PROVIDING AWARENESS, EXPLANATION AND CONTROL OF PERSONALIZED STREAM FILTERING IN A P2P SOCIAL NETWORK

A Thesis Submitted to the College of

Graduate Studies and Research

In Partial Fulfillment of the Requirements

For the Degree of Master of Science

In the Department of Computer Science

University of Saskatchewan

Saskatoon

By

Sayooran Nagulendra

PERMISSION TO USE

In presenting this thesis in partial fulfilment of the requirements for a Postgraduate degree from the University of Saskatchewan, I agree that the Libraries of this University may make it freely available for inspection. I further agree that permission for copying of this thesis in any manner, in whole or in part, for scholarly purposes may be granted by the professor or professors who supervised my thesis work or, in their absence, by the Head of the Department or the Dean of the College in which my thesis work was done. It is understood that any copying or publication or use of this thesis or parts thereof for financial gain shall not be allowed without my written permission. It is also understood that due recognition shall be given to me and to the University of Saskatchewan in any scholarly use which may be made of any material in my thesis.

Requests for permission to copy or to make other use of material in this thesis in whole or part should be addressed to:

Head of the Department of Computer Science

176 Thorvaldson Building

110 Science Place

University of Saskatchewan

Saskatoon, Saskatchewan

Canada

S7N 5C9

ABSTRACT

In Online Social Networks (OSNs), users are often overwhelmed with a huge amount of social data, most of which are irrelevant to their interest. Filtering of the social data stream is the common way to deal with this problem, and it has already been applied by OSNs, such as Facebook and Google+. Unfortunately, personalized filtering leads to “the filter bubble” problem where the user is trapped inside a world within the limited boundaries of her interests and cannot be exposed to any surprising, desirable information. Moreover, these OSNs are black boxes, providing no transparency for the user about how the filtering mechanism decides what is to be shown in the activity stream. As a result, the user trust in the system can decline. This thesis presents an interactive method to visualize the personalized stream filtering in OSNs. The proposed visualization helps to create awareness, explanation, and control of personalized stream filtering to alleviate “the filter bubble” problem and increase the users’ trust in the system. The visualization is implemented in MADMICA – a new privacy-aware decentralized OSN, based on the Friendica P2P protocol, which filters the social updates stream of users based on their interests. The results of three user evaluations are presented in this thesis: small-scale pilot study, qualitative study and large-scale quantitative study with 326 participants. The results of the small-scale study show that the filter bubble visualization makes the users aware of the filtering mechanism, engages them in actions to correct and change it, and as a result, increases the users’ trust in the system. The qualitative study reveals a generally higher proportion of desirable user perceptions for the awareness, explanation and control of the filter bubble provided by the visualization. Moreover, the results of the quantitative study demonstrate that the visualization leads to increased users’ awareness of the filter bubble, understandability of the filtering mechanism and to a feeling of control over the data stream they are seeing.

ACKNOWLEDGEMENTS

First and foremost, I would like to express my sincere gratitude to my research supervisor, Dr. Julita Vassileva, for providing me the opportunity to pursue Master's degree in Computer Science with full scholarship. Her tremendous support, guidance and continuous encouragement, helped me to carry out my research successfully and complete the thesis on time. Next, I would like to thank Dr. Ralph Deters and Dr. Gord McCalla for being part of my thesis committee and providing very constructive feedback during my thesis proposal defense, which was held in December 2013. I would like to further extend my gratitude to my external examiner, Dr. Anh Dinh, for providing valuable support and suggestions. I want to express my thanks to Gwen Lancaster, and other staffs at the department of Computer Science, University of Saskatchewan, for being very helpful throughout my studies.

I would like to thank my loving parents, Arunasalam Nagulendra and Jeyaranjani Nagulendra and my loving sister, Hinduja Nagulendra, for encouraging to pursue my dreams. Last but not least, I would like to thank my loving fiancée, Ilhavanchi Kanaganayagam, for her continuous support and motivation to successfully complete this thesis. I dedicate this thesis to her.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

DFRN	Distributed Friends & Relations Network
DOSN	Decentralized Online Social Network
FOAF	Friend-Of-A-Friend
HIT	Human Intelligent Task
OSN	Online Social Network
RDF	Resource Description Framework
TF-IDF	Term Frequency-Inverse Document Frequency
URI	Universal Resource Indicator

CHAPTER 1
INTRODUCTION

Today, with the enormous growth of Online Social Networks (OSNs) such as Facebook and Google+, millions of users are sharing activities with friends and followers creating an enormous stream of data in real-time. This data vary from personal news (such as what's on their mind, what they are doing, what they are thinking of) to global news (such as news about politics, science, sports, technologies, etc.). If we consider the social data stream of a single user from her friends, only a fraction of it is relevant and interesting, while the rest of the stream results in social data overload to the user. *Personalized stream filtering mechanisms aim at solving these challenges of social data overload by presenting the user with the most relevant content.* Social media sites, such as Facebook, Digg and YouTube, have already implemented personalized stream filtering, which presents the most relevant content to users while reducing the social data overload. However, these systems are black boxes and provide no transparency or explanation, so users do not have any idea about what are the activities that are hidden from the activity stream by the system and why they are hidden. As a result, the user trust in the system can decline. Moreover, as a result of successful personalization of the stream, in the long run, the user can be trapped inside a world within the limited boundaries of her interests. This is called "the filter bubble" problem. There are three key research questions that we aim to answer in this thesis.

1. Can visualization of the filter bubble be used as an effective technique to create user awareness, provide her with explanation and control of personalized stream filtering?

   The main purpose of personalized stream filtering is to reduce the social data overload by presenting only the relevant content. But showing what is hidden and filtered away from the stream can increase the social data overload problem. Therefore, the main challenge is to find an effective visualization technique that can be seamlessly integrated into the activity stream without contributing additionally to the social data overload. What is the right amount of detail to expose in the hidden filtered social data and its explanation? How do we organize these hidden filtered social data? What type of visualization is effective to display the hidden social data stream in an understandable way for the user? These issues can be explored through theoretical design and experiments with users.

2. Can a visualization of personalized stream filtering increase the user's trust in the filtering system?

   There is the possibility that some of the hidden filtered social data are being wrongly classified as undesirable. We believe that showing the hidden filtered social data using an effective method which does not cause the overload or undo the advantages of the filtering, will provide transparency of the personalized stream filtering to the user and increase the user acceptance of the system.

3. Can a visualization of personalized stream filtering alleviate "the filter bubble" problem?

   As the activity stream is personalized according to the user's interests, the user will ultimately only see activities related to her interest and will have no opportunity of discovering new interests. This will lead to "the filter bubble" problem where the user is trapped in a world filled with only items matching her interests. By exposing (some of the) hidden filtered social data, the user will become aware of the model that the system has of her, and may consciously decide to explore items from other areas by changing interactively her model and it will open the avenue for discovering new interests. As a result, the user would be able to come out of her filter bubble.

This thesis presents an interactive method to visualize the personalized stream filtering in Online Social Networks to create awareness, explanation, and control of personalized stream filtering to alleviate “the filter bubble” problem and increase the users` trust in the system.

1.1 Thesis Outline

The remainder of the thesis is structured as follows. A literature survey along with the research problem is presented in Chapter 2. Chapter 3 discusses the design and implementation of a P2P social network (MADMICA) that provides personalized filtering. Chapter 4 discusses the design and implementation of the proposed visualization. A small-scale pilot study to evaluate the visualization approach in MADMICA is described and the results and discussion are presented in Chapter 5. Chapter 6 presents a qualitative study which was carried out in order to understand in-depth the user perception of the filter bubble visualization i.e. what do users think about the visualization. Chapter 7 presents a large-scale quantitative study which was carried out to evaluate whether the users understand that the visualization provides awareness, explanation and control of filtering and the filter bubble. Finally, chapter 8 summarizes the contributions of this thesis, presents the conclusion and outlines directions for future work.

CHAPTER 2
BACKGROUND AND MOTIVATION

Today, social networks provide a global platform for people to share and collaborate with their friends and families. Facebook, Twitter and Google+ are currently the most widely used social networks. With the growth of mobile and web technologies, these social networks are growing rapidly and millions of users are sharing data with their friends and families. As of September 2013, Facebook has 1.15 billion number of users and 699 million daily active users [16]. 24% of the content that is shared on the internet is shared on Facebook [10]. 3.5 billion pieces of content are shared each week on Facebook [1], creating a stream of data that can overload any user. The social data overload problem is commonly solved by filtering out the irrelevant data. However, the filtering mechanisms used by social networks currently provide no transparency or explanation, so users do not have any idea about what social updates are filtered away from the social data stream by the system and why. As a result the user trust in the system can decline. Another problem related to filtering of the social data streams is the so-called "the filter bubble" problem, which can arise in the long run as the user sees only a skewed set of news that fit within the limited boundaries of her interests and misses potentially relevant and interesting news that have been classified by the system as not in the scope of interest of the user. This thesis proposes an interactive method to visualize the personalized stream filtering in Online Social Networks (OSN) to create user awareness, explanation, and control of the personalized filtering of their social data stream to alleviate "the filter bubble" problem and increase the users' trust in the system.

This chapter presents an overview of the related work in the areas of Recommender Systems (RSs), visualization of recommendations, the filter bubble problem and P2P Social Networks.

2.1 Recommender Systems and Information Filtering

Recommender Systems (RSs) are software systems which adapt to the needs of an individual user and provide personalized suggestions of most relevant information [46]. The personalized suggestions help users to make decisions on various types of items, such as what news items are interesting, what book to read, what movie to watch and so on. Information filtering systems can be considered as a type of recommender systems, which select from a stream of data (e.g. news, events, social updates, etc.) those that fit the scope of interest of the user. The difference between filtering and recommendation is that in filtering the irrelevant data are simply not displayed, i.e. remains hidden from the users, while in recommendation the relevant data is highlighted in some way (e.g. shown first in a list of search results, highlighted in a stream of data, etc.), but the irrelevant data is still available for the user to see.

Recommendation techniques have been applied to personalize the streams in online social networks such as Facebook, Google+ and Twitter [28, 30]. Facebook's edge rank algorithm is one of such filtering technique which presents a personalized stream of news and friends' status updates to the user by ranking every interaction on the site [25]. Tandukar & Vassileva [57] developed an interest-based filtering model for a decentralized OSN, which filtered out the irrelevant social data from the activity stream to reduce the social data overload of the user. Search engines also have implemented recommendations techniques to show the most relevant results first. For example, Google's PageRank algorithm is one of the most popular filtering algorithm used to personalize the web search results. The main techniques used in RSs [8] are: collaborative, content-based, demographic and knowledge-based recommendation.

2.1.1 Collaborative Recommendation Systems

Recommendations are generated using only information about rating profiles for different users. Users with a similar rating history as the current user are identified and used for recommending new information [14]. Here the users denote the individuals who provide ratings to items and also denote those who receive the recommendation of items. Items can be anything that can be rated by users such as books, movies, places, social updates, etc. Users rate the items using one of the rating methods such as scalar rating, binary rating and unary rating. When a user rates an item, the rating is stored in a database of historical user ratings to match each individual to others in order to create a "neighbourhood" of users with similar ratings. As discussed above, the concept behind collaborative filtering is the assumption that people with similar interests will rate items similarly, so if they have rated things similarly in the past, they will continue to do so in the future [48]. Thus once a "neighbourhood" of users with similar ratings have been defined, the items rated highly by some of these users in the neighbourhood, but not by others, can be recommended to them. Researchers have used collaborative filtering technique in different domains to build domain-specific recommender systems. GroupLens system is one of the first systems which generated recommendations for news using collaborative filtering [45]. MovieLens is a movie recommender system built using the techniques in GroupLens, to predict movies that the user might be interested in [13]. Another early collaborative system, called Ringo, was developed by Shardanad & Maes [49] to make personalized recommendations for music albums and artists.

2.1.2 Content-based Filtering Systems

Recommendations are generated using the history of interaction and the ratings previously given by one user. It uses the assumption that items with similar objective features will be rated similarly by a given user [48]. For example, if a user likes a webpage, which contains the phrase "social networks", she will like another webpage containing this phrase. An interest profile of a user is compared with item profiles of new items to generate recommendations. Interest profiles comprise mainly two different types of information: **a model of the user's preferences and a history of the user's interactions** with the recommendation system. A model of user preferences is created using the information about the items that the user is interested in and it is formulated as a function which predicts the likelihood of the user's interest in any item. A history of the user's interactions includes the information on the items that the user has viewed, rated and purchased. Moreover, it includes queries that the user has used to search for items he or she is interested in [41]. The history of user's interactions is used to provide some features of the system, such as to display the recently viewed, rated and purchased items, to avoid recommending the item repeatedly.

In addition, the history is also used as training dataset for machine learning algorithm that learns to predict new items that may be liked by the user. Machine learning algorithms for creating user models from history of user's interactions fall under classification learning. Classification learning is a technique used to classify the uncategorized samples into predefined classes [55]. Here the training dataset is categorized into two: items liked by the user and items not liked by the user. So the learning algorithm creates a function out of this training dataset which outputs an estimate of probability for liking an unseen item. There are many learning algorithms that use classification learning technique, such as decision trees, rule induction, nearest neighbor, and so on. Decision tree algorithm builds a decision tree by recursively partitioning the training data, in this case of recommender systems, into subgroups until those subgroups contain only instances of either items liked by the user or items not liked by the user, but not both of them. In rule induction, a set of rules are extracted from training dataset and used to recommend items. Unlike both decision trees and rule induction where training dataset is preprocessed, the nearest neighbor method simply stores all of its training data in memory and classification is done by comparing an unclassified item to all stored items using a similarity function and determines the nearest neighbor.

There are many systems which use content based technique to recommend items in diverse areas of interest. Mooney & Roy [32] developed a content based book recommender system which uses text categorization to recommend books to the users. Unlike other collaborative book recommender system, their system was able to effectively recommend unrated items and provide quality recommendations to users with unique interests. WebMate [11] developed by Chen and Sycara, learns user interest profiles when web searching and browsing, and provides personalized newspaper and search results using TF-IDF (Term Frequency – Inverse Document Frequency) machine learning technique.

Generally, content-based systems are mostly suited for recommending less frequently desired items to the users whereas collaborative systems are suited for recommending well-known items.

2.1.3 Demographic Filtering

In this method, the recommendations are generated based on a demographic profile of the user. The demographic data includes gender, age, ethnicity, knowledge of languages, employment status, and location, etc. Machine learning methods (most often classifications or decision trees) are applied to construct user models which give insight about what type of person likes a particular item. For example, if a particular food item is mostly liked by a particular age group of people, then it can be recommended to others who are in that age group and who do not know about that item. LifeStyle Finder is a recommender system which tries to use demographic data of user profiles to recommend websites and pages to users [27]. It gathers demographic data of user by involving them in a dialog guided by step by step wizard. The dialog contains questions about demographic data of user which is ultimately used to classify users into different life styles. Then based on this classification websites are recommended which provide services suitable to their lifestyles. This can be considered as a collaborative recommender because by matching the user demographic data, similar users are found and then the items are recommended.

2.1.4 Knowledge-based filtering

In this type of recommender systems, products are suggested based on inferences about a user's needs and preferences. One approach for reasoning can be based on the principles of case-based reasoning. In case-based reasoning, previous known cases similar to current case are retrieved and their solutions are applied to solve the current case. For example, the Wasabi Personal Shopper (WPS) works based on the principles of case-based reasoning to recommend items [9]. It provides recommendations and asks the users to examine items and respond with a feedback. The system builds a library of item-feedback cases, and uses this library to classify and recommend new items.

2.2 User Acceptance and Trust of Recommendations

Many researchers have worked on developing new RSs and improving the accuracy of their filtering algorithms. However the ultimate measure of success in this area is the user acceptance and trust of the recommendations and with respect to this measure there is still a lot of work that needs to be done [26]. The standard performance measures for RS focus on measuring the accuracy of the predictions i.e. how well the predicted ratings match actual ratings. But they cannot provide a valid method to test whether recommended data are valuable and previously unknown to the user. Providing a better user experience with RSs can increase user acceptance of recommendations. Recently, improving user experience has become one of the most important current areas of research in RSs. The RSs must adapt and understand the needs of the users at different stages and provide not only valuable recommendations to the users, but also, as proposed by Chen & Pu [43], explanation interfaces which are effective in building the users' trust in the recommendations. Previous research shows that explaining recommendations can increase the transparency of RSs and the users' trust in RSs [23, 63].

2.3 Explanations in Recommender Systems

Explaining the rationale behind the recommendation is an important aspect of recommender systems. Explanations provide users with a mechanism for handling errors that might come with a recommendation and help the users judge the accuracy of recommendations. When we consider how we accept the recommendations provided by other humans, we recognize that other humans are imperfect recommenders. In case of the recommendations suggested by a friend, we might consider the quality of previous recommendations by the friend or we may compare that friend's interests with our interests in the domain. However, if there is any doubt, justification of the recommendation is needed and we let the friend explain it. Then we can analyze the explanation and decide whether to accept the recommendation or not [47].

Tintarev and Masthoff [62] describe three motivations for explanations in recommender systems: (1) transparency, which exposes the underlying algorithm of the recommendation so that the user can trust the system; (2) trust, which enables the user to accept recommendation regardless of its correctness, and (3) scrutability, which enables the user to provide feedback on the recommendation to the system i.e. allow the users to tell the system that the recommendation is correct or incorrect, so that the system can improve the future recommendations (this also means the transparency and openness of the recommendations). Moreover, they also provide criteria which can be used to evaluate explanations in recommenders systems [61]. Bonhard et al.

conducted an experiment with participants simulating a movie recommender system to determine how different explanation parameters influence the usefulness of recommender systems [5]. Herlocker et al. [19], mention some benefits provided by explaining recommendations such as: justification, user involvement, education and acceptance.

Work related to explanations can also be found in many other domains such as psychology, philosophy and cognitive science. Johnson & Johnson [23] have done research on explanations in human-computer interfaces. Previous work on expert systems has also shown that explanations can provide considerable benefit [47]. Incorporating an explanation feature in recommender systems provides several benefits to users. It removes the black box from around the recommender system, and provides transparency.

2.4 Visualization of Recommendations

The way recommendations are presented is critical for the user acceptance of recommender systems. Visualization techniques can be deployed to provide an intuitive "at a glance" explanation for recommendations and can also motivate the user to accept the recommendation. Presenting the recommendations in a ranked list according to their recommendation score is the most simple and commonly used visualization technique.

Features like colour and font-size can be used to emphasize recommended items in a stream or list or items [65]. As shown in Figure 2.1, each item in a news stream has energy units associated with it and the recommended items have different colours and font sizes based on the number of energy units in order to achieve different levels of visibility for users. If units are high then the recommended item will have hot colours and large font size whereas when units are low, cool colours and small font sizes are used.

Figure 2.1: Visual appearance of recommended items in online community [65]

iBlogViz is a system to visualize blog archives. It uses many visual cues to represent the blog content and social interaction history with the blog entry which help to navigate the blog archive quickly and easily. Particularly, visual cues about the social response (comments) to the news can be used to help users navigate stream data quickly to find interesting news [21]. Figure 2.2 shows the visualization used in iBlogViz. In the upper part above the horizontal line, diamonds represent the link to the content of the blog entry and the length of the line indicates the total number of characters in the blog entry. The lower part has circles connected by lines which represent all comments for that entry and the total number of characters in all comments. Moreover, the size of the circle represents the total number of comments for that entry. This enables the users to find the interesting news which have more comments very quickly.

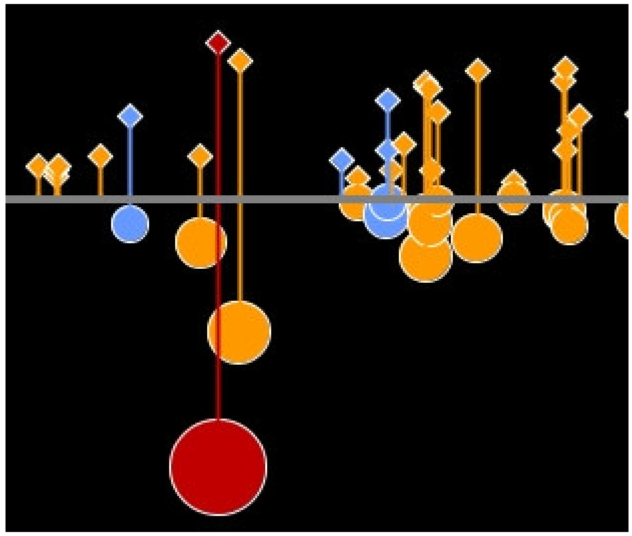


Figure 2.2: Visualization of entries in blog archive [21]

Webster & Vassileva [64] proposed an interactive visualization of a collaborative filtering approach in RSs that allows the user viewer to see the other users in her “neighborhood”, who are similar to her, and also to change manually to degree of influence that any of the other users can have on the recommendations of the viewer. Figure 2.3 shows the visualization of neighbors of user and the relationship with them and how they influence the recommendations made by the system. The current user is represented as a black dot. The inner blue circle shows the influence that the current user has on her neighbors, whereas the yellow region shows the influence of neighbors towards the current user.

Rings is a visualization of social data stream developed by Shi [50]. It helps the users of OSN to browse social data efficiently and find out the active users and the time pattern of their social updates. As shown in Figure 2.4, friends are represented as spirals and the number of social updates is shown using specific colours and sizes. In addition to this, different shades of colours are used to depict the popularity of social data. The position of spirals represents the elapsed time of social data.

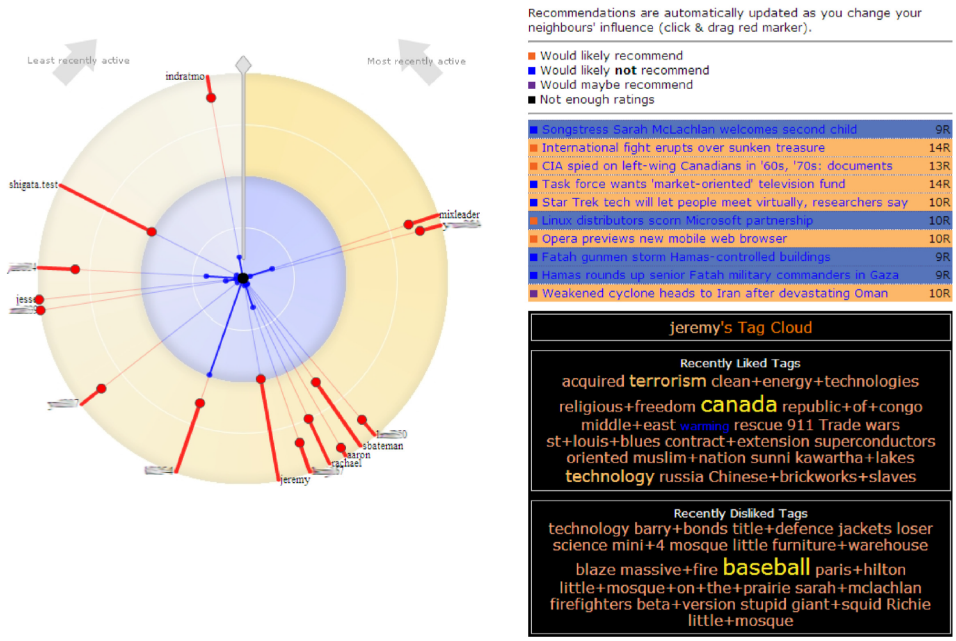


Figure 2.3: KeepUp Recommender Neighbor Visualization [64]

Figure 2.4: Rings Visualization of Social Data Stream, from Shi, Largillier & Vassileva [51]

2.5 The "Filter Bubble" Problem

As the activity stream is personalized according to the user's interests, the user will ultimately only see activities related to her interests and will have no opportunity of discovering items not related to her current interests, or developing new interests. This will lead to "the filter bubble" problem where the user is trapped in a world filled with only items matching her interests. "The filter bubble" is a term introduced by Eli Pariser [40] to denote a limited scope of information defined by the user's interests and isolated from anything that doesn't belong to this scope. This filter bubble can also be referred as "electronic village" – a term defined by McCalla in [31], which constrains the users' perspectives by giving them a partial view of the outside world.

Isolating the user in a filter bubble has its advantages and disadvantages. The main advantage is that it can help users get relevant information a lot faster while not causing social data overload. On the other hand, there are number of problems [40],

- **The problem of distortion:** the user has a distorted view of the content posted on the site or by the user's friends and does not know in what way it is biased. Users become less likely to be recommended information that is important, but not "likeable". For example, in Facebook it is very easy to click "like" for post about "a friend going on a vacation to Hawaii" but there is no "unlike" and users will not click "like" for a news post about an earthquake. So based on the user's feedback, one category of information (good news, e.g. vacations) will quickly become part of the filter bubble, but other important information (bad, but important news) will be filtered away.
- **The information equivalent of obesity:** Because of the users' tendency to give positive feedback, they will give feedback only to information items they are most compulsively attracted to. Using an analogy from food, users will be eating candy all the time, and the filter bubble leave users locked in a world consisting of information junk food.
- **The matter of control:** The growth of user knowledge will be greatly influenced by the algorithms and systems giving excessive power to the computer scientists who develop the personalization techniques. McCalla mentions about the impact of "electronic

village” in learning and teaching which could significantly limit the growth of user knowledge [31].

As an ever increasing proportion of users are using social networks to get any kind of news related information and nearly all OSN deploy information filtering to personalize their streams to users, the impact on how users consume the information and view the world is that it becomes harder for them to come out of their filter bubbles. According to Pariser, as the users are increasingly surrounded by the ideas with which they are already familiar and agree, while being protected from surprising information, or information that challenges their views, the filter bubble threats people’s open-mindedness and prevents learning. Psychologist Lowenstein mentions that the “curiosity is aroused when we are presented with an ‘information gap’” and Pariser suggests that the existence of curiosity, is based on awareness that something exists that is hidden or unknown [40]. Most of the personalization systems do not create awareness about what is being hidden from the user.

Resnick et al. [44] discuss the dangers of isolating users in filter bubbles and outline some strategies for promoting diverse exposure. They discuss two approaches to provide diverse exposure. One approach is to build diversity aware recommender systems and filtering mechanisms. As an example of this approach, Tandukar and Vassileva [56] developed an interest-based stream filtering technique, which allows for diversity exposure by allowing serendipitous important data to pass through the filter. The second approach is to provide tools and techniques that encourage users to consider diverse exposure. Munson has implemented a browser extension which displays the bias in a user’s online news reading over time, which will encourage users to seek the diverse exposure of news [33]. Though these kind of algorithmic approaches certainly find the most relevant content about what we are already interested in a more efficient manner, the human curators will make the actual diverse exposure possible in the system i.e. enabling the users to select what they want to see as well as what they do not want to see over the personalization presented by the algorithms. Facebook uses the edge rank algorithm to personalize the news feed based on likes and comments but what users are willing to like can be a poor measure of what they would actually like to see or what they need to see. For example, even though users do not click on news about Afghanistan much, they need to hear about it because there is an important war going on [59].

Recommendation and filtering techniques are used in any systems that deal with information such as all the social media sites, search engines, e-commerce sites, etc. Among these systems, social media sites are growing rapidly; particularly, Online Social Networks (OSNs) such as Facebook and Google + are having enormous growth rate for the past few years. Moreover, as mentioned previously, Facebook currently has 1.15 billion users and 699 million daily active users sharing 3.5 billion pieces of content each week, and it has influenced people's lives in many ways. [20]. Therefore, this thesis focuses on solving the filter bubble problem in online social networks. Particularly, Decentralized Online Social Networks (DOSNs) are the main target. DSONs have been proposed as solution for many problems found with centralized OSNs such as, they do not allow users much control over how their personal information is disseminated, they are incapable of reciprocal operation with other systems and they are centrally owned by one company. The next section provides an overview of the area of DOSN.

2.6 Decentralized Online Social Networks (DOSNs)

Decentralized OSNs, such as Diaspora or Friendica, have been proposed as an alternative to the currently dominant centralized OSNs, where people are forced to share their data with the central server operated by a company such as Facebook, Google+, etc., and thus lose their control and rights over it [3], [7], [42]. Decentralized OSNs are mostly based on peer-to-peer architectures. Peer to peer (P2P) is a type of distributed networked architecture in which individual nodes in the network act as both client and server. In contrast, centralized OSNs are based on a client/server architecture in which, logically there is only one central server that handles the requests from clients and serves the requested resources. Centralized OSNs store all data in one place logically (though it may be distributed physically) so it has a single point of control and failure. In addition to that, centralized OSNs have scalability issues, which need to be resolved by buying multiple data servers. For example, in the past, with the rapidly growing number of users in social networks such as Facebook and Google+, there have been many issues and challenges faced by the owners related to scalability of the system. In contrast, decentralized OSNs based on P2P architectures store data locally at each node, so they scale naturally with the growing number of nodes (users). In [66], the authors suggest as one of the main benefits of DOSNs that they give back to users the ownership and control of their data, which is a major step forward with respect to ensuring user privacy in information dissemination.

The decentralization, however, imposes constraints on developers of recommender systems for DOSN. For example, certain filtering approaches (e.g. collaborative filtering), are impossible, due to the unavailability of centrally stored user ratings, or traces. Content based recommendations are possible, but the need to delegate some of the personalization process tasks, to the individual nodes (clients) implies more complex software design for the peer-to-peer nodes, for example, the modeling of user interests, the filtering of irrelevant social data, securing the network from spammers, and ensuring that the users are connected [35]. Based on the physical architecture of the network, decentralized OSNs can be divided into two types: users host their social data in a trusted server (cloud) and users host both their social data and system in their own machine like a P2P application using P2P architecture.

In the first type where users hosting their social data on their own machine approach, all the social data of a user such as status updates, photos, contacts, messages, etc. are stored in user's machine which is then accessed via a URI using a secure method. In [66], authors have developed a framework which uses this type of decentralization and a user interface called Tabulator which is a data browser and editor for RDF data on the web. The framework follows FOAF (Friend-Of-A-Friend) standard specification to define the relationships and other social data of a user in a file which is stored on a trusted server (see Figure 2.5). FOAF also provides standard vocabulary which enables interoperability among various other social networks. In addition to that FOAF can be extended by other ontologies to accommodate more rich data about the user. In their framework, each user will have a web identity in the form of a URI which points to a trusted server or user's machine which has the FOAF file of the user. When the user wants to access her friends FOAF files she has to authenticate her identity with each server in order to access the FOAF file. This ensures that only your friends have access to your data. Authors also discuss about how the three popular applications in current social networking sites look in a decentralized version of it. As shown in figure 2.5, Personal Wall, Photo Album and News Feed have been hosted on user's trusted server and access control policies are defined to control who can view and edit them.

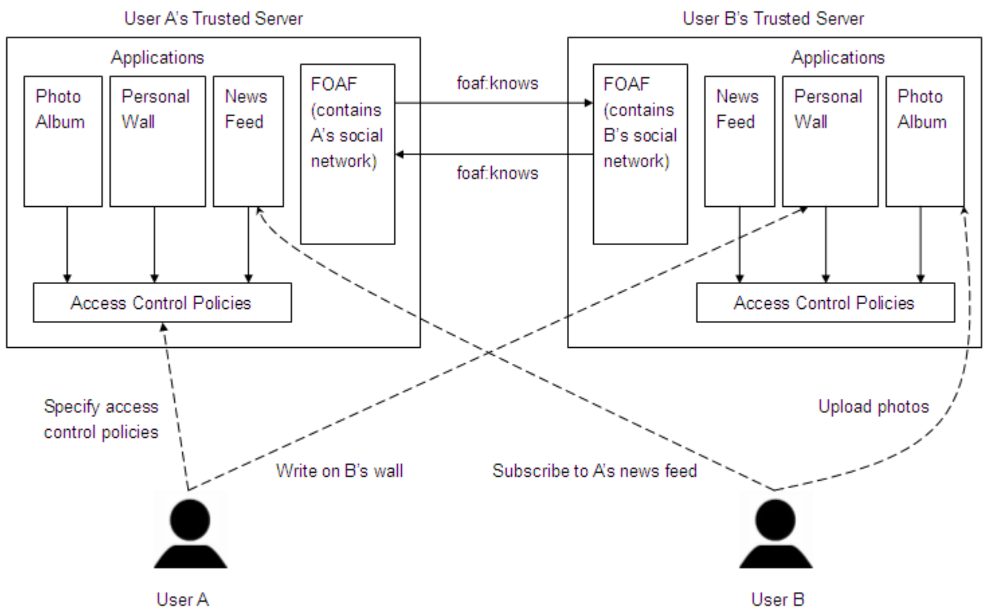


Figure 2.5: A framework of decentralized online social networking [66]

In the second approach, decentralized OSNs are implemented as P2P applications in a P2P architecture. P2P networks have been used in many other applications such as file sharing (e.g. Bit Torrent), messengers (Skype), etc. Here the users host their social data and application on their own nodes. All the individual nodes connect with each other and share resources. Since users' social data have been hosted in their own machine, the availability of this data depends on the online behaviour of them.

There exist many decentralized OSNs which are based on the above discussed types of decentralization such as Diaspora, Friendica, Appleseed, Safebook, Social Igniter, etc [2].

- **Diaspora:** It was started in 2010 as open source project and was nominated for "Best Social Network" in the 2011 Mashable.com awards. It uses Ruby as the programming language and Salmon protocol which a message exchange protocol running over HTTP designed to decentralize commentary and annotations made against newsfeed articles such as blog posts [15].

- **Friendica:** It was started in 2010 as an open source project, which uses PHP and DFRN protocol which is a distributed communications protocol which provides privacy and security of communications and also provides the basis for distributed profiles and making connections. In addition to that, it also uses OStatus which is an open standard for distributed status updates and OpenID for authentication [60].
- **AppleSeed:** It was started in 2004 as an open source project and was the first open source and distributed social networking platform. PHP is used as the implementation language and QuickSocial protocol for networking which is a simple, unified protocol for distributed social networking that uses HTTP and JSON to pass data between nodes [34].
- **Safebook:** It is a privacy preserving, open source OSN leveraging on real-life trust. It was proposed in 2009 by the Networking and Security Department in Eurecom and was funded by the Socialnets research project funded by the European Commission under the Information Society Technology. Python is used to implement it and it uses the standard P2P protocol for communication [12].
- **Social Igniter:** It was started in 2010 as an open source project and is a lightweight, simple to setup, easy to extend, social content management system. It uses the PHP language for implementation and WebFinger protocol which aims to provide information about people by their E-mail addresses. In addition, it also uses OpenID and OAuth [53].

After detailed analysis of aforementioned decentralized OSNs, we focussed on Friendica because of its unique features and active development over the other decentralized OSNs.

Friendica is a privacy-aware decentralized social networking platform which follows a P2P network architecture. Following are the features that are unique to Friendica compared to other decentralized OSNs [60],

- It can adapt to both types of decentralized architectures i.e. it allows users to host their data on cloud or host it completely as a P2P application.
- It is interoperable with other social networks where relationship can be made between two profiles from different social network.
- It supports administration on each node with easy to use administrator interface.
- It has a global member directory and the user can individually control whether to publish her profile which ensures the privacy of the user.
- It has a modularized plugin architecture which allows users to easily develop their own application/feature.
- It implements strong encryption between nodes.
- It allows users to disable anonymous profile viewing.
- It supports multiple profiles.
- It allows the user to set a time period for expiration of her post and the content is removed from all the servers if there is a copy.
- It has built-in support for OStatus federation (status.net, identi.ca, GNU-social, many others).
- It has built-in implementation of Diaspora protocols allows communication with any Diaspora member.
- It supports Email contacts and communications (two-way) via IMAP4rev1/ESMTP.

- It allows to import arbitrary websites and blogs into your social stream via RSS/Atom feeds.

2.7 Summary

As discussed above, some approaches for increasing the transparency and the users' trust in RSs involve explanations or visualizations making the mechanism of recommendations understandable or visible to the user. The user awareness is particularly important as a way to alleviate the filter bubble problem and increase the users' trust in the filtered stream. There haven't been approaches proposed to create awareness and visualize or explain the filter bubble problem while allowing users to nudge the bubble.

Recommendation systems have been applied massively in online social network sites and the filter bubble can have serious consequences on users' access to unbiased social information. The next chapters present a real world implementation of an interest based stream filtering approach, design of the visualization, results of a small scale exploratory user study to evaluate the usability of the visualization and users' understanding of the filtering approach and their trust in the system, a qualitative evaluation to in depth understand the user perceptions of the visualization, a large-scale quantitative study to evaluate the understandability of the visualization and finally a conclusion.

CHAPTER 3
MADMICA

MADMICA – an implementation of an interest based relationship filtering mechanism to filter out irrelevant social data in a decentralized OSN, based on the Friendica P2P protocol. The mechanism uses the interaction between users to construct interest-based models of the relationships between users, which act as filters while propagating social data related to a certain area of interest. The implementation of an interest-based stream filtering has been done in order to evaluate its usability and shortcomings and was used as the platform for implementing the filter bubble visualization. This chapter presents the implementation of interest-based stream filtering algorithm in MADMICA.

3.1 Introduction

MADMICA [35] is an implementation of a privacy-aware decentralized (peer-to-peer) OSN using the Friendica open source framework [29]. MADMICA implements an approach to filtering social data, according to a model of the strength of the user's interests in different semantic categories overlaid over a model of their social relationships, which was originally developed and evaluated in a simulation [56]. In essence, the filtering approach is based on a model of the user's interest in a finite set of categories of social data that is overlaid with a model of the strength of user interpersonal relationships (over each category). It consists of a matrix of relationship strengths (values between 0 and 1) between the user and each of her friends in different areas of interest. The model is updated based on implicit and explicit feedback from the user, based on the user actions over the social data (e.g. rating, commenting, forwarding or ignoring). The filtering of social data depends on the value of the strength of the relationship between the two users. The current relationship strength between a user and her friend in a given category is compared to a certain threshold value (currently a constant for all users in the OSN, but this could be personalized in the future) by the filtering algorithm to decide whether a new social update from this friend in the given category should be shown in the user's stream, or hidden. The news feed home page is quite similar to Facebook but user has to select a category when she wants to post an interesting news (see Figure 3.1).

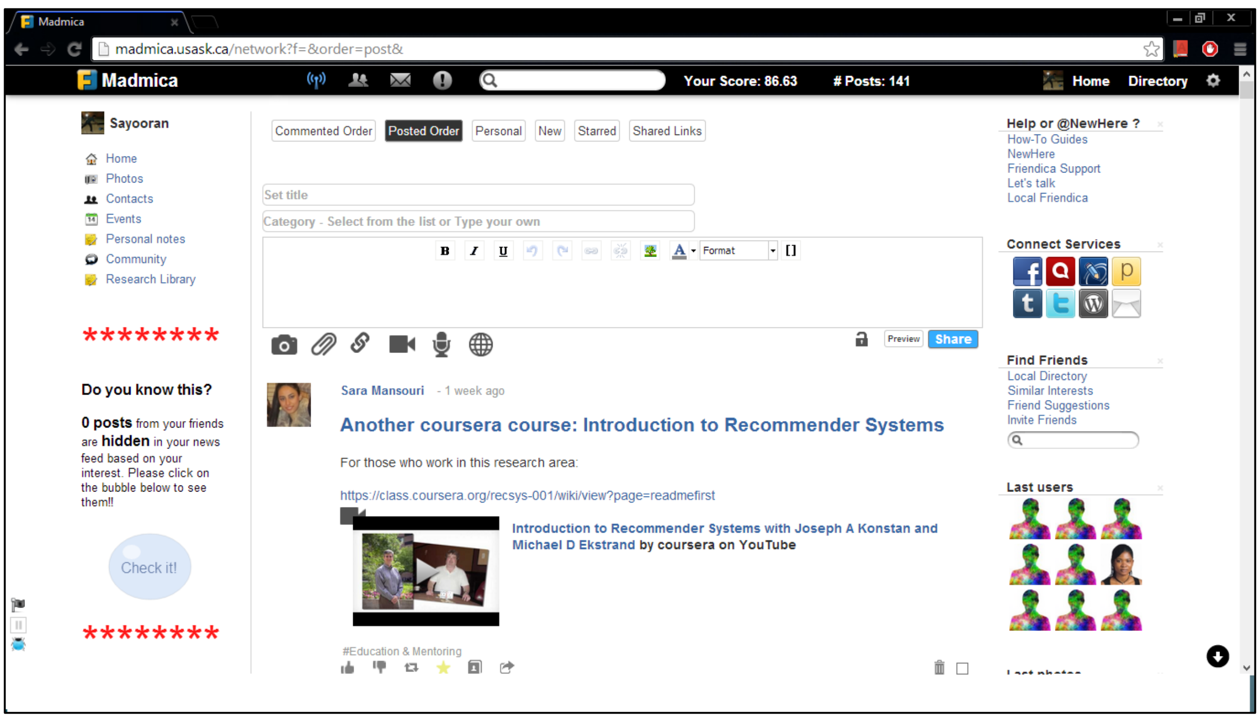


Figure 3.1: Screenshot of news feed home page in MADMICA

3.2 Architecture of MADMICA

MADMICA is built based on Friendica - a privacy aware decentralized social networking platform which follows P2P network architecture [35, 60]. The architecture comprises a collection of distributed nodes which are able to act as a client or server to other nodes in the network and communicate with each other on users` behalf. Each node is hosted in a server which is similar to a typical hosted blog. Friendica uses the Distributed Friends & Relations Network (DFRN) protocol to communicate and share information with peers in a decentralized manner [29]. Data are stored at the peers and the availability of the social data depends on the online behavior of peers. Moreover, data are accessed through URI. The data can be anything and Atom Syndication Protocol is used as a structural wrapper with several extensions to offer different data types and informational messages such as activity streams, threading, media, etc. MADMICA uses notification and polling methods to discover data among nodes. Recipients of targeted and timely data are notified, whereas newsfeed and public broadcasts are picked up by others polling the node from time to time.

Privacy of user data is well maintained in MADMICA using several aspects of the system. There are two types of communication between nodes - public communication and private communication. Secure handshake is made between nodes before each private communication. This ensures the user data are not exposed to third parties. Each node can be totally standalone and used by one individual for the ultimate privacy. Also each node can be used by many users or as a group. Since the nodes are distributed and own their data, no single company or individual can see all of the data in the network and mining this distributed data is very hard. The DFRN protocol specifies two types of information ownership. There can be an author who provided the communication and an owner who is the person on whose homepage or profile page the information was posted. The author can always delete her social data but has no control over distribution. The owner has control of distribution and allowing view rights. But redistribution of content is only allowed to the owner and it will not be redistributed to a third party by anyone who receives it from the owner. This ensures the full control of privacy to the owner of the content.

3.3 Interest Based Relationship Filtering in MADMICA

Tandukar and Vassileva proposed an approach for reducing the flow of irrelevant information in decentralized OSNs and evaluated the mechanism using a realistic Erlang simulation with 2318 nodes [56, 57]. The results showed that the filtering mechanism works and with the increasing number of social data passing through the network, the nodes learn to filter out irrelevant social data, while serendipitous important social data are able to bypass the filter.

As described in [56, 57], an interpersonal relationship model is used as a filtering mechanism for irrelevant information. The relationship model consists of a vector of relationship strengths between a user and her friends in different areas of interest. The filtering of social data depends on the strength of relationship between the two users, but this strength is contextualized according to a particular category of interest to which the social data item belongs to. The intuition behind this is that two people can be friends, but not share the same level of interest in different topics or categories and not trust each other’s judgment with regard to these categories. Therefore, a user may be interested to receive updates from her friend about, say fashion, but not in politics or health. On the other side, the same user may be interested to hear about health topics from another friend, yet, she may not be interested in her updates about fashion. While it adds complexity of representation and computation, it is advantageous to add an interest dimension to a relationship, since it allows the flexibility to filter both based on the source of the update and the category of interest. The strength of a relationship from one user to another in a given category of interest is based on previous interactions related to this category of interest. In general, to determine the area of interest of the shared information, users have to either tag their updates with the interest areas or the system has to extract semantics from the shared social update. However, allowing a full semantic categorization will complicate the system and there is no benefit at the algorithm level. That is why it is better to limit it to a certain fixed number (for example, 10 or so) of predefined general categories, similar to those used in Yahoo or other news sites, e.g. politics, news, technology, sports, health, fashion, living, art, games, humor, etc.

The relationship strength from one user (sender) to another (recipient) is updated using the feedback that the recipient has provided. The feedback is based on the actions of the recipient and influences differently the relationship strength. For example, if the social data are viewed and re-shared, this would increase significantly the relationship strength. Viewing and commenting or rating will also increase the strength. Just viewing and not doing anything else would slightly reduce the relationship strength in the category of the social update. The relationship strength is calculated using a simple formula for simulated annealing (reinforcement learning).

3.4 Implementation of MADMICA

MADMICA (http://madmica.usask.ca) is built with PHP, jQuery and MySQL technologies. The interest based relationship model filtering is implemented in MADMICA as a plugin. This ensures that the modularized plugin architecture of Friendica is preserved. So users of each MADMICA node have the ability to turn off the plugin so the filtering. As shown in Figure 3.2, each user hosts her MADMICA node instance. All the data which belong to the user reside in her node. Moreover, each user will have an interest based model for each of her friends/contacts and the model is stored in her own node. Since the model is used to filter only the stream of the user who owns the model, there is a reasonable benefit to maintain the model on her node by sacrificing some space and computational power. It also prevents the model from being tampered by malicious content creators to spam other users in the network with irrelevant information.

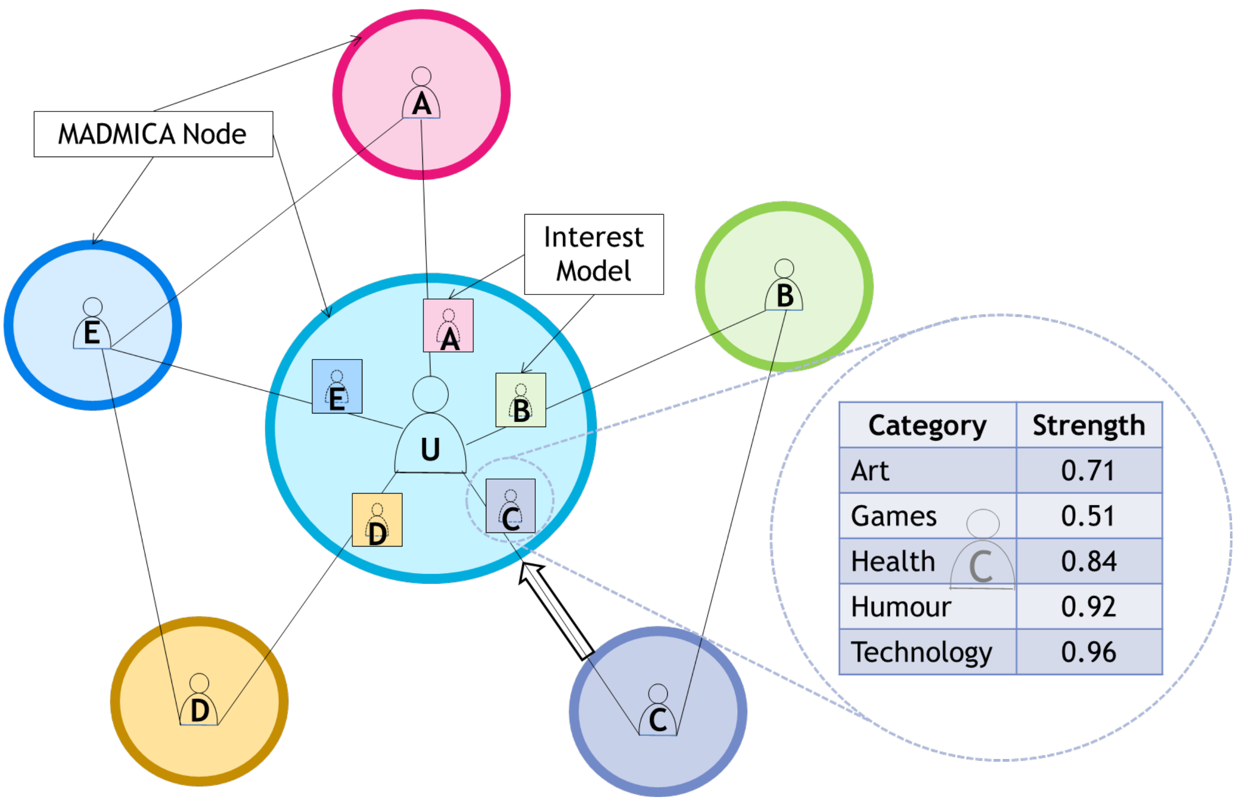


Figure 3.2: Interest based relationship model

Consider a scenario where user C wants to share with her friends something that she has found interesting in the category "technology". Each social data in MADMICA comprises four elements: heading, category, content and target audience. Category and content are mandatory fields. So, user C must provide a category by selecting one from the dropdown list or typing her own category. Once the content has been categorized, the user can share it with their friends. When the social data are shared, it will reach all the nodes of the user's friends. So in this case, the user U will receive the shared social data. Once the node of user U receives the social data, while generating the view of the stream for the user, the user model of the node originating the social data (user C) is checked to see whether to make the social data visible or hide the social data based on the relationship strength with that node in the category of the social data (see Figure 3.3). In this example, the interest based relationship strength of user C on technology is 0.96. So the social data will be made visible as the strength is greater than a threshold value which is currently set to 0.4. This threshold value has been found in the previous work on evaluating the interest based relationship model using a simulation by Tandukar and Vassileva [56].

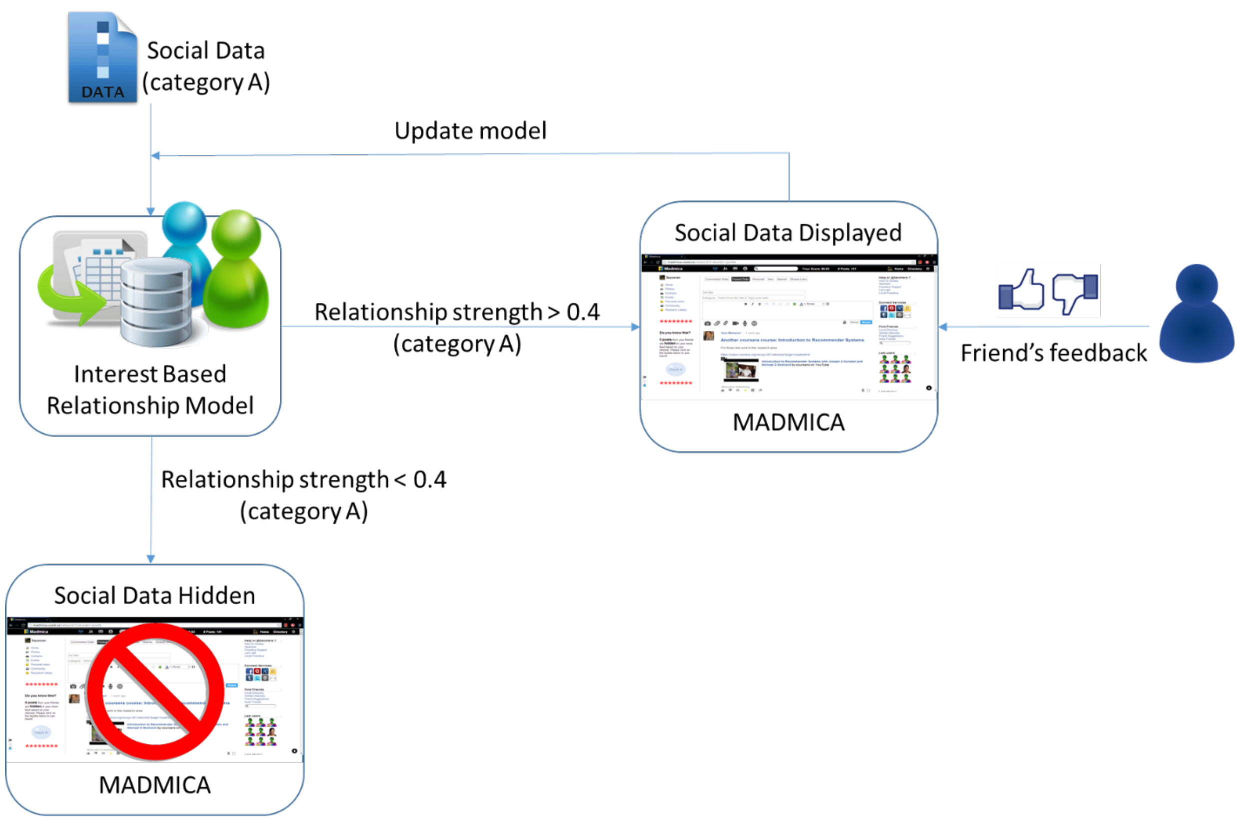


Figure 3.3: The interest-based filtering process

The relationship strength between one (sender) to another (recipient) user is updated using the feedback that the recipient has provided. Initially, the strengths of relationships with respect to each area of interest among all friends are set to 1. In the aforementioned example, the node of user U will send feedback to the node of C based on the actions of the user U. Different actions result in different feedback values and influence the relationship strength differently. Table 3.1 provides the feedback values for a range of actions the user can perform upon receiving the incoming social data. For example, if the social data are re-shared, the feedback value is high (0.9) and would increase significantly the relationship strength between U and C in the category of the data. Commenting will also increase the relationship strength. Just ignoring would reduce the relationship strength. By updating the relationship model, the strength of relationship for certain topics/areas will weaken and ultimately fall below the threshold and thus in the future social data by the friend with whom relationship strength on that topic is low will be filtered away (i.e. not shown in the stream). Thus gradually, over time, the system as a whole through interaction of all the nodes and implicit user feedback will learn about user interests and relationships and users will only see relevant social data from their friends with whom they have strong relationships.

Table 3.1: Categorization of Feedback

Type	Action	Feedback
Type 1	Share	0.9
Type 2	Comment	0.8
Type 3	Like	0.7
Type 4	Favorite	0.6
Type 5	Bookmark	0.5
Type 6	Ignore	0.4
Type 7	Dislike	0.3

The updating of relationship strengths happens after each interaction between the users and feedback for the social data. It is carried out asynchronously using AJAX requests from the stream. The strength of relationship between user U and user C for a given interest area/topic I is calculated using a simple formula for simulated annealing (reinforcement learning).

$$S_C^U(I) = \alpha * S_C^U(I)_p + (1-\alpha) * F \quad (1)$$

Here, $S_C^U(I)$ is the new strength of relationship from C to U, $S_C^U(I)_p$ is the previous strength of relationship for an interest area *I*. The parameter $\alpha \in [0, 1]$ is a learning rate of the system which should be considered high as possible to conserve relationships on basis of interest areas between users. It is important to note that $S_C^U(I) \neq S_U^C(I)$, i.e. the direction of relationship matters. The feedback that user C gives to user U for the social data that user C sent to user U is denoted by F, and its value varies from 0.3 to 0.9 as specified in Table 3.1.

3.5 Summary

This chapter presented the architecture of MADMICA and the implementation of interest-based stream filtering algorithm in MADMICA. The next chapter will discuss the design and implementation of filter bubble visualization using this MADMICA platform.

CHAPTER 4
FILTER BUBBLE VISUALIZATION

To achieve the goal of creating awareness, explanation, and control of personalized stream filtering in an OSN to alleviate the filter bubble problem and increase the users' trust in the system, we propose a visualization that metaphorically explains the filtering mechanism and provides means of control over certain parameters of the filtering for the users. This chapter presents the filter bubble visualization design and implementation in MADMICA.

4.1 Visualization Design

The visualization is based on a bubble metaphor to make the effect of the personalized stream filtering in OSNs more understandable for the users (see Figure 4.1). It divides the space of the screen in two parts - outside and inside the bubble. The items that are inside the bubble are visible for the user, those outside the bubble are those that have been filtered away and are invisible in the stream (but they are shown in the visualization). The visualization provides two alternative points of view: one focusing on the user's friends and one focusing on the categories of the social data originating from them in the OSN.

For example, Figure 4.1 shows a "category view", and in the bubble are the categories of social data that are not filtered away and the user is seeing in her stream (e.g. "Food & Health", "Education", "Cool Stuff", "News"). Outside the bubble are shown the categories of social data that are currently being filtered away from the user stream ("SOA", "Sports" and "Mobile"). The bubble shape design was chosen, not only because it fits well with the "filter bubble" metaphor, but because it is naturally scalable - can grow in size to accommodate more circles inside.

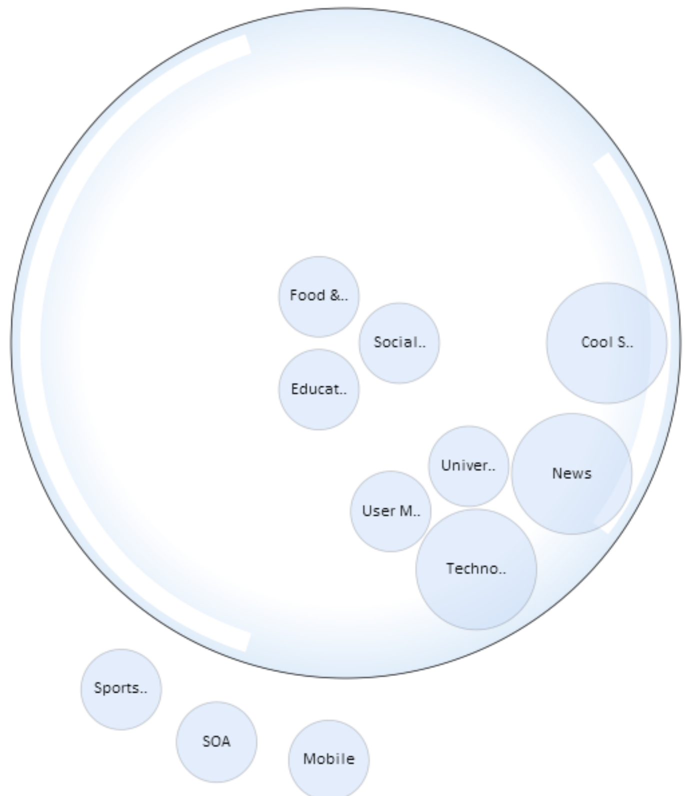


Figure 4.1: Filter bubble visualization – category view

The decision to visualize categories of social data rather than the data itself was made since it allows users to get a feeling of how the filtering mechanism works; as filtering is based on the content of the social update. Making the users aware of the different categories in which status updates are classified provides some transparency of the mechanism, which otherwise users won't be aware of. In essence, the category view summarizes what categories of social data the user is interested in and what categories of social data she has tended to ignore in her stream. In addition, the abstract category view scales better than showing the specific updates and does not lead to an overcrowded view and cognitive overload. Upon clicking on a circle representing a given category, a small pop-up window shows the list of social updates from the stream that belongs to the category. In this way, for example, by clicking on the "Mobile" circle shown in Figure 4.1, the user can see all the status updates from her OSN stream related to the "Mobile" category, that have been hidden from her. Thus we follow Shneiderman's [52] visualization design principle "overview first, details on demand".

The second view, called "friends view" (see Figure 4.2), shows in a similar way the bubble, but instead of circles representing categories of social data, the circles represent the user's friends who have posted the social data. If a friend's circle is inside the bubble, then the social data from that friend are visible in the user's stream, whereas if the friend circle is outside the bubble, the social data from that friend are hidden and not displayed in the stream.

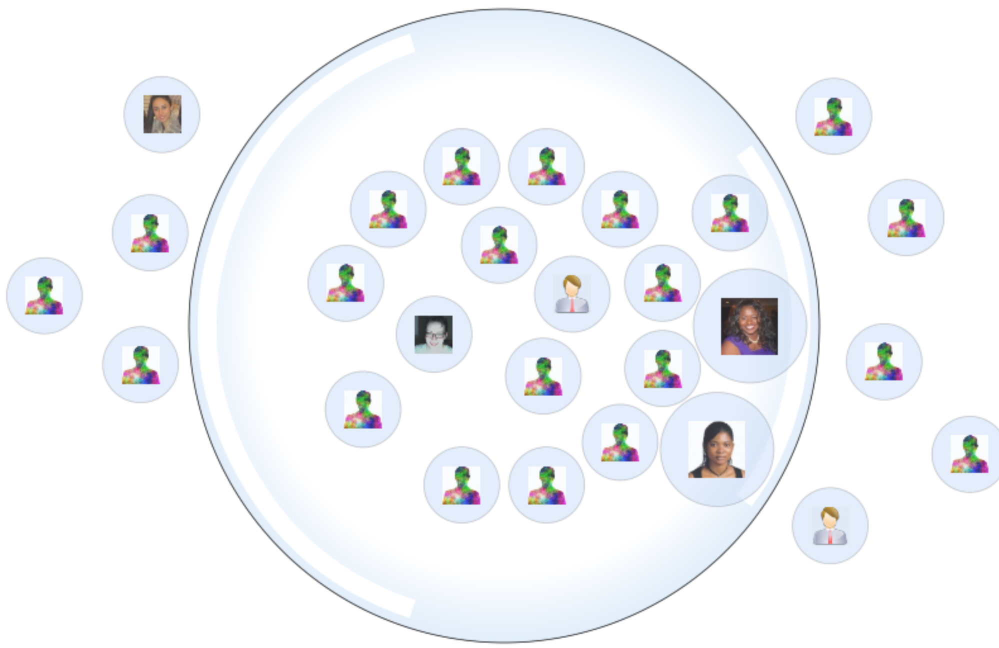


Figure 4.2: Filter bubble visualization – friends view

Since the filtering mechanism differentiates the filtered data both based on who the data comes from and the category of the data, the friends view displays the relationship that the user has with each of her friends with respect to a given category.

The user can select a specific category from a drop-down menu on the top of the screen (see Figure 4.3), and then see which of her friends will be inside the bubble for this category, i.e. who she is connected with respect to the chosen category. These are the people whose social updates in the selected category the user is seeing in her stream; the updates in this category of the other friends who are outside the bubble are being filtered away from the stream. In order to provide a better explanation of what is happening in the personalized stream, the position of category/friend circles represents the visibility of social data in your stream. For both views, the size of the category/friend circle denotes the number of social data items in a certain category by friends and it helps to understand how much social data are visible in the stream and how much data are hidden in the stream by the filtering mechanism and who is posting more social data and who is posting less.

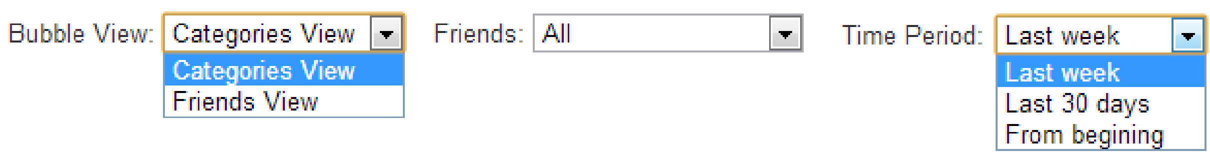


Figure 4.3: Filters on the visualization view

Another feature of the visualization design focuses on giving control of the personalized stream filtering to the users. This is done by allowing users to drag and drop the category/friend circles inside and outside the filter bubble. Dragging a category/friend circle inside the filter bubble enables the users to discover an interest area (category) which appears interesting or strengthen the relationship with a friend whose social data have been not visible in the stream. On the other hand, when users drag a category/friend circle outside the bubble, social data belonging to that category/from that friend will not appear in the stream anymore. This helps the users to get rid of uninteresting social data and also to avoid spammers who flood the stream with uninteresting and unwanted social data.

4.2 Implementation of Visualization

The technology used to implement the visualization is HTML 5 with jQuery. The code can be run by any device on a browser without any plugin and can be adjusted to fit any size screen in a graphically pleasing manner [24]. The visualization is implemented in MADMICA as a plugin. This ensures that the modularized plugin architecture of MADMICA is preserved. So the user of each MADMICA node has the ability to turn off the plugin.

Users are notified in a side menu next to their stream with a message "Do you know this? N posts from your friends are hidden in your news feed based on your interest. Please, click on the bubble below to see them!" This creates awareness to the users that filtering is happening in the stream and some social data are not shown in the stream. When users click on the small bubble icon, the visualization plugin is loaded. When loading the visualization, all shapes are generated on the HTML5 canvas using KineticJS framework according to the data retrieved from the database. The visualization view is updated instantaneously and it always shows the category/friend circles according to the newest value from the user's relationship model. The default view is category view. Stored procedures have been used in MySQL to speed up the loading of visualization with necessary data.

The visualization can be viewed based on three different filters: bubble view, friends/category, and time period (see Figure 4.3). This provides flexibility for the users to choose the desired view, and a time period of interest, since their interests in different categories and their relationships with friends are dynamic. The Bubble view filter consists of a dropdown menu that allows the user to select one of two views: category view and friends view. When the "category view" option is selected, a dropdown list is loaded in the Friends filter containing all the user's friends, so that she can individually select a friend and view all the semantic categories of social data that the user shares with this friend (i.e. shared interest) inside the bubble and those categories with respect to which the user and this friend do not share interest (outside the bubble). In this view (selected category and selected friend), the circles representing the categories will appear positioned either inside or outside the bubble, based on the relationship strength value with that friend on that category.

The bubble visualization is constructed on a kinetic stage class. It consists of three layers namely circleLayer, bigBubbleLayer, toolTipLayer (see Figure 4.4). circleLayer consists of either category/friend circles in a circleGroup. bigBubbleLayer has the outer big circle and its effects representing the bubble metaphor. toolTipLayer is used to provide tooltip support to the circles in the circleLayer. Before drawing the category/friend circles, all the category/friend circle data are retrieved using an AJAX call to the MADMICA server which is written in PHP (see Figure 4.5). Once the stage is loaded with all the components, circles are created using arrays based on the received bubble data from the server and then real circles are drawn referring to the created circle array. The code in Figure 4.6 shows the function which is used to create the circles.

```
var stage = new Kinetic.Stage({
                    container: 'container',
                    width: 1080, height: 950

              }),

     circleLayer = new Kinetic.Layer(),
     circleGroup = new Kinetic.Group(),
     tooltipLayer = new Kinetic.Layer(),
     bigBubbleLayer = new Kinetic.Layer();

     circleLayer.add(circlGroup);
     stage.add(bigBubbleLayer);
     stage.add(circleLayer);
     stage.add(tooltipLayer);
```

Figure 4.4: Components of bubble graphic

```
$con = mysql_connect($db_host,$db_user,$db_pass);
$dbs = mysql_select_db($db_data,$con);

$uid = $_REQUEST["uid"];
$period = $_REQUEST["period"];
$contacId = $_REQUEST["cid"];

$result = mysql_query("CALL GetCategoryBubble($uid, $contacId,
$period)");

$r = array();

if(mysql_num_rows($result)) {
        while($x = mysql_fetch_array($result, MYSQL_ASSOC))
                $r[] = $x;
                mysql_free_result($result);

        }
 mysql_close($con);

 echo json_encode($r);
```

Figure 4.5: PHP code to send the category bubble data to client

```
var circles = [];

function createCircles(c,r, radius, k, j, isCategory){

            for(var i=0; i<k;i++){

                    var x =  calculateCenterX(cx, radius,r,j);
                    var y = calculateCenterY(cy, radius,r,j);
                    j = j + c;

                    var valueToPush = {};
                    valueToPush["x"] = x;
                    valueToPush["y"] = y;
                    valueToPush["s"] = radius;
                    valueToPush["used"] = 0;

                    if(r >= 300){
                            valueToPush["out"] = 1;
                    }
                    else
                            valueToPush["out"] = 0;
                    circles.push(valueToPush);

            }
    }
```

Figure 4.6: Circle creation using arrays

To handle the problem of category name not fitting into the small circles, we display only a fraction of text inside the circle and the full text is shown as a tooltip when the mouse pointer is hovered over a category circle. To see the hidden social data for a particular category, the user has to click on the category circles which are outside the filter bubble and a pop-up window is loaded with the individual links to the social data (see Figure 4.7). By clicking each link on that menu, the user can see the individual social data items which have been hidden in the stream by the filtering mechanism.

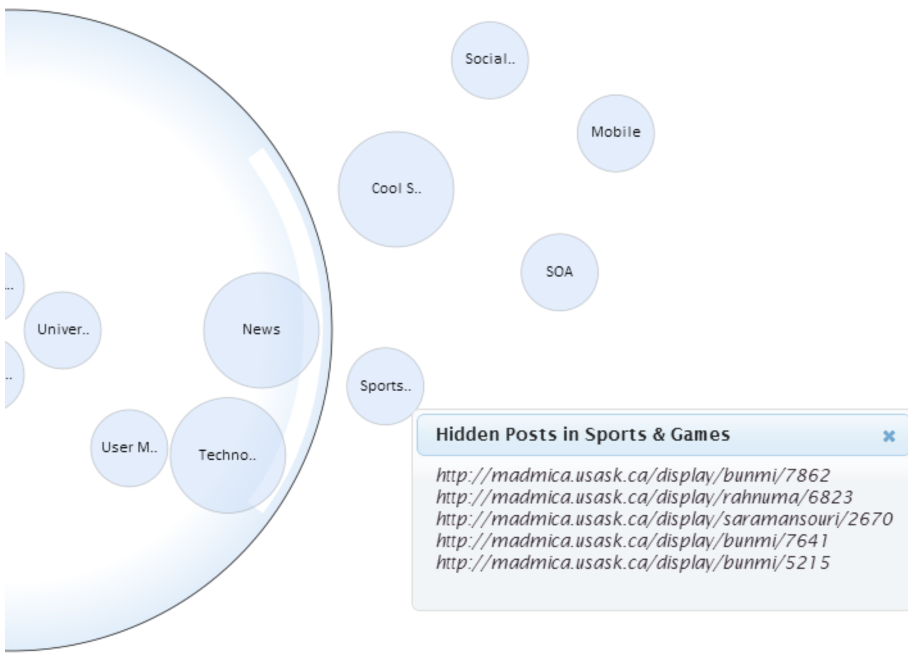


Figure 4.7: Screenshot of hidden posts pop-up window

The default option in the Friends filter is “All” which shows just all categories represented inside or outside the bubble depending on the average relationship strength in these categories across all the friends of the user. On the other hand, when the “friends view” option is selected in the Bubble view filter, the second filter changes to “Category”, allowing the user to pick a particular category of interest (the default is again “All”). The circles in this case represent the user’s friends and contain the friend’s avatar or photo, and the name of the friend appears as the mouse pointer is hovered over it (see Figure 4.2). In this case, the second filter shows a dropdown menu showing all the semantic categories available in MADMICA. The user can select a particular category and see which of her friends (from those who have posted social data in the selected category) are inside or outside her bubble. Both views can be generated based on a time period filter. This filter comprises several options: “from beginning”, “last 30 days” and “last week” as the dropdown list labels. By default “last week” is selected when the visualization is loaded which shows the categories/friends circles in/by which social data were generated during the last week.

To add control of the filtering, we have added the drag and drop feature so that users can drag a category/friend circle inside and outside the filter bubble to show or hide data from this category/friend. When dragging a category/friend circle inside the filter bubble, AJAX (Asynchronous JavaScript and XML) requests are generated from the visualization and the corresponding model values for the interest based relationships are updated in the database. Similarly, when dragging a category/friend circle outside the filter bubble, another set of AJAX requests are generated to save the data. To let the users know about the results of the drag and drop action, a message is displayed to the user informing about whether the social data will be made permanently visible or hidden based on the users’ action.

4.3 Summary

This chapter presented the design and implementation of interactive visualization. The visualization design is based on bubble metaphor and the implementation has been done in MADMICA. Next chapter presents the evaluation and results of the MADMICA system and the proposed interactive visualization.

CHAPTER 5
EXPERIMENT AND EVALUATION – PILOT USER STUDY

In this chapter the MADMICA system and the proposed interactive visualization are evaluated in a pilot user study and results are presented and discussed. The purpose of the small scale qualitative case study is to evaluate the usability and user acceptance of MADMICA system and the visualization and to see whether it achieves its goals of providing awareness, control and trust in the filtering mechanism of MADMICA. Unfortunately, due to the small scale of the study, the user experience with the privacy preserving aspects of the distributed architecture was not tested. This remains for future work. The subjects were 11 graduate students from our research lab who used the MADMICA system instead of Facebook to share interesting and research relevant links over a period of 3 weeks in March 2013. All participants were international students from various parts of the world (Asia and Africa), with computer science background and very familiar with social networks. Six were female and five were male. Moreover, gamification is used to motivate the participants to be active in the network throughout the experiment.

5.1 Hypotheses

The main goal of this small-scale user study was to find out if the visualization is usable, if it creates awareness of the filtering, understanding of the personalized stream filtering mechanism and ability to control it to alleviate the filter bubble problem and increase users' trust in the filtering. So the evaluation mainly aims at testing the following hypotheses.

1. The proposed visualization creates awareness, understanding and control of personalized stream filtering to alleviate the filter bubble problem.
2. The visualization increases the user's trust in the personalized stream filtering.

3. The visualization of filter bubble increases the users' experience with the system.

The another goal of this pilot study was to find out if the filtering mechanism reduced social data overload, if the users find the possibility to maintain ownership of their data attractive in principle and if the MADMICA system, as implemented, is usable.

5.2 Experimental Setup

Due to the small number of users and the fact that the users were lab students and knew each other well, privacy wasn't an issue, so for efficiency sake, we hosted only one MADMICA node to support all the participants. The study took place in a natural environment where users can use their own computers at their own convenient time (like using Facebook). Each user was asked to register at MADMICA and create a profile. Then they added each other as friends and shared anything they found interesting with their colleagues. We provided 11 semantic categories to classify the social data (the classification into one of the categories had to be done manually by the user when sharing something new with their friends), but allowed users to create their own categories (subject to approval by administrator in the experiment). The categories were chosen based on the main research areas in our lab, such as, education & mentoring, user modeling, mobile technologies, social computing, SOA, and common interest areas, such as food & health, news, sports & games, technology, university news and cool stuff.

At the end of the study, the participants were asked to answer two questionnaires. One questionnaire was related to the usability and user acceptance of MADMICA system and the second questionnaire was related to the filter bubble visualization. As this was a qualitative study, the questionnaire had mostly open ended questions, even though there were some closed ones. The open ended questions enable participants to provide free feedback and describe their own ideas or suggestions without any restriction. Responses for some of the closed questions were given on a

10-point Likert scale. Both types of questions focused on finding out about the user experience related to the proposed visualization and about the usability of the visualization. Moreover, they also focused on finding out about the user experience related to the implemented filtering approach and about the usability of the MADMICA system, as well as their opinions in principle of the idea of user control over their data, enabled by a system with P2P architecture. All of the 11 participants completed the final questionnaire. In addition to the questionnaires results, the usage of visualization of filter bubble was tracked by the system in order to collect data about users' actions on the bubble such as viewing the filter bubble visualization, dragging category/friend circle inside the filter bubble and dragging category/friend circle outside the filter bubble.

5.3 Gamification

To keep the participants engaged and motivated to be active in the network throughout the pilot study period, we provided monetary rewards for participation in the study, and gamified MADMICA using a simple scoring mechanism. The following game qualifying rules were met:

- At the end of the 3 weeks period, each participant had posted at least 50 social updates.
- At least 3 social updates had to be done per day by each participant for 21 days and a minimum total of 600 social data had to be produced by all participants at the end of the 3 weeks.
- A total reward of $600 was distributed among the participants, according to the participation score, computed using the following formulas:

$$Reputation = \frac{\sum_{friends}\sum_{categories} strength}{\#\ of\ friends * \#\ of\ categories} \quad (1)$$

$$Score = \#\ of\ posts * Reputation \quad (2)$$

Equation (2) is used to calculate the reputation of each user in MADMICA based on strength values of her relationships with her friends (from the viewpoints of her friends) over all categories. The scores are displayed in a leaderboard with ranking (see Figure 5.1).

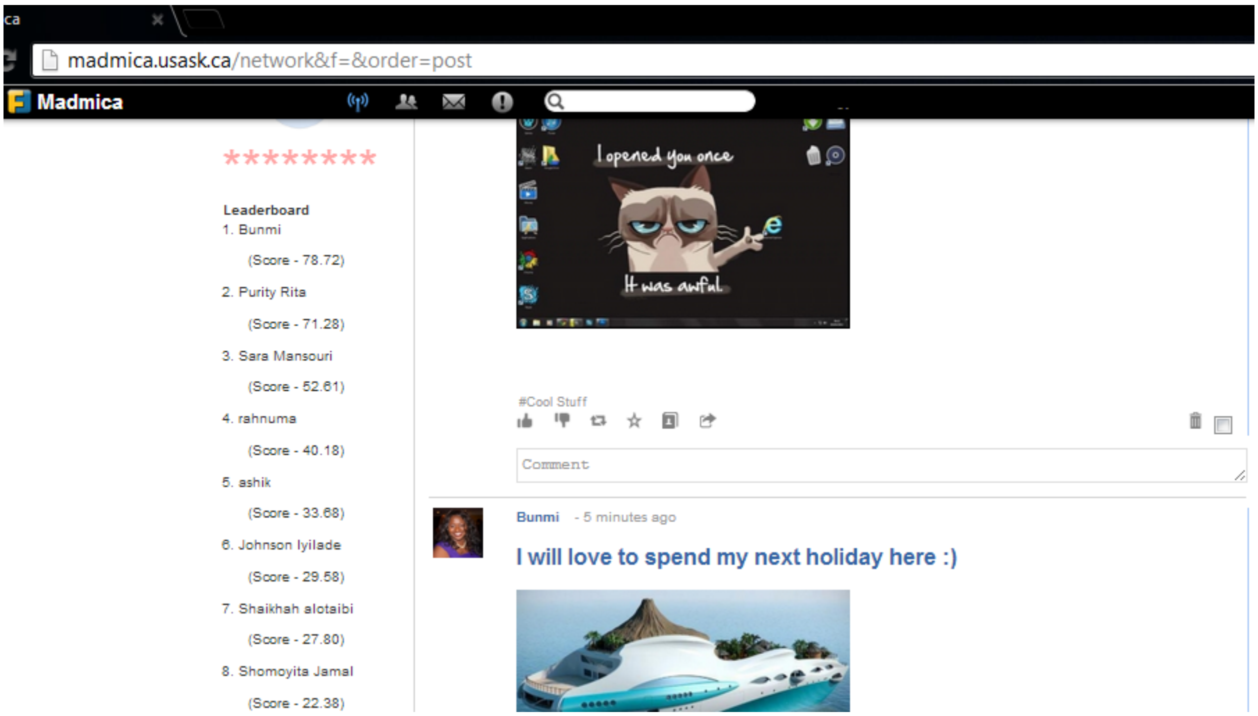


Figure 5.1: MADMICA home page with scoreboard

The SQL Procedure shown in Figure 5.2 is used to calculate the individual score for each user.

```
PROCEDURE friendicaDB.GetScore (IN userid INT(10))
BEGIN
        DECLARE NoOfPosts INT;
        DECLARE Reputation DOUBLE;

        SELECT IFNULL(avg(irm.rvalue), 0) INTO Reputation from ir_model irm
         INNER JOIN contact ct on irm.contact_id = ct.id
         INNER JOIN category c on irm.category_id = c.id
        INNER JOIN user ur on ct.nick = ur.nickname
         WHERE ct.nick in (select nickname FROM
        user WHERE uid = userid)
         and c.active = 1 and c.approved = 1 and ct.pending = 0 and ct.blocked = 0;


        SELECT IFNULL(COUNT(i.id), 0) INTO NoOfPosts FROM item i WHERE i.uid = userid
                                AND i.type ='wall'
                                AND i.id = i.parent
                                AND i.body != '' AND i.tag != ''
                                ORDER BY i.created DESC;
        SELECT  ROUND(NoOfPosts * Reputation, 2) AS score;
END
```

Figure 5.2: SQL Procedure for calculating the individual score

5.4 Results

The first few questions related to MADMICA aimed to evaluate the user awareness of the filtering and its impact (see Table 5.1). All of the 11 respondents agreed that they noticed the system filters some social data away and that the level of interestingness of the social data from their friends changed over time. Ten participants (91%) felt that the social data in their stream got more interesting with time, while one participant reported that the social data got less interesting. Table 5.2 shows the results for the question Q1 about whether at any point participants felt there were too many social data to keep track of or not. Three other close ended questions (Q2, Q3, and Q4) are related to usability of MADMICA. The results are summarized in Table 5.3.

Table 5.1: Closed Questions with Likert Scale

#	Question
Q1	Did you at any point feel there were too many social data to keep track of?
Q2	How easy it is to post a new social data in MADMICA?
Q3	How easy it is to select a category of your social data in MADMICA?
Q4	How easy it is to keep track with the social data of your friends?

Table 5.2: Results of Questions Related to Social Data Overload

Question	Too few social data			About normal				Too many social data		
	1	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*	*10*
Q1					1 (9%)	2 (18%)	7 (64%)			1 (9%)

Table 5.3: Questions Related to Usability of MADMICA

Question	Difficult									Easy
	1	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*	*10*
Q2				1 (9%)	1 (9%)		1 (9%)	3 (27%)	4 (36%)	1 (9%)
Q3				1 (9%)		1 (9%)	1 (9%)	2 (18%)	3 (27%)	3 (27%)
Q4		1 (9%)		1 (9%)	1 (9%)	1 (9%)	1 (9%)	2 (18%)	2 (18%)	2 (18%)

A set of open-ended questions gauged the users' trust and understanding of the filtering mechanism. The first question asked if the participant trusts a system that filters social data away from the stream (e.g. Facebook). Two participants (18%) responded yes, and two – no. Seven participants (64%) answered that their answer depends. In a follow-up question, out of those seven participants, five said that they would trust the system if they *had some means to control* the filtering mechanism, and other two said they would trust the system if they *understood* the mechanism. An open-ended question probed the participants' understanding on how the filtering works in MADMICA. Eight participants (72.73%) responded that it was based on their interaction with the social data in terms of likes, comments and re-sharing on different social data categories. Some of the answers state: "*I think based on my interaction in terms of likes and comments on different posts categories*", "*based on my posts and my interactions on others' posts*" and "*it filters away posts that I don't have much interests to*". The remaining three participants guessed that the filtering is done based on the categories/topics of the social data. Excerpts from the answers state: "*I think it filters the post depending on the topics of the posts*" and "*based on ones posts categories*". In reality, the filtering mechanism uses a combination of these two ideas - user interaction with the social data received from friends and the categories of the data, but none of the participants guessed correctly that the filtering depends on *which friends* the social data came from.

Another set of questions were aimed to evaluate the user experience of the MADMICA system as an alternative, privacy-protecting distributed social network and its usability. Eight participants (73.73%) stated that they have not tried to delete their data and did not have any concern about the privacy of their data hosted in the single MADMICA node during the experiment. But the remaining three other participants preferred to host their data (photos, files) on their own MADMICA nodes. To evaluate the experience with the system, the participants were asked an open-ended question about their favourite and least favourite features/things in MADMICA. Most of the participants reported that they liked: 1) the categorization of social data and the ability to see the social data by just clicking a category, 2) the interface that made it very easy to share links, videos, photos, and 3) the design layout. The least features by most of the participants were: 1) the limitation of being able to select only one category/tag per social data, and 2) the "dislike" button for social data (a feature of Friendica). To the question of finding whether something was obtrusive in MADMICA, most participants (81.82%) replied negatively, one participant said that categorizing a social data was obtrusive and another one said "sometimes posting was not easy".

Finally a question was asked to find what would encourage returning to MADMICA in the future. Most participants answered that the interesting, useful, relevant and exciting news found on their stream shared by their friends will be a motivation to use MADMICA continuously. Also some participants (18.18%) answered that the competitiveness and the point system might be another motivation factor to use MADMCIA in future. The following are some excerpts from the answers by the participants to the aforementioned question: "*nice gist from friends;*", "*interesting news posts/shares by my friends*", "*the news shared is so excited and related to our community*" and "*quality of posts*".

In addition to the questionnaire results, the experiment had system traced usage data of the visualization. Based on the tracked data, the number of users who performed actions on the visualization, such as clicking on bubble, dragging a category/friend circle inside, and dragging a category/friend circle outside was plotted for each day throughout the experiment (see Figure 5.3). The second questionnaire included a number of closed questions that we asked to get some quantitative data on important aspects of the visualization.

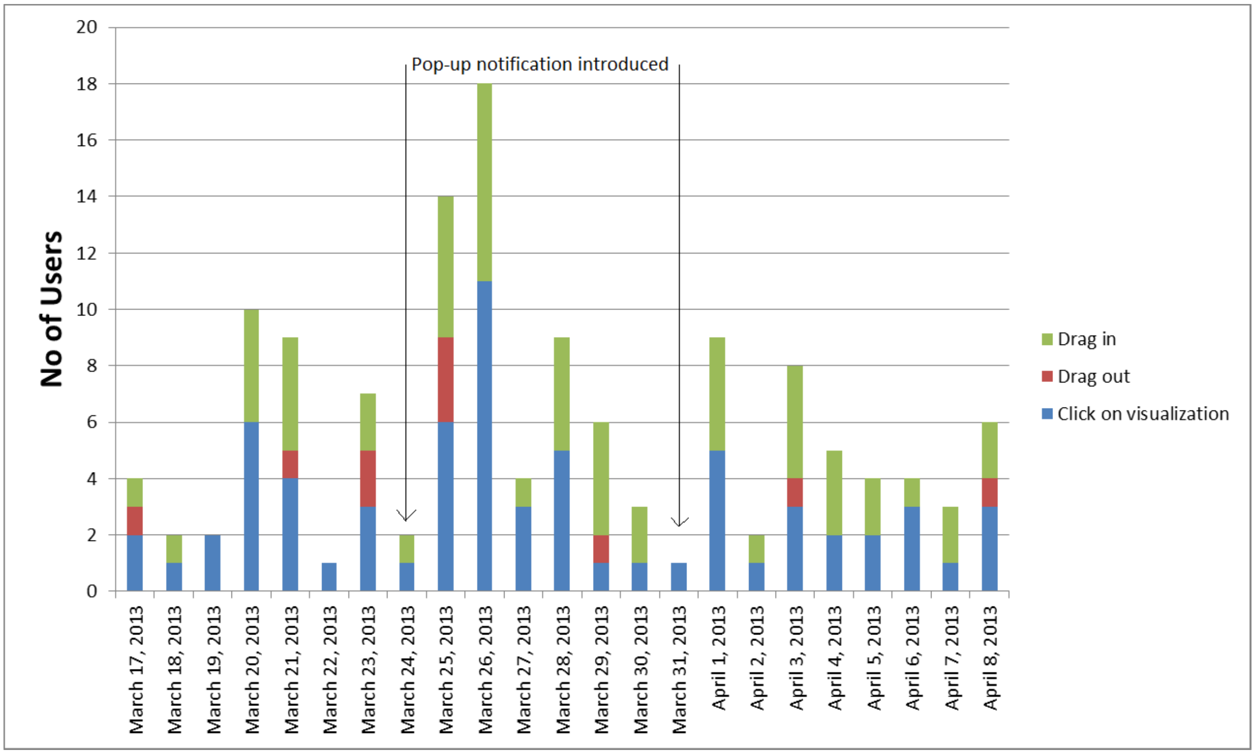


Figure 5.3: Number of users who accessed the visualization each day

Table 5.4 shows the list of closed questions with yes/no type answer aimed at finding whether the visualization has helped to create awareness, understanding, control and trust of personalized stream filtering and the results.

Table 5.4: Closed questions with yes/no answer

#	Question	YES	NO
Q1	Did you realize the system was filtering the posts from your friends away from your stream?	9 (82%)	2 (18%)
Q2	Does the visualization help you to understand how the filtering works?	8 (73%)	3 (27%)
Q3	Does the visualization give you a feeling of control over the stream of posts from your friends?	10 (91%)	1 (9%)
Q4	Does the visualization increase your trust in the filtering process?	9 (82%)	2 (18%)
Q5	Does the visualization help you trust the system more?	9 (82%)	2 (18%)

Another set of closed end questions focused on user experience of filter bubble visualization. The results are summarized in Table 5.5. On average, most of the participants answered above 5 in the scale.

Table 5.5: Results of closed questions related to user experience of filter bubble visualization (number of participants and percentage of participants who chose on a 10-level Likert-scale)

Question	**Results**									
Aesthetically pleasing	Very Low									Very High
	1	2	3	4	5	6	7	8	9	10
						1 (9%)	3 (27%)	3 (27%)	3 (27%)	1 (9%)
Friends view	Unhelpful									Helpful
					2 (18%)	1 (9%)	3 (27%)	3 (27%)	1 (9%)	1 (9%)
Category view	Unhelpful									Helpful
					1 (9%)		3 (27%)	3 (27%)	2 (18%)	2 (18%)
Awareness about hidden posts	Inadequate									Adequate
					2 (18%)	1 (9%)		4 (36%)	3 (27%)	1 (9%)
Arrangement of information on screen	Illogical									Logical
		1 (9%)			1 (9%)	1 (9%)	4 (36%)	2 (18%)	2 (18%)	1 (9%)
Manipulation of interest/friend circles (dragging in and out)	Difficult									Easy
				1 (9%)	1 (9%)	2 (18%)	2 (18%)	1 (9%)	2 (18%)	2 (18%)
Finding interest not inside your filter bubble			1 (9%)		1 (9%)	1 (9%)	2 (18%)	2 (18%)	3 (27%)	1 (9%)
Discovering new interests					2 (18%)		2 (18%)	2 (18%)	5 (45%)	
Discovering the interests of friends					1 (9%)	1 (9%)	3 (27%)	2 (18%)	4 (36%)	
Discovering the areas your friends are most interested			1 (9%)		2 (18%)		1 (9%)	3 (27%)	4 (36%)	

A set of close ended questions with Likert scale (1-10) shown in Table 5.6 were asked to evaluate the users' trust in the system. The results are summarized in Table 5.7.

Table 5.6: Closed questions for trust in the system with Likert scale

#	Question
Q6	Trust in the System before using the filter bubble:
Q7	Trust in the System after using the filter bubble:
Q8	Trust in the System after seeing the hidden posts:
Q9	Level of transparency in filtering provided by the system:

Table 5.7: Results of closed questions for trust in the system (number of participants and percentage of participants who chose on a 10-level Likert-scale)

#	**Very Low**									**Very High**
	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
Q6				2 (18%)	3 (36%)	1 (9%)	2 (18%)	1 (9%)	1 (9%)	
Q7					1 (9%)	2 (18%)	1 (9%)	4 (36%)	3 (27%)	
Q8				1 (9%)	2 (18%)			2 (18%)	4 (36%)	2 (18%)
Q9					1 (9%)		2 (18%)	5 (45%)	2 (18%)	1 (9%)

The second questionnaire also contained a set of questions aimed to evaluate the user awareness and understanding of personalized stream filtering and filter bubble visualization. Ten (91%) participants reported that they used the filter bubble visualization and one participant didn't use it. In a follow-up open-ended question, we asked "What do you think it represents?" nine out of ten participants who used the filter bubble visualization (90%) responded that they thought it represented their interest categories of social data that were displayed in their stream. Here are some of the excerpts from the answers: "*Shows my interests to different categories (category view) or to posts of friends (friends view)*", "*It represents my interest and posts I will receive", "It represents my interest category and that of others that is filtered from me"* and "*It reflects the interest a person showed in certain category of posts*". One participant mentioned specifically about the position of friend/category circles "*inside the bubble is the categories of the news I like* *while the hidden news belong to the categories outside the bubble, if friends view is selected, the same as category but for friends"* .

The second questionnaire contained open ended questions related to the user awareness and understandings of filter bubble visualization to solicit qualitative feedback. For the question "What do you think about the category view in the visualization?", three participants (27.27%) commented on what they understood about the category view: "*Category wise news/posts*" and "*I think category view is useful to visualize my choice of posts and help me to somewhat sort the posts I want to have a look on my wall.*" The remaining eight participants (72.73%) commented on the aesthetic aspect of the category view (*"nice, compact visualization", "It's good, easy to use"*).

For the question about what participants thought about the friends view, three participants (27.27%) reported that it was useful to avoid friends' social data in which they were not interested. Another three (27.27%) participants reported that they didn't use the friends view. Three participants (27.27%) said that it was a good and useful visualization. The remaining two participants (18.18%) said that it's an unnecessary view and they interpreted it wrongly. To a control question asking them to indicate a preference to one or the other view, all of the participants replied that they preferred the category view over the friends view. Five participants (45.45%) were happy with the current views and didn't suggest any other useful views. The remaining six participants (54.55%) suggested several other useful views, such as *"a mixture of both", "more subcategories! But I wonder about the trade-off with the simplicity", "time view! Popular view!", "By Date and week, and popular post -by like and comments"*, and so on.

Participants were asked about their perception about the position and size of category /friend circles in the filter bubble. For the position, seven participants (63.64%) believed that there is a meaning behind the position of the category/friend circle and all other participants (36.36%) didn't have any idea about the position of category/friend circle in the visualization. Similarly, we asked participants about the meaning of the varying size of a category/friend circle. Eight participants (72.73%) responded that there is a meaning to the size and out of those eight participants, two participants associated it with the volume of social data, two participants associated it with the interest level in the social data category, two participants associated it with the popularity of social data and the remaining two just responded with a "yes". The remaining three (27.27%) out of eleven participants didn't assume any meaning with the size of category/friend circle.

The last few questions in the series of open ended questions aimed at evaluating the controls given to the user in filter bubble visualization: whether they were actually used, considered useful and usable. The first question was about whether participants dragged the category/friend circles inside the bubble. Nine participants (82%) stated that they have dragged the category/friend circles from outside the filter bubble to inside the filter bubble while other two participants didn't do so. In a follow-up question, those who answered "yes" for dragging inside, were asked about the effect that they noticed after dragging a category/friend circle inside the filter bubble. Eight (88.89%) out of the nine participants said that there is an effect after dragging a category/friend inside the bubble. In particular, four participants out of those eight said that their interest areas expanded and more social data appeared in their stream. Only one participant out of those who tried dragging the circle inside said that there was no effect after the action. Similarly, a question was asked about dragging a category/friend circle outside the filter bubble. Four participants (36%) stated that they had tried dragging category/friend circle outside the filter bubble and noticed a change in their stream; particularly social data got filtered away. Other seven participants (63.64%) stated that they hadn't tried dragging a category/friend circle outside the filter bubble.

5.5 Discussion

The results of the first questionnaire show that the users had a good experience with the filtering mechanism; the participants found the social data in the stream becoming more interesting with time, as the system learned about their interests from their implicit feedback. On average, the participants felt that there were a normal number of social data in the stream at any time, which suggests that the filtering mechanism did not permit a social data overload. The participants seemed to have a partial understanding about how the filtering mechanism worked, and this is an area where further improvement is needed, since according to the results, the trust in a filtering system depend on both understating the mechanism, and more importantly, on being able to control it. Most of the participants (73.73%) did not show concern about the privacy of their data. This is natural in this pilot study, due to the small closed user community. But still there were some participants (27.27%) who had concern about privacy and wanted to host their data on their own MADMICA node. This shows that some participants are finding attractive the possibility of running their own MADMICA node to preserve the privacy of their data. This is encouraging for the future of decentralized social networks, yet one needs to keep in mind, that the participants in this study were computer scientists. A member of the general audience who is less technically oriented, may feel intimidated or incompetent to maintain her own data node. Generally, the usability feedback about the MADMICA system was positive, and showed that it was easy to use the system and keep track of posts in the stream.

The results of the second questionnaire show that participants are aware of the filtering. The following results provide enough evidence to support hypothesis 1: Most of the participants (80%) showed understanding about the representation of filter bubble visualization, knowing that the system is filtering their data stream (82%). The majority (73%) said that visualization helped them to understand the filtering. The results show that 63.64% of the participants believed that there is a meaning in the position of the category/friend circle with respect to the filter bubble, so it is evident that the majority understood the general metaphor of the visualization. More than 50% of the participants said visualization provided adequate awareness about the hidden social data. Even though eight participants (72.73%) responded that there is a meaning to the size of the circles, only two participants understood that the size denotes the volume of social data represented by the category or originated by the user represented by the circle. So the design needs improvement with respect to using size of the category/friend circles in the graphical language.

From the results of the open ended questions related to the category view and the friends view, we can see that the category view was more effective than the friends view in creating awareness and understanding of the personalized stream filtering and also the category view seems to be the most preferred view. So the friends view needs to be improved. The results to the open ended questions that aimed evaluating the control given to the user to manipulate the visualization and respectively, the filtering mechanism show that the participants felt they had a feeling of control over their stream and the filtering (91% of participants agreeing with closed question Q3 in Table 1). Thus it seems we have sufficient qualitative evidence in support of hypothesis 1.

Hypothesis 1 can also be supported by the results of the user actions graph (see Figure 5.3). The graph in Figure 5.3 depicts the user actions performed on the filter bubble visualization over the time period of the experiment. The beginning of the graph period can be marked as the learning phase where users get familiar with the drag and drop of category/friend circles. Then there is a sudden spike in user actions because we introduced a popup window to notify the users that social data are filtered away from the stream and introduce the visualization to gain back control of filtering. After one week, when the necessary awareness about the visualization has been created, the popup notification was turned off. Even after the notification has been turned off, from the graph in Figure 5.3, still we could see users checking the filter bubble visualization and dragging the circles in and out. This shows that the filter bubble visualization has been used to control the personalized filtering. Interestingly, most of the actions were “dragging in” categories or people, which means the participants counter-acted the filtering mechanism. There were a few “drag out” actions throughout the experiment and they were targeted at one particular participant who was the most active one in the group and was probably perceived as a spammer at a certain moments of high traffic by some of his/her friends.

The study results also provide evidence to support the hypothesis 2. The results of quantitative questions Q4 and Q5 show that visualization has helped to increase the users’ trust in the filtering mechanism and the system. In addition to that, comparing the results of Q6 and Q7 provides more clear evidence to support the hypothesis 2 i.e. most of the participants (63%) rated below 6 (more towards low) in scale for the trust in the system before using the filter bubble visualization. But after seeing the filter bubble visualization, 72% of participants rated above 6 in scale (more towards high) for the trust in the system. Moreover, 72% of participants have rated high (above 7 scale) their trust in the system after seeing the hidden posts provided by the visualization and most of the participants (72%) rated the level of transparency as high as possible (above 7). These results of Q8 and Q9 also support the hypothesis 2.

The results shown in Table 5.5 provide answers to questions about the general user experience with the system. They support hypothesis 3, because, 90% of participants have found that the filter bubble visualization is aesthetically pleasing by rating it above 6; 72% of participants have found that the friends view was helpful, 90% found that category view was helpful. In addition, 72% of participants rated that visualization has provided adequate awareness about hidden social data, 81% of participants found that the information on the screen was logically arranged, 63% of participants said dragging the category/friend circles in and out of the filter bubble was easy, 72% said finding an interest which is not inside their filter bubble was easy, 81% said discovering new interests and discovering the interests of friends were also easy and 72% said that discovering in which areas their friends are most interested was also easy. So all in all the results in Table 5.5 suggest that the user experience with the MADMICA was enhanced by the visualization. Moreover, results of finding an interest which is not inside your filter bubble and discovering new interests clearly shows that users are more aware of the filtering due to the visualization and are interested, able and willing to manipulate it to ensure that they will not be trapped inside a bubble world within the limited boundaries of their interests.

5.6 Summary

The results of the pilot study show that the filter bubble visualization makes the users aware of the filtering mechanism, engages them in actions to correct and change it, and as a result, increases the users' trust in the system. The next chapter presents a qualitative user study of the visualization which tries to understand the in-depth user perceptions about the visualization.

CHAPTER 6
EVALUATION: QUALITATIVE USER STUDY

The chapter presents a qualitative study which was carried out in order to understand in-depth the user perception of the filter bubble visualization i.e. what do users think about the visualization. Five (5) participants from different departments in the university took part in this study.

6.1 Experimental Setup

The study was carried out in a lab environment where users were given computers to use the MADMICA system and the visualization. The subjects were 5 university students (international students) from different fields of study such as public education, public health and statistics. They were recruited through a mailing list of potential subjects for HCI studies. First, the users were given some introduction to MADMICA and then about the filter bubble problem. After the introduction, users were given instructions to get familiar with the MADMICA newsfeed homepage and the filter bubble visualization for 10 minutes. Once they have explored the system, an interview was conducted. The interview consists of a set of tasks related to with 15 different views that are generated using the filter bubble visualization. They were asked to interact with the systems and think aloud, the users' actions were observed and recorded and the users' voice responses were recorded. The views in the questionnaire were generated to collect the perceptions about the visualization's main goals: providing awareness, explanation and control. Moreover, the views included both the category view and friends view.

6.2 Methods

The recorded users' voice responses were imported into NVivo software [36], which is a platform for qualitative research analysis. Then the voice responses were transcribed into text. With the help of the NVivo software, thematic analysis was carried out to identify the desirable and undesirable perceptions of the visualization. Thematic analysis categorizes qualitative data into themes. It encodes the qualitative information into codes that act as labels for sections of data [6]. The users' responses were coded by the researcher and the codes were grouped into three: position of circle, size of circle and drag action. While coding, the number of references for each code was also recorded i.e. the frequency of that code in the transcript of users' responses. Then based on the decision flow chart shown in Figure 6.1, the number of desirable references and undesirable references was calculated,

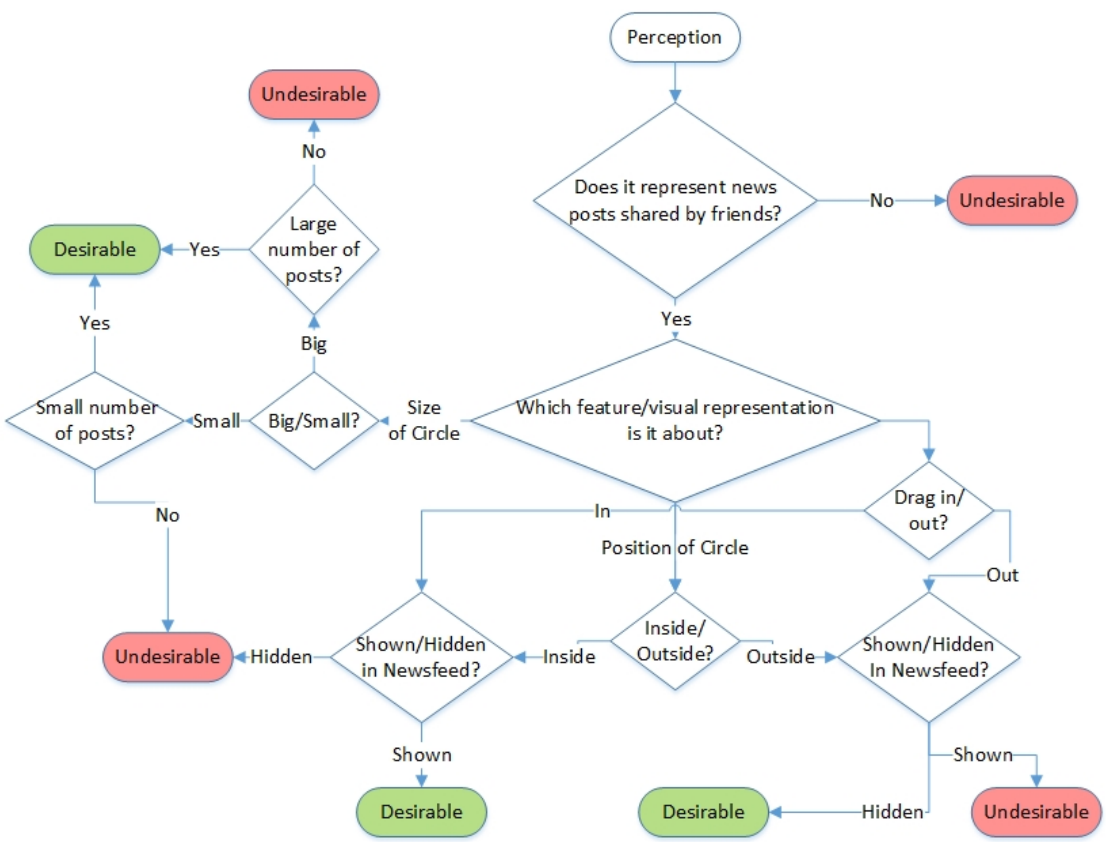


Figure 6.1: Flow chart of desirable/undesirable perceptions

6.3 Results

The thematic analysis results are summarized in Table 6.1. The desirability percentage for a perception category is calculated as the number of references that are desirable in that perception category divided by the total number of references for the position of circle visual representation multiplied by 100.

Table 6.1: Thematic Analysis Results

Feature/Visual Representation	**Perception Category**	**Sources (number of users)**	**References (desirable: undesirable)**	**Desirability percentage (%)**	**Undesirability percentage (%)**
Position of circle (friend/category)	Common interest	4	13 (10:3)	9.26 (10/108)	2.78
	Friends' interest	4	40 (16:24)	14.81	22.22
	Friends' sharing	5	25 (18:7)	16.67	6.48
	Interaction with newsfeed	1	3 (3: 0)	2.8	0
	User's interest	5	23 (19:4)	17.59	3.7
	Relationship	3	4 (2:2)	1.85	1.85
Size of circle	Number of posts	5	7 (6:1)	37.5 (6/16)	6.25
	Frequency of sharing	2	2 (2:0)	12.5	0
	Friends' interest	2	5 (4:1)	3.7	6.25
	Common interest	1	2 (0:2)	0	12.5
Drag action	Common interest	4	5 (4:1)	57.14 (4/7)	14.29
	Relationship	1	2 (0:2)	0	28.57

Regarding the position of circle visual representation, 108 total references were made i.e. users mentioned 108 times in all of their responses together regarding the position of circles relative to the contour of the big bubble. As shown in Table 6.1, the position of circles relative to the contour of the big bubble represents the user's interest, is the most referred (17.59 %) desirable perception category about the position of circle. Some excerpts from the transcript for the user's interest perception category follow: *"categories outside the bubble represent the posts that the user doesn't want to see", "categories inside the bubble represent my interests", "categories inside the bubble represent users main interests for the selected duration", "All the categories outside the bubble represent that none of user's friends posts are related", "categories outside the bubble* *represent the areas outside of my interest for that period",* and *"categories inside the bubble represent that the user wants to focus on them"*. The least referred (1.85%) desirable perception category regarding the position of circle is relationship. Some excerpts from the transcript for the least referred desirable perception follow: *"friend circle outside the bubble for a category doesn't mean that the user unfriended with that friend", "having some categories inside the bubble for last month for a friend might mean an acquaintance relationship",* and *"friend relationship is maintained regardless of user's friends are outside the bubble"*. Some excerpts from the transcript for the least referred desirable perception follows: *"friend circle outside the bubble for a category doesn't mean that the user unfriended with that friend", "having some categories inside the bubble for last month for a friend might mean an acquaintance relationship",* and *"friend relationship is maintained regardless of user's friends are outside the bubble".*

The most referred (22.22%) undesirable perception category is the friend's interest. But here only 4 users have referred this whereas in the most desirable perception all the 5 users referred it at some point in the transcript. Following are some excerpts from the transcript regarding the most undesirable perception: *"categories inside the bubble represent friend's interest and outside represents not interested"* and *"friend circle more in the middle more interest in the category selected"*. Like the least referred desirable perception, the least referred undesirable perception is relationship and the excerpt follows: *"friend circle outside the bubble represents unfriending".*

The number of posts related perception category is the most referred (37.5%) desirable perception for the size of the circle. Users perceive it as follows: *"bigger circle for friend/category represents more number of posts and small for less number of posts"*. As for the least referred (3.7%) desirable perception for the size of the circle, users perceive it as friend's interest i.e. *"larger circle for category means the selected friend has more interest on that category*". In case of undesirable perception category, the most referred (12.5%) one is common interest (*"small friend circle means more common interest between the user and friends"* and *"larger circle represents user has less interest on that friend"*) and the least referred (6.25 %) one is number of posts and friends' interest *("small circle means actually posted and big circle means less posted"* and *"bigger circle outside the bubble represents less interest of friend on that category"*).

There are two perception categories that emerged by the thematic analysis for drag action: common interest and relationship. Common interest is the most referred desirable perception category (57.14%) and there are no references for relationship on desirable perception. The excerpt from the transcript for common interest perceptions follows: *"dragging a category inside means to share more on that category with the friend", "dragging in may represent my future interest", "drag out because I don't want to have common interest", "drag out means lost interest in that category from that friend",* and *"drag all the friends outside the bubble means I want to ignore all the news from them".* The most and least referred undesirable perception category for the drag action are relationship (28.57%) and common interest (14.29%) respectively. Excerpt related to relationship is *"dragging outside a friend/category means unfriend"* and for the common interest is *"drag inside represents forcing the friend to take interest on that category".*

6.4 Discussion

The results of the qualitative data suggests that the subjects had both perceptions which are desirable and undesirable. Desirable perceptions (62.96%) regarding the position of circle had more references than undesirable perceptions (37.04%). This shows that most of the time the visualization users were aware of and had a good understanding about the filtering. In particular, the emergent codes such as common interests, friends' interest, friends' sharing, interaction with newsfeed, user's interest and relationship from the thematic analysis clearly show that the users have some understanding about the filtering.

On the other hand the users also had undesirable perceptions. This could be due to poor graphical language of the visualization and interface as a whole. For example, the reason that friends' interest was perceived as the most undesirable perception category, could be the poor label texts of the dropdown menus which are used to view the filter bubble in different dimensions. During the experiment, the labels for the first two dropdown menu were "Friend(s)" and "Category(s)". This creates a false perception when "Charlie" was selected in the menu "Friend(s)" and the category "All" was selected as in the menu "Category(s)" , that Charlie's interests were shown inside the bubble and what lies outside the bubble were not the interests of Charlie. The correct perception would have been that the user and Charlie share the interests shown as circles within the bubble, while the circles outside the bubble are Charlie's interests that the user doesn't share. To correct this misperception, as a result of the qualitative study, the labels were changed into "From Friend(s)" and "On Category(s)" (shown in Figure 6.2 depicting the updated version) before the quantitative study. In addition to that, a scalability issue of the visualization was resolved by restricting the selection to not allow choosing at the same time the options "All" from both menus "From Friend(s):" and "On Category(s):"

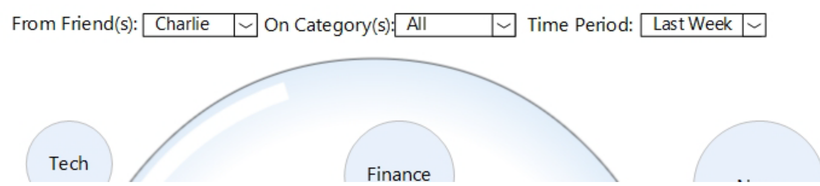


Figure 6.2: Updated version of the visualization

The size of the circle is another indicator for creating awareness about the filtering, i.e. a bigger size of the circle outside the filter bubble would let the users know that there are many posts that have been filtered out by the system on that category from that friend. Having 75% of the perceptions for size of the circle classified as desirable shows that it is intuitive enough to create an awareness about the filtering. The 25% of perceptions regarding the size of the circle classified as undesirable shows that the graphical language needs improvement. For example, it would be clearer if there is a number shown with the varying size. Moreover, the false perceptions of common interest for the size of the circle showed that users may have wrong perceptions about the meaning of the size of circles. For example, that the size of the circles represent the interests of the friends i.e. a smaller circle means that the friend has less interest on that category.

The drag action has 57.14% of desirable perceptions and 42.86% of undesirable perceptions. Despite the small difference, considering the number of users who referred the perception gives some clear indication that the majority of the participants (60%) were able to understand the control functionality of the filter bubble visualization. Though the perceptions were classified as desirable and undesirable, both of them helped to get more insight about the users perceptions about the visualization, improve the visualization and helped to prepare the questions and answers for the questionnaire of the quantitative study, presented in the next section.

6.5 Summary

The qualitative study reveals generally higher proportion of desirable user perceptions for the awareness, explanation and control of the filtering and the filter bubble provided by the interactive visualization. As a result of the qualitative study, the labels of the filters in the visualization were changed into "From Friend(s)" and "On Category(s)". The next chapter presents a large-scale quantitative study which was carried out to evaluate the understandability of the visualization and whether the users understand that the visualization provides awareness, explanation and control of filtering and the filter bubble. In addition to that, a study of evaluating the intuitiveness of the visualization by comparing it to the same interactive visualization provided with guided help, is also presented.

CHAPTER 7
EVALUATION: QUANTITATIVE USER STUDY

This chapter presents a large-scale quantitative study which was carried out to evaluate whether the users understand that the visualization provides awareness, explanation and control of filtering and the filter bubble. In addition to that, the study aimed to evaluate the intuitiveness of the visualization by comparing it to the same interactive visualization augmented with guided help. The study was conducted as an online survey and 326 participants from different parts of the world participated in the study.

7.1 Hypotheses

The goal of this user study was to find out if the visualization is understandable and intuitive, if it creates awareness and provides explanation of the personalized stream filtering mechanism and ability to control it to alleviate the filter bubble. So the evaluation aims at testing the hypotheses given in Table 7.1.

Table 7.1: Quantitative study hypotheses

Group	Hypothesis
Group 1 – No help provided	**H1A:** Users understand that the visualization provides *awareness* of the filtering and the filter bubble.
	H1B: Users understand that visualization provides *explanation* of the filtering and the filter bubble.
	H1C: Users understand that visualization provides *control* of the filtering and the filter bubble.
	H1D: Users understand the visualization and its functions.
Group 2 – Help text provided with the visualization	**H2A:** Users understand that the visualization provides *awareness* of the filtering and the filter bubble.
	H2B: Users understand that visualization provides *explanation* of the filtering and the filter bubble.
	H2C: Users understand that visualization provides *control* of the filtering and the filter bubble.
	H2D: Users understand the visualization and its functions.
Combined Group	**H3A:** Users from group 2 have more clear understanding that the visualization provides *awareness* of the filtering and the filter bubble than that of users from group 1.
	H3B: Users from group 2 have more clear understanding that the visualization provides *explanation* of the filtering and the filter bubble than that of users from group 1.
	H3C: Users from group 2 have more clear understanding that the visualization provides *control* of the filtering and the filter bubble than that of users from group 1.
	H3D: Users from group 2 have more clear understanding about the visualization and its functions than that of users from group 1.

7.2 Methods

This section presents the experimental setup of the quantitative study, the user participation and the tools used to evaluate the study.

7.2.1 Experimental Setup

The study was carried out as an online survey, which had the interactive visualization embedded into the survey, so that users could explore it and get some hands-on experience with it before answering the survey. Participants were randomly divided into two groups of 163 participants. The first group explored the filter bubble visualization by themselves without any help and the second group explored the visualization after reading through a help text (Time spent on the visualization with help text part of the survey was measured to ensure that they actually read the help text). The survey can be found in the Appendix C. After the participants consented to participate in the online survey, they were given some introduction about the MADMICA social network and the filter bubble problem in general. Both groups of participants were presented with a sample newsfeed homepage embedded in the survey, so that users could actually browse through the newsfeed without leaving the survey page. The sample newsfeed contained around 15 newsfeed items on 5 different categories such as Health, News, Movies, Music and Sports from five different friends named Alice, Bob, Charlie, Dave and Frank. The participants were given instructions to assume that the aforementioned people are their friends in MADMICA social network and to browse through the newsfeed homepage as they would do in Facebook. In addition to this, the newsfeed did not show around 7 posts out of those five categories from different friends i.e. the system filtered out some of the posts. Then the first group of users were presented with the interactive visualization exactly as in the MADMICA system and were instructed to explore the visualization without any help. The second group of users were presented with a help text that is similar to the help documentation found in any software, which explained the functionalities of the visualization and then they were asked to play with the visualization. Finally both groups were directed to the questionnaire to answer the questions.

7.2.2 User Participation

The online survey was conducted using Amazon Mechanical Turk (MTurk) which is a popular crowd-sourced participant pool. The data quality was ensured by placing attention check questions (ACQs) and restricting participation to MTurk workers with certain qualifications [39].

The suggested qualification among researchers to ensure data quality was to allow participants who have the HIT (Human Intelligent Task) Approval Rate (%) for all Requesters' HITs greater than or equal to 95 [39]. But we set even higher qualification to ensure the high data quality as follows: HIT Approval Rate (%) for all Requesters' HITs greater than or equal to 98% AND Number of HITs Approved greater than or equal to 5000. The data collection continued for 1 week and reached our target sample of 400 for both group together. The total time spent on the survey and the total time spent on the visualization were plotted for each participant response as shown in Figure 7.1 and Figure 7.2. Those who spent less than 10 minutes in the survey and less than 2 minutes in the visualization were assumed to have provided invalid responses and removed from the dataset to ensure the data quality. Then we analyzed the data again and checked the ACQ for validity and as a result, 163 valid responses for each group were collected (326 total valid responses). For each participant with a valid response, a compensation of 1$ was paid, which is a fairly high rate for an approximately 30-45 minutes long study on MTurk.

Figure 7.1: Total time spent on the survey

Figure 7.2: Total time spent on the visualization

7.2.3 Tools

The questionnaire contained 25 questions. The questions were grouped according to the metrics that they intend to measure. The metrics for understandability of the visualization are adapted based on the International Standards for Software Quality Evaluation [22]. Table 7.2 summarizes the metrics chosen for measuring the understandability of the visualization [22].

Table 7.2: Understandability Metrics

Metric Name	Purpose	Formula	Interpretation of measured value
Evident Functions	How many functions users were able to identify by exploring the visualization?	X = A / B A = Number of functions identified by the user B = Total number of actual functions	0<=X<= 1 Closer to 1.0 is better.
Function understand-ability	How many functions users were able to understand correctly by exploring the visualization?	X= A / B A= Number of functions whose purpose is correctly described by the user B= Number of functions available	0<=X<= 1 Closer to 1.0 is better.
Understandable input and output	Can users understand the input and the output of the visualization?	X= A / B A= Number of input and output data items which user successfully understands B= Number of input and output data items available from the visualization	0<=X<= 1 Closer to 1.0 is better
Completeness of description	How many functions are understood after reading the help text of the visualization?	X = A / B A = Number of functions understood B = Total number of functions	0<=X<= 1 Closer to 1.0 is better

Each group was individually tested for their understandability of the visualization. Moreover, a combined test was carried out to evaluate the intuitiveness by comparing the two groups' understanding. The understandability in group 1 is measured using the following metrics: Evident Functions, Function Understandability and Understandable Input & Output. Similarly, group 2 also uses most of the metrics used by group 1, but the Evident Functions metric is replaced with the Completeness of Description metric because of the help text that is provided with the visualization. There are 3 independent variables: awareness, explanation and control to assess the understandability of the visualization in both groups. Each of the independent variables was evaluated using the metrics given in Table 7.2 (first 3 metrics for group 1 and last 3 metrics for group 2) i.e. understandability of each independent variable was calculated. In addition to that, the overall understandability (referred as understandability hereafter) for each group was also calculated using the understandability metrics. Six (6) questions (2 Yes/No and 4 Multiple Choice Questions) were used to evaluate each of the independent variables. Altogether, there were 18 questions that were used to evaluate the overall understandability with 6 questions for each metrics in both groups. Our original hypotheses mentioned in section 7.1 were converted into the statistical form with the corresponding null hypothesis (see Table 7.3).

Table 7.3: Statistical Hypotheses

Hypothesis	H0 (null)	H1 (alternative)
H1A	$\mu_{\text{Awareness}} \leq 0.5$	$\mu_{\text{Awareness}} > 0.5$
H1B	$\mu_{\text{Explanation}} \leq 0.5$	$\mu_{\text{Explanation}} > 0.5$
H1C	$\mu_{\text{Control}} \leq 0.5$	$\mu_{\text{Control}} > 0.5$
H1D	$\mu_{\text{Understandability}} \leq 0.5$	$\mu_{\text{Understandability}} > 0.5$
H2A	$\mu_{\text{Awareness}} \leq 0.5$	$\mu_{\text{Awareness}} > 0.5$
H2B	$\mu_{\text{Explanation}} \leq 0.5$	$\mu_{\text{Explanation}} > 0.5$
H2C	$\mu_{\text{Control}} \leq 0.5$	$\mu_{\text{Control}} > 0.5$
H2D	$\mu_{\text{Understandability}} \leq 0.5$	$\mu_{\text{Understandability}} > 0.5$
H3A	$\mu_{\text{Awareness in Group 1}} = \mu_{\text{Awareness in Group 2}}$	$\mu_{\text{Awareness in Group 1}} \neq \mu_{\text{Awareness in Group 2}}$
H3B	$\mu_{\text{Explanation in Group 1}} = \mu_{\text{Explanation in Group 2}}$	$\mu_{\text{Explanation in Group 1}} \neq \mu_{\text{Explanation in Group 2}}$
H3C	$\mu_{\text{Control in Group 1}} = \mu_{\text{Control in Group 2}}$	$\mu_{\text{Control in Group 1}} \neq \mu_{\text{Control in Group 2}}$
H3D	$\mu_{\text{Understandability in Group 1}} = \mu_{\text{Awareness in Group 2}}$	$\mu_{\text{Understandability in Group 1}} \neq \mu_{\text{Awareness in Group 2}}$

As shown in Table 7.3, the mean value of understandability was considered for testing the hypotheses. According to Table 7.2, if the value of understandability is 0, then the users did not completely understand the visualization and if the value is 1, then the users completely understand the visualization. Moreover, the closer to 1, the users understand the visualization better. So the test mean value μ is set as 0.5 according to the scale of metrics used to measure the understandability. The null hypothesis was set as the mean value μ of understandability is less than or equal to 0.5 i.e. users do not have a good understanding about the visualization. The research hypothesis was set as the mean value μ of understandability is greater than 0.5 i.e. users do have a good understanding about the visualization compared to that of who are less than 0.5.

7.3 Results

This section presents the results of the quantitative study. Section 7.3.1 presents the results of the reliability test which is used to ensure the reliability of the questionnaire. Section 7.3.2 presents the results of normality test which is used test whether the data is normally distributed so that the chosen statistical test can be performed. Then the rest of the sections present the results of hypotheses tests.

7.3.1 Reliability Test

The internal consistency (reliability) of question items was measured using Cronbach's alpha. Higher value of a reliability coefficient (Cronbach's alpha) is associated with lower random error and greater measurement of the true score of the understandability. The acceptable value of Cronbach's Alpha should be the range of 0.70 to 0.95 [4]. The rules of thumb when considering Cronbach's Alpha value are as follows: greater than 0.9 means excellent, greater than 0.8 means good, greater than 0.7 means acceptable, greater than 0.6 means questionable, greater than 0.5 means poor, and less than 0.5 is unacceptable [18].

The measured values for Cronbach's alpha for both group are summarized in Table 7.4. Both the values are in the acceptable range i.e. the question items and scale are reliable.

Table 7.4: Reliability Statistics

Group	Cronbach's Alpha	Cronbach's Alpha Based on Standardized Items	N of Items
Group 1	.700	.698	18
Group 2	.755	.737	18

During the reliability test using SPSS statistics software, it suggests whether we can achieve more reliability by removing any of the question items. The suggested values are summarized in Table 7.5 and Table 7.6 for group 1 and group 2 respectively.

Table 7.5: Item-Total Statistics for Group 1

Question	Cronbach's Alpha if Item Deleted
AQ1	.699
AQ6	.700
AQ11	.670
AQ13	.666
AQ18	.675
AQ20	.680
EQ9	.710
EQ10	.728
EQ15	.681
EQ17	.673
EQ23	.679
EQ25	.687
CQ4	.693
CQ8	.695
CQ14	.679
CQ16	.693
CQ22	.676
CQ24	.676

Table 7.6: Item-Total Statistics for Group 2

Question	Cronbach's Alpha if Item Deleted
AQ1	.762
AQ6	.749
AQ11	.733
AQ13	.729
AQ18	.737
AQ20	.739
EQ9	.769
EQ10	.765
EQ15	.729
EQ17	.726
EQ23	.716
EQ25	.760
CQ4	.754
CQ8	.755
CQ14	.740
CQ16	.735
CQ22	.738
CQ24	.740

The question numbers are named with a prefix before the 'Q'. The prefix stands for the variable it is measuring (A: Awareness, E: Explanation, C: Control). For example, EQ10 means it is question number 10 which measures the variable - Explanation. Moreover, the questionnaire can be found in the Appendix C. As shown in Table 7.5, the highlighted row suggest that if question EQ10 is removed, the questionnaire could reach the maximum possible reliability of 0.728 which is not a considerable reliability improvement because to become "good" reliability, the value should reach 0.8. In addition to that, according to the Table 7.5, removing any other question item from the questionnaire will reduce the reliability below 0.7 which will fall into the unacceptable range. For the group 2, as shown in Table 7.6, the highlighted row suggests that the reliability value can reach its maximum possible value of 0.769 by removing the question item EQ9 but the improvement is not good enough, as before because considerable improvement in reliability can be seen if it reaches the "good" range of the coefficient. Unlike the group 1, a small improvement in reliability can be achieved by removing any question as shown in Table 7.6.

7.3.2 Normality Test

The assessment of the normality of the data is a prerequisite and essential to t-tests. The Normal Q-Q (Quantile - Quantile) plot for understandability was generated using SPSS (see Figure 7.3 and Figure 7.4).

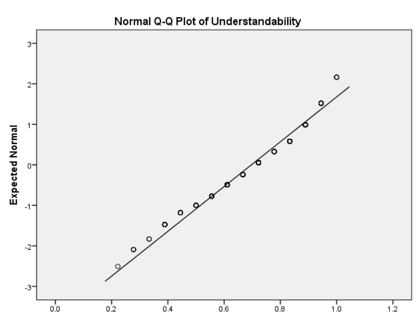


Figure 7.3: Normality Q-Q Plot of Understandability – Group 1

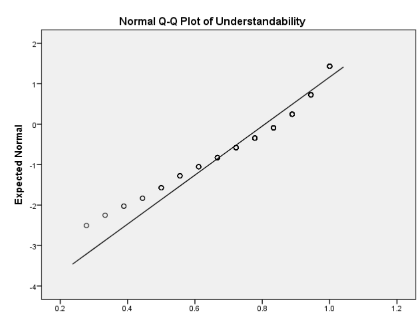


Figure 7.4: Normality Q-Q Plot of Understandability – Group 2

If the data are normally distributed, the data points will be close to the diagonal line. If the data points move away from the line in a non-linear way then the data are not normally distributed [58].

As we can see from the Normality Q-Q Plot shown in Figure 7.3 and Figure 7.4, the data are normally distributed in both groups because the data points stay close to the diagonal line.

7.3.3 Hypothesis Test – Group 1

One-sample t-test was used to determine whether the mean of a particular data set is different from the particular value. Before doing the t-tests, the following 4 assumptions were made: understandability is measured at the ratio level, the collected data are independent which means that there is no relationship between the observations, there are no significant outliers in the data, and the understandability is approximately normally distributed [37]. Then the t-tests were conducted for the 4 hypothesis tests in group 1 and the results are summarized in the Table 7.7.

Table 7.7: Hypothesis Analysis for Group 1

Test	Variable	Mean	2-tailed t	Degree of freedom (df)	1-tailed Critical t	1-tailed t < 2-tailed t	Means are in correct order	Alternative Hypothesis Accepted
1	**Awareness**	.7117	11.358	162	1.6543	**YES**	**YES**	**YES**
2	**Explanation**	.6176	6.953	162	1.6543	**YES**	**YES**	**YES**
3	**Control**	.7607	14.824	162	1.6543	**YES**	**YES**	**YES**
4	**Understandability**	.6967	13.884	162	1.6543	**YES**	**YES**	**YES**

The first t-test was conducted for the hypothesis H1A (null: μ Awareness ≤ 0.5, alternative: μ Awareness > 0.5) defined in Table 7.3. The Mean understandability of awareness ($M = 0.7117$, $SD = 0.2379$) was higher than the tested understandability value of 0.5, **a statistically significant mean difference of 0.21, 95% CI [0.18 to 0.25]**, $t(162) = 11.358$, $p < .001$. Similarly, the t-tests for hypothesis H1B (null: μ Explanation ≤ 0.5, alternative: μ Explanation > 0.5), H1C (null: μ Control ≤ 0.5, alternative: μ Control > 0.5) and H1D (null: μ Understandability ≤ 0.5, alternative:

μ Understandability > 0.5) were conducted and the results follow respectively, the Mean understandability of explanation about the filtering mechanism ($M = 0.6176$, $SD = 0.2159$) was higher than the tested understandability value of 0.5, **a statistically significant mean difference of 0.12**, **95% CI [0.08 to 0.15],** $t(162) = 6.953$, $p < .001$, the Mean understandability of control ($M = 0.7607$, $SD = 0.2246$) was higher than the tested understandability value of 0.5, **a statistically significant mean difference of 0.26**, **95% CI [0.23 to 0.30]**, $t(162) = 14.824$, $p < .001$ and the Mean understandability of visualization ($M = 0.6967$, $SD = 0.1808$) was higher than the tested understandability value of 0.5, **a statistically significant mean difference of 0.20, 95% CI [0.17 to 0.23]**, $t(162) = 13.884$, $p < .001$. In all four tests, there were a *statistically significant* difference between means ($p < .001$) and, therefore, we can reject the null hypotheses defined in Table 7.3, and accept the alternative hypotheses.

7.3.4 Additional Hypothesis Test on Graphical Language – Group 1

The key graphical language constructs of this visualization are,

1. The relative position of user's circles to the bubble (inside / outside)

2. The size of the users' circles (larger - more posts)

3. Dragging user circles in and out (showing / filtering away)

In addition to the above 3 constructs, another potential construct was identified from the qualitative study as follows: the position of circles inside the bubble (closer to the center or to the periphery). All the 3 other constructs were as part of each function of the visualization (providing awareness, providing explanation, and providing control) and were tested for statistical significance. In order to test whether users interpret this fourth construct or not, we included the answers based on this construct for two of the questions in the survey. During the analysis, we created a score for users based on how many out of the 2 questions they did not select this construct as an answer. Then the hypotheses were formed as follows: H0: $\mu_{Score} \leq 0.5$, H1: $\mu_{Score} > 0.5$. One sample t-test was conducted and the results are as follows: the Mean score for not selecting the graphical construct (M = 0.9571, SD = 0.1405) was much higher than the test score value of 0.5, **a statistically significant mean difference of 0.46, 95% CI [0.44 to 0.49]**, t (162) = 41.523, p < .001.There were a statistically significant difference between means (p < .001) and, therefore, we can reject the null hypothesis, and accept the alternative hypothesis.

7.3.5 Hypothesis Test – Group 2

One-sample t-test was conducted on the group 2 dataset. Before doing the t-tests, the following 4 assumptions were made: understandability is measured at the ratio level, the collected data are independent which means that there is no relationship between the observations, there are no significant outliers in the data, and the understandability is approximately normally distributed [37]. Then the t-tests were conducted for the 4 hypothesis tests in group 2 and the results are summarized in the Table 7.8.

Table 7.8: Hypothesis Analysis for Group 2

Test	Variable	Mean	2-tailed t	Degree of freedom (df)	1-tailed Critical t	1-tailed t < 2-tailed t	Means are in correct order	Alternative Hypothesis Accepted
1	**Awareness**	.7996	18.462	162	1.6543	**YES**	**YES**	**YES**
2	**Explanation**	.7403	14.413	162	1.6543	**YES**	**YES**	**YES**
3	**Control**	.8834	28.993	162	1.6543	**YES**	**YES**	**YES**
4	**Understandability**	.8078	23.802	162	1.6543	**YES**	**YES**	**YES**

The first t-test was conducted for the hypothesis H2A (null: μ Awareness ≤ 0.5, alternative: μ Awareness > 0.5) defined in Table 7.3. The Mean understandability of awareness (M = 0.7996, SD = 0.2072) was higher than the tested understandability value of 0.5, **a statistically significant mean difference of 0.30, 95% CI [0.27 to 0.33]**, t (162) = 18.462, p < .001. Similarly, the t-tests for hypothesis H2B (null: μ Explanation ≤ 0.5, alternative: μ Explanation > 0.5), H2C (null: μ Control ≤ 0.5, alternative: μ Control > 0.5) and H2D (null: μ Understandability ≤ 0.5, alternative: μ Understandability > 0.5) were conducted and the results follow respectively, the Mean understandability of explanation about the filtering mechanism (M = 0.7403, SD = 0.2128) was higher than the tested understandability value of 0.5, **a statistically significant mean difference of 0.24, 95% CI [0.21 to 0.27]**, t (162) = 14.413, p < .001, the Mean understandability of control (M = 0.8834, SD = 0.1689) was higher than the tested understandability value of 0.5, **a statistically significant mean difference of 0.38, 95% CI [0.36 to 0.41]**, t (162) = 28.993, p < .001 and the Mean understandability of visualization (M = 0.8078, SD = 0.1651) was higher than the tested understandability value of 0.5, **a statistically significant mean difference of 0.31, 95% CI [0.28 to 0.33]**, t (162) = 23.802, p < .001. In all four tests, there were a *statistically significant* difference between means (p < .001) and, therefore, we can reject the null hypotheses defined in Table 7.3, and accept the alternative hypotheses.

7.3.6 Hypothesis Test – Combined

Independent-samples t-test was conducted to compare the means between group 1 and group 2 on the same continuous dependent variable (understandability). Before doing the independent samples t-tests, the following 6 assumptions were made: understandability is measured on a continuous scale; the dataset consists of two categorical, independent groups; observations are independent, which means that there is no relationship between the observations in each group or between the groups themselves; there are no significant outliers; understandability is approximately normally distributed for each group; and the variances are homogeny.

The results for comparing the understandability of awareness of filtering and the filter bubble between group 1 and group 2 (Hypothesis - H3A) are summarized in Table 7.9 and Table 7.10.

Table 7.9: Group Statistics – Awareness

	Group	N	Mean	Std. Deviation	Std. Error Mean
Awareness	1	163	.7117	.23792	.01864
	2	163	.7996	.20718	.01623

Table 7.10: Independent Samples Test - Awareness

Independent Samples Test										
		Equality of Variances		t-test for Equality of Means						
									Interval of the	
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
Awareness	Equal variances assumed	2.621	.106	3.559	324	.000	.08793	.02471	.03932	.13655
	Equal variances not assumed			3.559	317.989	.000	.08793	.02471	.03932	.13655

H0: μ Awareness in Group 1 (No Help) = μ Awareness in Group 2 (With Help)

H1: μ Awareness in Group 1 (No Help) ≠ μ Awareness in Group 2 (With Help)

p-value (.106) > α (0.05) and so the variances are assumed to be equal. Using the 1st row of Table 7.10,

t =3.559, p-value < 0.001 => p-value < α (0.05) => Reject H0.

This test found that group 2 participants had **statistically significantly** higher understanding of awareness of the filtering and the filter bubble **(0.7996 ± 0.2072)** compared to that of group 1 participants **(0.7117 ± 0.2379)**, $t\,(324) = 3.559$, $p < .001$.

The results for comparing the understandability of explanation between group 1 and group 2 (Hypothesis - H3B) are summarized in Table 7.11 and Table 7.12.

Table 7.11: Group Statistics – Explanation

	Group	N	Mean	Std. Deviation	Std. Error Mean
Explanation	1	163	.6176	.21591	.01691
	2	163	.7403	.21284	.01667

Table 7.12: Independent Samples Test – Explanation

Independent Samples Test										
		Equality of Variances		t-test for Equality of Means						
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Interval of the Lower	Interval of the Upper
Explanation	Equal variances assumed	.078	.780	5.167	324	.000	.12270	.02375	.07598	.16942
	Equal variances not assumed			5.167	323.934	.000	.12270	.02375	.07598	.16942

H0: μ Explanation in Group 1 (No Help) = μ Explanation in Group 2 (With Help)

H1: μ Explanation in Group 1 (No Help) ≠ μ Explanation in Group 2 (With Help)

p-value (.780) > α (0.05) and so the variances are assumed to be equal.

Using the 1st row of Table 7.12,

t =5.167, p-value < 0.001 => p-value < α (0.05) => Reject H0.

This test found that group 2 participants had **statistically significantly** higher understanding of explanation of the filtering and the filter bubble **(0.7403 ± 0.2128)** compared that of group 1 participants **(0.6176 ± 0.2159)**, t (324) = 5.167, p < .001.

The results for comparing the understandability of control provided by visualization between group 1 and group 2 (Hypothesis - H3C) are summarized in Table 7.13 and Table 7.14.

Table 7.13: Group Statistics – Control

	Group	N	Mean	Std. Deviation	Std. Error Mean
Control	1	163	.7607	.22455	.01759
	2	163	.8834	.16885	.01323

Table 7.14: Independent Samples Test - Control

Independent Samples Test										
		Equality of Variances		t-test for Equality of Means						
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Interval of the Lower	Interval of the Upper
Control	Equal variances assumed	22.744	.000	5.576	324	.000	.12270	.02201	.07941	.16599
	Equal variances not assumed			5.576	300.814	.000	.12270	.02201	.07939	.16600

H0: μ Control in Group 1 (No Help) = μ Control in Group 2 (With Help)

H1: μ Control in Group 1 (No Help) ≠ μ Control in Group 2 (With Help)

p-value < α (0.05) and so the variances are NOT assumed to be equal.

Using the bottom row of Table 7.14,

t = 5.576, p-value < 0.001 => p-value < α (0.05) => Reject H0.

This test found that group 2 participants had **statistically significantly** higher understanding of control provided by visualization **(0.8834 ± 0.1689)**, compared that of group 1 participants **(0.7607 ± 0.2246)**, $t\ (300.8) = 5.576$, $p < .001$.

The results for comparing the overall understandability of visualization between group 1 and group 2 (Hypothesis - H3D) are summarized in Table 7.15 and Table 7.16.

Table 7.15: Group Statistics – Understandability

	Group	N	Mean	Std. Deviation	Std. Error Mean
Understandability	1	163	.6967	.18084	.01416
	2	163	.8078	.16509	.01293

Table 7.16: Independent Samples Test - Understandability

Independent Samples Test										
		Equality of Variances		t-test for Equality of Means						
									Interval of the	
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
Understandability	Equal variances assumed	1.467	.227	5.793	324	.000	.11111	.01918	.07338	.14884
	Equal variances not assumed			5.793	321.346	.000	.11111	.01918	.07338	.14884

H0: μ Understandability in Group 1 (No Help) = μ Understandability in Group 2 (With Help)

H1: μ Understandability in Group 1 (No Help) ≠ μ Understandability in Group 2 (With Help)

p-value (.227) > α (0.05) and so the variances are assumed to be equal.

Using the 1st row of Table 7.16,

t = 5.793, p-value < 0.001 => p-value < α (0.05) => Reject H0.

This test found that group 2 participants had **statistically significantly** higher overall understanding of the visualization **(0.8078 ± 0.1651)** compared that of group 1 participants **(0.6967 ± 0.1808)**, $t\,(324) = 5.793$, $p < .001$.

7.4 Discussion

This section discusses the results of the hypotheses tests. Section 7.4.1 and 7.4.2 discuss the results of the one sample t-tests of group 1 and group 2 respectively. Then the discussion of independent sample t-test for the combined group is presented in section 7.4.3.

7.4.1 Group 1 – No Help

The results of the quantitative study suggest that overall the users in group 1 had a good understanding about the visualization because the mean overall understandability (0.6967) of users was greater than the test mean (0.5 - which is the average of 0 - does not completely understand and 1 - completely understand). By comparing the means of variables awareness, explanation, and control, we can see that users have a better understanding (0.7607) about the control of filtering and the filter bubble provided by the visualization. This can be linked with the drag and drop feature of the visualization, which is very popular and commonly used action in many user interfaces and it is a very user friendly user interface construct. On the other side, the users' understanding about the visualization providing explanation to the filtering and the filter bubble has a lower value (0.6176). Though it is higher than 0.5, it clearly shows that the visualization has to be improved on this aspect. A possible improvement could be to provide some context sensitive help to the visual cues in the visualization. The overall understandability value of the visualization (0.6967) shows that the users had a good understanding about the visualization after exploring it for the first time without any help and it could be considered as an intuitive visualization.

Analyzing the t-test values gives us more insight into the understandability measures. As mentioned earlier, the understandability of visualization is calculated using the three variables Awareness, Explanation and Control. These three variables are understandability variables and are measured using the metrics presented in Table 7.2. The variables Awareness, Explanation and Control obtained a high 2-tailed value respectively 11.358, 6.953, and 14.824. These values are comparatively very high when compared with their relevant one-tailed t-test value, which is 1.65. This indicates that these three variables are a very good measure for the understandability of this visualization.

The additional test on graphical language results suggest that the users very rarely interpreted the position of circles inside the bubble (closer to the center or to the periphery) i.e. very few users selected it. A possible reason for this might be the nature of the question i.e. the question did not ask about this construct explicitly; the users might have only focused on the first 3 graphical constructs, which are intuitive and obvious. But it seems a useful construct and could be added as an improvement to the visualization in future.

7.4.2 Group 2 – With Help

The results of the quantitative study suggest that overall the users in group 2 also had a good understanding about the visualization. By comparing the means of variables awareness, explanation, and control, we can see that users have a better understanding (0.8834) about the control of filtering and the filter bubble provided by the visualization. Like in group 1, this can also be linked with the drag and drop feature of the visualization, which is very popular and commonly used action in many user interfaces and it is a user-friendly user interface construct. On the other side, the users' understanding about the visualization providing explanation of the filtering and the filter bubble has a lower value (0.7403) compared to the other variables (Awareness and Control). The reason for this might be the lack of completeness of the help text description. Though it is higher than 0.5, it clearly shows that the visualization and the help text have to be improved on this aspect. The overall understandability value of the visualization (0.8078) shows that the users had even better understanding about the visualization after exploring it with guided help.

Analyzing the t-test values gives us more insight into the understandability measures. As mentioned earlier, the understandability of visualization is calculated using the three variables Awareness, Explanation and Control. These three variables are understandability variables and are measured using the metrics presented in Table 7.2. The variables Awareness, Explanation and Control obtained a high 2-tailed value respectively 18.462, 14.413, and 28.993. These values are very high when compared with their relevant one-tailed t-test value, which is 1.65. With the high 2-tailed values in group 1, this group 2 values provides more support that these three variables are a very good measure for the understandability of this visualization.

7.4.3 Combined Group

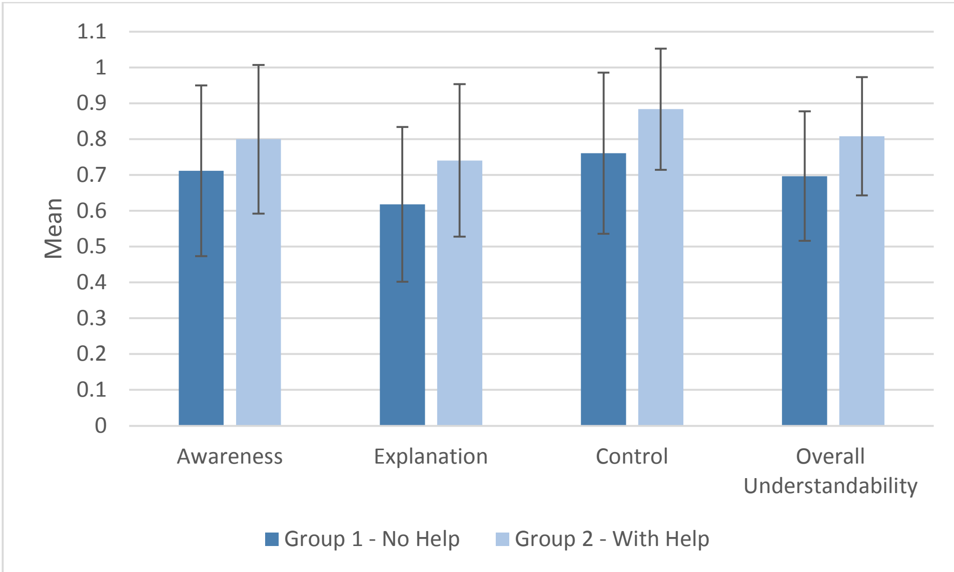


Figure 7.5: Understandability Chart - Group 1 vs Group 2

As shown in Figure 7.5, group 2 always had a better understanding of the visualization in all the understandability variables than that of the group 1 did and this difference in means is statistically significant from the independent-sample t-test results. This clearly shows the need for a context sensitive help support with the visualization. The percentage of increase in understandability of Awareness, Explanation, Control and Overall Understandability are 12.35%, 19.87%, 16.13% and 15.95% respectively. This clearly shows that the help text is useful. In both groups, the understanding that the visualization provides explanation about the filtering and the filter bubble, has the lowest value. This clearly shows that the visualization has to focus more on improving this aspect by all means. By looking at the overall understandability between the two groups, the intuitiveness of the visualization is not up to a very satisfying level. Overall, the visualization can be considered successful because all the understandability values are greater than 0.5 (in fact, all of them are greater than 0.6) shows that the users in the first group were able to understand the visualization and its functions after exploring the visualization without any help.

7.5 Summary

The quantitative study with 163 participants in each group (No help and with help) demonstrates that the visualization leads to an increased users' awareness of the filter bubble, understandability of the filtering mechanism and to a feeling of control over the data stream they are seeing. The next chapter presents a summary of the research, a list of contributions, and future directions of this research.

CHAPTER 8
CONCLUSION

This chapter provides a summary of the research, conclusion, a list of contributions and future research directions.

8.1 Summary of the Research

This thesis proposes an interactive method to visualize and manipulate the personalized stream filtering and the filter bubble in OSNs. A pilot study was conducted with eleven (11) participants from MADMUC Research Lab for three (3) weeks to evaluate the user acceptance and user experience with MADMICA and the proposed visualization. At the end of the pilot study, we identified that the interest based relationship filtering has the potential to reduce social data overload and also the filter bubble visualization has the potential to alleviate the filter bubble problem in OSNs and increase the users' trust in the system. Then a qualitative study was conducted to in-depth understand the user perceptions of the visualization. Most of the users were able to correctly interpret the graphical language of the visualization. Some of the undesirable perceptions helped to improve the visualization design. A large-scale quantitative study with 326 participants was carried out to evaluate whether the users understand that the visualization provides awareness, explanation and control of filtering and the filter bubble. In addition to that, the intuitiveness of the visualization was also tested. The results of quantitative study show that the visualization leads to increased users' awareness of the filter bubble problem in OSNs, understandability of the filtering mechanism and to a feeling of control over the filtering. Moreover, users showed a better understanding of the awareness and the control of the filtering mechanism than the explanation of the filtering mechanism. Even though we couldn't prove the intuitiveness of the visualization, the results of the understandability test for the group which explored the visualization without any help text show that the users understand the visualization and its functions even without the guided help. Nevertheless, the help provided with visualization improved the users' understanding of the visualization.

8.2 Conclusion

This thesis demonstrates that it is possible to alleviate the filter bubble problem in content-based stream filtering in a P2P social network. The proposed visualization created awareness and provided explanation of the filtering mechanism and the filter bubble and engaged the users in actions to control the filtering of news stream in a P2P social network leading to increased trust in the system.

8.3 Contributions

This thesis has the following contributions:

1. Real world implementation of an interest based stream filtering algorithm in a P2P Social Network. The implementation has been done as a plugin to Friendica P2P social network while preserving its modularized architecture.
2. A novel method to visualize the personalized stream filtering in Online Social Networks to create awareness, provide explanation, and control of personalized stream filtering and alleviate "the filter bubble" problem and increase the users' trust in the system. The visualization is based on a bubble metaphor to make the effect of the personalized stream filtering in OSNs more understandable for the users and it does not cause the overload or undo the advantages of the filtering.
3. Evaluation of an interactive visualization. Three user evaluations were conducted: small-scale pilot study, qualitative user study and a large-scale quantitative user study. The small-

scale pilot study involves mixed method of evaluation which comprises both qualitative and quantitative evaluation methods.

8.4 Future Work

Currently the user has to manually select the category out of a list of categories or enter their own category (which is subjected to admin approval) for the social data in MADMICA. The admin approval to user created/suggested category has been implemented in order to avoid the redundant categories, the overflow of categories in the system and to preserve effectiveness and performance of the interest based relationship model. Moreover, manual tagging of social data suffers from inconsistency and idiosyncrasy. In addition to that, humans are not consistent day to day or even minute to minute as social and personal context changes. As the feature of admin approval limits the categories in the system, there is a need for automated tagging and categorization of social data in order to solve this problem of limited number of categories while preserving the performance and effectiveness of the interest based relationship model. The automated categorization is expected to programmatically identify the tags for the social data except the photos and videos, categorize the social data based on an ontology and eliminate the burden of manual selection of category by users. Moreover, it should provide semantically rich representation of social data and should not cause an overflow of categories in the system. A starting point for this future work would be to look into the area of semantic data tag extraction. It can be implemented using OpenCalais web service [38].

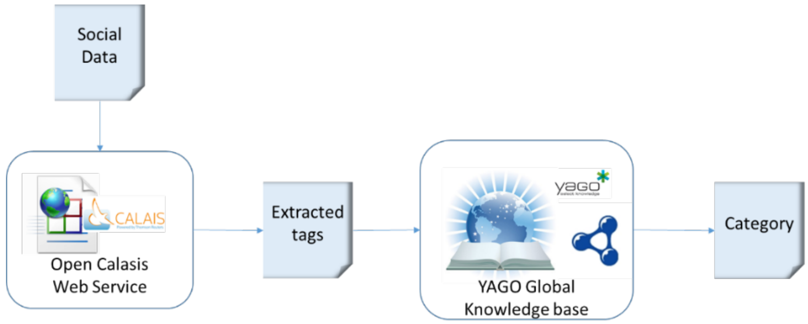


Figure 8.1: Automatic Category Extraction from Social Data in MADMICA

Figure 8.1 shows one of the possible implementation path for the automatic category extraction of social data in MADMICA. The social data can be analyzed and tags can be extracted using the OpenCalais web service. The resulted keywords/tags from the OpenCalais web service can be used to extract a general category using general purpose ontology YAGO. YAGO is a large ontology derived from Wikipedia and WordNet [54]. The category extracted from this process can be expected to provide semantically rich representation of social data while not causing the overflow of categories in the system.

RESEARCH PUBLICATIONS

1. Nagulendra, S., Vassileva, J.: Minimizing Social Data Overload through Interest- Based Stream Filtering in a P2P Social Network, Proc. IEEE International Conference on Social Computing, SocialCom'2013 (2013)

2. Nagulendra, S., Vassileva, J.: Providing Awareness, Understanding and Control of Personalized Stream Filtering in a P2P Social Network. 19th International Conference, CRIWG 2013. pp. 61–76 Springer Berlin Heidelberg (2013).

3. Nagulendra, S., Vassileva, J.: Understanding and Controlling the Filter Bubble through Interactive Visualization: A User Study. 25th ACM Conference on Hypertext and Social Media. Hypertext 2014. Submitted.

LIST OF REFERENCES

1. An Infographic: The Biggest Shift since the Industrial Revolution | TechnoBuffalo, Available online at: http://www.technobuffalo.com/2010/06/01/an-infographic-the-biggest-shift-since-the-industrial-revolution/. (accessed 5 Sept 2013)

2. A Standards-based, Open and Privacy-aware Social Web, http://www.w3.org/2005/Incubator/socialweb/XGR-socialweb-20101206/#Decentralized. (accessed 20 Aug 2012)

3. Berners-Lee, T., Hollenbach, J., Lu, K., Presbrey, J., Schraefel, M.: Tabulator redux: Browsing and writing linked data. Proc.1st Workshop on Linked Data on the Web (2008)

4. Bland, J. M., Altman, D. G.: "Statistics notes: Cronbach's alpha," pp. 570-572 (1997)

5. Bonhard, P. et al.: Accounting for taste. Proc. SIGCHI conference on Human Factors in computing systems - CHI ’06. p. 1057 ACM Press, New York, New York, USA (2006)

6. Boyatzis, R.E.:Transforming qualitative information: Thematic analysis and code development. Thousand Oaks, London, & New Delhi: SAGE Publications (1998)

7. Buchegger, S., Schiöberg, D., Vu, L.H., Datta, A.: PeerSoN: P2P social networking: early experiences and insights. Proc. The Second ACM EuroSys Workshop on Social Network Systems, pp. 46–52. ACM (2009)

8. Burke, R.: Hybrid Web Recommender Systems. The Adaptive Web, LNCS 4321. pp. 377–408 Springer-Verlag Berlin Heidelberg (2007)

9. Burke, R.: The Wasabi Personal Shopper: A Case-Based Recommender System. In: Proceedings of the 11th National Conference on Innovative Applications of Artificial Intelligence, pp. 844-849 (1999)

10. Chart of the day: How People Share Content on the Web, Available online at: http://www.businessinsider.com/chart-of-the-day-social-networking-sites-dominate-sharing-2009-7. (accessed 4 Aug 2013)

11. Chen, L. and Sycara, K. WebMate: a personal agent for browsing and searching. Proceeding of the 2nd international conference on autonomous agents and multi agent systems, (1998), 132–139

12. Cutillo, L. et al.: Safebook: A privacy-preserving online social network leveraging on real-life trust. IEEE Commun. Mag. 47, 12, 94–101 (2009)

13. Dahlen,B.J., Konstan,J.A., Herlocker,J.L., Good,N., Borchers,A., Riedl,J., Jump-starting Movielens: User benefits of starting a collaborative filtering system with "dead data". University of Minnesota TR 98-017 (1998)

14. Das, A.S. et al.: Google news personalization. Proceedings of the 16th international conference on World Wide Web - WWW ’07. p. 271 ACM Press, New York, New York, USA (2007)

15. Diaspora, https://joindiaspora.com/, (accessed 25 Aug 2012)

16. Facebook Investors, Available online at: http://investor.fb.com/releasedetail.cfm?ReleaseID=780093. (accessed 4 Aug 2013)

17. Garrett, J.J.: Elements of User Experience, The User-Centered Design for the Web and Beyond. Pearson Education (2010)

18. George, D., Mallery, P., SPSS for Windows step by step: A simple guide and reference. 11.0 update (4th ed.), Boston: Allyn & Bacon (2003)

19. Herlocker, J.L., Konstan, J. A., Riedl, J.: Explaining Collaborative Filtering Recommendations. Proc. ACM Conference on Computer Supported Cooperative Work CSCW 2000. pp. 241–250 ACM Press, New York, USA (2000)

20. How Social Media Influences People - Infographic | Social MagnetsSocial Magnets, http://www.socialmagnets.net/how-social-media-influences-people/. (accessed 11 Nov 2013)

21. Indratmo, Vassileva,J., Gutwin, C.: Exploring blog archives with interactive visualization. In: International conference on Advanced Visual Interfaces (2008)

22. ISO/IEC TR 9126-2:2003 – Product Quality – External Metrics, Available online at: http://www.iso.org/iso/iso_catalogue/catalogue_tc/catalogue_detail.htm?csnumber=22750. (accessed 2 Feb 2014)

23. Johnson, H., Johnson, P.: Explanation Facilities and Interactive Systems. Proc. 1st International Conference on Intelligent User Interfaces IUI'93. pp. 159–166 ACM (1993)

24. Kaela, S.: The Role of HTML5 and Flash in Web Design. (2012)

25. Kincaid, J.: EdgeRank: The Secret Sauce That Makes Facebook's News Feed Tick | TechCrunch, Available online at: http://techcrunch.com/2010/04/22/facebook-edgerank/. (accessed 6 July 2013)

26. Konstan, J.A., Riedl, J.: Recommender systems : from algorithms to user experience. User Modeling and User-Adapted Interaction. 22, 1-2, 101–123 (2012)

27. Krulwich, B.: Lifestyle Finder: Intelligent User Profling Using Large-Scale Demographic Data. In: Artificial Intelligence Magazine 18 (2), 37-45 (1997)

28. Kywe, S.M. et al.: A Survey of Recommender Systems in Twitter. Proc. 4th International Conference, SocInfo 2012, Lausanne, Switzerland, December 5-7, 2012. pp. 420–433 (2012)

29. Macgirvin, M.: DFRN – the Distributed Friends & Relations Network, Available online at: https://macgirvin.com/spec/dfrn2.pdf. (accessed 2 Aug 2012)

30. Manish, A. et al.: An Online News Recommender System for Social Networks. SIGIR-SSM. (2009)

31. McCalla, G.I.: "The Fragmentation of Culture, Learning, Teaching and Technology: Implications for the Artificial Intelligence in Education Research Agenda in 2010", International Journal of Artificial Intelligence in Education, 11, 2, 2000, pp. 177-196.

32. Mooney, R.J., Roy, L.: Content-based book recommending using learning for text categorization. Proc. fifth ACM Conf. Digit. Libr. - DL ’00. 195–204 (2000)

33. Munson, S.A., Resnick, P.: Presenting diverse political opinions. Proc. 28th international conference on Human factors in computing systems - CHI ’10. p. 1457 ACM Press, New York, New York, USA (2010)

34. Musiani, F.: When social links are network links: The dawn of peer-to-peer social networks and its implications for privacy. 4, 185–207 (2010)

35. Nagulendra, S., Vassileva, J.: Minimizing Social Data Overload through Interest- Based Stream Filtering in a P2P Social Network, Proc. IEEE International Conference on Social Computing, SocialCom’2013 (2013)

36. NVivo 10 for Windows, Available online at: http://www.qsrinternational.com/products_nvivo.aspx. (accessed 25 Feb 2014)

37. One-Sample T-Test using SPSS, Available online at: https://statistics.laerd.com/spss-tutorials/one-sample-t-test-using-spss-statistics.php. (accessed 2 Mar 2014)

38. OpenCalais Web Service API | OpenCalais, Available online at: http://www.opencalais.com/documentation/calais-web-service-api. (accessed 15 Nov 2012)

39. Paolacci, G., Chandler, J.: Running experiments on Amazon Mechanical Turk. 5, 5, 1–14 (2014)

40. Pariser, E.: The Filter Bubble: What the Internet Is Hiding from You. Penguin Press HC (2011)

41. Pazzani, M., Billsus, D.: Content-based Recommendation Systems. In: Brusilovsky, P., Kobsa, A., Nejdl, W. (eds.): The Adaptive Web: Methods and Strategies of Web Personalization, Lecture Notes in Computer Science, Vol. 4321. Springer-Verlag, Berlin Heidelberg New York (2007)

42. Pouwelse, J.A., Garbacki, P., Wang, J., et al.: TRIBLER: a social-based peer-to-peer system. Concurrency and Computation: Practice and Experience 20(2), 127–138 (2008)

43. Pu, P., Chen, L.: Trust Building with Explanation Interfaces. Proc. 11th International Conference on Intelligent User Interfaces IUI ’06. p. 93 ACM Press, New York, USA (2006)

44. Resnick, P., A. Munson, S., Garrett, R.K., Stroud, N.J., Kriplean, T.: Bursting Your (Filter) Bubble : Strategies for Promoting Diverse Exposure. Proc. Conference on Computer supported Cooperative Work CSCW 2013 Companion proc. pp. 95–100 ACM (2013)

45. Resnick,P., Iacovou,N., Suchak,M., Bergstrom,P., Riedl,J., GroupLens: An open architecture for collaborative filtering of netnews. Proceedings of 1994 Conference on Computer Supported Collaborative Work. pp. 175-186 (1994)

46. Resnick, P., Varian, H.R.: Recommender systems. Communications of the ACM. 40, 3, 56–58 (1997)

47. Ricci, F., Rokach, L., Shapira, B.: Introduction to Recommender Systems Handbook. Ricci, F., Rokach, L., Shapira, B., and Kantor, P.B. (eds.) In: Recommender Systems Handbook. pp. 1–35 Springer US, Boston, MA (2011)

48. Schafer, B., Frankowski, D., Herlocker, J., Sen, S.: Collaborative Filtering Recommender Systems. In: Brusilovsky, P., Kobsa, A., Nejdl, W. (eds.): The Adaptive Web: Methods and Strategies of Web Personalization. Lecture Notes in Computer Science, Vol. 4321. Springer-Verlag, Berlin Heidelberg New York (2007)

49. Shardanand,U., Maes,P., Social Information Filtering: Algorithms for Automating "Word of Mouth". Proceedings of ACM CHI '95. Denver, CO. pp. 210-217 (1995)

50. Shi, S.: Keeping Up with the Information Glut by Visualizing Patterns of Posting by Friends on Facebook, http://hdl.handle.net/10388/ETD-2011-09-139, (accessed 11 Feb 2013)

51. Shi, S., Largillier, T., Vassileva J.: Keeping Up with Friends' Updates on Facebook, 18th Int. Conf. on Collaboration and Groupware, Duisburg, Germany, Sept 16 - 19, (2012)

52. Shneiderman, B.: The Eyes Have It: A Task by Data Type Taxonomy for Information Visualizations. Proc. 1996 IEEE Symposium on Visual Languages. pp. 336–343 IEEE Comput. Soc. Press (1996)

53. Social Igniter, https://social-igniter.com/, (accessed 27 Aug 2012)

54. Suchanek, F.M. et al.: YAGO: A Large Ontology from Wikipedia and WordNet. Web Semant. Sci. Serv. Agents World Wide Web. 6, 3, 203–217 (2008)

55. Suh, S. C. Practical applications of data mining. Jones & Bartlett Publishers. (2012)

56. Tandukar U., Vassileva J.: Ensuring Relevant and Serendipitous Information Flow in Decentralized Online Social Network. Proc. AIMSA'2012, 15th biennial conference on AI Methods Systems, Applications, Springer Verlag, LNAI 7557, pp. 79-88. (2012)

57. Tandukar, U., Vassileva, J.: Selective Propagation of Social Data in Decentralized Online Social Network. Proc. UMAP 2011 Workshops, LNCS 7138. pp. 213–224 Springer-Verlag Berlin Heidelberg (2012)

58. Testing for Normality using SPSS, Available online at: https://statistics.laerd.com/spss-tutorials/testing-for-normality-using-spss-statistics.php. (accessed 2 Mar 2014)

59. The Filter Bubble: Algorithm vs. Curator & the Value of Serendipity | Brain Pickings, Available online at: http://www.brainpickings.org/index.php/2011/05/12/the-filter-bubble/. (accessed 15 Aug 2013)

60. The Internet is Our Social Network | Friendica, Available online at: http://friendica.com/. (accessed 11 Dec 2012)

61. Tintarev, N., Masthoff, J.: A Survey of Explanations in Recommender Systems. 2007 IEEE 23rd Int. Conf. Data Eng. Work. 801–810 (2007)

62. Tintarev, N., Masthoff, J.: Effective Explanations of Recommendations : User-Centered Design. Proc. 2007 ACM Conference on Recommender Systems RecSys'07. pp. 153–156 ACM (2007)

63. Wang, Y., Zhang, J., Vassileva, J.: Towards Effective Recommendation of Social Data across Social Networking Sites. Proc. Artificial Intelligence: Methodology, Systems, and Applications AIMSA'2010. pp. 61–70 Springer-Verlag Berlin Heidelberg (2010)

64. Webster, A., Vassileva, J.: The KeepUP Recommender System. Proc. 2007 ACM Conference on Recommender Systems RecSys '07. pp. 173–176 ACM, Minneapolis, Minnesota, USA. (2007)

65. Webster, A., Vassileva, J.: Visualizing personal relations in online communities. Proceedings of the adaptive hypermedia and adaptive web-based systems (AH'2006), June 21–23, pp. 223–233, Springer LNCS 4018, Dublin (2006)

66. Yeung, C.A., Liccardi, I., Lu, K., Seneviratne, O., Berners-Lee, T.: Decentralization: The future of online social networking. Proc. W3C Workshop on the Future of Social Networking Position Papers (2009)

APPENDIX A

PILOT STUDY QUESTIONNAIRE AND SUMMARY OF RESPONSES

-

Madmica Questionnaire - 1

* Required

1. **Did you notice a change in the level of interest you had in the posts of your friends in time? ***

   Mark only one oval.

   - ( ) Yes
   - ( ) No

2. **If yes to the above question,**

   Mark only one oval.

   - ( ) The posts got more interesting
   - ( ) The posts got less interesting

3. **Did you at any point feel there were too many posts to keep track of? ***

   Mark only one oval.

	1	2	3	4	5	6	7	8	9	10	
Too few posts	( )	( )	( )	( )	( )	( )	( )	( )	( )	( )	Too many posts

4. **Did you notice that the system filters some posts away? ***

   Mark only one oval.

   - ( ) Yes
   - ( ) No

5. **Do you trust a system that filters posts away from the stream? ***

   Mark only one oval.

   - ( ) Yes
   - ( ) No
   - ( ) Depends

6. **If "Depends"**

Mark only one oval.

- On whether I understand the mechanism of filtering
- On whether I have some means to control
- On the number of posts
- Mostly I wouldn't care

7. **How do you think the system decides what to filter away? ***

8. **Name your three favourite things about Madmica, and your three least favourite? ***

9. **What would encourage you to return to Madmica in the future? ***

10. **Was anything too obtrusive in Madmica? ***

11. **Have you ever tried to delete your updates (data) or had concern about the privacy of your data hosted in Madmica? If yes, would you prefer to host your data in your server to avoid privacy risks?** *

12. **How easy to post a new item in Madmica?** *
Mark only one oval.

	1	2	3	4	5	6	7	8	9	10	
Difficult											Easy

13. **How easy to select a category of your post in Madmica?** *
Mark only one oval.

	1	2	3	4	5	6	7	8	9	10	
Difficult											Easy

14. **How easy it is to keep track with the posts of your friends?** *
Mark only one oval.

	1	2	3	4	5	6	7	8	9	10	
Difficult											Easy

Madmica Questionnaire - 2

Please use the visualization diagram provided on the left to answer the questions below.

* Required

1. **Did you use the filter bubble visualization? ***
   Mark only one oval.
   - Yes
   - No

2. **If yes, what do you think it represents?**

3. **What do you think about the category view in the visualization? ***

4. **What do you think about the friend view in the visualization? ***

5. **Which view do you prefer? ***
   Mark only one oval.
   - Category View
   - Friend View

6. **Can you think of other views that could be useful?** *

7. **What do think about the position of category/friend circle? Does it have a meaning?** *

8. **What do you think about the size of the category/friend circle? Does it have a meaning?** *

9. **Did you try dragging a category/friend circle from the outside to the inside of the bubble?** *
Mark only one oval.

- Yes
- No

10. **If yes, did you notice an effect after of doing this?**

11. **Did you try dragging a category/friend circle from the inside to the outside of the bubble?** *
Mark only one oval.

- ( ) Yes
- ( ) No

12. **If yes, did you notice an effect after of doing this?**

13. **Why does a box full of links appear when you click on a category/friend circle which is outside the bubble? Can you guess which of these links were more recently posted?** *

14. **Did you realize the system was filtering the posts from your friends away from your stream?** *
Mark only one oval.

- ( ) Yes
- ( ) No

15. **Does the visualization help you to understand how the filtering works?** *
Mark only one oval.

- ( ) Yes
- ( ) No

16. **Did you notice that system orders a coffee when you post an item?** *
Mark only one oval.

- ( ) Yes
- ( ) No

17. **Does the visualization give you a feeling of control over the stream of posts from your friends?** *
Mark only one oval.

- ( ) Yes
- ( ) No

18. **Does the visualization increase your trust in the filtering process?** *
Mark only one oval.

- ( ) Yes
- ( ) No

19. **Does the visualization help you trust the system more?** *
Mark only one oval.

- ( ) Yes
- ( ) No

Trust in the System

20. **Trust in the System before using the filter bubble:** *
Mark only one oval.

	1	2	3	4	5	6	7	8	9	10	
Very Low	( )	( )	( )	( )	( )	( )	( )	( )	( )	( )	Very High

21. **Trust in the System after using the filter bubble:** *
Mark only one oval.

	1	2	3	4	5	6	7	8	9	10	
Very Low	( )	( )	( )	( )	( )	( )	( )	( )	( )	( )	Very High

-

22. **Trust in the System after seeing the hidden posts:** *

Mark only one oval.

	1	2	3	4	5	6	7	8	9	10	
Very Low	◯	◯	◯	◯	◯	◯	◯	◯	◯	◯	Very High

23. **Level of transparency in filtering provided by the system:** *

Mark only one oval.

	1	2	3	4	5	6	7	8	9	10	
Very Low	◯	◯	◯	◯	◯	◯	◯	◯	◯	◯	Very High

Filter Bubble Visualization

24. **Aesthetically pleasing** *

Mark only one oval.

	1	2	3	4	5	6	7	8	9	10	
Very Low	◯	◯	◯	◯	◯	◯	◯	◯	◯	◯	Very High

25. **Friends View** *

Mark only one oval.

	1	2	3	4	5	6	7	8	9	10	
Unhelpful	◯	◯	◯	◯	◯	◯	◯	◯	◯	◯	Helpful

26. **Category View** *

Mark only one oval.

	1	2	3	4	5	6	7	8	9	10	
Unhelpful	◯	◯	◯	◯	◯	◯	◯	◯	◯	◯	Helpful

27. **Awareness about hidden posts** *
Mark only one oval.

	1	2	3	4	5	6	7	8	9	10	
Inadequate											Adequate

28. **Arrangement of information on screen** *
Mark only one oval.

	1	2	3	4	5	6	7	8	9	10	
Illogical											Logcial

29. **Manipulation of interest (categories) / friend circles (dragging in and out)** *
Mark only one oval.

	1	2	3	4	5	6	7	8	9	10	
Difficult											Easy

30. **Finding an interest which is not inside your filter bubble** *
Mark only one oval.

	1	2	3	4	5	6	7	8	9	10	
Difficult											Easy

31. **Discovering new interests** *
Mark only one oval.

	1	2	3	4	5	6	7	8	9	10	
Difficult											Easy

32. **Discovering the interests of friends** *
Mark only one oval.

	1	2	3	4	5	6	7	8	9	10	
Difficult											Easy

33. **Discovering in which areas your friends are most interested** *

Mark only one oval.

	1	2	3	4	5	6	7	8	9	10	
Difficult											Easy

Sayooran N

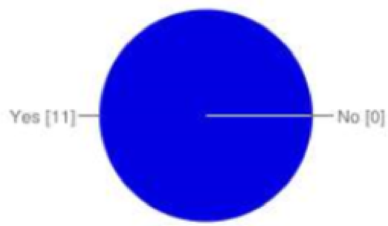
Edit this form

11 responses

View all responses

Summary

Did you notice a change in the level of interest you had in the posts of your friends in time?

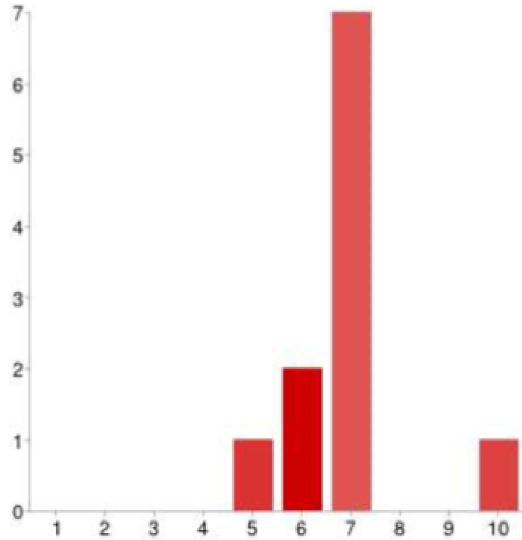


Yes	**11**	100%
No	**0**	0%

If yes to the above question,

The posts got more interesting	**10**	91%
The posts got less interesting	**1**	9%

Did you at any point feel there were too many posts to keep track of?

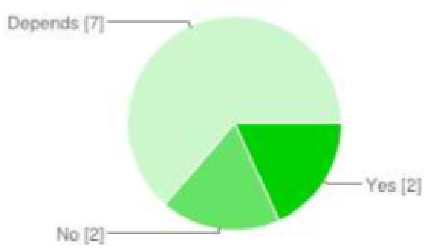


1	**0**	0%
2	**0**	0%
3	**0**	0%
4	**0**	0%
5	**1**	9%
6	**2**	18%
7	**7**	64%
8	**0**	0%
9	**0**	0%
10	**1**	9%

Did you notice that the system filters some posts away?

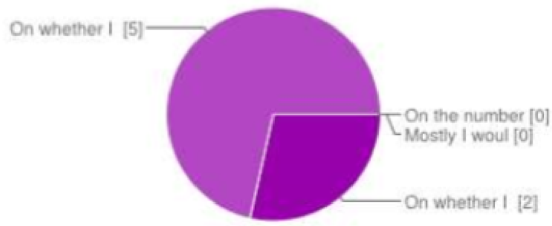
Yes	**10**	91%
No	**1**	9%

Do you trust a system that filters posts away from the stream?



Yes	**2**	18%
No	**2**	18%
Depends	**7**	64%

If "Depends"



On whether I understand the mechanism of filtering	**2**	29%
On whether I have some means to control	**5**	71%
On the number of posts	**0**	0%
Mostly I wouldn't care	**0**	0%

How do you think the system decides what to filter away?

based on my interest by topic I think based on my interaction interns of likes and comments on different posts categories. based on ones posts categories. based on my posts and my interactions on others' posts My behavior in the system and what I chose to filter out using the bubble based on my activities and interests it filters away posts that I don't have much interests to It filters the area that I don't show interest in; I don't comment on the posts of that area. No idea, maybe based on my clicks? I think it filters the post depending on the topics of the posts.

Name your three favourite things about Madmica, and your three least favourite?

favourite features: tagging; dislike (but a bit ambiguous); categorization (but still needs to be expanded on) least favourite features: comments come as spam on my wall; hard to figure out the original posts that people comment on; attaching links is not consistent 1. Easy interface for use 2. Leader board. 3. Hides unwanted post (Filter bubbles) 1. It was not easy to categorise a post. 2. Not able to comments and likes some of the person post in my list. 3. Sometimes seems too much post to keep track with. favorite things: importing titles with pictures and videos by using the link only interface design like button least favorite: no automatic sign in needs more tags doesn't support HTTPS links 3 fav things: 1. Leaderboard 2. I can see the last likes by the users. 3. Easy sharing 3 least fav: 1. I was confused about the "Introduction" system. 2. The bubble of hidden posts sometimes irritates me. 3. "Community" as it was confusing as well. I like filtering, point system, categorization. Did not like the lack of user control in filtering and no transparency on how points are awarded, would also like to control the number of notifications I receive in my mail Favorite: The filter bubble (which prevent information overload), the interface Least favorite: the loading speed three favorite things: The

Bubble The ability to see just the category that you want, buy choosing that category from Left side column. The Posts which where related to our Lab three least favorite: It hides uninterested posts too early! You can not edit your posts. The Meaningless posts just to gain points make the social media uninteresting!   Three favorites: 1- the bubble that shows my interests 2- the customization of the bubble, so I can drag and drop according to my interest 3- the leaderboard lease favorites: 1- the dislike button, I feel I don't like the news but I like sharing the news and find it useful. I prefer to comment in this case instead of like/dislike 2- the way of displaying updates (shaikhah likes .... status) because it prevent me to see the original posts first 3- one tag is allowed   Visualization Regular email on updated Easy to link with likeminded people No very easy to post sometime   Favourite: 1. filtering mechanism 2. design layout 3. flexibility in uploading post Least Favourite: 1. 2. icons 3. editing post   Favorite: close community, Least favorite: UI, doesn't work on IE,

What would encourage you to return to Madmica in the future?

nice gists from friends; competitiveness   More contents   share video links   interesting news posts/ shares by my friends   Larger community maybe?   I guess its privacy and interest-based filtering   the news shared are so excited and related to our community   If my friends continues to use it, to keep a touch with them.   quality of posts and the point system   To share things that is related to my friend list or I found that useful for them.   Sharing posts!

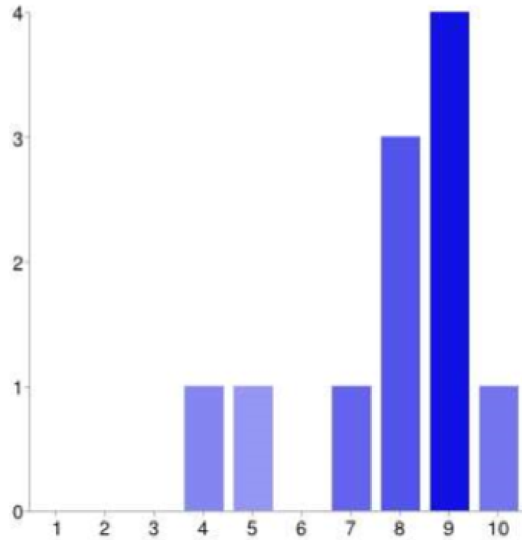
Was anything too obtrusive in Madmica?

no   some timing posting was not easy   No.   comment and likes on other people's posts. I prefer seeing original posts on my news feeds and not comments from all and sundry   No   Not that I noticed   categorise a post.   nope

Have you ever tried to delete your updates (data) or had concern about the privacy of your data hosted in Madmica? If yes, would you prefer to host your data in your server to avoid privacy risks?

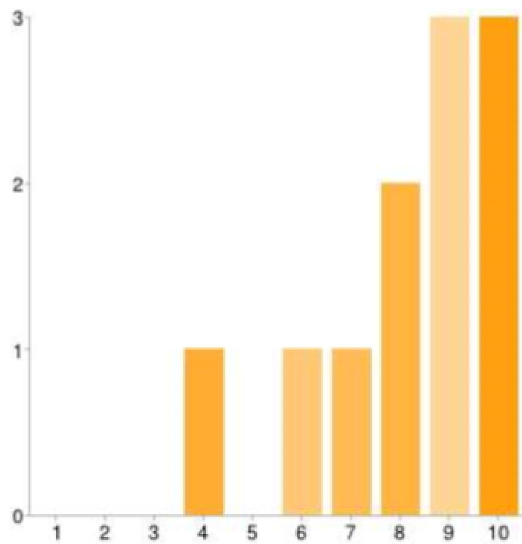
no   No.   no, no privacy concern   Sure.   No   no   no but I prefer to host my data in my server   may be my pictures

How easy to post a new item in Madmica?



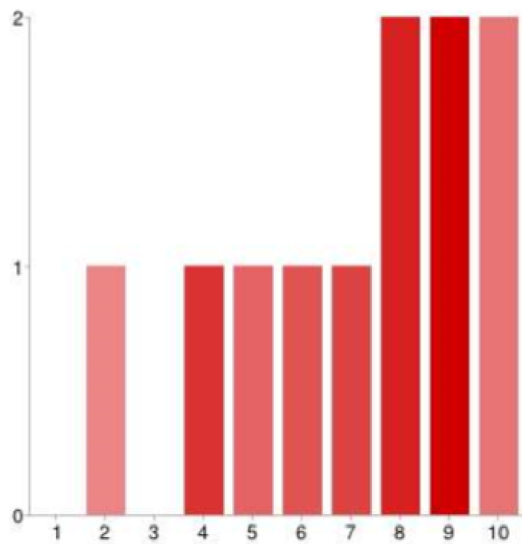
1	**0**	0%
2	**0**	0%
3	**0**	0%
4	**1**	9%
5	**1**	9%
6	**0**	0%
7	**1**	9%
8	**3**	27%
9	**4**	36%
10	**1**	9%

How easy to select a category of your post in Madmica?



1	**0**	0%
2	**0**	0%
3	**0**	0%
4	**1**	9%
5	**0**	0%
6	**1**	9%
7	**1**	9%
8	**2**	18%
9	**3**	27%
10	**3**	27%

How easy it is to keep track with the posts of your friends?



1	**0**	0%
2	**1**	9%
3	**0**	0%
4	**1**	9%
5	**1**	9%
6	**1**	9%
7	**1**	9%
8	**2**	18%
9	**2**	18%
10	**2**	18%

Number of daily responses

Sayooran N

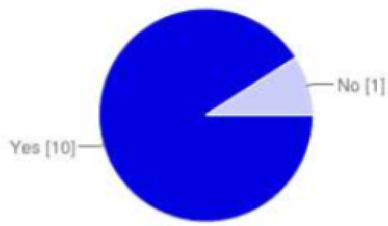
Edit this form

11 responses

View all responses

Summary

Did you use the filter bubble visualization?



Yes	**10**	91%
No	**1**	9%

If yes, what do you think it represents?

Shows my interests to different categories (category view) or to posts of a friends (friend view) my friend's and my interests It reflects the interest a person showed in certain category of posts. inside the bubble is the categories of the news I like while the hidden news belong to the categories outside the bubble, if friend view is selected, the same as category but for friends What'll be displayed on my stream Thing that made visible from us automatically or we have make them invisible manually. It represent my interest and posts i will receive It represents my interest category and that of others that is filtered from me for prioritizing the items in my news feed.

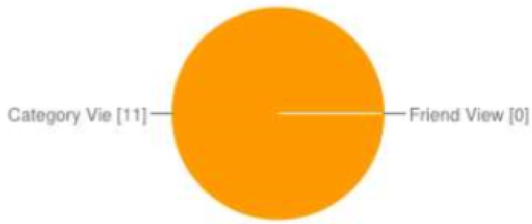
What do you think about the category view in the visualization?

nice, comapct visualization it was ok It interesting helps to avoids the posts you do not like It's good, easy to use good good, but some categories are repeated I think category view is useful to visualize my choice of posts and help me to some what sort the posts I wan to have a look on my wall. Category wise news/posts I liked it It's good, There should be more categories It's awesome

What do you think about the friend view in the visualization?

didn't use it I didn't use it That's quite useful, sometimes we find the posts of a friend uninteresting, so we could filter them. It will be interesting too,I can avoid some friends post that is not interested to me.   i find it interesting   good   I used Catgory view   Same as above   Friend view is good in customizing my wall and to refrain from receiving posts I am not interested to.   wasn't that necessary for me. there was no need to filter friends since they are few in number.   What my friends are watching

Which view do you prefer?



Category View	**11**	100%
Friend View	**0**	0%

Can you think of other views that could be useful?

no   Category and friends looks good   time view! Popular view!   sub-categories, because the listed categories are too limited   By Date and weekh, and popular post -by like and comments   a mixture of both   No   none for now   Their should option for user to generate categories.   more subcategories! but I am wonder about the tradeoff with the simplicity

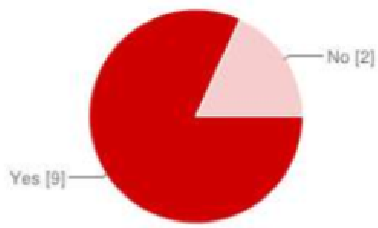
What do think about the position of category/friend circle? Does it have a meaning?

No.   I don't think so, but it is really in a clear position. it may be in the same position as groups in Facebook!   I think positions indicates my involvement to a particular category or my interaction to a friend.   the closest to the center are the categories that are more interesting   yes   yes, it's good, it can help one sort by friend and category simultaneously   The most recent one are in the center   No idea!   Never thought about it

What do you think about the size of the category/friend circle? Does it have a meaning?

The most used one is the biggest one   more interest about the category of larger size   No.   Bigger size most likely to indicate my interest and my interaction to categories and friends respectively.   yes   I didn't notice that   I think the volume of each bubble presents the amount of data inside.   Yes, the popularity or activity level   nope   Maybe bigger ones have more posts?

Did you try dragging a category/friend circle from the outside to the inside of the bubble?

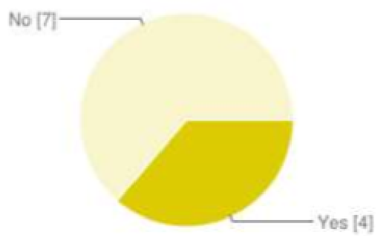


Yes	**9**	82%
No	**2**	18%

If yes, did you notice an effect after of doing this?

yes. yes, the friend and category became part of my interest circle posts of that category/friend present to me again. Did not notice carefully.. :(yes Yes. Items belong to that bubble will be displayed on my stream yes, more news appear

Did you try dragging a category/friend circle from the inside to the outside of the bubble?



Yes	**4**	36%
No	**7**	64%

If yes, did you notice an effect after of doing this?

yes yes, the category became part of my interest circle Yes more news hidden

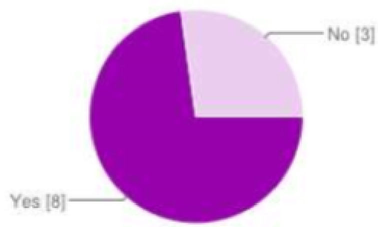
Why does a box full of links appear when you click on a category/friend circle which is outside the bubble? Can you guess which of these links were more recently posted?

yes. I didn't notice that I am not viewing those links on my wall. yes No idea Haven't noticed this feature. never clicked on the bubbles no idea To see the posts that are related to that that friend or category. The topper the link is the newer it is. the first appear in the list is the most recent Not sure

Did you realize the system was filtering the posts from your friends away from your stream?

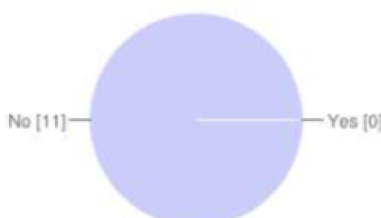
Yes	**9**	82%
No	**2**	18%

Does the visualization help you to understand how the filtering works?



Yes	**8**	73%
No	**3**	27%

Did you notice that system orders a coffee when you post an item?

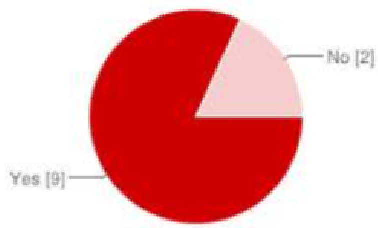


Yes	**0**	0%
No	**11**	100%

Does the visualization give you a feeling of control over the stream of posts from your friends?

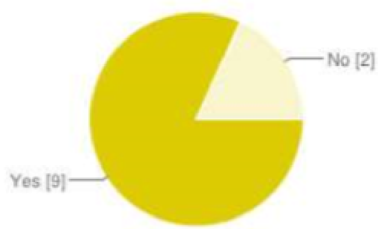
Yes	**10**	91%
No	**1**	9%

Does the visualization increase your trust in the filtering process?



Yes	**9**	82%
No	**2**	18%

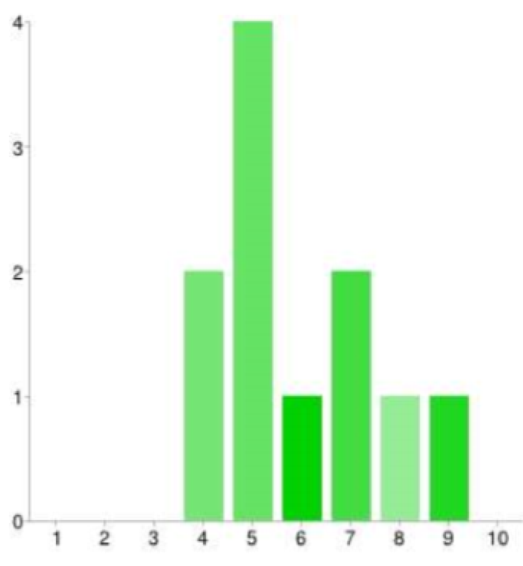
Does the visualization help you trust the system more?



Yes	**9**	82%
No	**2**	18%

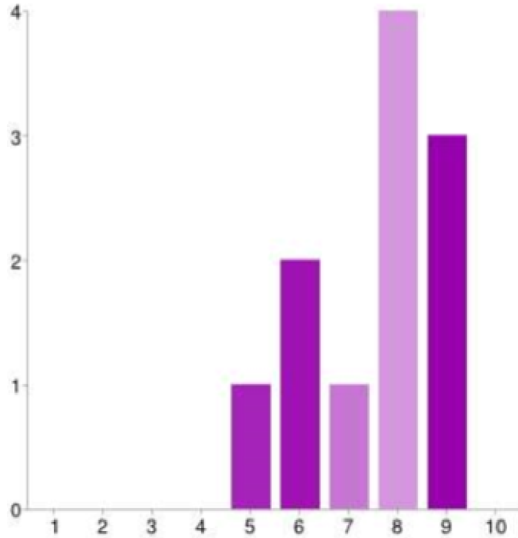
Trust in the System

Trust in the System before using the filter bubble:



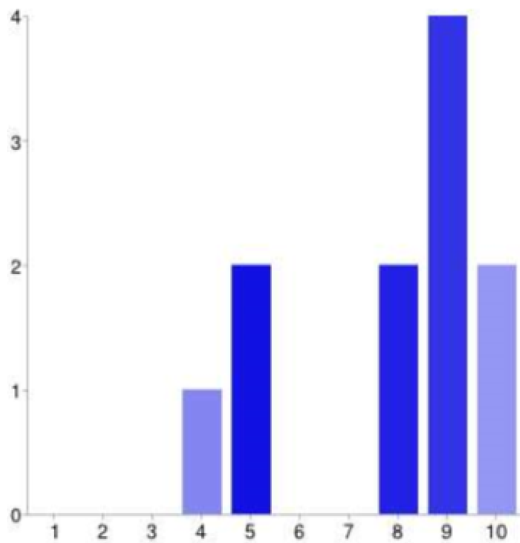
1	**0**	0%
2	**0**	0%
3	**0**	0%
4	**2**	18%
5	**4**	36%
6	**1**	9%
7	**2**	18%
8	**1**	9%
9	**1**	9%
10	**0**	0%

Trust in the System after using the filter bubble:



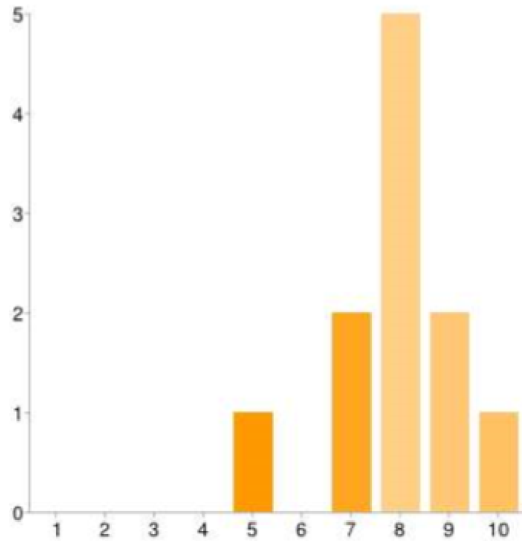
1	**0**	0%
2	**0**	0%
3	**0**	0%
4	**0**	0%
5	**1**	9%
6	**2**	18%
7	**1**	9%
8	**4**	36%
9	**3**	27%
10	**0**	0%

Trust in the System after seeing the hidden posts:



1	**0**	0%
2	**0**	0%
3	**0**	0%
4	**1**	9%
5	**2**	18%
6	**0**	0%
7	**0**	0%
8	**2**	18%
9	**4**	36%
10	**2**	18%

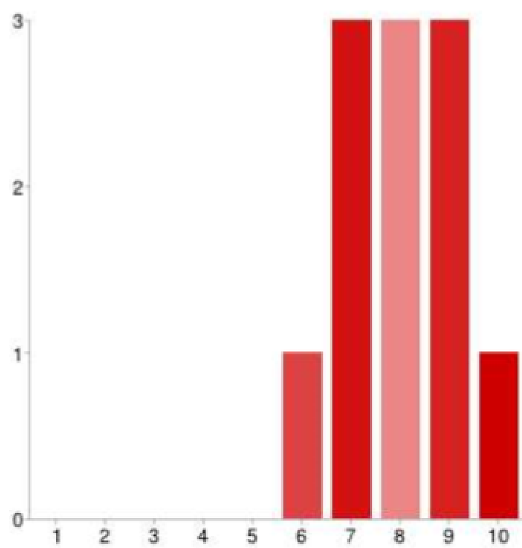
Level of transparency in filtering provided by the system:



1	**0**	0%
2	**0**	0%
3	**0**	0%
4	**0**	0%
5	**1**	9%
6	**0**	0%
7	**2**	18%
8	**5**	45%
9	**2**	18%
10	**1**	9%

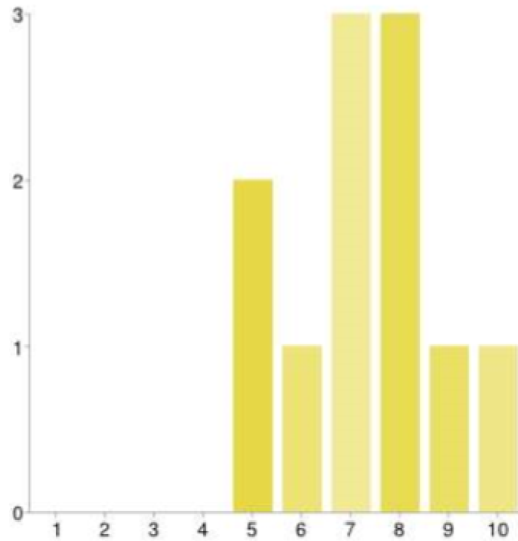
Filter Bubble Visualization

Aesthetically pleasing



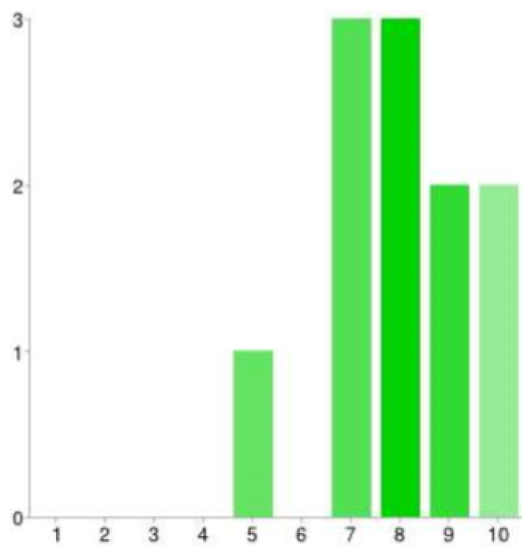
1	**0**	0%
2	**0**	0%
3	**0**	0%
4	**0**	0%
5	**0**	0%
6	**1**	9%
7	**3**	27%
8	**3**	27%
9	**3**	27%
10	**1**	9%

Friends View



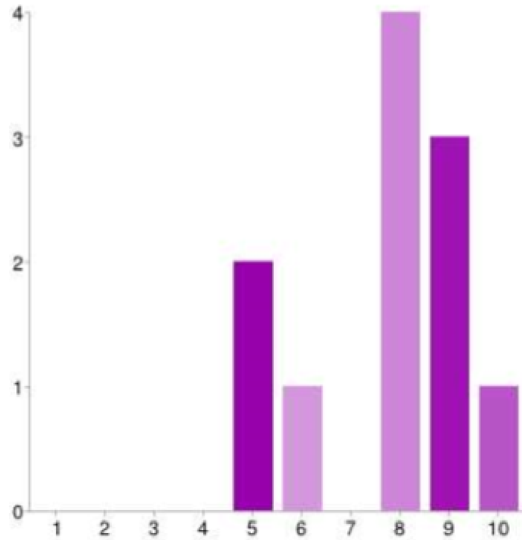
1	**0**	0%
2	**0**	0%
3	**0**	0%
4	**0**	0%
5	**2**	18%
6	**1**	9%
7	**3**	27%
8	**3**	27%
9	**1**	9%
10	**1**	9%

Category View



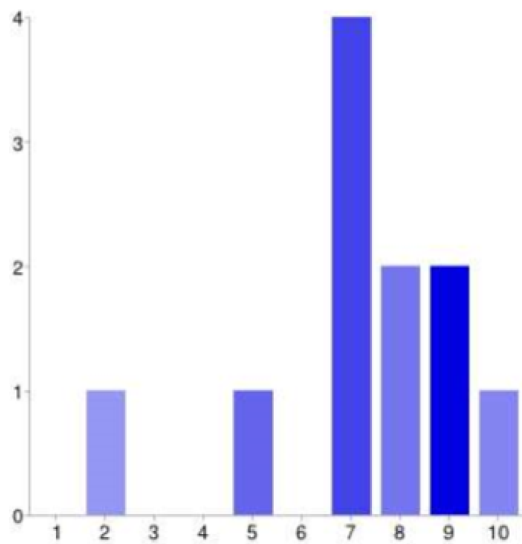
1	**0**	0%
2	**0**	0%
3	**0**	0%
4	**0**	0%
5	**1**	9%
6	**0**	0%
7	**3**	27%
8	**3**	27%
9	**2**	18%
10	**2**	18%

Awareness about hidden posts



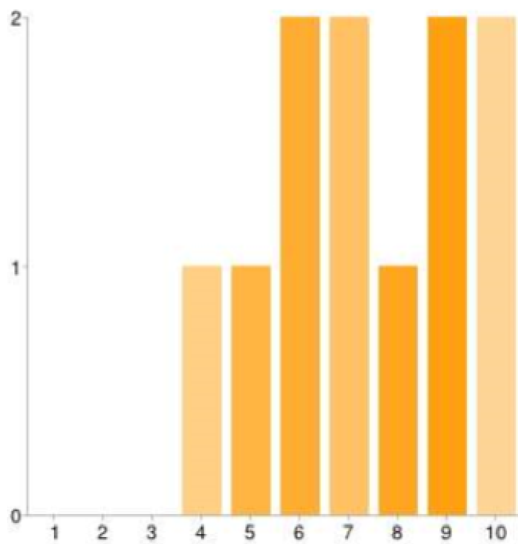
1	**0**	0%
2	**0**	0%
3	**0**	0%
4	**0**	0%
5	**2**	18%
6	**1**	9%
7	**0**	0%
8	**4**	36%
9	**3**	27%
10	**1**	9%

Arrangement of information on screen



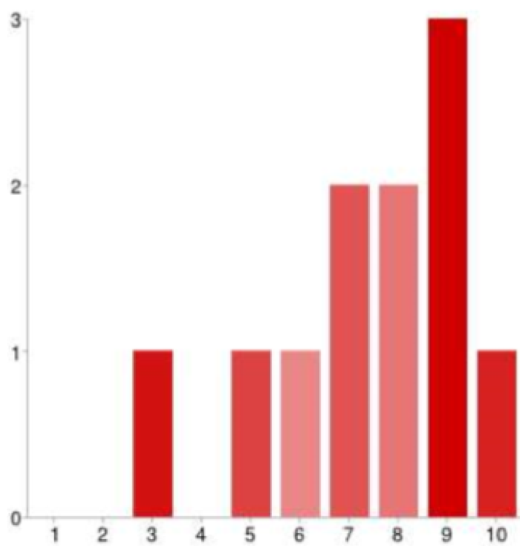
1	**0**	0%
2	**1**	9%
3	**0**	0%
4	**0**	0%
5	**1**	9%
6	**0**	0%
7	**4**	36%
8	**2**	18%
9	**2**	18%
10	**1**	9%

Manipulation of interest (categories) / friend circles (dragging in and out)



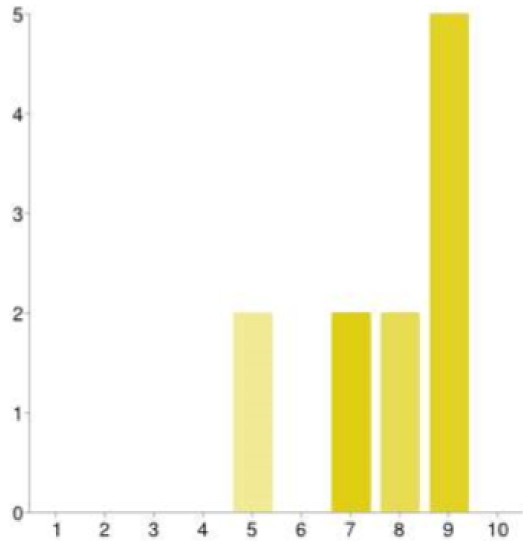
1	**0**	0%
2	**0**	0%
3	**0**	0%
4	**1**	9%
5	**1**	9%
6	**2**	18%
7	**2**	18%
8	**1**	9%
9	**2**	18%
10	**2**	18%

Finding an interest which is not inside your filter bubble



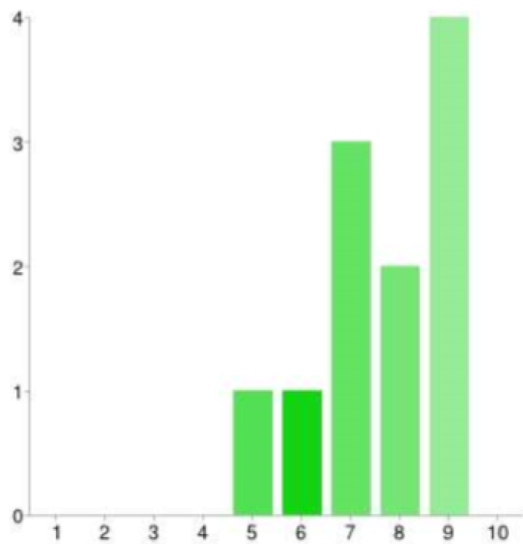
1	**0**	0%
2	**0**	0%
3	**1**	9%
4	**0**	0%
5	**1**	9%
6	**1**	9%
7	**2**	18%
8	**2**	18%
9	**3**	27%
10	**1**	9%

Discovering new interests



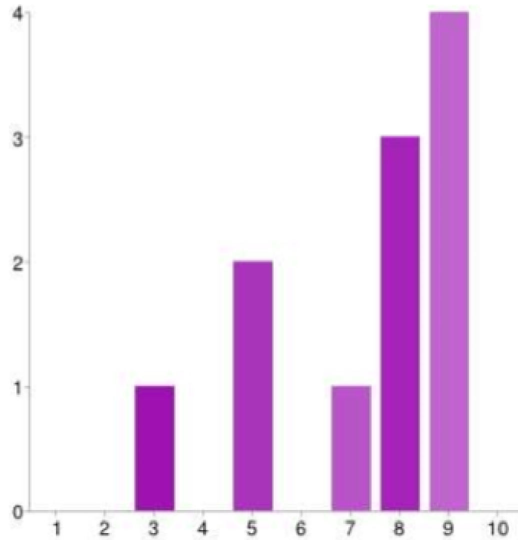
1	**0**	0%
2	**0**	0%
3	**0**	0%
4	**0**	0%
5	**2**	18%
6	**0**	0%
7	**2**	18%
8	**2**	18%
9	**5**	45%
10	**0**	0%

Discovering the interests of friends



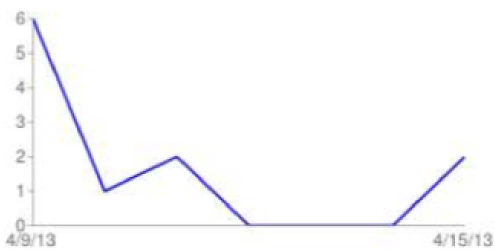
1	**0**	0%
2	**0**	0%
3	**0**	0%
4	**0**	0%
5	**1**	9%
6	**1**	9%
7	**3**	27%
8	**2**	18%
9	**4**	36%
10	**0**	0%

Discovering in which areas your friends are most interested



1	**0**	0%
2	**0**	0%
3	**1**	9%
4	**0**	0%
5	**2**	18%
6	**0**	0%
7	**1**	9%
8	**3**	27%
9	**4**	36%
10	**0**	0%

Number of daily responses



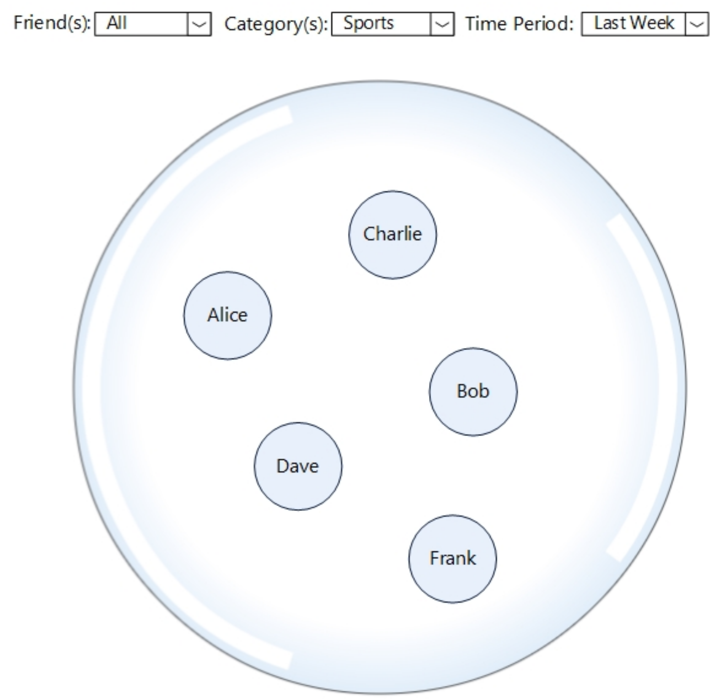
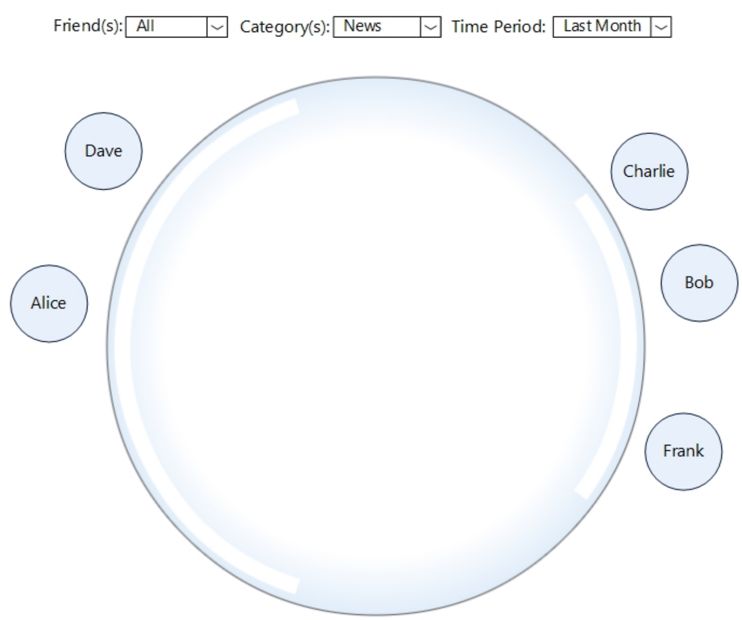
APPENDIX B

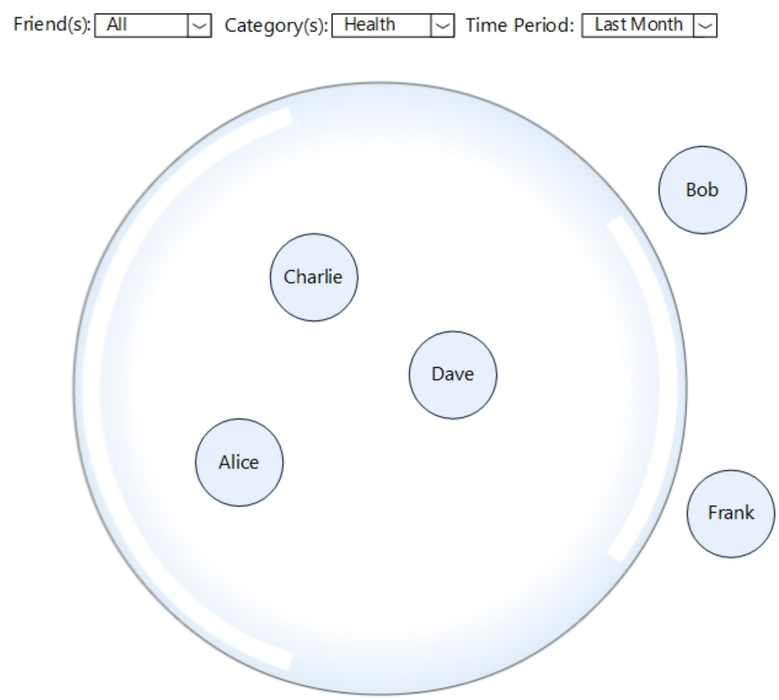
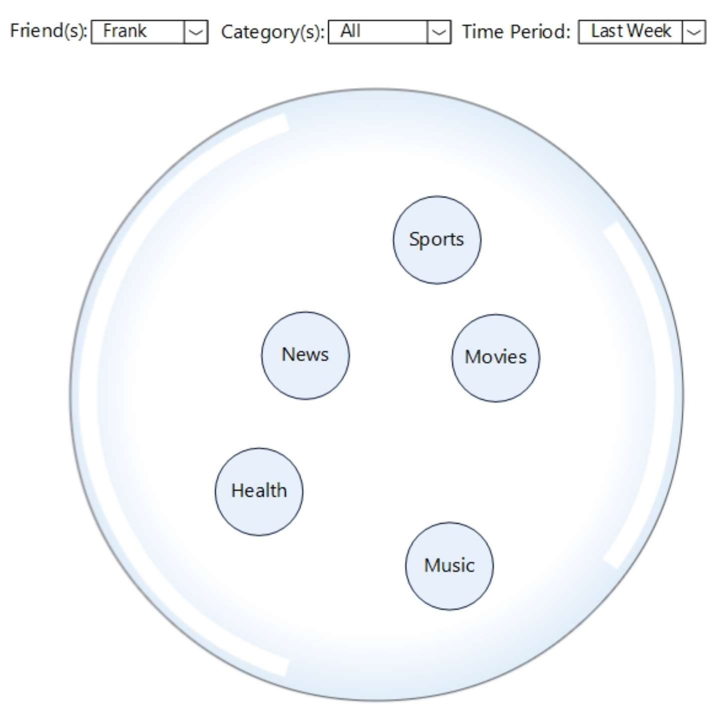
QUALITATIVE STUDY QUESTIONNAIRE AND RESPONSES

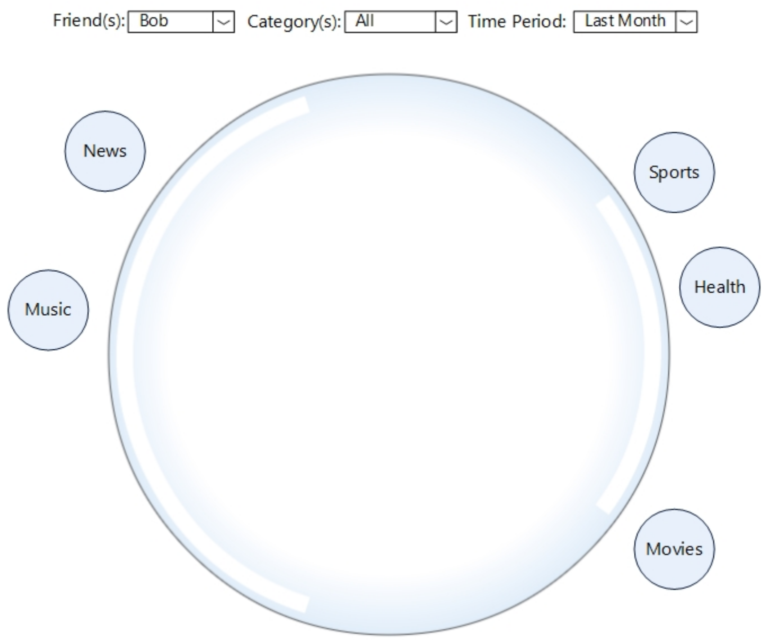
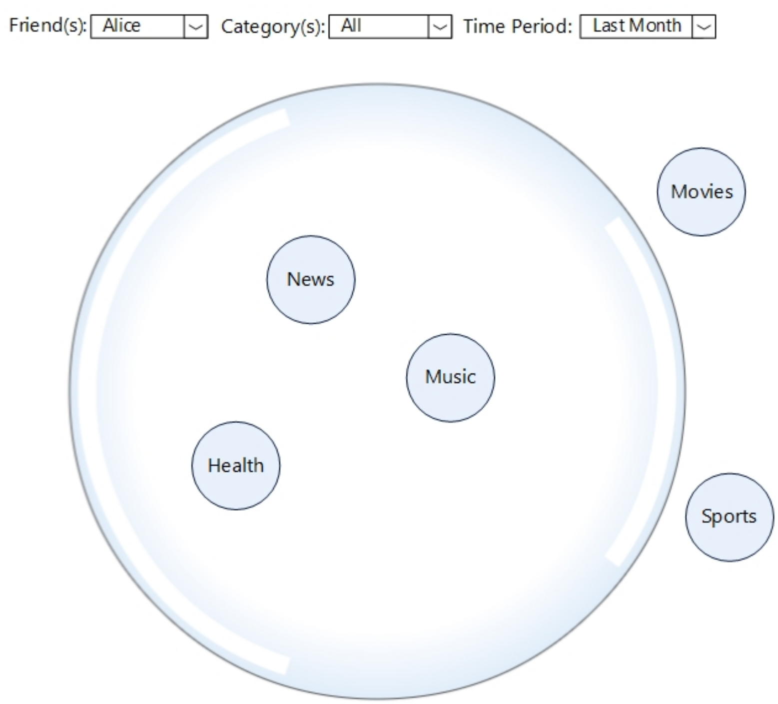
Getting Familiar with the System

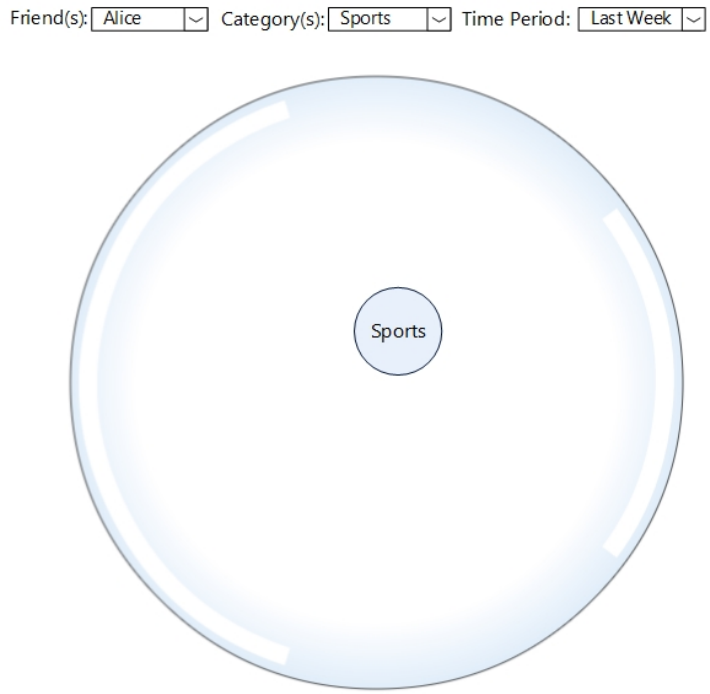
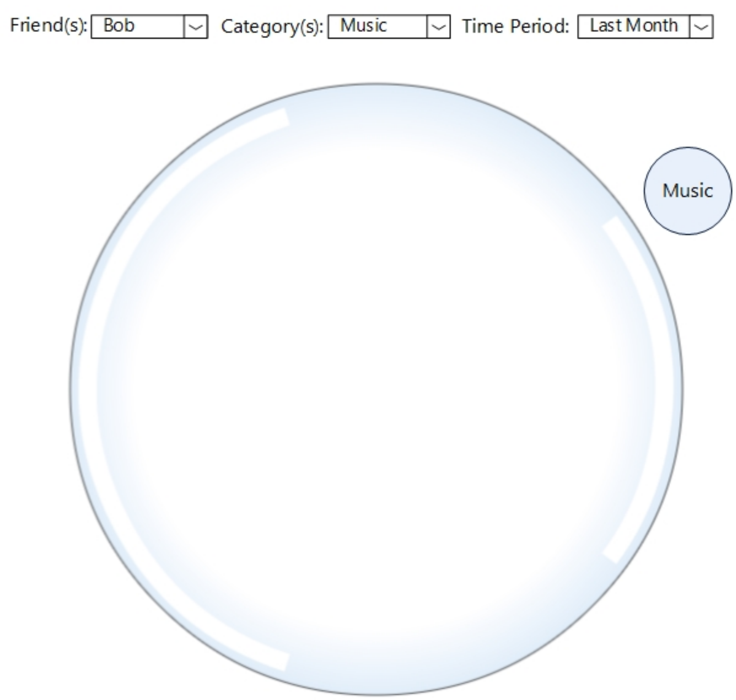
MADMICA is an online social network similar to Facebook. MADMICA system customizes and filters the news posts shared by your friends based on categories of your interests (e.g. news, sports, music, movies, health, etc.) and the relationships with your friends. In the long run, you will ultimately see the news feed related to **your interests only** and will have no opportunity of discovering news feed not related to your current interests, or developing new interests. This will lead to “the filter bubble” problem in which you are trapped in a world filled with only news feed matching your interests. We are trying to overcome this problem by creating a visualization of the filter bubble where all the news feed are organized into categories and friends.

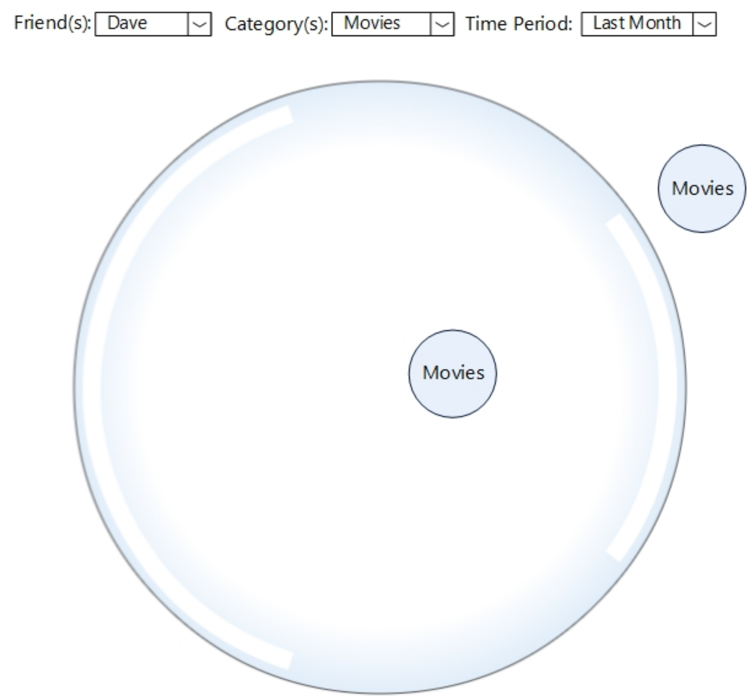
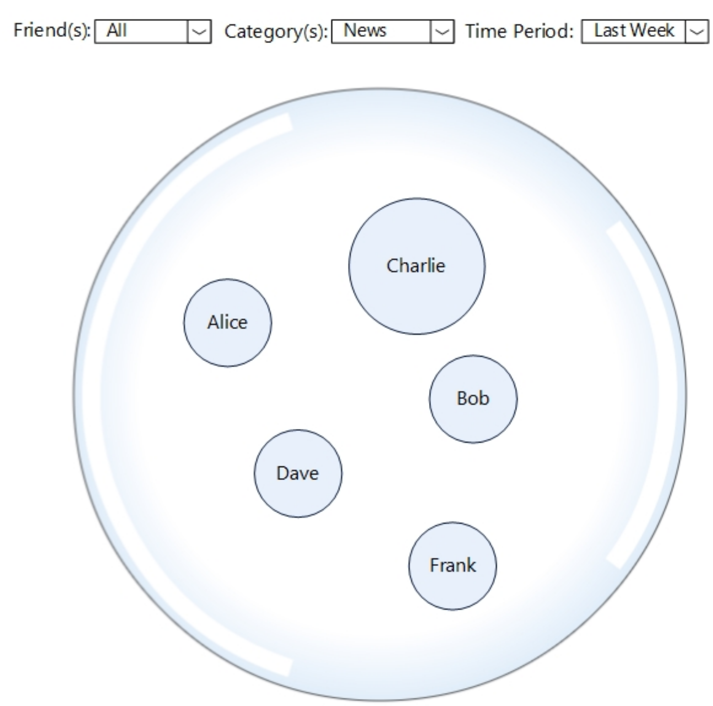
To start with, first get familiar with the news feed homepage in MADMICA. To get to know more about what the system has done to your news feed, click the bubble icon on the left side of the news feed homepage and then you will be redirected to the filter bubble visualization. You will be given 10 minutes to use the visualization and get familiarize with it. Once you have an understanding about the visualization, please answer the questionnaire in the next page.

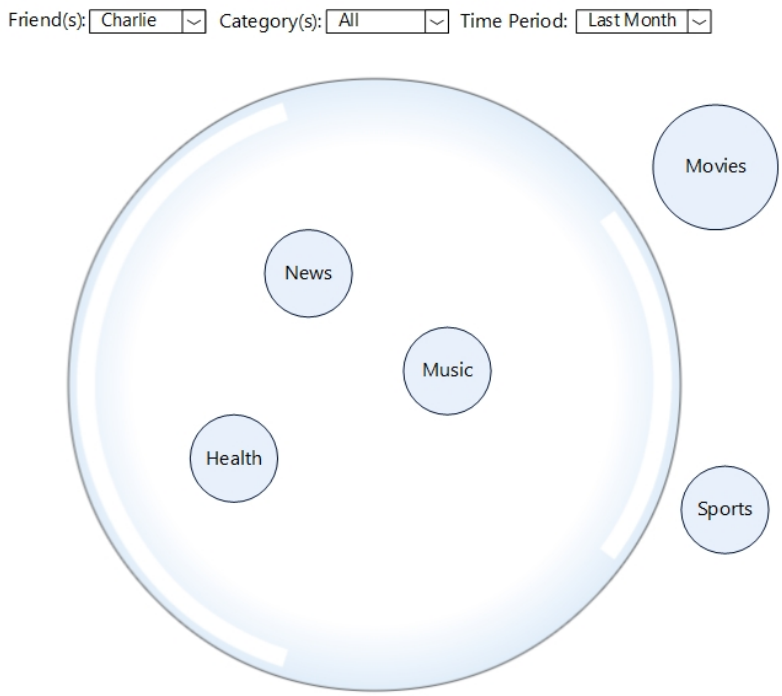
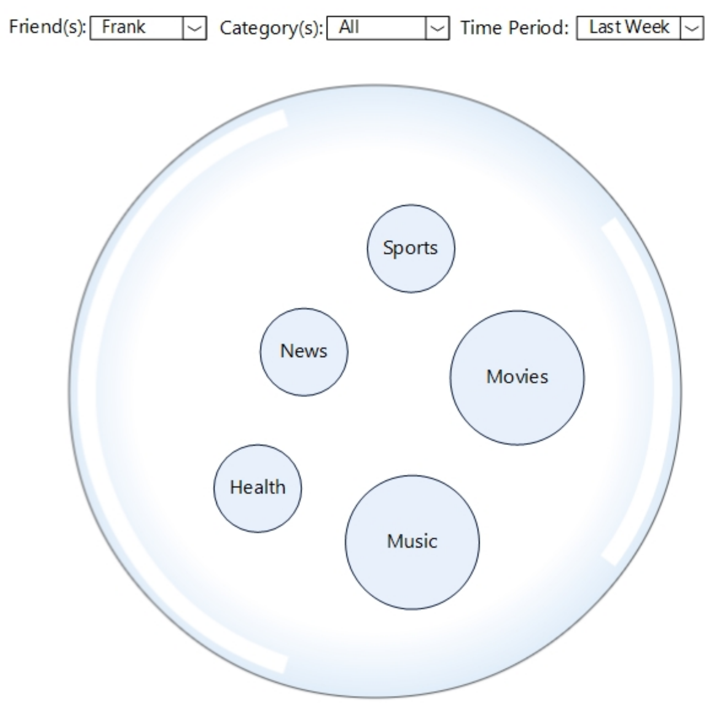
<table>
<tr><th>No</th><th>View</th><th>Question</th></tr>
<tr><td>1</td><td>
</td><td>What do you understand about this view?</td></tr>
<tr><td>2</td><td>
</td><td>What do you understand about this view?</td></tr>
</table>

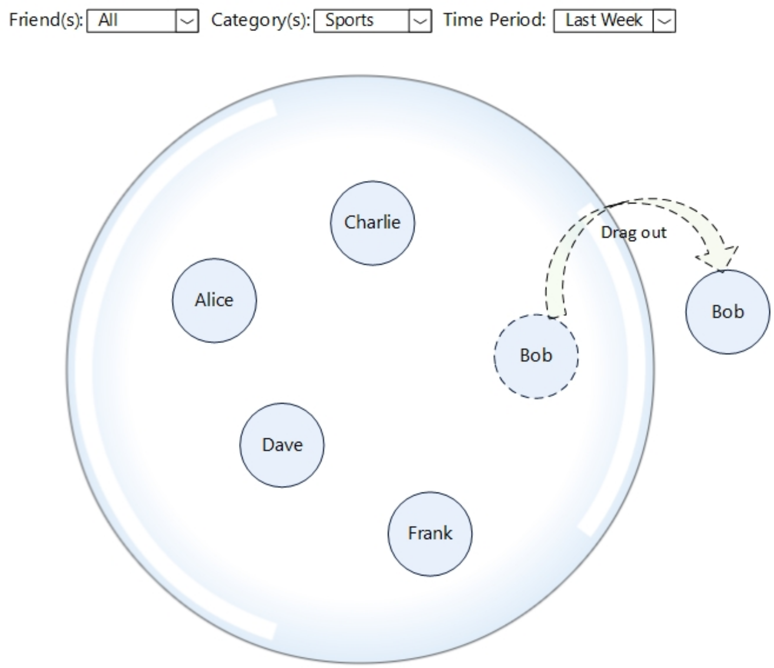
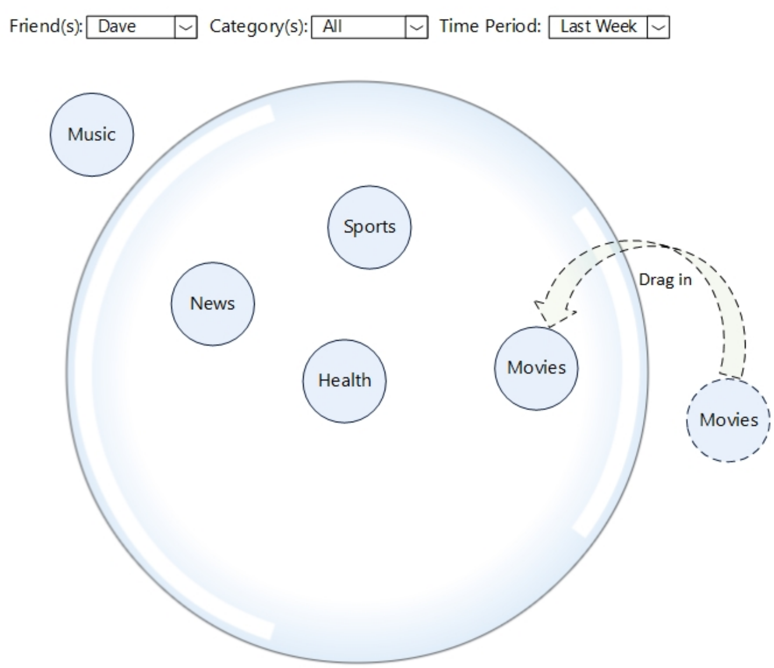
3		What do you understand about this view?
4		What do you understand about this view?

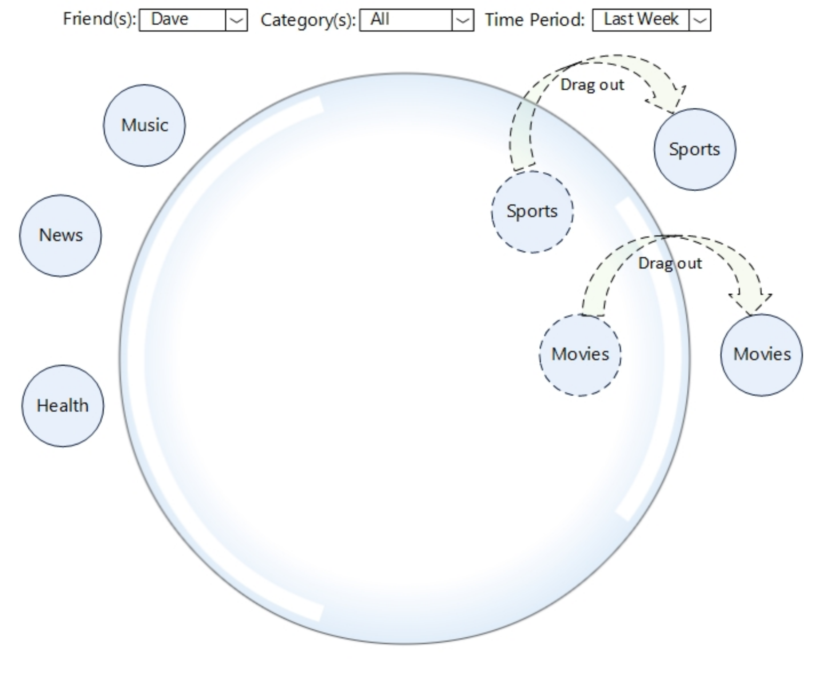
5		What do you understand about this view?
6		What do you understand about this view?

7		What do you understand about this view?
8		What do you understand about this view?

9		What do you understand about this view?
10		What do you understand about this view?

11		What do you understand about this view?
12		What do you understand about this view?

13		You drag Bob outside the filter bubble. What does this action really mean? What do you understand about the resulting view?
14		You drag Movies category inside the filter bubble. What does this action really mean? What do you understand about the resulting view?

15		You drag both Sports and Movies outside the filter bubble. What does this action really mean? What do you understand about the resulting view?

Qualitative Study – User Responses (Audio Transcript)

Participant 1

View	Timespan	Content
1	0:00.0 - 0:42.6	My friends shared a category of sports. They are big fans of sports. Just like me. They post some news about sports on homepage. When they post about sports news I would get informed by interest bubble and then I can reply.
2	0:42.8 - 1:10.7	I think these friends, they don't really care about news. So even though I like news. They don't like it. So they post anything about news, I would not get, they will not show in my homepage.
3	1:10.9 - 2:02.5	I think Charlie, Dave and Alice, they are in the interest bubble. So they care about health. And they go to gym. And they like working out/ anything about health. But Bob and Frank, they don't really care about health. They may be interested in other categories. But I still friend them. Not because of health but because of other interest categories.
4	2:04.4 - 2:40.5	Frank, he may be sharing many interest with me. Sports, News, Health, Movies, Music. I will share any information in this categories with him. And he will share with me as well.
5	2:40.5 - 3:20.5	So Last month, Bob doesn't post anything of all category of my interests.
6	3:22.2 - 4:10.9	Last month Alice post something about News, Music, Health and I saw them and I replied to Alice. And the system considered that I was interested in News, Health, and Music as well as for Alice. May be Alice posts something about Movies or sports, but I did not reply. So the system considered that I was not interested in Movies and sports.
7	4:10.9 - 4:52.8	Alice went work out and may be post some pictures on working out and I saw that and I replied to that and we had some interaction on homepage in Last week.
8	4:52.8 - 5:19.1	For the last month, Bob didn't post anything about music. Nothing else.
9	5:19.1 - 6:32.2	Two movies. May be under the movies there are categories like comedy movies. Action movies. So the system put the category of movies I like into the filter bubble and filter the category of movies I don’t like outside the bubble.
10	6:32.2 - 7:07.9	Last week these friends shared something about News with me. And the biggest one is Charlie and I guess Charlie posted most pieces of News on homepage or I replied Charlie most about News.

11	7:07.9 - 8:15.9	Last month, Charlie may share something about Music, News, and Health on homepage with me. She didn't anything about Sports and less for Movies.
12	8:15.9 - 9:02.9	In this bubble, Movies and Music are the two biggest bubbles. Other three are basically the same size. So I guess, last week Frank shared something about Movies and Music with me more than News, Sports and Health.
13	9:02.9 - 10:06.9	If I drag Bob out, maybe I think he said something about sports that freaked me out and made me mad. And I don't want to see anything about sports from Bob. So I dragged him out of the filter bubble. So I don't need to read his message about Sports on homepage.
14	10:06.9 - 11:05.3	Last week I found that Dave may be has same taste with me in Movies so I thought if I drag Movies category inside the filter bubble, maybe we can share more about movies and we could hang out together and we can ask each other, hey when are you available to watch movie on homepage.
15	11:06.5 - 12:28.5	Because I saw Music, News, and Health outside of filter bubble, so I guess I dragged these categories out before and then I drag sports and movies out. May be also categories is All, I don't want to know anything about Dave on what he did on last week and it will not show in my homepage.

Participant 2

View	Timespan	Content
1	0:00.0 - 0:37.2	It would seem that all these users that appear inside the bubble, they actually because centered small bubbles, means that they actually posted like I would have to say not just one or two posts but actually a lot of posts and they are really into sports just last week. The bigger the bubble is the less posts we have and smaller and kind of centered the more about the topic.
2	0:37.1 - 1:11.0	They didn't make any posts about the category during last month. That's why none of them appear inside the bubble. It's kind of filtered out.
3	1:11.0 - 2:25.0	Three of the users in inside the bubble. So basically I was getting only posts from them about that topic. And the rest I didn't get any posts during last month. I think the duration has to play a factor in this.
4	2:25.0 - 2:54.1	For Frank, he has a lot of common interest that I have at the same time. Since most of the categories are inside the bubble. These are kind of main interests right now.

5	2:54.1 - 3:50.1	Bob, there is nothing in common with him. I don't have anything in common. None of his posts are actually related to me at all. All of the categories are outside the bubble. So his posts will be filtered out.
6	3:50.1 - 4:27.7	Have some kind of stuff in common and kind of discuss the same posts that is inside the bubble basically News, Music and Health. Not too long and some kind of recent since the period is last month. So it might be just alike the relationship with this user would be like friend acquaintance.
7	4:27.7 - 6:36.0	Alice would be posting about sports during last week. This is very specific. Just discussing one category.
8	6:36.0 - 6:58.6	Same as for 7th view. So it's targeting just one subject and category. For this user, it's been filtered out. It's out of the area of my interest. So his interest is not same as the mine.
9	6:58.6 - 7:48.6	This one I would have to say, inside the category of movies, we just share certain stuff not everything like type of genre of movies the he like. We don't same kind of movies that's why we have one outside and one inside.
10	7:48.6 - 8:22.7	All of them have contributed to News /talked about News. But the people that I have more in common are the people with small bubbles. The larger the bubble is of less interest to me. Small bubbles contribute more than larger one. It might be the opposite. But from what I saw from the newsfeed, the smaller the bubble more posts about whatever the interest is.
11	8:23.6 - 9:39.7	I only have three interest in common with him. Other two categories are just filtered out and any posts he put about these two I wouldn't see because I don't have the common interest.
12	9:39.7 - 10:04.1	Everything Frank would post, I would see.
13	10:04.1 - 11:19.3	Either Bob no longer posts about that. Type of posts he post about Sports doesn't fall to my interest. That's why dragged out of filter bubble.
14	11:19.3 - 14:09.6	Right now I have interest about the movies he post about. He actually have interest about movies that I post about.
15	14:09.6 - 14:45.3	This is basically I want to tell the guy that I don't have any common interest with him. I lost interest with Sports and Movies.

Participant 3

View	Timespan	Content
1	0:00.0 - 0:36.6	Charlie, Alice, Dave, Bob and Frank share the same bubble of sports during last week. They have same interest on Sports.
2	0:36.6 - 1:03.2	Among all the friends the bubble indicates news thing, none of them are interested in news during last month.
3	1:03.2 - 1:22.1	Since last month the bubble represents the health that is the interest topic for Charlie, Alice and Dave. But not for Bob and Frank.
4	1:22.1 - 1:41.2	This time the bubble represents a person called Frank. For that particular person, Health, News, Sports, Movies and Music are of interest since last one week.
5	1:41.2 - 2:06.7	It means for Bob, Neither Music nor News, Sports, Health and Movies is the topic of interest. It's something else which is not mentioned here.
6	2:06.7 - 2:32.0	For Alice, the bubble mainly represents what you are interested. What lying outside the bubble means you are not interested for that particular period?
7	2:32.0 - 2:43.4	For Alice, since last week, she is following or I can say the interest newsfeed is sports.
8	2:43.4 - 2:56.9	For last one month, Bob is not interested in music maybe it's out of the bubble of interest for Bob.
9	2:56.9 - 3:26.8	Sometimes Dave is interested in Movies and sometimes not.
10	3:26.8 - 3:53.8	Charlie, Alice, Dave, Bob, and Frank are interested in News. They share the same common interest since last week.
11	3:53.8 - 4:06.5	For Charlie, since last one week, out of all categories his interests were News, Health and Music. Two things he was not interested or less interested are Movies and sports.
12	4:06.5 - 4:14.1	For Frank, since last one week, he is following sports, news, health, music and movies.
13	4:14.1 - 6:00.1	Bob was actually interested. Did I unfriend him. He was no more my friend. He lost interest.
14	6:00.1 - 6:19.6	Interested in sports, news, and health. Not interested on Music. But I dragged him and made him take interest in movies.

15	6:19.6 - 7:33.7	Dave was interested in two things since last one week. How can I drag interest out of somebody? I unfriend that guy.

Participant 4

View	Timespan	Content
1	0:01.3 - 1:08.0	For the friends, we will see something related to sports from last week.
2	1:08.0 - 1:30.4	If I drag all the circles outside the bubble, that means I ignore all the news from these people since last month.
3	1:30.4 - 2:08.6	Bob and Frank out of big bubble, that means I ignored something about health they shared since last month.
4	2:08.6 - 2:53.9	For Frank, what I want to focus on are sports, music, movies, health, news.
5	2:53.9 - 3:18.1	I ignore his sharing about sports, news, music, health and movies.
6	3:18.1 - 3:47.0	I only want to hear sharing from Alice about Health, News and Music.
7	3:47.0 - 4:25.6	For specific person, specific category, I will see only about sports news from Alice since last week.
8	4:25.6 - 4:53.9	I ignore Bob's Music posts. The music posts were out of my interface from Bob.
9	4:53.9 - 6:27.0	Both Movies? I don't know... May be for long run...
10	6:27.0 - 7:09.3	I would see news from Alice, Dave, bob, frank and Charlie. Bigger size is related to number of posts or frequency of sharing news.
11	7:09.3 - 8:11.5	Similar as last view. Difference is Movies are larger size. He shares more about Movies. Less about other categories.
12	8:11.5 - 8:21.5	----
13	8:21.5 - 8:49.4	All share something about sports. I only want to see others posts instead of Bob.
14	8:49.4 - 9:25.9	I want to see all the sharing about News, Sports, health and movies from Dave.
15	9:25.9 - 9:55.3	I don't want to see any sharing from Dave. So I drag him out.

Participant 5

View	Timespan	Content
1	0:00.0 - 1:05.5	People are more in the middle. So they are more interested in the category selected. Bubbles are same size, so they is no difference between who is more interested. More outside less interested.
2	1:05.5 - 1:55.2	Those people are not interested on the selected category. Because they are all outside the bubble.
3	1:55.2 - 2:48.5	In the last month, those three people mention more on the category selected or more interest in the category selected. Those two people may have interest in the category selected. But just last month they have less interest.
4	2:48.5 - 3:10.6	In last week, this person has the most interest in the five small bubbles selected because category is all.
5	3:10.6 - 4:32.5	Bob did have interest in five small bubbles in last month. There are some possibilities either he has nothing updated in his page or he has changed his interest.
6	4:32.5 - 5:03.1	Her main focus during last month was those three bubbles. And mentioned less about two bubbles outside.
7	5:03.1 - 5:31.2	In the last week she just focused on sports. She is interested in sports during last week.
8	5:31.2 - 5:58.5	Bob has interest in music last month but he mentioned nothing about music. He has changed his interest to other things.
9	5:58.5 - 7:42.0	Dave is interested in movies for last month. Perhaps half, half. No idea why there are two movies. Half month he mentioned lot about movies. Other half month mentioned less.
10	7:42.0 - 7:59.9	So those five people are interested in news. Big bubble means Charlie has most interest in news. Those other people are just ok.
11	7:59.9 - 9:08.2	In the last month, Charlie has more interest in Health, News, and Music. Because he has five interest. But last month he doesn't mention that at all. But movies has bigger one. So I guess this one used to be the most interested one. But last month just ignored.
12	9:08.2 - 9:41.8	So last week Frank is most interested in Music and Movies. But he was still interested in health, news, and sports but not as interested as music and movies.

13	9:41.8 - 10:28.5	I drag Bob out, because I think, we are both interested in sports but I have different opinions with him. I love badminton and he is interested in hockey. I don't want to see his opinions about his sports interest.
14	10:28.5 - 11:56.2	He was not interested in Music last week. He was not interested in movies last week. But I think I do have some similar things in movies and I just want to add this in the bubble. Perhaps next week we can have same interest in that.
15	11:56.2 - 12:48.7	Dave was interested in sports, movies. But he was not interested in health, news and music. But I guess I just don't want to have similar interests with him.

APPENDIX C

LARGE-SCALE QUANTITATIVE STUDY QUESTIONNAIRE

Providing Awareness and Control of Personalized Newsfeed Filtering in Social Networks

UNIVERSITY OF SASKATCHEWAN

You are invited to participate in this study aiming at evaluating the understandability of an interactive visualization used to provide awareness, understanding and control of personalized newsfeed filtering in Social Networks.

Time to Complete This Survey: ~ 30 minutes

Title of Study: Providing Awareness and Control of Personalized Newsfeed Filtering in Social Networks

Researcher(s): Dr Julita Vassileva, Department of Computer Science, jiv@cs.usask.ca

Sayooran Nagulendra, Department of Computer Science sayooran.nagulendra@usask.ca

Purpose and Procedure: The goal of this study is to evaluate the understandability of an interactive visualization used to provide awareness, understanding and control of personalized newsfeed filtering Social Networks. The study may contribute to the research area of Recommender Systems. To achieve this, we have designed a set of questions that we need you to respond to.

Confidentiality: Once you sign the consent form (by clicking next on this page), we will not require any personal identifiable information such as name, email, etc.

Dissemination of Results: Aggregated results derived from this study will appear in a Masters thesis and articles published in peer reviewed conferences and scientific journals.

Right to Withdraw: Your participation in this study is voluntary, and you may withdraw from the study for any reason, at any time, without penalty of any sort, however, you will not receive the compensation for participation.

Consent to Participate: I have read and understood the description provided; I have had an opportunity to ask questions and my questions have been answered. I consent to participate in the Survey, understanding that I may withdraw my consent any time during the study. **By clicking Next button I consent to participate in this survey**. Use **only Firefox or Chrome** browser to complete this survey.

There are 25 questions in this survey

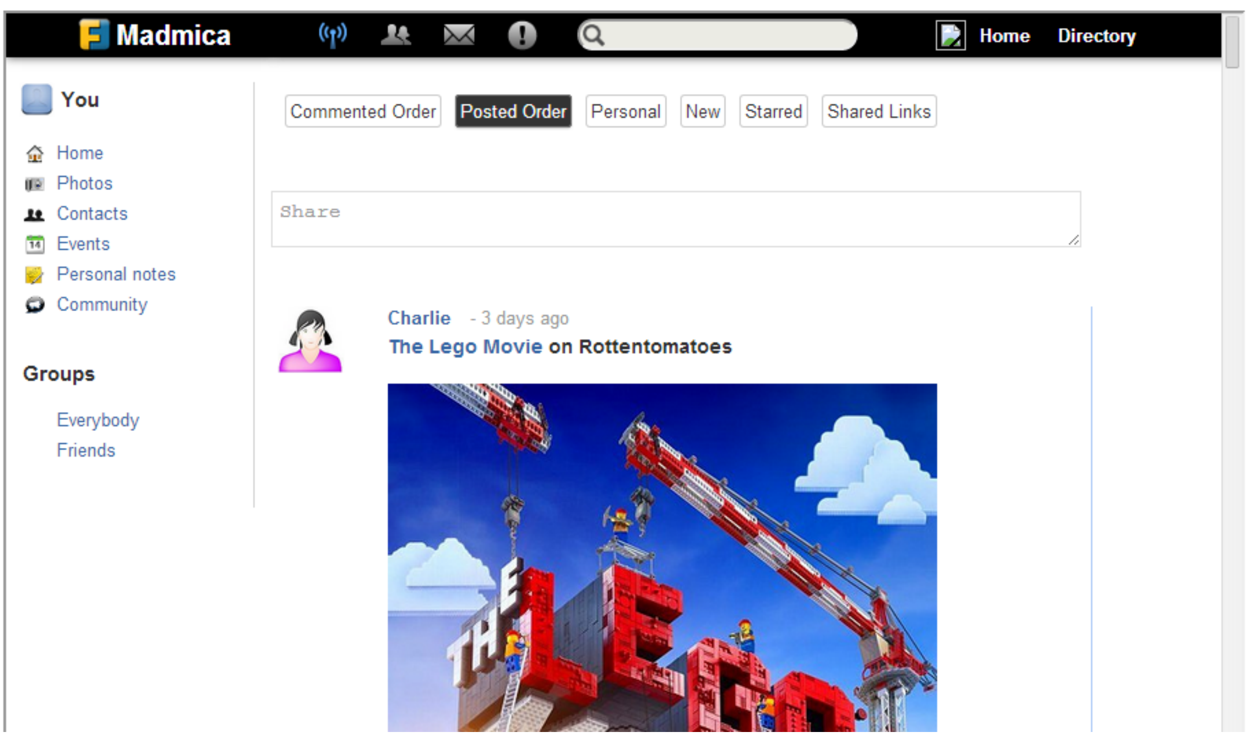
<<This section is only visible for the survey group without Help Text >>

Introduction

MADMICA is an online social network similar to Facebook. MADMICA system customizes and filters the newsfeed (posts shared by your friends) based on categories of your interests (e.g. news, sports, music, movies, health, etc.) and the relationships with your friends. In the long run, you will ultimately see the newsfeed related to your interests only and have reduced opportunity of discovering news not related to your current interests, or developing new interests. This will lead to "the filter bubble" problem in which you will be encapsulated in a bubble of your comfort, seeing only newsfeed related to your interests, and being spared of anything else. We are trying to overcome this problem by creating a visualization of the filter bubble where all the newsfeed (including the filtered newsfeed) from your friends are organized into categories and friends.

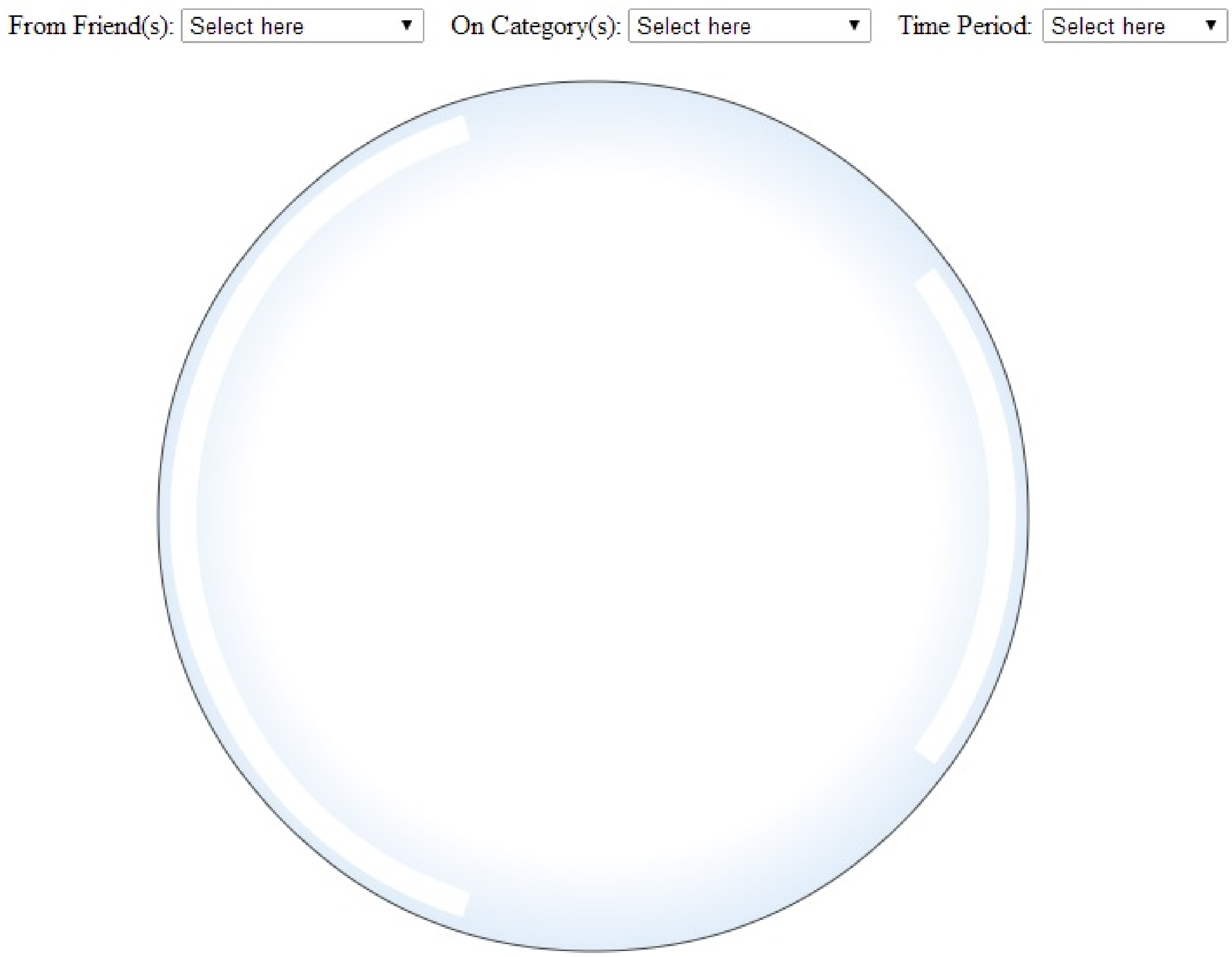
The Visualization displays the posts from your friends organized into categories and friends as circles inside a big bubble as well as outside the big bubble. Depending on the context, you can perform many actions on these category/ friend circles such as click, drag and drop, etc. Each action may represent a functionality of the visualization. In addition to that, there are three filters: From Friend(s), On Category(s) and Time Period, provided with the visualization to create different views.

Assume that you have friends named Alice, Bob, Charlie, Dave and Frank in MADMICA social network. On a particular day your newsfeed homepage in MADMICA looks like the following, (please, make sure to use the vertical scroll bar in the below newsfeed homepage to see all the stories posted by friends)



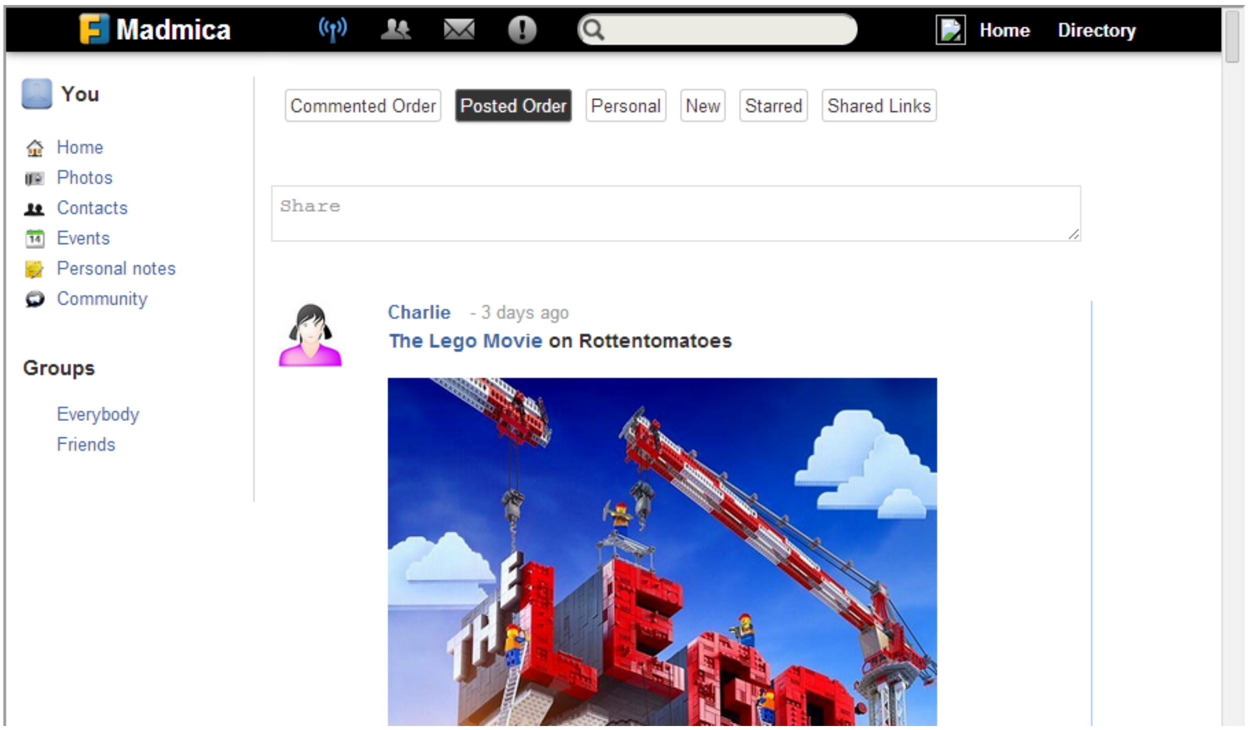
<<This section is only visible for the survey group without Help Text >>

Explore the following visualization by selecting from the dropdown menus, clicking and so on and try to get an idea of its representation by comparing with your newsfeed homepage above. Then try to identify the functionalities (what can be done using it) of the following visualization and note them down for your reference in order to answer further questions.



<<This section is only visible for the survey group with Help Text >>

Assume that you have friends named Alice, Bob, Charlie, Dave and Frank in MADMICA social network. On a particular day your newsfeed homepage in MADMICA looks like the following, (please, make sure to use the vertical scroll bar in the below newsfeed homepage to see all the stories posted by friends)



<<This section is only visible for the survey group with Help Text >>

Read the following help text from the visualization system and try them out with the interactive visualization provided on this page and answer the questions in the next page.

1. What is Filter Bubble?

   MADMICA is an online social network similar to Facebook. MADMICA system customizes and filters the newsfeed (posts shared by your friends) based on categories of your interests (e.g. news, sports, music, movies, health, etc.) and the relationships with your friends. In the long run, you will ultimately see the newsfeed related to your interests only and have reduced opportunity of discovering news not related to your current interests, or developing new interests. This will lead to "the filter bubble" problem in which you will be encapsulated in a bubble of your comfort, seeing only newsfeed related to your interests, and being spared of anything else. We are trying to overcome this problem by creating a visualization of the filter bubble where all the newsfeed shared by your friends are organized into categories and friends.

2. Why some of the circles are inside the filter bubble and others are outside the filter bubble?

   Circles inside the bubble represent the categories of interests or friends who shared posts that were shown in your newsfeed homepage. On the other hand, circles that are outside the bubble represent the categories of interests or friends who shared posts that were hidden (filtered out by system) in your newsfeed homepage.

3. Can I make the hidden newsfeed posts visible in my newsfeed homepage?

   Yes. This can be done by dragging the circle of the category/friend back into the filter bubble. For example, suppose you are interested in receiving posts with the "Health" category from Alice, but the filter bubble shows that category's circle outside of the bubble. By dragging the "Health" circle inside the bubble it will add that category back to your interest list and you will see the posts with that category by that particular friend in your newsfeed homepage in the future.

4. Can I tell the system that I have lost interest on a particular category of posts shared by my friend and filter out those posts in future?

   Yes. Drag the circles of your interest categories outside the filter bubble. This will ensure that you will not receive any posts related to that category by that friend in future.

5. Why does the size of circles vary?

   The size denotes the number of posts. If the number of posts is smaller than 5 then the size would be small and if the number of posts is greater than or equal to 5, then the size of the circle would be big.

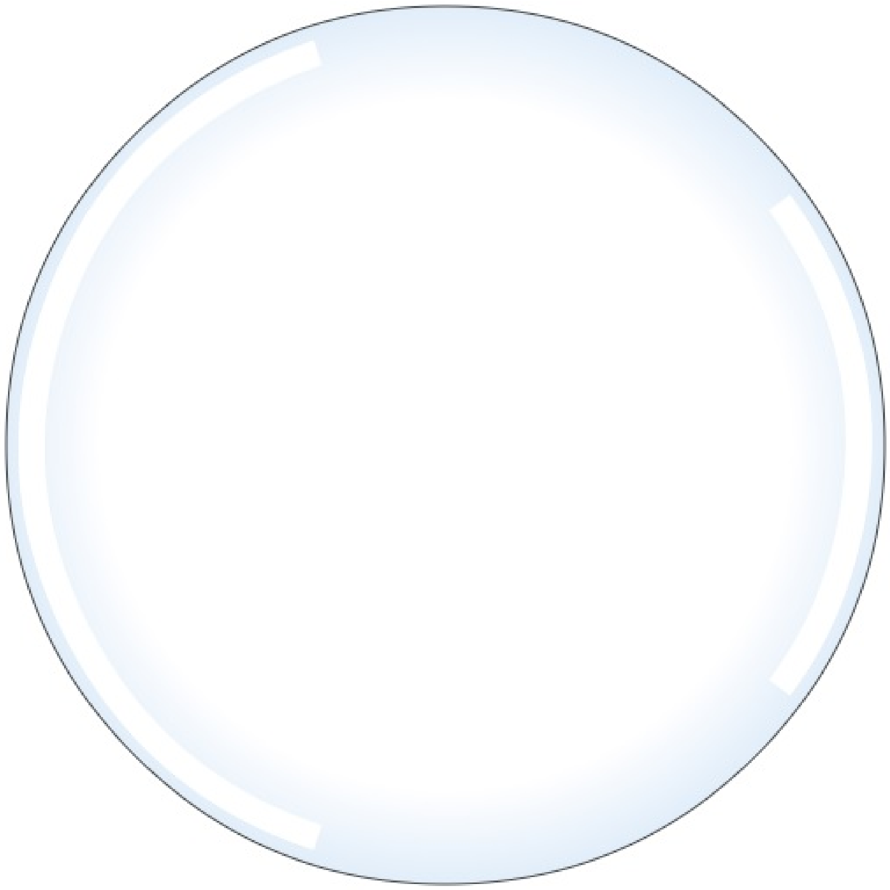
6. Can I see the newsfeed that were filtered out by the system (hidden newsfeed) individually?

   Yes. Individual posts that were hidden on a category shared by a friend can be found by clicking the circles that are outside the filter bubble.

<<This section is only visible for the survey group with Help Text >>

Play with this interactive visualization and try the functionalities that you understood from the above help text. You may also need the above newsfeed homepage together with the visualization to understand the functionalities.

From Friend(s): Select here ▼ On Category(s): Select here ▼ Time Period: Select here ▼



<<This section is only visible for the survey group without Help Text >>

Indicate whether the following are functions of the filter bubble visualization that you explored in the previous page or not. (Yes = Is a function/ No = is NOT a function)

1. View the friends who shared some posts that were shown in your newsfeed homepage over a time period.

- Yes
- No

2. View all the categories of posts that you shared with your friends over a time period

- Yes
- No

3. View all the friends who unfriended with you over a time period

- Yes
- No

4. Drag and drop a category circle from inside the big bubble to outside the big bubble.

- Yes
- No

5. View the total number of posts that you shared with your friends over a time period

- Yes
- No

6. View the categories of posts shared by a friend that were filtered out by the system over a time period

- Yes
- No

7. View the posts shared by an alien from Mars over a time period

- Yes
- No

8. Drag and drop a friend circle from outside the big bubble to inside the big bubble.

- Yes
- No

9. View your categories of common interests with your friend over a time period

- Yes
- No

10. View a specific post (link to that post) from a friend on a category which was filtered out by the system.

- Yes
- No

<<This section is only visible for the survey group with Help Text >>

Indicate whether the following statements about the visualization are True or False (Yes = True/ No = False)

1. Circles inside the big bubble represent the categories of posts that were shared by your friends or friends who shared posts on a category and you saw them in your newsfeed homepage.

- Yes
- No

2. Circles inside the big bubble represent categories of posts that you shared with your friends over a time period

- Yes
- No

3. Circles inside the big bubble represent all the friends who unfriended with you over a time period

- Yes
- No

4. Drag and drop a category circle from inside the big bubble to outside the big bubble means that you tell the system that you don't like to see that category of posts in your newsfeed homepage in future.

- Yes
- No

5. Total number of posts that you shared with your friends over a time period is displayed inside the big bubble

- Yes
- No

6. Circles outside the big bubble represent categories of posts or friends who shared some posts that were filtered out (hidden) by the system over a time period

- Yes
- No

7. Big bubble shows the newsfeed shared by an alien from Mars over a time period

- Yes
- No

8. Drag and drop a friend circle from outside the big bubble to inside the big bubble means that you want to see the posts shared by that friend in your newsfeed homepage in future.

- Yes
- No

9. Categories inside the bubble represent common interests with your friend

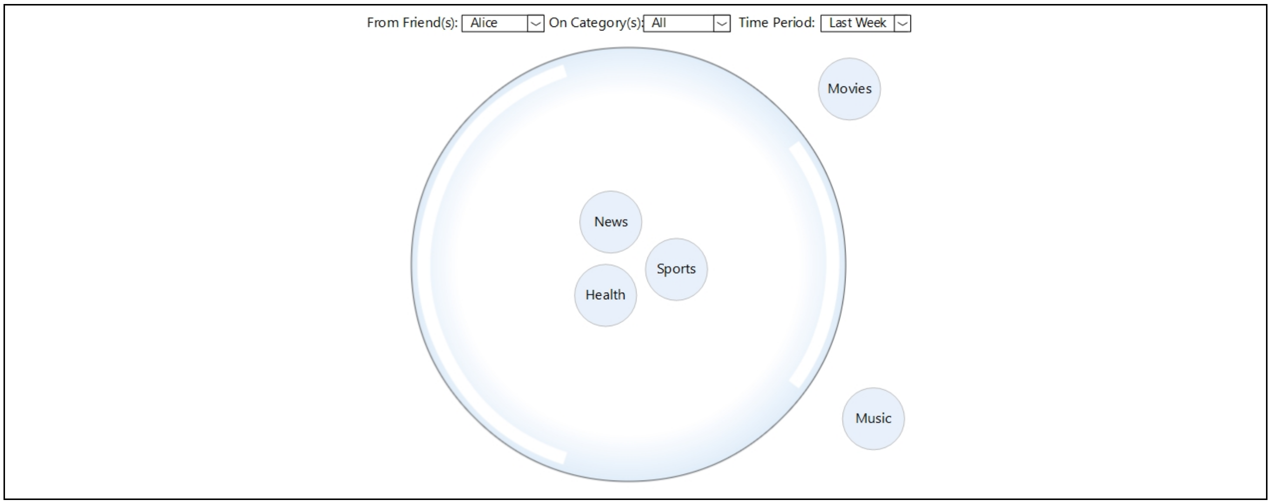
- Yes
- No

10. A bigger circle for friend/category represents a greater number of posts and a smaller circle represents a smaller number of posts.

- Yes
- No

<<This section is visible for both survey groups>>

11. You selected From Friend(s): Alice, On Category(s): All, Time Period: Last Week. Your filter bubble looks like the following. What functionality of the visualization does this action best represent?

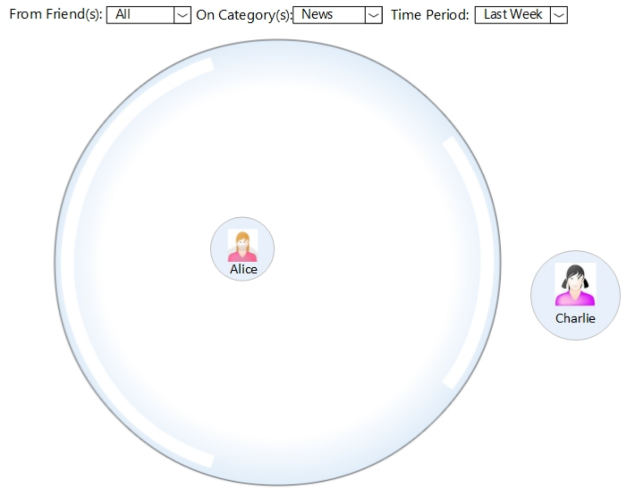


Choose one of the following answers

- View the categories of posts shared by Alice that were filtered out by the system
- View the categories of posts shared by Alice that were shown in your newsfeed homepage
- View the categories of posts shared by Alice that were shown in your newsfeed homepage or filtered out by the system
- None of the above

<<This section is visible for both survey groups>>

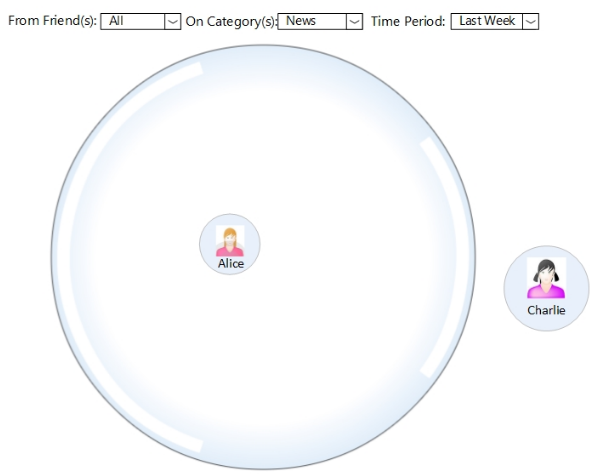
12. You selected From Friend(s): All, On Category(s): News, Time Period: Last Week. Your filter bubble looks like the following. What functionality does this action best represent?



Explain it in your own words.

<<This section is visible for both survey groups>>

13. You selected From Friend(s): All, On Category(s): News, Time Period: Last Week. Your filter bubble looks like the following. What functionality does this action best represent?

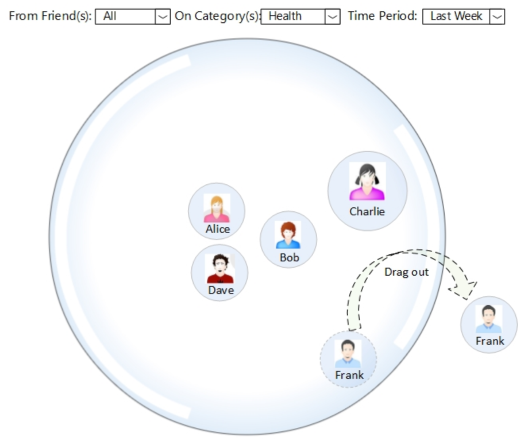


Choose one of the following answers

- View the friends who shared some posts on category News that were shown in your newsfeed homepage
- View the friends who shared some posts on category News that were filtered out by the system
- View the friends who shared some posts on category News that were shown in your newsfeed homepage or filtered out by the system
- None of the above

<<This section is visible for both survey groups>>

14. You selected From Friend(s): All, On Category(s): Health, Time Period: Last Week. Then you Drag and drop Frank from inside the big bubble to outside the big bubble as shown below. What functionality does this action best represent?

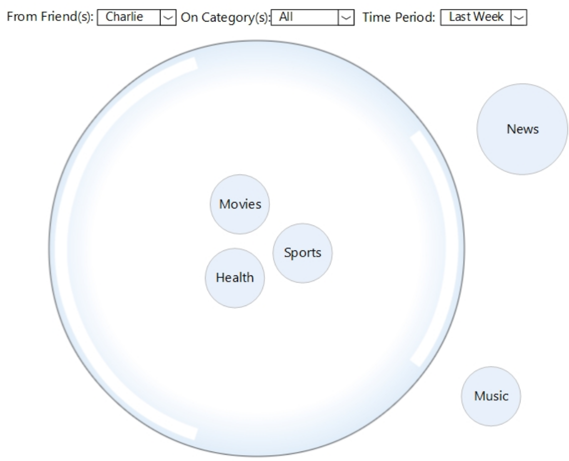


Choose one of the following answers

- To tell the system to NOT to filter out the health related posts from Frank from now on
- To tell the system to remove Frank from my friends list
- To tell the system to filter out (hide) the health related posts from Frank from now on
- To tell the system to add Frank to my friends list on health category

<<This section is visible for both survey groups>>

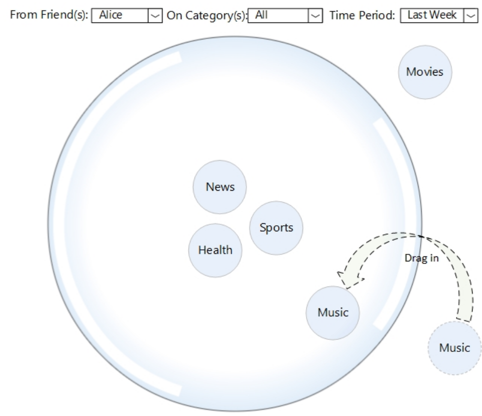
15. You selected From Friend(s): Charlie, Category(s): All, Time Period: Last Week. Your filter bubble looks like the following. What functionality does this action best represent?



Choose one of the following answers

- View the categories of posts that you shared with Charlie on your wall during last week
- View your categories of interest among the categories of posts shared by Charlie during last week
- View the total number of posts that you shared with Charlie during last week
- View Charlie's most favourite interest category by comparing the positions of circles inside the bubble

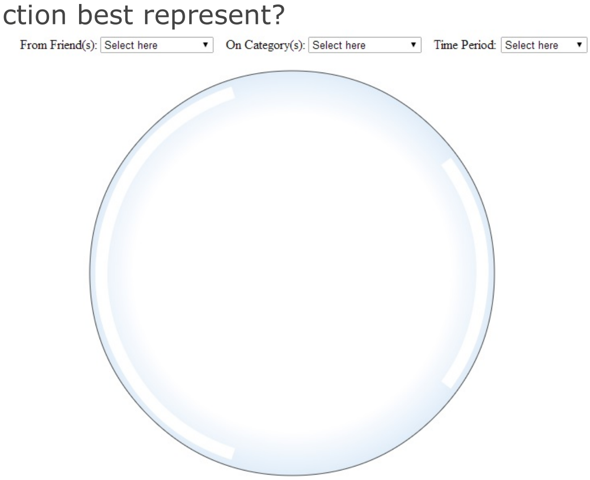
<<This section is visible for both survey groups>>
16. You selected From Friend(s): Alice, On Category(s): All, Time Period: Last Week. Then you Drag and drop Music category circle from outside the big bubble to inside the big bubble as shown below. What functionality does this action best represent?



Choose one of the following answers

- To tell the system NOT to filter out the posts from Alice on Music category from now on
- To tell the system to remove Alice from my friends list
- To tell the system to filter out the posts from Alice on Music category from now on
- To tell the system to add Alice to my friends list

<<This section is visible for both survey groups>>
17. Select From Friend(s): Alice, On Category(s): All, Time Period: Last Week. Click on a category circle outside the bubble. View the individual links on the popup menu. What functionality does this action best represent?

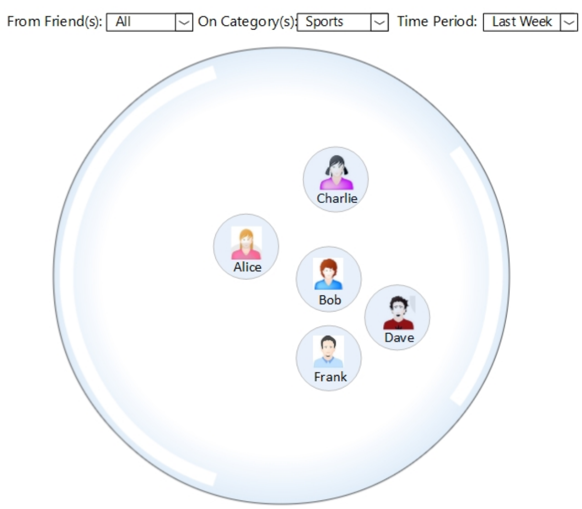


Choose one of the following answers

- To view the individual posts on that category which Alice shared with you that were shown in your newsfeed homepage during last week
- To view the individual posts on that category that you shared with Alice during last week
- To view the individual posts on that category that you shared with Alice and filtered out by the system during last week
- To view the individual posts on that category which Alice shared with you that were filtered out by the system during last week

<<This section is visible for both survey groups>><<This section is visible for both survey groups>>

18. Which of the following best describes your filter bubble shown below?

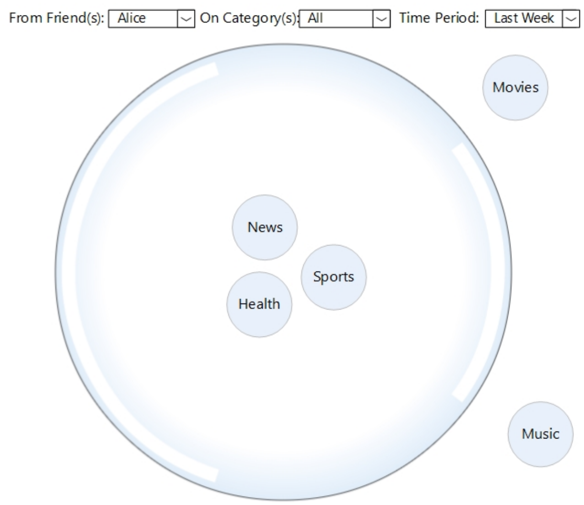


Choose one of the following answers

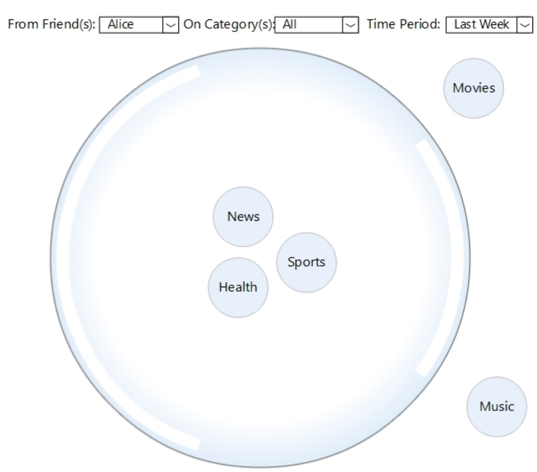
- Circles inside the big bubble represent your friends who shared some posts on your interest of Sports during last week and you saw them in your newsfeed homepage
- Friend circles inside the big bubble represent that you shared some posts on Sports with them during last week
- Circles inside the big bubble represent your friends who shared some posts on your interest of Sports during last week and you ignored them in your newsfeed homepage
- Friend circles that are more closer to the center of the big bubble represent that they have more interest on the Sports category

<<This section is visible for both survey groups>>

19. Explain your filter bubble shown below in your own words.



<<This section is visible for both survey groups>>
20. Which of the following best describes your filter bubble shown below?

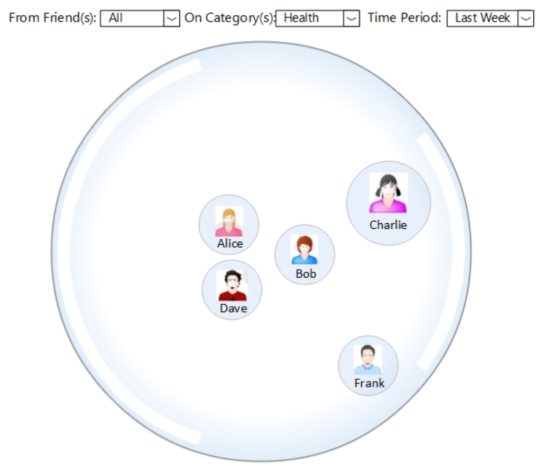


Choose one of the following answers

- Category circles inside the big bubble represent that you shared some posts on them with Alice during last week
- Circles inside the big bubble represent your interest categories of posts shared by Alice during last week and you saw them in your newsfeed homepage
- Circles inside the big bubble represent your interest categories of posts shared by Alice during last week and the system filtered out (hide) them
- Circles outside the big bubble represent that Alice shared less number of posts on that categories during last week

<<This section is visible for both survey groups>>

21. Which of the following best describes your filter bubble shown below?



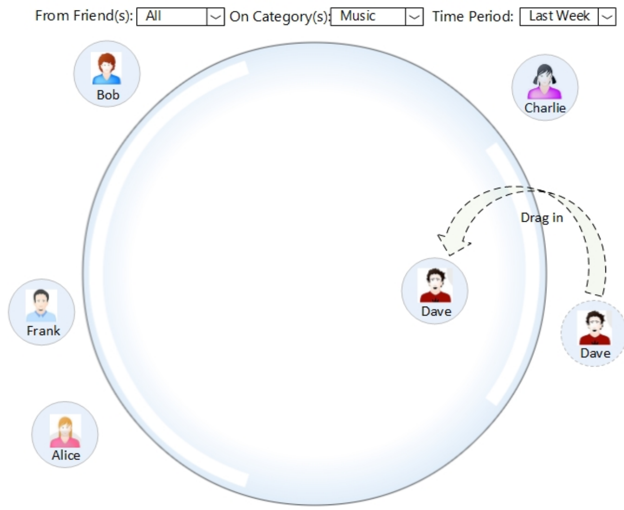
Choose one of the following answers

- Output view represents a soap bubble
- Output view represents categories shared by alien came from Mars
- Output view represents aliens who came from a different galaxy
- Output view has five friend circles inside the big bubble

<<This section is visible for both survey groups>>

---

22. You drag and drop Dave from outside the big bubble to inside the big bubble. Which of the following best describes the action and resulting view of your filter bubble as shown below?

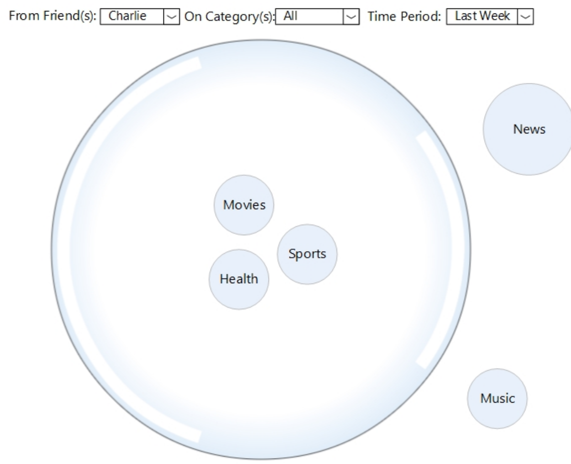


Choose one of the following answers

- ● It represents Dave was not interested in Music and by dragging him inside the bubble you force him to take interest on music
- ● It represents that you got interest in seeing Dave's posts on music and you want to tell the system NOT to filter out Dave's posts on music from now on
- ● It represents you unfriended Dave and you don't want to see his posts on Music anymore
- ● None of the above

<<This section is visible for both survey groups>>

23. Which of the following best describes your filter bubble shown below?

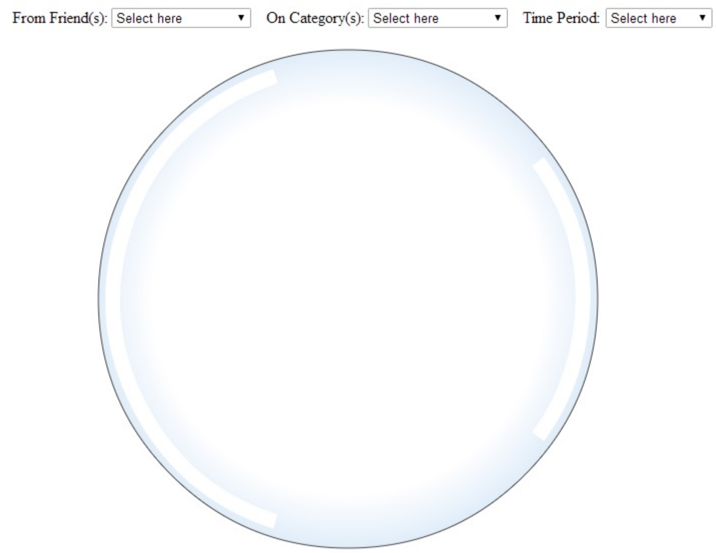


Choose one of the following answers

- ● System filtered out (hide) posts on Music and News shared by Charlie because you did NOT show interest (ignored) on Charlie's posts on Music and News in the past.
- ● System filtered out (hide) posts on Music and News shared by Charlie because Charlie had no interest on Music and News in the past.
- ● Music and News Categories are outside the big bubble because Charlie did NOT share any posts on Music and News categories
- ● Bigger circle for News category represents that Charlie shared very few posts on that category during last week.

<<This section is visible for both survey groups>>

24. Input Select, From Friend(s): Alice, On Category(s): All, Time Period: Last Week. Now you drag and drop Sports circle from inside the big bubble to outside the big bubble. Which of the following best describes the drag action and the resulting output view?



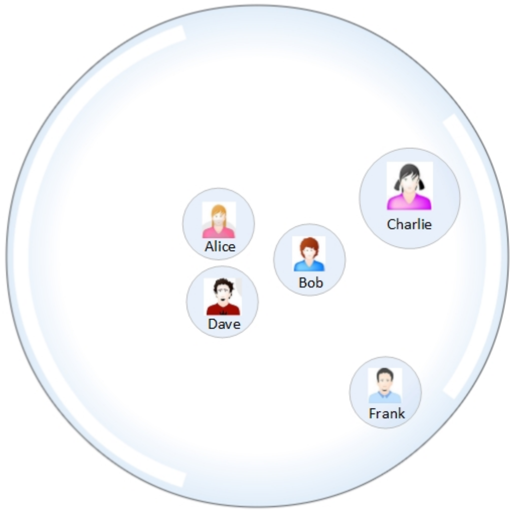
Choose one of the following answers

- Dragging Sports outside the big bubble represents you are forcing Alice to quit interest on Sports and she will not share anything on Sports anymore.
- Dragging out represents that you lost interest on Sports related posts from Alice and you don't want to see any future posts on Sports by Alice in your newsfeed homepage anymore.
- Dragging out represents that you unfriended Alice and she is not your friend anymore.
- None of the above

<<This section is visible for both survey groups>>

25. Which of the following best describes your filter bubble shown below?

From Friend(s): All On Category(s): Health Time Period: Last Week



Choose one of the following answers

- You saw the posts on health shared by All of your friends because your friends are interested in them and they interacted with their posts frequently in the past by liking, commenting and re-sharing them.
- The position of Charlie's circle shows that she is not interested in the category anymore and will stop posting in this category from now on.
- You saw the posts on health shared by All of your friends because you are interested in them and showed your interest with health category of posts in the past by frequently liking, commenting and re-sharing them.
- None of the above

Link to User Responses

Group 1 (Without Help Text):
http://www.amazon.com/gp/drive/share?ie=UTF8&s=JjCtG5VdSMcqJSuM3H0azE
Group 2 (With Help Text):
http://www.amazon.com/gp/drive/share?ie=UTF8&s=Kr0TtctmS0Ikub7bQFYjBM