

# Visualizing Group User Behaviors for Social Network Interaction Design Iteration

Zhenyu Gu<sup>(✉)</sup>, Jia Ming Yu, Zhanwei Wu, and Zhan Xun Dong

Shanghai Jiao Tong University, Shanghai, China  
zygu@sjtu.edu.cn

**Abstract.** Considering the popularity of UCD methods in recent years, it's no surprise that User behavior data analysis has become an important tool in design process. Behavior tools based on data mining technology, such as Flurry and Google Analytics, is widely used in web-based applications to support quantitative user research. However, information visualization provided by those tools is usually adapted to business other than design needs, which could be hardly used by designers. In this paper, successful experience from CRM system is analyzed, relations between user behavior and pattern is studied according to design factors. A prototype visualizing user data gathered from a social photo app is developed to integrate user behavior visualization into interaction design iteration process. User experiments are conducted to evaluate the prototype system. Results shown, interactive visualization on real-time user data could help to promote design iteration.

**Keywords:** Quantitative analysis · Information visualization · Design iteration

## 1 Introduction

With the development of social network and pervasive computing technology, social products that integrate social networking functions into daily activities are becoming more and more popular. However, in the initial design investigation or later design iterations of these products, traditional small-scale user test is usually problematic, including biased sample, difficulty in surveys for massive data or incapability of static questionnaires to reflect dynamic changes in social networking activities. Thus, big data processing techniques are introduced by some researchers and developers, to optimize interaction procedure through the analysis of social behavior data [1]. In addition, Cloud Computing technology is adopted to collect large-scale user behavior data and preference [2].

However, it is extremely difficult to find out the correlation between user behavior patterns and the design variations in the massive user data, which hinders the development of these applications. Research has shown that proper visualization enables designers to find the implicit correlations in big data and understand the trends for further design improvements [3].

In this paper, we take the interaction design of a social digital photo frame as an example to show how to perform visualization analysis on collected large-scale user

data and apply the results to the interface optimization. In the end, we testified the effectiveness of the method through the designers’ practical real use.

## 2 Data Definition and Preprocess

The loading capacity of social applications can be evaluated through the number of hits of users captured by the server side. For user behavior analysis, the interaction activeness among users can also be evaluated with the user access behavior. Through the analysis of the user behavior data, social habits in different dimensions can be revealed, including daily online time and access frequency. In addition, these data can be used to enrich the node graph visualizing the social connections among the users with extra dimensions.

### 2.1 Definition of Dataset

Based on graph theory, user behavior data in social network can be divided into nodes set  $U$  and links set  $L$ . The nodes set is user information  $u_i \in U$ , which can be described as  $\{N_u, N_a, T, A\}$ . Wherein,  $N_u$  is the ID number assigned in the social network,  $T$  is the device type used by the user,  $N$  is the nickname of the user and  $A$  is the activation time of the user. The detail definition of nodes set is given as followed:

- User ID in social network ( $N_u$ ) is unique number of every user, which is the identification of each user.
- User Device type ( $T$ ) is the device which user login the social network, for example mobile, tablet and PC.
- User nickname ( $N_a$ ) is a kind of text tag, which may imply the relationship between users.
- User Activation time ( $A$ ) is the first time of user to login this social network.

Links set  $l_{ij} = (u_i, u_j)$  describes the relation between user  $u_i$  and user  $u_j$ . The relation here can be divided into two layers  $f_{ij}$  and  $s_{ij}$ . The  $f_{ij}$  is relation in static data layer, while  $s_{ij}$  is the one in dynamic data layer. The difference between these two data layers is described as followed (in Fig. 1):

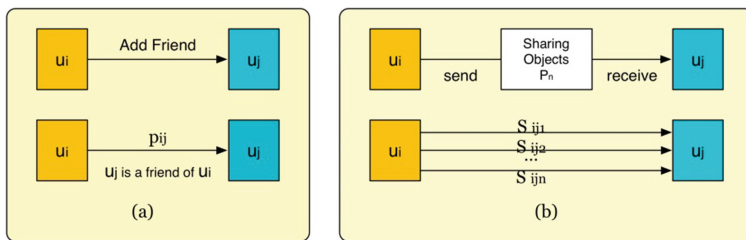


Fig. 1. Links based on contact layout; Links based on share action

- The link  $f_{ij}$  in static data layer is mainly based on the followers and friend list of users. In social network, the  $f_{ij}$  indicates the only directed line from  $u_i$  to  $u_j$  which means  $u_i$  add  $u_j$  as friend and the only value of  $f_{ij}$  is the create time of relation.
- The link  $s_{ij}$  in dynamic data layer makes share resources as primary key. Since some users will allow strangers to get their sharing content, through mining the commend records and related operation records of the same content,  $s_{ij}$  can indicate the content P shared from user  $u_i$  to user  $u_j$ . The P can be described as a collection of resources  $\{p_1, p_2, p_3 \dots p_n\}$ . Wherein,  $p_i$  can be described as  $\{P_N, P_T, U_T, N_F, N_T, T_G\}$ , in which  $P_N$  is the name of the resource,  $P_T$  is the created time of the resource,  $U_T$  is upload time,  $N_F$  is the ID of user who shares the resource,  $N_T$  is the ID of user who receives the resource and  $T_G$  is the tag added while sharing.

### 2.2 Data Preparation

Since social applications interact with server frequently, most of the user operation records can be analyzed from server log files. According to the data structure defined above, the raw data need to be got should contain user information, user relations and operation records. Among those, user information and user relations are both synchronized to the server. And the user operation records can be analyzed from the server log files of HTTP requests, since the parameters carried in the web request package may indicate the operator, operation time as well as other key elements.

In the visualization module, the target data is divided into relational data and line items. For efficient analysis, database is combined by line database and graph database. The line database stores user operations records, which are sorted by operation time, and the graph database stores user basic information as well as friend relations (in Fig. 2).

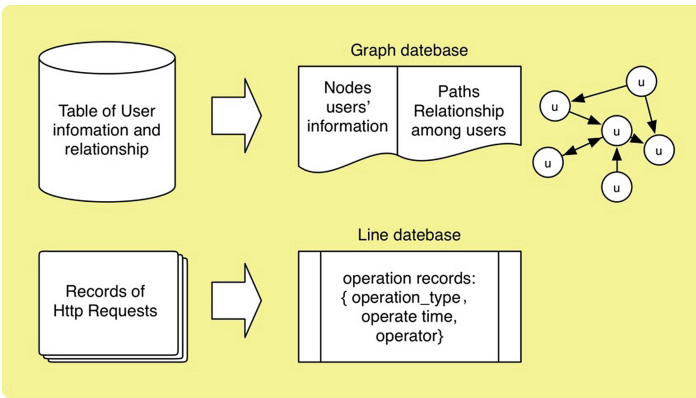


Fig. 2. The workflow of data preprocess

### 3 Visualization Solutions for User Behavior Data

It has been proved that the information visualization is an effective way of perception, because it can complement and strengthen the psychological perception with visual factors [5]. In prediction of intelligent transportation, the visualization techniques are applied to display the real-time status of traffic so as to identify potential congested roads quickly and make scheduling decisions timely [6]. In design process, in order to plan specific design goal, the designer will define an abstract character (Persona) from user research as well as depict the storyboard to describe the scenarios [7]. Then the interaction process will be defined clearly based on them.

The three processes above are familiar to designers. In many research, it is shown that association fresh knowledge to familiar things can improve the ability of understanding new knowledge. Next paragraph will mention a classification process of user data according to the design factors above.

To reflect the importance of the user behavior analysis of application, the sample data is from a family photo sharing social network whose online users are 4773. Different from stand-alone applications, the user behavior data not only contains line operation records of single user, but also involves associated operation records between users.

#### 3.1 Frequency-of-Using Visualized by Heat-Map Matrix Diagram

Heat-map usually labels the weight in graphic by using different markers, such as attractiveness and click times. In the diagram, the weight is usually preformed by the brightness and hue of color.

In social network applications, frequency of accesses to SNS (social network sites), total time spent on SNS per user and number of online users over time are major variables of statistics. These variables can reflect user viscosity of the application, peak period of online users and etc. To designers, these variables can be linked to the user habits as defined in the model of persona.

In order to adapt to the visualization of complex multi-dimension data, we learn from heat matrix [8] to visual the interrelation among short-time distance, long-time distance and amount of information. This form was indicated helpful for designers to understand data from the macroscopic point of view through user interviews.

#### 3.2 Connections Among Multi-devices Visualized with Bipartite Diagram

Bipartite graph, also known as bigraph, is a kind of special model in graph theory. Bipartite graph is a graph whose undirected vertices can be split into two disjoint sets (A, B) and two vertices  $i$  and  $j$  associated in same edge are in each of the two different vertex set A and B in figure [9].

For multi-terminal applications, understanding the proportion and correlation ratio of different terminals can help people to judge the trend of product in the market. The distribution and active number of all the members in social application are both

important for analysis, which can assist marketing to make better decisions and can also provide information to enrich the usage scenarios (Fig. 3).

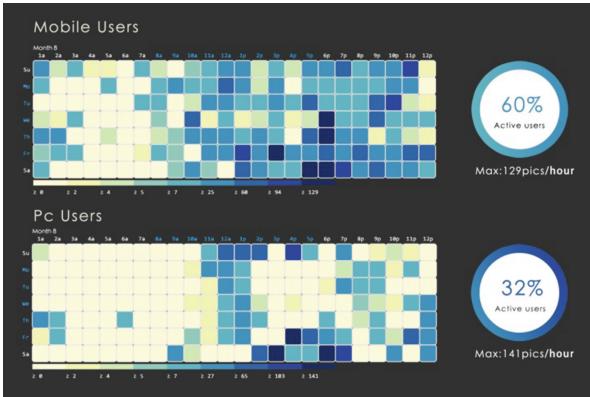


Fig. 3. User sharing frequency (The top is the mobile user and the bottom is PC user)

In the visualization, in order to describe relevance of terminal connections and photo-sharing amount, these are all visualized in the form of bipartite graph and added interactive mode to display association between two datasets (in Fig. 4). In the interview, some designers said that they could find correlation among data much more easily by combining interactive animation.

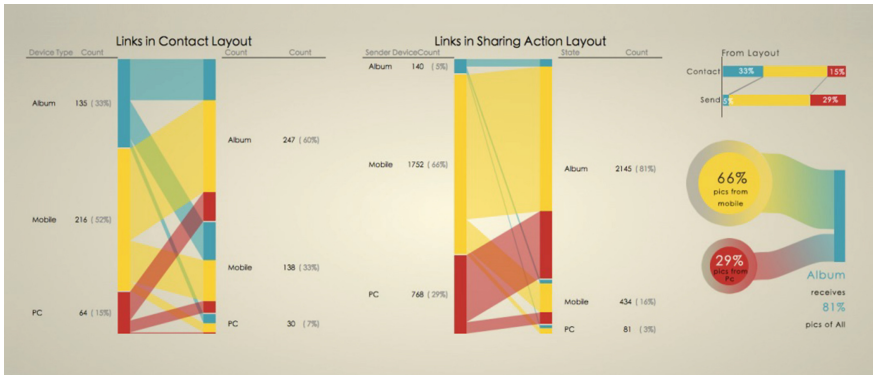


Fig. 4. Connections among multi-terminals

### 3.3 Using Paths and Interests of the Users

User traffic is usually used as a judgment standard of page attractiveness in Web site design such as conversion rate measuring the attraction of webpage content. In design, frequency of every operation can be mapped into the step defined in interaction process

model. In order to get better express of interrelation between operation depth and usage frequency, the design of visualization combines the Sankey diagram [10], flow diagrams and tree diagram into it in order to display the usage amount of each operation and user difference between two operations (in Fig. 5).

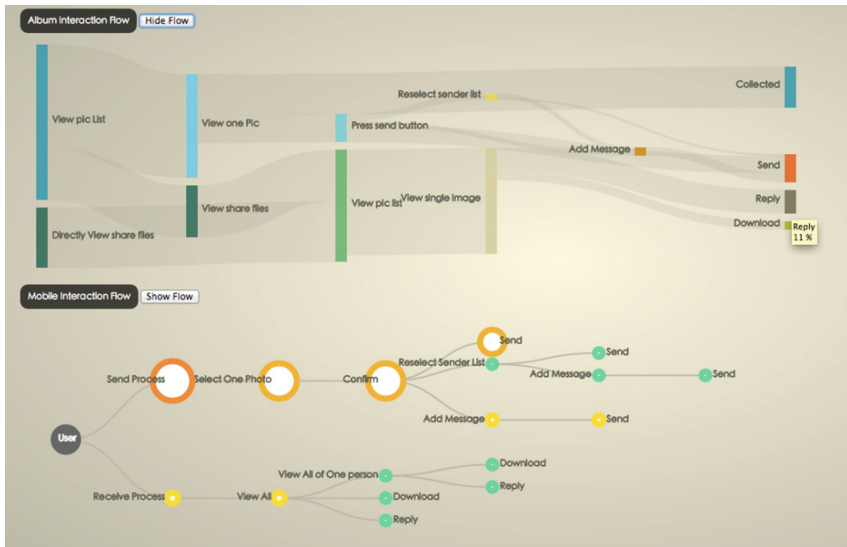


Fig. 5. Tree diagram and flow diagram of operations

Flow diagram is usually used to display quantity change. The connections in flow diagram are described as strip graphics whose height infers the value of weight. Tree diagram has strong meaning of levels, which can describe the sequence of operations directly. Sankey diagram, also called sankey energy split graph or Sankey energy balance diagram, belongs to a specific type of flow diagram. The width of branches in the diagram corresponds to the size of data flow and the width of beginning branches equals to the width of end branches, which performs the balance of energy. Interviews of users shows that the form of tree diagram is much more fit to display the unidirectional and hierarchy of operations directly and the flow diagram combining the Sankey diagram is conducive for users to understand the variation of user amount between different levels, which makes the diagram hold context properties.

## 4 Applying the Visualization Solutions

Based on the analysis of the three visualization diagrams above, the following will associate UI elements in actual App with user elements. The target application is a family photo-sharing platform, which contains three different devices (mobile, digital albums and pc). There are 4773 online users using this application currently, and the earliest user started to use the app since September 2013.

#### 4.1 Comparing Behavioral Characteristics Among Multi-terminals

In Fig. 6, the proportion of mobile users and pc users is 2:1, while the sharing peak of pc users is higher than the one of mobile users. Through the distribution of the nodes in diagram, it can be inferred that the sharing frequency of mobile users is much more decentralized than the one of pc users and the average sharing amount of mobile users is not very high. The sharing frequency of pc users is much more concentrated and the amount of photos shared once is higher.

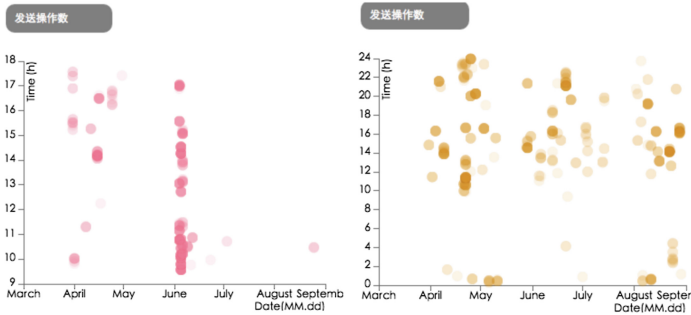


Fig. 6. Sending frequency of PC user (left), Sending frequency of mobile user (right)

Moreover, under the circumstances of same sending activity (total photos shared), a mobile user and a pc user are chosen to do the track comparison. By the comparison, we found that:

- (1) Mobile users usually share pictures instantly and single sharing amount is lower, but the frequency of sharing is higher.
- (2) The frequency of pc users is lower, but the sharing strength is higher, which means pc users have strong demand of batch sharing.

#### 4.2 The Flow of Content Sharing Among Multi-terminals

In Fig. 4, contrast to the ratio of user connections and photo sharing amount, we can find that the largest number of connections exist between mobile user and album users, and the second largest number of connections exist between album user and mobile user in user relations layer. And in sending action layer, most of the photos are sent from mobile users, and album users receive 80 % of those photos, so most of feedback design should be applied between mobile users and album users.

#### 4.3 Conversion Rate of Operations

Conversion rate is mainly used to evaluate the attractiveness of web-page design currently and in app design it can also be used to measure usability. Compared to the statistics of single operating amount, conversion rate can reflect the characteristics of

process better. In interaction process diagram (Fig. 5), the operations with lower conversion rate are: replying messages when receiving photos, collecting photos and downloading photos. Among those, the low conversion rate of downloading photos is caused by automatic downloading function, so it cannot be optimized as target operation necessarily. Besides, although replying message and collecting photos do not affect the process of photo previewing and playing, these two operations belong to feedback design, which means the single operation may involve multi-users.

The feedback design mentioned above is mostly applied between mobile users and album users. The status of low feedback causes the problem that implementation rate of related operations in different devices is also lower. By the flow chart of users, we found that the depth of feedback operations is all 3 or 3 above, which means users have to go through at least 3 pages to read or commend a photo since they see the notification of new photos. Such design is obviously not conducive for the high feedback from users.

## 5 Test and the Efficacy

In order to verify the helpfulness of the solution, six students with design education background are invited to use this visualization tool. After a certain learning period to understand these visualization diagrams above, they all gave some solutions to improve the interaction design of this application. The main improvement advices given are:

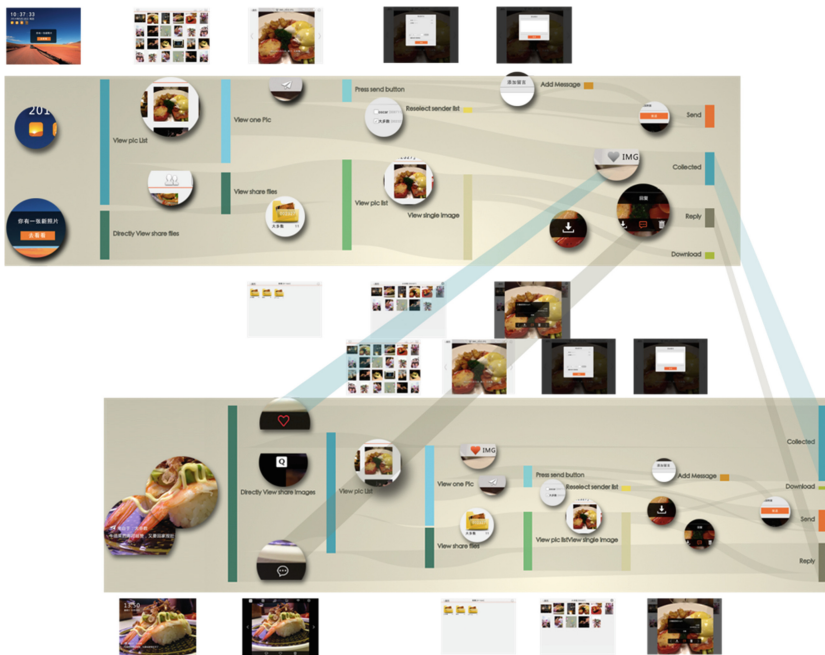


Fig. 7. Comparison of interaction process before and after optimization



- (1) On the basis of original user habits, the feedback design in the process should be simplified. In order to allow album users to make high feedback to others, such operations such as photos commending and collecting should be migrated from the current depth that is deeper than level 3 to the position of level 1.
- (2) In order to stimulate the album users user feedback operations, the messages caused by such operations are all displayed in the front-page automatically.

According to the advices above, the app for album users is redesigned, and some of the users were invited to use the new version. The new user behavior data collected from these test users are visualized by the same solution in paragraph 3. Compared to the diagrams driven by the old user data, the conversion rate of feedback operations is obviously higher than before (in Fig. 7).

## 6 Conclusion

Visualizing data by the demands of target users is one of the measurements to enhance the readability of diagrams. In this paper, the target demands of visualization are found from the design elements in the model of persona, scenario and interactive process and the demo was developed on the basis of an online photo social application. Then, students with design education background were invited to give improvement advices about interaction design by reading the visualization graph. And the data collected from small test sample is used to evaluate the effectiveness of the improvement advices.

From the usage scope of this visualization solution, the effective visualization solutions can inspire designers to give more reasonable optimization advices. While due to the need of data-driven, it is not suitable to be used for preliminary research but more suitable for design optimization and evaluation in later stages of design. In the respect of application process, the capture module and storage module of the user behavior data should be designed during the development of the application because the visualization solutions need to collect a large amount of the data as a precondition. Currently, the visualization solution provided in this paper is only based on a common design process. With deeper research of optimization target in different devices, the visualization solution can be much more refined and targeted.

## References

1. Benevenuto, F., et al.: Characterizing user behavior in online social networks. In: Proceedings of the 9th ACM SIGCOMM Conference on Internet measurement Conference, ACM (2009)
2. Xu, M., Walton, J.: Gaining customer knowledge through analytical CRM. *Ind. Manage. Data Syst.* **105**(7), 955–971 (2005)
3. Keim, D.A.: Information visualization and visual data mining. *IEEE Trans. Vis. Comput. Graph.* **8**(1), 1–8 (2002)
4. Balachandran, A., et al.: Characterizing user behavior and network performance in a public wireless LAN. In: ACM SIGMETRICS Performance Evaluation Review, vol. 30, no. 1. ACM (2002)

5. Meirelles, I.: *Design for Information: An Introduction to the Histories, Theories, and Best Practices Behind Effective Information Visualizations*. Rockport Publishers, Beverly (2013)
6. Uchida, M., Shirayama, S.: Formation of patterns from complex networks. *J. Vis.* **10**(3), 253–255 (2007)
7. Kankainen, A.: UCPCD: user-centered product concept design. In: *Proceedings of the 2003 Conference on Designing for User Experiences*. ACM (2003)
8. Wilkinson, L., Friendly, M.: The history of the cluster heat map. *Am. Stat.* **63**(2), 179–184 (2009)
9. Dulmage, A., Mendelsohn, N.: A structure theory of bi-partite graphs. *Trans. [8] Roy. Soc. Can. Sec. 3.* **53**, 1–13 (1959)
10. Schmidt, M.: The Sankey diagram in energy and material flow management. *J. Ind. Ecol.* **12** (1), 82–94 (2008)