Three Essays on Dynamics of Online Communities

Armin Ashouri Rad

Dissertation submitted to the faculty of the Virginia Polytechnic Institute and State University in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

in

Industrial and Systems Engineering

Hazhir Rahmandad, Chair

Konstantinos P. Triantis

Navid Ghaffarzadegan

Mohammad Mahdian

March 16, 2016

Blacksburg, VA

Key words: Dynamics of opinion, opinion space, attitude measuring, online communities, online behaviors, users' activity

Three Essays on Dynamics of Online Communities

Armin Ashouri Rad

ABSTRACT

Essay #1: Reconstructing Online Behavior through Effort Minimization

Data from online interactions increasingly informs our understanding of fundamental patterns of human behavior as well as commercial and social enterprises. However, this data is often limited to traces of users' interactions with digital objects (e.g. votes, likes, shares) and does not include potentially relevant data on what people actually observe online. Estimating what users see could therefore enhance understanding and prediction in a variety of problems. We propose a method to reconstruct online behavior based on data available in many practical settings. The method infers a user's most likely browsing trajectory assuming that people minimize effort exertion in online browsing. We apply this method to data from a social news website to distinguish between items not observed by a user and those observed but not liked. This distinction allows us to obtain significant improvements in prediction and inference in comparison with multiple alternatives across a collaborative filtering and a regression validation problem.

Essay #2: Measuring Individual differences: A Big Data Approach

The amount of behavioral and attitudinal data we generate every day has grown significantly. This era of Big Data has enormous potential to help psychologists and social scientists understand human behavior. Online interactions may not always signify a deep illustration of individuals' beliefs, yet large-scale data on individuals interacting with a variety of contents on specific topics can approximate individuals' attitudes toward those topics. We propose a novel automated method to measure individuals' attitudes empirically and implicitly using their digital footprints on social media platforms. The method evaluates content orientation and individuals' attitudes on dimensions (i.e. subjects) to explain individual-content ratings in social media, optimizing a pre-defined cost function. By applying this method to data from a social news website, we observed a significant test-retest correlation and substantial agreement in inter-rater reliability testing.

Essay #3: Social Media and User Activity: An Opinion-Based Study

An increasing fraction of social communications is conducted online, where physical constraints no longer structure interactions. This has significantly widened the circle of people with whom one can interact and has increased exposure to diverse opinions. Yet individuals may act and respond differently when faced with opinions far removed from their own, and in an online community such actions could activate important mechanisms in the system that form the future of the outlet. Studying such mechanisms can help us understand the social behaviors of communities in general and individuals in particular. It can also assist social media outlets with their platform design. We propose models that capture the changes in individuals' activities in social media caused by interacting with a variety of opinions. Estimating the parameters of the models using data available from a social news website (Balatarin) as a case study, we extracted mechanisms affecting the communities on this platform. We studied the effect of these mechanisms on the future formation and the lifecycle of the platform using an agent-based simulation model. Having examined the effect of biased communities on the social media, the results imply that individuals increase their online activity as a result of interacting with contents closely aligned to their own opinion.

Acknowledgements

I would like to express my special appreciation and deepest gratitude to my advisor Dr. Hazhir Rahmandad, you have been a tremendous mentor for me. I would like to thank you for all your support and guidance over the past years. Thank you for the countless hours you have invested in me over these years. Every discussion we had inspired me to go further and see more clearly. I fully enjoyed your organized mind, genius ideas and deep insights. Your life style, personality and professionalism have always amazed me. I could not have imagined having a better advisor for my Ph.D. Thank you Hazhir for constantly teaching me and believing in me. I wish you all the best in your life and career.

I would also thank the other committee members for their support, brilliant comments and suggestions. Dr. Kostas Triantis, thank you for your great feedbacks, you made me think more about my research and its applications in industry. Dr. Mohammad Mahdian, thank you for your supports, you are a role model for me as someone who connects research and industry. Dr. Navid Ghaffarzadegan, thank you for all your support, especially in the last few months of my work, and thank you for trusting me with your research project. I have really enjoyed working with my committee and highly value each of your contributions to this effort.

A special thanks to Balatarin community and organizers, Mehdi Yahyanejad (Balatarin founder), Mani (organizer) and many others, who provided me with data and answered numerous questions. My very sincere thanks to Sara, my best friend who always inspires and encourages me to keep going forward. I also would like to extend my gratitude to Prof. Alinaghi Mashayekhi who supported me in getting this far. I should thank the National Science Foundation, Division of Social and Economic Sciences for providing financial support and assisting me in continuing my education.

Finally, I take this opportunity to express the profound gratitude from my deep heart to my beloved family for their support, encouragement and patience during my pursuit of the Doctorate. My wonderful mother, your love, courage and strength truly inspire me. My brother and sister, your support and love cannot be expressed enough in words; and in memory of my father, whose belief in education has brought me this far.

Contents

List of figures:

List of tables:

1. Chapter 1- Introduction:

1.1 Problem context:

As of now the online social networking application Facebook has more than 1.4 billion monthly active users who spend on average 21 minutes per day on the application[1]. At the same time, every minute, three hundred hours of content were uploaded to Youtube[2]. Instagramers share over 80 million photos and videos each day[3], and about 500 million Tweets are sent every day[4]. Social media platforms have become an integral part of the daily lives of billions of people. They are changing lives of millions of people around the world and transforming people's habits, expectations and lifestyle. Not long ago the era of social media, as we understand it today, started with "Open Diary" (OpenDiary.com founded on 1998, shut down 2014), an early social networking site that brought together online diary writers into one community. Then growth of high-speed internet access further added to the popularity of the concept, leading to the creation of social networking sites such as MySpace (in 2003) and Facebook (in 2004). Nowadays, there are hundreds of social network websites, some support pre-existing (in real life) connections, others help strangers connect based on shared interest, activities or political views. Even with the same technological features, the cultures that emerge around social network websites are varied [1, 2]. Many scholars have examined social networks in order to understand the culture, users' engagement and practices [3-5].

However, very few researches studied the variety of opinion in social media users and the effect of that on online behaviors of individuals and the future opinion formation in social media platforms. Such research needs an extensive amount of data on users' behavior and interactions within the social media and methods to automatically measure individual opinions. Here in the first essay, we propose a method that enables us to collect different types of data on users' activity in social network based on reconstructing the history of a website. Then we propose a novel method for quantifying users' opinions and attitudes in social networks and study the variety of opinions that exist in a social news website. Finally, using a simulation model we study the future opinion formation of an online community and evaluate the effect of biased communities on the life cycle of social media platforms.

1.2 Research contributions:

1.2.1 Essay #1

Studying the dynamics of online communities and opinion formation of individuals in such environments needs historical data on users' activity. However, data gathered from social networks and other online sources usually lacks information on the context in which activity has happened. For instance, social networks usually collect data on user-object interactions (i.e. liking a post on Facebook) but rarely collect data on the other content on the page (e.g. friends' profile page, pages of groups and homepage in Facebook) that are shown side by side and/or competing with the object of interest. Besides, very few social networks keep data on user-object exposure (i.e. who has seen what object, even if they had no interaction with that object), therefore we usually do not know what people have actually viewed

[1] http://www.businessinsider.com/how-much-time-people-spend-on-facebook-per-day-2015-7
[2] http://expandedramblings.com/index.php/youtube-statistics/
[3] https://www.instagram.com/press/?hl=en
[4] http://www.internetlivestats.com/twitter-statistics/

online, therefore we can't tell if a user didn't like an object, or simply did not see it, should they have no interaction with the object. These create a major concern when attempting to estimate users' opinions. In the first essay we overcome some of these shortcomings in data by taking advantage of available data on user-object interactions, algorithmic information on the underlying filtering and sorting designs of social networks, and the idea that users minimize their browsing effort.

1.2.2 Essay #2

We express ourselves in interactions with socially generated digital contents in forms such as "like", "share", "vote", "retweet", and "pin". These interactions form valuable data for understanding individual personalities, behaviors and their attitudes and opinions toward different issues. Online interactions may not always signify a deep illustration of individuals' beliefs, but large scale data regarding interactions with variety of contents on specific topics can provide a good estimate of individuals' attitudes toward those topics. In general, attitudes can be evaluated explicitly, but on sensitive issues fear of judgment and tendency to appear well adjusted, among other factors, could bias the responses. Therefore, to measure attitude of individuals toward such issues researchers developed methods that measure the attitudes indirectly or implicitly such as implicit association test and word fragment test. Despite all the advantages in implicit methods, current implicit attitude measuring methods also have many drawbacks. Implicit measures could be affected by psychological factors such as self-perception and self-observation. Environmental factors such as context and prior exposure can also lead to biased measurement of individuals' attitude. Therefore these measures show limited stability across multiple measurements. Finally, existing explicit and implicit methods for opinion measurement are expensive and not easily scalable to measuring individuals' opinions in large online social networks and over time. In the second essay we propose a novel automated method to estimate individuals' attitude empirically and implicitly using available user-object interaction data in social media. Proposed method resolves many drawbacks in the current implicit methods, is scalable, and reliable over time.

1.2.3 Essay #3

Social media as a category of Internet-based applications that allow the creation and exchange of user generated content [2], started a new era for the internet. User generated contents are the fundamental blocks of these platforms and users activity determines the success, lifecycle and impact of a social media. However, users generate contents representing their opinions and points of view, consequently content generated by a user could seem biased from the perspective of others and that perception of bias could affect their actions and activity in the social media. Such actions could activate important mechanisms in the system that forms the future of the outlet. Studying such effects and mechanisms could help us understand the social behaviors of communities in general and individuals in particular. It could also assist social media outlets in platform design. In this essay we use opinion data extracted using the method proposed in the second essay and define regression models that estimate changes in individuals' activities in social media, caused by interacting with variety of opinions. We mathematically evaluate the models for a social news website (Balatarin) and extract the mechanisms that affect the communities in the website. We then simulate the platform and study the effect of communities' biasness on the future opinion formation and lifecycle of the platform.

References:

1. Ellison, N.B., *Social network sites: Definition, history, and scholarship.* Journal of Computer-Mediated Communication, 2007. **13**(1): p. 210-230.
2. Kaplan, A.M. and M. Haenlein, *Users of the world, unite! The challenges and opportunities of Social Media.* Business horizons, 2010. **53**(1): p. 59-68.
3. Castells, M., *The culture of real virtuality: The integration of electronic communication, the end of the mass audience, and the rise of interactive networks.* The Rise of the Network Society: With a New Preface, Volume I, Second edition With a new preface, 2010: p. 355-406.
4. Kumar, R., J. Novak, and A. Tomkins, *Structure and evolution of online social networks*, in *Link mining: models, algorithms, and applications*. 2010, Springer. p. 337-357.
5. Benevenuto, F., et al. *Characterizing user behavior in online social networks*. in *Proceedings of the 9th ACM SIGCOMM conference on Internet measurement conference*. 2009. ACM.

2. Chapter 2- Reconstructing Online Behavior through Effort Minimization

2.1 Abstract

Data from online interactions increasingly informs our understanding of fundamental patterns of human behavior as well as commercial and social enterprises. However, this data is often limited to traces of users' interactions with digital objects (e.g. votes, likes, shares) and does not include potentially relevant data on what people actually observe online. Estimating what users see could therefore enhance understanding and prediction in a variety of problems. We propose a method to reconstruct online behavior based on data available in many practical settings. The method infers a user's most likely browsing trajectory assuming that people minimize effort exertion in online browsing. We apply this method to data from a social news website to distinguish between items not observed by a user and those observed but not liked. This distinction allows us to obtain significant improvements in prediction and inference in comparison with multiple alternatives across a collaborative filtering and a regression validation problem.

Keywords: Effort minimization, Social media, Online behavior, Collaborative filtering

2.2 Introduction

Increasingly, online social networks are integral to our lives and how we interact with others. People express themselves in their interactions with socially generated digital objects (i.e. product listings, postings, comments, stories, songs, tweets) in different forms such as "click", "buy", "like", "vote", "retweet", "share", and "digg". These digital interactions form a valuable source of data for understanding human behavior. User-object interaction data over time could help predict user preferences for new items, facilitate demand forecasting and product placement, allow for better customization of user interfaces and menus, and enable the study of social influence and network dynamics, among others [1-9].

However, data gathered from social networks and other online sources often suffer from a drawback. We usually do not know what people have actually viewed online[5]; And with the increasing trends of customization and filtering, the menu from which people choose could be as important in shaping their choices as their tastes [10]. For example on a social news website the fact that a story did not get a positive vote from a user could imply that the user did not like that story or that he/she simply did not see it. This problem is most acute where expression of a choice can only be seen as a positive signal (i.e. binary choices such as "like" button in Facebook) compared to multi-category ratings that embed more information in a data point, such as Netflix movie ratings [11]. The problem is exacerbated within websites that show the same item on multiple pages (i.e. similar items on Facebook homepage and friends' wall), or different items on the same page for different users, because the number of choice sets that could have led to an interaction are large.

The inference implications of this challenge are potentially significant: most object-user interaction matrices are very sparse, i.e. only a small fraction of cells are non-zero [12]. Thus there are many

[5] Software firms who own a specific platform may collect and log some of this data based on tracking the specific locations where users interact with objects. However, this data is often proprietary and also does not include all relevant contextual variables (e.g. characteristics of the neighboring objects at the time of interaction may be missing).

additional data-points, and much information to be acquired about an individual's taste, in treating zero cells as objects seen but not reacted to, vs. objects not seen. Imagine a user who has viewed 1000 items, voting positively for 10 of those items, vs. one who has viewed 20 and has voted for the same 10 items. In the absence of data on which items have actually been viewed, the user-object interaction matrices will be identical for both users, but the users may have very different propensities to vote in general, and tastes for particular items. The challenge also extends to inference about the inherent quality of an object. Early positive votes increase the propensity of other users to vote for a story [6, 13], pushing it to the more popular categories on a website, and leading to even more exposure and votes. The resulting dynamics can promote specific objects (e.g. songs on a music sharing website) arbitrarily and decouple common measures of quality (e.g. views, positive votes) from the underlying quality [14].

A few methods have been developed to address this challenge using user-object data directly. These methods on the numbers of interactions for users and objects in the user-object matrix to estimate the likelihood that lack of expression signifies lack of interest or lack of viewing (impression). Total number of friends who voted for a story [15], number of retweeter's followers [16], average user activity, and story popularity [17] are some of the proposed proxies to evaluate the exposure of an item to those users that didn't express their choice in interaction with that object. However, these methods implicitly assume the likelihood of an impression directly scales with the total number of interactions for each individual and/or object. This assumption confounds the number of impressions with the effects of taste similarity between user and object, object's inherent quality, user's overall activity propensity, and social influence. All these mechanisms may increase the number of interactions, yet to enable theoretical understanding as well as better prediction and policy design we may need to tease out their independent effects. Impressions, on the other hand, often result from the user interface design and customization algorithms combined with user's likelihood to go online. As a result the existing methods may be biased when used to make inference about the objects' quality, user-object taste similarity, user activity, or social influence, among others [18].

Another set of methods, largely developed in recommendation systems literature, use other sources of data, besides the user-object matrix, to control for users' interest or activity patterns [19]. These methods do not attempt to resolve the problem of unknown impressions, rather, by collecting and including other relevant variables enhance the predictive power of the algorithm. For example, Jawaheer and colleagues use play-count (the number of times a track is played) to improve an online music recommendation system [20]; Purchase time is used to improve a recommendation system for an online wallpaper store [21]; and Kim and colleagues incorporate social network relationships data to enhance a digital store recommender system [22]. Other examples include users' behavioral patterns in purchasing items (in a product recommendation system [23, 24]), watching habits (in a TV show recommender [25]), and dwell time (in a joke recommendation system [26]). While these methods do not directly address the inability to distinguish negative evaluation from lack of impression, when such additional data sources area available it is beneficial to include those sources in predictive models.

In this paper, we develop a method for inferring, through history reconstruction and effort minimization, the likelihood of individuals observing objects they have not responded to, based only on data about what they have responded to (e.g. voting data over time). Additionally our method provides information on likely location of the item on the page and other contextual variables at the time of impression. We first provide an overview of the method and the empirical context in which we test it. We then discuss the results of this test. Finally, we show the benefits of this method in two validation applications. We show that by including the resulting estimates for impressions we can significantly improve the predictive power of a simple collaborative filtering algorithm and a regression model.

2.3 Method and Result

Our method explicitly estimates the impressions and separates those from the number of interactions for a user or object. The core idea is to build the history of the online platform and estimate the users' browsing trajectory based on the idea that users minimize their browsing effort. Reconstructing user activity using web log files [27, 28], clickstream data [29, 30], mouse tracking [31] and eye tracking [32] are common in studying visitors' behavior and learning their interests [29, 33]. However, this type of reconstruction rests on having access to complete log/clickstream data or working with data from static websites where each interaction could only have happened in a single location. The majority of current social news data are generated in dynamic websites where the user could have interacted with an object on multiple locations, and the fact that an interaction has happened does not let us know where it has happened, which is needed for fully reconstructing the user activity.

At the heart of our method is approximating the most likely user behavior based only on the interaction trace that is publicly available. This approximation is rooted in the observation that people conserve cognitive effort in general, and in their browsing activity in particular [34]. We define alternative pathways users could have taken given their observed expression (i.e. votes for similar items which could have been viewed on different pages), and find the most likely pathway, i.e. the one that minimizes user's effort. Using these estimates we can identify the likelihood and location that each object has been observed by each user, thus addressing the original challenge. Our method consists of three key steps. First, we identify the potential content that individuals could have been exposed to, i.e. recreate the content history of social network. Next, we estimate the content most likely the individual has observed based on their actual behavior pattern and effort minimization principle. Finally, we estimate the likelihood of observing each object based on the results in the previous step and the individual's activity pattern. We next introduce our empirical setting and discuss the steps in the context of application to this data.

2.3.1 Empirical Setting and Data

We use data from Balatarin, the largest social news website (examples of social news sites include Reddit and Diggs) for the Persian-speaking community. On Balatarin users can post links to different news items, websites, blog posts, or multimedia content (i.e. videos, pictures, sounds). We call these links stories. Users are also able to read other user's stories and vote or comment on them. Since its inception in August 2006 Balatarin has gathered over 56,500 registered members, two and a half million stories, sixty five million votes and several million comments. Balatarin sorts and ranks news based on popularity of the story (current number of votes), time of publish, and time of promotion in case of stories promoted to the "hot stories" page. This system is similar to Reddit and also resembles popularity and time-based sort and filtering options common to a wide range of applications, from Netflix and Audible to Amazon and Yelp.

At any point in time a story could be found in a few different places on Balatarin, and therefore the fact that a user has voted for a story does not identify the location where the interaction has happened. Specifically, Balatarin has a promotion system that promotes popular stories with votes more than a specified threshold to its "first" page. It also gives visitors the option of reading first page stories or the recently published (i.e. not-promoted) stories. In the default first page, stories are sorted by the time of promotion (i.e. the most recent promoted story placed at the top of the first page), but users also have the choice of sorting promoted stories of last day, week or month by the descending number of votes. Balatarin sorts not-promoted stories based on their publish time in recently published stories page; however, users are also able to sort them by descending number of votes. We call these options of sorting stories (chronologically by publish (or promotion) time or descending number of votes) ordering pages. Each of these ordering pages are further broken down into multiple subpages each accommodating 25 stories. Finally, a user can view all stories in an ordering page or focus on a specific category (political, economics, sports, social, etc.). In summary, a story cannot be seen in both first page and recently published page at the same time, but in each of these pages it could be seen in more than one ordering page. Promoted stories are active for five days in the first page ranking, and not-promoted stories stay one day in recently published stories page. Figure 2-1 provides a screenshot of Balatarin's user interface and shows where users can interact with and customize the content.

Figure 2-1. Screenshot of Balatarin

We have access to publicly available Balatarin data on stories (including Story ID, posting User's ID, Time of Posting, user identified Story Category) and votes (Story ID, Time of Vote, voting User's ID). This is the type of social network data publicly available in many settings. We do not have access to data on where (which ordering page, and subpage) votes have been casted or any other information on what stories have been viewed by the users. In fact the Balatarin Inc. does not collect that data either.

2.3.2 History Reconstruction

The first step in our algorithm is history reconstruction. Most online social networks use algorithms of ranking and filtering to present the content they deem most appealing to their users and prevent information overload. Personalized filtering algorithms filter contents for each user individually based on available data on user's interest. For example Facebook, twitter and Pinterest feed users with posts from their friends or people, pages and groups they follow. Many other social networks use filtering (and ranking) algorithms that provide the same choices of content to every user. For example Reddit does this by ranking the posts based on popularity, or time of publish, and filters out unpopular or old stories. Recreating these individualized or site-specific histories is the first step of our method.

Two types of data are needed to recreate this history. First, qualitative/algorithmic information on the underlying filtering (or ranking) algorithm of a social network is required, and often available from an application's publicly available information, or could be estimated by reverse engineering the logic from direct observations. In the case of Balatarin we explained these algorithmic rules briefly above, and a more detailed algorithmic representation is provided in the appendix (Figure 2-4). The second source of data is the user-object interactions over time. User-post liking in Facebook (i.e. who liked which post and when) is an example of such user-object interactions data. This data is needed to reconstruct the history of dynamic websites because usually users' activities on the network inform the sorting and filtering algorithms at work in populating different locations (e.g. stories with more votes may be put on the top of a social news website). Time stamps identifying the specific time at which an interaction has occurred are also needed for history reconstruction because rebuilding history includes a time dimension to tracks when an object (Story in Balatarin's case) is liked or promoted as well as where it could show up as a result of its state (e.g. current number of votes) at any point in time. These two data sources (algorithmic information and user-object interaction history) typically suffice to recreate the history of a social network, though if page generation uses random functions (e.g. in selecting a subset of stories to show) the reliability of the recreated history would go down.

We inferred underlying algorithm of Balatarin by investigating the published rules of Balatarin and confirmed them with observation of the site, and used them to recreate Balatarin's history. Specifically, using the data on posting and voting times we calculate the status (i.e. the number of votes, promotion status, location on each ordering page, etc.) of stories over time and rebuild the history of Balatarin in a simulation environment. In this environment when a story is published (an event that can be read from our dataset), it is sent to the top of the recently published stories page (sorted by publish time) and the last place in recently published stories sorted by (descending) number of votes ordering page. The simulation rebuilds the state transitions for stories (e.g. promotion of a story) using actual vote data, and updates the ordering pages and subpages accordingly. For example if a story gets enough votes to be promoted to the first page, it will be removed from the not-promoted orderings pages and placed at the top of the first page immediately, pushing all the other stories in that ordering page one position down (and thus resetting the content in different subpages). Similarly, the ordering pages for the promoted stories of last day, week and month sorted by descending number of votes get updated each time a vote is casted to a promoted story. We remove stories from the ordering pages based on their respective lifecycle. Thus, using the history reconstruction simulation we can find whether a story has been promoted at any time and what position it would have occupied on feasible ordering pages and subpages.

2.3.3 Estimating Browsing Trajectories

The simulation environment provides us with the feasible locations for an object at the time a user interacts with that object in our database (e.g. the ordering pages and subpages that contain a story at the time the story gets a vote from a user). In the second step of our algorithm we start from these feasible locations and find the interaction pattern for each user that minimizes her browsing effort.

Effort can be measured based on the amount of scrolling and the number of clicks that a user should undertake in each potential browsing trajectory. In case of Balatarin, we find the ordering page that minimizes user's effort in reaching each of his/her consecutive voted stories, given the previous ordering page the user has likely visited. This dynamic program can be solved for each user to find the ordering page (and subpage) in which the user has likely voted for each of the stories.

Given the navigation options available to a Balatarin user, we specify three terms in the cost function that represents user efforts. First, a change in the ordering page requires the user to move the mouse to the top menu of the site and choose a different ordering page. This variable is captured as $D_{Ordering}(i, i+1)$ which is one if a change in ordering page is required from story i to story $i+1$. The effort needed for such move (the penalty of this move in the cost function) is represented in *p3.* A second term measures the changes in subpages, $D_{Subpage}(i, i+1)$, which requires a click from the user, and its cost (*p2*). The distance between two subsequent votes on an ordering page, $D_{Story}(i, i+1)$, counts the stories between the two and captures the cost of scrolling up or down. We minimize the sum of these costs, i.e. total effort by each user (u) by selecting the ordering pages in which the user could vote for each story over all the stories she has voted for (N_u). The set of feasible ordering pages at the time of each interaction is taken from the results of history reconstruction step.

$$Cost(u) = \sum_{i=1}^{N_u} p1 * D_{Story}(i, i+1) + p2 * D_{Subpage}(i, i+1) + p3 * D_{Ordering}(i, i+1) \quad (1)$$

After exploring the effect of different penalty parameters on the overall performance of the algorithm we set them at one unit for each story between two consecutively voted stories ($p1$), five units for changing the subpage ($p2$), and 200 units for changing ordering page ($p3$). The results are rather insensitive to moderate changes to these cost parameters. We solve the optimization problem for each user separately using a greedy search method.

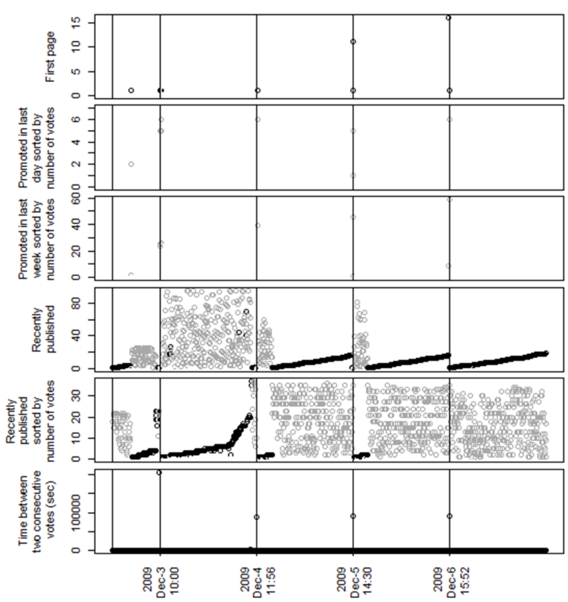


Figure 2-2-a. Sub-pages

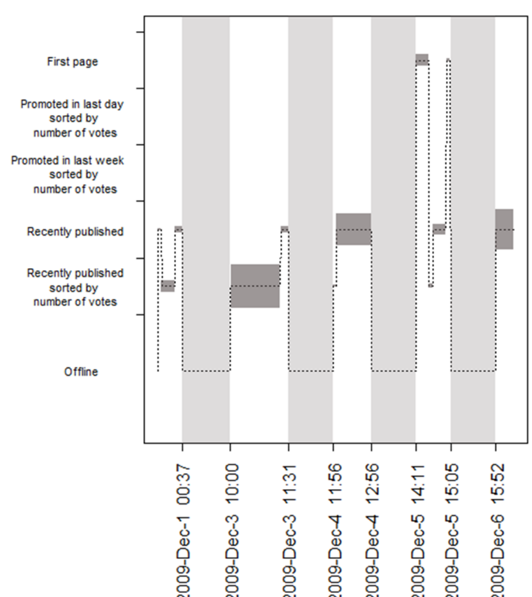


Figure 2-2-b. Optimal browsing trajectory

Figure 2-2. Feasible sub-pages (a) and inferred browsing history (b) for a sample user.

Figure 2-2 represents feasible ordering pages and subpages (panel a) and inferred browsing history (panel b) for a sample user. In panel a, we observe the feasible vote locations on five possible ordering pages and their multiple subpages (Y axes) for each story the user has voted on. Time between votes is also measured (lower panel) to distinguish different browsing sessions. Each dot identifies one feasible ordering page (and subpages) a voted story could have occupied at the time of a vote. The color codes demonstrate the result of the optimization, where black dots show the ordering page/subpage the votes are mostly likely cast. For this user the most likely browsing trajectory is summarized in panel b. It appears that the user has largely spent her/his time on the recently published pages, alternating between the chronological and popularity sorting across different browsing sessions. S/he has also visited the first page in a few instances. In this panel the area of each rectangle represents the time the user has spent on that ordering page with logarithmically scaled inter-session intervals (due to the longer offline intervals).

Table 2-1 provides an overview of the percentage of votes casted on different ordering pages for all the users, found through our algorithm. Most votes are cast on the recently posted page with chronological (default) sorting (32%), followed by first page (Promoted, chronologically sorted; 21%) and recently published stories sorted by number of votes (14%). The algorithm is unable to make a precise categorization for 9% of the stories, e.g. when a user casts a single vote in a session and thus the cost for movement between different votes is not defined.

Ordering page	**Vote Percentage**
First page (Promoted, Promotion time)	21%
Promoted, Most voted last day	6%
Promoted, Most voted last week	2%
Promoted, Most voted last month	<1%
Recently published, Posting time	32%
Recently published, Votes	14%
Other pages	15%
Not categorized by the algorithm	9%

Table 2-1. Percentage of votes on different ordering pages

2.3.4 Estimating Impressions and Locations

Once the users' browsing histories are recreated through effort minimization, we can estimate, among other things, the likelihood that an object is seen by a user but not interacted with, e.g. stories seen but not voted for in Balatarin. We arrive at these estimates based on assumptions on how likely it is for a story in the neighborhood of a voted story to be seen. For example, one may assume that when a user votes for a story, other stories in that subpage are also seen by the user. Or we can use a simple function (e.g. exponential decay with half-life α) to calculate the probability of exposure based on the distance (i.e. links) between voted and not voted stories in the same subpage. Those assumptions can then be calibrated based on the predictive power of the model in specific applications, e.g. by changing α to maximize the ability of the model to predict the likelihood of the next vote given the previous known votes.

2.4 Validation and Inference

In order to assess the effectiveness of our method we develop two prediction and inference problems and compare our method's performance against available alternatives. First, we develop a collaborative filtering model that predicts pairs of user-story votes in Balatarin using a weighted matrix factorization method [17]. We compare the predictive power of this collaborative filtering model by comparing weights that represent impression likelihoods (extracted from our behavior reconstruction method) and the available alternatives from the literature. In a second set of comparisons we compare our method to alternatives in regressions that estimate the impact of different story characteristics on its likelihood of getting a vote.

2.4.1 Collaborative Filtering Test

Recommendation systems provide personalized recommendations for products or services using various knowledge discovery techniques and are widely popular across various online platforms [35]. Collaborative filtering is one of those techniques that has proved effective in diverse recommendation systems [35-38] and is defined as "a method of making automatic predictions about the interests of a user by collecting taste information from many users" [39]. Using a common data structure, in which users and objects (e.g. products, movies, stories) are separately identified, collaborative filtering predicts interest of users in objects based on taste information obtained from users' rating on other items [40, 41]. One of the most established methods for collaborative filtering factorizes the user-object interaction matrix ($R_{n\times m}$) to users' and objects' taste matrices ($U_{n\times d}$ and $V_{m\times d}$) by minimizing a predefined cost function ($C(U,V)$; See equation 2)[42-44]. Existing matrix factorization methods can be extended to the case where ratings are zero-one based (e.g. "like" button in Facebook and "votes" in Balatarin; thus $R_{i,j}$ elements are binary) and observations are sparse.

However, with zero-one matrices distinguishing negative examples from missing values is both critical for improved prediction and a hard task [17]. The existing solution in the literature is to weigh the error terms (matrix W in equation 2) in collaborative filtering cost function to decrease the effect of missing data on estimation [17]. Two alternative weighting methods have previously been proposed, both using explicit user-object rating data to extract implicit weightings. User oriented weighting assumes if a user has more votes, s/he is more likely to have seen an items s/he didn't vote for ($W_{ij} = \frac{\sum_k R_{ik}}{\max(\sum_k R_{.k})}$) [17]. Item oriented weighting assumes if an item has fewer votes, the zero cells for this item are more likely missing items ($W_{ij} = 1 - \frac{\sum_k R_{kj}}{\max(\sum_k R_{k.})}$) [17]. We will compare the effectiveness of these two weighting methods with the ones coming from our method (discussed below). In all cases voted items are seen by definition, so we set $W_{ij} = 1$ where $R_{ij} = 1$.

$$C(U,V) = \sum_{i=1}^{n}\sum_{j=1}^{m} W_{ij}(R_{ij} - U_{i.}V_{j.}^{T})^2 + \lambda\big(\|U\|_F + \|V\|_F\big) \qquad (2)$$

To avoid over-fitting we include a regularization cost item using Frobenius norm ($\|.\|_F$) of user and story taste vectors with the regularization weight of λ. Minimizing cost function of this convex optimization problem using nonlinear conjugate gradient algorithm [45] we can estimate values of users' and stories' tastes. These in turn can inform predictions for how much a user may like a story s/he has not seen.

We use two alternative weighting schemes that build on our behavior reconstruction method. First, we assume if a user votes for a Balatarin story in a sub-page, all stories in that subpage are also seen by the user (W_{ij} = 1). Stories placed in other sub-pages are given a small weight (0.05) to account for the possibility that the user has browsed those subpages but has not voted in them yet[6]. The second scheme uses a negative exponential function to quantify the weight for each story based on the distance, in number of intervening stories, from each set of two consecutively voted stories ($d_{k,j}$ and $d_{k,j+1}$) on the subpage (Equation 3). If no other stories are voted on the subpage, distance until the beginning/end of previous/next subpage is used, and for subpages with several voted stories the cumulative weight based on all the stories, capped at one, is used. This formulation reflects the increased likelihood of other stories being viewed when multiple are voted in a page.

$$W_{ij} = Min\left(1, \sum_{k=1}^{N-1} \frac{d_{k,j} e^{-\theta d_{k+1,j}{}^2} + d_{k+1,j} e^{-\theta d_{k,j}{}^2}}{d_{k,j} + d_{k+1,j}}\right), \quad where\ N = \begin{matrix} Number\ of\ votes\ user\ i \\ casted\ in\ subpage \\ containing\ story\ j \end{matrix} \qquad (3)$$

The parameter θ defines the slope of decrease in weight by getting further from previous/next voted story and we set it at θ = 0.01 without calibration (to avoid giving unfair advantage to our method). Note that different parameters could be used for upward vs. downward weight change, and these parameters can be calibrated for specific applications.

With the four alternative weighting schemes (user based, story based, and the two building on our method), we set out to test the predictive power of the alternative collaborative filtering algorithms. Specifically, we develop sequential training and testing datasets from Balatarin, estimate the collaborative filtering algorithms using the training set and compare the predictive performance of the four methods using the test set. Balatarin data from October to December 2009 is chunked into four hour intervals. For each test interval the training data includes items from beginning of time until that test interval. After test is conducted on that interval, we push forward the test interval by four hours and re-estimate the four models to conduct another test. A four dimensional taste vector (*d*=4) is estimated for each user (U) and story (V), along with a fixed effect parameter for each user and each story that captures the inherent attractiveness of story and the intrinsic tendency of the user to vote. For every story (j) that appears on the sub-page after a story that is voted by a user (i) we calculate $U_{i.}V_{j.}^T$, predict a vote if this vector is above 0.5 and a no-vote otherwise. This test design is conservative on two fronts, providing a strong test of our method. First, by only considering in prediction comparison the stories that follow one that has been voted for, the value of our method for predicting votes for stories further away (and thus less likely to be seen by the user) is not observed. Moreover, to run a fair comparison we forego the use of additional features only available through our method (such as the number of votes a story has at the time of an impression, or its location on the page). Therefore our method's value may well be higher for the more realistic prediction problems.

Table 2-2 reports the results from this binary classification test, including prediction summary (true positive (TP), true negatives (TN), false positives (FP), and false negatives (FN)) and alternative

---

[6] Combining with user and story oriented methods, more complex weighting schemes could scale this fixed probability for impression in other pages based on user or story activity; but to assess our method alone, without combination with other approaches, we use the (potentially inferior) fixed probability.

performance metrics. Precision reports the fraction of correctly predicted votes ($TP/(TP+FP)$), accuracy is the fraction of total votes correctly predicted ($(TP+TN)/(TP+TN+FP+FN)$), and recall (also called sensitivity) is the probability of detecting a vote ($TP/(TP+FN)$) [46]. F_β ($= (1+\beta^2).(precision.recall/(\beta^2.precision+recall)$) offers an aggregate measure of accuracy that combines recall and precision [47]. The predictions that use our behavior reconstruction method are closer to the User-oriented method but dominate it across the board because our method has both larger numbers of true positives and true negatives. Story oriented method grossly under-estimates the likelihood of voting, thus getting more negatives right, at the expense of missing the large majority of votes. If one's goal is to correctly predict votes (rather than non-votes) then our method provides a notable improvement over the alternatives. Given that the matrix factorization methods for collaborative filtering are among the best performing alternatives with potentially limited room for improvement [42, 48], observing improvements of multiple percentage points, in a conservative test, provides additional evidence about the value of behavior reconstruction and incorporation of estimated impression patterns.

In fact, practical applications would benefit from including additional features only available from our method (i.e. location of story on the page and the number of votes it has at the time of each impression). If we include those factors, adjust the θ to be asymmetric ($\theta_1 = 0.01, \theta_2 = 0.02$), and use smaller time windows (200 seconds instead of 4 hours; to prevent considerable change in the values of dynamic features such as location and number of votes), our predictions in the same task improve significantly (Recall of 0.93, F_1 of 0.62, and Accuracy of 0.88, see Table 2-4 in the appendix).

Method	F_1	F_{10}	Recall	Precision	Accuracy	True Positives (TP)	True Negatives (TN)	False Positives (FP)	False Negatives (FN)
User oriented	0.32	0.26	0.26	0.42	0.69	28,404	245,791	38,637	82,362
Story oriented	0.11	0.06	0.06	0.70	0.73	6,541	281,618	2,810	104,225
Behavior reconstruction, all stories in voted pages exposed to user	0.37	0.28	0.28	0.57	0.74	31,234	259,571	24,857	79,532
Behavior reconstruction, exposure weighting formula (3)	0.41	0.33	0.33	0.54	0.73	36,741	253,145	31,283	74,025

Table 2-2. Comparison metrics

We also compare the alternative learning rates, i.e. how the predictions improve by additional data. Figure 2-3 shows the F_1 measure for the four alterantive weighting methods as a function of the minimum number of votes from a user used in estimating her taste. Not only our method starts from a higher base performance, but also it shows faster learning, i.e. its predictions improves faster with additional data, offering significant advantage when more data is available.

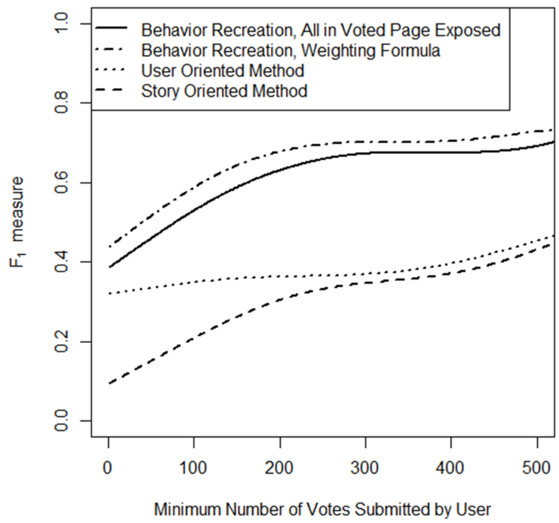


Figure 2-3. Improvements in F_1 measure with additional data.

2.4.2 Regression Test

To assess the value of behavior reconstruction for inference, we conduct a separate test to predict story votes based on its exposure and other characteristics. Specifically, we run Poisson regressions estimating the number of votes each story obtains in subsequent time intervals of 200 seconds each for the first 30 minutes after the story is posted. We use two parallel Poisson regression models in which we set the exposure parameter of poison regression in two different ways. First, behavior reconstruction allows us to use the number of times the story is seen by users (calculated, with the same two methods as the previous test). In the absence of our method the best choice is to use the total activity on Balatarin's (i.e. the total number of votes in the interval) as exposure. Keeping the other independent variables equal, we compare the accuracy of regressions to assess the value of our method for understanding the causes of voting in this setting.

Controlling for exposure, we include the following independent variables in our regressions: stories' categories, their promotion status, their current number of votes, and their place in each of the ordering pages at the beginning of the time interval (Equation 4):

$$\ln(E(\#votesThisInterval)) = \beta_0 + \beta_1.promoted? + \beta_{2-7}.subpage + \beta_{8-13}.location + \beta_{14}.PriorVotes + \beta_{15}.\log(PriorVotes) + \beta_{16-22}.category + \ln(exposure) \qquad (4)$$

Table 2-3 provides regression coefficients for the key independent variables (all statistically significant; dummy coefficients for subpage, location, and category are not shown) as well as measures to compare the performance of the alternative methods. Overall accuracy is the ratio of predicted most likely number of votes (in integer numbers) exactly matching the actual number of votes a story gets in the interval. This metric is not very sensitive because the majority of stories receive no vote and are predictive to receive fewer than 0.5 vote (so that the closest integer is 0). Positive accuracy is a more discriminating measure and calculates the same concept over all the stories with a positive number of votes (i.e. excludes the zero-vote stories). F-measure for regression and Akaike Information Criterion [49] provide other metrics for goodness of fit.

Our two methods perform significantly better than the basic exposure formula across all the measures. They predict the correct number of votes as much as eight times more, and the F-measure improves by a factor of four. Moreover, the exponential function for weighting the likelihood of impression adds significant value in terms of prediction quality and goodness of fit, outperforming the simpler heuristic (i.e. all stories in a subpage with a voted story are seen) across the board. The improvements are pretty significant in magnitude, in fact a perfect model (i.e. correctly predicting the mean of the underlying Poisson process) would still show some error due to the inherent randomness of the Poisson generating processes. Potentially more importantly, the use of behavior reconstruction changes the inferences made about causal effects. Specifically, the directions of the effects switch for the effect of story promotion and logarithm of current votes. Using our method one can infer that stories on the promoted pages are *less* likely to get a vote, if they are seen. It is likely that the base regression gives promotion a positive coefficient because more people visit the promoted pages and thus vote there. Behavior reconstruction allows us to tease this effect apart from the tendency to value promoted stories and provides more reliable inference. Similarly, our method predicts a more modest, and decreasing, return on how much previous votes influence the likelihood of getting a new vote, i.e. the social influence effect in voting [6]. Here, the correlation between having more votes and being presented in more visible areas on Balatarin creates a bias in the base case estimates that our method is able to correct for.

Independent Variable	Behavior reconstruction, stories in voted pages exposed to user	Behavior reconstruction, exposure weighting formula (3)	Overall activity as exposure
Intercept	7.343	16.09	-6.753
Promoted or not	-0.649	-1.172	0.0927
Number of votes	0.0086	0.0056	0.0065
Ln(number of votes)	-0.0414	0.247	-0.263
Performance Metrics			
Overall accuracy	86.03%	86.60%	85.66%
Positive accuracy	27.38%	33.29%	4.49%
F-measure	0.50	0.58	0.14
AIC	8,128,656	6,432,560	11,349,977

Table 2-3. Regression coefficients and performance measures

2.5 Discussion and limitations:

In this paper we propose a generic procedure for deriving users' online behavior from the data on their interactions with objects on a social network and apply it to data from a social news website. Based on the idea that people conserve their efforts in their online behavior, we develop an optimization approach to estimate user browsing, the locations they have visited, and the objects they have viewed. These estimates allow one to distinguish between objects seen but not reacted to and those not seen, a major improvement over common user-object interaction data that does not distinguish between the two. Moreover our method enables the collection of many contextual data items about the object at the time it is seen (e.g. location and number of votes), a valuable resource for enhanced prediction and inference. Behavior reconstruction would enable tracking other valuable information such as duration of online sessions, preference for different pages, and preference for interacting with different parts of the page. Such data can be used to enhance prediction in recommendation systems and other applications. Comparisons with alternative methods in the context of collaborative filtering and regression provide ample evidence for the value of this method for improved prediction and more reliable inference in the context of typical social network data. At its core, this method rests on a cognitively motivated assumption, that users minimize their efforts in their online interactions. This assumption may be violated for interactions initiated by non-human actors (e.g. online robots). Yet, the significant prediction and inference enhancements that result from the method provides further empirical support for this core assumption in the case of humans. In fact, the method is conceptually and algorithmically rather simple; resulting efficiency gains point to the value that considering human cognition and psychology can bring into the design of algorithms. We hope this example stimulates more information systems research that leverages psychological principles in algorithm design.

This method can be applied in diverse problems. By estimating users' online browsing behavior, one can study how individuals build habits, explore a website, and change their browsing patterns over time, on the one hand informing new website designs, and on the other hand facilitating a better understanding of user psychology online. Estimates of impression likelihood can enhance demand forecasting in online markets [1, 50], content customization in social networks [51-53], opinion estimation for online group dynamics, and many other prediction and inference applications with sparse interaction data. Further precision in inference and prediction may result from reconstruction of contextual factors (such as the number of votes and the location on the page at time of impression) otherwise not available in typical archival data.

Behavior reconstruction method can, in theory, be beneficial in any setting where we only observe one type of interaction. For example browsing patterns in physical supermarkets could be reconstructed based on the purchased items and the minimum walking path, providing additional insight into what drives purchase decisions in break and mortar stores. Practical feasibility of different applications depends on a few factors. Data availability is an important consideration. Our method is limited to applications where user-object interaction data with time-stamps is available and the algorithms for site customization can be reconstructed. Researchers with access to data on the location of interactions between users and objects (e.g. data scientists working within social network firms) can skip effort minimization, the second step of the method. Nevertheless, they will benefit from the history reconstruction (for extracting contextual variables not otherwise available, such as the characteristics of the neighboring objects) and impression likelihood estimation (step three). Computational costs for history reconstruction and optimization may become prohibitive if an object can show up in thousands of alternative locations. Thus the most viable candidates are data from websites with moderate levels of customization such as social news (e.g. Reddit) and review (e.g. Yelp) sites, rather than individually customized applications which call for reconstructing a huge number of alternative ordering pages in parallel (e.g. Facebook). Finally, the coding effort to implement this method may be non-trivial, as it entails simulating the history of a website and all its interactions. Fortunately a single simulation is all that is needed for most applications; indeed we did not face computational challenges in conducting this research on ordinary laptops. Despite these limitations we hope the current behavior reconstruction method can be of value to data scientists for diverse prediction and inference applications.

Appendix A

Detailed Algorithm

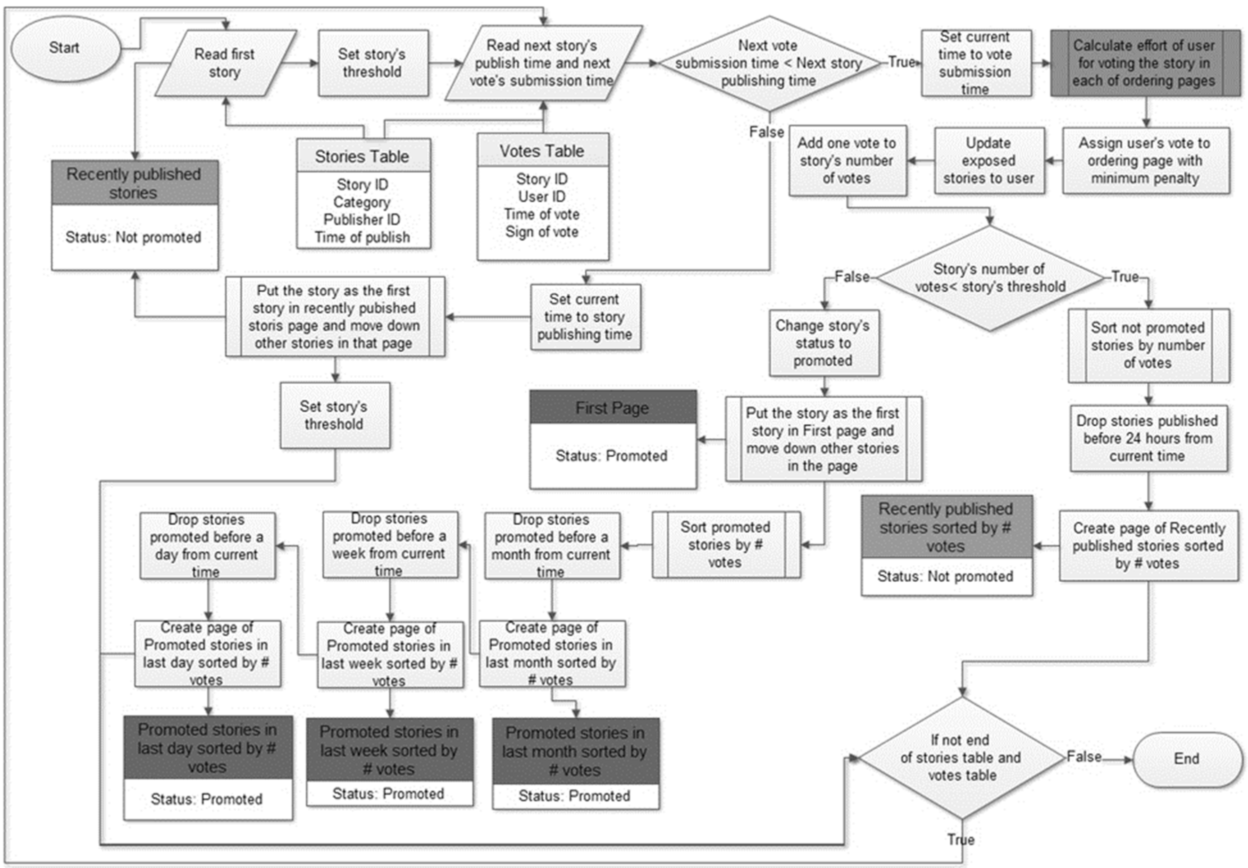


Figure 2-4. Balatarin's history reconstruction algorithm

History reconstruction algorithm in case of Balatarin's website works based on two databases: stories and votes. The former consists of: category of story, ID of story, story's publisher ID, and time of publish. Votes database includes: ID of voter, ID of voted story, Time of vote, and the sign of vote (based on Balatarin's rules, users are only allowed to cast negative votes for stories that they find inappropriate, as a result we ignore negative votes here). The history reconstruction process is summarized in Figure 2-4. Starting from the first published story, we add it as the top story in recently published stories page. Next, we compare the publish time of the next story and time of next vote. If the publish time of the next story is prior to the time of the next vote, we add that story at the top of the recently published stories page and push down other stories in that page. If the next vote comes before next story, we first calculate the effort of user to vote for that story in different possible ordering pages and set the user's current browsing page to the ordering page with the minimum effort. Next, we update the current number of votes for the respective story and compare the number of votes with the promotion threshold in the story's category. If story's vote passes the threshold, we change the story's status to promoted and push it as the first story in the first page. With each casted vote we also update other relative ordering pages (e.g. we sort not-promoted stories based on descending number of votes in

Recently Published Stories Sorted by Number of Votes ordering page). Finally, for each ordering page, we compare the time of publish for each story with the ordering page's lifecycle and drop the stories that passed the lifecycle. This process will continue until reaching the end of our database(s).

Collaborative filtering results with small time window:

The comparisons in the body of the paper do not utilize all of the benefits of history reconstruction, because we wanted to have a fair comparison with other algorithms which do not benefit from extra predictive variables and more parameters which could be tuned for enhanced performance. We thus ran another test to assess the potential upsides of history reconstruction alone. In conducting this test we used a shorter prediction time window and asymmetric weighting parameter θ. We also included the location of stories and the number of stories' votes in predicting user's preferences. The results are reported below.

Method	F_1	F_{10}	Recall	Precision	Accuracy	True Positives	True Negatives	False Positives	False Negatives
Behavior recreation using extracted features	0.62	0.93	0.94	0.47	88%	9,221,602	73,337,399	10,598,438	550,664

Table 2-4. Collaborative filtering metrics with 200 seconds time window, $\theta_1 = 0.01$ and $\theta_2 = 0.02$

References:

1. Linden, G., B. Smith, and J. York, *Amazon. com recommendations: Item-to-item collaborative filtering.* Internet Computing, IEEE, 2003. **7**(1): p. 76-80.
2. López-Nores, M., et al., *MiSPOT: dynamic product placement for digital TV through MPEG-4 processing and semantic reasoning.* Knowledge and Information Systems, 2010. **22**(1): p. 101-128.
3. Lerman, K., *Social networks and social information filtering on digg.* arXiv preprint cs/0612046, 2006.
4. Leskovec, J., D. Huttenlocher, and J. Kleinberg. *Predicting positive and negative links in online social networks*. in *Proceedings of the 19th international conference on World wide web*. 2010. ACM.
5. Lazer, D., et al., *Life in the network: the coming age of computational social science.* Science (New York, NY), 2009. **323**(5915): p. 721.
6. Muchnik, L., S. Aral, and S.J. Taylor, *Social influence bias: A randomized experiment.* Science, 2013. **341**(6146): p. 647-651.
7. Aral, S. and D. Walker, *Identifying influential and susceptible members of social networks.* Science, 2012. **337**(6092): p. 337-341.
8. Liang, T.-P., H.-J. Lai, and Y.-C. Ku, *Personalized content recommendation and user satisfaction: Theoretical synthesis and empirical findings.* Journal of Management Information Systems, 2006. **23**(3): p. 45-70.
9. Ghoshal, A., S. Kumar, and V. Mookerjee, *Impact of Recommender System on Competition Between Personalizing and Non-Personalizing Firms.* Journal of Management Information Systems, 2015. **31**(4): p. 243-277.
10. Pariser, E., *The filter bubble: How the new personalized Web is changing what we read and how we think*. 2011: Penguin.
11. Demiriz, A., *Enhancing product recommender systems on sparse binary data.* Data Mining and Knowledge Discovery, 2004. **9**(2): p. 147-170.
12. Lee, J.-S., et al., *Classification-based collaborative filtering using market basket data.* Expert Systems with Applications, 2005. **29**(3): p. 700-704.
13. Aral, S., *The problem with online ratings.* MIT Sloan Management Review, 2014. **55**(2): p. 47.
14. Salganik, M.J., P.S. Dodds, and D.J. Watts, *Experimental study of inequality and unpredictability in an artificial cultural market.* science, 2006. **311**(5762): p. 854-856.
15. Ver Steeg, G., R. Ghosh, and K. Lerman. *What stops social epidemics?* in *ICWSM*. 2011.
16. Hodas, N.O. and K. Lerman. *How visibility and divided attention constrain social contagion*. in *Privacy, Security, Risk and Trust (PASSAT), 2012 International Conference on and 2012 International Confernece on Social Computing (SocialCom)*. 2012. IEEE.
17. Pan, R., et al. *One-class collaborative filtering*. in *Data Mining, 2008. ICDM'08. Eighth IEEE International Conference on*. 2008. IEEE.
18. Marlin, B., et al., *Collaborative filtering and the missing at random assumption.* arXiv preprint arXiv:1206.5267, 2012.
19. Oard, D.W. and J. Kim. *Implicit feedback for recommender systems*. in *Proceedings of the AAAI workshop on recommender systems*. 1998.
20. Jawaheer, G., M. Szomszor, and P. Kostkova. *Comparison of implicit and explicit feedback from an online music recommendation service*. in *proceedings of the 1st international workshop on information heterogeneity and fusion in recommender systems*. 2010. ACM.
21. Lee, T.Q., Y. Park, and Y.-T. Park, *A time-based approach to effective recommender systems using implicit feedback.* Expert systems with applications, 2008. **34**(4): p. 3055-3062.

22. Lee, C.-H., Y.-H. Kim, and P.-K. Rhee, *Web personalization expert with combining collaborative filtering and association rule mining technique.* Expert Systems with Applications, 2001. **21**(3): p. 131-137.
23. Núñez-Valdéz, E.R., et al., *Implicit feedback techniques on recommender systems applied to electronic books.* Computers in Human Behavior, 2012. **28**(4): p. 1186-1193.
24. Cho, Y.H. and J.K. Kim, *Application of Web usage mining and product taxonomy to collaborative recommendations in e-commerce.* Expert systems with Applications, 2004. **26**(2): p. 233-246.
25. Hu, Y., Y. Koren, and C. Volinsky. *Collaborative filtering for implicit feedback datasets*. in *Data Mining, 2008. ICDM'08. Eighth IEEE International Conference on*. 2008. IEEE.
26. Yin, P., et al. *Silence is also evidence: interpreting dwell time for recommendation from psychological perspective*. in *Proceedings of the 19th ACM SIGKDD international conference on Knowledge discovery and data mining*. 2013. ACM.
27. Facca, F.M. and P.L. Lanzi, *Mining interesting knowledge from weblogs: a survey.* Data & Knowledge Engineering, 2005. **53**(3): p. 225-241.
28. Munk, M., J. Kapusta, and P. Švec, *Data preprocessing evaluation for web log mining: reconstruction of activities of a web visitor.* Procedia Computer Science, 2010. **1**(1): p. 2273-2280.
29. Alves, R., et al. *Clickstreams, the basis to establish user navigation patterns on web sites*. in *Fifth International Conference on Data Mining, Text Mining and their Business Applications*. 2004.
30. Clark, L., et al., *Combining Ethnographic and Clickstream Data to Identify User Web Browsing Strategies.* Information Research: an international electronic journal, 2006. **11**(2).
31. Freeman, J.B. and N. Ambady, *MouseTracker: Software for studying real-time mental processing using a computer mouse-tracking method.* Behavior Research Methods, 2010. **42**(1): p. 226-241.
32. Pan, B., et al. *The determinants of web page viewing behavior: an eye-tracking study*. in *Proceedings of the 2004 symposium on Eye tracking research & applications*. 2004. ACM.
33. Berendt, B., et al., *The impact of site structure and user environment on session reconstruction in web usage analysis*, in *WEBKDD 2002-Mining Web Data for Discovering Usage Patterns and Profiles*. 2003, Springer. p. 159-179.
34. Kuo, F.-Y., et al., *An investigation of effort–accuracy trade-off and the impact of self-efficacy on Web searching behaviors.* Decision Support Systems, 2004. **37**(3): p. 331-342.
35. Li, Y., L. Lu, and L. Xuefeng, *A hybrid collaborative filtering method for multiple-interests and multiple-content recommendation in E-Commerce.* Expert Systems with Applications, 2005. **28**(1): p. 67-77.
36. Herlocker, J.L., et al., *Evaluating collaborative filtering recommender systems.* ACM Transactions on Information Systems (TOIS), 2004. **22**(1): p. 5-53.
37. Han, P., et al., *A scalable P2P recommender system based on distributed collaborative filtering.* Expert systems with applications, 2004. **27**(2): p. 203-210.
38. Pathak, B., et al., *Empirical analysis of the impact of recommender systems on sales.* Journal of Management Information Systems, 2010. **27**(2): p. 159-188.
39. Wang, J., *Encyclopedia of data warehousing and mining*. 2006: IGI Global.
40. Sahoo, N., et al., *Research Note-The Halo Effect in Multicomponent Ratings and Its Implications for Recommender Systems: The Case of Yahoo! Movies.* Information Systems Research, 2012. **23**(1): p. 231-246.
41. Roh, T.H., K.J. Oh, and I. Han, *The collaborative filtering recommendation based on SOM cluster-indexing CBR.* Expert Systems with Applications, 2003. **25**(3): p. 413-423.
42. Koren, Y., R. Bell, and C. Volinsky, *Matrix factorization techniques for recommender systems.* Computer, 2009. **42**(8): p. 30-37.

43. Takács, G., et al., *Scalable collaborative filtering approaches for large recommender systems.* The Journal of Machine Learning Research, 2009. **10**: p. 623-656.
44. Abernethy, J., et al., *A new approach to collaborative filtering: Operator estimation with spectral regularization.* The Journal of Machine Learning Research, 2009. **10**: p. 803-826.
45. Nocedal, J. and S.J. Wright, *Conjugate gradient methods*. 2006: Springer.
46. Olson, D.L. and D. Delen, *Advanced data mining techniques [electronic resource]*. 2008: Springer.
47. Rijsbergen, C.J., *Information Retrieval, 2nd ed.* 1970: Butterworths, London, UK.
48. Su, X. and T.M. Khoshgoftaar, *A survey of collaborative filtering techniques.* Advances in artificial intelligence, 2009. **2009**: p. 4.
49. Sakamoto, Y., M. Ishiguro, and G. Kitagawa, *Akaike information criterion statistics.* Dordrecht, The Netherlands: D. Reidel, 1986.
50. Mild, A. and T. Reutterer, *An improved collaborative filtering approach for predicting cross-category purchases based on binary market basket data.* Journal of Retailing and Consumer Services, 2003. **10**(3): p. 123-133.
51. Konstas, I., V. Stathopoulos, and J.M. Jose. *On social networks and collaborative recommendation*. in *Proceedings of the 32nd international ACM SIGIR conference on Research and development in information retrieval*. 2009. ACM.
52. Cai, X., et al., *Collaborative filtering for people to people recommendation in social networks*, in *AI 2010: Advances in Artificial Intelligence*. 2011, Springer. p. 476-485.
53. Susarla, A., J.-H. Oh, and Y. Tan, *Social networks and the diffusion of user-generated content: Evidence from YouTube.* Information Systems Research, 2012. **23**(1): p. 23-41.
54. Putnam, R.D., *Bowling alone: The collapse and revival of American community*. 2001: Simon and Schuster.

3. Chapter 3 - Measuring Individual Differences: A Big Data Approach

3.1 Abstract

The amount of behavioral and attitudinal data we generate every day has grown significantly. This era of Big Data has enormous potential to help psychologists and social scientists understand human behavior. Online interactions may not always signify a deep illustration of individuals' beliefs, yet large-scale data on individuals interacting with a variety of contents on specific topics can approximate individuals' attitudes toward those topics. We propose a novel automated method to measure individuals' attitudes empirically and implicitly using their digital footprints on social media platforms. The method evaluates content orientation and individuals' attitudes on dimensions (i.e. subjects) to explain individual-content ratings in social media, optimizing a pre-defined cost function. By applying this method to data from a social news website, we observed a significant test-retest correlation and substantial agreement in inter-rater reliability testing.

Keywords: Implicit attitude measuring, Big data, Social media

3.2 Introduction

The amount of digital data we generate every day has reached a stage where Facebook alone has more than 300 petabytes of data stored about us, with an incoming daily rate of 600 terabytes[7]. We send more than three hundred thousand tweets, upload more than three hundred hours of video on YouTube, and 'like' more than a million pictures and four million posts on Instagram and Facebook per minute. An increasing fraction of our social interactions occurs online in social networks, social news websites, forums, and other internet-based media through posting, sharing, commenting, and other forms of digital interaction. Based on the GWI Social report[8], as of 2015 a typical internet user spends on average 1.77 hours per day on social networks, while younger generations spend a lot more time than that (2.68 hours for 16-24 years old and 2.16 for 25-34s) on social media. Online social interactions also shape our moods [1], affect our behavior [2, 3] and impact our non-virtual world socially [4] and even politically [5]. We express ourselves in interactions with socially generated digital objects (posts, comments, videos, sounds and pictures) in forms such as "like", "share", "vote", "retweet", and "pin". These interactions form valuable data for understanding human personalities, behaviors and their attitudes and opinions toward different issues and subjects.

Attitudes, according to Blum and Naylor [6], are beliefs, feelings and action tendencies toward an idea (or object, people, etc.), which can facilitate or hinder actions (see reasoned action [7, 8]); given their central role in human action, measuring attitudes is critical for social scientists. Measuring attitudes can help us understand and predict individuals' behavior in society regarding an issue (e.g. see the ABC model of attitude [9]). As a result, knowledge of a population's attitude can guide the planning and implementation of social policy, or design of better products. Additionally, it can provide policy makers with insights into public responses to various policies before they are implemented. Although attitudes can be evaluated explicitly (i.e. by asking people directly about their opinion on a subject), factors such as fear of social judgment and tendency to appear well adjusted, unprejudiced and open minded could affect the measurement dramatically and could lead us to biased presumptions and deductions about

[7] https://code.facebook.com/posts/229861827208629/scaling-the-facebook-data-warehouse-to-300-pb/

[8] http://insight.globalwebindex.net/social

the subjects. In order to avoid that, researchers tend to measure attitudes indirectly. Different implicit methods have been proposed in the literature, such as: priming procedures [10], which evaluate what is activated from the memory by the presentation of some object; Implicit Association Tests (IAT) [11], which evaluate the strength of an association between two concepts using the latency in responding to them as a single unit; and Word fragment, Go/No-Go Association Task (GNAT), among others [12]. Comparing implicit methods to explicit ones, implicit methods are less affected by intentional efforts to deceive [13]; they are more valid for unconscious attitudes and controversial/sensitive subjects [14] and are more direct than explicit ones (since, unlike implicit measures, explicit measures must go through the conscious processing system) [13]. Yet current implicit attitude measuring methods also have many drawbacks. Lemm et al. [15] showed that self-perception could affect measured attitudes toward homosexuality. Asendorpf's study [16] illustrated how self-observation changes measured shyness when the subjects are involved in an interaction with an attractive stranger of the opposite sex. Rudman et al. showed the effect of prior exposure on implicit measuring by exposing subjects to violent rap music before a racial prejudice test [17]. Wittenbrink [18] illustrated how implicit measures could be sensitive to context by changing the background pictures of Black/White faces in a racial prejudice test. Low reliability of response-latency measures, low inter-item consistency (correlation over subjects) and low stability (correlation overtime) are some other drawbacks of the current implicit attitude measuring methods [14]. Additionally, we are still not sure what implicit attitude evaluation methods (such as IAT) are actually measuring [19-21]. Even using implicit measuring methods, the fact that we need to evaluate the subjects' attitudes in a controlled and staged environment would change their behavior. Moreover, designing and implementing the tests and recruiting subjects for an experiment is a costly process; therefore, old-fashioned experiments are not very scalable. Finally, with current implicit measuring methods, collecting continuous time series data (i.e. for studying changes in the attitudes of individuals) is not an easy task, since we need to contact the subjects continuously and hope for their participation. Using available individuals' online interactions data however, can help researchers to collect data continously without bothering the subjects, and avoid unnatural behavior produced due to artificiality of the attitude measuring methods setting. It may also reduce the cost, enhance the scalibility of the experiments and overcome other drawbacks of the implicit attitude measuring methods.

Although online interactions (such as posts, likes and votes) may not always signify a deep illustration of individuals' beliefs, large-scale data on individuals' interactions with a variety of contents on specific subjects (and in different circumstances) can cumulatively represent individuals' attitudes toward the subjects. For instance, Kosinski et al. [22] show that some private traits and attributes are predictable from people's digital records. They use logistic and linear regressions on Facebook Likes data and effectively predict personal attributes such as sexual orientation, religious and political views, along with individual traits such as intelligence, happiness and drug abuse. Linking available data with other individuals' characteristics (e.g. psychological and sociological characteristics) can assist us even further in understanding a particular society. Rentfrow et al. [23], for instance, used data available on over more than 1.5 million subjects (from multiple sources including but not limited to Gosling-Potter Internet Personality Project, Rentfrow-Potter Music Preference Project and MyPersonality Facebook application) to study three psychological profiles (Friendly & Conventional, Relaxed & Creative, and Temperamental & Uninhibited) and their geographical distributions in the U.S. They also studied the political views, economic status, social values, and health status of people in different sections of the U.S. They found that people who live in Middle America are more friendly and conventional and conservative in social values. They have moderately high levels of extraversion, agreeableness and conscientiousness, moderately low neuroticism, and very low openness. They have comparatively low levels of education, wealth, economic innovation, and social tolerance and tend to be politically conservative and religious, with unhealthy lifestyles. Due to the availability of data, in some cases computer-based personality evaluations and traits judgments even outperform those of humans. Youyou et al. [24] compared the accuracy of human and computer-based personality judgments on a sample of 86,220 volunteers who completed a personality questionnaire. They used LASSO linear regression to extract the Big Five personality features from subjects' Likes on Facebook, and compared their accuracy with human personality judgments (obtained from the participants' Facebook friends) and self-ratings on some key criteria (such as self-other agreement, inter-judge agreement, and external validity). They found that computer-based judgments correlate more strongly with participants' self-ratings than the average human judgments (r=0.56 to 0.49). Their automated computer-based traits judgments outperformed human judges in 12 of their 13 criteria and worked even better than self-rated personality judgments in areas such as substance use and political attitudes. The results of mentioned studies (and other studies in the field) pushed our understanding of this field to a new level. Yet, the lack of a step-by-step methodology for attitude measuring based on the online interaction data (that considers individuals' characteristics and contents' properties in the interactions) is notable.

In this study we propose a novel automated method to measure individuals' attitudes (toward different issues) empirically and implicitly using the available interaction data in social media. The proposed method is reliable in evaluating attitudes and resolves many of the drawbacks of the current implicit methods. Based on user-object interaction data in social media, the method projects and evaluates users' attitudes on dimensions that explain users' behavior in interacting with online objects (i.e. liking, voting or rating the objects) by optimizing a pre-defined cost function. We then extract the underlying meaning of the dimensions and align evaluated attitude values using linear regressions. We apply the proposed method to publicly available data from a social news website, validate the proposed method and identify its characteristics compared to alternative implicit methods.

3.3 Method

Measuring attitudes using individuals' online interaction data is a fairly new topic in psychology literature; however, to some extent, it has been studied before in other fields with different terminology. Opinion mapping methods in the field of computer science, for example, measure the opinions of individuals (on different issues) and have the same concepts as attitude measuring in psychology. On this subject, Opinion Space [25] and EU Profiler [26] are two tools developed to measure and map opinions on multidimensional space, based on answers provided by individuals to some predefined questions. Both use principal component analysis (PCA) and reduce the dimensionality of the data collected from individuals (on a continuous or discontinuous Likert scale) to two dimensions. These methods require original surveys to be administered and therefore time series data collection, which requires multiple measures of the same person over time, is challenging.

In data visualization literature, Multi-Dimensional Scaling (MDS) [27] as a method of visualizing the level of similarity across individuals (or objects) in a dataset, has been used to map individuals (or objects) in n-dimensional "ideology" (i.e. perception, opinion, or attitude) space. Perceptual mapping techniques [28], for instance (as an application of MDS in marketing research), are well-known for visually displaying the perceptions of customers about products. NOMINATE [29], a tool used in political science literature, maps individuals (e.g. members of Congress) on an ideology space (e.g. left-right or liberal-conservative spectrum) using a MDS method on individual-object rate data (e.g. the House's vote on a bill). In short, MDS methods take a (object-object or user-user) distance matrix (e.g. $\delta_{i,j}$ as the distance between user i and j) as the input and estimate ideology values that explain the distance matrix (e.g. $\delta_{i,j} \approx \|x_i - x_j\|$, where x_i and x_j are ideology values for users i and j). Using NOMINATE, Barberá et al. [30] measured political ideology (i.e. the extent of one's being a democrat, independent or republican in political orientation) based on individuals' connections in social network. They assumed that the probability of a connection between two users in a given network is negatively correlated to their distance in a "latent ideological space". They applied their method to data available from Twitter's followers-followee network and (using correspondence analysis) measured users' "political ideology" on one dimension. Network data (i.e. friends' network data on Facebook or follower-followee network data on Twitter) could represent a community (of likeminded people) to which individuals belong. However, it is not very reliable for attitude (opinion or ideology) measuring, since the magnitude and the structure of the network may vary based on the individual's connectivity (i.e. the number of users to whom an individual is connected, which could affect the individual's amount of activity, openness, etc. on the social media). For instance, a political journalist (or anyone who is curious about politics) may be connected to (i.e. may "follow" in the case of Twitter) many different users, even from other political parties, just to get their news and opinions. Moreover, the probability of having a connection between two individuals is not only dependent on their political leanings, but also on many other factors which will confound any unidimensional estimate from this data. Besides, user-online objects' interaction data (i.e. Likes on Facebook, retweets on Twitter, etc.) is much richer than network data and includes information on individuals' attitudes on multiple topics and with a higher degree of accuracy because individuals decide on liking or retweeting more based on their opinions on the content of the object than other confounding factors. On a similar subject, Bond and Messing measured the political ideology of Facebook users based on the political Facebook pages they Liked [31]. They used a utility matrix derived from the Likes data (as the input of NOMINATE) and decomposed the utility matrix to an "ideology measure for the pages" using the singular value decomposition method. Then, by averaging the scores for the political pages that the user had Liked, they derived an estimation of the "users' ideological location". MDS methods in general are based on distance matrices and, when applied to user-object interaction datasets, they treat the users and objects as independent entities. For instance, Bond and Messing [31] explained that to extract a user's ideological location directly from the data (i.e. instead of by averaging the pages' ideological values for each user) they needed to derive the user's distance matrix and decompose it separately (which they avoided due to technical difficulties in decomposing a large user-user distance matrix). In addition to this, MDS methods do not provide any specific technique for inferring the dimension's definition. For instance, Poole and Rosenthal (in their application of NOMINATE [29]) just assumed that the dimension (in one-dimensional space) was representative of the political left-right spectrum (which is probably the first explanatory factor in the House's vote on a bill in most cases).

Here we propose a method to measure and map the attitudes (i.e. ideology, opinion, etc.) of individuals using their user-online object interaction data on social media. The proposed method uses existing user-object interaction data directly and does not rely on survey results. Thus, it is suitable for collecting time series data on the attitudes of people (at an individual level) for any application (e.g. attitude change through time, individuals' reactions when interacting with people of different opinions). Additionally, unlike network data, user-object interaction data represents opinion of individuals toward the contents of objects; it is not related to network's structure nor to individual's connectivity. Unlike MDS methods, proposed method measures the attitudes of users and the contents (i.e. objects) point of view at the same time and treats them as related entities. Besides, the method comes with a technique for interpreting the (attitude space) dimensions in detail. The method can also extract and control for the effects of exogenous variables on attitude measuring. It is highly flexible and can be used on different types of data (e.g. binary voting, continuous grading, 5-star ratings, Likert-based rating) by modifying the cost function and the choice of the optimization algorithm.

Based on user-object interaction data found in social media, we can get an idea of individuals' preferences, tastes and attitudes on different subjects (i.e. by looking at a Netflix user's ratings we can estimate his/her tastes regarding movie genres). Recommendation systems (i.e. the Netflix movie recommendation system) use such user-object interaction data and make their recommendations based on the idea that users with similar tastes rate objects (roughly) similarly. Matrix factorization methods, which are most commonly applied to recommendation systems, aim to explain individuals' rating criteria by characterizing both objects and individuals on factors inferred from the rating patterns [32]. Here we use the idea of recommendation algorithms and develop a method to empirically measure and map the attitudes of individuals with regard to different matters, using their online interaction data. While the core idea of our method is closely related to recommendation systems, our method has a few novel features that contribute to this literature: A) We design and extend a method for collaborative filtering that is scalable to large datasets and applies to binary data (e.g. when individuals just like, or retweet, rather than rating an object on a continuous scale, such as in movie ratings). B) We control for various non-opinion factors that influence individuals' interaction patterns. C) We develop and test a method for extracting meaningful opinion dimensions from the otherwise abstract opinions estimated using existing methods. Below we first introduce the basic collaborative filtering method using by many recommendation systems, and then go into the details of our method.

One of the most commonly implemented (matrix factorization based) recommendation system methods is collaborative filtering, which is defined as "a method of making automatic predictions about the interests of a user by collecting taste information from many users" [33]. Collaborative filtering starts with a common data structure, in which users and objects (e.g. products, movies, stories) are separately identified. It then moves forward based on a simple assumption: users who similarly rate a set of objects have similar tastes/opinions to each other, compared with other users who do not show such similarity in their ratings. Thus, these methods use matrix factorization techniques to form taste vectors for users and objects. By estimating the elements of a k-dimensional taste vector assigned to each user and object, matrix factorization collaborative filtering techniques find the taste/opinion values that minimize the difference between the observed and expected ratings. In essence, this procedure estimates a taste vector for each user and each object, so that similarly rated objects and corresponding raters (users) have only a small distance between their taste vectors. As a result, users with a low degree of similarity in rating have a larger distance between their taste vectors than those with greater similarity. While such 'k-dimensional taste vectors' are abstract and algorithmic by design, we hypothesize that they could be transformed into positions on a meaningful opinion space. So having estimated the taste vectors we should find the underlying human-understandable concepts corresponding to each dimension in the space and then map the attitudes of users toward those concepts.

In computer science and machine learning literature, the idea of measuring taste has been viewed in the context of predictive models (e.g. to predict which movie a user would like, considering his/her previous ratings), rather than focusing on the underlying concept of taste vectors or measuring attitudes on pre-specified dimensions. Thus, most of the methods that are used by computer scientists produce taste values as a sub-product of their process and do not attempt to map those values to individuals' real attitudes toward the object's actual characteristics. In a movie rating website, for instance (e.g. Netflix), users rate movies based on their attitudes toward different characteristics of the movies (i.e. genre, acting, storyline, etc.). Recommendation systems will then estimate a taste vector for each user and each movie based on those ratings. However, those taste vectors only show the relative closeness of the users' tastes, rather than investigating the underlying meaning of the values in the taste vector. For example, three users could have taste vectors of [3, 4, -1], [2.8, 4.2, -0.9] and [-2, -3, 1]; these values indicate that the first two users have closer opinions compared to the third one. As a result, the first user probably likes movies that the second user rated more highly than did the third user; however, those values do not indicate the attitudes of those users toward movie characteristics (i.e. whether any of them like movies of the horror genre more than others). Additionally, due to the predictive purpose of machine learning methods, the underlying reason for the ratings is ignored in those methods. For instance, when a user rates the movie Fight Club as 4.5 out of 5, this could indicate that the user likes the horror genre or movies directed by David Fincher or movies in which Brad Pitt acts (or any of those factors could have affected the user's rating). Recommendation systems ignore the underlying reasons that cause users to rate a movie highly or poorly. They merely attempt to identify users with similar tastes (without even knowing what that underlying taste is) in order to recommend movies rated highly by the other users. As a result, a movie recommendation system would recommend a range of movies to the user, from Forrest Gump (which has almost no similarity to the previously rated movie) to Se7en (which has the same genre, director and leading actor) based on the ratings of 'similar' users.

There are other challenges in using the user taste vector produced by the current collaborative filtering methods. First, the taste vector estimated using such methods of matrix factorization does not distinguish between opinion-based drivers of the rating and other factors that may affect it. Factors such as the popularity of the content subject [34], content producer (e.g. the actor or director of a movie, the publisher of a story or the content, etc.) [35], level of exposure, level of user's activity, level of content's attractiveness and many other factors can affect the rating beside the actual opinion of the user on the content's characteristics (e.g. the genre or the storyline of a movie). Second, such methods usually create the taste vector of each user on many dimensions (Facebook's news feed recommendation system, for instance, uses more than 100,000 dimensions) but, as previously mentioned, they do not map these dimensions to any concept (or object's characteristics) and simply use them to evaluate the distance between the contents (or the users). While having 100,000 dimensions is useful for predicting user's preferences, it is not informative about the underlying meanings of those dimensions. Second, current methods mix non-responses and negative responses (e.g. whether the user did not rate a movie because s/he did not like it or has not watched it), thus reducing their accuracy [36]. Third, the design of current methods is focused on continuous scale for rating (while many of the behavioral data we have is binary, e.g. Facebook Likes, etc.). Finally, rotation insensitivity in the matrix factorization method (which will be discussed later) will make recommendation methods, in their current form, ineffective in measuring individuals' attitudes.

Building on the underlying collaborative filtering algorithms, we develop a method that measures the attitudes of individuals, but is tailored to resolve the abovementioned drawbacks. In the proposed method we focus on the more difficult case of binary choices (however, it can be used on continuous data or other integer ratings such as 5-stars) using a logistic-based cost function. We use a likelihood based cost function, which can be scaled for very large problems or different types of available data. Incorporating various control variables in the function, we tease-out the non-opinion drivers of ratings. We resolve the effect of non-responses by weighting those based on the probability of exposure (discussed in the first essay). We also provide a technique to map the extracted taste dimensions of individuals to real-world concepts that affect their ratings and validate the results using human evaluators.

Before going into details of our algorithm we use a simple example to build intuition about the method and illustrate common collaborative filtering algorithms and how we address the various challenges discussed. Consider three individuals (Dan, Simon, and Matilda) who have rated four movies (Fight Club, Die Hard, Pride and Prejudice, and The Silence of The Lambs) based on their tastes on a 5-point scale, as follows:

	Fight Club	**Die Hard**	**Pride and Prejudice**	**The Silence of The Lambs**
Dan	5	2	1	4
Simon	3.5	5	1	1
Matilda	2.5	1	5	2

Table 3-1. Sample user-object rating matrix

Let us call this rating matrix R and try to factorize it into two taste vectors in such a way that $R_{3\times 4} = U_{3\times k}.V_{4\times k}{}^{T}$, where U is the taste vector of the individuals and V is the same for movies. Note that the column number (k) is not fixed for the taste vectors and is based on the number of factors which we think could affect individuals' ratings of movies (though here limited data would exclude large k); we can set different numbers of dimensions for factorization. Here, we assume that the individuals rated the movies based only on their interest in various genres (horror, action or romance). Factorizing R into two 3-dimensional taste vectors, *one of the results* could be the following:

$$U = \begin{bmatrix} 2 & 1 & 0.5 \\ 0.5 & 2.5 & 0.5 \\ 1 & 0.5 & 2.5 \end{bmatrix}, V = \begin{bmatrix} 2 & 1 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 2 \\ 2 & 0 & 0 \end{bmatrix},$$

where, in our example, $R_{3\times 4} = U_{3\times 3}.V_{4\times 3}{}^{T}$ perfectly. Knowing (from prior knowledge not reflected in the data) the genre of these movies (i.e. Fight Club is a horror action movie; Die Hard is an action movie; Pride and Prejudice is in the romance genre; and The Silence of The Lambs is a horror movie), we can guess that the columns in V reflect the score (i.e. taste) of movies in the horror, action and romance genres, respectively; and the columns in U show the taste of individuals in each of these genres in the same order. So it can be said that Dan really likes horror movies (compared to the others), and he also enjoys action movies, but he is not a fan of romance movies. Simon loves action movies but does not like horror or romance movies. Matilda, on the other hand, loves romance movies; she enjoys horror movies too, but is not a fan of action movies. Thus, as a result of our factorization, we are able to assign a number to the attitude of each individual toward each movie genre.

Note that matrix factorization is not always as trivial as it appeared here. In real-world problems, rating matrices can rarely be factorized perfectly; for instance, we cannot factorize the R matrix, shown above, in 2-dimensional taste vectors perfectly (i.e. there are no U and V such that $R = U.V^T$). Additionally, the size of R could increase based on the multiplication of individuals and objects, so large data applications will be more complex. When dealing with such problems, the matrix factorization process usually requires the definition of a cost function to get to a fairly close estimate of the unknown taste vectors, and the use of a proper optimization algorithm. The cost function should be defined in such a way that it leads the optimization algorithm toward the best possible taste vectors that form a rating matrix close to R. For the case shown in the example, $f(U,V) = (R - U.V^T)^2 = \sum_{i=1}^{3}\sum_{j=1}^{4}(R_{i,j} - U_{i.}V_{j.}{}^T)^2$ could be a good choice of cost function, since minimizing $f(U,V)$ pushes U and V toward points that reduce the distance between the actual rating matrix (i.e. R) and the estimated one ($\hat{R} = U.V^T$).

A common complication in matrix factorization is handling non-rated items (i.e. non-response). This occurs when there are items that have not been rated (or not even viewed) by some individuals. In our movie rating example, if Fight Club had not been rated by Simon ($R_{2,1} = 0$) then we should not consider $R_{2,1}$ in the optimization process, otherwise (i.e. if we put $R_{2,1} = 0$) the optimization process would push Simon's taste value toward not liking Fight Club at all (rating it zero), which is not implied by data since he had not seen the movie. In fact, in real-world problems, we usually deal with very sparse rating matrices with over 90% of cells in R matrix being zero due to non-observation (very few people rate every object). To deal with non-response cases we can weight non-rated user-object pairs with zero (or with small values if there is a chance that the individual did not rate the object because s/he did not like it) in the cost function to nullify (or reduce) the effect of non-response cases in the estimation process. In our example, by adding the exposure weighting matrix to the cost function we obtain $f(U,V) = \sum_{i=1}^{3}\sum_{j=1}^{4} W_{i,j} \times (R_{i,j} - U_{i.}V_{j.}{}^T)^2$, where $W_{2,1} = 0$ and $W_{i\neq 2, j\neq 1} = 1$, which means that Dan's rating on Fight Club has no effect on the estimations of U and V.

Another issue that has to be noted in defining the cost function for matrix factorization is the upper/lower bounds in ratings. In our example we assume that the ratings are bounded between 1 and 5, which means that we limit individuals in their ratings. Bounding the ratings may seem logical in the application; however, it adds an unreasonable limitation to our cost function. For instance, in our example Simon rated both Pride and Prejudice and The Silence of The Lambs as one, but in reality it may be that he dislikes romance movies more than horror movies but is unable to illustrate his taste because of the lower bound. The current cost function $f(U,V)$ forces the optimization process to push $U_{2.}$, $V_{3.}$ and $V_{4.}$ to the point where $U_{2.}.V_{3.}{}^T = U_{2.}.V_{4.}{}^T = 1$, but with the above assumption, the value for $U_{2.}.V_{4.}{}^T$ should be less than $U_{2.}.V_{3.}{}^T$ since Simon dislikes romance movies (such as Pride and Prejudice) more than horror ones (The Silence of The Lambs). We can deal with this issue by using cost functions that do not limit $U.V^T$ to upper and lower bounds; one example is the Sigmoid function, the application of which will be discussed later.

Beside the individual's opinion, many other factors could affect the ratings on the objects. For instance, some individuals tend to rate higher than others, while some objects are more attractive than others (they attract higher ratings regardless of the point of view). Such factors can be considered in the cost function using fixed effect variables. In our movie example, adding a fixed effect to individuals and movies taste vectors could be done by adding a column of ones to the taste vectors. Thus, Dan's estimated rating on Fight Club on our 3-dimensional taste space would be $U_{1.}.V_{1.}^{\ T}$ = U_{11} + V_{11} + U_{12} × V_{12} + U_{13} × V_{13} + U_{14} × V_{14}, in which U_{11} and V_{11} are the fixed effects for Dan and Fight Club, respectively. A higher U_{11} implies that Dan rates higher on average (compared to other individuals in the R matrix) and a high V_{11} shows that Fight Club is more attractive compared to the other movies rated. Aside from the individual factors (i.e. fixed effects), some of the objects' time-varying properties could also affect the ratings. Again, in our movie example, people may rate some of the movies higher because of their popularity (i.e. a current movie's high average rating) regardless of personal taste. This effect can also be captured using factor variables. The factor variable for movie popularity could be considered in the cost function as $f(U,V,\beta) = \sum_{i=1}^{3}\sum_{j=1}^{4}(R_{i,j} - U_{i.}V_{j.}^{\ T} + \beta \times r_j)^2$, where r_j is the popularity of the movie (i.e. the movie's current average rating) and β should be estimated by the optimization algorithm along with U and V. Although factor effects could vary between different individuals (i.e. a currently high average rating may affect some individuals more than others), for simplicity we assign only one coefficient to each factor (i.e. we estimate the general effect of the factor variable on the ratings).

Lastly, the matrix factorization process is rotation-insensitive, which means that by factorizing a single rating matrix multiple times (optimizing it from different starting points) we could end up with different taste matrices each time. These different pairs of matrices (U and V) are all products of a single pair transformed (rotated and/or scaled) around the origin. In other words, when we factorize a matrix such that $R \sim U.V^T$ by multiplying the right-hand side of the equation in any (invertible) matrix of A and its inverse, we have $R = U.V^T = (U.A).(V(A^{-1})^T)^T = U'.V'^T$. This means that our actual underlying taste concepts (i.e. in our movie example case, different genres) may not be aligned with the axis of the factorized taste vectors. To adjust the axis of the taste vectors to the underlying concepts, we need to first guess the underlying concepts that would affect the ratings (genres, in our example), then score some of the objects on those concepts manually, rotate the automatically-scored taste vectors until they correspond to the manually scored taste vectors (using correlation, regression or similar methods), find the rotation matrix that best fits, and finally transform estimated U and V using this rotation so that the estimated taste vectors align with concepts we can relate to. We will discuss this technique further in the next section.

We first extend the existing matrix factorization methods (introduced above) to the case where ratings are zero-one/vote based (e.g. Likes on Facebook, or Votes on Reddit) and observations are sparse (not every user has seen every object). The details of this method are discussed below but, in short, we use a novel method for identifying those objects a user has likely observed (introduced in the previous essay and with more detail in [36]) and then define a maximum likelihood estimation procedure for estimating user and object taste vectors.

As discussed, the main idea is to decompose (i.e. factorize) a user-object rating matrix (i.e. $R_{m \times n}$, where $R_{i,j}$ is the rate user i assigned to object j) into (dot) product of two (users' and objects') taste matrices (i.e. $R_{m \times n} = U_{m \times k}.V_{n \times k}^{\ T}$ where $U_{i.}$ is user i's taste vector and $V_{j.}^{\ T}$ is the taste vector of object j (in a k dimensional opinion space)). To decompose (user-object voting) binary matrices, we use a simple logistic probability model. Let us assume that a user with taste vector $U_{i.}$ votes for an object (movie, music, story, etc.) with taste vector $V_{j.}$ with probability $F(U_{i.}.V_{j.}^{\ T}) = \frac{1}{1+e^{-U_{i.}.V_{j.}^{\ T}}}$ which, in a one-dimensional space, means: 1) for a neutral user ($U_{i.}$ = 0) (or neutral object ($V_{j.}$ = 0)) there is a 50% chance of voting regardless of the object (or user) taste; 2) for a very biased user ($U_{i.}$ = ∞ or $U_{i.}$ = −∞)

we have almost 100% chance of voting for objects biased in the same direction ($V_{j.} > 0$ for $U_{i.} = \infty$ and $V_{j.} < 0$ for $U_{i.} = -\infty$); and almost no chance of voting for objects biased in the other direction (same goes for a very biased object); 3) for a normally biased user ($0 < U_{i.} \ll \infty$ or $-\infty \ll U_{i.} < 0$) we have more than 50% chance of voting for objects biased in the same direction ($V_{j.} > 0$ for $U_{i.} > 0$ or $V_{j.} < 0$ for $U_{i.} < 0$) and less than 50% chance for those biased in the other direction ($V_{j.} < 0$ for $U_{i.} > 0$ or $V_{j.} > 0$ for $U_{i.} < 0$).

Based on the likelihood of the proposed probability function, we can define a cost function that maximizes the likelihood of the observed R matrix by optimizing the taste vectors for users and objects. Let us assume that we have a matrix ($R_{m\times n}$ where m indicates the number of users and n the number of objects) which shows who (which user) voted for what (which objects) as binary data ($R_{i,j} = 1$ if user i voted for object j and $R_{i,j} = 0$ otherwise). Also, let us assume that we have the probability of observation (exposure) as another matrix $W_{m\times n}$, where $W_{i,j}$ shows the probability of user i having observed object j ($W_{i,j} = 1$ when $R_{i,j} = 1$ since the user has definitely observed any object s/he has voted for). Thus, we can formulate the likelihood of voting as $L = W_{i,j} \times F\left(U_{i.}.V_{j.}^{T}\right)$, which is the product of exposure probability and voting probability. Then, the likelihood of not voting is $1 - L = 1 - W_{i,j} \times F\left(U_{i.}.V_{j.}^{T}\right) = 1 - W_{i,j} + W_{i,j} \times \left(1 - F\left(U_{i.}.V_{j.}^{T}\right)\right)$. The later formulation is more insightful: when object j does not receive any vote from user i, it means that either user i observed object j but did not like it ($W_{i,j} \times \left(1 - F\left(U_{i.}.V_{j.}^{T}\right)\right)$ or user i did not observe object j ($1 - W_{i,j}$). Therefore, given the voting and observation data ($R_{m\times n}, W_{m\times n}$), the likelihood function (for each user-object pair) is:

$$\mathcal{L}\left(U_{i.}, V_{.j} \middle| R_{i,j}, W_{i,j}\right) = \begin{cases} F\left(U_{i.}.V_{j.}^{T}\right) & where\ R_{i,j} = 1 \\ 1 - W_{i,j} + W_{i,j} \times \left(1 - F\left(U_{i.}.V_{j.}^{T}\right)\right) & where\ R_{i,j} = 0 \end{cases} \qquad (1)$$

Note that if user i votes for object j, then we know that s/he observed the item ($W_{i,j} = 1$); therefore, $\mathcal{L}\left(U_{i.}, V_{.j} \middle| R_{i,j} = 1\right) = W_{i,j} \times F\left(U_{i.}.V_{j.}^{T}\right) = F\left(U_{i.}.V_{j.}^{T}\right)$, and the log-likelihood function (for each user-object pair) is:

$$\log\left(\mathcal{L}\left(U_{i.}, V_{.j} \middle| R_{i,j}, W_{i,j}\right)\right) = \begin{cases} \log\left(F\left(U_{i.}.V_{j.}^{T}\right)\right) & where\ R_{i,j} = 1 \\ \log\left(1 - W_{i,j} + W_{i,j} \times \left(1 - F\left(U_{i.}.V_{j.}^{T}\right)\right)\right) & where\ R_{i,j} = 0. \end{cases} \qquad (2)$$

Since individual's preferences are embedded in the taste vectors, the votes of different users are independent once controlling for those preferences. Therefore, the cost function (which maximizes the log-likelihood values) can be defined by adding up the negative log-likelihood value of each user-object pair, on all users ($i = 1,2, \dots m$) and all objects ($j = 1,2, \dots, n$), given voting and observation data ($R_{m\times n}$ and $W_{m\times n}$):

$$\text{Cost Function} = -1 \times \sum_{i=1}^{m} \sum_{j=1}^{n} \Big(R_{i,j} \times \log\left(F\left(U_{i.}.V_{j.}^{T}\right)\right) + \left(1 - R_{i,j}\right) \times \log\left(1 - W_{i,j} + W_{i,j} \times \left(1 - F\left(U_{i.}.V_{j.}^{T}\right)\right)\right)\Big) \qquad (3)$$

Minimizing the above cost function (equation (3)) by estimating the taste vectors, we find the optimum taste vectors for users and objects. Estimated taste vectors maximize the likelihood of observed user-object voting, based on voting data ($R_{m\times n}$) and observation data ($W_{m\times n}$).

Now, to extract and capture the effect of factor variables, we can redefine the input parameter of our probability function (i.e. $F(.)$) as $U_{i,.}V_{j,.}^{T} + \beta \times Q_{i,j}$ where $Q_{i,j}$ is the factor variable value of object j when user i voted for it (or, in the case of non-rated objects, the factor variable value of the object the last time the user saw it). For instance, in the case of capturing the effect of a movie's popularity on the ratings in our example, $Q_{1,2}$ was the average rating of Die Hard when Dan rated the movie. We estimate the value of the β coefficients using the optimization process (along with optimizing U and V). Note that multiple factor variables can be added to the probability function for other influential factors.

The proposed cost function is non-linear, smooth (has derivatives of all orders on $U, V \in \mathbb{R}$) and continuous but not convex (with infinite optimum solutions, which are the transformed versions of each other, as discussed above). It is high-dimensional, computationally expensive and its variables (in practice) need to be bounded (details provided in the appendix). Yet, this optimization problem features a special structure that simplifies our task tremendously: all local optima for this optimization reach the same payoff function, and the optimal solutions are transformations of each other. Therefore, simple gradient search methods can find a global optimal solution, from which all of the other optimal solutions are reachable. Based on these key characteristics of the problem and our studies on different optimization algorithms, we chose the limited memory BFGS (L-BFGS) [37] optimization algorithm for our case study (more details provided in the appendix). Note that other optimization algorithms may work better for other cases based on the defined cost function.

After optimizing the cost function on the taste vectors, we have to look for underlying concepts (i.e. characteristics of the objects) that resulted in such ratings. Basically, each element of the objects' taste vectors (e.g. each dimension in the 3-dimensional movie taste vector) could be an indicator of a real-world concept (e.g. one type of genre in each element). If we were lucky and the estimated taste vectors perfectly aligned with real world concepts we cared about, then the estimated value of the element for each object would show the extent of a related concept in the object (e.g. $V_{4,1}$ = 2 is the extent of horror in The Silence of The Lambs). On the other hand, the value of each element in the individual taste vector would have shown the attitude of the person toward that concept (e.g. $U_{1,1}$ = 2 represents the attitude of Dan toward liking the horror genre (compared to others in the example)). As previously discussed, the axis of optimized taste vectors generally do not align to the meaningful concepts and, as a result, guessing the underlying concept of each element is not always straightforward. Next, we propose a technique to extract the underlying concepts of the taste vectors and adjust the axis to those concepts.

To adjust the taste vectors' axis and extract the underlying concepts of the taste vectors, we have to first identify a menu of potential real world concepts/dimensions that would affect the ratings. These concepts depend on the category of objects; for instance, in the case of rating movies of different genres, the quality of the acting and that of the storyline, as well as different genres. The number of potential dimensions we consider should exceed k, the number of estimate taste dimensions, so that we can identify at least k relevant dimensions out of those. Then we need to score some of the objects on those concepts manually. The number of manually-scored objects should be significantly more than the number of concepts we are scoring on. We relate each dimension of the estimated taste vectors to one (or more) of the concepts by comparing the manual scores and the estimated taste vector values of the objects. We start with one of the estimated dimensions (in a k-dimensional opinion space) and rotate the taste vectors around it one degree at a time (other increments could be used; given the automated calculations in this step one degree rotation is easy and accurate). Then, for each dimension of the rotated taste vector (except for the rotation axes), we can compare the manual-scores with the rotated values based on a fitting metric (i.e. R^2) of a linear regression. The linear regression predicts the (rotated) taste values (as its dependent variable) using the manual-scores (as independent variables). Comparing a fitting metric of the linear regressions (on all of the dimensions and all of the rotated degrees), we choose the best fit (e.g. with the highest R^2) as the base rotation and rotate both individuals and objects' taste vectors to that angle. The respective concept(s) with a significant (or the most significant) coefficient will be assigned as the concept(s) of the fixed dimension (i.e. the dimension chosen with the highest R^2 value). Next, the fixed dimension from the previous step becomes our rotation axes for the next iteration and we re-do the rotation and the regression comparison process around it. We continue this process until we find the best angle for all of the dimensions (for a k-dimensional opinion space this takes (at least) $k-1$ iterations). Note that in more than 2-dimensional space, we may need to iterate on this process until it converges. In such cases, we can compare R^2 with its previous turn and choose the best fit as the proper angle and move to the next dimension with the second highest R^2. Additionally, we may have multiple significant coefficients in a regression, which would mean that all of those concepts are partially incorporated in the automatically estimated dimension of interest. Since the number of dimensions we chose for the factorizing process is smaller than the number of concepts we consider in manual scoring the factorization process may have combined multiple concepts in one dimension of the taste factor. At the end of this process we have identified the rotations needed to align the automatically estimated taste vectors with real world concepts, and the concepts that map into each of those dimensions (an example of the technique provided in the appendix).

Having estimated the individual tastes on different dimensions, one issue remains: how to use this method over time and stay consistent on the underlying dimensions that are estimated? Having rating data in multiple time windows (e.g. weekly ratings of the same individuals on different movies), one can use the proposed method on each time window and measure the attitude in the shape of a time series. Aside from the consistency evaluation of the method over time (i.e. test-retest measurement), we can use the time series to study any change in individuals' attitudes. Instead of using the above method on each time window to adjust the concept of taste vectors on the axis, we can use bridging stories (stories that fall in both two subsequent time windows, the taste of which are estimated in the first window and fixed as a parameter in the second) to keep the rotation matrix consistent between different time windows. In theory, by fixing (i.e. assigning the value manually instead of optimizing) one of the objects' (or the users') taste vector through the optimization process, all of the other taste vectors will be transformed in such a way as to stay consistent with the fixed taste vector. For instance, in our movie rating example, if we set the Fight Club taste vector values fixed to $V_{1.} = [2, 1, 0]$, after matrix factorization (using the optimization process), we will end up with the same values we had in V and U (for all of the movies and individuals), optimizing from any starting point. Using this property of matrix factorization, one can transfer one (or more) object(s) with the taste vector values and ratings to the next time window's optimization as fixed values and obtain time series taste vectors with a consistent rotation matrix. Then, the axis adjustment process can be performed on the whole batch of taste vectors. In our movie rating case, for instance, if the same individuals rated some other movies, to keep the rotation (and the respective underlying concepts) persistent in the matrix factorization of new ratings, we can add Fight Club ratings (i.e. 5, 3.5 and 2.5 for Dan, Simon and Matilda, respectively) to the new R matrix and set its taste vectors to estimated values from the previous optimization (i.e. $V_{1.}$ = [2, 1, 0]); as a result, the new taste vectors would keep the same underlying concepts and the same rotation.

In the next section we apply the proposed method to data available from a social news website (called Balatarin) and derive its users' attitudes on two dimensions. We find the underlying meaning of the dimensions and align the attitude values using the described method. We apply the attitude measuring process for 22 consecutive months on a weekly basis and extract the time series of users' attitudes. We then compare the proposed method with some other implicit attitude measuring methods on inter-rater and test-retest metrics using a survey.

3.4 Case Study

As a case study we use data from Balatarin, the largest social news website (examples of social news sites include Reddit and Digg) for the Persian-speaking community. On social news websites such as Balatarin, users can post links to different news items, websites, blog posts, or multimedia content (i.e. videos, pictures, sound clips). Here we call these links stories. Users are also able to read other users' stories and vote or comment on them. Since its inception in August 2006, Balatarin has gathered over 56,500 registered members, two and a half million stories, sixty-five million votes and several million comments. Balatarin promotes popular stories (those with more than a specific threshold) to its First Page (as hot stories). Users are also able to sort and rank stories based on their popularity (current number of votes), time of publication, and time of promotion in the case of stories promoted to the First Page; we refer to these ranking options as ordering pages. Each of these ordering pages is further broken down into multiple subpages, each accommodating 25 stories. Stories in Balatarin are also categorized into different subjects such as politics, sport, art, etc. Most Balatarin users are politically active and for our study we focus only on political stories. Considering the political situation of Persian-speaking countries (such as Iran), people would be wary of responding to explicit questions about their political orientation, which makes our case interesting in terms of implicit attitude measuring.

In our case study users' votes for stories are in binary form (i.e. the user either votes for a story positively or not at all)[9] and are considered as a rating. Applying the proposed method, we need a user-story voting (R) matrix, an exposure weighting (W) matrix and the stories' factor variables. Different features of stories can affect users' voting, so we extract those effects using factor variables. These factors include: the number of a story's votes, the story's promotion status (i.e. promoted to the First Page or not), the ordering page in which the user reads the story and the place of the story (i.e. which subpage and row) in the ordering page are included as factor variables. In addition to those, the publisher of the story (i.e. the user who posted the story) on Balatarin has an effect on the voting. Our observation and qualitative interaction with the Balatarin community shows that some users tend to follow and vote for stories published by particular users. We capture this effect in another factor variable, called the follower variable, which is defined as the fraction of stories posted by publisher of the story (i.e. followee) that are voted for by the user (i.e. follower). There is also the potential for strategic voting, whereby users vote for stories posted by others in a reciprocal tit-for-tat fashion. We

[9] In Balatarin users also can vote negatively for stories but, based on Balatarin's rules, negative votes should only be cast in limited conditions, such as insults or cases of violating copyright. Thus, we limited our analysis to positive votes.

extract that effect using a new factor variable (gratitude variable) as the fraction of stories published by the user and voted for by the publisher of the story.

We have access to publicly available Balatarin data on stories (including story ID, publisher ID, time of posting, story category) and votes (voted story ID, time of vote, voter ID). This is the type of social network data publicly available in many settings. We did not have access to data on where (which ordering page, and subpage) votes were cast or any other information on which stories were viewed by the users. Also, we did not have direct data on the number of votes a story had or its promotion status when the user was exposed to the story. In fact, Balatarin Inc. does not even collect any of that data. The data, along with follower variable and gratitude variable values, were extracted using a novel history reconstruction method which we developed and introduced in a previous essay and [36]. Adding the discussed factor variables to the cost function, the following is the objective function that we maximize to factorize the user-story voting matrix:

$$\begin{aligned}\text{Cost Function} = -1 \\ \times \sum_{i=1}^{m}\sum_{j=1}^{n} R_{i,j} \times \log\left(F\left(U_{i.}.V_{j.}^{T} + \beta \times Q\right)\right) + \left(1 - R_{i,j}\right) \\ \times \log\left(1 - W_{i,j} + W_{i,j} \times \left(1 - F\left(U_{i.}.V_{j.}^{T} + \beta \times Q\right)\right)\right) \qquad (4)\end{aligned}$$

where,

$$F\left(U_{i.}.V_{j.}^{T} + \beta \times Q\right) = 1/\left(1 + e^{-(U_{i.}.V_{j.}^{T} + \beta_1 \times Gratitude\ variable_{i,j} + \beta_2 \times Follower\ variable_{i,j} + \beta_3 \times Story's\ vote_{i,j} + \beta_4 \times Story's\ place_{i,j} + \beta_5 \times Story's promotion\ status_{i,j} + \beta_{6-11} \times Story's\ ordering\ page_{i,j})}\right). \qquad (5)$$

In the above formula $R_{i,j}$ represents the vote of user i on story j, where $R_{i,j} = 1$ if user i voted for story j and $R_{i,j} = 0$ otherwise. R matrix is based on the direct user-story voting data we have on Balatarin. $W_{i,j}$ is the exposure weight of user i on the story j (i.e. the probability of story j being seen by user i). When a user votes for a story we assume that all of the stories in the corresponding subpage are exposed to the user (i.e. $W_{i,k\ in\ subpage\ of\ j} = 1$ when $R_{i,j} = 1$). $U_{i.}$ and $V_{j.}$ represent the taste vector of user i and story j. For our study we consider 2-dimensional taste vectors (i.e. $U_{n \times k}$ and $V_{m \times k}$ where n and m represent the numbers of users and stories, respectively, and $k = 2$). $Story's\ vote_{i,j}$ is the number of votes story j had when it was exposed (the last time) to user i (note that users may see each story multiple times). $Story's\ ordering\ page_{i,j}$ represents six dummy variables, one for each ordering page (if user i sees story j in the l^{th} ordering page, the respective $Story's\ ordering\ page_{i,j} = 1$ and all the others are zero). $Story's\ place_{i,j}$ shows the place (subpage and row, i.e. $25 \times subpage + row$) of story j when exposed to user i (in the ordering page). $Story's\ promotion\ status_{i,j} = 1$ if story j is promoted (to the First Page) when seen by user i, and equals zero otherwise. $Gratitude\ variable_{i,j}$ and $Follower\ variable_{i,j}$ indicate the gratitude and follower effects, previously discussed, for story j and user i.

Optimizing the defined cost function (equation 4) based on the data we collected using our history reconstruction algorithm from Balatarin, we estimated the taste vectors of users and (political) stories (and factor variables (β_{1-11})) on weekly time windows from August 2006 until May 2008 on a 2-dimensional taste space. Taste vector values were estimated for all of the active users and active stories (i.e. users who cast votes and stories that received votes) in each time window and 10% of stories transferred from the previous time window to the next one as the bridge stories, to keep the rotation and underlying concepts of the axis consistent. This process provided us with 6,151 user-week taste vectors and 287,327 story taste vectors. The estimation was conducted on 2.1 GHz (16 cores) computer and took approximately 2,000 hours of computation.

To find the underlying concepts of the dimensions and adjust the corresponding axis, we developed a multivariate linear regression model that assesses the impact of manually-scored predefined characteristics of (a few selected) stories (as the independent variables) on estimated taste values of those stories (as the dependent variable). We looked at two types of concepts, those related to the content of stories and those related to their tone and style. In the case of Balatarin, based on our focus on political news and our understanding from political situation of Persian-speaking countries and also reviewing extreme stories (with extreme taste values), we defined several content characteristics related to Iran's politics. These characteristics are: 1) Opposing the reformist party vs. supporting the party 2) Supporting the actions of Basij[10], Sepah[11] and the ministry of intelligence in Iran vs. supporting students' activities, human rights, freedom of speech and opposing political imprisonment 3) Opposing George Bush, U.S., Israel, England and Saudi Arabia's actions and policies toward Iran, supporting Palestinian resistance, opposing U.S. (possible) war declarations against Iraq and Afghanistan (and Iran) vs. supporting U.S., Israel or England's actions and policies toward Iran 4) Supporting Shia Islam's ideological view and opposing Mojahedin[12] 5) Supporting Iran's nuclear program 6) Supporting (previous and current supreme leaders) Khomeini and Khamenei 7) Supporting (former president) Ahmadinejad and his government. To simplify the discussion, we categorized these characteristics as supporting or opposing the Iranian government (see Table 3-2). A second set of characteristics on which we rated the stories related to the style and preference of the story and user. These personal preferences include: 1) the user's preference for reading news rather than the personal opinions of other users and the corresponding issue for stories (i.e. if it is a news item or an opinion piece); 2) the user's preference for links containing rich media (i.e. picture, video clip or sound clip); 3) the user's preference for posts on ordinary (e.g. related to daily life) subjects over news; 4) the user's preference for informal language over formal language in Balatarin posts; 5) the user being a fan of rumors (non-confirmed news) in Balatarin posts; and 6) the user's preference for reading funny posts over serious ones. For the sake of simplicity, here we label as yellow stories those stories with personal opinion, containing rich media, with ordinary subjects, informal language, from unreliable sources and funny posts (see Table 3-3).

We asked two raters with knowledge of Balatarin and Persian politics to rate 100 stories (selected randomly from 90% most extreme taste values, i.e. those placed out of 90% level of a 2-dimensional Gaussian distribution with mean zero and covariance matrix found in the estimated story tastes) on the 13 characteristics above (ratings from -2 to 2 on each question). The two raters had inter-rater reliability of 0.80 (Krippendorff's alpha [38]) on characteristic of the content (considered as reliable [39]) and 0.32 on style and preference of the story (not very reliable). We combined the scores by averaging each on the raters.

[10] Iranian volunteer force of Islamic government loyalists
[11] Army of the guardians of the Islamic revolution in Iran
[12] An anti-government group in diaspora

Following the algorithms description, we then rotated the taste values around the center, one degree at a time, we fit the first dimension of the rotated automatically estimated taste values for the 100 sample stories and selected the rotation degree with the best fit, based on R^2 values (regression results for the best fit are provided in Table 3-2). We included only the factors related to political leanings (7 questions) as independent variables, so that we identify the rotation that gives us the axis with the most relevant content on political leanings. Given that we only have two dimensions, once one is fixed, the other is also identified, so we can conduct another regression and find out how much the second dimension relates to various stylistic preferences. The results of the regression analysis on the second dimension of taste vector over the 6 related characteristics are provided in Table 3-3.

Supporter of the Iranian government	Opposition to the Iranian government	Coefficient	p-value
Opposing reformists	Supporting reformists	-0.9840	0.00
Supporter of the actions of Basij, Sepah and the ministry of intelligence in Iran	Supporting students' activities, human rights, freedom of speech and opposing political imprisonment	-0.5815	0.65
Opposing the actions and policies of George Bush, the U.S., Israel, England and Saudi Arabia toward Iran, supporting Palestinian resistance, opposition to U.S. (possible) war declarations against Iraq and Afghanistan (and Iran)	Supporting the actions and policies of the U.S., Israel or England toward Iran	-0.5802	0.04
Supporting Shia Islam's ideological view and opposing Mojahedin	-	-1.008	0.19
Supporting Iran's nuclear program	Opposition to Iran's nuclear program	-0.8515	0.15
Supporting Khomeini and Khamenei	Opposition to Khomeini or Khamenei	-0.2833	0.57
Supporting Ahmadinejad and his government	Opposing Ahmadinejad or his government	-0.5815	0.05

Table 3-2. The regression coefficients on the first dimension: All characteristics ranged from -2 for extreme support to 2 for extreme opposition to the Iranian government.

Fan of Yellow stories	Fan of Serious stories	Coefficient	p-value
Preferring posts with personal opinions on Balatarin	-	-0.4927	0.36
Preferring posts containing rich media	-	-2.2790	0.00
Preferring posts with ordinary (e.g. related to daily life) subjects	Preferring news	1.0230	0.63
Preferring informal language in posts	Preferring formal language in posts	0.7087	0.00
Being a fan of rumors (non-confirmed news) in Balatarin posts	Preferring news from reliable sources	0.0912	0.71
Preferring funny posts on Balatarin	Preferring serious posts on Balatarin	0.0379	0.89

Table 3-3. The regression coefficients on the second dimension: Preferring posts with personal opinions and posts with rich media scored as binary (one for posts with personal opinion and zero otherwise; one for links including rich media and zero otherwise), others ranged from -2 for extreme fans of yellow stories to 2 for extreme fans of serious stories.

The results of the two regression analyses on the manually-scored stories (for the best fit based on R^2 after rotating the taste vectors (highest R^2 = 0.17 at -30°)) are shown in Table 3-2 and 3-3. While the multiple items related to support for Iranian government and preference for yellow stories are highly correlated, the regressions highlight more specific variables that better relate to the preference for votes, and thus are picked up by the automatically estimated taste vectors. Specifically, three of the concepts explain the individual ratings more than others in the first dimension (1: Opposing reformists; 2: Opposing the actions and policies of George Bush, U.S., Israel, England and Saudi Arabia toward Iran, supporting Palestinian resistance, opposing U.S. (possible) war declarations against Iraq and Afghanistan (and Iran); and 3: Supporting Ahmadinejad and his government). On the second dimension two concepts are more explanatory (1: Preferring posts containing rich media; 2: Preferring informal language in posts.).

The coefficients in Table 3-2 imply that the first element of the taste values for users and stories that support the Iranian government lies on the positive side of the first axes and for the opposition to the Iranian government the taste value (of the first dimension) is placed on the negative side of the axes. Based on the coefficients in Table 3-3, yellow stories and fans of those lie on the negative side of the second dimension, while serious ones are placed on the positive side.

3.5 Results:

After finding the best rotation angle and the most explanatory underlying concept for each axis, we rotated the taste values and aligned them to the new X and Y axes in the attitude space. Thus, we mapped the median attitude values (for those weeks that the user was active) for 6151 Balatarin users on a 2-dimensional attitude space (Figure 3-1), where each point is representative of one user's attitude value.

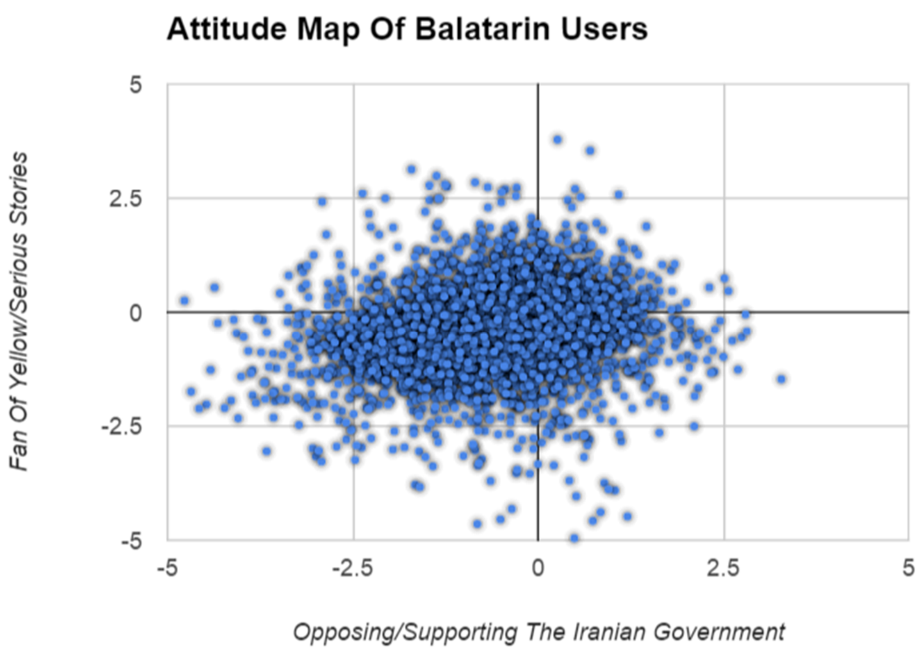


Figure 3-1. Attitude map of Balatarin users

In the figure, the X axis represents the attitude of each user (and its extent) toward opposing or supporting the Iranian government (from negative to positive, respectively). The Y axis shows the attitudes of users in terms of being fans of yellow or serious stories (the negative and positive sides of the axis, respectively). Those with (in absolute) higher values have a more extreme attitude compared to those that are close to zero (i.e. neutral users).

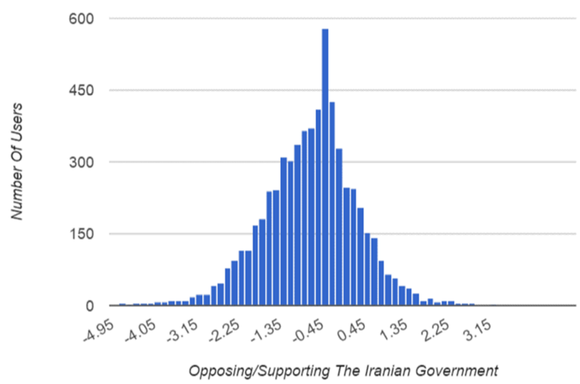


Figure 3-2-a. Distribution of Users' Attitudes Toward the Iranian Government ($\mu = -0.68, \sigma^2 = 0.94$)

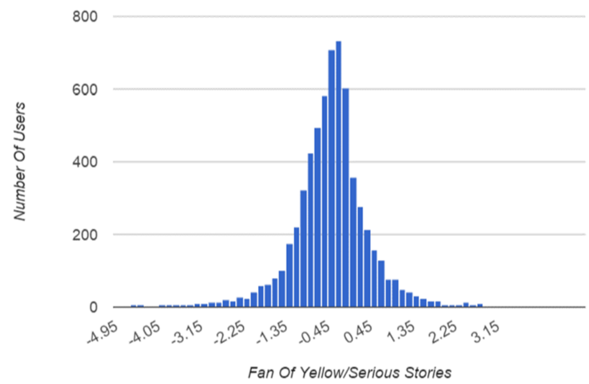


Figure 3-2-b. Distribution of Users' Attitudes Toward Yellow/Serious Stories ($\mu = -0.38, \sigma^2 = 0.57$)

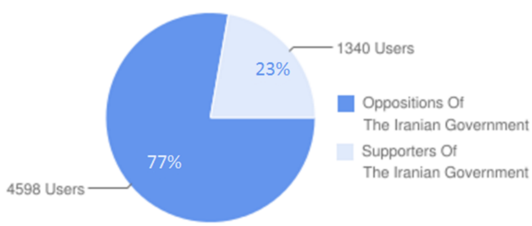


Figure 3-2-c. Pie Chart on Opposing/Supporting the Iranian Government

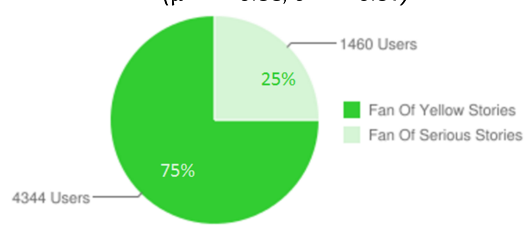


Figure 3-2-c. Pie Chart on Fan of Yellow/Serious Stories

Figure 3-2. Distribution of Users' Attitudes toward each dimension

Figure 3-2 illustrates the distribution of users' attitudes toward the Iranian government (Figure 3-2-a) and yellow/serious stories (Figure 3-2-b). The left skewness in both distributions shows that the majority of Balatarin users oppose the Iranian government and are fans of yellow stories (as the pie charts in Figures 3-2-c and 3-2-d illustrate).

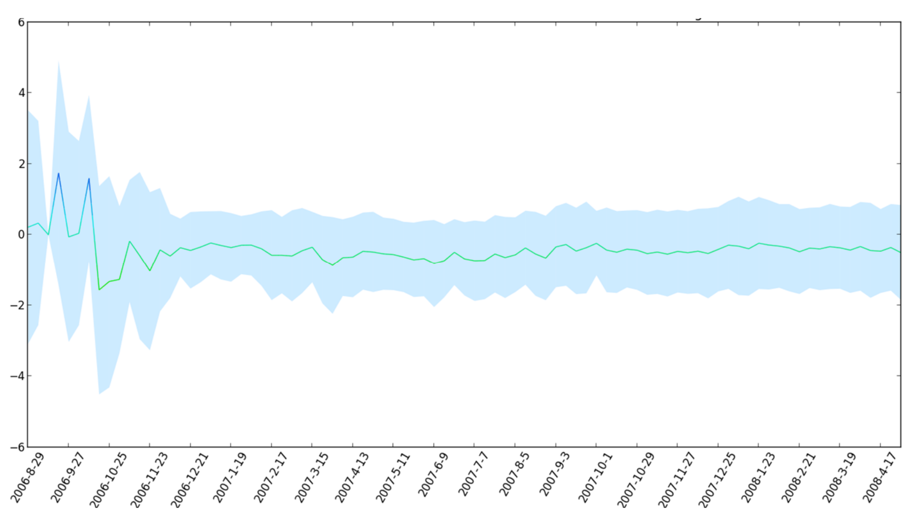


Figure 3-3-a. Mean and standard deviations of users' attitudes toward the Iranian Government over time

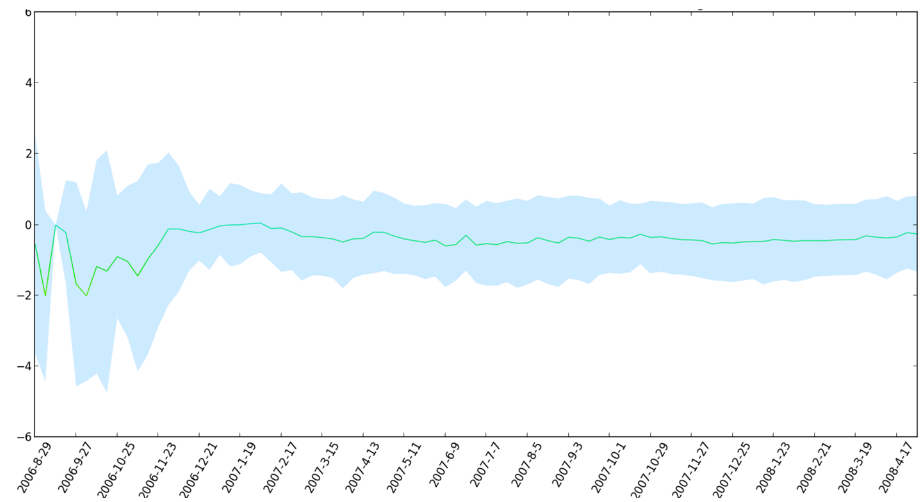


Figure 3-3-b. Mean and standard deviations of users' attitudes toward yellow posts over time

Figure 3-3. Mean and standard deviations of users' attitude toward each dimension

Figure 3-3-a and 3-3-b show the mean and standard deviations of users' attitudes toward the Iranian government and yellow/serious stories on Balatarin, respectively, over time. On both graphs the mean and standard deviations are calculated based on the attitudes measured on a weekly basis. The high deviation of the attitude values in the first month is mainly caused by the low number of users and low activities on Balatarin in its early stages. Starting from 2007, the mean and variance on both dimensions remain fairly steady (i.e. implying the stability of opinions as well as the method for measuring attitudes through multiple time windows).

To formally evaluate the reliability of method we conduct two analyses. First, calculating the correlations between estimated opinion values of the same individuals (over one and two weeks) as the test-retest measure, we get r = 0.66, r = 0.65 on one week and r = 0.55, r = 0.53 on two weeks for the first and second dimensions, respectively. Compared to the test-retest results Bosson et al. provided for implicit attitude measurement methods [40], our method (over one week) is placed right after Implicit Association Test (IAT with r = 0.69) and on top of the other six methods (Supraliminal: 0.08, Subliminal: 0.28, Stroop task: -0.05, ISES: 0.38, Initials-preference task: 0.63 and Birthday-preference task: 0.53). Compared to the test-retest results of different IAT studies provided by Egloff et al. [41], our method outperforms seven out of eight (except for Bosson et al. study) studies (ranged from 0.32 to 0.69). Next, we validate the proposed method using a survey in which we asked subjects to categorize selected stories as supporting or opposing the Iranian government and as yellow or serious stories.

Second, to evaluate the external validity of the proposed method, a small survey was conducted to assess the algorithms ability in categorizing stories on each dimension. Five human raters completed a questionnaire, in which participants were asked to categorize 20 stories selected from Balatarin in terms of supporting or opposing the Iranian government, and another set of 20 stories on being serious or yellow stories. For each set an explanation provided a definition of the dimension on which the stories were to be categorized. This explanation was based on the characteristics shown in Table 3-2 and 3-3 which we found most relevant in understanding each of the two dimensions. For the first opinion dimension stories were selected randomly from those with a taste value outside the 10-90 percentile on the first dimension and a taste value between 40%-60% percentiles on the second dimension. Therefore we had selected stories that according to the algorithm should be clearly in one of the two categories of pro/anti-government. We used a similar approach for the second dimension to sample from stories that are either yellow or serious. The surveys were designed in such a way that each story is repeated for multiple subjects to make the inter-rater consistency calculation possible as well. Accuracy of results are provided in Table 3-4 and the inter-rater measure calculated (using Cohen's kappa [42] metric) equals 0.64 on the first dimension and 0.68 on the second one (which shows substantial, though imperfect, agreement between subjects). Interestingly, the correlation among subjects was $r = 0.66$ compared to the correlation between subjects and the algorithm $r = 0.77$, which suggests the algorithm performs significantly ($p = 0.03$) better than human subjects in assessing the underlying dimensions.

	Subject 1	**Subject 2**	**Subject 3**	**Subject 4**	**Subject 5**
Accuracy on the 1st dimension	95%	100%	85%	80%	95%
Accuracy on the 2nd dimension	90%	75%	90%	80%	95%

Table 3-4. Survey accuracy results

3.6 Discussion and conclusion

The popularity of social media has rapidly increased over the past decade. Major social websites such as Facebook and Twitter collect a huge amount of data on individuals' behaviors, personal preferences, social interactions and attitudes, among many other things. This era of Big Data has enormous potential for helping psychologists and social scientists understand human behavior. Social media and Big Data can help researchers to collect behavioral information without sampling human participants and explicit data collection. This type of data collection is invisible to users and reduces self-reporting errors. In addition, it gives psychological scientists the ability to measure individuals' behavior continuously without relying on their self-input. Besides, such large data samples can achieve the statistical power that laboratory studies lack [43]. Due to the sheer size of social media databases, collecting representative samples is easier compared to traditional recruitment methods [44]. Such a large amount of data may also reveal aspects of behavior that are small in magnitude and are not observed with smaller samples [45].

In this paper we proposed a method to empirically and implicitly measure the attitudes of individuals toward different issues using their interaction data on social media. The proposed method maps the attitudes of individuals and views embedded in online content with which individuals interacted on a multidimensional attitude space using only individual-content rating data available from many social media outlets (i.e. Likes on Facebook). The method is flexible for considering various factors that could affect individuals' ratings of objects; it is scalable for handling a large amount of data and can be conducted to extract time series attitude values on an individual level. The proposed method comes with techniques for deriving the underlying concepts of each dimension in the attitude space and keeping the dimensions fixed and aligned between different time windows. We applied the method to data available from a social news website called Balatarin for more than six thousand active individuals, using more than 22 months of data in weekly time windows. We validated our model based on a test that categorized Balatarin's stories, based on the extracted taste of the content, in four different groups of attitudes by means of a survey. The results show a high correlation between the test- retests and a high value in the inter-rater metric derived from the human rater survey.

Comparing the proposed method to prior implicit attitude measuring methods, the proposed approach is not based on response-latency and is therefore free of response-latency error. Social media users are exposed to different expressions and viewpoints on different subjects in a variety of contexts. For example, Balatarin users read different news articles from a variety of outlets, along with the opinions of others on each topic. The proposed method considers all users' responses (e.g. their votes on different posts) and extracts their attitudes (toward the subject), generalizing the responses in different contexts. This generalization improves the sensitivity to context in comparison with other implicit attitude measuring methods. In addition, the method evaluates individuals' attitudes based on their responses within a time window (a weekly time window in our case study). Thus, unlike previous attitude measuring methods that collect data at one point in time and capture individuals' attitudes in that moment, the proposed method generalizes data collected over time and evaluates individuals' attitudes in that time horizon, which makes it more robust regarding the effect of prior exposures. Furthermore, interaction with various contents and users (with different opinions and beliefs) improves individuals' self-perception (toward different subjects) and adjusts the self-observation of the individual (i.e. help individual to make up his/her mind on the subject), which controls the effects of these psychological factors in the proposed attitude measuring method. The proposed method takes advantage of data provided by individuals on social media of their own volition; thus, it greatly reduces the cost of data collection (i.e. designing and implementing instruments, recruiting subjects, insuring high response rates, etc.) and scaling. By using data available on social media, the behavioral characteristics of individuals can be measured continuously without the need to persuade the subjects to complete the surveys and without reliance on their recollections. Additionally, even with implicit measurement techniques, the current attitude evaluation methods still need to be conducted in a laboratory environment, which can influence users' responses and can result in biased measurement [46]. Lastly, unlike the implicit association test (IAT), the proposed method provides a technique to derive the underlying concepts being measured. It specifies those concepts that have a significant effect on individuals' ratings and evaluates the effect of each concept empirically (using a regression model).

The method has its limitations as well. For instance, when we use data collected from social media we do not have control over what factors are discussed and thus could be identified in the data, where as in laboratory experiments we design the questions. Many exogenous factors could affect individuals' behavior, consequently affecting social media data and our experiment results, and we may not be able to observe and control for all those factors. Moreover, computational costs could become a concern if millions of users and hundreds of millions of stories are to be analyzed. Finally, many insights come from direct observation and interaction with human subjects; online data lacks such richness and may offer limited cues on the mechanisms underlying the measured trends [45].

Maybe one of the most important properties of social media data is its availability over time on an individual level, which makes time series analysis possible for each subject. In our case, attitude data over time can be derived for each social media users and form time series data for each subject. Such time series are a valuable source of data for researchers and can be used to study the changes and formation of attitudes over time. In online communities specifically, by using attitude time series analysis for each user we can study the way in which interacting with a variety of opinions gradually changes individuals' attitudes toward topics and results in different formations of communities on social media (i.e. consensus, polarization and plurality) [35, 47]. Attitude time series data on social media users can also help researchers to understand individuals' reactions when interacting with other opinions. This can help researchers to understand the cause of the rise and fall of activities on social media platforms. Such studies can be used to design better social media platforms that increase users' engagement and social media's lifecycle (as well as that of all the surrounding tools and companies), helping the economy and increasing the welfare of the society.

Appendix B

An example on the axis adjustment technique

Let's assume by factorizing matrix R in our movie rating case we get the following taste vectors (note that $U'.V'^T = U.V^T = R$):

$$U' = \begin{bmatrix} 2.04 & 0.93 & 0.46 \\ 1.39 & 1.68 & -1.40 \\ 0.45 & 2.41 & 1.20 \end{bmatrix}, V' = \begin{bmatrix} 2.16 & 0.5 & 0.25 \\ 0.86 & 1 & -1.5 \\ -0.5 & 1.73 & 0.86 \\ 1.73 & 0 & 1 \end{bmatrix}$$

To adjust the taste vectors' axis (and extract the underlying concept of each dimension), we first assume (i.e. guess) that the users rated the movies based on four different genres: horror, action, romance and mystery. Now we manually score Pride and Prejudice, Fight Club and Die Hard on the extent of each genre as below:

	Horror	**Action**	**Romance**	**Mystery**
Fight Club	5	3	0	5
Die Hard	1	5	2	2
Pride and Prejudice	0	1	5	0

Table 3-5. Manually scored movies

Holding the Y-axis (i.e. second column in U' and V') as the rotation axis, we rotate V' one degree at a time using $V'' = V'.R_y$ where R_y is the rotation matrix around the Y-axis defined as below:

$$R_y = \begin{bmatrix} \cos(\theta) & 0 & \sin(\theta) \\ 0 & 1 & 0 \\ -\sin(\theta) & 0 & \cos(\theta) \end{bmatrix}$$

We then calculate the correlation between (each of) the manually scored genres (e.g. [5,1,0] for the horror genre) with the first and the third rotated taste vector (i.e. first and third column of the $V'' = V'.R_y$ matrix). Note that here in this example with only three rated movies we don't have enough data points to run a regression, so we use a simple correlation as the goodness-of-fit measure. An example (only for three rotation angles) of the correlation results provided below:

Rotation angel	**-π/6**		**-π/3**		**-π/4**	
	1st taste vector	**3rd taste vector**	**1st taste vector**	**3rd taste vector**	**1st taste vector**	**3rd taste vector**
Horror	**0.98**	-0.38	0.64	-0.71	0.86	-0.56
Action	0	-0.97	-0.62	-0.81	-0.32	-0.91
Romance	-0.80	0.74	-0.25	0.94	-0.56	0.86
Mystery	0.91	-0.57	0.46	-0.85	0.74	-0.73

Table 3-6. Correlation matrix example rotating around Y-axis

As the result shows, the highest amount of correlation is between the horror genre and the first taste vector at $-\pi/6$ rotation angle. Therefore, we assign the horror genre to the first taste vector, set the rotation around the Y-axis at $-\pi/6$ degree for V'' and hold the X-axis as the rotation axis for the next round. Now rotating V'' around the X-axis one degree at a time using $V''' = V''.R_x$ where R_x is the rotation matrix around the X-axis defined as below:

$$R_x = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos(\theta) & -\sin(\theta) \\ 0 & \sin(\theta) & \cos(\theta) \end{bmatrix}$$

we calculate the correlation between the manually scored genres with the second and the third rotated taste vector. This time we get the following results for the correlations:

Rotation angel	-π/6		-π/3		-π/4	
	2nd taste vector	3rd taste vector	2nd taste vector	3rd taste vector	2nd taste vector	3rd taste vector
Horror	-0.45	-0.51	0.18	-0.65	-0.01	-0.57
Action	0.78	-0.93	**1.00**	-0.86	0.97	-0.91
Romance	0.02	0.83	-0.59	0.91	-0.41	0.87
Mystery	-0.25	-0.68	0.39	-0.80	0.19	-0.74

Table 3-7. Correlation matrix example rotating around X-axis

Based on the results, the second taste vector correlates perfectly with the action genre at $-\pi/3$ rotation angle so we assign action genre to the second taste vector and set the rotation around the X-axis at $-\pi/3$ degree for V'''. To find the underlying concept of the third vector we can simply compare the correlation between the third column of V''' and the scores of the romance and mystery genre (which are 0.91 and -0.80 respectively) and assign the genre with the highest correlation to the third taste vector. Note that in this example we only use three manually scored movies and consequently the correlation values are high for (almost) all of the genre-vector pairs. In real world problems we have to score more objects manually (e.g. more than 30 for running regressions) and most likely we will end up with (more) significant differences between goodness-of-fit measures.

Choosing the optimization algorithm

The problem of high dimensionality in collaborative filtering (and other recommendation systems) has been discussed in the literature [48, 49]. The case of opinion mapping is not an exception and, as a result, the ability to handle high dimensional problems is an important factor for choosing the proper optimization algorithm. The proposed cost function is non-linear, smooth (has derivatives of all orders on $U, V \in \mathbb{R}$) and continuous but not convex; therefore, local optimization algorithms could fall in local optimums. Besides, the proposed cost function (and its gradient) is computationally expensive (due to the existence of the logarithm, exponential and dot product calculations for each pair of user-object) and as a result the optimization algorithm needs to converge with minimum function calls. Also, due to precision issues, the cost function variables need to be bounded; therefore, supporting constraints is another key factor in selecting the optimization algorithm.

In the case of matrix factorization problems, the gradient descent algorithm is perhaps the simplest technique to implement; however, it requires a high number of function calls and the convergence could be very slow [50]. The conjugate gradient method is faster in convergence for such problems but it is more complicated to implement and the convergence of the conjugate gradient (and other gradient-based methods) is very sensitive to the choice of step size, which can be an inconvenience in the case of large applications [50]. Newton's optimization method, on the other hand, requires the Hessian matrix, which is computationally expensive in our case. However, quasi-Newton methods, which use the approximation of the Hessian matrix, are computationally cheaper. BFGS, as one such quasi-Newton method, supports constraints on the variables. Its limited-memory version [37], which sits between BFGS and the conjugate gradient, computes and keeps the low-rank version of the (approximate) Hessian instead and, as a result, needs far less memory and computation time in high-dimensional problems. Limited memory BFGS (L-BFGS)'s good performance on non-linear and non-convex problems is also proven. Therefore, based on its properties, the L-BFGS algorithm is a good choice to optimize the cost function of opinion estimation here.

Limited precision and variables' boundary

In theory, U and V vectors could be any real number ($-\infty < U < \infty, -\infty < V < \infty, U, V \in \mathbb{R}$), and the range of optimization should be open to all real numbers. However, due to precision issues, we need to put some constraints on these parameter values. A common and important issue in our case is the precision of calculations in the scripting language. In order to solve the precision problem, we need to set a range for U and V that keeps the $|U_{i.}.V_{j.}{}^{T}|$ small enough to avoid the overflow problem. Thus, we can set a symmetric boundary for k dimensional U and V such that $-\varepsilon < U_d, V_d < \varepsilon$ for all $d = 1,2,\dots,k$ and $B_L < U_{i.}.V_{j.}{}^{T} < B_U$. Here, B_L is the lower bound in which $\frac{1}{1+e^{-U_{i.}.V_{j.}{}^{T}}}$ equals the smallest positive (non-zero) value supported by the scripting language (i.e. Python) and $B_U$is the upper bound in which $e^{-U_{i.}.V_{j.}{}^{T}}$ (exponential function in the scripting language (i.e. exp() in Python)) returns its smallest positive (non-zero) value. Knowing the largest and the smallest positive (non-zero) numbers that the scripting language supports (i.e. 1.79e+308 and 2.22e-308, respectively in Python), we can calculate such boundaries. In our case (assuming Python to be the scripting language), for a two-dimensional opinion vector ($k = 2$) boundaries are calculated as $-18.82 < U_d, V_d < 18.82$.

References:

1. Coviello, L., et al., *Detecting emotional contagion in massive social networks.* PloS one, 2014. **9**(3): p. e90315.
2. Steers, M.-L.N., R.E. Wickham, and L.K. Acitelli, *Seeing Everyone Else's Highlight Reels: How Facebook Usage is Linked to Depressive Symptoms.* Journal of Social and Clinical Psychology, 2014(33): p. 701-731.
3. Feinstein, B.A., et al., *Negative social comparison on Facebook and depressive symptoms: Rumination as a mechanism.* Psychology of Popular Media Culture, 2013. **2**(3): p. 161.
4. Pempek, T.A., Y.A. Yermolayeva, and S.L. Calvert, *College students' social networking experiences on Facebook.* Journal of Applied Developmental Psychology, 2009. **30**(3): p. 227-238.
5. Bond, R.M., et al., *A 61-million-person experiment in social influence and political mobilization.* Nature, 2012. **489**(7415): p. 295-298.
6. Blum, M. and J. Naylor, *Training and learning.* Industrial psychology: Its theoretical and social foundations. Harper & Row, New York, 1968: p. 237-275.
7. Sheppard, B.H., J. Hartwick, and P.R. Warshaw, *The theory of reasoned action: A meta-analysis of past research with recommendations for modifications and future research.* Journal of consumer research, 1988: p. 325-343.
8. Fishbein, M., *A theory of reasoned action: some applications and implications.* 1979.
9. Breckler, S.J., *Empirical validation of affect, behavior, and cognition as distinct components of attitude.* Journal of personality and social psychology, 1984. **47**(6): p. 1191.
10. Wittenbrink, B., C.M. Judd, and B. Park, *Evidence for racial prejudice at the implicit level and its relationship with questionnaire measures.* Journal of personality and social psychology, 1997. **72**(2): p. 262.
11. Greenwald, A.G., D.E. McGhee, and J.L. Schwartz, *Measuring individual differences in implicit cognition: the implicit association test.* Journal of personality and social psychology, 1998. **74**(6): p. 1464.
12. Fazio, R.H. and M.A. Olson, *Implicit measures in social cognition research: Their meaning and use.* Annual review of psychology, 2003. **54**(1): p. 297-327.
13. De Houwer, J., *What are implicit measures and why are we using them.* The handbook of implicit cognition and addiction, 2006: p. 11-28.
14. Cunningham, W.A., K.J. Preacher, and M.R. Banaji, *Implicit attitude measures: Consistency, stability, and convergent validity.* Psychological science, 2001. **12**(2): p. 163-170.
15. Lemm, K.M., *Personal and social motivation to respond without prejudice: Implications for implicit and explicit attitude and behavior*. 2001, ProQuest Information & Learning.
16. Asendorpf, J.B., R. Banse, and D. Mücke, *Double dissociation between implicit and explicit personality self-concept: the case of shy behavior.* Journal of personality and social psychology, 2002. **83**(2): p. 380.
17. Rudman, L.A. and M.R. Lee, *Implicit and explicit consequences of exposure to violent and misogynous rap music.* Group Processes & Intergroup Relations, 2002. **5**(2): p. 133-150.
18. Wittenbrink, B., C.M. Judd, and B. Park, *Spontaneous prejudice in context: variability in automatically activated attitudes.* Journal of personality and social psychology, 2001. **81**(5): p. 815.
19. van Ravenzwaaij, D., H.L. van der Maas, and E.-J. Wagenmakers, *Does the name-race implicit association test measure racial prejudice?* Experimental psychology, 2015.
20. Klauer, K.C., et al., *Process components of the Implicit Association Test: a diffusion-model analysis.* Journal of personality and social psychology, 2007. **93**(3): p. 353.

21. Kaufman, S.B., *Does the implicit association test (IAT) really measure racial prejudice? Probably not.* Psychol Today, 2011.
22. Kosinski, M., D. Stillwell, and T. Graepel, *Private traits and attributes are predictable from digital records of human behavior.* Proceedings of the National Academy of Sciences, 2013. **110**(15): p. 5802-5805.
23. Rentfrow, P.J., et al., *Divided we stand: Three psychological regions of the United States and their political, economic, social, and health correlates.* 2013.
24. Youyou, W., M. Kosinski, and D. Stillwell, *Computer-based personality judgments are more accurate than those made by humans.* Proceedings of the National Academy of Sciences, 2015. **112**(4): p. 1036-1040.
25. Faridani, S., et al. *Opinion space: a scalable tool for browsing online comments*. in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. 2010. ACM.
26. Trechsel, A.H. and P. Mair, *When parties (also) position themselves: An introduction to the EU Profiler.* Journal of Information Technology & Politics, 2011. **8**(1): p. 1-20.
27. Kruskal, J.B., *Multidimensional scaling by optimizing goodness of fit to a nonmetric hypothesis.* Psychometrika, 1964. **29**(1): p. 1-27.
28. Hauser, J.R. and F.S. Koppelman, *Alternative perceptual mapping techniques: Relative accuracy and usefulness.* Journal of Marketing Research, 1979: p. 495-506.
29. Poole, K.T. and H. Rosenthal, *A spatial model for legislative roll call analysis.* American Journal of Political Science, 1985: p. 357-384.
30. Barberá, P., et al., *Tweeting From Left to Right Is Online Political Communication More Than an Echo Chamber?* Psychological science, 2015: p. 0956797615594620.
31. Bond, R. and S. Messing, *Quantifying Social Media's Political Space: Estimating Ideology from Publicly Revealed Preferences on Facebook.* American Political Science Review, 2015. **109**(01): p. 62-78.
32. Koren, Y., R. Bell, and C. Volinsky, *Matrix factorization techniques for recommender systems.* Computer, 2009(8): p. 30-37.
33. Wang, J., *Encyclopedia of data warehousing and mining*. 2006: IGI Global.
34. Muchnik, L., S. Aral, and S.J. Taylor, *Social influence bias: A randomized experiment.* Science, 2013. **341**(6146): p. 647-651.
35. Rad, A.A., H. Rahmandad, and M. Yahyanejad, *Estimating the Dynamics of Individual Opinions in Online Communities.*
36. Rad, A.A. and H. Rahmandad, *Reconstructing online behaviors by effort minimization*, in *Social Computing, Behavioral-Cultural Modeling and Prediction*. 2013, Springer. p. 75-82.
37. Liu, D.C. and J. Nocedal, *On the limited memory BFGS method for large scale optimization.* Mathematical programming, 1989. **45**(1-3): p. 503-528.
38. Hayes, A.F. and K. Krippendorff, *Answering the call for a standard reliability measure for coding data.* Communication methods and measures, 2007. **1**(1): p. 77-89.
39. Krippendorff, K., *Reliability in content analysis.* Human Communication Research, 2004. **30**(3): p. 411-433.
40. Bosson, J.K., W.B. Swann Jr, and J.W. Pennebaker, *Stalking the perfect measure of implicit self-esteem: The blind men and the elephant revisited?* Journal of personality and social psychology, 2000. **79**(4): p. 631.
41. Egloff, B., A. Schwerdtfeger, and S.C. Schmukle, *Temporal stability of the implicit association test-anxiety.* Journal of Personality Assessment, 2005. **84**(1): p. 82-88.
42. Cohen, J., *Weighted kappa: Nominal scale agreement provision for scaled disagreement or partial credit.* Psychological bulletin, 1968. **70**(4): p. 213.

43. Jaffe, E. *What Big Data Means For Psychological Science*. 2014; Available from: http://www.psychologicalscience.org/index.php/publications/observer/2014/july-august-14/what-big-data-means-for-psychological-science.html.
44. King, D.B., N. O'Rourke, and A. DeLongis, *Social media recruitment and online data collection: A beginner's guide and best practices for accessing low-prevalence and hard-to-reach populations.* Canadian Psychology/Psychologie canadienne, 2014. **55**(4): p. 240.
45. Gomez-Marin, A., et al., *Big behavioral data: psychology, ethology and the foundations of neuroscience.* Nature neuroscience, 2014. **17**(11): p. 1455-1462.
46. Wallaert, M., A. Ward, and T. Mann, *Explicit control of implicit responses: Simple directives can alter IAT performance.* Social psychology, 2010. **41**(3): p. 152.
47. Rahmandad, H. and M. Mahdian. *Modeling polarization dynamics in online communities*. in *Proceedings of the 29th International Conference of the System Dynamics Society*. 2011. Citeseer.
48. Billsus, D. and M.J. Pazzani. *Learning Collaborative Information Filters*. in *ICML*. 1998.
49. Sarwar, B., et al., *Application of dimensionality reduction in recommender system-a case study*. 2000, DTIC Document.
50. Lee, D.L., *Improving the capacity of complex-valued neural networks with a modified gradient descent learning rule.* Neural Networks, IEEE Transactions on, 2001. **12**(2): p. 439-443.

4. Chapter 4 - Social Media and User Activity: An Opinion-Based Study

4.1 Abstract

An increasing fraction of social communications is conducted online, where physical constraints no longer structure interactions. This has significantly widened the circle of people with whom one can interact and has increased exposure to diverse opinions. Yet individuals may act and respond differently when faced with opinions far removed from their own, and in an online community such actions could activate important mechanisms in the system that form the future of the outlet. Studying such mechanisms can help us understand the social behaviors of communities in general and individuals in particular. It can also assist social media outlets with their platform design. We propose models that capture the changes in individuals' activities in social media caused by interacting with a variety of opinions. Estimating the parameters of the models using data available from a social news website (Balatarin) as a case study, we extracted mechanisms affecting the communities on this platform. We studied the effect of these mechanisms on the future formation and the lifecycle of the platform using an agent-based simulation model. Having examined the effect of biased communities on the social media, the results imply that individuals increase their online activity as a result of interacting with contents closely aligned to their own opinion.

Keywords: Social media, User activity, Opinion measuring, Agent based simulation

4.2 Introduction:

Nowadays, social media outlets form an important part of our society. As of 2015 Facebook (with more than 1.44 billion monthly active users) has more active users than China's population. Added to that, over one billion active users on YouTube, 302 million on Twitter, 300 million on Instagram and 190 million on LinkedIn show the size of the outlets we are dealing with today. Although with such huge numbers of customers the future of these outlets seems promising, most of them rose from the ashes of ancestral social media forums equally popular in their own time. Facebook became popular after the fall of MySpace, and YouTube started after shareyourworld.com closed. In the past decade many other social media platforms have bloomed, grown and then fallen. Researchers have studied different aspects of these rises and falls, yet the effect of the variety of opinion on social media life cycle has remained almost untouched.

In one project, Cannarella and Spechler developed a SIR (susceptible, infected and recovered) model for explaining MySpace users' activity trends and predicted the future fall of Facebook [1]. Torkjazi et al. studied the activity of MySpace users and discussed the different factors behind users' departure from social networks [2]. They explained that social networks are vulnerable to new fashions and that it is important to innovate and add new features to the platform to keep the users interested in a social media platform. They also discussed the importance of linking users to other likeminded individuals when the platform grows, but did not provide further explanation or details on this matter. Wu et al. [3] studied the relationship between the user's arrival and departure on social networks respecting network topology and discussed the effect of active friends in a social network on its users' departure rates. Giliette, in an interesting article in Bloomberg BusinessWeek [4], explained some of the technical, managerial, financial and social issues that could lead to social media outlet failure. He explained how a lack of technology (e.g. spam filtering, in the case of MySpace), poor choice of advertisement services (e.g. inappropriate advertisements in MySpace) and bad publicity (e.g. a court case about inappropriate contents for children on MySpace) could create bad user experiences and push a social media platform to the edge of failure. Garcia et al. in another empirical research, studied the effect of resilience (i.e. the ability of a community to withstand changes) on the fall of social networks. They studied the effect of events such as friends leaving a social network on users' departures, based on the network topology [5]. Despite all of the research that has been done on changes in social media users' activities, the effect of exposure to a variety of opinions is a fairly new topic. In this essay we study this aspect of social media activity and produce evidence regarding the effect of (consumed) contents' opinions on individuals' activities and, consequently, the rise and fall of social media.

Social media platforms and especially online social networks (OSNs) began a new era for the internet, in which ordinary users generate content and consume content provided by others. In such an environment, user-generated content is the fundamental building block (of OSNs and social media outlets) and consequently the amount (and trend) of user activity determines the success and failure of platforms. It is natural to assume that users produce content (i.e. write stories, share news, etc.) based on their own viewpoints and opinions on a variety of topics, and content generated by one user may be repulsive from the perspective of others, which could affect their actions and activity in the social media. Individuals may act and respond differently when faced with opinions much different from their own, and in an online community such actions, from more activity to leaving the community, could determine the future of the platform. The direction of those effects are not clear ex ante. For example, after reading an article in favor of republicans in the U.S., democrat users of Twitter may post more contents (i.e. news, articles, pictures, videos, etc.) on Twitter promoting the democrats' point of view. On the other hand, one may argue that democrats (on average) may actually visit Twitter less frequently after reading contents opposing their viewpoints. Such mechanisms could affect Twitter to the point that it loses some of its users. Studying such effects and mechanisms can help us understand the social behaviors of communities (e.g. republicans and democrats on Twitter in the example) in general and individuals in particular, and can also assist social media outlets in policy making (by predicting their future situation) for future increases or decreases in the numbers of users. In this research we propose a method to evaluate and capture such mechanisms. We used opinion data extracted using a novel opinion mapping method (introduced in the previous essay) and estimate regression models that capture changes in individuals' activities on social media (i.e. posting, commenting and revisiting rates) caused by interacting with a variety of opinions. We applied the proposed method to a case study (a social news website called Balatarin) and mathematically evaluated the models. Extracting the mechanisms affecting the communities in the platform under study, we were able to simulate and predict the future formation of those communities. Conducting sensitivity analysis on communities' bias and extremeness of opinion, we extracted the effect of each parameter on the future formation of the outlet. Finally, based on our analysis, we propose different policies that can help social media outlets increase their popularity and extend their active life.

4.3 Methodology:

We developed a novel method to measure individuals' opinions on different issues based on user-online object (i.e. content, story, news, video, etc.) interaction data. Further details on the opinion measurement method are provided in the previous essay, but in short we assume that there is a higher probability of an individual voting for a story (or rating it more highly) that represents an opinion supporting that of the individual (on the topic) compared to contents that opposes his/her opinion.

Building on that idea, we extracted the opinions of users and content on different topics (based on the rating data) by factorizing the user-object voting matrix into two (user and object) opinion vectors. For instance, in a movie rating website, if we put the data on user-movie ratings in the format of a matrix R (where $R_{i,j}$ represents the rating user i assigned to movie j), using factorization methods we can factorize R into taste (i.e. opinion on genres and other factors that affected users' rating of the movies) vectors of users and movies $R \sim U.V^T$ in such a way that U and V represent the tastes of users and movies, respectively (e.g. $U_{i.}$ could be representative of user i's interest in the action genre and then $V_{j.}$ shows the extent of action in movie j, and $R_{i,j}$ as a result is the rating user i assigned to movie j considering the action genre).

In our case study we use data from Balatarin, a well-known Persian social news website. Balatarin, as one of the first Persian-speaking social media platforms (and considering that its servers are placed outside of Iranian borders) became an environment in which users could freely discuss the Iranian government and its policies. For example, in 2009 during the Iranian protests after the presidential election, Balatarin was used as a platform for coordinating the protests. Consequently, a large portion of Balatarin's contents are today concentrated on Iran's government and its political issues. Balatarin users can post stories (i.e. links containing news, videos, pictures, etc., which are here considered online objects), comment and vote for each other's stories. Stories recently published on Balatarin can be viewed on an ordering page (called recently published stories ordering page) in which stories are sorted according to their time of publication (as default). The lifecycle of the stories on the recently published stories ordering page is one day, after which stories are removed from that page. Recently published stories receiving a sufficient number of votes are promoted to the first page of Balatarin (another ordering page), where stories are sorted according to the promotion time. The lifecycle of stories on the first page is five days, after which promoted stories are removed from the first page. Users of Balatarin can label their stories with a particular category (i.e. political, social, sport, etc.) but in this study we specifically focus on the political contents published on the website and limit our analysis to those.

To map the opinions, we treat data on user-story votes as an interaction matrix (R) and factorize it into opinion vectors of users and stories (U and V respectively, for user and story). Details of this algorithm are discussed in the previous essay, and summarized here. We define a utility function in which the user gains more utility reading stories which convey an opinion close to his/her own and would vote for those with higher probability. We developed a cost function (for factorization process) that estimates opinion vectors of users and stories to factorize the interaction matrix. A simplified version of the factorization cost function in our case can be defined as:

$$Cost = -1 \times \sum_{i=1}^{n} \sum_{j=1}^{m} R_{i,j} \times Sigmoid\left(U_{i.}.V_{j.}^{\ T}\right) + \left(1 - R_{i,j}\right) \times \left(1 - Sigmoid\left(U_{i.}.V_{j.}^{\ T}\right)\right) \qquad (1)$$

where $R_{i,j} = 1$ if user i votes for story j and $R_{i,j} = 0$ otherwise. $U_{i.}$ and $V_{j.}$ are the opinion vectors for user i and story j, respectively, and $Sigmoid(x) = 1/(1 + e^{-x})$. The user-story interaction matrix (R) is given and extracted from Balatarin's data and here we minimize the cost function by optimizing the opinion vectors (i.e. U and V for each user and each story). Votes on Balatarin are binary and optimizing the cost function will estimate $U_{i.}$ and $V_{j.}$ in such a way that $Sigmoid\left(U_{i.}.V_{j.}^{\ T}\right)$ (i.e. the probability of user i voting for story j due to the closeness of the user's opinion to that of the story) increases for voted user-story pairs ($R_{i,j} = 1$) and ($1 - Sigmoid(U_{i.}.V_{j.}^{\ T})$) (i.e. the probability of user i not voting for story j due to the distance between the user's opinion and that of the story) increases for the other cases ($R_{i,j}$ = 0). Thus, in a one-dimensional space opinion vectors, users and stories biased toward the same direction (i.e. $U_{i.} > 0$ and $V_{j.} > 0$ or $U_{i.} < 0$ and $V_{j.} < 0$) have more than 50% probability of voting (i.e. user i votes for story j with a probability higher than 50%) and those biased toward different directions (i.e. $U_{i.} > 0$ and $V_{j.} < 0$ or $U_{i.} < 0$ and $V_{j.} > 0$) have less than 50% probability of voting.

In our case study each opinion vector contains two opinion dimensions (U_1 and U_2 for users and V_1 and V_2 for stories) and one fixed effect (U_3 for users and V_3 for stories). The fixed effect in the user's opinion vector represents the user's interest in voting in general compared to others (i.e. there is a higher value for those who vote more frequently) and for the story's opinion vector the fixed effect shows the attractiveness of the story (i.e. there is a higher value for stories that attract more votes regardless of the opinion). Using a (more complex) cost function derived from (1) (which controls for several other factors that could affect voting such as the previous votes of a story) we estimated the opinion values of users and stories on Balatarin in weekly time windows (more details have been provided in the previous essay). By studying stories' estimated opinion values, we found that in the 2-dimensional opinion vectors we defined, the first dimension (i.e. U_1 for users and V_1 for stories) represents support for, or opposition to, the Iranian government (on positive and negative sides of the axis, respectively). The second dimension (i.e. U_2 for users and V_2 for stories) represents a mixture of other properties of stories that affect the voting of Balatarin users, in which having rich media (i.e. video, picture and sound), expressing a personal opinion or non-confirmed news, having an ordinary (related to daily life) subject (i.e. not news) and having funny contents are the most explanatory features in this dimension. Here, for simplicity we refer to stories containing rich media, personal opinions, non-confirmed news, with ordinary or funny contents as yellow stories. We refer to the other group of stories (those that do not contain rich media or personal opinion and contain confirmed news and serious contents) as serious stories. The opinion values for yellow stories (and for fans of those stories) lie on the negative side of the second opinion dimension, while the opinion values for serious stories (and for fans of those) lie on the positive side (more details are provided in the second essay).

To study the effect of interacting with different opinions on individuals' activities on social media, we developed regression models that measure the effect of reading stories on Balatarin (based on the story's and user's opinions regarding the Iranian government and being a fan of yellow/serious stories) on its users' activities. In the regression models, as individuals' activities on social media, we focused on posting stories on Balatarin, revisiting the website and commenting on the stories. As mentioned, we estimated users' and stories' opinion values in weekly time windows.

To study the effect of interacting with different opinions on users' activities we measured the change in the activity (i.e. the number of stories posted by the user, the number of times the user went online (i.e. revisiting) and the number of comments s/he posted on Balatarin) in two separate weeks 23 weeks apart, and estimated the effect of opinions on that change based on the cumulative utility (defined below) users had gained from reading stories within those 23 weeks.[13] We move the time window one week at a time for each user and recorded the marginal utility (as an independent variable) and the change in the activity (as a dependent variable value) for each set to fit the regression model. For

[13] The number of "in between weeks" was optimized using F-test on the first regression model (posting rate) and we chose the number of in between weeks providing the most explanatory model. Note that since the number of variables (users and time-fixed effects) and the number of datasets change in models with different between weeks, we cannot use R^2, AIC or log-likelihood value for this comparison.

instance, we calculated the change in the number of published comments between the first week and the 25th week (i.e. $user\ i\ number\ of\ comments_{at\ week\ 25} - user\ i\ number\ of\ comments_{at\ week\ 1}$) and measured the sum of utility gained in weeks 2 to 24 from reading stories. Then we moved the time window one week forward and calculated the change between 2nd and 26th week and measured the utility in weeks 3 to 25 and so forth. We continued collecting the data in such a manner for all of the active weeks of all of the active users.

In each of the regression models we set a random effect for the user's baseline activity change, to measure the average change in the user's activity based on factors independent of what s/he reads. We also controlled for the number of weeks that had passed since the user had joined the website, to assess the effect of being a newcomer vs. old-timer in the community on users' activities.

We categorized each user-story pair into one of the 16 different groups based on the sign of the opinion of the story and user with regard to the Iranian government and their being yellow or serious. For instance, if we assume user i, who is a supporter of the government ($U_{i,1} > 0$) and is a fan of yellow stories ($U_{i,2} < 0$), read story j which opposes the Iranian government ($V_{j,1} < 0$) and is yellow ($V_{j,2} < 0$), then user i - story j pair lies in the group that represents $U_1 > 0, V_1 < 0, V_2 < 0, U_2 < 0$ (i.e. 8th group in Table 4-1). We measured the marginal utility the user gained from reading each story using $sigmoid(U_1 \times V_1 + U_2 \times V_2 + V_3) - 0.5$ (where U_1 and U_2 are the user's opinion values and V_1 and V_2 are the story's opinion values and V_3 is the attractiveness of the story, i.e. the story's fixed effect) and calculated cumulative gained utility on each of the groupings as one independent variable in regression models. This leads to a total of 16 coefficients that capture how various pairs of user-story types may have a different type of impact on the user's activity. Here the marginal utility gained by the user from reading a story is the result of comparing (i.e. subtracting) the total utility user gains from reading the story ($sigmoid(U_1 \times V_1 + U_2 \times V_2 + V_3)$) with the utility s/he gained from reading a completely neutral story (i.e. $sigmoid(0) = 0.5$ meaning the story has no preference (bias) toward any of the groups (supporter/opponent, yellow/serious) in either dimension, and the story's attractiveness is zero (i.e. the story is neither attractive nor unattractive in obtaining votes)). Thus, for example, when the user reads an adverse story (e.g. a serious supporter reads a yellow opposition story with zero attractiveness), the result of the subtraction is negative, indicating that the user does not gain his/her expected utility by reading that story (in a sense, the user gains less utility by reading that story compared to when s/he reads a story that is not about the government at all (i.e. a neutral story)). In other words, considering the time and energy the user spends in reading the story, s/he gains less utility than his/her minimum expectation. Thus, here the marginal utility implies the difference between gained and expected utility (which could be either negative or positive). Thus, for example, the regression model for revisiting is defined as:

$$\begin{aligned} &user\ i\ number\ of\ revisits_{at\ week\ t+25} - user\ i\ number\ of\ revisits_{at\ week\ t} \\ &\quad = C \\ &\quad + \beta_{l=1\ to\ 16} \sum_{k=t+1}^{t+24} \sum_{\substack{j \in\ stories\ that \\ user\ i\ read \\ in\ week\ k\ and \\ belongs\ to \\ opinion\ group\ l}} \left(sigmoid\left(U_{i,1} \times V_{j,1} + U_{i,2} \times V_{j,2} + V_{j,3}\right) - 0.5\right) \\ &\quad + Posts_{\substack{in\ weeks \\ between \\ t+1\ to\ t+24}} + Comments_{\substack{in\ weeks \\ between \\ t+1\ to\ t+24}} + \alpha_i + T_i \qquad (2) \end{aligned}$$

where $user\ i\ number\ of\ revisits_{at\ week\ t+25} - user\ i\ number\ of\ revisits_{at\ week\ t}$ represents the change in the number of times user i revisited the website between weeks t and $t + 25$, C is the intercept of the model, α_i is the user random effect for user i and T_i is the time random effect (i.e. for the number of weeks that have passed since user i joined the website). $Posts_{\substack{in\ weeks \\ between \\ t+1\ to\ t+24}}$ and $Comments_{\substack{in\ weeks \\ between \\ t+1\ to\ t+24}}$ are the number of stories and comments user i published in the weeks between $t + 1$ and $t + 24$. We extracted the effect of opinions by summing up the marginal utility (i.e. $sigmoid\left(U_{i,1} \times V_{j,1} + U_{i,2} \times V_{j,2} + V_{j,3}\right) - 0.5$) the user gained by reading stories in each group, where β_l coefficient captures the effect of reading stories in group l on the user's change in activity (we have $l = 1\ to\ 16$ one for each opinion group). We used the same model for the change in the number of posted stories and published comments; we removed $Posts_{\substack{in\ weeks \\ between \\ t+1\ to\ t+24}}$ and $Comments_{\substack{in\ weeks \\ between \\ t+1\ to\ t+24}}$ for the posting and commenting models respectively.

Data on posting stories and comments are directly available on Balatarin's data sets. However, to extract the data on the stories each user read and revisiting data we used a novel history reconstruction method introduced in the first essay. In short, we recreated the history of each action (i.e. voting, posting, commenting, getting online, etc.) of Balatarin users based on a heuristic effort minimization model and the ranking algorithm of Balatarin. Thus, we can recreate a snapshot of the website (i.e. stories in each sub-page and ordering page, promotion status of stories, etc.) at any time a user has voted for a story and estimate the stories s/he read based on his/her location (i.e. ordering page, sub-page and place in the sub-page where the user voted for the story) on the website. More details on this method are provided elsewhere [6]. Using the collected data from Balatarin we estimated all of the coefficients in the defined regression models; in the next section we discuss the factors that predict users' activities on Balatarin. Then we simulated the effect of these mechanisms on Balatarin's communities using an agent-based model.

4.4 Regression Results and Mechanisms:

Using the history reconstruction and opinion estimation methodology we introduced in the previous sections, we collected the data needed for our regression models and estimated the coefficients in the models. Table 4-1 presents the estimated coefficient for each regression model along with the p-values.

	Posts		Online rate		Comments	
	Estimate	p-value	Estimate	p-value	estimate	p-value
(Intercept)	0.80862	0.00	-0.09011	0.00	0.60149	0.23
β_1: $U_1 > 0, V_1 > 0, V_2 > 0, U_2 > 0$	0.00863	0.00	-0.00108	0.00	-0.01027	0.00
β_2: $U_1 > 0, V_1 > 0, V_2 > 0, U_2 < 0$	-0.00095	0.27	-0.00001	0.52	0.00091	0.27
β_3: $U_1 > 0, V_1 > 0, V_2 < 0, U_2 > 0$	0.00029	0.14	0.00000	0.82	0.00465	0.02
β_4: $U_1 > 0, V_1 > 0, V_2 < 0, U_2 < 0$	0.00547	0.01	0.00000	1.00	0.00687	0.01
β_5: $U_1 > 0, V_1 < 0, V_2 > 0, U_2 > 0$	0.00138	0.00	-0.00002	0.37	0.00044	0.60
β_6: $U_1 > 0, V_1 < 0, V_2 > 0, U_2 < 0$	0.00121	0.04	0.00180	0.00	-0.01575	0.00
β_7: $U_1 > 0, V_1 < 0, V_2 < 0, U_2 > 0$	0.00118	0.00	0.00242	0.00	-0.02072	0.00
β_8: $U_1 > 0, V_1 < 0, V_2 < 0, U_2 < 0$	-0.00076	0.13	0.00011	0.17	0.00051	0.61
β_9: $U_1 < 0, V_1 > 0, V_2 > 0, U_2 > 0$	0.00162	0.00	-0.00009	0.23	0.00018	0.78
β_{10}: $U_1 < 0, V_1 > 0, V_2 > 0, U_2 < 0$	0.00015	0.95	0.00224	0.00	-0.00956	0.00
β_{11}: $U_1 < 0, V_1 > 0, V_2 < 0, U_2 > 0$	0.00141	0.00	0.00131	0.01	0.00182	0.08
β_{12}: $U_1 < 0, V_1 > 0, V_2 < 0, U_2 < 0$	-0.00103	0.00	0.00021	0.11	-0.00546	0.01
β_{13}: $U_1 < 0, V_1 < 0, V_2 > 0, U_2 > 0$	0.00626	0.01	-0.00015	0.16	-0.01184	0.00
β_{14}: $U_1 < 0, V_1 < 0, V_2 > 0, U_2 < 0$	-0.00105	0.34	-0.00042	0.04	0.00083	0.25
β_{15}: $U_1 < 0, V_1 < 0, V_2 < 0, U_2 > 0$	0.00078	0.67	0.00005	0.36	0.00034	0.43
β_{16}: $U_1 < 0, V_1 < 0, V_2 < 0, U_2 < 0$	0.0077	0.00	0.00003	0.45	0.00011	0.85
Comments	-0.00065	0.00	0.00000	0.77	-	-
Posts	-	-	0.00002	0.20	0.01167	0.00

Table 4-1. Estimated coefficient and P-value for the regression model

There are different mechanisms that can be derived from the estimated coefficients for users' activities, which will be discussed below.

4.4.1 Posting Rate Mechanisms:

The positive (significant) values estimated for coefficients β_1, β_4, β_{13} and β_{16} indicate that all the opinion groups post more after reading stories with the same viewpoint as their own. Hence, for all the opinion groups, gaining utility by reading stories with the same point of view as theirs increases their posting rate. On the other hand, reading stories with contrary opinions (e.g. serious supporters reading yellow opposition stories) reduces the posting rate for all the groups except yellow oppositions (i.e. β_6, β_7 and β_{11} estimated to be positive and significant, while β_{10} is insignificant). Note that a greater difference between the opinion of the user and that of the stories s/he reads results in a further decrease in his/her (utility and) posting rate. In other words, when a serious supporter reads a yellow extreme opposition (or extremely yellow opposition) story, his/her posting rate decreases more than when s/he reads a yellow neutral opposition story. Additionally, serious users (supporters and oppositions) post more stories reading serious opponent (opposition and supporter respectively) story if it increases their utility (i.e. if the utility gained from the seriousness of the content overcomes the decrease in utility caused by opposing/supporting the government), and decreases it otherwise (β_5 and β_9 estimated to be positive and significant). In other words, reading serious stories that are somehow more neutral (i.e. not extreme in opposing/supporting the government) increases the posting rate of serious users. However, reading serious extreme opponent stories reduces serious users posting rates, implying that they become discouraged in sharing their opinions when they read stories with opinions differing widely from their own with regard to the government (i.e. when they feel the opinions of their audiences are very different from theirs). β_{12} is estimated to be negative, which means that yellow oppositions post more stories after reading yellow extreme supporter stories. Unlike in the previous cases, here yellow oppositions confront yellow extremely adverse opinions by posting more stories.

A negative significant coefficient for a number of published comments indicates that the posting rate decreases when a user publishes more comments. This could be the result of users becoming annoyed and discouraged when they participate in discussions on their posted stories, and consequently reducing their posting. A positive significant intercept indicates that on average the number of posts increases (regardless of what users have read on the website) between the two intervals. This change in the posting rate varies between different users ($\sigma_{users' \, random \, effect}$ = 0.9522), but the number of weeks that have passed since the user started using the website (i.e. the time random effect) has no significant effect on his/her posting rate.

Conditional R-squared and marginal R-squared of the regression model are 0.41 and 0.14, respectively, meaning that the model explains 41% of the variability in the data, while the fixed effects alone explains 14% of the variability.

4.4.2 Online (Revisiting) Rate Mechanisms:

Most of the coefficients in the online rate regression model are not significant, which implies that what users read on the website has limited effect on their likelihood of revisiting the website. We will discuss those that are significant here. Negative β_1 signifies that reading likeminded stories decreases serious supporters' online rates. This means that they return less frequently after reading those stories. β_6, β_7, β_{10} and β_{11} are all estimated to be positive and significant, which means that all of the opinion groups revisit the website less often after reading stories from users with adverse opinions (e.g. serious supporters reading stories from yellow oppositions). This could be the result of users becoming frustrated by reading stories that are incompatible with their opinions. A negative significant β_{14} implies that yellow oppositions return to the website more often upon reading highly serious opposition stories.

A negative significant intercept indicates that users' online rates decrease on average over time, but the scale varies (slightly) based on the number of weeks that have passed since the date the user joined the website ($\sigma_{time \, random \, effect}$ = 0.05943). There is no variation in the change of online rates between different users. Note that posting is conditional on user being online, and consequently positive and negative intercept for posting and online rate (respectively) implies that over time users revisit website less frequently but in each online session they post more stories. Additionally, the number of published posts and comments has no effect on the online rate.

The conditional and marginal R-squared are both less than 0.01, which means that the model explains less than 1% of the variability in the data. Therefore, the change in users' online rates is mostly affected by other exogenous variables that are not considered in our model.

4.4.3 Commenting Rate Mechanisms

Negative β_1 and β_{13} imply that serious users (both supporters and oppositions) decrease their contributions to discussions when they read serious stories in agreement with them (either opposing or supporting the government). This means that, for serious users, reading stories with the same viewpoint as theirs decreases their willingness to participate in discussion. A positive β_3 shows a decrease in the comment rate for serious supporters when they read highly yellow supporter stories. A positive significant β_4 means that yellow supporters comment more after reading stories with the same viewpoint as their own. Negative β_6 and β_7 indicate that serious and yellow supporters of the government become more engaged in discussions after reading yellow and serious (respectively) oppositions' stories. β_{10} is also negative and significant, meaning that yellow oppositions comment more after reading serious supporter stories. A negative β_{12} implies that yellow oppositions engage in more discussion after reading stories from yellow extreme supporters. β_{13} is negative and significant, which means that serious oppositions comment less after reading stories with the same opinions as theirs.

The estimated intercept value is not significant, so there is no constant change in the number of comments over time. However, the change in the number of comments varies for different users and time since joining the website ($\sigma_{user\ random\ effect}$ = 0.8305 and $\sigma_{time\ random\ effect}$ = 0.8861). Also, the number of posts the user publishes has a positive effect on his/her number of comments. This could be due to the users' contributions in discussions on their posted stories. Conditional R-squared of the online rate regression model is 0.13, while its marginal R-squared is 0.05 (i.e. the model explains 13% of the variability in the data, while fixed effects alone explain 5% of the variability).

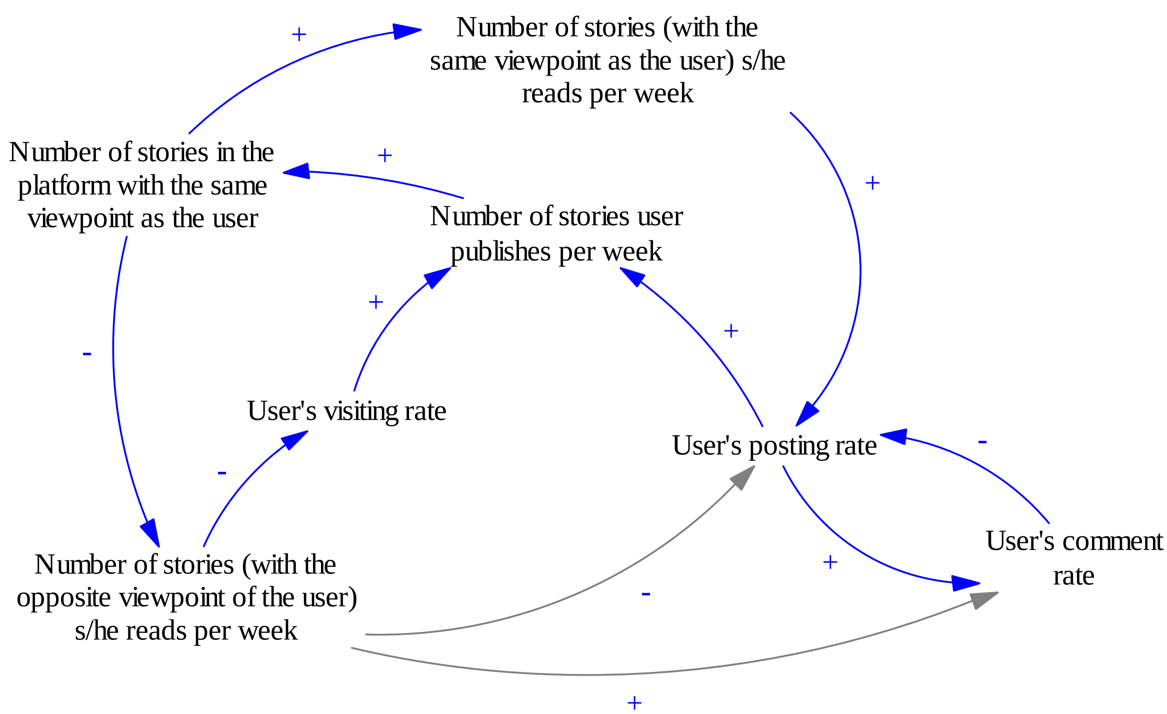


Figure 4-1. Different mechanisms form around user reading opposite/agreeing opinions

Figure 4-1 represents some of the mechanisms that come from users reading stories with the same viewpoint as/opposite viewpoint of their own. In short, all of the opinion groups post more after reading stories with the same viewpoint as their own. All of the groups (blue connections in the figure) visit the website less frequent and three out of four (gray connections in the figure) post less often but publish more comments reading contrary opinions. Additionally, users publish more comments after posting stories but post less after publishing comments. Finally, users on average post more often but revisit less frequent over time. Next, we develop an agent-based simulation model and study the formation of Balatarin's communities as a result of users' interactions with different opinions.

4.5 Simulation Model:

The regressions above estimated the impact of consuming different types of opinion on an individual's activity on Balatarin. However, in a system where individuals affect each other, the outcome is hard to assess based on individual-level regression equations. To study and understand the evolution of such a nested system, we develop an agent-based model that simulates the behaviors and actions of the users and the formation of communities as a result of the users' interactions with the opinions of others.

We model a simplified version of the social news websites in which users can get online on the website, publish posts and comment on other posts. Here (for simplification) we ignore the voting mechanism and assume a system that keeps posted stories for one day and sorts the stories based on their time of publication (similar to the ordering page of recently published stories on Balatarin). We consider two different general states for the user: online and offline. An online state equates to visiting the website; users can also read stories, post new stories and comment on stories they read. Figure 4-2 shows the states and the transitions between states for each simulated user.

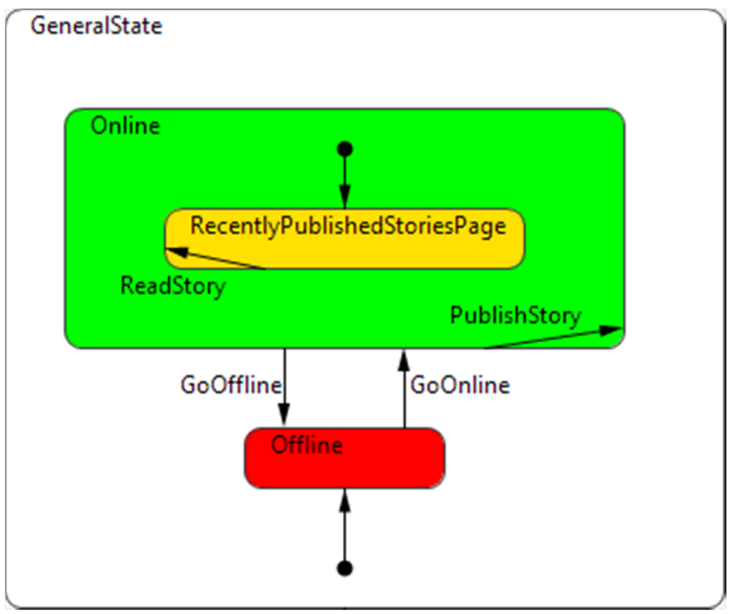


Figure 4-2. Users' states and transitions

Users, as agents in the simulation, change their state from offline to online through the *GoOnline* transition with the rate of *OnlineRate*. In the online state, users can either publish a new story (using the *PublishStory* transition) or visit the (*RecentlyPublishedStoriesPage*) ordering page and read stories (through the *ReadStory* transition). Users go offline after spending a certain amount of time (with the rate *OnTimeAverage*) on the website through the *GoOffline* transition. Table 4-2 shows the transitions' triggers and the pseudo-codes for each transition.

Transition Name	Trigger	Pseudo Code
GoOnline	OnlineRate (rate)	UserNumberOfGettingOnlineThisWeek++;
ReadStory	ReadingRate (rate)	Story s =SearchForNewStory(); SetActivity(s.StoryOpinionDim1,s.StoryOpinionDim2,s.StoryAttractiveness); CommentForStory(CommentRate); SetStoryAsRead(s);
PublishStory	PostingRate/OnTimeAverage (rate)	AddStory(UserOpinionDim1,UserOpinionDim2); UserNumberOfPublishedStoriesThisWeek++; OnlineRate+= PostsEffectOnOnlineRate; CommentRate+=PostsEffectOnCommentRate; SetBoundaries();
GoOffline	OnTimeAverage (timeout)	EmptyReadStories();

Table 4-2. Transitions' triggers and the pseudo-codes

Users can visit the website on a daily basis with the rate *OnlineRate.* When the user is online, s/he can read new stories (i.e. starting from the top of the ordering page, the user looks for a story s/he has not read before through *SearchForNewStory*) with the rate of *ReadingRate* and his/her activity (*OnlineRate*, *CommentRate* and *PostingRate*) updates based on the opinion and the attractiveness (fixed effect) of story (*StoryOpinionDim1, StoryOpinionDim2, StoryAttractiveness*) s/he reads (through *SetActivity* procedure). Then s/he decides whether to comment on the story based on his/her *CommentRate* (see *CommentForStory* in the appendix). Finally, the story will be marked as read for the user (using the *SetStoryAsRead* procedure).

Users can also publish stories with the rate $\frac{PostingRate}{OnTimeAverage}$, where *OnTimeAverage* is the duration for which the user browses the website and *PostingRate* is the average number of stories the user publishes in that duration. When the user posts, his/her story will be added to the system with the same opinion of the publisher (through *AddStory* procedure) and will be placed at the top of *RecentlyPublishedStoriesPage*. Then the *OnlineRate* and *CommentRate* will be updated based on the effect they get from posting a story. After each update in activity, we check the rates for any violation in the boundaries[14] (in *SetBoundaries* procedure). Finally, after a certain amount of time (*OnTimeAverage*), the user leaves the website and the memory assigned to the story s/he read will be released (through *EmptyReadStories*). The parameter values for user agents are provided in Table 4-3 (*ReadingRate* and *OnTimeAverage* estimated from available data from Balatarin and for other parameters, we assume a normally distributed (i.e. standard normal) community). Later we will study the effect of these parameters on the formation of communities:

Parameter Name	Parameter Value
ReadingRate	44/hour
OnTimeAverage	38 minutes
UserActivity	$Normal\ (\mu = 0, \sigma = 1)$
UserOpinionDim1	$Normal\ (\mu = 0, \sigma = 1)$
UserOpinionDim2	$Normal\ (\mu = 0, \sigma = 1)$

Table 4-3. User agent parameters and values

[14] We assume that the user will not publish more than 100 stories, post more than 500 comments or visit the website more than 5 times per day. These rates cannot be negative either.

SetActivity updates the *PostingRate*, *OnlineRate* and *CommentRate* of the user based on the opinion of the story s/he read using the following formulas:

$$PostingRate = PostingRate + OpinionEffectOnPostRate_{opinion\ group\ i} \times (sigmoid(UserOpinionDim1 \times StoryOpinionDim1 + UserOpinionDim2 \times StoryOpinionDim2 + StoryAttractiveness) - 0.5) \qquad (3)$$

$$OnlineRate = OnlineRate + OppinionEffectOnOnlineRate_{opinion\ group\ i} \times (sigmoid(UserOpinionDim1 \times StoryOpinionDim1 + UserOpinionDim2 \times StoryOpinionDim2 + StoryAttractiveness) - 0.5) \qquad (4)$$

$$CommentRate = CommentRate + OppinionEffectOnCommentRate_{opinion\ group\ i} \times (sigmoid(UserOpinionDim1 \times StoryOpinionDim1 + UserOpinionDim2 \times StoryOpinionDim2 + StoryAttractiveness) - 0.5) \qquad (5)$$

where index $OppinionEffect_{\_opinion\ group\ i}$ indicates the effect (i.e. coefficient) of the story on the rate based on the user's and the story's opinion group (i.e. supporting/opposing the Iranian government and being a fan of yellow/serious stories). In formulas 3, 4 and 5 we basically calculate the marginal utility value that the user gains by reading the story (i.e. $sigmoid(UserOpinionDim1 \times StoryOpinionDim1 + UserOpinionDim2 \times StoryOpinionDim2 + StoryAttractiveness) - 0.5$), as in the regression models, and derive the effect of that marginal utility on the user's activity change by multiplying the marginal utility value by the $OpinionEffect$ coefficients (i.e. $OppinionEffectOnPostRate$, $OppinionEffectOnOnlineRate$, $OppinionEffectOnCommentRate$) in the respective opinion group.

For simplification, in our simulation model we assumed that there was only one ordering page (*RecentlyPublishedPage*) so that when a story is posted it is placed at the top of this ordering page. The lifecycle of stories on this ordering page (*LifeCycleOfStoryInRecentlyPublishedPage*) is one day. Figure 4-3 shows the state and transitions of stories.

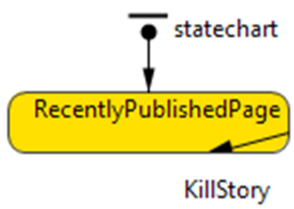


Figure 4-3. State and transitions of stories

KillStory transition removes the story from the ordering page (and the system) when it reaches the end of its lifecycle (i.e. one day for a recently published ordering page). Table 4-4 shows the parameters of story objects.

Parameter Name	Parameter Value
StoryAttractiveness	*Normal* ($\mu = 0, \sigma = 1$)
StoryOpinionDim1	Publishers' *UserOpinionDim1*
StoryOpinionDim2	Publishers' *UserOpinionDim2*

Table 4-4. Parameters (and parameter values) of story agents

The results of the simulation model and parameter variation will be presented in the next section.

4.6 Simulation Results:

In this section, we present our simulation of a symmetric (i.e. unbiased) community (with mean zero in starting opinions) and discuss the evolution over time of such a community considering the mechanisms described in the previous sections. Then, by changing the parameters of the simulation, we are able to study the behavior and the future formation of communities in other (simulated) social media environments, those results will be presented afterward.

4.6.1 Simulating an Symmetric Community

First, we ran the simulation on 500 users with standard normally distributed opinions (on 2 dimensions) and stories' attractiveness (μ=0,σ=1) for 1000 days.[15] The total numbers of weekly posts and the average (weekly) posts in each quadrant (i.e. supporter of the Iranian government and fan of serious stories, supporter of the Iranian government and fan of yellow stories, oppositions of the Iranian government and fan of serious stories, oppositions of the Iranian government and fan of yellow stories) are provided in Figure 4-4 and Figure 4-5. The results show that the total and average numbers of posts (for all of the opinion groups) have bell-shaped trends (with different skewness and peak values for each of the quadrants). Therefore, in a symmetric community, users in all opinion groups start losing interest in activity on the website (i.e. revisiting the website and posting stories) after reaching a peak point. This means that if we assume that no new user will join the community, a platform with such a structure is doomed to failure in a few years.

By examining the post rate and online rate (Figure 4-14 and Figure 4-15, provided in the appendix), we can see that the post rates, on average, increase (with a goal-seeking trend) over time (in all the quadrants), but a constant decline in the online rate limits the number of posts published by users. In other words, over time users tend to post more stories during each of their online sessions; however, the number of their online sessions declines. The interaction of these two effects leads to the nonlinear relationship observed in total activity, with an initial increase and a longer-term decline. For all the opinion groups, bell-shaped trend of the average weekly posting is a result of increase in posting at first (caused by reading stories with the same opinion of the group and constant increase in posting rate over time (i.e. positive intercept for posting rate)) and then decrease in visiting (caused by reading more stories with adverse opinion and constant decrease in online rate (i.e. negative intercept for online rate)). However, skewness and the peak of the average weekly post varies between different groups, due to the role of other mechanisms in the system. For instance, among users who are fan of serious stories, those who support the government (unlike the oppositions) visit the website less frequent after reading stories with the same opinion as their own. As a result, the peak of the average weekly posts for that group ends up lower than that of the oppositions (and their posting rate also declines faster the other group). Such mechanisms can influence the future formation of the opinion in the platform; we will discuss more of those mechanisms in the next section.

[15] We conducted a sensitivity analysis on the number of users in the community and with 500 users (and more) consistent results were obtained. Basically, based on the structure of the simulated social media site (on which stories have a one-day lifecycle and are sorted based on their time of publication), for consistent results we need to have enough new stories for each user to read on each of his/her online sessions. The *ReadingRate* and *OnTimeAverage* of users determine the number of new stories we need in the system for consistent simulation results and, considering the parameter values we have (Table 4-3), 500 users generate enough new stories for all users to read. The community fizzles due to lack of content for significantly smaller user bases.

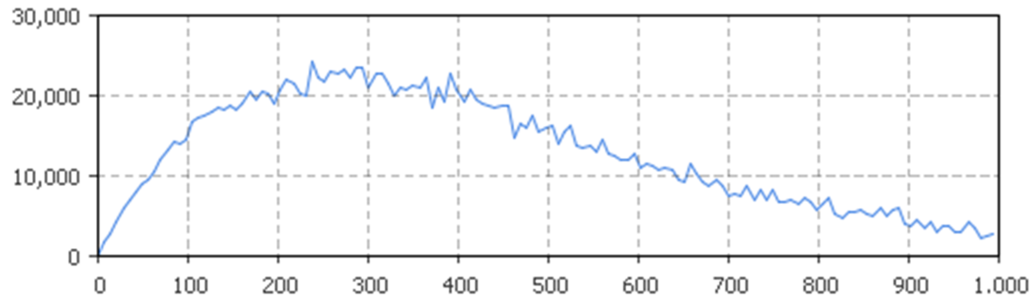


Figure 4-4. Total number of weekly posts simulated for a symmetric community over 1000 days

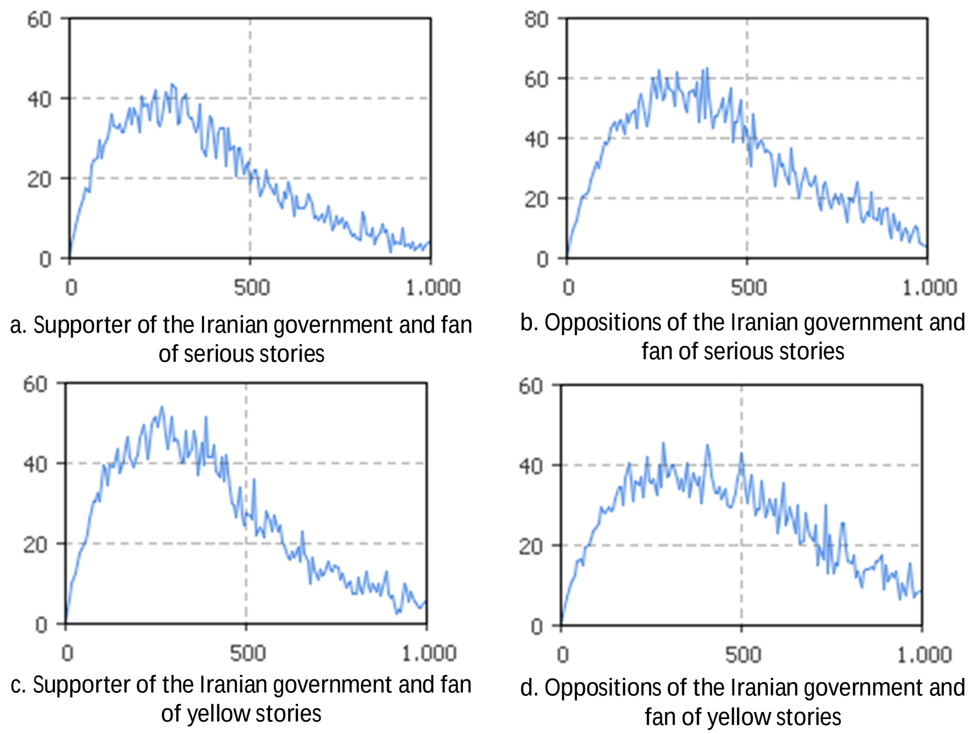


Figure 4-5. Average weekly posts in each quadrant simulated for a symmetric community over 1000 days

The total and average numbers of published comments (see Figure 4-16 and Figure 4-17, provided in the appendix) have the same skewed bell-shaped trends. Again, the comment rate follows a (goal-seeking) rising pattern (see Figure 4-18 in the appendix) and the online rate caps the number of published comments. Therefore, over time, users also publish more comments during each online session, but in total the number of their comments decreases due to their participating in fewer online sessions.

In general, the decreasing online rate in all of the groups (see Figure 4-15 in the appendix and the estimated negative intercept for the online rate regression) can be interpreted as users becoming tired of the community, regardless of their opinion and that of the community. Furthermore, the fact that the number of posts and comments is estimated as being non-significant (in the online rate regression) shows that even users' activity rates on the website have little effect on their eventually becoming tired of the community and ultimately leaving it.

4.6.2 Parameter Variation

So far we simulated a symmetric community, meaning the user opinions were distributed with the mean 0 on the two dimensions. Online communities could be asymmetric (i.e. biased) toward an opinion for different reasons. The popularity of a community within a specific group of users with specific mindsets, subsets of initial users who invite and attract likeminded individuals to join the community, among other reasons, could form a community biased towards one quadrant of the opinion space or another. Here we study the effect of community bias on its users' activity rates and the evolution of the community.

4.6.2.1 Communities Asymmetric in Supporting/Opposing the Iranian Government

As our first experiment, the effect of an imbalanced number of users supporting versus opposing the Iranian government on the website on the future formation of the communities on the website was studied. In order to do this, we changed the mean of users' opinions in the first dimension in the range of $[-1, 1]$ (from users mostly opposing the Iranian government to mostly supporting the government) with a step-size of 0.1 and ran a simulation with 500 users for each.

The result of the simulation for the total weekly number of posts published (see Figure 4-6 with highlights and Figure 4-19 in the appendix without highlights) indicates that more asymmetric communities ($\mu = 1.0, 0.9, -1.0, -0.9$, highlighted as blue for supporters and red for oppositions of the Iranian government in Figure 4-6) post more stories compared to symmetric ones ($\mu = 0.0, -0.1, 0.1$, highlighted as purple). The result shows that, in general, communities biased toward oppositions of the Iranian government post more stories compared to those biased toward supporting the government. However, compared to symmetric communities, both supporters and oppositions of the government ultimately post more stories.

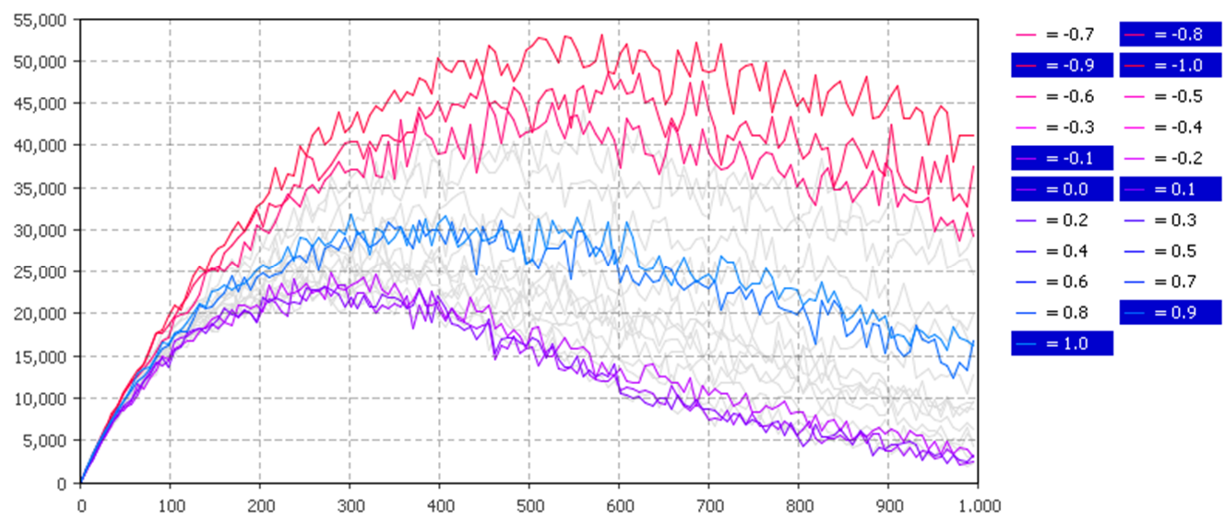


Figure 4-6. Total number of weekly posts simulated for differently biased (in terms of supporting/opposing the Iranian government, red lines for the oppositions and blue lines for the supporters of the government) communities over 1000 days (more biased communities highlighted)

Further investigating the cause of such an outcome, the average number of posts in each opinion group (see Figure 4-20 in the appendix) illustrates that both supporters and oppositions publish more posts when the community is biased toward them. However, serious oppositions of the Iranian government (on average) post more stories (when the community is biased toward them) compared to serious supporters.

The average posting rate for serious supporters decreases significantly in the communities biased toward the oppositions (see Figure 4-21 in the appendix), which indicates that serious supporters become discouraged (from posting) when reading opposition stories (since β_5 and β_7 (both estimated to be positive) have a negative effect on the posting rate when the user reads the opinions of the opposition (extreme opposition in the case of β_5) stories). Therefore, serious supporters lose ground when they interact with oppositions, although reading stories that support their own opinions pushes them to post more (to support opinions that match theirs, since β_1 is estimated to be positive). On the other hand, the posting rate of yellow oppositions increase in communities biased toward supporters. Based on the positive estimated value for β_{16} and the negative estimated value for β_{12}, yellow oppositions become encouraged to post more when they read sympathetic stories (i.e. yellow opposition stories), but their posting rate also increases when they read yellow stories that extremely support the government. The former could be a case of supporting similar opinions (by posting more stories in line with the opinion) and the latter could show a confrontational act by yellow oppositions in front of yellow extreme supporters. In the other two groups of opinion (i.e. yellow supporters and serious oppositions), the biasness of the community has a minor effect on their posting rate, but a change in their online rates affects the average weekly number of posts.

Online rates (see Figure 4-22, provided in the appendix) have a declining trend in all groups and the slope increases (i.e. the online rate decreases faster) in communities biased toward adverse opinions. For serious supporters, the biasness has a minor effect on the online rate, compared to others.

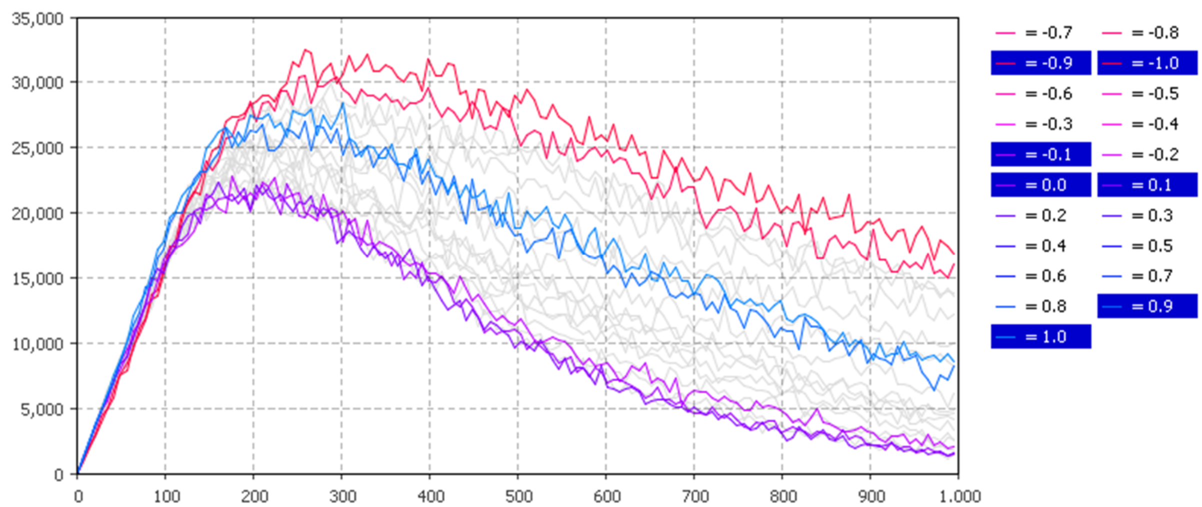


Figure 4-7. Total number of weekly comments simulated for differently biased (in terms of supporting/opposing the Iranian government, red lines for the oppositions and blue lines for the supporters of the government) communities over 1000 days (more biased communities highlighted)

A study of the effect of the community bias on weekly published comments indicates that communities biased toward oppositions or supporters of the government generate more comments ($\mu = 1.0, 0.9, -1.0, -0.9$ highlighted in Figure 4-7 with blue and red) compared to symmetric communities ($\mu = 0.0, 0.1, -0.1$ highlighted in purple in Figure 4-7).

While yellow supporters become encouraged to discuss more when they read stories closely aligned with their opinion (β_4 estimated to be positive in the comment rate regression model), serious supporters and serious oppositions feel less motivated to contribute to the discussions when they read stories conveying opinions the same as theirs (β_1 and β_{13} estimated to be negative). This could mean that they do not see much value in discussing the issue with people with whom they share common ground. Moreover, in three out of four groups (except for serious oppositions, where β_{11} is not significant), the result of the regression shows that users tend to comment more when they read stories conflicting with their opinion (i.e. β_6, β_7 and β_{10} estimated to be negative). This could be a result of users' tendency to defend their standpoint in front of adverse opinions. Yet, an examination of the average comments (see Figure 4-25 in the appendix) and comment rates (see Figure 4-24 in the appendix) shows that all groups post more comments when the community is biased in their favor. Considering the coefficient values in Table 4-1, this behavior is mainly caused by the positive effect of posting on a number of published comments, meaning that users tend to contribute to discussions that take place on their published stories. As a result, when communities are biased toward a group, since that group tends to post more stories, they will end up commenting more on their published stories and will therefore have more discussion in communities biased toward them.

4.6.2.2 Communities Asymmetric in Being a Fan of Yellow/Serious stories

Likewise, the effect of an imbalanced number of users being fans of yellow or serious stories on activity was studied by changing the mean of users' opinions in the 2nd dimension in the range of $[-1, 1]$ (users mostly fans of yellow stories to mostly fans of serious stories) with 0.1 step-size.

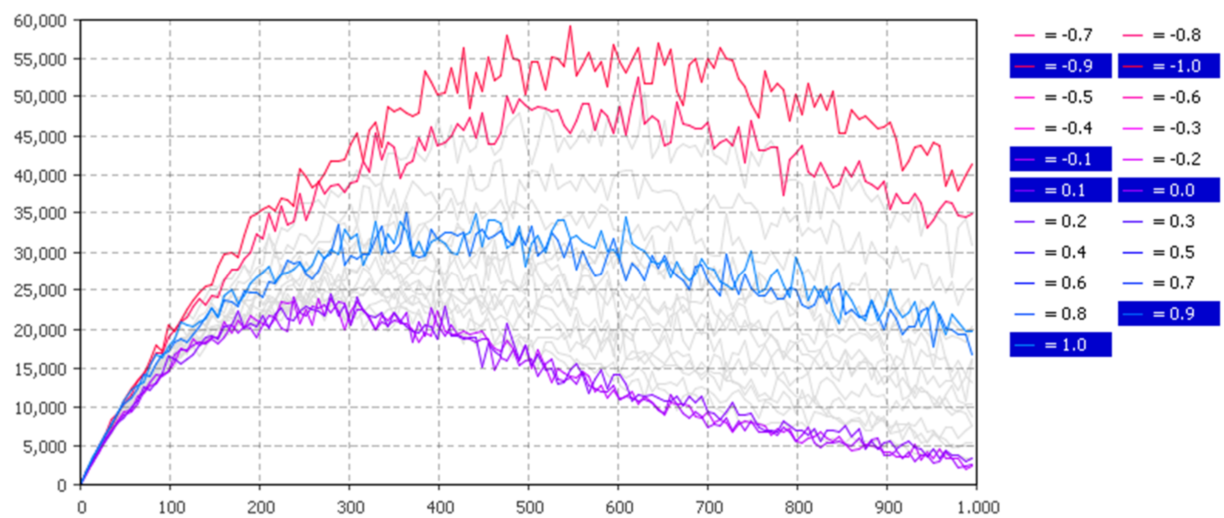


Figure 4-8. Total number of weekly posts simulated for differently biased (in terms of being a fan of yellow/serious stories, red lines for the fans of yellow stories and blue lines for fans of serious stories) communities over 1000 days (more biased communities highlighted)

As Figure 4-8 (and Figure 4-26 without highlights in the appendix) shows, asymmetric communities (toward yellow/serious stories) generate more posts than symmetric ones. This behavior mostly comes from the change in the online rates (see Figure 4-28 in the appendix), indicating that users' online rates drop more slowly when the community is biased in favor of them. This means that serious users tend to visit the website more often when there are more serious contents on the website, and the same is true for yellow users with more yellow contents on the website.

The total number of published comments follows the same trend, meaning that users discuss more in asymmetric communities (see Figure 4-29 in the appendix). In this case, beside the online rates, a change in the comment rates also has a significant effect on the discussions and (as Figure 4-30 in the appendix shows) in most cases (except for serious supporters whom biasness has a minor effect on their comment rate) users tend to contribute to discussions more in communities that are biased (in terms of being a fan of yellow/serious stories) toward them. Again, an increase in the number of posted stories and as a result contributing to discussions on posted stories also has a major effect on increases in commenting.

4.6.2.3 Communities' Extremeness in Supporting/Opposing the Iranian Government

Next we studied the effect of having a more extreme opinion (toward the Iranian government) on the formation of Balatarin's communities, by changing the standard deviation of users' opinions in the first dimension in the range of [0.1, 2.0) (neutral to extreme supporter/oppositions of the government) with 0.1 step-size.

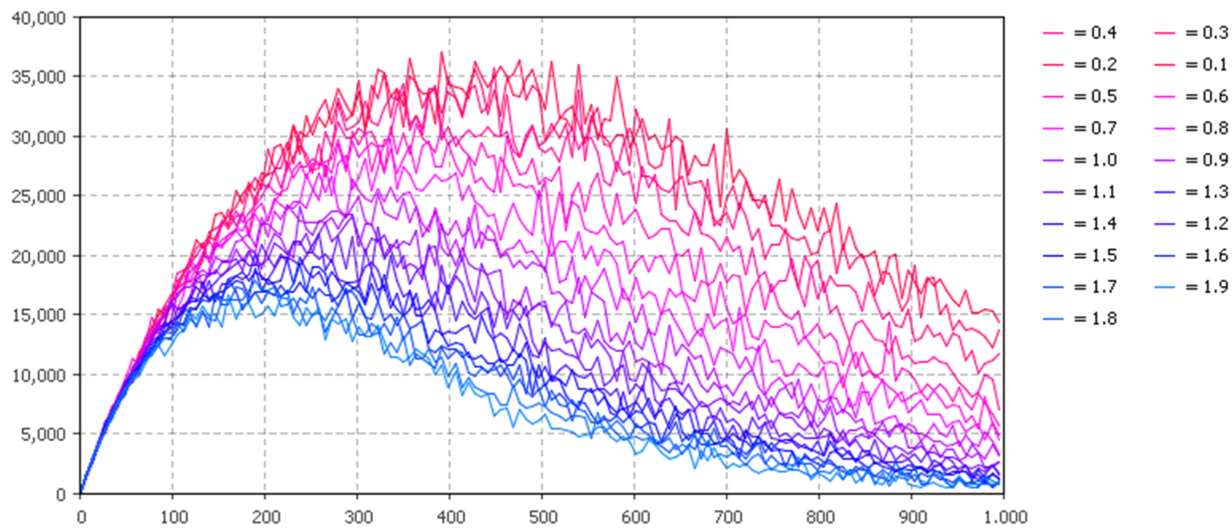


Figure 4-9. Total number of weekly posts simulated for neutral/extreme communities (red lines for more neutral communities and blue lines for more extreme communities) over 1000 days

The result of the total number of published stories shows a decrease (shown as blue lines in Figure 4-9) in posting when users are more extreme (in terms of either supporting or opposing the government). The change in the average weekly posting rate is not significant in this case (see Figure 4-31 in the appendix), meaning that the behavior is mainly caused by the change in the online rate. The average online rate (see Figure 4-32 in the appendix) indicates that in all groups, users tend to visit the community less often when the community's opinion regarding supporting or opposing the government is more extreme. This is due to the fact that the online rates (in all groups of opinions) decrease when users interact with stories contrary to their point of view (e.g. when serious supporters read a yellow opposition story), since β_6, β_7, β_{10} and β_{11} are estimated to be negative. When communities are more extreme, the effect of these coefficients increases and, as a result, online rates decay faster. This could mean that users cannot remain in communities consisting of individuals with opinions extremely different from their own.

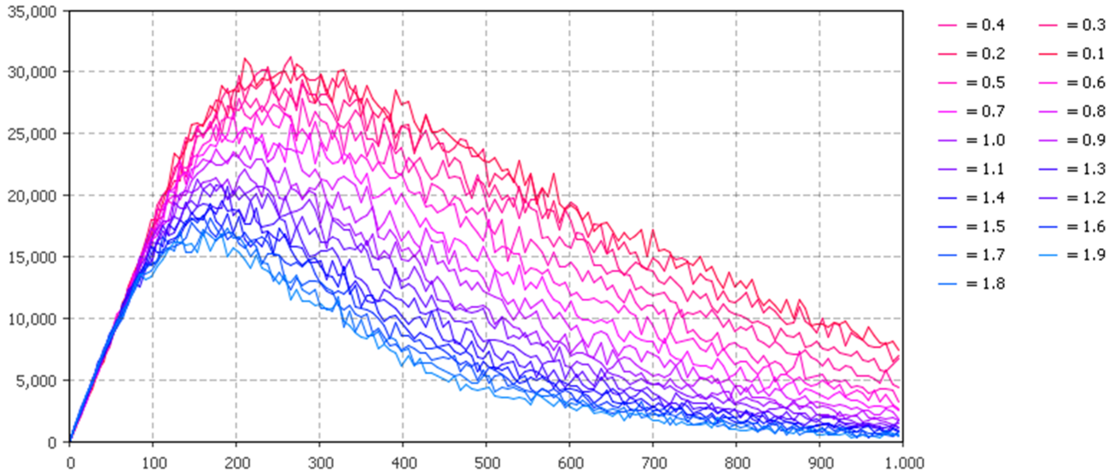
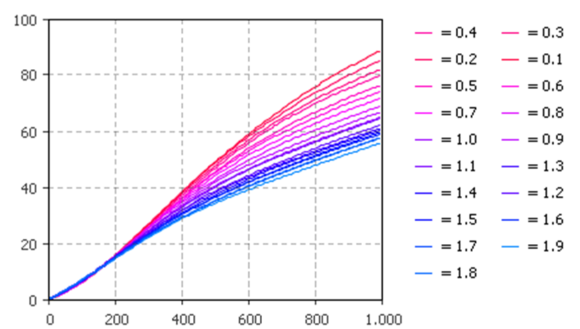
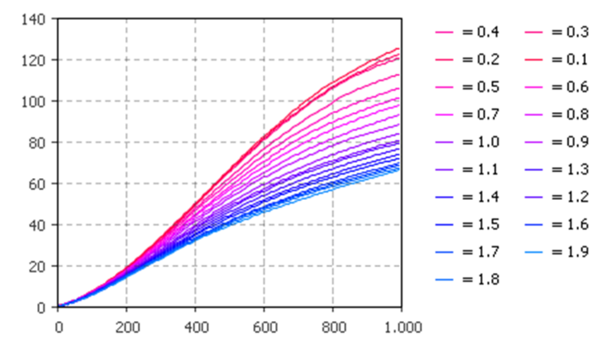


Figure 4-10. Total number of weekly comments simulated for neutral/extreme communities (red lines for more neutral communities and blue lines for more extreme communities) over 1000 days

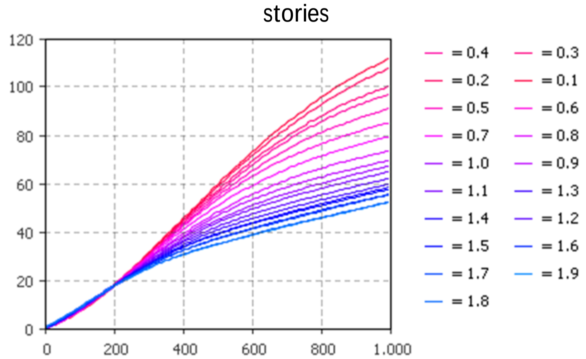
The total number of published comments follows the same pattern in which users have more discussion in more neutral communities (Figure 4-10). Based on the coefficient estimated for changes in the comment rates, users tend to discuss more when they read stories with adverse opinions (β_6, β_7 and β_{10} are negative). However, here (similar to the first experiment) more comments are published as the result of users contributing to discussions in response to their published posts. Thus, posting more stories in neutral communities forces users to discuss more (on their posted stories) and increases their comment rates in such communities compared to more extreme ones (Figure 4-11).



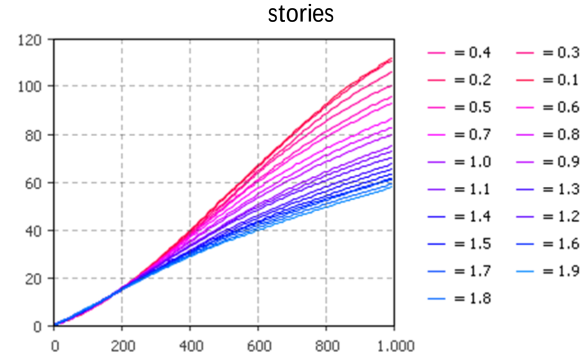
a. Supporter of the Iranian government and fan of serious stories



b. Oppositions of the Iranian government and fan of serious stories



c. Supporter of the Iranian government and fan of yellow stories



d. Oppositions of the Iranian government and fan of yellow stories

Figure 4-11. Average comment rates in each quadrant simulated for neutral/extreme communities (red lines for more neutral communities and blue lines for more extreme communities) over 1000 days

4.6.2.4 Attractiveness of stories

Finally, conducing sensitivity analysis on the attractiveness of the stories (changing the mean of stories' fixed effect in the range of [−1.0,1.0] (less attractive to more attractive) with a step-size of 0.1) highlights that the total number of published posts and comments are higher in communities with more attractive stories (Figure 4-13 and Figure 4-14). Posting rates are not sensitive to stories' attractiveness (see Figure 4-33 in the appendix); however, the change in online rates is highly significant (Figure 4-34 in the appendix). This means that reading high quality stories will not encourage users to post more, but will make them visit the website more often. In the case of serious oppositions and yellow supporters, high quality stories (with μ close to one, shown in Figure 4-34 in the appendix) can even increase the online rate, persuading users to return to the website more often and breaking the pattern of activity decline.

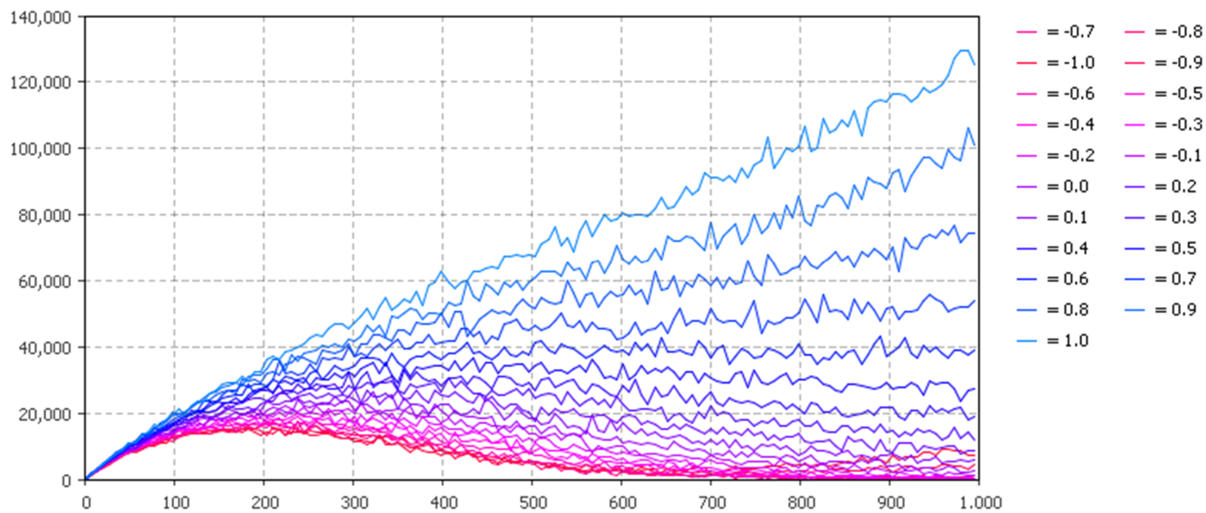


Figure 4-12. Total number of weekly posts simulated for communities generating high/low quality contents (red lines for communities generating low quality contents and blue lines for communities generating high quality contents) over 1000 days

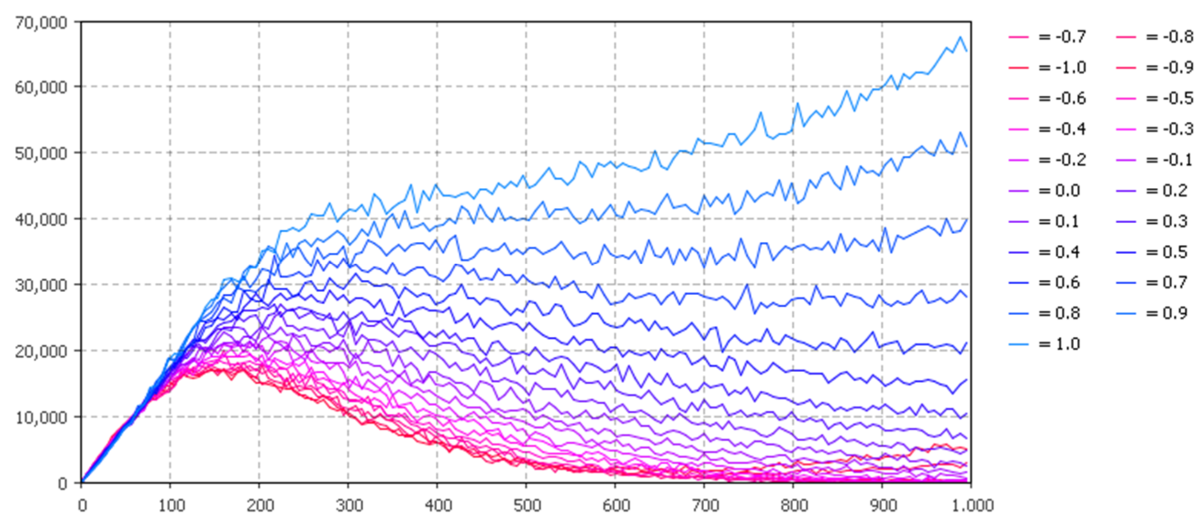


Figure 4-13. Total number of weekly comments simulated for communities generating high/low quality contents (red lines for communities generating low quality contents and blue lines for communities generating high quality contents) over 1000 days

Comment rates, however, are highly sensitive to the attractiveness of stories (Figure 4-35 in the appendix) and increase for high quality stories in all groups. Therefore, users (regardless of their opinions) tend to discuss higher quality stories more.

4.7 Discussion and Conclusion:

Social media platforms play an important role in people's lives today. From messaging applications such as Viber, Telegram and Line, to picture sharing platforms such as Pinterest and Instagram and social networks such as Facebook, they all affect our daily lives to a great extent. We use social media for contacting our family and friends, to share our ideas [7] and spread the word about our causes [8][9], services or products [10][11]. The effect of social media is such that (on average) we now spend around 7-10% of our lives on social media[16]. Regardless of its positive [12][13] or negative [14][15][16] effects, maintaining a social media site (that would be worth billions of dollars and would cost millions of dollars per month) and keeping it sustainable and active is very important for its shareholders, employees, users, as well as other companies, services and communities built up around it (e.g. social network advertisement partners, applications that depend on the social media platform (e.g. Facebook's quick login button is used in many other applications), social network communities (e.g. Facebook fan pages, groups, etc.)). Furthermore, there exists evidence suggesting that using social media as an internal communication tool [17] in organizations can increase internal engagement between employees and help in generating new ideas across teams [18]. Recently, many organizations have developed their own internal social networks (e.g. Nestle has its own internal version of Facebook [19]) or have started using existing solutions (such as Yammer (owned by Microsoft, internal social network of eBay and SuperVlu), Chatter (internal social network of Dell), Slack, Salesforce, M.S. Sharepoint, etc.) to create such platforms. Thus, studying users' (as the core part, content generator and driving force of the social media) behaviors and their reaction in interacting with each other in online communities plays an important role in finding the cause of the rise and fall of social media platforms and thus in keeping that part of the technological economy moving.

In this research, we studied the effect of individuals' interactions with a variety of opinions in social media on their online activities and consequently on the formation and future of the platform. We estimated different mechanisms that influence individuals' behaviors online, and studied the effect of these mechanisms on the lifecycle of social media through an agent-based simulation model. Finally, we examined the effect of different compositions of users on individuals' activity and the future of the social media. Although our data extraction algorithm, opinion estimation method and models on changes in users' activity (as the three main components of this research) were customized for the types of data we collected from Balatarin, all of these components have generic approaches and can be tailored and applied to other social media platforms.

Our results indicate that individuals' reactions to others' opinions vary based on the subject of the argument (e.g. the reactions of supporters of the Iranian government to oppositions' stories are different from those of fans of serious stories to reading yellow stories) and on the opinions of the individuals regarding the subject (e.g. supporters' reactions to oppositions' stories are different from oppositions' reactions to supporters' stories). We found that, in general, individuals increase their activity on social media upon interacting with contents that closely match their own opinions; however,

[16] GWI Social report 2015: Typical internet users spend on average 1.77 hours per day on social networks, while younger generations spend a lot more than that (2.68 hours for 16-24-year-olds and 2.16 for 25-34s) on social media.

reading extreme content could still have a negative effect on individuals' activity levels. Furthermore, the quality of the content (i.e. the attractiveness of the stories) plays an important role in keeping individuals interested and active on the outlet. Therefore, filtering low quality content through various sorting mechanisms could help to keep social media platforms active.

Besides the nuanced impacts of consumed opinions on users' activities, some practical implications follow our results. First, it seems that keeping users entails engaging them with content that are closely aligned to the users' opinions (i.e. by creating clusters of likeminded users and feeding them with content generated among themselves). Most of the social network platforms today use recommendations and ranking algorithms that work based on the proximity of opinions. Some extract users' opinions directly (using methods such as collaborative filtering) and feed the users with content that have proven interesting to other likeminded people (e.g. Netflix and Pandora recommend movies and songs to users based on the tastes of other users), while others use different proxies to feed users with content closely aligned to their own opinions (e.g. Facebook feeds its users with content shared by their friends or pages in which they are interested). Both cases create plurality in the communities, which, based on our results, increases the lifecycle of the outlets. Yet, ranking algorithms in both cases treat all of the opinion groups the same which may not be optimal. For example there is no difference in the ranking of oppositions' stories for supporters and supporters' stories for oppositions, although they may react differently to adverse opinions. Our study, however, shows that different opinion groups could be treated according to their reactions, in order to maximize users' activity rates.

The usage of social media and the focus of its user-generated contents could be different across various platforms or even countries and cultures. Facebook, for example (in most countries), is known as a forum for sharing personal opinions, feelings, stories, pictures and videos. However, in Japan (due to the Japanese culture, which emphasizes good public expressions, and because of Facebook's insistence on real-name-only accounts), people use Facebook to show their professional face and share their resume, work experience and business connections on it (somewhat like LinkedIn). In another example, Telegram - a popular texting platform (and one of the few that is not blocked by the government) - in Iran is used by Iranians as a channel for sharing (political) news, funny contents and even as a product and service advertisement platform. Thus, different communities (formed to discuss specific subjects, such as the Iranian government, political news, etc.) and the ways in which they use different social media outlets should not be ignored in the platform's structure or feeding algorithms. The results of our study also show that the best way for social media platforms to increase their lifecycle is by focusing on more neutral users and keeping them away from contents conveying extreme opinions. Finally, attracting new users (even considering scaling technical difficulties) can always help a social media outlet to grow and provide its users with more fitting and high quality contents.

A more fundamental challenge to the society relates to the impact of these filtering algorithms on users opinions (which we did not model here). If users are shown content consistent with their opinions, they may find few opportunities to be challenged, and to revise their opinions. Such filter bubbles may hurt democratic conversations critical for sifting through complex social problems, building understanding across ideological gaps, and solving big challenges that require people of different opinions to come together. To the extent that single social media companies are forced to create filter bubbles to thrive, the society may be facing a prisoners dilemma in which no firm dares to promote cross-boundary communication which is in the interest of everybody.

Appendix C

Appendix 1

The simulation's procedures, inputs and pseudo code are provided below in Table 4-5.

Procedure/function Name	Inputs	Pseudo Code
SearchForNewStory	-	**for**(**Story** s : get_Main().StoriesInRecentlyPublishedPage) **if**(!ReadStoriesID.contains(s.StoryID)) **return** s;
SetStoryAsRead	**Story** s	ReadStories.add(s); ReadStoriesID.add(s.StoryID);

AddStory	**double** UserOpinionDim1
	double UserOpinionDim2
		Story s=get_Main().add_story();
		s.StoryOpinionDim1=UserOpinionDim1;
		s.StoryOpinionDim2=UserOpinionDim2;
		s.PublisherID=UserID;
		get_Main().LastPublishedStoryID++;
		s.StoryID=get_Main().LastPublishedStoryID;
		get_Main().StoriesInRecentlyPublishedPage.add(0,s);
		s.DateOfPublish=getTime();
		SetStoryAsRead(s);

Sigmoid	**double** x	**return** 1.0/(1+Math.exp(-x));
SetBoundaries	-	PostingRate=Math.min(Math.max(0,PostingRate),100/day()); OnlineRate=Math.min(Math.max(0,OnlineRate),5/day()); CommentRate=Math.min(Math.max(0,CommentRate),500/day());

EmptyReadStories	-	**for**(**int** i=0; i<ReadStories.size(); i++) {
		Story s=ReadStories.get(i);
		if((getTime()-s.DateOfPublish)>1*day()){
		ReadStories.remove(s);
		ReadStoriesID.remove(i);
		}
		}

CommentForStory	CommentRate	**if**(uniform(0,1)<(CommentRate/(ReadingRate*OnTimeAverage))){
		UserNumberOfCommentsPostedThisWeek++;
		PostingRate=PostingRate+CommentsEffectOnPostRate;
		OnlineRate=OnlineRate+CommentsEffectOnOnlineRate;
		SetBoundaries();
		}

SetActivity	**double** StoryOpinionDim1
	double StoryOpinionDim2
	double StoryAttractiveness
		if(UserOpinionDim1>=0 && StoryOpinionDim1>=0 && StoryOpinionDim2>=0 && UserOpinionDim2>=0){
		PostingRate=PostingRate+OpinionEffectOnPostRate[0]*(sigmoid(
		UserOpinionDim1*StoryOpinionDim1+
		UserOpinionDim2*StoryOpinionDim2+
		StoryAttractiveness)-0.5);
		OnlineRate=OnlineRate+OppinionEffectOnOnlineRate[0]*(sigmoid(
		UserOpinionDim1*StoryOpinionDim1+
		UserOpinionDim2*StoryOpinionDim2+
		StoryAttractiveness)-0.5);
		
		CommentRate=CommentRate+OppinionEffectOnCommentRate[0]*
		(sigmoid(
		UserOpinionDim1*StoryOpinionDim1+
		UserOpinionDim2*StoryOpinionDim2+
		StoryAttractiveness)-0.5);
		}
		//…
		//same goes for other cases
		SetBoundaries();

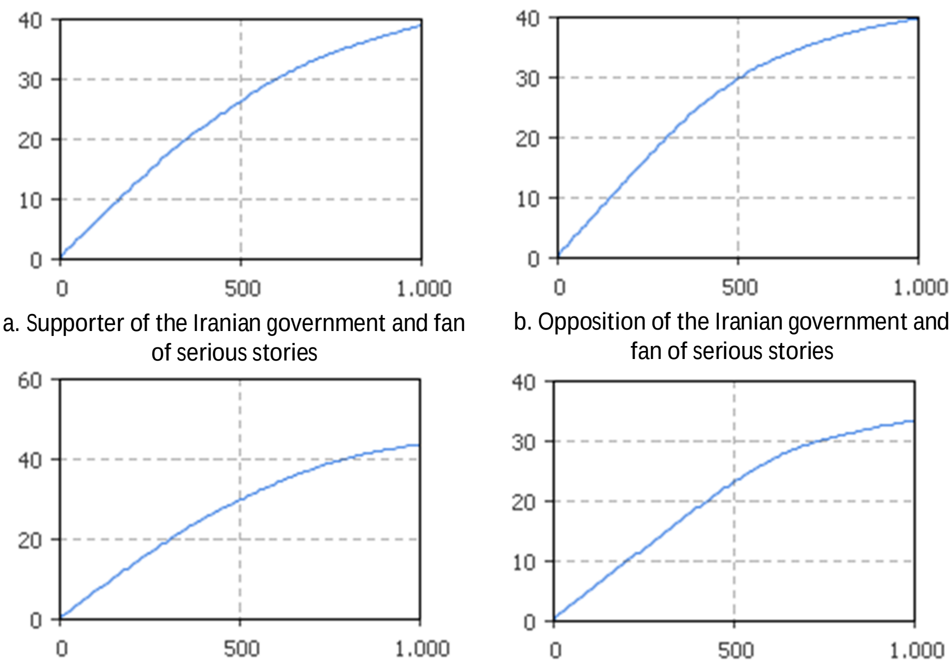
Table 4-5. Procedures/functions and pseudo-codes

The *CommentForStory* procedure determines whether the user will comment on the story he has read by using the following decision making rule:

$$user\ comments\ for\ the\ story : \begin{cases} True, & uniform(0,1) < (CommentRate / (ReadingRate \times OnTimeAverage)) \\ False, & otherwise \end{cases}$$

In the formulation above, $ReadingRate \times OnTimeAverage$ indicates the average number of stories the user reads in each online session and $CommentRate$ is the number of comments s/he publishes on average. As a result, an increase in the *CommentRate* will increase the probability of the user commenting on the story. ($uniform(0,1)$ is a uniformly random generated number (between 0 and 1)).

Appendix 2



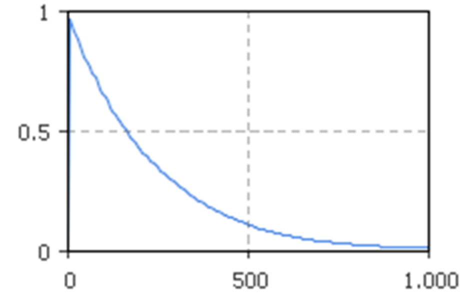
a. Supporter of the Iranian government and fan of serious stories

b. Opposition of the Iranian government and fan of serious stories

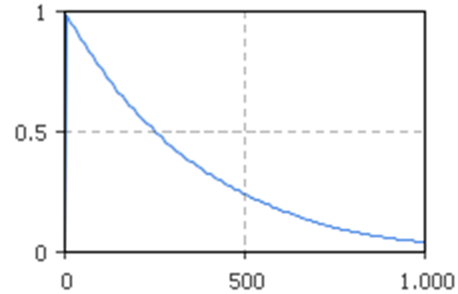
c. Supporter of the Iranian government and fan of yellow stories

d. Opposition of the Iranian government and fan of yellow stories

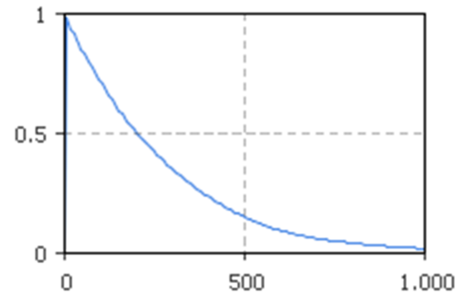
Figure 4-14. Average weekly post rate on each quadrant simulated for a symmetric community over 1000 days



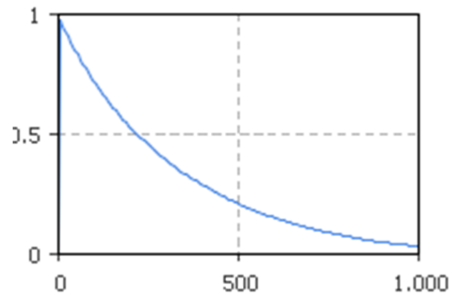
a. Supporter of the Iranian government and fan of serious stories



b. Opposition of the Iranian government and fan of serious stories



c. Supporter of the Iranian government and fan of yellow stories



d. Opposition of the Iranian government and fan of yellow stories

Figure 4-15. Average weekly online rates on each quadrant simulated for a symmetric community over 1000 days

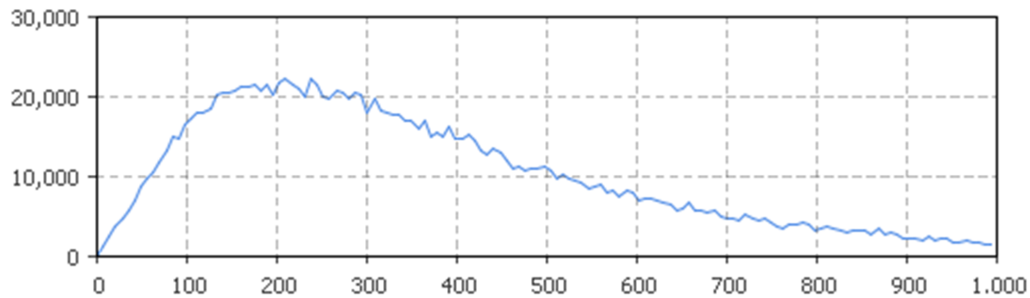
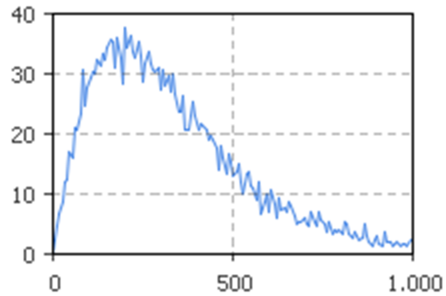
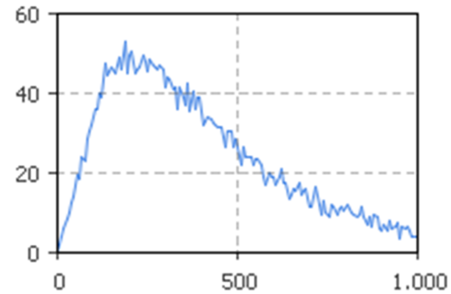


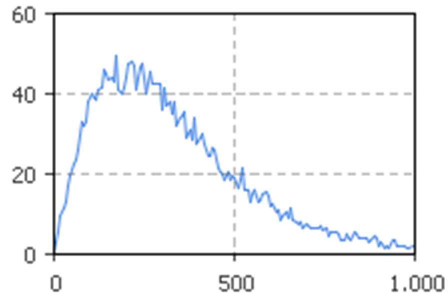
Figure 4-16. Total number of weekly comments simulated for a symmetric community over 1000 days



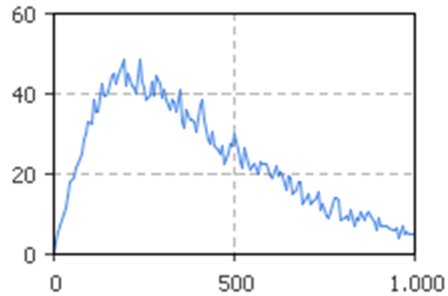
a. Supporter of the Iranian government and fan of serious stories



b. Opposition of the Iranian government and fan of serious stories

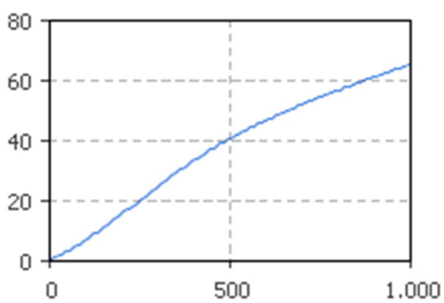


c. Supporter of the Iranian government and fan of yellow stories

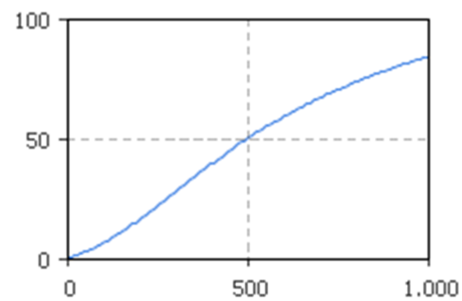


d. Opposition of the Iranian government and fan of yellow stories

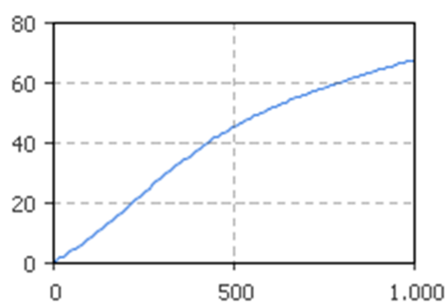
Figure 4-17. Average weekly comments on each quadrant simulated for a symmetric community over 1000 days



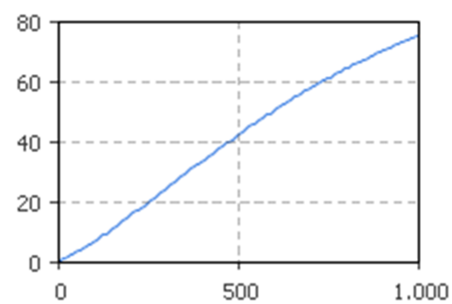
a. Supporter of the Iranian government and fan of serious stories



b. Opposition of the Iranian government and fan of serious stories



c. Supporter of the Iranian government and fan of yellow stories



d. Opposition of the Iranian government and fan of yellow stories

Figure 4-18. Average weekly comment rates on each quadrant simulated for a symmetric community over 1000 days

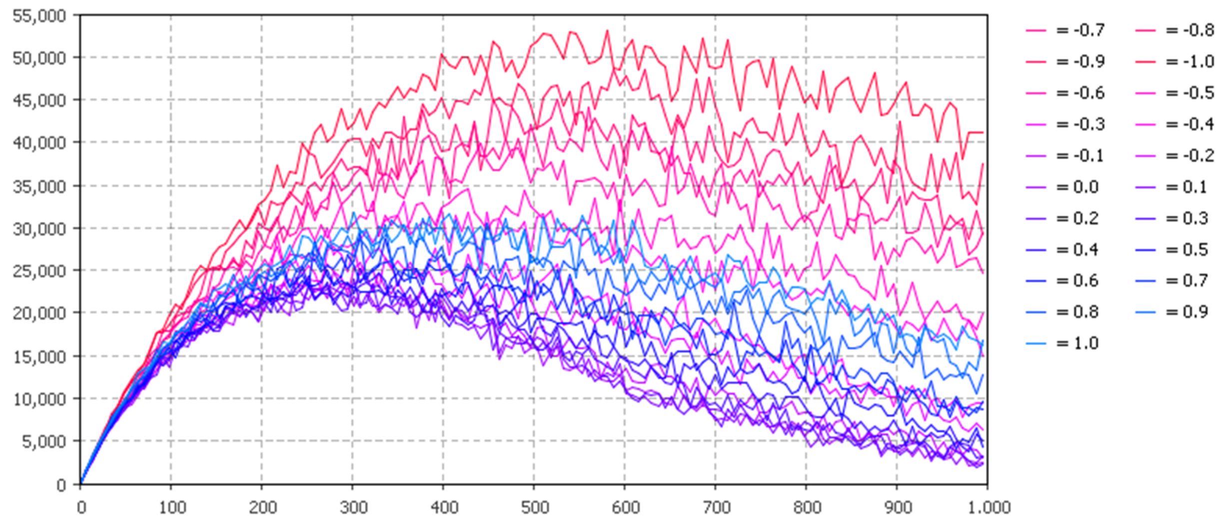
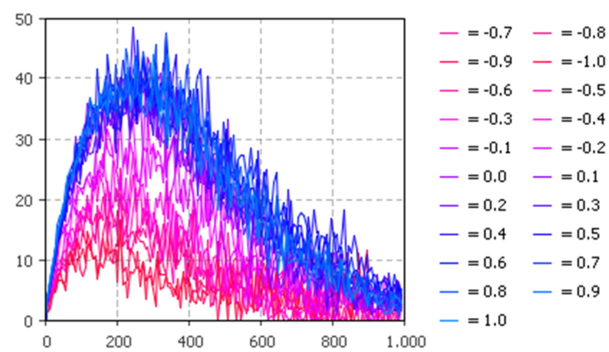
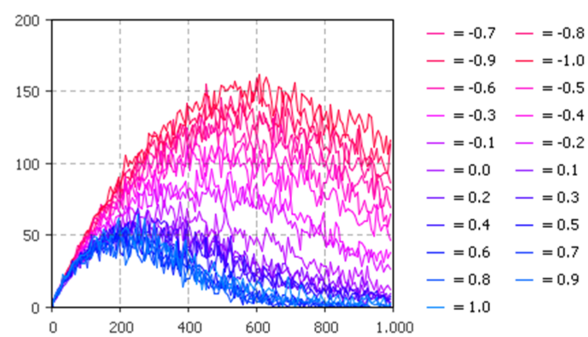


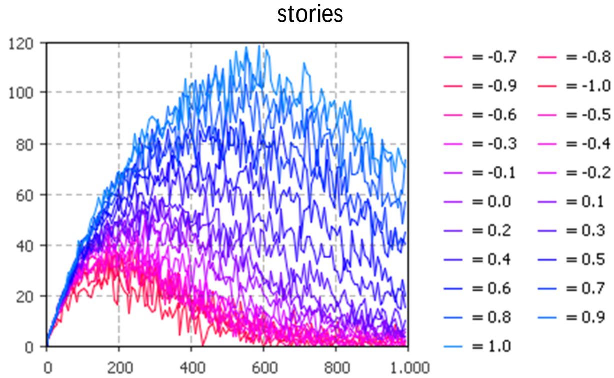
Figure 4-19. Total number of weekly posts simulated for differently biased (in terms of supporting/opposing the Iranian government) communities over 1000 days (red lines for oppositions and blue lines for supporters of the government)



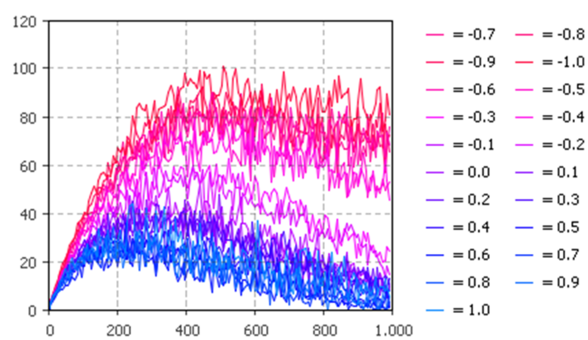
a. Supporter of the Iranian government and fan of serious stories



b. Opposition of the Iranian government and fan of serious stories

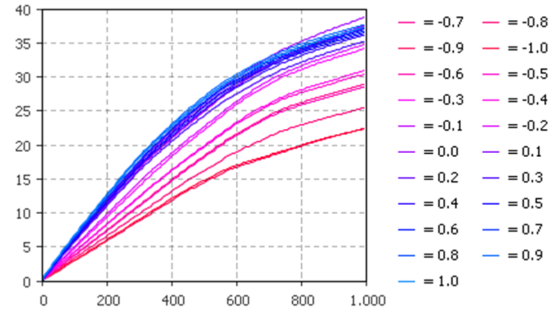


c. Supporter of the Iranian government and fan of yellow stories

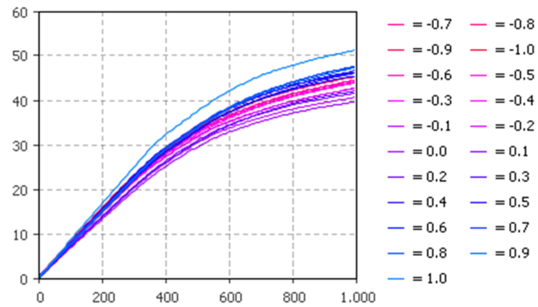


d. Opposition of the Iranian government and fan of yellow stories

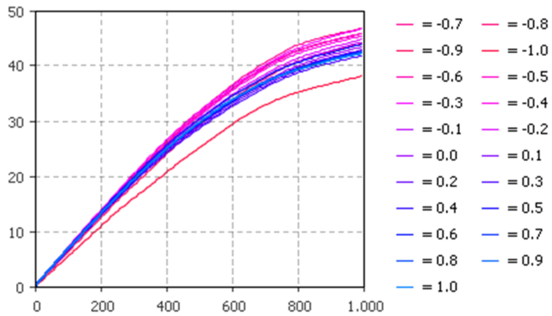
Figure 4-20. Average weekly posts on each quadrant simulated for differently biased (in terms of supporting/opposing the Iranian government) communities over 1000 days (red lines for the Oppositions and blue lines for the supporters of the government)



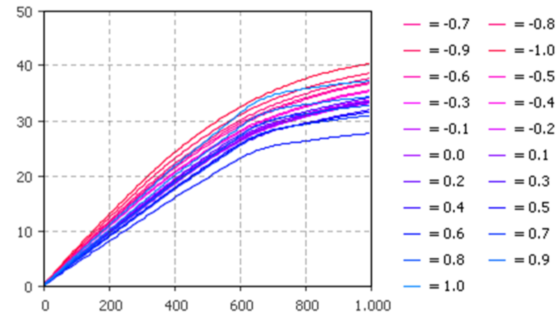
a. Supporter of the Iranian government and fan of serious stories



b. Opposition of the Iranian government and fan of serious stories

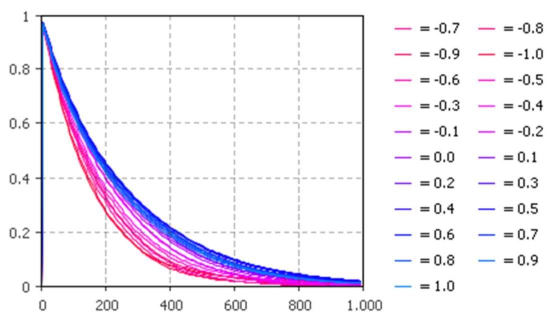


c. Supporter of the Iranian government and fan of yellow stories

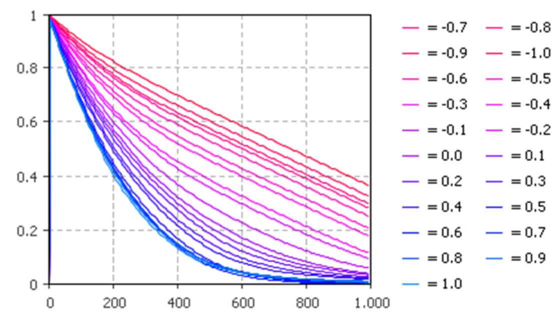


d. Opposition of the Iranian government and fan of yellow stories

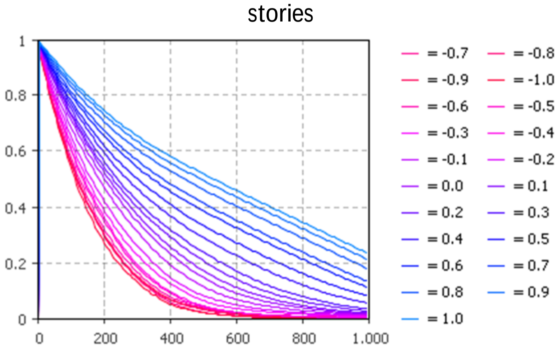
Figure 4-21. Average post rates on each quadrant simulated for differently biased (in terms of supporting/opposing the Iranian government) communities over 1000 days (red lines for the oppositions and blue lines for the supporters of the government)



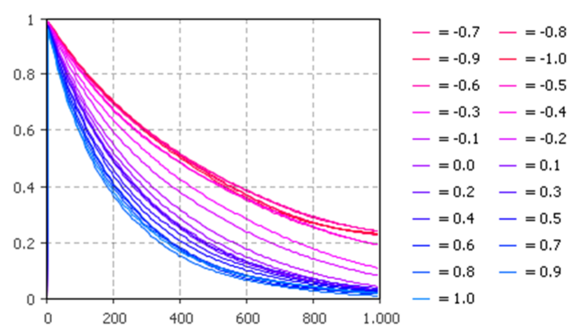
a. Supporter of the Iranian government and fan of serious stories



b. Opposition of the Iranian government and fan of serious stories



c. Supporter of the Iranian government and fan of yellow stories



d. Opposition of the Iranian government and fan of yellow stories

Figure 4-22. Average online rates on each quadrant simulated for differently biased (in terms of supporting/opposing the Iranian government) communities over 1000 days (red lines for the oppositions and blue lines for the supporters of the government)

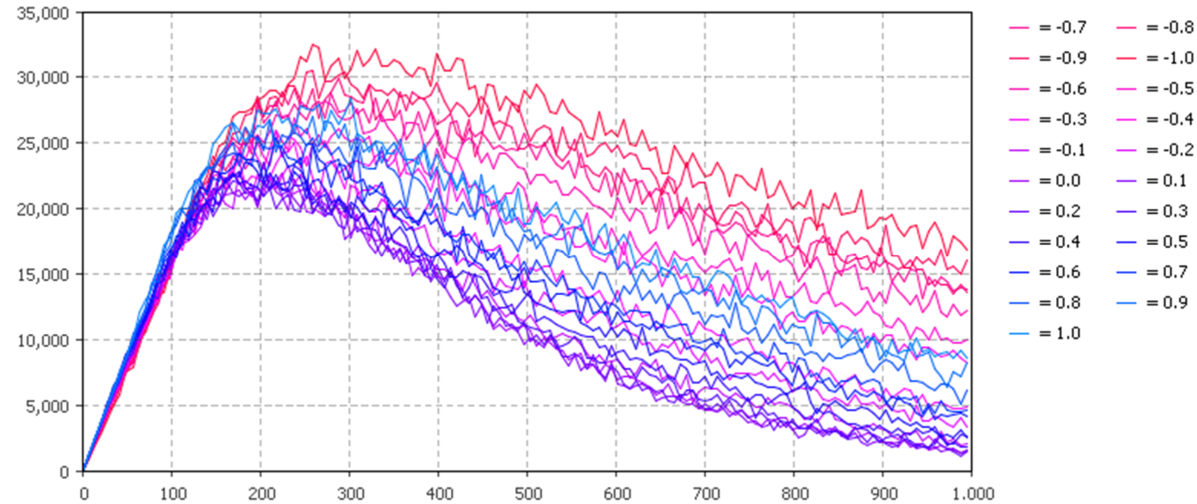
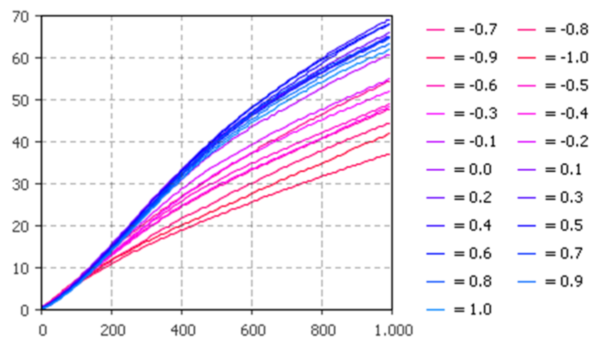
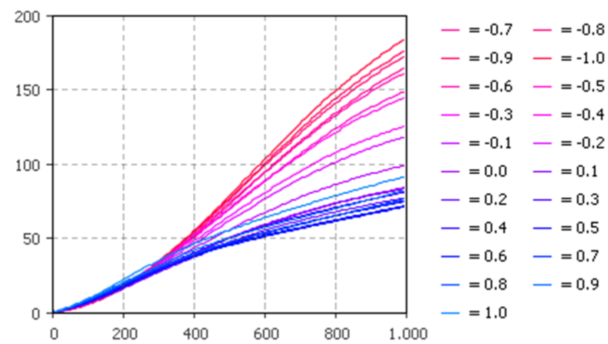


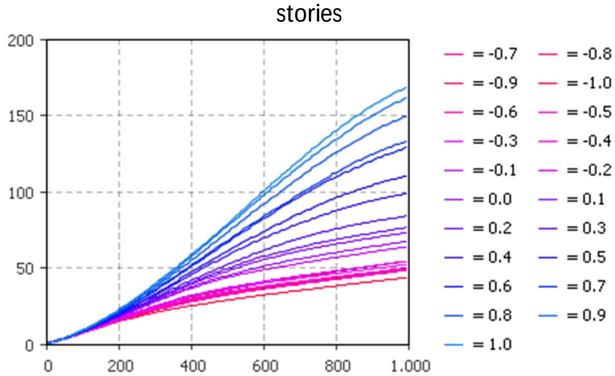
Figure 4-23. Total number of weekly comments simulated for differently biased (in terms of supporting/opposing the Iranian government) communities over 1000 days (red lines for oppositions and blue lines for supporters of the government)



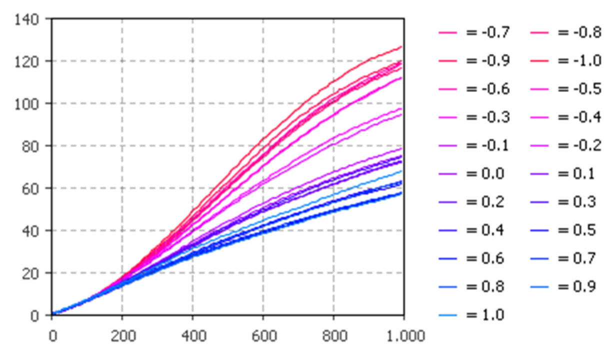
a. Supporter of the Iranian government and fan of serious stories



b. Opposition of the Iranian government and fan of serious stories

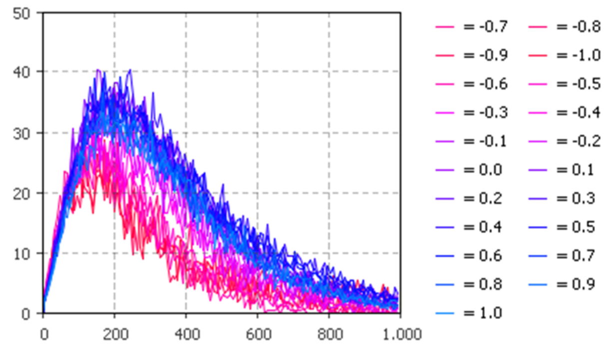


c. Supporter of the Iranian government and fan of yellow stories

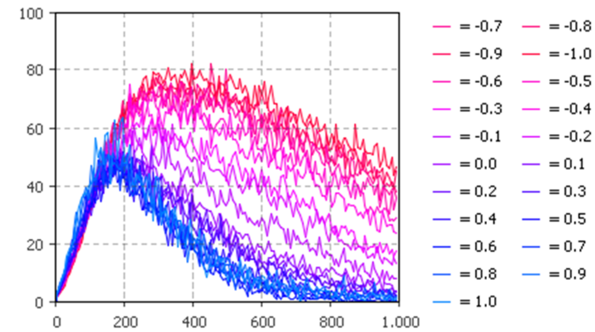


d. Opposition of the Iranian government and fan of yellow stories

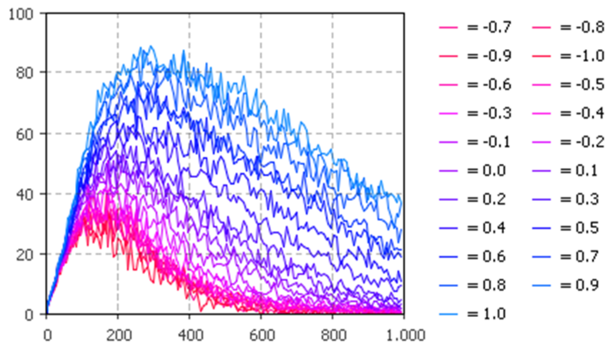
Figure 4-24. Average comment rates on each quadrant simulated for differently biased (in terms of supporting/opposing the Iranian government) communities over 1000 days (red lines for the opponents and blue lines for the supporters of the government)



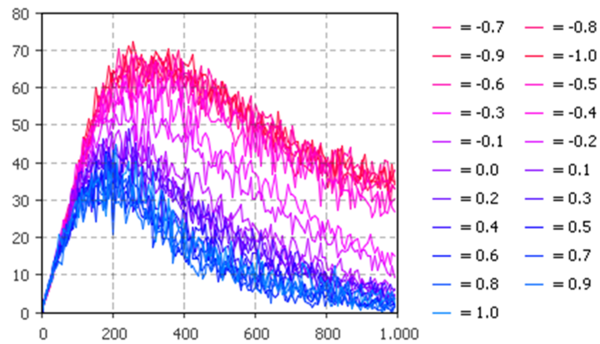
a. Supporter of the Iranian government and fan of serious stories



b. Opposition of the Iranian government and fan of serious stories



c. Supporter of the Iranian government and fan of yellow stories



d. Opposition of the Iranian government and fan of yellow stories

Figure 4-25. Average weekly comments on each quadrant simulated for differently biased (in terms of supporting/opposing the Iranian government) communities over 1000 days (red lines for the oppositions and blue lines for the supporters of the government)

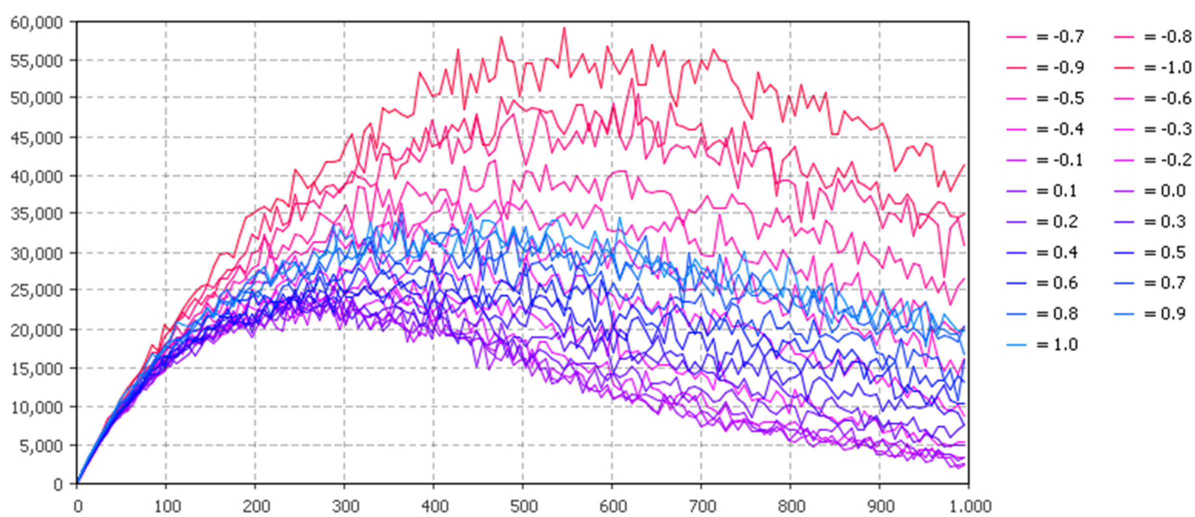
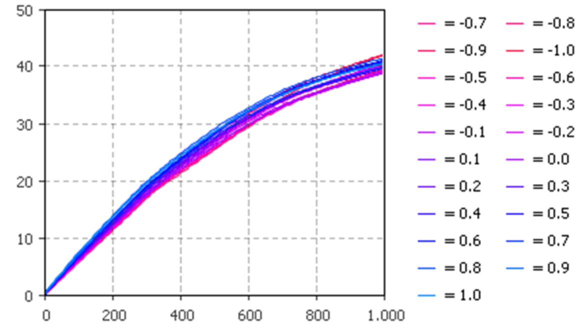
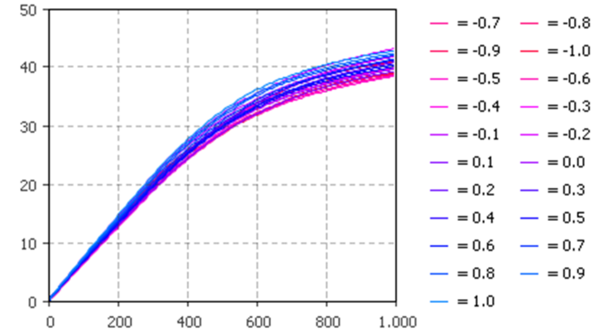


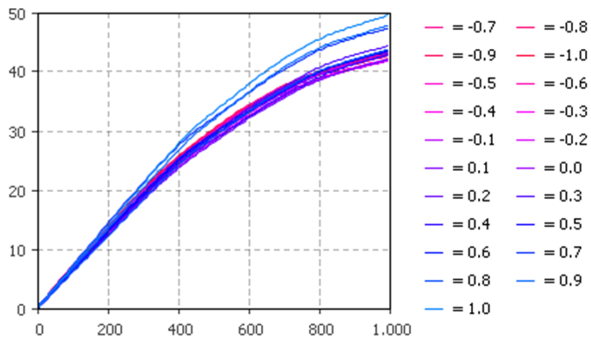
Figure 4-26. Total number of weekly posts simulated for differently biased (in terms of being a fan of yellow/serious stories) communities over 1000 days



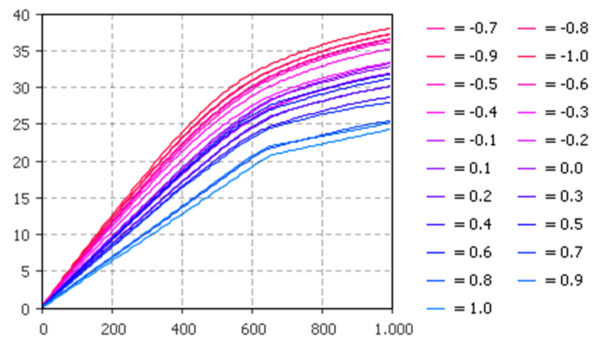
a. Supporter of the Iranian government and fan of serious stories



b. Opposition of the Iranian government and fan of serious stories

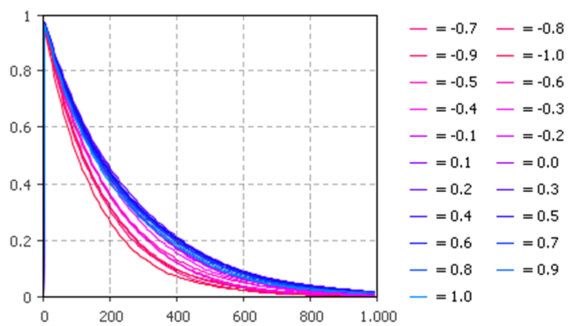


c. Supporter of the Iranian government and fan of yellow stories

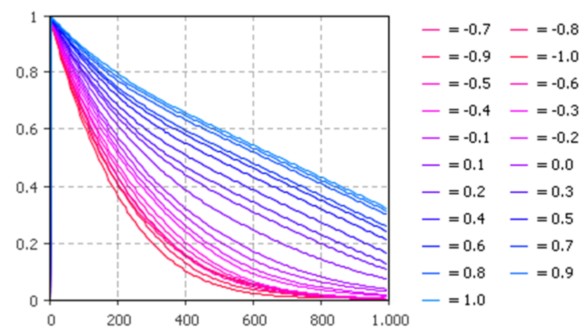


d. Opposition of the Iranian government and fan of yellow stories

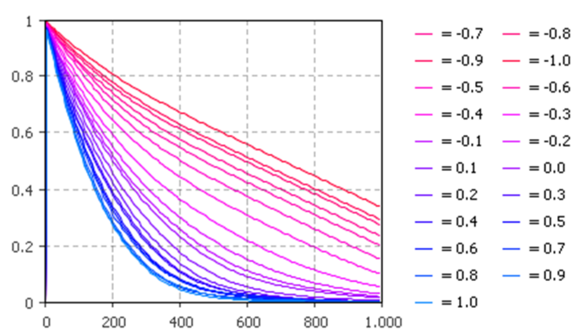
Figure 4-27. Average post rates on each quadrant simulated for differently biased (in terms of being a fan of yellow/serious stories) communities over 1000 days (red lines for fans of yellow stories and blue lines for fans of serious stories)



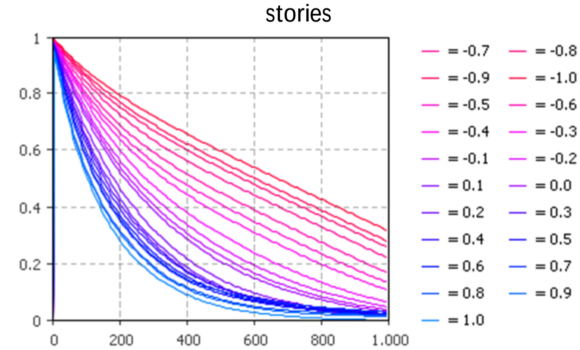
a. Supporter of the Iranian government and fan of serious stories



b. Opposition of the Iranian government and fan of serious stories



c. Supporter of the Iranian government and fan of yellow stories



d. Opposition of the Iranian government and fan of yellow stories

Figure 4-28. Average online rates on each quadrant simulated for differently biased (in terms of being a fan of yellow/serious stories) communities over 1000 days (red lines for fans of yellow stories and blue lines for fans of serious stories)

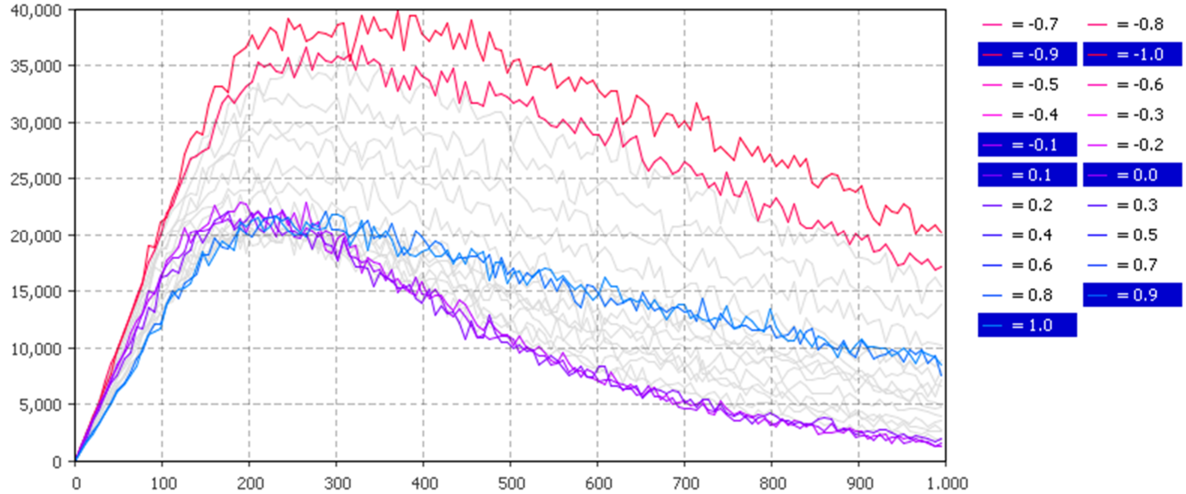
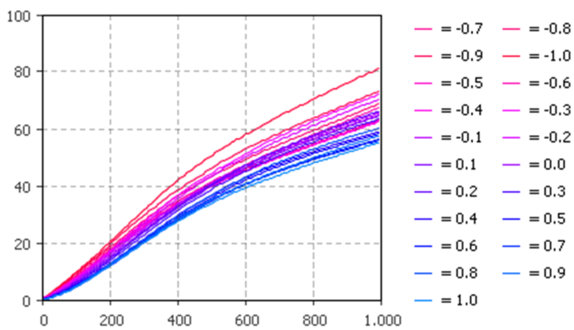
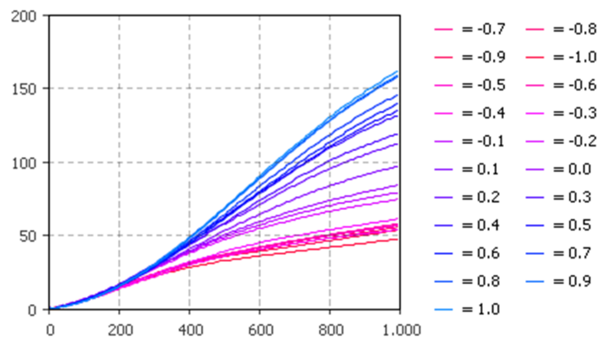


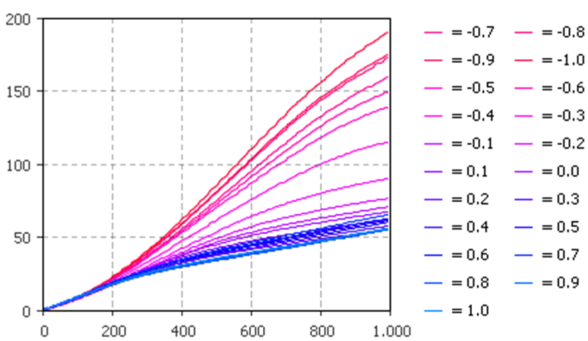
Figure 4-29. Total number of weekly comments simulated for differently biased (in terms of being a fan of yellow/serious stories) communities over 1000 days (red lines for fans of yellow stories and blue lines for fans of serious stories)



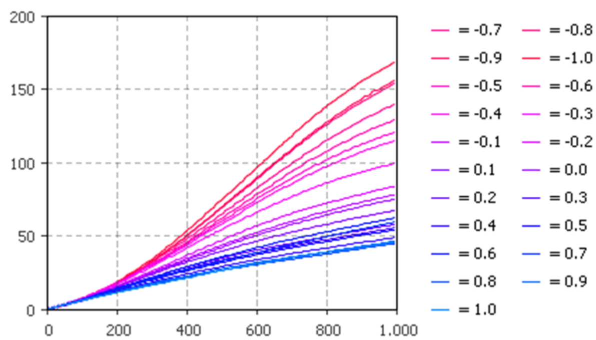
a. Supporter of the Iranian government and fan of serious stories



b. Opposition of the Iranian government and fan of serious stories

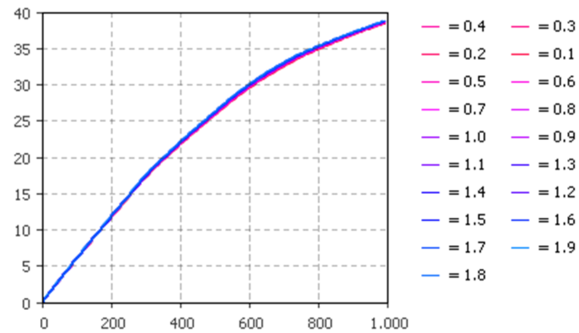


c. Supporter of the Iranian government and fan of yellow stories

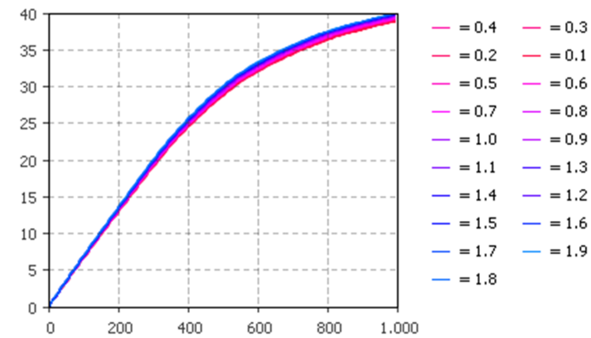


d. Opposition of the Iranian government and fan of yellow stories

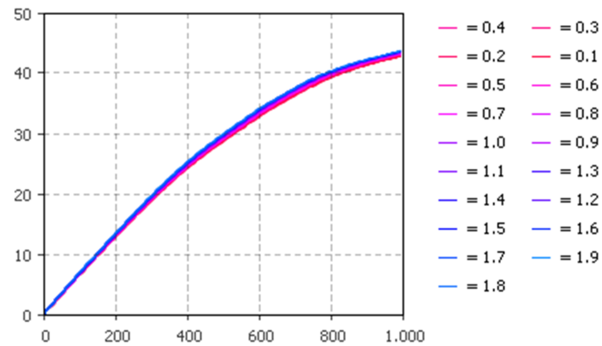
Figure 4-30. Average comment rates on each quadrant simulated for differently biased (in terms of being a fan of yellow/serious stories) communities over 1000 days (red lines for fans of yellow stories and blue lines for fans of serious stories)



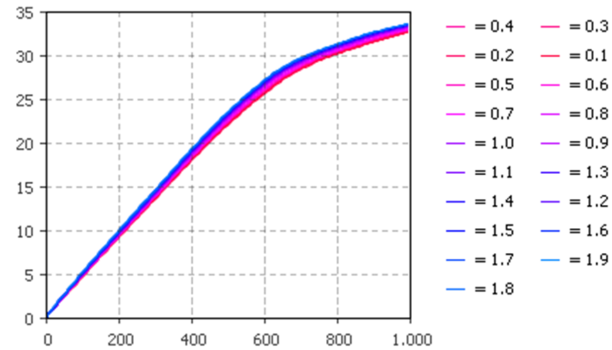
a. Supporter of the Iranian government and fan of serious stories



b. Opposition of the Iranian government and fan of serious stories

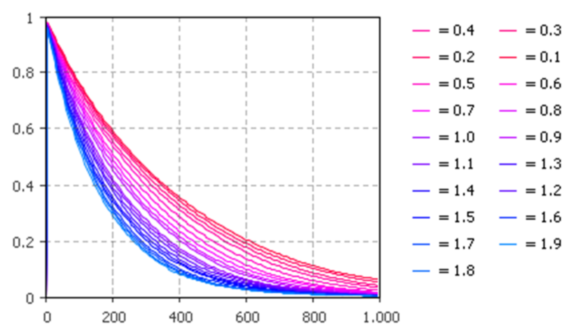


c. Supporter of the Iranian government and fan of yellow stories

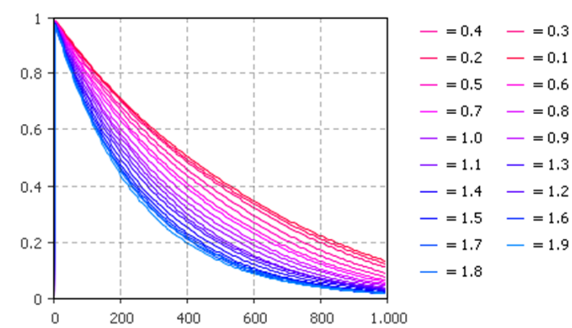


d. Opposition of the Iranian government and fan of yellow stories

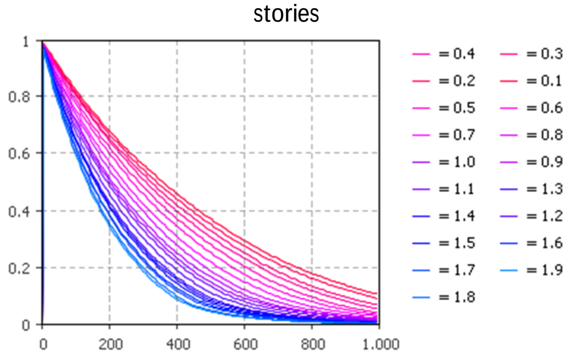
Figure 4-31. Average post rates on each quadrant simulated for neutral/extreme communities (red lines for more neutral communities and blue lines for more extreme communities) over 1000 days



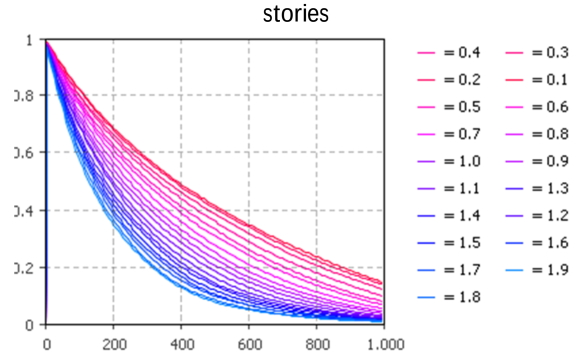
a. Supporter of the Iranian government and fan of serious stories



b. Opposition of the Iranian government and fan of serious stories

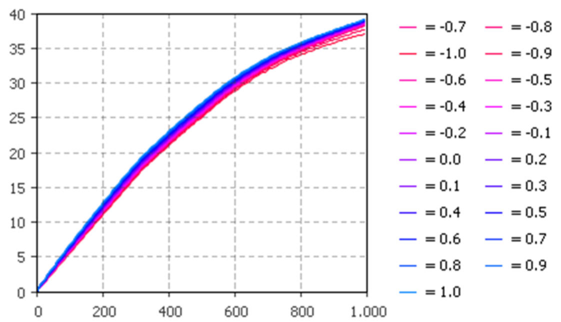


c. Supporter of the Iranian government and fan of yellow stories

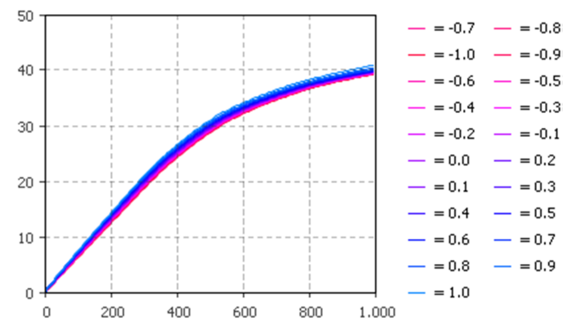


d. Opposition of the Iranian government and fan of yellow stories

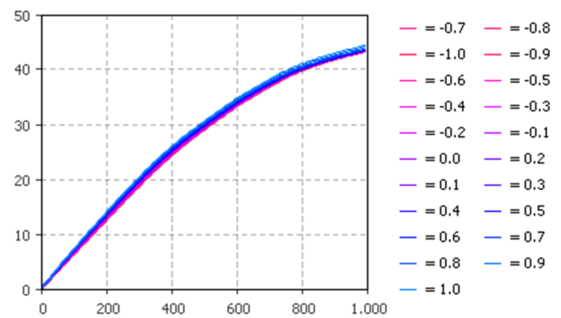
Figure 4-32. Average online rates on each quadrant simulated for neutral/extreme communities (red lines for more neutral communities and blue lines for more extreme communities) over 1000 days



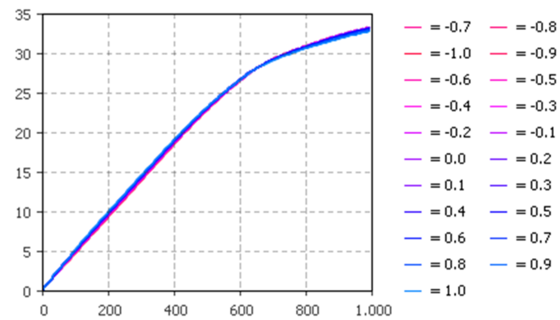
a. Supporter of the Iranian government and fan of serious stories



b. Opposition of the Iranian government and fan of serious stories

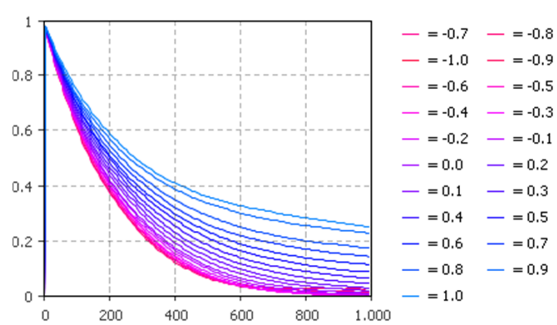


c. Supporter of the Iranian government and fan of yellow stories

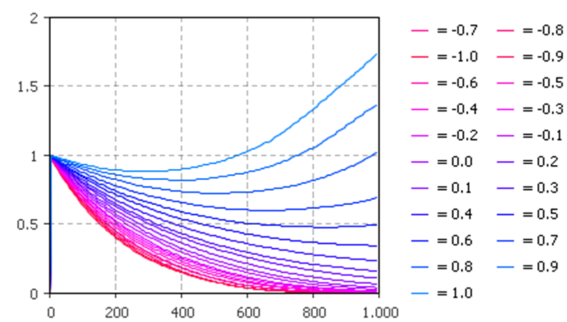


d. Opposition of the Iranian government and fan of yellow stories

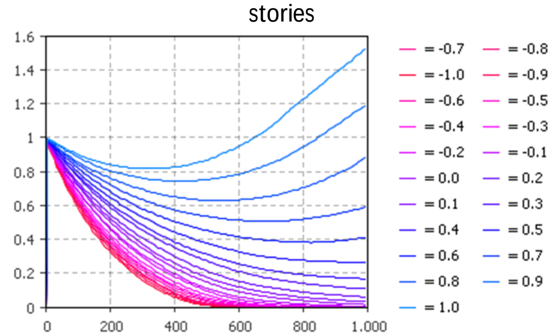
Figure 4-33. Average post rates on each quadrant simulated for communities generating high/low quality contents (red lines for communities generating low quality contents and blue lines for communities generating high quality contents) over 1000 days



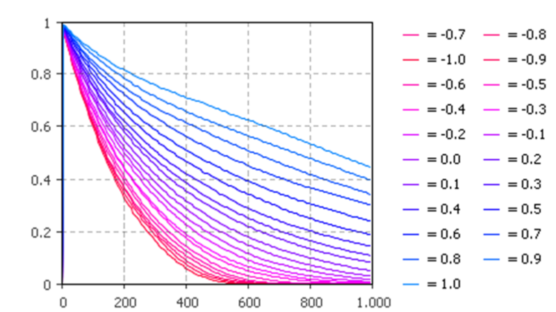
a. Supporter of the Iranian government and fan of serious stories



b. Opposition of the Iranian government and fan of serious stories

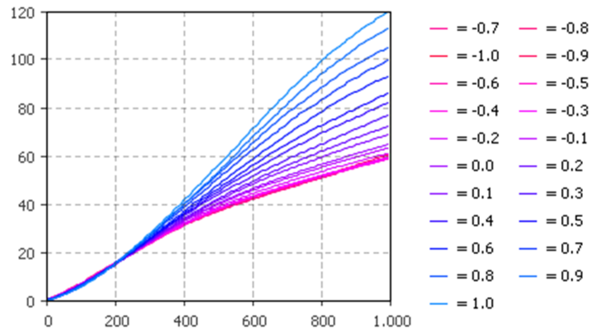


c. Supporter of the Iranian government and fan of yellow stories

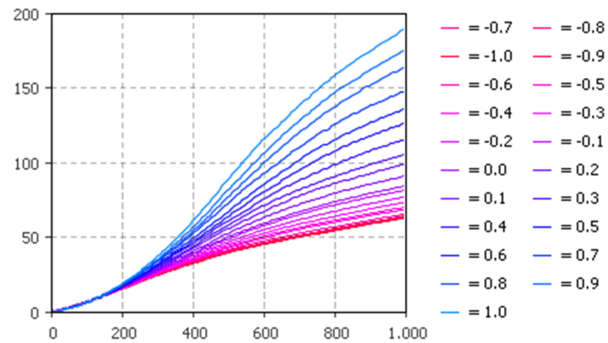


d. Opposition of the Iranian government and fan of yellow stories

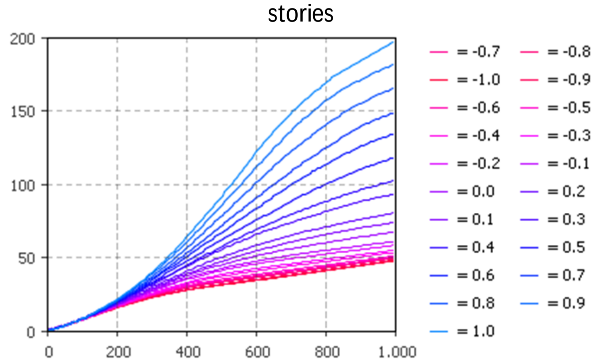
Figure 4-34. Average online rates on each quadrant simulated for communities generating high/low quality contents (red lines for communities generating low quality contents and blue lines for communities generating high quality contents) over 1000 days



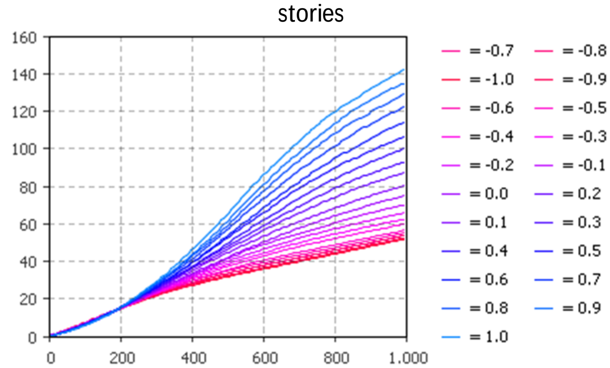
a. Supporter of the Iranian government and fan of serious stories



b. Opposition of the Iranian government and fan of serious stories



c. Supporter of the Iranian government and fan of yellow stories



d. Opposition of the Iranian government and fan of yellow stories

Figure 4-35. Average comment rates on each quadrant simulated for communities generating high/low quality contents (red lines for communities generating low quality contents and blue lines for communities generating high quality contents) over 1000 days

References:

1. Cannarella, J. and J.A. Spechler, *Epidemiological modeling of online social network dynamics.* arXiv preprint arXiv:1401.4208, 2014.
2. Torkjazi, M., R. Rejaie, and W. Willinger. *Hot today, gone tomorrow: on the migration of MySpace users*. in *Proceedings of the 2nd ACM workshop on Online social networks*. 2009. ACM.
3. Wu, S., et al. *Arrival and departure dynamics in social networks*. in *Proceedings of the sixth ACM international conference on Web search and data mining*. 2013. ACM.
4. Giliette, F., *The rise and inglorious fall of myspace.* 2011.
5. Garcia, D., P. Mavrodiev, and F. Schweitzer. *Social resilience in online communities: The autopsy of friendster*. in *Proceedings of the first ACM conference on Online social networks*. 2013. ACM.
6. Rad, A.A. and H. Rahmandad, *Reconstructing online behaviors by effort minimization*, in *Social Computing, Behavioral-Cultural Modeling and Prediction*. 2013, Springer. p. 75-82.
7. Bakshy, E., et al. *The role of social networks in information diffusion*. in *Proceedings of the 21st international conference on World Wide Web*. 2012. ACM.
8. Jalali, M.S., et al., *Information diffusion through social networks: The case of an online petition.* Expert Systems with Applications, 2016. **44**: p. 187-197.
9. Ashoury, A., O. Herrera-Restrepo, and H. Zhang. *The Diffusion of a Petition in a Social Network*. in *63rd Annual Conference and Expo of the Institute of Industrial Engineers*. 2013.
10. Tucker, C.E., *Social networks, personalized advertising, and privacy controls.* Journal of Marketing Research, 2014. **51**(5): p. 546-562.
11. Trusov, M., R.E. Bucklin, and K. Pauwels, *Effects of word-of-mouth versus traditional marketing: findings from an internet social networking site.* Journal of marketing, 2009. **73**(5): p. 90-102.
12. Haythornthwaite, C., *Social networks and Internet connectivity effects.* Information, Community & Society, 2005. **8**(2): p. 125-147.
13. Rad, A.A., H. Rahmandad, and M. Yahyanejad, *Estimating the Dynamics of Individual Opinions in Online Communities.*
14. Shore, J. *Social Media Distractions Cost U.S. Economy $650 Billion* 2012; Available from: http://mashable.com/2012/11/02/social-media-work-productivity/#3uiH6pYA5iqU.
15. Campbell, M.A., *Cyber Bullying: An Old Problem in a New Guise?* Australian journal of Guidance and Counselling, 2005. **15**(01): p. 68-76.
16. Skeels, M.M. and J. Grudin. *When social networks cross boundaries: a case study of workplace use of facebook and linkedin*. in *Proceedings of the ACM 2009 international conference on Supporting group work*. 2009. ACM.
17. Gose, C. *Internal communications: the rise and fall of internal social networks*. 2013; Available from: http://www.rmgnetworks.com/blog/bid/215425/Internal-communications-the-rise-and-fall-of-internal-social-networks.
18. Wadee, Z. *Facebook your boss: using social media in internal communications*. 2013; Available from: http://www.theguardian.com/careers/careers-blog/facebook-employers-encourage-social-media.
19. Parekh, R. *Internal Affairs: Social Media at the Office*. 2012; Available from: http://adage.com/article/digital/internal-affairs-social-media-office/237207/.

5. Conclusion

As a society, social media impacts our daily lives in ways that we could have never imagined before. In a short time tools such as fax machines and cassettes that had been widely used for leading revolutions like fall of Berlin walls (1989) and Iran revolution (1979) replaced by social media in Iranians Green Revolution (2009) [1] and Arab Spring (2010) [2]. In response to disasters, nonprofits used social media to mobilize rescue efforts after Haiti earthquake (2010) [3] and to collect donation after tsunami in Japan (2011) [4]. People used Safety Check feature provided by Facebook to let their friends know they are safe after Nepal earthquake (2015) [5] and Paris attacks (2015). Social media also highly affecting elections in the U.S. and around the world. In the U.S. 2012 presidential election Obama widely used social media to organize the supporters [6] and spent more than $40 million in digital campaigning. In countries like Iran and Egypt, where political choice and speech are limited, organizers use social media to attract supporters. Companies use social media for marketing and making personal connection with the customers. Nabisco for instance introduced their Oreo Instagram account on a Super Bowl commercial (costing on average $4 million for a 30 second spot) in 2013. Oreo's social media marketing goes to the extent that, having a team for real-time advertisement for the Super Bowl, they responded in less than 5 minutes to Super Bowl blackout in 2013 advertising their product on twitter (Figure 5-1). Startups use social media for fundraising (i.e. crowdfunding) and finding their target population (e.g. kickstarter.com, Indiegogo.com, gofundme.com, etc.). In a way, social media makes a more level playing field for the small businesses.

Figure 5-1. Online advertisement examples

In the past decade, researchers from different fields studied many aspect of social media, from its structures [7] to its effects on individuals' behavior and its influences on the global society [8, 9]. Yet, because of the complexity exists in such huge outlets, connecting the dots between different aspects of social media remained as a gap in the literature. System science provides a framework for dealing with such complexity. By simplifying the reality and dividing it in small modules, system science helps us to mathematically model each module. It helps us to understand the relation between the modules and the effect modules have on each other (and on the system as a whole) through feedback loops. It provides us with tools and methods to quantify these relations and study the consequence of change in baseline, each module or parameter on the system's outcome. Here we aimed to connect the dots and study the dynamics of individuals' behavior in a social media considering the structure of the platform, using the system science approach. We studied the effect of interacting with variety of opinions exist in a social media community, on social media users' behavior and activities; then investigated mechanisms that could affect the lifecycle and the future opinion formation of the community. To do so, we divided the problem in three main parts: 1) extracting the data, 2) measuring individuals' opinion (attitude) and 3) deriving the dynamics and empirically modeling the system (i.e. social media). We provided methods and techniques to deal with each of the sub problems in a separate essay.

For the first essay (chapter 2) we proposed a procedure to derive individuals' online behavior from generic user-object interaction data available from many social media platforms. We used the method to study the behavior and preferences of a social news website (i.e. Balatarin) users and to extract time based users browsing data (e.g. users' online time in each session, stories they read in a time interval, etc.) and objects (stories) property data (e.g. number of votes stories had at each point in time, promotion status of stories, etc.) we needed for the next step. We developed an optimization approach that estimates user browsing behavior (e.g. users' location and stories s/he read at each point in time) based on the idea that people conserve their efforts (e.g. number of clicks or scrolling to reach a story they read) in their online browsing. Proposed procedure could improve predictive power of tools (e.g. recommendation systems) and models (e.g. collaborative filtering models) that are based on users browsing or objects properties data. It also allows researchers to convert user-object interaction data from binary vote/no-vote type to vote/no-vote/not seen type which also enhances the predictive power of models that feed on user-object interaction data (e.g. collaborative filtering and many other content recommendation models). The method can also be applied in variety of problems that deal with browsing behaviors such as online ad positioning, enhancing website designs, etc. Proposed method is limited to the cases where time based user-object interaction data is available. For social media in which an object can show up in many (e.g. thousands) locations the method could be computationally expensive. For cases in which social media ranking algorithm contains random choices or considers previous viewed objects, the method would suffer lack of accuracy. Coding implementation for the method could be non-trivial for cases that consider many different factors in content ranking algorithm.

In the second essay (chapter 3) we proposed a method that implicitly and empirically measures individuals' attitude (i.e. opinion) toward different issues based on their interaction data in social media. The method uses user-object interaction data (i.e. users votes or rate on online objects) to map individuals (and objects) on multidimensional opinion space. The method comes with techniques to derive underlying concepts of the opinion dimensions. It also controls the effect of other factors that can affect the ratings. Using the data we extracted and collected from Balatarin using our proposed history reconstruction procedure in the first essay, we applied the method on Balatarin users and mapped their opinion on a 2-dimensional space through time. We validated the method using a survey in which subjects categorized selected stories from Balatarin in different groups of opinion. Our results show high test-retest correlation and high inter-rater value. Proposed method is free of response-latency errors, it derives individuals' attitude generalizing his/her responses in a time period so it is less sensitive to the context and is more robust toward prior exposure. The method helps researchers to measure individuals' attitude using available data on social media and as a result highly reduces the cost of data collection. The method is scalable and also gives researcher the ability to evaluate individuals' attitude in time series manner. Proposed method can help researcher to study the change and formation of individuals' attitude through time. Researchers can also use the method to study the reaction of individuals in interacting with other opinions in social media. However, proposed method can only measure attitude on issues that have effect on ratings in the social media. Besides, the method depends on the social media data, therefore, it is limited to those subjects that actively use such outlets.

In the third essay (chapter 4) we studied the effect of individuals' interaction with different opinions in social media on their online activities. We developed regression models that estimate the change in online rate, posting rate and commenting rate of individuals based on the opinion of objects they interacted with. Based on the data we extracted from Balatarin using our history reconstruction procedure (chapter 2) and the opinion values we estimated based on our attitude mapping method (chapter 3), we evaluated the changes in Balatarin users activity using the proposed regression models. We then studied the mechanisms caused by these changes and developing an agent based simulation model we estimated the future formation and lifecycle of Balatarin. Finally we investigated the effect of biased online communities on the social media users' activity and future form of the outlet. The results imply that reactions to opinions varies based on the subject and the opinion of individual on the subject. Interacting with likeminded people increases online activities in general, however, viewing extreme contents could decrease the activity. The result of our study signifies the importance of feeding users with contents close to their opinion in keeping the social media active. The result also indicates that different opinion groups react differently in interacting with each other. Consequently, to keep users' activity optimum, social media platforms have to feed different opinion groups with variety of contents respective to their reactions. Finally based on our results, filtering low quality contents and those with extreme opinions could extend the lifecycle of the outlet.

References:

1. Burns, A. and B. Eltham, *Twitter free Iran: An evaluation of Twitter's role in public diplomacy and information operations in Iran's 2009 election crisis.* 2009.
2. Howard, P.N., et al., *Opening closed regimes: what was the role of social media during the Arab Spring?* Available at SSRN 2595096, 2011.
3. Yates, D. and S. Paquette, *Emergency knowledge management and social media technologies: A case study of the 2010 Haitian earthquake.* International Journal of Information Management, 2011. **31**(1): p. 6-13.
4. Gao, H., G. Barbier, and R. Goolsby, *Harnessing the crowdsourcing power of social media for disaster relief.* IEEE Intelligent Systems, 2011(3): p. 10-14.
5. Brengarth, L.B. and E. Mujkic, *WEB 2.0: How social media applications leverage nonprofit responses during a wildfire crisis.* Computers in Human Behavior, 2016. **54**: p. 589-596.
6. West, D.M., *Air Wars: Television Advertising and Social Media in Election Campaigns, 1952-2012*. 2013: Sage.
7. Kietzmann, J.H., et al., *Social media? Get serious! Understanding the functional building blocks of social media.* Business horizons, 2011. **54**(3): p. 241-251.
8. Kaplan, A.M. and M. Haenlein, *Users of the world, unite! The challenges and opportunities of Social Media.* Business horizons, 2010. **53**(1): p. 59-68.
9. Segerberg, A. and W.L. Bennett, *Social media and the organization of collective action: Using Twitter to explore the ecologies of two climate change protests.* The Communication Review, 2011. **14**(3): p. 197-215.