

Published in: Sommerer, C., Mignonneau, L. & King, D., 2008. *Interface Cultures: Artistic Aspects of Interaction* 1st ed., Transcript.

COMMENT FLOW: VISUALIZING COMMUNICATION ALONG NETWORK PATH

DIETMAR OFFENHUBER, JUDITH DONATH,
MIT MEDIA LAB

Abstract

Social networks are abstract organizational structures that help us understand the relationships among a group of interconnected individuals. Much recent research has focused on understanding the structure of these networks. Yet the network itself is a conceptual topology. The key is the activity that flows along the network paths: the support offered, the information given, the gossip exchanged. Our goal in this research is to develop ways of representing social networks that embody the individuality of the nodes and links. By incorporating this information, we can advance our understanding of the social dynamics of the network, including the roles of the individual members and the temporal changes in the network structure. We have designed and implemented a flexible tool for the content driven exploration and visualization of a social network. Building upon a traditional force-directed network layout, our system shows the activity and the information exchange between nodes, taking the sequence and age of the messages into account. This project serves both as an illustration of one approach to the general problem of individualized network visualization and as an example of the practical uses of such representations.

Introduction

In the last couple of years, online social networking platforms have become widely popular. In these services, participants create self-descriptive profile pages that often include their tastes and preferences, and a number of links to other members of the same service, usually called “friends”. The purpose is to meet new people, make new business contacts by taking advantage of network effects.

Social networks are abstract organizational structures that help us understand the relationships among a group of interconnected individu-

als. Much recent research has focused on understanding the structure of these networks, identifying patterns such as bridges, structural holes, etc. and on developing visualizations for these often complex entities.

Typical network diagrams are detailed in their structure, but generic in their representation of the nodes and links. Yet it is the differences among the people (nodes) and their relationships (links) that create the specific structure of each network and that determine the strength and significance of the ties.

The first part of the paper outlines the key issues in visualizing social networks with individuated nodes and links. The second part describes a specific research project - the visualization of communication patterns in an online social network site - carried out within this program.

This project serves both as an illustration of one approach to the general problem of individuated network visualization and as an example of the practical uses of such representations. In the mySpace service (the networking site used for this research) network-only visualization methods are no longer sufficient to meaningfully represent the community structure. Numerous commercial profiles, fake/spam/celebrity profiles and widgets such as automatic friend adders result in a huge number of connections, many of which carry little information about a person's actual social ties and behavior. The average mySpace user has around 131 friends, but there are also profiles with over a million "friends". By going beyond the "skeleton" of network connectivity [1] and looking at the flow of information between the individual actors we can create a far more accurate portrait of online social life. Comparing the first generation of social networking sites to the services that are currently popular, we find that today's social software platforms are more content-centric than their earlier counterparts, which were more concerned with the structural representation of the social network. In recent social software platforms such as Flickr, Facebook, mySpace, the social network serves the purpose of providing context information for the different kinds of activities.

In the last couple of years, online social networking platforms have become widely popular. In these services, participants create self-descriptive profile pages that often include their tastes and preferences, and a number of links to other members of the same service, usually called "friends". The purpose is to meet new people, make new business contacts by taking advantage of network effects.

Social networks are abstract organizational structures that help us understand the relationships among a group of interconnected individuals. Much recent research has focused on understanding the structure of

these networks, identifying patterns such as bridges, structural holes, etc. and on developing visualizations for these often complex entities.

Typical network diagrams are detailed in their structure, but generic in their representation of the nodes and links. Yet it is the differences among the people (nodes) and their relationships (links) that create the specific structure of each network and that determine the strength and significance of the ties.

The first part of the paper outlines the key issues in visualizing social networks with individuated nodes and links. The second part describes a specific research project - the visualization of communication patterns in an online social network site - carried out within this program.

This project serves both as an illustration of one approach to the general problem of individuated network visualization and as an example of the practical uses of such representations. In the mySpace service (the networking site used for this research) network-only visualization methods are no longer sufficient to meaningfully represent the community structure. Numerous commercial profiles, fake/spam/celebrity profiles and widgets such as automatic friend adders result in a huge numbers of connections, many of which carry little information about a person's actual social ties and behavior. The average mySpace user has around 131 friends, but there are also profiles with over a million "friends". By going beyond the "skeleton" of network connectivity (Barabási and Crandall 2002) and looking at the flow of information between the individual actors we can create a far more accurate portrait of online social life. Comparing the first generation of social networking sites to the services that are currently popular, we find that today's social software platforms are more content-centric than their earlier counterparts, which were more concerned with the structural representation of the social network. In recent social software platforms such as Flickr, Facebook, mySpace, the social network serves the purpose of providing context information for the different kinds of activities.

As Donath and Boyd described (Donath and Boyd 2004), most traditional social networking sites rely on a similar model for interpersonal connection that suffers from a problematic simplification. The links between people are mutual and public, but also unnuanced and decontextualized: no distinction is made between the connection to a close relative and a complete stranger. All friend links are equal, and usually equally visible to public. The information is not very reliable - extra information is required to acknowledge the highly differentiated nature of human relationships. Feld used the term 'focus' for the different, sometimes incompatible partitions of social relationships (Feld 1981).

As a result, people browsing public profiles of strangers in order to meet new people do not know whether the interesting profile they encounter represents a real person at all - it might as well be a commercial disguised as a person. There is also active spamming

activity, the fake profiles don't wait to be discovered - members of sites receive numerous invitations, often accompanied by a message written in a personal tone, that are in fact automatically generated spam. The users have little awareness of the nature of the social space they move through, no means of understanding who these other people are they meet / making sense of the space and the people they meet there.

One way of solving this problem is by focusing on the communicative activities taking place in the social network. By looking at their flow of conversation, we can understand who these strangers are.

Related Work

The visualization of social networks has a long history dating back to Jakob Moreno's work on sociometry in the 1930ies(Freeman 2000). Essentially, there are two ways for the graphical representation of social networks: first, the arrangement of the participating actors in the rows and columns of a square matrix, its cells containing information about the relationship between the corresponding actors. The second and more popular way is the graphical representation of the social ties between the participating actors in the form of a node – link diagram. Each node stands for an individual, while their relationships are expressed through connecting lines. While matrix representations have the advantage of an occlusion-free and compact view on the network, node link diagrams are glanceable and easy to understand. However, depending on the complexity of the network, they tend to produce a layout problem.

A number of toolkits have been developed for this purpose, JUNG (O'Madadhain, Fisher et al.), Touchgraph (Shapiro 2002), Guess(Adar 2006) are systems for the display of network structures. The Vizster toolkit (Heer and Boyd 2005) is able to handle large graphs and has been applied for visualizing the community structures in the Friendster(<http://www.friendster.com>) network. However, in spite of the number of existing tools, very few of them are designed for the interactive manipulation of large graphs with several thousand Edges at reasonable frame rates, Working with the data from mySpace, this is a necessity. A second requirement was the possibility of displaying media content, activities and flows on top of the graph. Because of these two requirements, we decided to develop a custom tool designed to fulfill these purposes. Our software builds on top of the conceptual framework of the se-

maspace prototype (Offenhuber and Dirmoser) The software developed for Comment Flow is not restricted to social networks, but designed as a compact tool for exploring, editing and displaying large networks, and thus fills the gap between extensive modular frameworks and compact but limited graph drawing applications.

Conversations in mySpace

Contemporary Social Networking Sites offer a number of ways for their users to interact with each other. The services usually allow the exchange of private messages, but most of the interaction takes place in the public: sharing media, poking, exchange of small gifts or public messages on each other's profiles. The most popular way to communicate with a friend is to post a comment on the friends profile page, where it can be read by all friends or everyone in the network(Amanda Lenhart 2007).

Among the available social networking platforms, we picked the mySpace service for illustrating our concept. The reason for this is not only its present volume, but also the problems and challenges that it poses for the understanding and visualization of social networks:

The cost of adding new connections is very low – a click on a link on the profile page is enough to send a request, after approval through the profile owner the connection is made. In contrast, services such as the professionally oriented LinkedIn(<http://www.linkedin.com>) make this process harder by making sure that the people seeking the connection actually know each other. The low barrier for establishing connections results in a very high number of connections per profile - a median value of 1311¹, whereas profiles with hundreds of thousands connections are not uncommon.

The mySpace network shows many structural differences to real-world social network. As a result of the prolific number of a profile's connections, the mySpace network has a very small diameter. For example, the shortest distance between any two users is usually two, since every new profile starts with a friend connection to Tom Anderson, the president of mySpace.

The role and identity of a profile owner is ambiguous. A profile may represent an individual person, but not necessarily. Fake profiles of celebrities and historical personalities are as common as profiles that are

¹ As determined from our sample of 27000 profile pages, excluding profiles from the music section (bands have a friend count up to 10 times higher than standard profiles, resulting in a median value of 1186 from a sample of 6700 band profiles)

spam-generating advertisements(Aaron Zinman 2007),(Roush 2006). A Friend connection to a fake profile is a way to specify personal tastes and preferences in an implicit way and as such meaningful information. However, since the network does not differentiate between real humans and fake profiles, commercials and organizations traditional network analysis is not useful in that case.

MySpace does not offer an API or tools to facilitate the exploration of the personal networks. The interface could be described more properly as a people directory rather than a social networking tool.

With a large number of friends and a limited amount of space on the front page of a profile, visibility is an important issue. Only a few connections can be displayed, a dedicated space, the “top-n box” allows the profile owner to select which friends should be featured on the front page of the profile.

Another way to appear on the front page of a profile is through a posting in the comment space. This part of the profile is designed similar to a guest book - visitors can leave short personal messages and images as entries. The Comment space however does not discriminate between friends and follows a strict chronological order. Since posting a comment is a convenient way of being visible on the front page of other profile pages, many of these comments have the character of commercials. However, on personal profile pages, the comment section is often used for ongoing conversations among friends.

Looking at these conversational postings, we see that the guest book format has a number of characteristics: First, the communication process is non-collocated - the answer might be instantaneous or follow weeks later. Second, the communication does not take place in the same space - the messages are dispersed over multiple profiles, making it harder to follow the conversation for someone who is not involved. Third, the postings are usually visible to the public. This encourages other friends to join in and comment on an existing conversation. Instead of a traditional one to one interaction, this turns into a hybrid conversation with multiple thematic threads and participants at the same time. However, the threads are not extensive, usually they disappear after two or three bounces. Comparing the incoming communication with the comments on the pages of friends allows for reconstruction for a conversation.

Privacy

Most social networking sites have privacy filters implemented to a different degree. mySpace allows setting a profile to private. However, the default setting is a public profile, and a minority of users takes the step of changing that. Furthermore, its interface is designed in a way so

that it is not immediately obvious which parts of the profile are visible to the public and which are not. But not only the own profile settings but also those of the top linked friends contribute to online identity of the profile owner. *"social status, political beliefs, musical taste, etc, may be inferred from the company one keeps"*(Donath and Boyd 2004)

Most users know that their comments are visible to the public, but when embedded into the profile page within a mass of other comments, this fact is easily tolerated. It is like having a conversation in a crowded restaurant, surrounded by chatter. In a similar way, the fragmentation of messages across multiple profile pages offers, if not privacy protection, at least some level of comfort. However, the situation might change if the available information is extracted and displayed in a different context. As Jeffrey Rosen puts it: *"when intimate information is removed from its original context and revealed to strangers, we are vulnerable to being misjudged on the basis of our most embarrassing, and therefore most memorable, tastes and preferences."*(Rosen 2001) A tool like the one that we are proposing changes the visibility of personal information in its social neighborhood. If deployed and accepted by the users, the tool would certainly affect the way in which people communicate within the network, since it could possibly portray the users in an awkward light. One way to deal with this issue is to only visualize the structure and temporality while hiding the actual comments. In order to read them, the user has to navigate to the specific profile, where they are displayed in their original context.

Usage Scenario

Many users made the experience of accepting friendship from a stranger and ending up with hundreds of spam messages on their comment boards. Naturally, before accepting an offer, one wants to know whether this person is a real person or a spam bot. Our tool can be used to visually assess communication behavior. Looking at the comment sections, three parameters seem especially meaningful:

1. The temporality of the network - the age of the messages, the frequency of communication. Is a profile constantly updated?
2. One vs. two-way communication - is it a conversation or is it one way broadcasting? Could this provide clues whether people really know each other?
3. Quantity of information - is it a one-time greeting of a newly added friend or actually a conversation?

The temporality is expressed through the opacity of the nodes. Based on the age of the last message posted by a specific profile, its transpar-

ency is increased until becoming invisible if older than an adjustable threshold age.

The directionality can be determined through the way how the participating profiles are visually grouped together by the layout algorithm: nodes with one directional communication are clustered around the central node to which they are connected, while nodes with bidirectional exchange are unconstrained.

These two properties combined make characteristic communication patterns visible. A cluster of other profiles that did not receive a reply to their message surrounds each profile. The amount of unidirectional comments is a first hint at the type of the profile: certain profile types, such as band or celebrity profiles have comment sections usually have a very high rate of incoming comments that remain unanswered. Personal profiles tend to show a lower rate of incoming comments, but these are more frequently answered and often grow into an ongoing conversation. Finally, fake profiles used for commercial purposes usually uniformly broadcast a large number of messages into their network, without receiving any reply. With the visual output of our tool, it is possible to get an idea about the posting behavior of an unknown profile and its activity within its proximate neighborhood. For a more detailed view of the interaction between profiles the marks along the edges can be examined.

The quantity of information is obvious through the marks along the edge, each of them representing a single comment. If desired, they can be animated along the edge in order to clarify their direction in a very dense network. By zooming in and displaying the actual text can follow conversations, otherwise spread across multiple profiles, in the same space.

Design and Implementation

The Comment Flow software was specially designed for the interactive manipulation of very large networks. The tool is designed to be generic, not limited to the application on myspace. Comment flow was written in java, utilizing the JOGL library for OpenGL accelerated graphics. It uses a force directed layout algorithm, a simplified and computationally less expensive version of the Fruchterman/ Rheingold spring-embedder (Fruchterman and Reingold 1991). It uses a clustering approach that discriminates directed and undirected edges, in our case corresponding to the directionality of exchange.

The network can be explored both in 2d and 3d space. Zoom, pan and orbit (in 3d) are the main methods of navigation; the view can be automatically centered on a node selected from a list or from the visible area. Large network structures can be partially rebuilt. To support visual identification of profiles, the profile pictures are extracted and displayed

on the node. Comments are displayed as visual elements with text labels and distributed along the connecting edge. To suggest the direction of the exchanged, they are placed on the same side of edge as the asymmetrical arrowhead. In order to resolve any confusion about the flow direction, the comments can be animated along the edge. The neighborhood of the selected node is highlighted through a color gradient based on the topological distance, facilitating the visual identification of personal networks. The transparency of a node indicates the age of the last activity, helping to identify the most active parts of the network.

Data acquisition in Comment Flow usually starts with a single profile, continued by a successive expansion of the network by extracting the comment network from the code of the profile page. For each identified profile a new node is created and the exchange of comments in both directions determined. A single profile page shows typically up to 50 comments, the number of new nodes increases exponentially with each successive step. At distance 4 from the original node the network usually has more than 15000 nodes. However, this network is very sparse with an edge count only slightly higher than the number of nodes, omitting many interconnections that might exist between the nodes.

Conclusion

This paper presents a visualization system for the representation of communication within social networking sites. The visualizations generated by "Comment Flow" serve as a meaningful map of the social activities within online social networks. It creates a more differentiated picture of individual relationships than the list of uniform "friendships" offered by social network platforms. We believe that this is a step towards the next generation of social network visualization, where nodes and links are highly differentiated and associated with diverse media formats. It shows a viable path to the general problem of representation of online identity and its context in social network visualization.

References

- Aaron Zinman, J. D. (2007). "is Britney Spears Spam?" Proceedings of Fourth Conference on Email and Anti-Spam, Mountain View, California.
- Adar, E. (2006). "GUESS: a language and interface for graph exploration." Proceedings of the SIGCHI conference on Human Factors in computing systems: 791-800.
- Amanda Lenhart, M. M. (2007). "Teens, Privacy & Online Social Networks." Pew Internet & American Life Project.
- Barabási, A. L. and R. E. Crandall (2002). Linked: The New Science of Networks, Perseus Publishing.

- Donath, J. and D. Boyd (2004). "Public Displays of Connection." BT Technology Journal **22**(4): 71-82.
- Feld, S. L. (1981). "The Focused Organization of Social Ties." American Journal of Sociology **86**(5): 1015.
- Freeman, L. C. (2000). "Visualizing Social Networks." Journal of Social Structure **1**(1): 4.
- Fruchterman, T. M. J. and E. M. Reingold (1991). "Graph Drawing by Force-directed Placement." Software- Practice and Experience **21**(11): 1129-1164.
- Heer, J. and D. Boyd (2005). "Vizster: Visualizing Online Social Networks." InfoVis 2005 IEEE Symposium on Information Visualization.
- O'Madadhain, J., D. Fisher, et al. The JUNG (Java Universal Network/Graph) Framework.
- Offenhuber, D. and G. Dirmoser. "SemaSpace", <http://residence.aec.at/didi/flweb>.
- Rosen, J. (2001). The Unwanted Gaze: The Destruction of Privacy in America, Alfred A. Knopf.
- Roush, W. (2006). "Fakesters: On MySpace, you can be friends with Burger King. This is social networking?" TECHNOLOGY REVIEW-MANCHESTER NH **109**(5): 72.
- Shapiro, A. (2002). "TouchGraph Project." Sourceforge. Net Open Source Repository.

Appendix: Figures

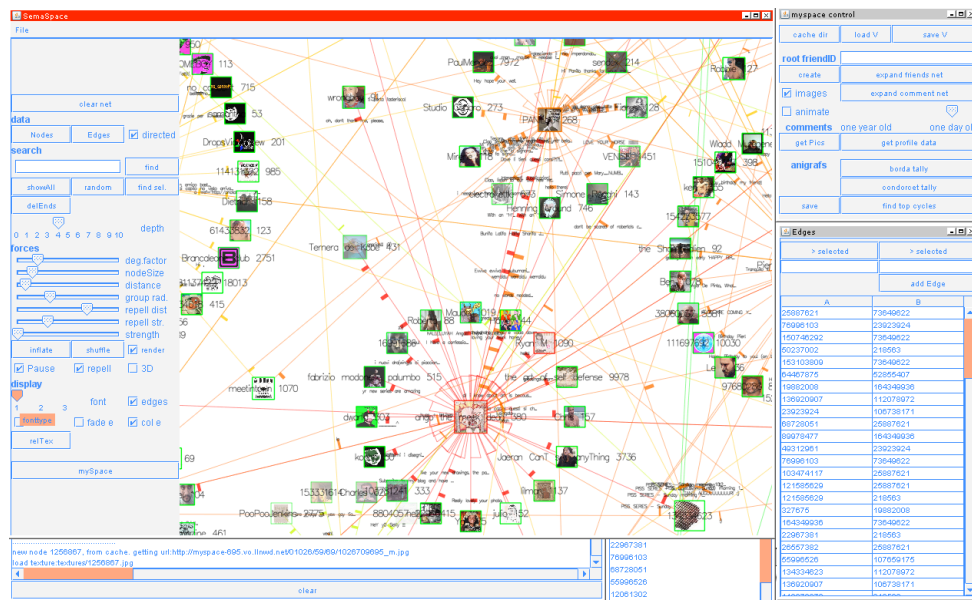


Figure 1: Screenshot of the Comment Flow application

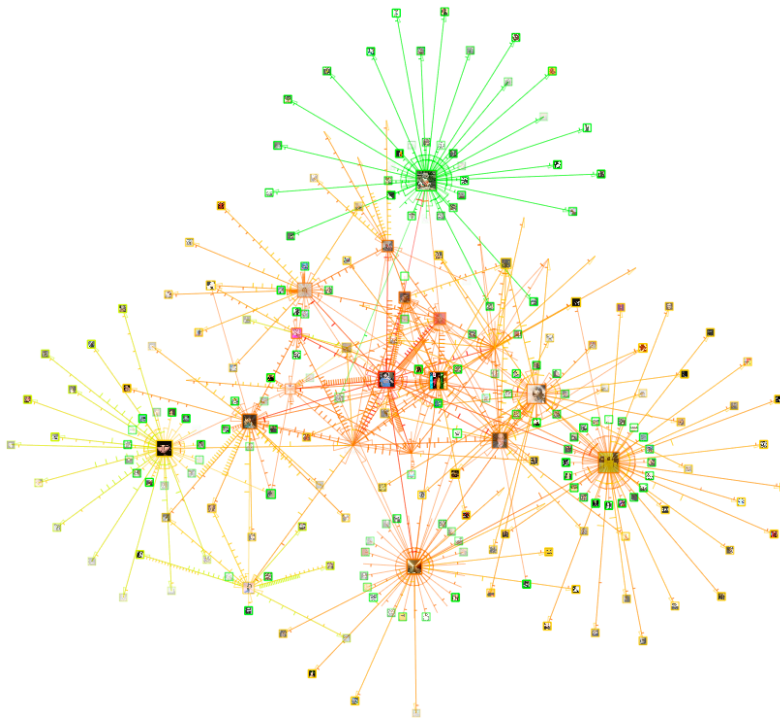


Figure 2: A distance 3 network. The edge colors indicate the topological distance from the selected node, visually supporting the interactive exploration of the network.

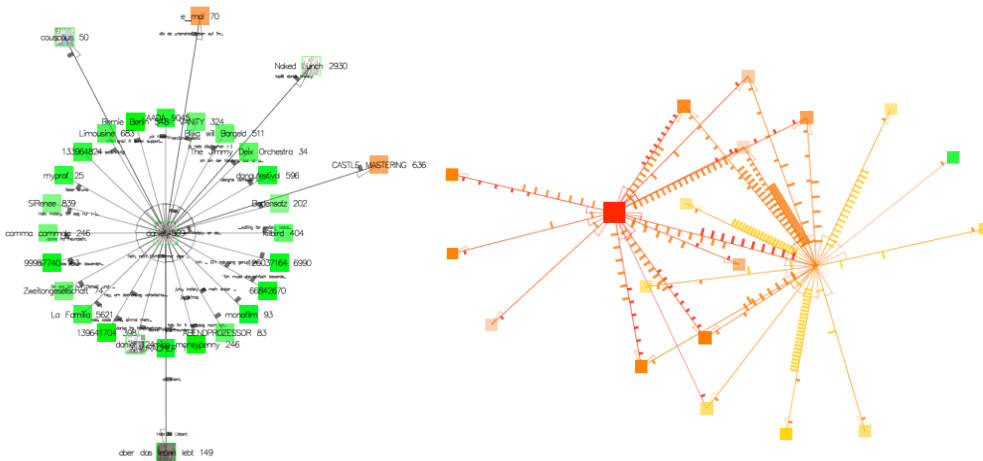


Figure 3: The visual pattern of one-way and two way message exchange. On the left side we see a band profile with a number of incoming posts, but very few replies. On the right side we see a group of profiles using the comment section for extensive conversation. In our application, leaf nodes are grouped around their connected node and appear as circular clusters. Applied to the myspace data, the clustered nodes represent such instances of one-way communication.

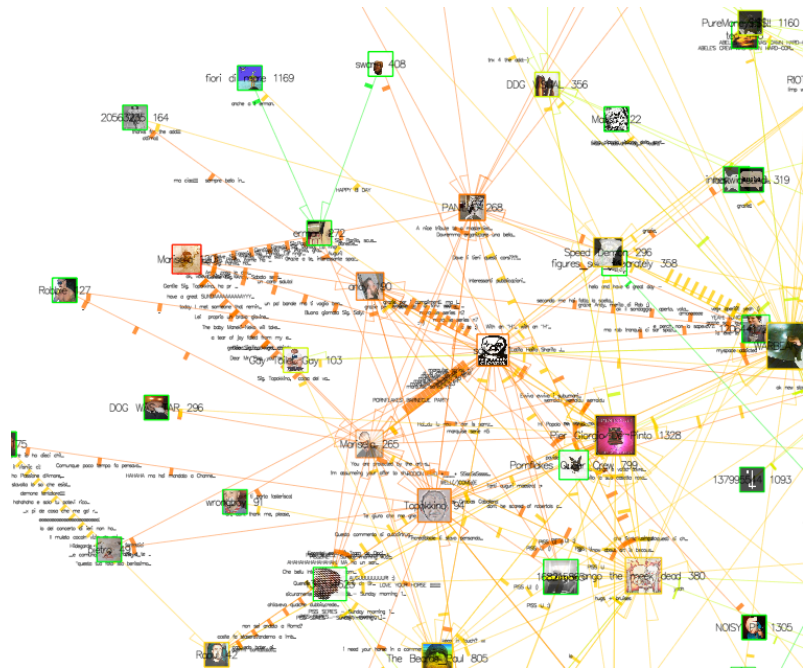


Figure 4: Closeup, each mark along the edge corresponds to a message posted in the comment space of a profile.

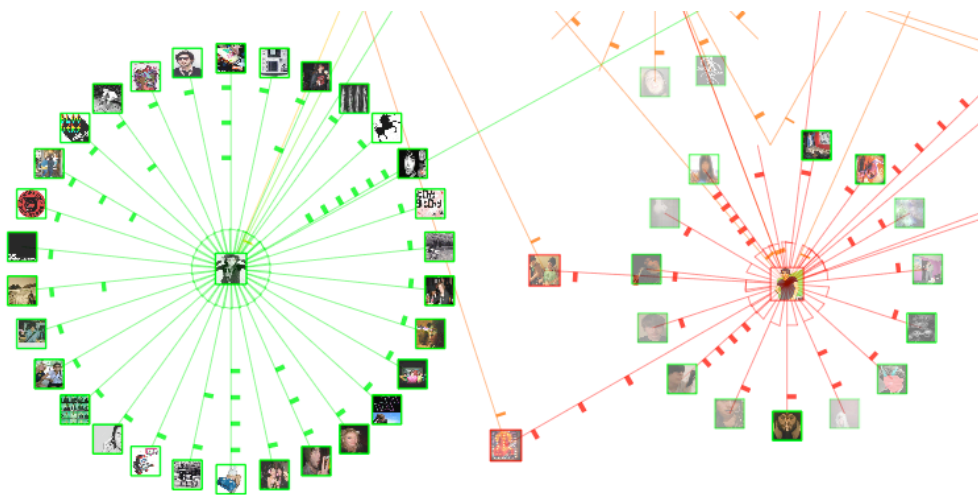


Figure 5: The age of the most recent message sent by a profile is indicated through the opacity of the corresponding node