

The Visualization of Mass Information in Social Network with a Holistic View

Xin Li

Department of Computer Science and Technology, Tsinghua University, Beijing, P. R. China
lixinsemail@gmail.com

Xia Zhang

School of Journalism and Communication, Tsinghua University, Beijing, P. R. China
zhangxia.tsinghua@gmail.com

Peng Cui

Department of Computer Science and Technology, Tsinghua University, Beijing, P. R. China
cuijp@mail.tsinghua.edu.cn

Zhiyong Fu

Department of Information Art and Design, Tsinghua University, Beijing, P. R. China
fuzhiyong@tsinghua.edu.cn

Shiqiang Yang

Department of Computer Science and Technology, Tsinghua University, Beijing, P. R. China
yangshq@mail.tsinghua.edu.cn

Baoguo Cui

School of Journalism and Communication, Tsinghua University, Beijing, P. R. China
bqcui@mail.tsinghua.edu.cn

In this paper, we propose visualization for mass information in the social network with a holistic view, which focused on representing the effect of user communication. With the Rapid Expansion of social network, it is necessary to provide a holistic view to the users and researchers to understand the effect of information communicated in the network. The social network users will be visualized in 3-dimension space, with 2-dimension graph layouts delivering various characteristic of the network. The layout is mainly based on user's role in the network, but not the relationship of users in the traditional way. By visualizing the disseminating and stop of information flow among users, we can distinguish users by their effect and contribution in communication among the network. The visualization combines technology, content and expression. We can then study the social network by the performance of the network structure situation and information flow.

Information flow. Holistic view. Communication effect. 3-Dimension visualization. Social network.

1. INTRODUCTION

The social network grows fast globally, with tons of users connecting together and massive information flowing around. Graph visualization is an effective method to understand complicated information involved (Jia *et al.* 2009). The current situation is that the traditional network visualization runs on an excessive way, because of the number of users and complicated connections. The traditional graphic drawing is based-on point-line model to present users and relations. It is appropriate for small scale visualization, like one user perspective (Heer & Boyd 2005) or categories (Kim *et al.* 2009). When the large data, e.g. web scale data, are input to the system, the graph will be full of points and lines. We can access to only little knowledge from the overloading graphic, since the information is too much to be recognized. It seems that there is a loser stuck in a maze, and we need to provide an overall map, but not a list of every detail about which two corners are connected.

We lay out a large number of users in a 3-dimensional space based-on their effects to

communication and the whole network with a holistic view. Every user can be seen as a system, and we are concerned with the input, output and the interface. Corresponding to the social network, these three factors are the numbers of visited blogs, effectively shared blogs and friends of a user. By this model, we can place all the users onto a 3-dimensions space. The relationship appears whenever they take an effective action, i.e. a user visits a blog which their friends share or publish. Through these relationships, the blog-visiting actions converge into information flow which performs as comets crossing sky.

The number of friends reflects the influence of a user in a social network. More friends provide higher possibility for information communication. The activity of users is shown by their visiting and sharing blogs as a user's information receiving and sending behaviour.

Using the different axis projection, we can obtain the user communication activity plot, user influence strength plot and receiving information plot. The users in social network are like stars in the sky, and

the plots disclose the user condition as star maps. The tail of a comet will show the visitors volume of a blog. We can understand how different kinds of blogs flow in the network and how users intercept the information, which should be the negative key points for dimension information (Skold 2008). The visualization fits for human perceiving process. We name this visualization system as Pitaya.

2. BACKGROUND

2.1 Related Work

To study and present the social behaviour in a network by visualization, it can be traced back to the study based-on email communication. Drexel visualizes the selected email, topic and the relationship between users by different functions in a small scale network (Jia *et al.* 2008). With the rise of social network sites, the visualization for cyber social behaviour research is mainly focused on SNS (Hao *et al.* 2010). The social network visualization analysis tool is used for analysing blogs data and is concerned about the user relationship structure, in a thousand nodes weight network (Nan 2009). Visualization in a 3D space helps to show the complex data in social network. The 3D social network and links visualization takes the social network user number, child-network, child-child network as three axes (Ding 2007). The links present the document exchange with time standing for the time (Du 2009, Roam 2009). The colour, shape and axis play an important role in the interface presentation (Zhu *et al.* 2006). Vizster selects a single user as the core of visualization. The friends are placed by their relationships, which are shown as links, between each other and also to the core user. It works for the end-user exploration (Heer & Boyd 2005).

There is a situation of social network should be considered seriously, which is SNS with a very large scale. And the visualization of SNS also faces the large scale network beside a small community (Batagelj & Mrvar 1998). To make the visualization for large network clearly, clustering the users into small group is an effective method (Perer & Shneideman 2006). The clustering can be based on multiple factors, such as user relationships, activities and information dimension. The colour still plays a significant role in presenting the multiple data for assisting nodes placement.

2.2 Motivation

The traditional visualization of network structure consists of nodes (the users), and links (the connection between users). The node-link structure meets the reorganization of nets. When the amount of uses is below hundreds level, this network structure gives an excellent performance. However,

when we are dealing with dozens thousands of nodes, it is shown many nodes and links cannot be distinguished. The meaningful links becomes the noise of visualization and lose their function (Marlow 2006). The links among nodes always stand for the relationship between users. There will be longer link, if the strength of user connection is low. The close relationship is presented as short link. The longer link takes more performance space as well as visual attention. So the visual effect of links is against their meaning, which means the information performance does not match user attention.

3. SOCIAL NETWORK ANALYSIS

3.1 The necessity of general visualization

It has been widely accepted that social media is a kind of social property which contains a large number of user's information. The social media could potentially bring many opportunities, e.g. branding communication, advertise marketing. However, in practical, connection between each other does not mean communication. Even when people are communicating, there is likely useless information. Most of the people in the friends list are inactive for any effective information. To express ideas and advertisement, it is necessary to define the real useful communication within the social media.

Although there is huge amount of information in the social media, it is not completely random. In this paper, we use info graphics method, which take each person as a unit to study browsing and sharing information. The info graphics is visualization of complex data and through the visualized shape, demonstrate the relationship among clients and information exchange. Info graphics have drawn increasingly interests in worldwide. For example, digg, friendfeed, flickr all have their visualization program. However, it is still a relatively new area in China.

3.2 Mass Communication prospective

Pitaya separates the users and information. How does the social medium form, and how do users gather to form a small community. The traditional Symbolic Interactionism believes itself generated from interaction with others and primary group generated from the interaction. People are constantly adjusting themselves through the keeping variable environment, exchanging information and sharing symbol. This theory works well for the micro study of communication. However, it was found the theory is not perfect for the social media. This is because (a) in the internet, interpersonal communication is not just interaction and (b) interference information, e.g. viral

marketing A.D. could participate into the Mass Communication. On the other hand, the method used to study Mass Communication, such as quantitative analysis does not apply for social media, as it over stresses the “majority decision” principle. Therefore, Pitaya separates user’s scale and information to build a new model. It uses people as the node of information dissemination. Reproduce interpersonal relationship and information flow in a micro scale. Meanwhile, consider environmental effect for the information from the whole situation. The establishment of three dimensions could clearly show the relationship in the information flow as well as the influence and closeness among users.

3.2.1. Friend – the core of social media

There are some important theories in the social media, such as rule 150. As the low requirement of social media, one can easily add hundreds of friends in very short time even when this people are not connected in real life at all. In fact, the active friends are far less than 150. Many are inactive (shown as grey points in Pitaya). Most of interactions occurred within a relatively small scale and many are friends in real life. Based on this phenomenon, 1-9-90 consider :in most online communities, 90% of users are lurkers who never contribute, 9% of users contribute a little, and 1% of users account for almost all the action (Lampe & Johnston 2005).

User participation often more or less follows a 90-9-1 rule: 90% of users are lurkers (i.e., read or observe, but don’t contribute), as shown in Figure 1. 9% of users contribute from time to time, but other priorities dominate their time.

1% of users participates a lot and account for most contributions: it seems as if they don’t have lives because they often post just minutes after whatever event they just comment.

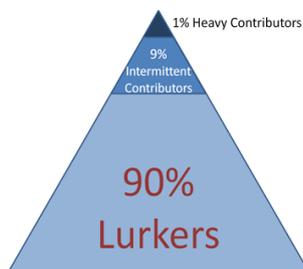


Figure 1: Lurkers map

3.2.2. Share — enhancing the social network

An active social network requires positive interactions. We suggest there is a small community belonging to the social network, which is the core of activation. According to rule 28, we could assume 80 % activation is generated by 20 % users. These users are defined as active core

of the social network and form the “Social Pandora Box”. This connection behaviour (you need to explain the connection is between what) means a very active sharing activity, including filtering the interest points, selecting and sharing topics. These users publish information, leading to feedbacks from their friends and therefore, enhance their connection and tighten their community. Though the interaction and sharing among users, we expect to observe how a new connection forms within the social network and how the users affect others.

3.2.3. Visit — expanding the social network

In contrast to the active users, there is a group of people who obtain information from social network but do not share information. They are in the boundary of the social network and weakly connected with their friends. To expand the social network, these users need to be encouraged to share information which can be achieved by a more efficient premium and penalty mechanism. Social network also obeys the Matthew effect, i.e. active users will become more active and inactive users will become more inactive (Lampe *et al.* 2006). For those users who only visit but not share information, no matter how weak connections they have built up, their contribution to the social network is limited (Watss 2002).

Pitaya reveals the whole social network from these three aspects mentioned above. Relationship itself is not valuable. It only becomes useful when information is flowing inside it. Hence, instead of studying relationship, we focus on which kind of users provide most contribution to the social network and accelerate the information flow. In the following section, we will discuss the meaning of visual interfaces which are constructed by different elements.

3.3 Observation

We place a large number of users into the *Friend Number – Visit Number – Share Number* three-dimension social space. The viewpoint is set to be outside of user location, by which we achieve the goal of holistic observation. To get more analysis value image, we need to place the perspective carefully. Borrowing front view, side view, bottom view from the mechanical drawing, we can watch the users along the three axes. We can get three kinds of images, and in each of them, there are two parameters playing the major roles.

The Friends-Visit, Visit-Share and Share-Friend images are more suitable for Orthogonal projection, and the third axis information is used for organizing the covering between users. If the view is not along the axes, we should select the nature visual effect. To render 3D objects into 2D screen, the

perspective projection is an intuitive method, which can show the three-dimensional characteristic of objects by cooperating with rotation. The red, green, blue and alpha of colour are always the assistant factors of visualization.

The social network is large and complex system. The users of social network are the input/output interface. The value of the system can be reflected by the social behaviours among the users. Our intervention is the information communication social behaviour, as well as the blogs information translating among the interfaces.

We define a visiting blog events happens in the system when user A visits a blog by sharing or publishing from user B. Publishing is a kind of sharing from users themselves, so we define that it is an effective sharing event of user B. The effective sharing event presents the output property of interface, as well as the visiting event stands for input property. The friend number of one user is the possibility to visit or effective sharing behaviour, as well as the capability of input and output. By different view to projecting the same space, we can get a serial of interconnected result images. The visit event number and effective share event number are the result of output and input. Two kinds of results plus the possibility, the three planes present the meaning as blow:

- Friend-Visit View: System Output Plane.
- Friend-Share View: System Input Plane.
- Visit-Share View: Output/Input Comparison

Based-on the users' distribution in the three maps, we can control the visit and effective sharing events effects for input and output capability and comparison in the images. On this basis, we can visualize the information flow, and then get the sensitivity of users to the selected information.

4. SYSTEM DESIGN

4.1 Visual Effect Design

70% of nerve endings and 40% cortex are related in human's brain. Visualization uses this ability to make people easily understand the trend, model and unusual information of data. Compared with text, visualized information is easier to be understood and accepted. Whole visualized figures allow normal users find disciplines and relation which is hard to be achieved with other method. Mobilized visualization could make users notice the variation of information. We studied using many ways to demonstrate data. There are three parameters we mainly considered about: transparency, colour, and size, as shown in Figure 2. The purpose of visualization:

- (1) Allows users response immediately after input data and will not get confused when two different kinds of information mixed together;
- (2) Displays static data (the data no longer increase) and dynamic data (data increase with time) at the same time;
- (3) Focus on useful data instead of mixing background with inactive grey users. (or: instead of background and inactive grey users).

We illustrate more information though visual experiences to describe the topic by three steps: 1. Found visualization element (transparency, colour, and size); 2. Filter extra information and construct a three dimension model (the contrast of background colour); 3. Insert topic (the plan for colouring different topic)

In the figure, each user is represented by a point. When there is no topic defined, 10% transparent grey points are distributed on the net and arranged according to number of friends, share and times of browsed. When inserting different topics, the points will be coloured differently. As the number of sharing and browse increases, the point will turn from 10% transparent to non-transparent, meanwhile, the size will be enlarged.

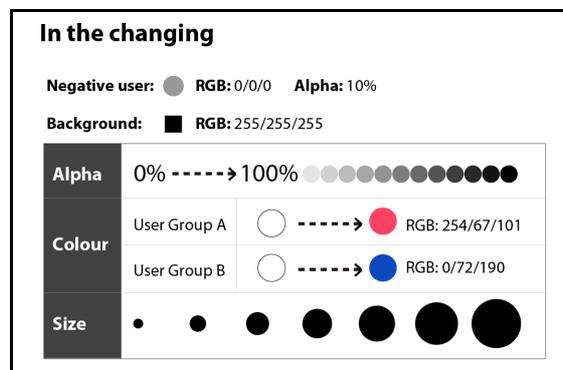


Figure 2: Elements

4.2 System Structure

This visualization system is built by the data including users and visiting events information, on the basis of social network holistic analysis. The three-dimension space is developed by DirectX SDK, which supports multiple rendering effect and interaction interface (Zhong *et al.* 2009, Fan *et al.* 2011).

4.2.1. Data

The data comes from the biggest social network sites, Renren, in China. It includes fifty thousand users' information and these user visiting blogs logs in four weeks. There are three parts of data:

Relationship: friends of each other.

Visiting log: who visits whose blog from whom shares and when
 Topic model: every blogs has been valued by different topics.

4.2.2. Network Structure

From every visiting log, we can get the visiting blogs number and effective sharing blogs number in this period. The social relationship supports the friend number of everyone. Based on these original data, we can place the users into the 3D space, as shown Figure 3. For a clearer performance, we colour the visit number as red and share number as blue. The colour information can help us distinguish the users' location in static picture. Because in the social network the user properties differ widely, the users gather around the point of origin and visit axis. The distribution situation means that a few of users having a large number of friend or strong input effect to the social network. The difference of visiting number is relatively small, which prove that the capability of social network outputting information to a single user is limited. This also indicates the information a user received is limited in a period.

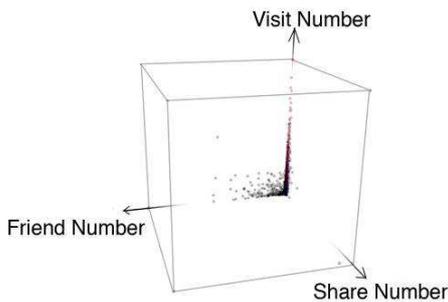


Figure 3: Original 3D network structure

The Figure 3 can be abstracted as a user distribution model as Figure 4 (a). It only takes a little space of the whole display area. We hope the distribution model is like Figure 4 (b) as possible, which uses the space much more effectively.

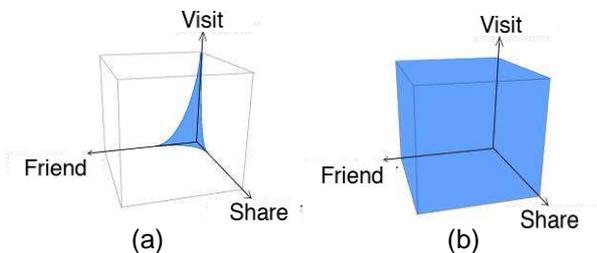


Figure 4: User Distribution Model

We try to scale the Figure 4(a) to the Figure 4 (b). Meanwhile, we also filter some edge and noise data. Because the visualization data is parts of the social network, but not the whole. There is some connection between the inside and outside of our data. We delete some users on the edge of the

social space. In the scaling process, the absolute position has been changed, but the relative location of the users remains. With five times of rooting transformation with different fixed points, we get the final holistic user distribution in 3D social space, as shown in Figure 5. The Figure 5 (a) is the perspective image, in which most of user nodes are colourful and take much display space, and (b), (c), (d) are orthogonal projection from three axes. The users are almost in the whole cube. The value of red colour corresponds to the share number, while the blue colour corresponds to the visit number. The three axes data has been normalized and limited in the cube. In the Figure 5 (c), users are mostly around the line 'Visit Number = Share Number', which means the ratio of input effect to output effect is nearly 1. In the Figure 5 (b) (d), The Friend number also corresponds to the visit and share with a shifting. The users are at the middle of Friend Number Axis, because the number of users, who have few friends, is small.

4.2.3. Information Flow

Lots of user visiting events makes the information flow. The information in social networks flows along the friend relationship cross users. The information flow can be reflected by two kinds of view: process view and effect view.

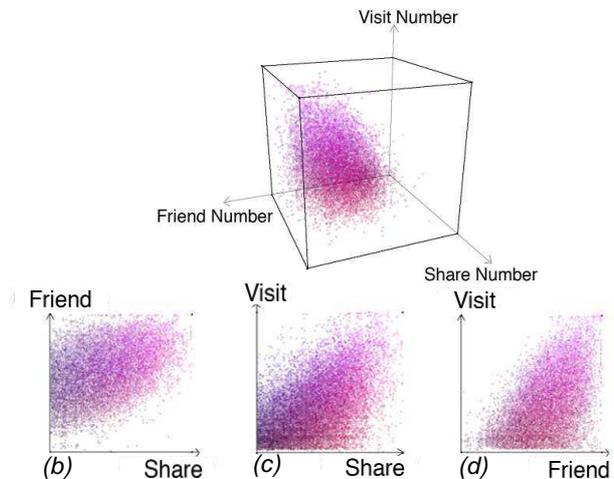


Figure 5: Users in 3D social space

From a process aspect, similar with traditional visualization model, the communication is to form a link between two users, i.e. connection in the figure. The benefit of this model is when there is less nodes, it has a clear structure. Basically, it uses connection between users as a parameter. However, when the number of nodes increases, the lines are over concentrated which makes the system too complicated. To deal with massive data, we find a method to keep a clear structure of users while effectively eliminate the overload of information. Traditional classification for users of social media converges to the closely connected users and separates the weakly linked user. As a

result, each user can only take a small visual space. Our method is opposite to the traditional way. In a three dimension space, we separate friends from each other. Thus, each small category represents a group of users around this coordinate. Meanwhile, in the individual network, the friends of users are distributed in the same network structure.

4.2.4. Clustering Method

We make every user in the space move step by step. In every step, the movement of one is determined by the distance to the users around it. As shown in Figure 6, we make three kinds of thresholds: Min, R1, R2. The x-axis is the distance of neighbouring users to the moving user. The y-axis is the additional movement of the moving user from the around user effect. The direction of movement is toward to the moving user. If the neighbouring user is a friend of the moving user, the additional movement is following the green line. If the around user is not a friend, the additional movement is following the blue polyline. In every step, the final movement of every user is the average of the additional movement. The Figure 7 shows the clustering result by the 5th, 10th, 15th, 20th step. After the 40th step, the system has been stabilized. The final result is not only the three dimension view of the whole network but also overlapping of individual community. Figure 8 shows the clustering of non-related users.

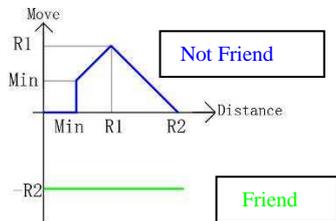


Figure 6: Distance-Move plot

When we only consider the result of visualized information flow, the pathway and procedure of information is less important. The information paths connect the groups of users, as shown in Figure 9. We draw a light line, when a visiting event happens. The dark area means that the information flow is dense. The key is the variation of starting point of information. Each information flow represents the level of interests of users. Using each information flow as a unit, one can find the preference of different users for different information (Hao *et al.* 2010). This allows us to investigate the sensitivity and activity of information in the social network from a whole perspective. Therefore, it provides a reference for developing social network applications and studying information communication.

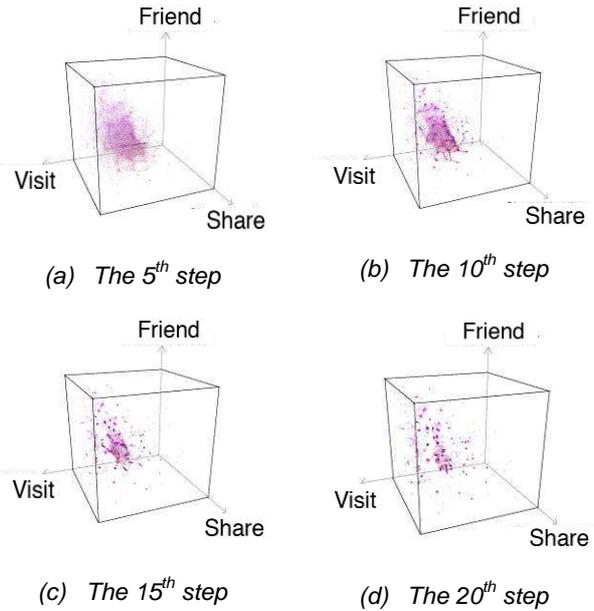


Figure 7: Clustering Processing

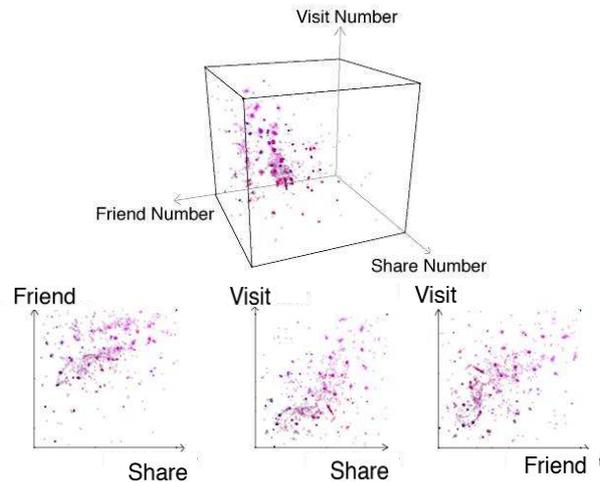


Figure 8: Clustering Result: The 40th step

4.2.5. Interaction & Interface

In this paper we use touch screen as interaction interface device. On the edge of the touch screen, users can switch visual angle. Using finger to drag in the centre, people can (1) rotate the three dimension space and observe from different angle; (2) zoom in or out by multi touch. Interaction interface combined with display interface enable a natural experience, shown as Figure 10. Users can focus on analysis of the visualization system, instead of paying much attention on operate the device. As the touch screen devices become more and more popular, users are used to this interaction process. Because this system is design for analysing large amount of data, this process needs to have multiple persons interacting at the same time. The touch screen is able to provide the result. Furthermore, participants could see the operating procedure and therefore provide an optimized interaction experience.

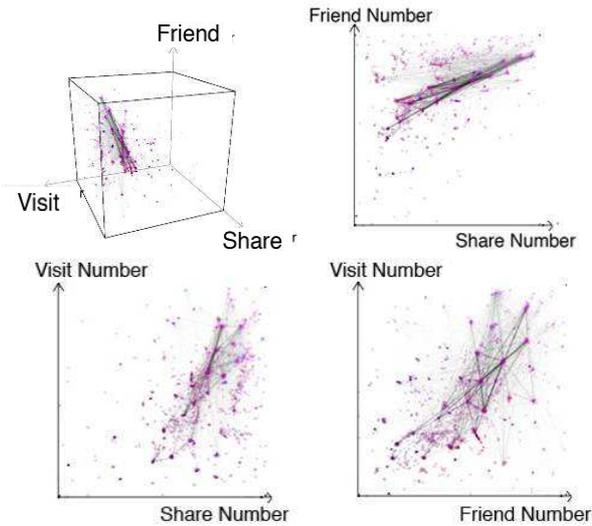


Figure 9: Information Flow

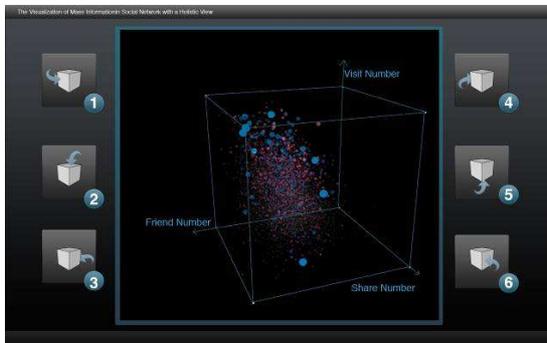


Figure 10: User Interface

5. VISUALIZATION RESULT

5.1 Description

5.1.1. Communication Analysis

Group A (Figure 11 (b), blue) is a report of traffic accident in China. Group B (Figure 11 (a), red) is workplace topic. The time scale is set to 4 weeks.

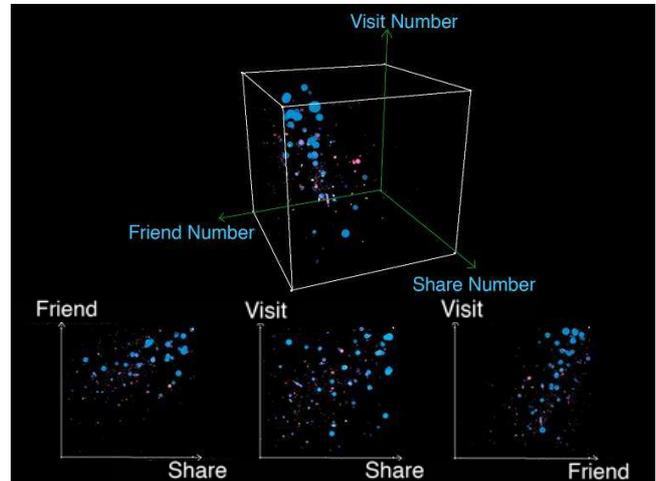
- (1) Before insert any data, 10% grey even sized points are distributed in the net.
- (2) After the topics have been inserted, the grey points start to change colour and become less transparent along with increasing the size.
- (3) When certain amount of data is inserted, the coloured points become more compressed and the size changes significantly.
- (4) When all the data is inserted, the points are no longer changing and form a steady relationship between users based on the topics.

In this example, we select two topics of blogs to visualize their effect to the users. In the Figure 11, (b) is the normal user layout, and (a) is the result after clustering. With clustering, the more special

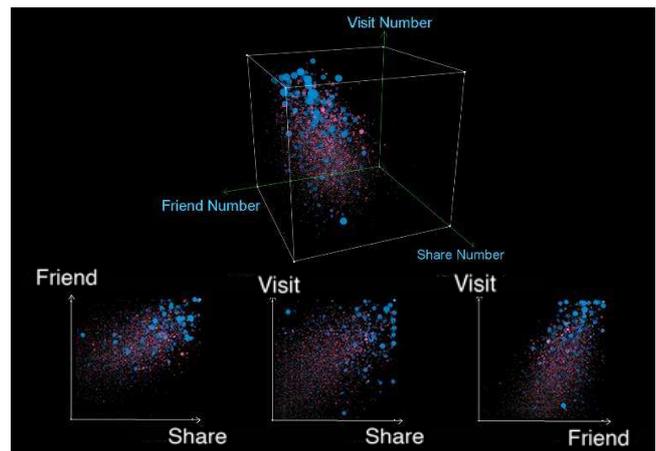
users are more noticeable. Most of the normal small nodes has been clustered together to be considered by groups.

5.1.2. Topic Result: Three aspects of Pitaya

According to the appearance of different aspects, we made the following conclusions:



(a) Clustered Visualizing



(b) Normal Visualizing

Figure 11: Topic Comparison Result

(1) Visit-friends:

More friends lead to larger visiting number. Most of visiting data are located at the high number of friend side. Compare these two groups of data, one may find people pay much more attentions on the workplace topic (group b, green). However, within this topic, it rarely forms a focus incident. In contrast, with the report of traffic accident in China (group a, blue), users are not only very interested but also immediately gathers a large number of users and forms a focus incident.

(2) Share-friends:

From this aspect, it is observed the number of friend does not relate to sharing. It mainly depends on the topic of the posts. Less posts about the

workplace topic (group B, red) is shared than traffic accidents (group A, blue). This trend also applies to the focus incident.

(3) Visit-share:

The top left of the figure is the active users of the net. This group of people are also known as "opinion leader". They are more likely to browse and share information. Visiting and sharing information are mainly depended on the topic. People are more interest in the focus incident.

6. CONCLUSION

By developing a visualization tool, we control the dynamic performance by naturally touching gesture and combine human-computer interaction together. The process of visualization is to show the mass data by art path, which will be a fantasy sky map with real metaphor. The data is from a real social network site in China, so the results reasonable for future research and design. We build a visualization model for large scale social network, which also fits the cognition of image as a whole for people. By the model, we are able to show the social network in a holistic view to understand SNS as a whole, including topics comparison, information flow and a new clustering method. The overall situation may show the information communication effect to be basics for future research and development in the social network. Distinguishing the useful information enables us to obtain more precise results, e.g. people's preference in virtual world, the trend of an incident spreading. The visualization work in this paper is not only analysing data, but also incorporating entertaining and participatory factors. As a result, this project is not only applied for personals and companies, it can also be developed to be a social game and be promoted to common users.

7. ACKNOWLEDGMENT

This work is supported by National Natural Science Foundation of China, No.60933013 and No. 61003097; National Basic Research Program of China, No. 2011CB302206; China Postdoctoral Science Foundation, No. 20100470285, and National Key Project Series, No. 2011ZX01042-001-002.

8. REFERENCES

Jia, Y., Garland, M. and Hart, J.C. (2009) Hierarchical Edge Bundles for General Graphs.

Heer, J. and Boyd. D. (2005) Vizster: Visualizing online social networks, Computer Science Division, University of California.

Kim, S., Han, H., Jeon, K., Jung, H. and Yan, H.Y. (2009) Visualization of Information Flows in a Very Large Social Network.

Skold, M. (2008) Social Network Visualization. Royal Institute of Technology.

Hao, H-X., Wu, L-D. and Song H-C. (2010) Study on visualization of network community and its links. Computer Engineering and Applications.

Ding, X-D. (2007) Design and Implementation of social network visualization and analysis tools for network inspection. Shanghai Jiao Tong University.

Wang, N. (2009) Empirical Study on Relativity Between Interpersonal Interaction And Social Capital, Based on Campus SNS Network. Shandong University.

Du, J. (2009) Chinese Network Communication Research. Zhejiang University Press.

Roam, D. (2009) The Back of the Napkin. Citic Press.

Zhu W-Z, Chen C-M and Allen, R-B. (2006) Visualizing an Enterprise Social Network from Email. ACM Press.

Jia Y., Hoberock J., Garland M. and Hart J. (2008) On the Visualization of Social and other Scale-Free Networks. Visualization and Computer Graphics.

Batagelj V. and Mrvar A. (1998) Program for Large Network Analysis. Connections.

Adam P. and Ben S. (2006) Balancing Systematic and Flexible Exploration of Social Networks. IEEE Transactions on Visualization and Computer Graphics.

Zhong Y., Li X., Fan M-M. and Shi Y-C. (2009) Doodle space: painting on a public display by cam-phone. International Multimedia Conference. ACM Press.

Fan M-M., Li X., Zhong Y., Tian L., Shi Y-C. and Hao W. (2011) Surprise Grabber: A Co-located Tangible Social Game Using Phone Hand Gesture. CSCW 2011, ACM Press.

Watts D. (2002) A Simple Model of Global Cascades on Random Networks. In Proc. National Academy of Sciences.

Marlow C. (2006) Linking without thinking: Weblogs readership and online social capital formation. International Communication Association.

Lampe C. and Johnston E. (2005) Follow the (slash) dot: effects of feedback on new members in an online community. ACM Press.