criminal suspects by making false statements and finally the criminal suspects were all acquitted.



(a) Case A: H=15 (b) Case A: H=15 (c) Case A: H=15

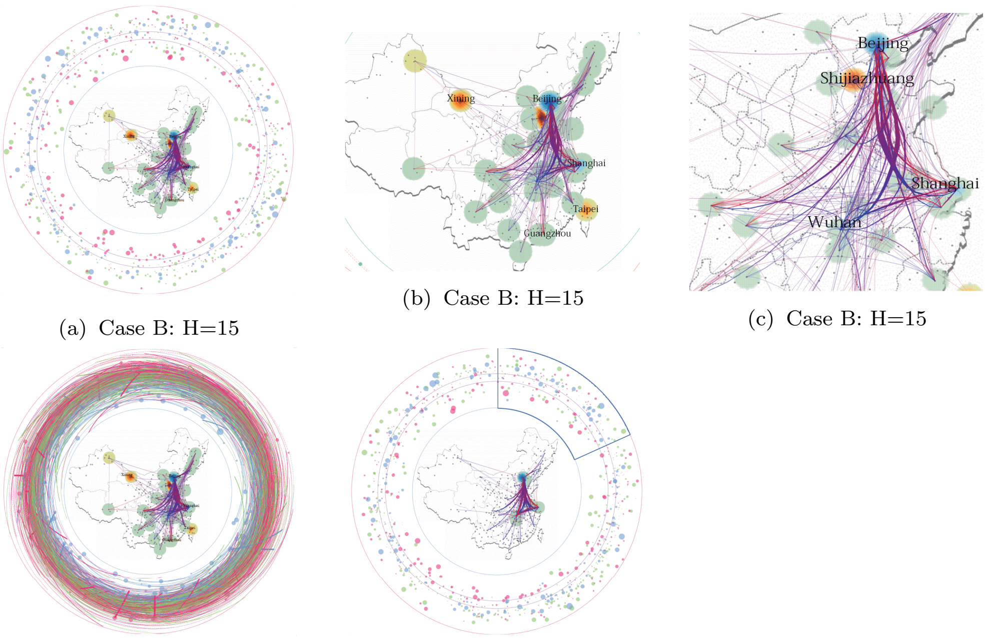
(d) Case A: H=15 (e) Case A: H=15

Figure 6. Visualization of Case A.

Here, we visualize micro-blog sentiment changes, tendency, the distribution of sentiment with the regions and retweeting connections among cities.

4.1 Visual analysis of Case A

The results of Case A show that the case has several active cities such as Beijing, Hongkong and Shanghai illustrated in Fig. 6(a), Fig. 6(b) and Fig. 6(c). Beijing is the major source place of original messages. Most users in other cities retweeted messages posted by users in Beijing. A large number of users in Hongkong retweeted the micro-blog posts Beijing and Wuhan. Users in Nanning impact great influence on microbolg users. Users' sentiments were main positive and negative under the influence of Case A. Only few of them were neutral. A lot of users' sentiment transited from positive to negative. Meanwhile, many transited from negative to positive. A lot of positive users' sentiment transited to neutral. But few transited from negative to neutral. That some strong negative users' sentiment transited to strong positive showed that these people are easily express strong sentiment than others. Users' sentiment mainly located near neutral band. That is to say, most of the users did not express strong sentiment. Figure 6(b) is the detail of geographical analysis. Figure 6(c) shows many trajectories of sentiment transition. The main transition trend is from the positive direction to negative direction. There still existed some users whose sentiment transited from negative to positive sharply. Besides, the low density of the lines indicates that the case does not cause much social concern. Many users think it as a commercial speculation and are not willing to pay attention to it. Figure 6(d) shows the visual analysis of data within a time period. We could conclude that most cities like Beijing, Guangzhou, Shanghai, etc. showed strong negative influence and only Nanning show strong positive sentiment. Figure 6(e) also shows the interaction that we could select a city and explore sentiments distribution of users in right this city. We could see that sentiments of Nanning users were mainly related to positive. Only few were related to neutral and negative.



(a) Case B: H=15

(b) Case B: H=15

(c) Case B: H=15

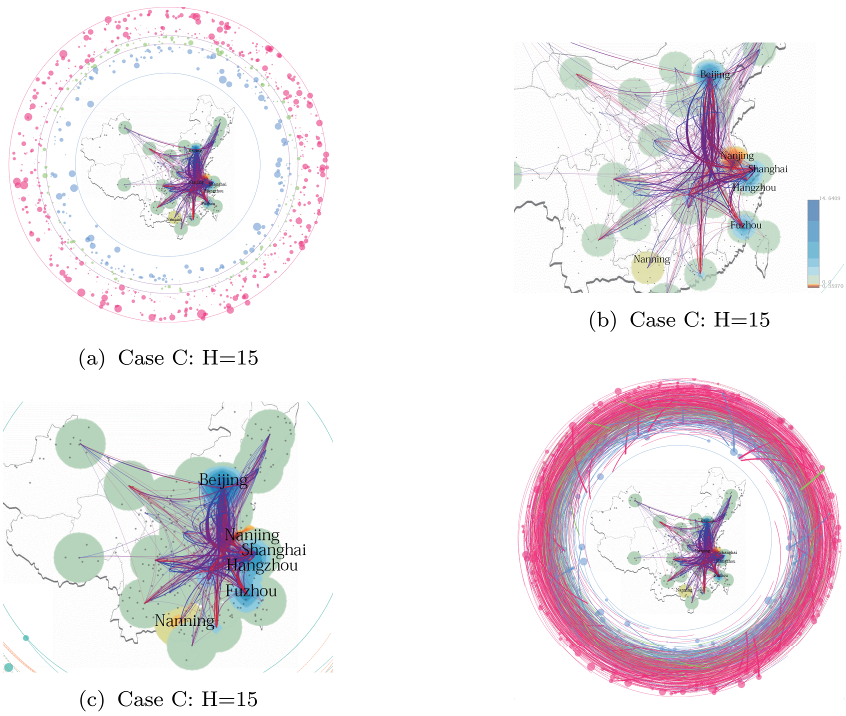
(d) Case B: H=15

(e) Case B: H=15

Figure 7. Visualization of Case B.

4.2 Visual analysis of Case B

According to Fig. 7(a) and Fig. 7(c), we can see that for Case B retweeting connections among Beijing, Shanghai and Changsha are very strong. Xining, Beijing, Fuzhou, Taipei, etc. had great influence on other cities. Users' sentiment mainly transited from positive to negative. A few positive transited to neutral. And the users in Kunming expressed negative sentiments clearly. The outer ring shows that users' sentiments mainly transited from positive to negative under the influence of Case B and finally became negative. Because an event similar to Case B also happened in Kunming, local users expressed more sympathy for Miss Deng and condemned the official more strongly than other cities. The negative sentiment shown in the schematic view of sentiment influence is stronger than Case A. Users' sentiments distributed more separately and did not shrink near neutral band compare with Case A. Users expressed stronger sentiments and there were more users who expressed negative sentiments than Case A. Figure 7(e) is the visual analysis of data within a time frame. At this time, active cities were Beijing, Shanghai and Changsha. Only Beijing had great influence. A few sentiments transited from positive to negative or from negative to positive sharply. And there were no neutral sentiments which transited sharply.



(a) Case C: H=15

(b) Case C: H=15

(c) Case C: H=15

(d) Case C: H=15

Figure 8. Visualization of Case C.

4.3 Visual analysis of topic C

Figure 8(a) is the final state of sentiment transition. We could clearly see the distribution of positive, negative and neutral sentiments. The range of sentiment change was very large and sentiment reached maximum and minimum under the influence of Case C. Positive and negative are the main sentiments. A lot of people showed strong positive sentiment. And there were more positive users. Only a few users showed strong negative sentiment. The inner rings of the results of Case C indicate that the case had much more retweeting connections than previous two cases. This case has the largest number of active cities in the three cases. Active cities like Beijing, Shanghai, Nanjing, etc. had really great influence and much more retweeting connections. There were much more trajectories than previous two cases in Fig. 8(d). Some sentiments transited from positive to negative, from negative to positive and from positive to neutral sharply.

This case caused much controversy and was debated over the Chinese legal system and also attracted academic attention. The schematic view of sentiment influence in the lower right corner shows that public sentiments to this event are very strong.

Geographical layout of public sentiment is discussed in Sec 4.3. The multi-level parameter H can be adjusted by keyboard to reveal details of different levels. As illustrated in Fig. 6(b) or Fig. 7(b), the layout of public sentiment is shown in a low level when H is set 15. We just see the overall trend of distribution of public sentiment. If the value of H is halved to 7, some difference appears.

By zoom and pan operations we can see local details. In Fig. 8(b), the public sentiment in Nanjing is negative while the public sentiments in the cities near to Nanjing such as Shanghai, Hangzhou are positive. When H is relatively large and sentiment influences for cities are wide, the negative sentiment in Nanjing will be offset by the stronger positive sentiments in the cities close to Nanjing. Comparing Fig. 8(b) and(c), the distribution edges in Fig. 8(b) is more even and clear than Fig. 8(c)'s. According to the discussion in Sec 4.2, we reduce the degree of edge bundling when the view window is zoomed into.

4.4 Comparison of different methods

Previous research on micro-blog data concluded some major methods, such as statistics, timeline, stream river, geography based method and combination of these methods. Our method is a combination of geography based method and a timeline-like method. It could visualize data changes over time, value distribution, geography-based propagation patterns (summary layout) and the influence of each city.

Table 1 Comparison of similar methods

	Value	Social relationship	Geographic information	Intuition	Comprehensive analysis
Statistics[25,26,28,29]	+++++			++	+
Timeline[19,28]	++++			+++	++
Map[29,30]	+++	+++	+++++	++++	+++
Our Method	++++	+++	+++++	++++	+++++

Description: We concluded some indexes and evaluate them with a range from null to five +. It provides better analysis with more +. Value: whether this method supports value operation.

Compared to statistical method, our method can provide an intuitive result and mine the data deeply. Our method could provide a timeline but also their distribution according to their value. Steam river is in fact a timeline method. It could visualize a little more information and has a very beautiful visual representation. But its analyzing ability is still limited. Geographical method can show the pattern. Sometimes it has to combine with a timeline. While our method provides an intuitive model and combines them together. Summary of comparisons is listed in table 1.

5. Conclusion and Further Work

This paper presents a time-varying visual analysis of micro-blog sentiment based on electron cloud model and a space-varying visual analysis of micro-blog sentiment using kernel density estimation and edge bundling.

We conducted qualitative comparison with other methods. In the future, we hope to conduct much more detailed comparison or quantitative comparison.

Acknowledgement

This paper was supported in part by Natural Science Foundation of China under Grant no. 61532002 and no.61272199, National High-tech R&D Program of China (863 Program) under Grant no. 2015AA016404, the Specialized Research Fund for the Doctoral Program of Higher Education under Grant 20130076110008, and the open funding project of State Key Laboratory of Virtual Reality Technology and Systems, Beihang University under Grant No. BUAA-VR-15KF-14.

References

[1] Nagy A, Valley CMS, Stamberger J. Crowd Sentiment Detection during Disasters and Crises. Proc. of the 9th International ISCRAM Conference. 2012. 1–9.

[2] Nielsen FÅ. A new ANEW: Evaluation of a word list for sentiment analysis in microblogs. arXiv preprint arXiv. 2011. 1103–2903.

[3] Dasgupta S, Ng V. Topic-wise, sentiment-wise, or otherwise?: Identifying the hidden dimension for unsupervised text classification. Proc. of the 2009 Conference on Empirical Methods in Natural Language Processing. 2009. 580–589.

[4] Park J, Kim H, Cha M, Jeong J. Sentiment analysis on bad news spreading. Proc. of the ACM CSCW Workshop on Design, Influence, and Social Technologies (DIST). 2012.

[5] Chew C, Eysenbach G. 2010. Pandemics in the age of twitter-content analysis of tweets during the 2009 H1N1 outbreak. PLoS ONE, November 2010, 5(11): 1–13.

[6] Holten D, Van Wijk JJ. Force-Directed Edge Bundling for Graph Visualization. Computer Graphics Forum. Blackwell Publishing Ltd. 2009, 28(3): 983–990.

[7] Cui W, Zhou H, Qu H, Wong PC, Li X. Geometry-based edge clustering for graph visualization. IEEE Transactions on Visualization &Computer Graphics, 2008, 14(6): 1277–1284.

[8] Gansner ER, Hu Y, North S, Scheidegger C. Multilevel agglomerative edge bundling for visualizing large graphs. Pacific Visualization Symposium (PacificVis), 2011 IEEE. 2011. 187–194.

[9] Hurter C, Ersoy O, Telea A. Graph bundling by kernel density estimation. Computer Graphics Forum. Blackwell Publishing Ltd.2012, June, 31(3): 865–874.

[10] Cao N, Lin YR, Sun X, Lazer D, Liu S, Qu H. Whisper: Tracing the spatiotemporal process of information diffusion in real time. IEEE Transactions on Visualization and Computer Graphics. IEEE. 2012, 18(12): 2649–2658.

[11] Gloor PA, Krauss J, Nann S, Fischbach K, Schoder D. Web science 2.0: Identifying trends through semantic social network analysis. International Conference on Computational Science and Engineering, 2009. CSE'09. IEEE. 2009, l(4): 215–222.

[12] Diakopoulos N, Naaman M, Kivran-Swaine F. Diamonds in the rough: Social media visual analytics for journalistic inquiry. 2010 IEEE Symposium on Visual Analytics Science and Technology (VAST). IEEE. 2010. 115–122.

[13] Xiao D, Ruoyu S, Qitu H, Xiaoling Z. A computer simulation method on the electronic probability distribution of hydrogen atom. 2011 7th International Conference on Digital Content, Multimedia Technology and its Applications (IDCTA). IEEE. 2011. 1–4.

[14] Marcus A, Bernstein MS, Badar O, Karger DR, Madden S, Miller RC. Twitinfo: Aggregating and visualizing microblogs for event exploration. Proc. of the 2011 Annual Conference on Human Factors in Computing Systems. 2011. 227–236.

[15] Bollen J, Mao H, Pepe A. Determining the public mood state by analysis of microblogging posts. Proc. of the Alife XII Conf. MIT Press. 2010.

[16] Choi D, Kim P. Sentiment analysis for tracking breaking events: A case study on twitter. Intelligent Information and Database Systems. Springer Berlin Heidelberg. 2013. 285–294.

[17] Turner. User-directed sentiment analysis: visualizing the affective content of documents. Proc. of the Workshop on Sentiment and Subjectivity in Text (SST '06). Association for Computational Linguistics. Stroudsburg, PA, USA. 2006. 23–30.

[18] Bohr N. LIX. On the effect of electric and magnetic fields on spectral lines. The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science, 1914, 27(159): 506–524.

[19] Wang CB, Xiao Z, Liu YH, Xu YR. SentiView: Sentiment analysis and visualization for internet popular topics. IEEE Transactions on Human-Machine Systems, 2013, 43(6).

[20] Kppel M, Schler J. The importance of neutral examples for learning sentimen$t$.Computational Intelligence, 2006, 22(2): 100–109.

[21] Ding X, Liu B, Yu PS. A holistic lexicon-based approach to opinion mining. Proc. of the 2008 International Conference on Web Search and Data Mining(WSDM '08). ACM. New York, NY, USA. 2008. 231–240.

[22] Taboada M, Brooke J, Tofiloski M, Voll K. Lexicon-based methods for sentiment analysis. Comput. Linguist. 2011, 37(2): 267-307.

[23] Wilson T, Wiebe J, Hoffmann P. Recognizing contextual polarity in phrase-level sentiment analysis. Proc. of the Conference on Human Language Technology and Empirical Methods in Natural Language Processing (HLT '05). 2005.

[24] Pang B, Lee L. A sentimental education: Sentiment analysis using subjectivity summarization based on minimum cuts. Proc. of the 42nd Annual Meeting on Association for Computational Linguistics (ACL '04). 2004.

[25] Krishnamurthy B, Gill P, Arlitt M. A few chirps about twitter. Proc. of the First Workshop on Online Social Networks. ACM. 2008. 19–24.

[26] Huberman BA, Romero DM, Wu F. Social networks that matter: Twitter under the microscope. Available at SSRN 1313405. 2008.

[27] Honey C, Herring SC. Beyond microblogging: Conversation and collaboration via Twitter. 42nd Hawaii International Conference on System Sciences(HICSS'09). IEEE. 2009. 1–10.

[28] Hughes AL, Palen L. Twitter adoption and use in mass convergence and emergency events. International Journal of Emergency Management, 2009, 6(3): 248–260.

[29] Fang Y, Chang MW. Entity Linking on Microblogs with spatial and temporal signals. Transactions of the Association of Computational. 2014.

[30] Han J, Kong LY. Interrelation between micro-blog hotspots and Urban spatial. Micro-Blog and Spatial Network 49th ISOCARP Congress. 2013.