

# Twitviz - Exploring Twitter network for your interests

**Denzil Ferreira**

University of Madeira  
Department of Mathematics and Engineering  
lizned.arieref@gmail.com  
twitter.com/liznedarieref

**João Rodrigues**

Carnegie Mellon University  
Human-Computer Interaction Institute  
joao.pedro.rodrigues@gmail.com  
twitter.com/jprod

**Maria Freitas**

University of Madeira  
Department of Mathematics and Engineering  
carmofreitas@gmail.com  
twitter.com/omrac

**Vitor Ferreira**

University of Madeira  
Department of Mathematics and Engineering  
vitor.dominor@gmail.com  
twitter.com/archimast

## ABSTRACT

In this paper we explore a tool that allows you to explore your social network on Twitter. Twitviz, is a desktop application that allows users to explore their social networks on Twitter based on their interests. Twitviz provides two main tools in the form social network graphs: the Keyword view displays relationships between keywords of interest and users tweeting about them; and the user's egocentric Network view that emphasizes on reciprocal followers. We believe these visualizations provide an easy and useful way of identifying both known and unknown users with related interests, thereby promoting expansion and specialization of their social network.

## Author Keywords

twitter, keywords, interests, visualization, awareness, social, networking.

## INTRODUCTION

Recent years have witnessed the impacts of distributed content sharing (Wikipedia, Blogger), social networks (Facebook, MySpace, Twitter), sensor networks and pervasive computing[2]. By participating on social networks, we have the potential to learn more about yourselves and others, with the help of the right tools to do it[1,2]. Nowadays, many people can no longer imagine a life without such applications[3].

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

UMa 2009, June 19, 2009, Funchal, Madeira, Portugal.

But social network studies go back to 1930's[4], where these were employed as an analytical tool in sociological research. Resorting to color, spatial position, size and shape, huge quantities of information were laid out for better understanding.

Visualization techniques have been used for researchers in different areas: social network analysis, chemistry, biomedical/genomics research, genealogies, internet networks, citation networks, diffusion networks, analysis of texts, data-mining[5].

We can reach everyone everywhere all the time, instantly. As a result, the differences between personal and remote communication patterns are diminishing[3]. So, the 1:1 conversation scheme known from traditional remote communication will more and more be complemented with group interaction, much as when socializing in the real world. For efficient group communication, however, an N:N solution is required. Social networks should facilitate group communication and offer support for collaboration[1,2,3].

One aspect that usually is neglected by social networking services is the lack of support for initial group formation process[3]. Usually they rely on already existing social networks databases and connect the new user to the same users they have already connected on other social networks.

Twitter in particular, provides poor support for discovery of communities of interest beyond following a public user that represents a company, event or group of interest. Ad-hoc discovery of topics of interest and their followers is limited to the keyword searching tools. Users have appropriated the "hashtags" for keyword searching as a way to categorize their tweets and share their interests to their community of followers

Motivating social networks is facilitating the discovery of people, connections and communities to promote increased awareness of community structure and information exposure, while preserving and enhancing the user social capabilities.

Social network exploration is usually done by massive graph structures of social connectivity and typical web interfaces to these social networks remain relatively impoverished, showing only the network connections of users in a linear list on a webpage.

Twitviz helps the user on this initial group formation, not by presenting him with the users he already knows, but by allowing him to explore new people that share the same interests.

## TWITTER

Even with all means of communication available to us nowadays, we still wonder what someone is doing.

Whether we need to tell a friend or group of friends what we're doing or where we are, whether we want to know what someone is doing and where he is, whether we want to reach out to a diversity of opinions on a subject or just say something we read, thought or experienced, to different people around the world, twitter makes it possible in an easy way.

Twitter is a free social networking and micro-blogging service that enables its users to exchange short messages, called tweets. This can be done through the twitter website or through its widgets and mobile applications. The connections made on Twitter are based on Following other users or seeing their private posts and on having Followers or others that see your private posts.



**Fig 1. A Twitter profile page**

Users can choose to make their tweets public or available to any Twitter member or private, being visible only to their followers.

Since its creation in 2006 by Jack Dorsey, Twitter has gained an incredible and unexpected worldwide notability and popularity. The several news posts and users comments support this:

*Twitter is the first thing on the web that I've been excited about in ages.* - Jason Kottke, Blogger

*Twitter is the telegraph system of Web 2.0.* - Nicholas Carr, Author and Technologist

*Incredibly useful* - wired

*Twitter is on its way to becoming the next killer app.* - TIME Magazine

*Suddenly, it seems as though all the world's a-twitter.* - Newsweek

*When I first started doing it, I thought, 'geez, not another website to worry about updating and checking', but now I'm glad I did it.* - point\_chevalier, livejournaler

In 2009, Twitter won a Webby Award in the "Breakout of the Year" category. On February 2009, Compete.com blog entry, ranked twitter as the third most used social network, with a number of 6 million monthly visitors and 55 million visits. In March 2009, Nielsen.com blog ranked Twitter as the fastest-growing site in the Member Communities category for February 2009. Twitter had a monthly growth of 1382%, next to the 240%, growth of Zimbio, followed by Facebook with an increase of 228%. However, according to this source, Twitter has a 40% retention rate of users.

## Micro-Blogging

Micro blogging is a form of multimedia blogging that allows users to send brief text updates or micromedia such as photos or audio clips and publish them, either to be viewed by anyone or by a restricted group which can be chosen by the user. These messages can be submitted by a variety of means, including text messaging, instant messaging, email, digital audio or the web. The content of a micro-blog differs from a traditional blog in that it is typically smaller in actual size and aggregate file size. A single entry could consist of a single sentence or fragment or an image or a brief, ten second video. But, still, its purpose is similar to that of a traditional blog[10].

Although not as popular as Twitter, there are other micro-blogging services, that followed Twitter approach: Jaiku, Tumblr, Beemood, Toorl, Second Brain within others. The similarities between these websites and Twitter can be clearly identified.



Fig 2. Micro-Blogs

### Why and What people tweet?

A university of Maryland study published in 2007 [7] captured 1,348, 543 tweets from 76,177 members over a two-month period. The researchers analyzed both the content and network structure of their sample. One of the outcomes was the identification of four common user intentions prompting members to use Twitter: *Daily, Chatter, Conversations, Sharing Information, and Reporting News.*

Twitter saw its initial purpose of sharing “What am I doing” statements, diverge to several other ones, such as alerting to events (conferences, birthdays, embrace causes), request for information, political campaigning, polls, games, social support, creative expression, short conversations, commenting on news, commenting on reading materials and conferences. With an open API, Twitter engages users on its growth, improvement and diversification.

The reasons that support such an increasing adoption of Twitter and that distinguishes it from other social websites, or regular blogging, seems to be its simplicity of use due to the informal and short amount of information exchange and the faster mode of communication. By encouraging shorter posts, it lowers users requirements of time and thought investment for content generation. Emotionally, people seem to want to achieve a level of cyberspace presence, to feel another layer of connection with friends and the world [8].

It seems that users get really addicted to Twitter or do not use it and or even do not understand it at all. Some consider Twitter the next-generation social network with the potential to replace other social network sites, including the

mighty Facebook, yet others see it as a supplement to these other services[9].

Organizations use Twitter to advertise their news and products as well as providing a mean of publicity and customer service. Twitter acts almost as a replacement for the RSS feeds.

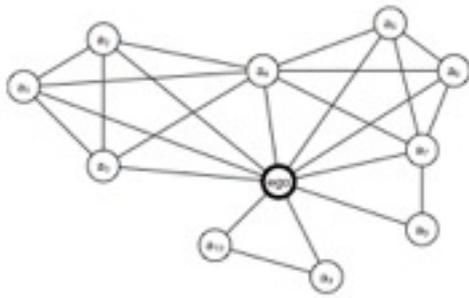
### PREVIOUS WORK

Exploring social networks is nothing new[1,2,3]. The social networks change according to what’s being more used today and the tools users use to interact with the social network.

Miluzzo et al[1] argues that the most common text messages people send each other today are “where are you?”, followed by “what are you doing?”. Their work with CenceMe relies on using sensing components (e.g., accelerometers, GPS, proximity sensors, microphone, camera, etc) to find a new way to answer these questions. The availability of these components allow the rapid development of people-centric sensing applications and systems. Visualizing a social networking when using a mobile device places challenges of screen display space, processing power for relationship data mining, energy management, connectivity to the network and storage of information. CenceMe allowed the users to share their location, photos and text-messaging but not visualize the social network they belong.

Ganekar et al.[2] suggests that their platform of Micro-Blog can be a deployable tool for sharing, browsing, and querying global information. By allowing the users to query for information using their phones (questioners) and getting answers from other users (answerers), they soon develop relationships that are location based. The answerers share their location, allowing the questioner to visualize on Google Maps from where the answers come. Their visualization shows Social Collaboration by connecting those who need services to those who can offer them.

Grob R. et al[3] presents a community-aware mechanism that allows to efficiently select contacts in order to address them as a group. Arguing that mobile versions of online social networks do not exploit the full potential of the underlying technology, they propose a desktop utility (Cluestr) that enhances communication among the members of a group. The visualization chosen to facilitate the group communication was ego-graphs.



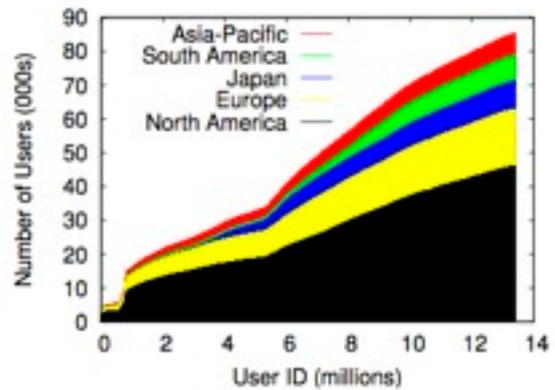
**Fig 3. Sample ego-graph**

A user's ego-graph can contain a significant amount of community information, depending on what the size of the nodes mean, the distance between the nodes, the thickness of the edges and coloring applied.

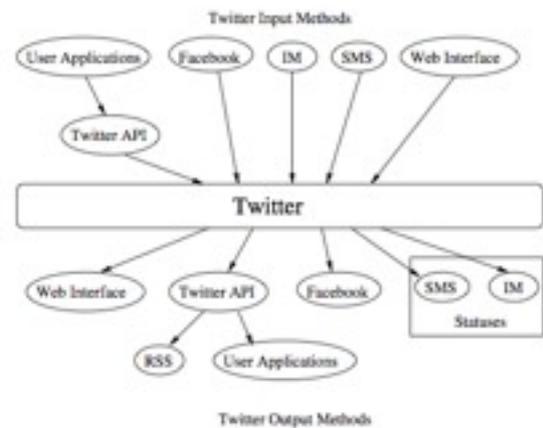
Heer J. et al[4] Vizster uses ethnographic research using Friendster as social network. He employs ego-graphs to explore articulated connections between a user and his friends. This also allows us to get higher level patterns of community. They follow the philosophy of "start with what you know, then grow".

Batagelj et al[5] developed Pajek, a framework whose tools are used for analysis and visualization of social networks. This tool has been used for visualizing marriage relationships, network connections, dictionary relationships, normalization of news networks and block-modeling. As Danziger, M.[6] claims, most visualizations today are mostly made towards scientific community. Regular users should be able to explore the data they produce. By providing accessible social visualizations, users can become aware of behaviors, capabilities and influence they produce or induce towards other users.

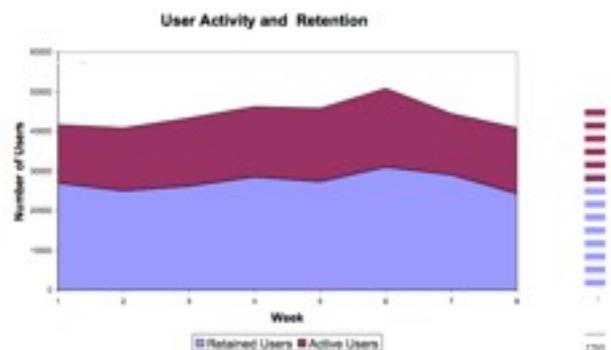
In accordance to Krishnamurthy et al.[11] studies, higher popular users (those who have many followers) update their status more often than those who (likely passively) update their status more often than those who (likely passively) follow more users. Users who have more than 250 followers send many more status updates than those who follow more than 250 users.



**Fig 4. Twitter users growth by geography**



**Fig 5. Twitter input and output methods**



**Fig 6. User activity and Retention**

A user is considered active during a week if he or she has posted at least one post during that week and an active user is considered retained for the given week, if he or she reposts at least once in the following X weeks.

## TWITVIZ

### Design goals

Given the previous considerations, the goal of our application is to promote an increased user's awareness of one's social network, in this case, Twitter. Concretely, we aim to provide the users of Twitter a social networking Twitter client, which they can readily use to expand their network of people with common interests (friends). Through this perspective, they are building their social network, based on Twitter.

Consequently, our Twitter client focuses on a limited set of tasks, performed in the context of the larger activity of building their social network. Their role while facing the application coincides with their role when accessing Twitter through their web interface: a user with an individual account to access the Twitter service and post small and quick messages or updates to anyone who is interested.

We based our design of the user interface of the application on supporting the aforementioned activity and related tasks and the user's role towards the application.



Fig 8. Home panel



Fig 8. Home panel

Relating to the user's role as a Twitter user, he/she can quickly post updates or messages to the Twitter service, as well as seeing the latest 20 updates from friends or users he/she follows, through the home area (Fig 8).



Fig 9. Information area

He/she can also decide to follow/unfollow another user's updates through the information area (Fig 9). This is the part of the application which makes it a basic Twitter client, though it does not support the communication facilities available through the Twitter service (direct messaging).

In order to support the aforementioned activity and related tasks, the application UI possesses two other areas at the center and the left (Fig 8,9): the editable keyword list and the visualization tabs. These facilities support our simple and basic approach towards allowing the user to expand his social network of friends. We have considered that one criterion people usually resort to finding interesting friends and acquaintances are common interests. For example, when two people are dating, one simple way of establishing a connection is sharing interests and activities and thereby allowing each of them to find a common ground and relate to each other.

### Design approaches

The user supplies his interests in the form of simple, representative keywords through the keyword list and then the application queries the Twitter service for updates containing some keyword. From those updates, the application retrieves the originating Twitter users and

builds a connection between the keyword in question and the retrieved users.

The way we present this information and also social network information to the user is closely similar to the approach Heer and Boyd[4] follow in Vizster, towards visualizing online social networks, that is, a visual approach. Both the relationships between keywords and related users and the social network the user is present in are presented through visual graphs (similar to data structures of the same name). In both of them, the user “discovers” each potential friend by clicking on the node representing him. Then the relevant potential friend’s details are presented in the information area, where the user also has the option to follow him, thereby expanding his social network. Both the graphs are effectively node-link representations, where nodes represent keywords (central nodes) or users (peripheral nodes) and links represent the relationship between them. In the case of the social network graph, every node represents a Twitter user and the links represent both friendship and following relationships. However this graph takes an egocentric approach, meaning the user logged on to the application is at the center and each additional level represents a farther level of friendship. This allows for a personal perspective, thereby allowing the user to start his exploration from a familiar, safe place. We would like to add that the social networking graph allows for an additional approach: the user may visually discover new friends by discovering friends of friends or people who possess the same interests, but of whom he/she does not have knowledge of.

### **Visualization design**

The layout of both networks is determined by using a force-directed algorithm, in which graph elements are positioned based on a physics simulation of interacting forces: nodes repel each other and are kept together by springs which are the edges.

The visualization supports different manipulation techniques to efficiently and easily interact with and navigate it. Basic interaction is done through simple mouse operations. As mentioned before, clicking on a node makes the relevant potential friend’s details appear on the right. The user can also move a node by the traditional clicking-and-dragging mechanism. Upon release, it is repositioned by application of the above algorithm.

The user can navigate the visualization space by both manual and automated panning and zooming. Panning is performed by dragging the background of the display with the left mouse button. Manual zooming is performed by most traditional methods like the middle scrolling mouse button or holding down the left mouse button down.

### **Architectural/Implementation considerations**

Regarding implementation decisions, we chose to implement the application using the Java platform, not only the language, but also other included technologies like the Swing framework for a quick and simple UI design, while

supporting standard GUI conventions and mechanisms. The Java platform itself has as a main advantage its inherent portability, allowing the application to be executed in several platforms, including mobile ones (although the application would have to be further optimized, in terms of performance and resource consumption).

The application is built upon two major APIs/frameworks: Prefuse and Twitter4J. Prefuse allowed for a powerful and yet simple visualization design (the mechanisms described before are available by default). Twitter4J encapsulated the Twitter HTTP API in a JAVA API, dispensing us from worrying with several low-level details related to the HTTP protocol and request/response parsing.

All core visualization data is recorded and obtained from files, using the GraphML format, which is an XML-based file format specification widely used for graph visualizations. This feature is also available through the prefuse toolkit.

We should note that our application does not take into account all user profiles present on Twitter, seeing as the Twitter service hides some of them, for privacy reasons, if the owning user so specifies.

The querying of Twitter users connected with user-specified keywords is done through the Twitter Search API, which supports more general and complex searching queries than the keyword-based query we take in our application.

### **Limitations**

Unfortunately, our application has some restrictive limitations.

Firstly, the Twitter service limits the number of requests (calls) any given IP or account on behalf of a user can make, because of performance reasons. This, in turn, limits the expansion of both graphs. We request to be whitelisted for development purposes, giving as many requests as 20000 per hour instead of the default 100 per hour. Nevertheless, for end-users, the same solution does not apply.

Secondly, without using a dedicated GraphML generator, the application must rely on pre-existing and pre-formatted XML files.

Thirdly, the removal of nodes is not a simple operation, using the embedded data and visual structures in the prefuse toolkit. While traversing the nodes of the graph, using for example an iterator provided by those data structures, if one node is removed, the entire collection is rearranged in such a way that illegal access operations occur. We were unable to find a workaround and consequently, the application is broken regarding node removal.

Lastly, since this application aims to be a proof-of-concept for the goals stated at the beginning, it has not been thoroughly optimized and it might be to some degree memory-intensive, as the graphs grow in complexity and dimension. Another factor which adds to this fact is the

choice of the Java programming language, because it requires the application to be run on top of a virtual machine.

Content twittered can be generated without centralized oversight, and their content can reflect (intentionally or not) inaccurate information. The word matching using regular expressions for acquiring twitters of interest doesn't take into account twitter text semantics. For example, we could be interested in fruit apple and get results from Apple Inc.

Another limitation of using Twitter API, is its lack of support for filtering language, meaning that we might get a lot of false positives, specially if we are interested in a global word, like jQuery and get results from people in Chinese.

## CONCLUSION

We started Twitviz with the objective of allowing users to establish new connections with complete strangers. Another objective we had was to allow the user to know exactly how many users of the Following list were following the user back. Because Twitviz is also a client of Twitter, we wanted to allow the user to update his status and see what his followed users have been twitting.

After the user successful login, using his Following list and reverse checking if the followed user has the user on his Following list, we can determine if the connection between both users is bidirectional. By coloring the nodes and edges of the visualization with green and yellow, users can distinguish followed users that follow them back. Information about the user is gathered upon clicking on the followed user, giving the possibility to the user to cancel his following status.

This simple visualization gives a visual interpretation on how the users connections are established.

If we had implemented a relationship historical database, we could highlight the users that had stopped/started following us, since last login.

Another approach for differentiating the bidirectional connections from unidirectional could be by using the Prefuse Grouping API. This could be used to visually group nodes together.

To reduce the number of API calls done, we could have allowed the user to add a new friend by his username directly, instead of first getting information about the user. But, since we wanted to less error prone, we let the user first check if he selected the right user.

By allowing people to search and save bookmarks of words mentioned in tweets, people may find other Twitter users that they are more likely to follow. This allows people to increase their networks and relationships and allows the identification of networks within users and their tweets content.

Since we keep a history of twitted text from public line and search results, every time we add a new keyword, we check

in the history if the new keyword is matched by someone and establish a connection between the user and the keywords.

As we can't delete keywords (which are nodes on a GraphML file), due to a restriction on Prefuse GraphML API, we should have used a relational database for storing keywords and users. This would have allowed us to do faster queries on the nodes, reducing the amount of memory and file I/O operations Twitviz requires.

## FURTHER WORK

There is still room for improvement in our social network visualization.

- On the keyword and network visualization, the selected user should be highlighted.
- The background color of the feedback messages should display different color according to the importance of the message, light green for successful actions and light red for errors.
- Feedback that the application is actually doing something when we wait for the response on Twitter's API.
- Although we can zoom in and out using the scroll of the mouse, we should provide a button that allows the same thing.
- Provide a toggle button for displaying or hiding the labels.
- The Twitter Search API sometimes returns results that do not relate to the typed keyword. This is a problem of the Twitter API.
- Twitviz is dependent on the rate limits imposed by Twitter. In order to increase the rate limits, the user should request Twitter to whitelist his account to better explore Twitviz.
- Removing keywords from the graph does not work properly. Prefuse GraphML API doesn't support the removal of parent nodes. Every time we remove a child node, the node index count gets changed, meaning that the parent index gets changed at runtime, causing an exception. An alternative to GraphML files should be considered, like using a relational database as MySQL.
- Every time we start the application, it should run an update on the keyword results, so we can have recent results from the keyword list.
- A ratio of the people we follow against those that follow you back.

Another visualization idea is, to merge both visualizations, meaning that the strangers that appear as a result of a searched keyword, if they are related with any of our Following list, they should be connected to our friend.

## REFERENCES

1. Miluzzo, E; Lane, N.; Fodor, K.; Peterson, R.; Lu,H.; Musolesi, M.; Eisenman, S.; Zheng, X.; Campbell, A.: Sensing Meets Mobile Social Networks: The Design,

- Implementation and Evaluation of the CenceMe Application. *SenSys'08, November 5-7, 2008, pp. 337-350.*
2. Gaonkar S.; Li, J.; Choudhury, R.; Cox, L.; Schmidt A.: Micro-Blog: Sharing and Querying Content Through Mobile Phones and Social Participation. *MobiSys'08, June 17-20, 2008, pp. 174-186.*
  3. Grob, R.; Kuhn, M.; Wattenhofer,R; Wirz, M.: Cluestr: Mobile Social Networking for Enhanced Group Communication. *GROUP'09, May 10-13, 2009, pp. 81-90.*
  4. Heer, J.; Boyd,D.; Vizster: Visualizing Online Social Networks.
  5. Batagelj, V.; Mrvar, A.: Pajek - Analysis and visualization of large networks. *March 4, 2003.*
  6. Danziger, M.: Information Visualization for the People. *Boston University, 2008*
  7. Hazlewood, W.; Makice, K.; Ryan, W. ; Twitterspace: A Co-developed Display using Twitter to Enhance Community Awareness
  8. Zhao, D.; Rosson, M.; How and Why People Twitter: The Role that Micro-blogging Plays in Informal Communication
  9. Zeichick, A.; A-Twitter over twitter
  10. <http://en.wikipedia.org/wiki/Micro-blogging>
  11. Krishnamurthy, B.; Gill, P.; Arlitt, M.; A Few Chirps About Twitter