SocialVis

Improving Social Network Knowledge through Information Visualization

Sofia Penaforte

Instituto Superior Técnico
sofia.penaforte@tecnico.ulisboa.pt

Abstract. The popularity of Social Network Sites (SNS) like Facebook is generating high quantities of personal information. However, such information is scattered across different concepts (posts, chats, comments, etc). This dispersed information hinders users to have an aggregated perspective of their social networks. Although there is related work using visualization techniques to provide better insights to users about their social networks, none of them provides a multi-dimensional approach: (1) the visualizations focus on the network itself and are not user-centered; (2) each research work only focus on a particular narrow set of dimensions about the social network. This thesis argues that a set of multi-dimensional user-centered visualizations can improve users' insights about their social networks. We implemented such visualizations in a prototype called SocialVis, which provides four main views: (1) global view of the network; (2) shared groups; (3) shared likes; (4) interactions. All views spawn different dimensions of the social network. They are also interconnected and their visualizations react to filters that the user activates when interacting with them. Preliminary tests, usability tests and case studies show that SocialVis enables users to answer complex questions about their social networks that would not be possible while using only Facebook.

Keywords: Social Network Sites, Visualizations, Facebook, User-centered

1 Introduction

The current growing popularity of Social Network Sites (SNS), such as Facebook, is generating large quantities of personal information. However, the same SNS provide such personal information in a way that is intended for immediate consume. This format hinders users to have an aggregated perspective of their social network. In other words, users can easily answer questions like 'what did you talk about with a particular contact last night?', but will have a difficult time while identifying the most discussed topics in the past three years with that person.

Generally, the behavioral patterns and relationship trends between users and their contacts are not immediatelly accessible. The information is there, but is scattered along many different concepts that are not intended for analitical usage.

Although there is related work using visualization techniques to provide better insights to users about their social networks, none of them provides a multi-dimensional approach: (1) the visualizations focus on the network itself and are not user-centered; (2) each research work only focus on a particular narrow set of dimensions about the social network.

This thesis states that, having a pre-defined set of visualizations that can leverage in transformed aggregated view of the data originated by SNS, users are empowered to have better insights about their social networks.

Regarding document structure, this section presents a brief introduction about this research problem and our work towards it. Section 2 discusses related work and identifies the main issues that need to be addressed to enable users to be more knowledgeable about their social networks. Section 3 presents our solution proposal, which is evaluated with tests and case studies discussed in Section 4. Finally, the document concludes with Section 5.

2 Related Work

There are related works that focus on letting users know about topological characteristics of their social networks. These characteristics show how the social network members are related among themselves, which connections are shared, and what are the existing degrees of separation.

The Sociogram concept, introduced by Jacob L. Moreno in 1930 [1], consists on a social visualization used to map social networks. These kind of visualizations focus on revealing two main kinds of organizational patterns: (1) social groups, which consist on groups of people that have strong connections, and (2) social status, which consists on people that are related in a similar way [1].

Heer and Boyd [2] have created *Vizster*, which is a tool that focus on providing users a view of their social networks using interactive sociograms. In these sociograms, each member of the social network is represented by his name and photo.

Sociograms allow users to explore how the members of their social networks are connected among themselves. Two types of techniques are used to provide such information: (T1) connectivity highlighting and (T2) connection view. Through the connectivity highlighting technique (T1), users can select a node and the nodes with a maximum of two degrees of separation are also highlighted. The network is presented in a egocentric perspective, focusing on the individual and her most direct friends. Such approach follows the philosophy of Shneiderman [7], saying that "start with what you known, and grow on that". This philosophy attemtps to reduce the user's perception effort in depicting information from the sociogram. With the connection view technique (T2), the user can know how two particular members of her network are connected, revealing any mutual friendships that may exist.

Perer and Shneiderman [6] allows users to analyze their networks, combining both perspectives of social visualization and statistical data analysis. Although not applied to Social Network Sites, the combination of social visualization and

statistical data analysis appears to be trending [8, 6] as more and more social network visualization systems use graph representations to implement their features [3].

NodeXL [8] is an extension for Microsoft Excel that allows to perform analysis on social networks. This tool integrates network analysis metrics and graph creation algorithms. The tool can be used to analyze data of SNS like Facebook and Twitter, making easier the identification of key-contacts, documents, groups and events.

Nardi et al. [4] created *ContactMap*, which consists on a tool that attempts to model the contact network of the user, showing who is more important in both his personal and work lives.

Other related work, known as *PeopleGarden* [11], allows users of online forums to have a data portrait about how their community behaves. This research work uses flowers as visual metaphors to represent individual data portraits, and gardens as the combination of those individual data portraits.

Viégas et al. [10] proposed another work known as *Themail*, which provides a visualization that allows users to identify their relationships throught the capture of their email history. The tool focus on helping users with two main questions: (Q1) what kind of subjects are discussed with which email contacts; and (Q2) in what way does my conversations with that person differ from the ones I have with other persons.

Also concerning the analysis of email history, the *PostHistory* [9] work pretends to show the user, what is his interaction level along a year time frame. This analysis shows if the user consumes or produces more messages, and who are the recipients of those messages.

Concluding, there are a lot of research works that use visualization techniques to provide more insights to users about their social behavior, whether in email, foruns or social networks sites. However, none of them focus on the egocentric aspect of the information, i.e. in what way does it concern the user using the application. None of these research works provide user-centered visualizations that can show for instance who are the contacts that the user most interacts with, or share the same favorite musical artists. Also, the visualizations normally have limited interactivity and interconnection, meaning that the user cannot for example ask filtered questions like: from the contacts that are female and single, who do I share the most favorite tv shows with?

3 Solution

The goal of this research work is to study information visualization techniques that allow users to easily have an egocentric perspective of patterns about their social network in an holistic view. Essentially, the thesis states that providing a set of interactive visualizations that are user-centered, can in fact enhance the user knowledge about his social network. Such enhancement allows users to answer questions that they simply could not by just using the Facebook site.

The SocialVis [5] is therefore composed by three main views, which comprehend a set of visualizations that the user can interact with, and that are interconnected. Visualizations are interconnected in the sense that, if the user by interacting with a particular visualization applies a particular filter, then all other visualizations adapt to that filter as well. Hence, if the user selects the users living in a particular country, then all other visualization will only show contacts from that country.

3.1 Global View

In Figure 1, an example of the global view is shown. In such view, the user can identify the distribution of gender crossed with relationship status of her social network. By clicking on a particular combination, a filter is applied, and all visualizations will refresh to show only contacts with the same gender and relationship status that the filters correspond to.

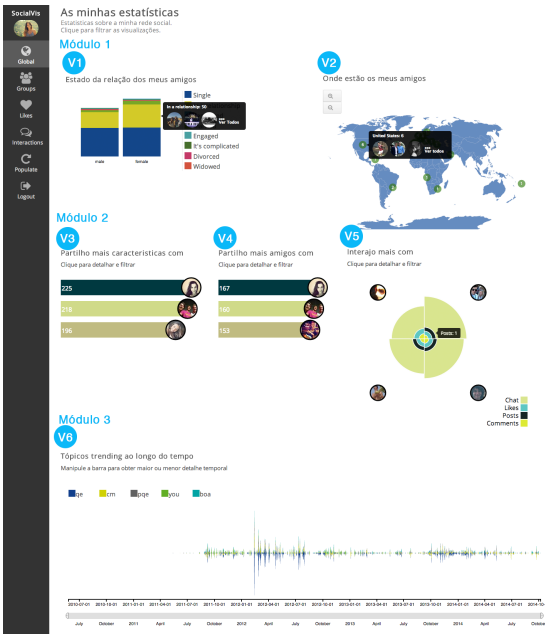

Fig. 1. Global View: Global Overview of the Social Network

The world map essentially shows the geographic distribution of the contacts through the world. The user can also click in a particular country or city, and apply a filter to that geographical zone.

In this same view, there are two bars visualizations that essentially show the top 3 contacts having more things in common (i.e. likes, friends and groups),

and the top 3 contacts with whom the user share more friends with. Both these visualizations react to mousehover, highlighting the mousehovered contact in all other visualizations present in that view. More to the right, there is a visualization that focus on showing the 4 contacts with the most interactions (i.e. chat, likes, posts and commentaries). By clicking on a particular type of interaction, the user is redirected to the interactions view that show that particular kind of interaction in detail, i.e. overtime and in an aggregated way.

At the end of this Global View, a streamgraph appears to show the five most discussed topics within a particular temporal window. This window can be adjusted through a slider mechanism that allows the user to redefine the interval of time to some particular granularity of interest (the most granular window is the day).

3.2 Shared Characteristics View

The second and third views in SocialVis regard the shared interests of the user and her contacts. First, we have the Shared Groups view (see Figure 2) that works in a very similar way to the Shared Likes view. Hence, no need to explain both views individually.

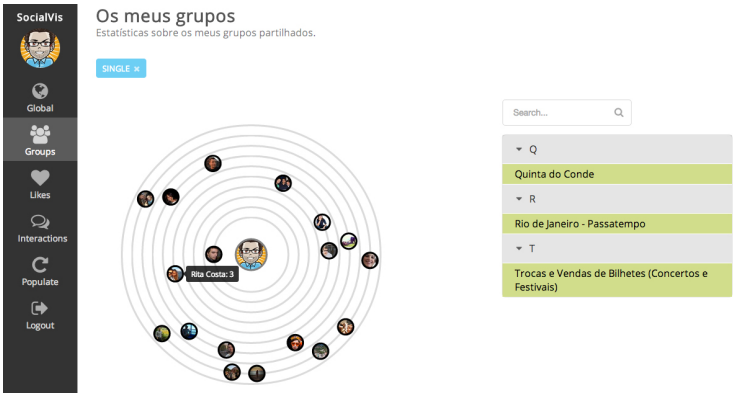

Fig. 2. Shared Groups View

The shared groups view focus on showing the user's contacts in a radial visualization, where the contacts sharing more groups appear closer to the user (which stands in the center of the radial layout), and users that share less groups appear far away from the center. By mousehovering a contact in this radial visualization, the number of shared groups appears as a tooltip, and the list of groups on the right is momentarily filtered during the duration of the mousehover event.

3.3 Interactions View

This is the last view provided by SocialVis, which depicts the user's interactions both grouped and overtime in detail. It is also possible to select a second contact from the contact list on the right and compare both contacts. This view ignores the filters as it would not make sense comparing two users while restraining a set of other characteristics actively filtered.

In Figure 3, it is illustrated the two visualizations that compose the interactions view. The first visualization shows the interactions with a particular contact overtime. It is possible to select a particular type of interaction among the available: chat, commentaries, posts and likes. Note that the interactions' bidirectional nature is shown in this visualization, as it differentiates the interactions initiated by the contact (top part of the graphic) from the interactions initiated from the user herself (bottom part of the graphic).

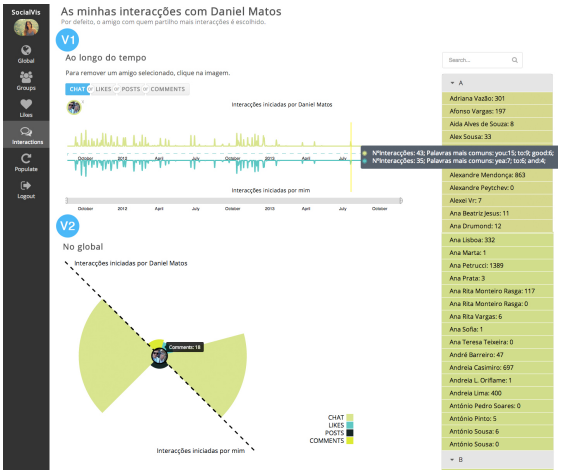

Fig. 3. Interactions View: Single Contact Perspective

Figure 4 illustrates the interactions view when showing a comparison between two contacts. In this case, both visualizations adapt to consolidate the different types of interactions (i.e. chat, likes, commentaries and posts) and provide two color lines to distinguish the contacts being compared.

Also the bottom visualization, adapts to globally identifying who starts the most interactions, independently from the interaction type.

3.4 Implementation

SocialVis is a web application that is built using PHP and MySQL, with the Laravel MVC framework, and D3.js for the visualizations. The filters mechanism was implemented using PHP sessions that would save a list of active filters

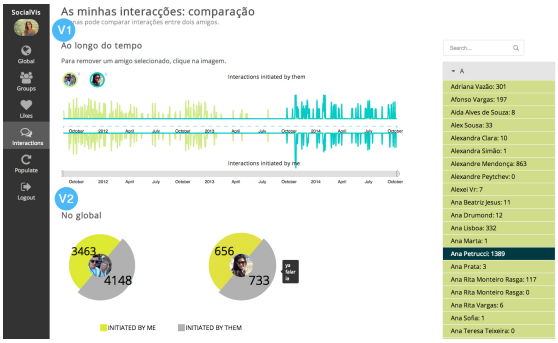

Fig. 4. Interactions View: Comparison Perspective

to be applied to all queries that visualizations would make to the server. Limitations regarding the Facebook API lead to throttling solutions as Facebook sets a maximum of 3600 requests per hour. This limitation in addition with the quantity of data of a relatively active user (with more than 500 friends), that is registered in Facebook for some time (more than 2 years), could take a complete day to download all the necessary data for SocialVis to begin displaying its visualizations. To overcome this limitation, a mechanism of jobs was implemented, and each job would be saved into the database, and handled according to its priority.

4 Results

In the preliminary tests conducted with the first SocialVis prototype, all five users demonstrated surprise and satisfaction while viewing the kind of information they could learn about their social networks. In particular, all of them have said that SocialVis showed them facts they would not be able to know through the Facebook site. Four of these five participants showed interest in continuing to use SocialVis in a daily basis.

With the feedback gathered from the preliminary tests, SocialVis was refactored, and a second improved version of the prototype was developed. This new version would support interconnected visualizations with the possibility for users to apply filters while interacting with each one of the visualizations.

4.1 Usability Tests

For the usability tests, 20 participants that have an advanced level of proficiency with Facebook were selected. Most users reported through a survey that they access Facebook in a daily basis. The tests consisted in 8 different tasks, using the data of an external user. Before the test, each participant had a debrief session of 5 minutes, where each visualization and interaction was explained. Aftewards, they were be given 10 minutes to explore the application freely on

their own before starting the test. For each task, a maximum of 5 minutes was established.

Given specific tasks with unambiguous responses, their rationale follows:

- **Task 1:** Check if the user can identify, through the Global View, the contacts with whom he interacted the most.
- **Task 2:** Check if the user can identify the most discussed topics in the global view, independently of any user.
- **Task 3:** Check if the user can identify which contacts are living in a particular country.
- **Task 4:** Check if the user can identify common interests with a particular person.
- **Task 5:** Check if the user can identify which contacts belong to a particular group.
- **Task 6:** Check if the user can identify who is a conversation starter with a particular contact.
- **Task 7:** Check if the user can identify the most discussed topics with a particular person.
- **Task 8:** Check if the user can identify, between two contacts, with whom does he interacts the most.

In Table 1, the results from the usability tests performed using the tasks described above are depicted.

	Time (in seconds)				**Errors**				**Task Completion**
Task	$\bar{x}$	σ	**Max**	**Min**	$\bar{x}$	σ	**Max**	**Min**	**%Success Rate**
Task 1	26.3	14.21	60	8	0.1	0.31	1	0	100%
Task 2	107.45	80.08	300	37	0.55	0.83	3	0	90%
Task 3	18.3	7.02	35	10	0	0	0	0	100%
Task 4	9.75	2.47	14	4	0	0	0	0	100%
Task 5	27.5	21.97	110	13	0.25	0.79	3	0	100%
Task 6	33.2	23.02	117	9	0.15	0.67	3	0	100%
Task 7	100.35	61.98	300	35	0.6	1.43	5	0	95%
Task 8	36.25	16.40	69	17	0	0	0	0	100%

Table 1. Usability Tests Analysis

According to the table above, both tasks 2 and 7 are the worst performing in both time and success rate parameters. This is mainly because those tasks were the most complex ones, but also because the interaction pattern with a slider caused difficulties to some users that revealed not being used to that particular user interface component. In a general way, all other tasks were relatively successful regarding that users only got 5 minutes of pre-explanation about how

SocialVis works, and 10 minutes to use it freely before beginning the test. Another point important to refer is that four of the 20 participants did not want to use the optional 10 minutes to explore the application before begining the test, either because they were short on time, or because, as IT experts, they were extra-confident about themselves. In such particular cases, participants had difficulties in recalling how some of the visualizations work.

5 Conclusions

Nowadays, the popularity of Social Network Sites (SNS) like Facebook originates an excessive amount of data that users find difficult to consume. This is mainly because such data is provided within a particular local and focused context: you known when someone sends you a message, or you can view the latests posts from your connections one by one. However, when users try to answer questions like: what is the most trending subject over the past five months with a particular contact? or who are the three female single connections with whom I share more favorite tv shows? Although that would be possible if one would count them manually, it is a very difficult task. The information is right there, but is not provided to the user in a way that she can efficiently answer such analitical questions.

This document presented an extended abstract of the research work behind SocialVis. The thesis of SocialVis states that using a set of interactive visualizations that are interconnected, users can have better insights about the facts behind their social networks.

Tests with users and case studies showed that such users could answer difficult questions concerning their social network, questions that would be unanswerable while using only the Facebook site.

Regarding future work, SocialVis could benefit from a re-implementation of the filtering system in the client-side. Users with many friends (more than 500) and an intensive usage of the Social Network Site (e.g. chat, commentaries, posts) could actually have their user experience diminished by the responsiveness of the visualizations given the currently implemented filter mechanism.

A final future work pointer could be towards a complete re-implementation of the SocialVis application within a Desktop version that could run locally in the user's computer. This is mainly to tackle the privacy issue that users have while an application like SocialVis gathers all their user social network data online. Maybe with a local Desktop version of SocialVis, users would be less reluctant to have their data downloaded to their own computers.

Bibliography

[1] Freeman, L.C.: Visualizing social networks. Journal of social structure 1(1), 4 (2000)

[2] Heer, J., Boyd, D.: Vizster: Visualizing online social networks. In: Information Visualization, 2005. INFOVIS 2005. IEEE Symposium on. pp. 32–39. IEEE (2005)

[3] Henry, N., Fekete, J., McGuffin, M.: Nodetrix: a hybrid visualization of social networks. Visualization and Computer Graphics, IEEE Transactions on 13(6), 1302–1309 (2007)

[4] Nardi, B.A., Whittaker, S., Isaacs, E., Creech, M., Johnson, J., Hainsworth, J.: Integrating communication and information through contactmap. Commun. ACM 45(4), 89–95 (Apr 2002), http://doi.acm.org/10.1145/505248.505251

[5] Penaforte, S., Guerreiro, J., Gonçalves, D.: Socialvis: Aumentar o conhecimento da rede social através de visualização de informação (2013)

[6] Perer, A., Shneiderman, B.: Integrating statistics and visualization: case studies of gaining clarity during exploratory data analysis. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. pp. 265–274. CHI '08, ACM, New York, NY, USA (2008), http://doi.acm.org/10.1145/1357054.1357101

[7] Shneiderman, B.: The eyes have it: a task by data type taxonomy for information visualizations. In: Visual Languages, 1996. Proceedings., IEEE Symposium on. pp. 336–343 (1996)

[8] Smith, M.A., Shneiderman, B., Milic-Frayling, N., Mendes Rodrigues, E., Barash, V., Dunne, C., Capone, T., Perer, A., Gleave, E.: Analyzing (social media) networks with nodexl. In: Proceedings of the Fourth International Conference on Communities and Technologies. pp. 255–264. C&T '09, ACM, New York, NY, USA (2009), http://doi.acm.org/10.1145/1556460.1556497

[9] Viégas, F., Boyd, D., Nguyen, D., Potter, J., Donath, J.: Digital artifacts for remembering and storytelling: posthistory and social network fragments. In: System Sciences, 2004. Proceedings of the 37th Annual Hawaii International Conference on. pp. 10 pp.– (2004)

[10] Viégas, F.B., Golder, S., Donath, J.: Visualizing email content: portraying relationships from conversational histories. In: Proceedings of the SIGCHI conference on Human Factors in computing systems. pp. 979–988. ACM (2006)

[11] Xiong, R., Donath, J.: Peoplegarden: creating data portraits for users. In: Proceedings of the 12th annual ACM symposium on User interface software and technology. pp. 37–44. UIST '99, ACM, New York, NY, USA (1999), http://doi.acm.org/10.1145/320719.322581