Three Facets of Online Political Networks: Communities, Antagonisms, and Polarization

by

Mert Ozer

A Dissertation Presented in Partial Fulfillment
of the Requirement for the Degree
Doctor of Philosophy

Approved September 2019 by the
Graduate Supervisory Committee:

Hasan Davulcu, Chair
Huan Liu
Arunabha Sen
Yezhou Yang

ARIZONA STATE UNIVERSITY

December 2019

ABSTRACT

Millions of users leave digital traces of their political engagements on social media platforms every day. Users form networks of interactions, produce textual content, like and share each others' content. This creates an invaluable opportunity to better understand the political engagements of internet users. In this proposal, I present three algorithmic solutions to three facets of online political networks; namely, detection of communities, antagonisms and the impact of certain types of accounts on political polarization. First, I develop a multi-view community detection algorithm to find politically pure communities. I find that word usage among other content types (i.e. hashtags, URLs) complement user interactions the best in accurately detecting communities.

Second, I focus on detecting negative linkages between politically motivated social media users. Major social media platforms do not facilitate their users with built-in negative interaction options. However, many political network analysis tasks rely on not only positive but also negative linkages. Here, I present the SocLSFact framework to detect negative linkages among social media users. It utilizes three pieces of information; sentiment cues of textual interactions, positive interactions, and socially balanced triads. I evaluate the contribution of each three aspects in negative link detection performance on multiple tasks.

Third, I propose an experimental setup that quantifies the polarization impact of automated accounts on Twitter retweet networks. I focus on a dataset of tragic Parkland shooting event and its aftermath. I show that when automated accounts are removed from the retweet network the network polarization decrease significantly, while a same number of accounts to the automated accounts are removed randomly the difference is not significant. I also find that prominent predictors of engagement of automatically generated content is not very different than what previous studies point

out in general engaging content on social media. Last but not least, I identify accounts which self-disclose their automated nature in their profile by using expressions such as bot, chat-bot, or robot. I find that human engagement to self-disclosing accounts compared to non-disclosing automated accounts is much smaller. This observational finding can motivate further efforts into automated account detection research to prevent their unintended impact.

ACKNOWLEDGMENTS

First and foremost, I would like to thank my advisor Dr. Hasan Davulcu, who trusted me and supported me in all of my endeavours during these last 5 years. His endless energy and result oriented approach motivated me to deliver for every milestone of this journey. I am also thankful to my committee members; Dr. Huan Liu, Arunabha Sen, and Yezhou Yang. The classes I have taken from them and their valuable feedback significantly helped me to shape this dissertation.

I can not be more grateful for my parents Bahriye, and Tayfur Ozer, and my little sister Goknil Ozer. They were always a Skype call away to share the happiness of achievements and to uplift when things were not looking so promising. Without their endless passion for education and progress, I would not be where I am right now.

I have worked with many talented colleagues; Saud Alashri, Sultan Alzahrani, Swetha Baskaran, Niharika Bollapragada, Betul Ceran, Pankaj Chabra, Anuj Ghandi, Chinmay Gore, and Nyunsu Kim. Discussing research ideas with Amin Salehi and seeing him working in his cubicle every day was a major motivating factor during these years. Finally, I would like to thank my dearest colleague Mehmet Yigit Yildirim. Without him, without his collaboration and endless support, this journey would not succeed the way it is now. I will miss our daylong study sessions in the coffee shops and libraries all over the Southwest.

Last but not least, I would like to give the most special thanks to Candice Eisenfeld and Sydney. I am the luckiest person to meet you on my first days in Arizona. Without your endless love, support, and goofy jokes I would not be able to succeed in my Ph.D. journey. Cheers to happier days ahead.

TABLE OF CONTENTS

Page

CHAPTER Page

CHAPTER Page

LIST OF TABLES

LIST OF FIGURES

Figure Page

Chapter 1

ONLINE POLITICAL NETWORKS & OVERVIEW OF MY CONTRIBUTION

1.1 Introduction

Politics, by its definition, is a relational phenomenon (Victor *et al.*, 2016). With the recent explosive growth of social media platforms, large amounts of relational data have become available for researchers. The availability of large scale data for digitized political involvement opens an array of dimensions to study (Barberá, 2015; M Bond *et al.*, 2012; Conover *et al.*, 2011a). In this proposal I present my contributions on the three aspects of online political networks; namely community, negative link detection, and measuring the impact of automated accounts on political polarization. I develop novel algorithms for community detection(Ozer *et al.*, 2016) and negative link detection(Ozer *et al.*, 2017, 2018) tasks separately, and design set of experiments on observational data to detect the impact of automated accounts on political polarization. Out of the two major social media platforms, I had to limit the experimental validation of my efforts to Twitter datasets. Unfortunately, accessing Facebook data is not straightforward (Tufekci, 2014).

1.1.1 Community Detection

Community detection is a fundamental task in social network analysis (Girvan and Newman, 2002). A community (Girvan and Newman, 2002) can be defined as a group of users that (1) interact with each other more frequently than with those outside the group and (2) are more similar to each other than to those outside the group. Utilizing community detection algorithms to detect online political camps has

attracted many researchers (Tang *et al.*, 2012; Sachan *et al.*, 2012; Ruan *et al.*, 2013). In this work, I propose three non-negative matrix factorization frameworks to exploit both user connectivity and content information in Twitter to find ideologically pure communities in terms of their members' political orientations.

Twitter presents three types of connectivity information between users: follow, retweet and user mention. In this dissertation, I do not use follow information since follow relationships correspond to longer-term structural bonds (Myers *et al.*, 2014) and it remains challenging to determine if a follow relationship between a pair of users indicate political support or opposition. Furthermore, it has been observed that neither user retweets nor user mentions always indicate endorsement in Twitter (Tufekci, 2014). However in the political sub-domain of Twitter, it has been shown that retweets tend to happen between like-minded users rather than between members of opposing camps (Conover *et al.*, 2011a).

Using both connectivity and content information for community detection in social networks has been a popular approach among many researchers' prior works (Pei *et al.*, 2015; Ruan *et al.*, 2013; Sachan *et al.*, 2012; Tang *et al.*, 2012). In (Tang *et al.*, 2012), Tang et al. propose a general framework for integrating multiple heterogeneous data sources for community detection. Tang's work does not pay attention to identifying the endorsement subgraph of the connectivity graph. In (Sachan *et al.*, 2012) Sachan et al. propose an LDA-like social interaction model by representing user connectivity as a document alongside message content. This approach also does not discriminate between positive or negative user engagement. In (Ruan *et al.*, 2013), Ruan et al. propose to use a filtered graph to eliminate ambiguous interactions by checking content similarity in the user's neighborhood. In this formulation, only local content patterns are taken into consideration whereas in my formulations I incorporate the global content patterns into my optimization framework.

Pei et al. in (Pei *et al.*, 2015) also model the problem as nonnegative matrix tri-factorization problem which factorizes user-word, tweet-word and user-user matrices into lower rank representations of users and tweets while regularizing it with user interaction and message similarity matrices. They build user-user connectivity matrix by utilizing the structural follow relationships which do not capture dynamic political context-sensitive engagement. They treat all user mentions and retweets identically and without any discrimination for endorsement. Their framework also lacks word similarity regularization.

I develop and experiment with three nonnegative matrix factorization frameworks: MultiNMF, TriNMF, DualNMF, which incorporate connectivity alongside different types of content information as regularizers. After experimenting with different dimensions of user content and different types of induced connectivity networks I discovered that incorporating more information does not necessarily yield higher clustering performance. Highest quality clustering is achieved through endorsement filtered connectivity based on methods I develop in Section 2.1 alongside user-word matrix based content regularization. My DualNMF framework gives purity scores around 88%, adjusted rand index around 75% and NMI around 67%. It improves all of the other baseline methods significantly as presented in Section 2.3 and it also improves over the NMTF framework developed recently by Pei et al. (Pei *et al.*, 2015) by 8% in purity, 47% in ARI and by up to 60% in NMI metrics. Proposed endorsement filtered sub-graph of user mentions and retweets also improves all baseline methods in almost all of the experimental setups by up to 109% in NMI, 71% in ARI and 17% in purity.

The contributions of this proposal in community detection task can be summarized as follows:

- I start with *retweets without edits* as indicators of positive endorsements between users and utilize Heider's P-O-X triad balance theory (Heider, 1958) to

incorporate selected "structurally balanced" *edited retweets* and *user mentions* into a weighted undirected connectivity graph as additional indicators of positive endorsements.

- I develop algorithms which incorporate users' content information in my community detection frameworks to overcome the sparse nature of Twitter connectivity networks. I break down Twitter message content into three categories; words, hashtags and URLs, and design experiments to measure the performance contributions of each category. Proposed Non-negative Matrix Factorization (NMF) algorithms use user-word, user-hashtag and user-domain matrices to be factorized into lower rank user vector representations while regularizing over user connectivity and content similarity to map users into their respective communities.

1.1.2 Implicit Negative Link Detection

Beyond any doubt, social media has become a prominent platform for people to express their political stances and opinions for more than a decade (Ausserhofer and Maireder, 2013). It developed into a medium for politicians and political organizations to interact with the public (NACOS, 2013). To name a few, 44th President of the United States, Barack Obama makes an appearance on a Reddit Ask Me Anything, 45th President Donald Trump constantly utilises Twitter for his political messaging, many grassroot organizations mobilise their movements on Twitter and Facebook. Consequently, online social networks more and more are becoming an active field of study for political analysis tasks.

Many researchers have extensively studied the nature of online political networks (Conover *et al.*, 2011a,c; Johnson and Goldwasser, 2016; Ozer *et al.*, 2016). Most of

the existing works utilise platform-specific positive interactions between users such as share and like in Facebook or retweet and like in Twitter to infer insights from and model political activities in such social media platforms. (Conover *et al.*, 2011a) present how platform-specific positive interactions in Twitter shows a polarised behaviour in which one side does not retweet or like the other side's contents.

Major online social media platforms, however, do not provide its users options to state negative opinions in the form of a simple click such as "dislike" which might convey opposition or disagreement towards each other. Nonetheless, many political analysis tasks need the information of rivalries, resentments between political actors to get a complete picture of the online political landscape. This very nature of major social media platforms limit the capabilities of researchers studying online political networks effectively. Many researchers usually choose to study the online social networks where explicit negative links are available to them such as Epinions, Slashdot or Wikipedia instead (DuBois *et al.*, 2011; Leskovec *et al.*, 2010a; Yang *et al.*, 2012). Certainly, these online platforms are not the hotspots where people participate to express their political views through.

Therefore, I focus on inferring the implicit negative links between users of online political networks. I aim to detect the link's negative nature, when any form of an overall disagreement, opposition or hostility is present between two social media users. It is a challenging problem due to two main reasons. First, there is no readily available online political network dataset in which negative links are explicitly present between its users. Therefore, the developed model must be unsupervised. Second, there is no simple predictor of negative links such as "dislike" in major social media platforms where the main body of the online political activity resides. However, opportunities are unequivocally present as well. Recent works in the social media mining research (Tang *et al.*, 2015; Liu *et al.*, 2016) show that negative sentiment in

the textual interaction between users is a good predictor of the negative link of those two users. Moreover, certain social psychology phenomenons such as social balance or social status theory are proven to be helpful in predicting negative links in certain network configurations(Leskovec *et al.*, 2010b).

In this work, I first propose a nonnegative matrix factorization framework SocLS-Fact that combines signals from sentiment lexicon of words, platform-specific positive interactions and social balance theory to detect implicit negative and positive links in online political networks. I do not focus on the accuracy of the positive links since it is already a well studied problem and simple good predictors are already available. Additionally, I extend my SocLS-Fact framework to online settings to allow the integration and analysis of newly acquired data in a computationally efficient manner. Through this extension, it becomes convenient to run SocLS-Fact on a much smaller dataset without compromising effectiveness by utilizing previously detected implicit links to calibrate the model.

I discuss two applications where detected implicit negative links can be employed to give a better understanding of the underlying political configuration of the target dataset. The first application is presented to show the added value of the detected implicit negative links in community detection tasks. The second application is proposed to show the informativeness of the detected implicit negative links related to polarisation patterns between political groups.

The main contributions of the paper are,

- Proposing SocLS-Fact an unsupervised model for implicit negative link detection in social media platforms where platform-specific negative interactions or negative links between users are not present.

- Introducing an online extension for SocLS-Fact to dynamically incorporate newly observed data while refraining from retraining the whole dataset.
- Showing the added value of the negative links in community detection tasks for online political networks.
- Providing two human-annotated online political network datasets for further research interest.

1.1.3 Measuring the Polarization Impact of Automated Accounts

Social media has been one of the most prominent mediums in political communication for the last decade. Its wide accessibility, ease of use, and reach out capacity have attracted millions to participate in political discussions in these socio-technical systems. People organized protest movements (Theocharis *et al.*, 2015; Varol *et al.*, 2014), toppled down authoritarian regimes (Tufekci and Wilson, 2012) with social media in their action toolkit. Social media has also become instrumental in campaigning for underrepresented issues and communities with hashtag activism. It lead to a stronger voice and awareness in mainstream media as in the cases of #metoo (Manikonda *et al.*, 2018), and #blacklivesmatter (Carney, 2016).

The act of retweeting undeniably plays a crucial role in these information dissemination and campaign building processes on Twitter (Boyd *et al.*, 2010). When Twitter users want to re-post or share some other users' content in their own profile, they simply use the retweet functionality of the platform. This simple mechanism has given users capability to share posts of others they like with their own followers. Its use, however, has reached beyond a simple intent to share when analyzed at broader scale. Many scientific arguments have been hypothesized around this functionality. Scholars present numerous anecdotal findings showing the connection between the use

of this functionality and the political homophily (Colleoni *et al.*, 2014; Conover *et al.*, 2011b) whereas Barbera et al. (Barber *et al.*, 2015) present evidence for the cases where cross-camp interactions are present such as in the Boston bombing, Winter olympics, and Super Bowl events. However, the overall polarization in major political issues on Twitter has risen between 10% and 20% (Garimella and Weber, 2017) over the last decade.

Positive and controversial aspects aside, wide accessibility of social media also attract malicious use of these platforms at multiple levels. Researchers and data journalists investigate and report several cases of abuse including but not limited to the issues of cyberbullying (Hosseinmardi *et al.*, 2015), anti-vaccination (Mitra *et al.*, 2016), ISIS propaganda (Farwell, 2014), and white supremacist (O'Callaghan *et al.*, 2013) propaganda. In the majority of the misuse cases, automated accounts (a.k.a. bots) are found to be playing a significant role as well (Broniatowski *et al.*, 2018; Ferrara, 2017; Stella *et al.*, 2018). In this work, I focus on automated accounts' role in political polarization on Twitter retweet networks. I quantify the polarization impact that automated accounts induce and its textual, emotional, and behavioral correlates. To the best of my knowledge, this is the first work that tackles the problem of measuring the polarization impact of automated activity on social media.

My investigation is in two-folds; on synthetically generated networks and on a real-world social media network. First, I set up synthetically generated network scenarios to evaluate the robustness of my experimentation logic. Second I focus on a Twitter dataset that span the time period of tragic Parkland school shooting and its aftermath. In synthetic scenario, I (1) produce polarized user networks, (2) quantify the polarization, and (3) measure the impact of random node removals on polarization. I find no evidence that random removals significantly affect the polarization measurements on synthetically generated networks. This finding motivates us to em-

ploy a similar removal experiment on Twitter dataset. Using the Twitter dataset, I build retweet networks at hashtag and aggregate levels. I show that removing automated accounts from retweet networks significantly reduces the polarization of retweet networks while random removals do not.

Polarization effect in retweet network is prevalent due to the automatically generated content's appeal without any doubt. If automated activity was not getting any traction from other users, I would not observe any significant change in the polarization of the retweet network. To carry out an investigation on textual and user profile characteristic correlates of engagement that automated accounts attract, I develop a zero inflated negative binomial regression task on retweet count. I compare the predictors with earlier studies of engaging content on social media and find that the engagement correlates of automated account tweets are closely overlapping with previous findings. I also find that the use of the word "they" play a positive role in gaining higher retweet counts which, to the best of my knowledge, was not explored before.

Lastly, I conduct a similar removal experiment with only self-identifying automated accounts. I find that the polarization impact vanishes on the automated accounts that self-identify their automated nature in their profiles. These self-identifications are apparent to human users either at profile name or *screen_name* level. Pairwise engagement ratio of human controlled accounts with self-identifying automated accounts (0.1154) is overtly lower than the engagement with undisclosed automated accounts (1.3001). I believe that this observational finding can further motivate the efforts put into automated activity detection on social media research in alleviating their unintended impact.

1.2 Previous Literature

1.2.1 Community Detection

Since the introduction of the modularity metric by Newman in (Newman, 2006), plenty of modularity based community detection methods have been proposed in the literature (Fortunato, 2010; Blondel *et al.*, 2008; Clauset *et al.*, 2004; Waltman and van Eck, 2013). I employ Blondel et al. (Blondel *et al.*, 2008) and Clauset et al. (Clauset *et al.*, 2004) works as baseline algorithms to compare with mine due to their wide popularity among practitioners. A general drawback of these algorithms, when they are applied to Twitter networks, is that due to the sparse nature of the connectivity they end up with an artificially large number of communities.

Non-negative Matrix Factorization

Non-negative Matrix Factorization(NMF) algorithms by Lee et al. (Lee and Seung, 2000) and Lin et al. (Lin, 2007) have been extensively used and extended for different variations of community detection problems. Cai et al. (Cai *et al.*, 2011) introduced GNMF algorithm to incorporate Laplacian graph regularization to the standard NMF algorithm which assumes data points are sampled from a Euclidean space which is not the case usually for real-world applications. Gu et al. (Gu and Zhou, 2009) further incorporate local learning regularization to NMF which assumes that geometrically neighboring data points are similar to each other, and should be in the same cluster. For co-clustering purposes Ding et al. (Ding *et al.*, 2006) propose non-negative matrix tri-factorization with orthogonality constraints. Shang et al. introduce graph dual regularized NMF algorithm in (Shang *et al.*, 2012) by claiming that not only observed data but also features lie on a manifold.

1.2.2 Implicit Negative Link Detection

I survey link prediction, sentiment classification and dynamic network modeling methods proposed for similar line of research to mine in social media mining literature.

Link prediction in social media is an extensively studied problem. Its precedings can be traced back to the structuralist social psychology studies (Heider, 1958) that became popular in early 20th century. Link prediction studies standing out as most related to my problem definition are (Kunegis *et al.*, 2013; Leskovec *et al.*, 2010a; Tang *et al.*, 2015; Yang *et al.*, 2012). (Leskovec *et al.*, 2010a) propose a framework that predicts the sign of user links in online networks. They train classifiers using certain triad configuration and graph features to learn from existing data in which both explicit positive and negative links are present. (Yang *et al.*, 2012) make use of explicit negative links through items that users comment to rather than using direct negative links between users. Signed bipartite graph of users and items is used to infer connectivity patterns among users. In their prediction model, they accommodate the principles of balance and status from social psychology theory.

However, these methods are not capable of being trained for major social media platforms (i.e. Twitter, Facebook) due to the nonexistence of explicit negative links or platform-specific negative interaction capabilities of users in those platforms. To address this limitation, (Kunegis *et al.*, 2013) present an approach to predict negative links when only positive links are available explicitly. They further investigate the added value of negative links when they are predictable to a certain extent by using only properties of the positive links and not using any additional information such as textual content. However, they experiment only with Slashdot and Epinions datasets in which negative links or interactions between users are explicitly available. How generalizable their approach for other major social media platforms such as Face-

book or Twitter, in which no platform-specific negative interaction is available, is not discussed. In (Tang *et al.*, 2015), Tang et al. introduce a supervised classification scheme to predict the negative links among missing links assuming that in many social media platforms, negative links are indirect and implicit. They use negative sentiment polarity of textual interactions between user pairs to synthetically generate the negative labelled links. This method also relies on experiments conducted only on Slashdot and Epinions datasets. On the other hand, my framework stands out as it is proposed for the social media platforms that do not provide any platform-specific negative interaction capabilities to their users.

Second line of research related to my work is sentiment classification in social media. (Hu *et al.*, 2013b) propose a supervised sentiment classification model which takes advantage of connected text messages having similar sentiment labels. (Hu *et al.*, 2013a) further investigate whether emotional signals such as emoticons can be incorporated in order to infer the sentiment classes of the tweets in Twitter. To credit the informative value of the overall sentiment of the textual interactions between users for predicting the polarity of the user link, (Hassan *et al.*, 2012) propose a supervised classification framework. It considers all textual interactions of the user pairs' and learn relevant sentiment features from human annotated prior user link polarities. However, it does not use any platform-specific interaction types which are vastly available on many social media platforms. (West *et al.*, 2014) develop a model that combinatorially optimises the agreement between the sentiment class of user pairs' textual interaction and the polarity label of the explicit user link. They make use of Wikipedia, and U.S. Congress dataset, in which explicit negative links or platform specific negative interactions are available. My work differentiates itself from aforementioned others in the literature by using platform-specific positive interactions, a sentiment lexicon of words and socially balanced triads for detection.

The last line of research related to my work is dynamic network modeling methods. (Aktunc *et al.*, 2015) propose an extension of well-known modularity based smart local moving algorithm for dynamic networks. The main goal of the modeling is community detection. The work concerns with only explicitly defined links on networks. (Mankad and Michailidis, 2013) propose a non-negative matrix factorization approach for modeling dynamic networks. They also utilise the concept of temporal smoothing as my online framework does. However they do not take any other form of interaction in networks other than explicit links. On a similar line of research, (Yu *et al.*, 2017) propose modeling dynamics of networks by using temporal matrix factorization. They investigate the modeling options of temporal unfolding of networks by factorizing different snapshots of networks into one constant and one time-varying matrix. Similar to my modeling, they also use a decay function to weight importance of previous snapshots in temporal order. For predicting links in dynamic networks, (Zhu *et al.*, 2016) propose using a temporal latent space. They assume two users who are located closer in a temporal latent space is likely to form link in the next snapshot. The concept of temporal smoothing also plays an important role in their dynamic network modeling. My work stands out from aforementioned four works by focusing on implicit links rather than explicit ones as majority of social media platforms do not allow explicit negative links. Moreover, these previous efforts do not involve incorporating textual interactions, sentiment signals or social balance theory.

1.2.3 Measuring the Polarization Impact of Automated Accounts

My work is inspired by the previous research on online networks, political polarization, and prevalence and impact of automated activity on social media platforms. In this section, I briefly discuss my connection points to previous literature on these three subjects.

In the Logic of Connective Action (Bennett and Segerberg, 2012), Segerberg et al. suggest explaining action networks in three broad categories; self-organizing connective action networks, organizationally enabled connective action networks, and organizationally brokered collective action networks. They characterize these three network types on a spectrum of organizational coordination, one extreme being little to no organizational coordination and the other extreme being strong to full organizational coordination. Recent history has seen an upsurge of the first two types partially thanks to the advancement of communication technologies. As discussed by numerous scholars previously(Cleaver, 1998; Castells, 1996; Halavais and Garrido, 2003), Zapatista movement of early 90s epitomize the connective action phenomenon. In (Halavais and Garrido, 2003), Garrido et al. characterize EZLN's online network presence (`http://www.ezln.org/`) and how it shapes the international support network of the group.

Since then, researchers conduct multitudes of observational studies around connective action networks. Agarwal et al. (Agarwal *et al.*, 2014) analyze the role of Twitter in occupy protests by suggesting a theoretical framework and analyzing Twitter streams during the protests. Several other studies present evidence for the positive role and pitfalls of social media use in community building(Waitoa, 2013; Clark, 2014; Vigil-Hayes *et al.*, 2017). Tufekci points out the fragile and ephemeral nature of these social media fueled connective action networks in multiple anecdotes (Tufekci, 2017). She marks the impetus role of social media in large crowds getting together without as much effort as it would take with more traditional grassroots campaigning. Regardless of its role in connective action, new information technologies are proven to have a measurable impact through natural experiments on voter turn-out (M Bond *et al.*, 2012), political choice of undecided voters (Epstein and Robertson, 2015), or money donation (Bimber, 2001).

With the advent of social media and its relevancy in political communication, a phenomenon called automated account (a.k.a. bot account) came under a spotlight. Recently, numerous studies disclose their existence and impact on social media. Varol et al. (Varol *et al.*, 2017) characterize the detection of these types of accounts and their interactions with human controlled accounts on Twitter. Stella et al. (Stella *et al.*, 2018) study the automated accounts' behavior in the 2017 Catalan independence referendum. They show that automated accounts deliberately target central hubs with inflammatory content for traction from the general public. Ferrara et al. (Ferrara, 2017) disclose a flock of social bots in a misinformation campaign during the 2017 presidential election season in France. Shao et al. (Shao *et al.*, 2018) present evidence of higher activity by automated accounts in spreading low credibility news sources on Twitter. Very recently, Lou et al. (Lou *et al.*, 2019) develop a model of information spreading with agents having limited attention and how automated activity can easily overshadow deliberate democratic exchange of information on these social media platforms.

Retweeting is a widely adopted action form of Twitter users in the platform for encouraging political participation such as donation or protest (Boyd *et al.*, 2010). Political polarization in this behavior has been found to be an imminent component. Early studies by Adamic et al. (Adamic and Glance, 2005) explore the divided nature of republican and democrat blogs in the blogosphere. Along the same vein, Conover et al. (Conover *et al.*, 2011b) identify the polarized nature of retweet networks of Twitter among democrats and republicans in the U.S.. Weber et al. (Weber *et al.*, 2013) also identify a similar polarized behavior among secularist and islamist Twitter users in Egypt. To quantify the level of polarization, Garimella et al. (Garimella *et al.*, 2018) suggest a random walk based polarization metric for political hashtags in retweet networks of Twitter.

I study these three prevalent phenomenon; namely, online networks (connective action), political polarization, and automated activity jointly on Twitter in the unfolding of Parkland school shooting event. I aim to measure the impact of automated activity on polarization of the endorsement (retweet without edit) networks.

Chapter 2

COMMUNITY DETECTION

2.1 Structural Balance of Retweet and Mention Graph

Since P-O-X triad balance theory proposed by Heider in (Heider, 1958), structural balance of signed networks has been studied extensively. Heider proposed that in a signed triad, only two combinations of eight possible sign configurations are possible for a triad to be structurally balanced. Those are the following cases;

1. three positive edges,

2. one positive and a pair of negative edges.

In other words, there cannot be any structurally balanced triad having only one negative edge. I adopt this social theory for the Twitter user connectivity networks, by assuming that "retweets without edits" imply political endorsement or an unambiguous positive edge (Wong *et al.*, 2016). However, when a retweet is edited, it has already been shown that (Boyd *et al.*, 2010), it does not necessarily mean endorsement anymore. Moreover, user mentions do not imply endorsements either. For these reasons, I only consider retweets without edits as positive edges. For the rest of the user actions, corresponding to retweets with edits and users mentions, it is hard to detect positivity or negativity of the edges.

In certain triad configurations, retweets with edits and user mentions can be identified as positive edges with the help of Heider's triad structural balance (TSB) rules. Since I do not have unambiguous negative edges, the second case is not applicable. However, since I have some positive edges to begin with, I can employ Heider's first

case (i.e. three positive edges), to infer that in the presence of a triad with a pair of positive edges, the third edge can also be labeled as positive. An example configuration with a pair of positive edges is shown in Figure 2.1. In this case, TSB rule is applicable and would allow us to infer that any user mention or retweet with an edit edge connecting the lower pair of users in the triad is indeed a positive edge. By employing this inference mechanism I identify the endorsement filtered user connectivity network.

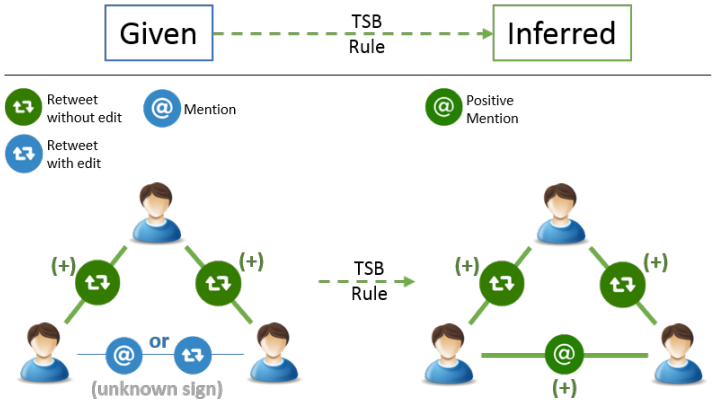


Figure 2.1: An Example Application of TSB Rule. Inferring a Positive Link Between Two Users If They are Both Connected to a Third User with a Positive Link.

2.2 Proposed Methods

I propose three methods for clustering politically motivated users in Twitter namely; MultiNMF, TriNMF and DualNMF. For MultiNMF method I use document term representation of user-word, user-hashtag and user-domain matrices to be factorized and regularize the factorization problem with the user connectivity graph, cosine similarity matrices of words, domains and hashtag co-occurrence matrix. For TriNMF method I use only user-word and one of user-hashtag or user-domain ma-

trices and regularize over user connectivity and cosine domain similarity or hashtag co-occurence matrix. For DualNMF method I factorize user-word matrix into two non-negative lower rank matrices while regularizing it with user connectivity and cosine word similarity. Before going into the details of the three algorithms I present notation in Table 2.1. In this work, instead of using only full user retweet and mention

Table 2.1: Notation

$\mathbf{X}_{uw}$	user x word	counts of words used by users
$\mathbf{X}_{uh}$	user x hashtag	counts of hashtags used by users
$\mathbf{X}_{ud}$	user x domain	counts of distinct domains used by users
$\mathbf{R}$	user x user	adj. matrix of retweet without edit graph
$\mathbf{M}$	user x user	adj. matrix of mention and retweet with edit graph
$\Delta\mathbf{M}_w$	user x user	adj. matrix of mentions and retweet with edits completing retweet without edit triads weighted by retweet without edit edges
$\mathbf{C}$	user x user	any combination of user connectivity graphs
$\mathbf{H_{sim}}$	hashtag x hashtag	hashtag co-occurence matrix
$\mathbf{D_{sim}}$	domain x domain	domain similarity matrix
$\mathbf{W_{sim}}$	word x word	word similarity matrix
$\mathbf{U}$	user x cluster	cluster assignment matrix of users
$\mathbf{H}$	hashtag x cluster	cluster assignment matrix of hashtags
$\mathbf{D}$	domain x cluster	cluster assignment matrix of domains
$\mathbf{W}$	word x cluster	cluster assignment matrix of words

network I offer three types of user connectivity regularizers as follows;

- $\mathbf{R} + \mathbf{M}$: It is the adjacency matrix of the full retweet and mention graph. If there exists both retweet and mention edges between two users, weights are summed up.

- $\mathbf{R} + \Delta\mathbf{M}_w$: It is the adjacency matrix of the union of retweet and mention graphs in which mention edges and retweet with edits either complete a missing link in a triad of retweet without edit or already correspond to a retweet without edit edge. The ones that complete a missing link in a triad of retweet without edit are weighted by the multiplication of the weights of two retweet without edit edges in the triad. $\Delta\mathbf{M}_w$ can be defined formally as;

$$\Delta\mathbf{M}_w = \{(i, j, \mathbf{M}_{ij}(\mathbf{R}_{ij} + \sum_{k=1}^{N} \mathbf{R}_{ik}\mathbf{R}_{kj}))\}$$

For word similarity and domain similarity regularizers I make use of cosine similarity. For hashtag similarity I build similarity matrix by making use of co-occurences of hashtags in tweets. If two hashtags occur in the same tweet, I assume that those two hashtags are similar.

2.2.1 MultiNMF with Multiple Regularizers

To incorporate usage of both hashtags and domains of shared url links by users, I propose an NMF framework which has the following objective function;

$$\begin{aligned}
\mathbf{J}_{\mathbf{U},\mathbf{H},\mathbf{D},\mathbf{W}} =& \parallel \mathbf{X}_{uw} - \mathbf{U}\mathbf{W}^T \parallel_F^2 + \parallel \mathbf{X}_{uh} - \mathbf{U}\mathbf{H}^T \parallel_F^2 \\
&+ \parallel \mathbf{X}_{ud} - \mathbf{U}\mathbf{D}^T \parallel_F^2 + \alpha Tr(\mathbf{U}^T L_{\mathbf{C}} \mathbf{U}) \\
&+ \gamma Tr(\mathbf{H}^T L_{\mathbf{H}_{sim}} \mathbf{H}) + \theta Tr(\mathbf{D}^T L_{\mathbf{D}_{sim}} \mathbf{D}) \\
&+ \beta Tr(\mathbf{W}^T L_{\mathbf{W}_{sim}} \mathbf{W}) \\
&s.t. \quad \mathbf{U} \geq 0, \mathbf{H} \geq 0, \mathbf{D} \geq 0, \mathbf{W} \geq 0
\end{aligned} \tag{2.1}$$

where $L_{\mathbf{C}}$ is the Laplacian matrix of adjacency matrix of user connectivity graph defined as $D_{\mathbf{C}} - \mathbf{C}$ and $D_{\mathbf{C}}$ is the matrix which contains the degree of each user node in its diagonals. $L_{\mathbf{H}_{sim}}$, $L_{\mathbf{D}_{sim}}$ and $L_{\mathbf{W}_{sim}}$ follow the same definition for hashtags and words. Due to the very fuzzy multi-class nature of words, hashtags and domain names, I do not include orthogonality constraints for matrices $\mathbf{U}, \mathbf{H}, \mathbf{D}, \mathbf{W}$, which usually result in more precise clusters for co-clustering tasks. It is easy to see that the proposed objective function is not convex for $\mathbf{U}, \mathbf{H}, \mathbf{D}$ and $\mathbf{W}$, hence I develop an iterative algorithm which tries to find a local minima by updating each matrix iteratively as follows;

$$\mathbf{U} \leftarrow \mathbf{U} \odot \sqrt{\frac{\mathbf{X}_{uw}\mathbf{W} + \mathbf{X}_{uh}\mathbf{H} + \mathbf{X}_{ud}\mathbf{D} + \alpha L_{\mathbf{C}}^{-}\mathbf{U}}{\mathbf{U}\mathbf{W}^T\mathbf{W} + \mathbf{U}\mathbf{H}^T\mathbf{H} + \mathbf{U}\mathbf{D}^T\mathbf{D} + \alpha L_{\mathbf{C}}^{+}\mathbf{U}}} \tag{2.2}$$

$$\mathbf{H} \leftarrow \mathbf{H} \odot \sqrt{\frac{\mathbf{X}_{uh}^T\mathbf{H} + \gamma L_{\mathbf{H}_{sim}}^{-}\mathbf{H}}{\mathbf{H}\mathbf{U}^T\mathbf{U} + \gamma L_{\mathbf{H}_{sim}}^{+}\mathbf{H}}} \tag{2.3}$$

$$\mathbf{D} \leftarrow \mathbf{D} \odot \sqrt{\frac{\mathbf{X}_{ud}^T\mathbf{D} + \theta L_{\mathbf{D}_{sim}}^{-}\mathbf{D}}{\mathbf{D}\mathbf{U}^T\mathbf{U} + \theta L_{\mathbf{D}_{sim}}^{+}\mathbf{D}}} \tag{2.4}$$

$$\mathbf{W} \leftarrow \mathbf{W} \odot \sqrt{\frac{\mathbf{X}_{uw}^T\mathbf{U} + \beta L_{\mathbf{W}_{sim}}^{-}\mathbf{W}}{\mathbf{W}\mathbf{U}^T\mathbf{U} + \beta L_{\mathbf{W}_{sim}}^{+}\mathbf{W}}} \tag{2.5}$$

where $L_{ij}^{+} = (|L_{ij}| + L_{ij})/2$ and $L_{ij}^{-} = (|L_{ij}| - L_{ij})/2$. $\odot$ represents element-wise multiplication and $\frac{[\cdot]}{[\cdot]}$ represents element-wise division. Derivation of update rules can be seen in the Appendix of (Ozer *et al.*, 2017). Complexity of the method can be inferred as $\mathcal{O}(i(uwk + uhk + udk + u^2k + h^2k + d^2k + w^2k))$ when complexity of multiplying any X matrix with any of $\mathbf{U}, \mathbf{H}, \mathbf{D}, \mathbf{W}$ is considered to be $\mathcal{O}(uwk)$, $\mathcal{O}(uhk)$, $\mathcal{O}(udk)$ and multiplying any of Laplacian matrices L with any of $\mathbf{U}, \mathbf{H}, \mathbf{D}, \mathbf{W}$ is taken as $\mathcal{O}(u^2k)$, $\mathcal{O}(h^2k)$, $\mathcal{O}(d^2k)$ or $\mathcal{O}(w^2k)$ where i is the number of iterations, u is number of users, h is the number of hashtags, d is the number of domains, w is the number of words and k is the number of clusters. The general algorithmic framework is given at the end of methodology in Algorithm 1.

2.2.2 TriNMF with Three Regularizers

To incorporate usage of hashtags or domains of shared url links solely, I propose a new NMF framework which has the following objective function.

$$\begin{aligned}\mathbf{J}_{\mathbf{U},\mathbf{H},\mathbf{W}} =& \parallel \mathbf{X}_{uw} - \mathbf{U}\mathbf{W}^T \parallel_F^2 + \parallel \mathbf{X}_{uh} - \mathbf{U}\mathbf{H}^T \parallel_F^2 \\ &+ \alpha Tr(\mathbf{U}^T L_{\mathbf{C}} \mathbf{U}) + \gamma Tr(\mathbf{H}^T L_{\mathbf{H}_{sim}} \mathbf{H}) \\ &+ \beta Tr(\mathbf{W}^T L_{\mathbf{W}_{sim}} \mathbf{W}) \\ &s.t. \quad \mathbf{U} \geq 0, \mathbf{H} \geq 0, \mathbf{W} \geq 0\end{aligned} \tag{2.6}$$

where $L_{\mathbf{C}}$ is the Laplacian matrix of user connectivity defined as $D_{\mathbf{C}} - \mathbf{C}$ and $D_{\mathbf{C}}$ is a diagonal matrix which contains the degree of each user in its diagonals. $L_{\mathbf{H}_{sim}}$ and $L_{\mathbf{W}_{sim}}$ follows the same definition for hashtags and words. After applying the same procedure followed in Section 2.2.1, I get updating rules as follows.

$$\mathbf{U} \leftarrow \mathbf{U} \odot \sqrt{\frac{\mathbf{X}_{uw}\mathbf{W} + \mathbf{X}_{uh}\mathbf{H} + \alpha L_{\mathbf{C}}^{-}\mathbf{U}}{\mathbf{U}\mathbf{W}^T\mathbf{W} + \mathbf{U}\mathbf{H}^T\mathbf{H} + \alpha L_{\mathbf{C}}^{+}\mathbf{U}}} \tag{2.7}$$

$$\mathbf{H} \leftarrow \mathbf{H} \odot \sqrt{\frac{\mathbf{X}_{uh}^T\mathbf{U} + \gamma L_{\mathbf{H}_{sim}}^{-}\mathbf{H}}{\mathbf{H}\mathbf{U}^T\mathbf{U} + \gamma L_{\mathbf{H}_{sim}}^{+}\mathbf{H}}} \tag{2.8}$$

$$\mathbf{W} \leftarrow \mathbf{W} \odot \sqrt{\frac{\mathbf{X}_{uw}^T\mathbf{U} + \beta L_{\mathbf{W}_{sim}}^{-}\mathbf{W}}{\mathbf{W}\mathbf{U}^T\mathbf{U} + \beta L_{\mathbf{W}_{sim}}^{+}\mathbf{W}}} \tag{2.9}$$

Note that this update rules can be obtained by setting $\mathbf{D}$, $\mathbf{D}_{sim}$ and θ equal to 0 in Equations 2.2, 2.3, 2.5. Complexity of the method can be calculated by omitting the costs of operations done over matrices $\mathbf{X}_{ud}$, $\mathbf{D}$ and $L_{\mathbf{D}_{sim}}$. The complexity of the method is $\mathcal{O}(i(uwk + uhk + u^2k + h^2k + w^2k))$.

2.2.3 DualNMF with Two Regularizers

To use only user word matrix as user content and regularize factorization with user connectivity and keyword similarity, inspired by (Yao *et al.*, 2014), I present

DualNMF objective function as;

$$\begin{aligned} \mathbf{J}_{\mathbf{U},\mathbf{W}} =& \| \mathbf{X}_{uw} - \mathbf{U}\mathbf{W}^T \|_F^2 + \alpha Tr(\mathbf{U}^T L_{\mathbf{C}} \mathbf{U}) \\ &+ \beta Tr(\mathbf{W}^T L_{\mathbf{W}_{sim}} \mathbf{W}) \\ &s.t. \quad \mathbf{U} \geq 0, \mathbf{W} \geq 0 \end{aligned} \tag{2.10}$$

After following the same procedure introduced in Section 2.2.1, I can get the update rules for $\mathbf{U}$ and $\mathbf{W}$ as;

$$\mathbf{U} \leftarrow \mathbf{U} \odot \sqrt{\frac{\mathbf{X}_{uw}\mathbf{W} + \alpha L_{\mathbf{C}}^{-}\mathbf{U}}{\mathbf{U}\mathbf{W}^T\mathbf{W} + \alpha L_{\mathbf{C}}^{+}\mathbf{U}}} \tag{2.11}$$

$$\mathbf{W} \leftarrow \mathbf{W} \odot \sqrt{\frac{\mathbf{X}_{uw}^T\mathbf{U} + \beta L_{\mathbf{W}_{sim}}^{-}\mathbf{W}}{\mathbf{W}\mathbf{U}^T\mathbf{U} + \beta L_{\mathbf{W}_{sim}}^{+}\mathbf{W}}} \tag{2.12}$$

Complexity of the method can be inferred as $\mathcal{O}(i(uwk + u^2k + w^2k))$ after omitting the extra operations done over matrices $\mathbf{X}_{uh}$, $\mathbf{H}$ and $D_{\mathbf{H}_{sim}}$ in the previous method. The general algorithm can be summarized as the application of the related update

Algorithm 1 NMF Algorithms

Input:$\{\mathbf{X}_{uw}, \mathbf{X}_{uh}, \mathbf{X}_{ud}, \mathbf{C}, \mathbf{H}_{sim}, \mathbf{D}_{sim}, \mathbf{W}_{sim}, \alpha, \beta, \theta, \gamma\}$

Output: $\mathbf{U}$

1: Initialize $\mathbf{U}, \mathbf{H}, \mathbf{D}, \mathbf{W} > 0$

2: **while** $\Delta residual > threshold$ **do**

3: Update $\mathbf{U}$ by using one of Equations 2.2, 2.7, 2.11

4: Update $\mathbf{H}$ by using one of Equations 2.3, 2.8

5: Update $\mathbf{D}$ by using Equation 2.4

6: Update $\mathbf{W}$ by using one of Equations 2.5, 2.9, 2.12

7: **end while**

8: Assign user i to community j where $j = argmax_j \mathbf{U}_{ij}$.

rules to the matrices $\mathbf{U}, \mathbf{H}, \mathbf{D}, \mathbf{W}$. For MultiNMF with multi regularizers method,

equations 2.2, 2.3, 2.4, 2.5 are applied. For TriNMF with three regularizers method, equations 2.7, 2.8, 2.9 are applied and **D** matrix is not included in calculations. For DualNMF method, equations 2.11 and 2.12 are applied and **H** and **D** matrices are not included in calculations. I make implementations of all three algorithms publicly available [1] .

2.3 Experiments and Results

2.3.1 Data Description

I make use of a pair of publicly available [2] political Twitter datasets to evaluate my methods. These datasets are user lists of 419 British political figures from four major political parties in the UK, namely; Conservative and Unionist Party, Labour Party, Scottish National Party, Liberal Democrats and others, and 349 major Irish political figures from seven political parties; Fianna Fail, Fine Gael, Green Party, Sinn Fein, United Left Alliance, Independents. Several statistics for the datasets are shown in Table 2.2.

For the UK and Ireland data, I crawl all of the tweets sent from the accounts of given user id lists. In order not to be heavily influenced by the extremely polarized election season, I only used tweets dated after May, 7 2015, which was the election day in the UK. To balance the share of number of tweets from each user I limit the number of tweets to 200 per user.

For each dataset, same preprocessing method is followed. First, words occurring less than 20 times and stop words are eliminated. After eliminating word features, users and tweets that lack content are also eliminated. Hashtags and domains that appear only once are not taken into consideration either. Statistics shown in Table

[1] `http://www.public.asu.edu/~mozer/ASONAM2016Code.tar.gz`

[2] Users' Twitter id lists can be obtained from http://mlg.ucd.ie/aggregation/index.html

Table 2.2: Characteristics of the UK and Ireland Datasets

	UK	Ireland
# of Tweets	19,947	14,656
# of Retweets	1,566	7,088
# of Mentions	4,956	22,072
# of Words	10,766	7,973
# of Hashtags	945	986
# of URL Domains	946	634
# of Users	233	258
# of Baseline Communities	5	7

2.2 show the numbers after preprocessing.

2.3.2 Evaluation Metrics

To evaluate the methods, I make use of three well known clustering quality metrics, namely; purity, adjusted rand index(Hubert and Arabie, 1985) and normalized mutual information(Strehl and Ghosh, 2002).

Purity can be formally defined as;

$$Purity = \frac{1}{n}\sum_{i=1}^{k} max_j |C_i \cap l_j|$$

where k is the number of communities found, n is the number of instances, l_j is the set of instances which belong to the class j, and C_i is the set of instances that are members of community i.

Adjusted Rand Index (Hubert and Arabie, 1985) can be formally defined as;

$$ARI = \frac{RI - E[RI]}{max(RI) - E[RI]}$$

$$RI = \frac{s + s'}{\binom{n}{2}}$$

s is the number of pairs which belong to both same ground-truth class and identified community. s' is the number of pairs which belong to both different ground-truth classes and identified communities. It evaluates the similarity of ground-truth class labels and clustering result.

Normalized Mutual Information can be formally defined as;

$$NMI = \frac{\sum_{j=1}^{|l|} \sum_{i=1}^{|C|} P(j,i) log\Big(\frac{P(j,i)}{P(i)P(j)}\Big)}{\sqrt{H(l)H(C)}}$$

where, $H(l)$ and $H(C)$ are the entropy of class and community assignments of l and C. $P(j,i)$ is the probability that randomly picked user has class label j and belongs to the community i while $P(j)$ gives the probability of randomly picked user to be in class j and similarly $P(i)$ to be in community i.

2.3.3 Baseline Algorithms

As a baseline to evaluate the performance of using both connectivity and content information, I design experiments with connectivity-only and content-only clustering methods.

For connectivity-only method, I use Louvain (Blondel *et al.*, 2008) and CNM (Clauset *et al.*, 2004) algorithms utilizing modularity optimization over user adjacency matrix. Modularity is defined as:

$$Q = \frac{1}{2m} \sum_{ij} (A_{ij} - \frac{k_i k_j}{2m}) \delta(c_i, c_j) \tag{2.13}$$

where $\delta(c_i, c_j)$ is the Kronecker delta symbol, c_i is the label of the community to which node i is assigned, and k_i is the degree of node i.

For content-only approach, I experiment with k-means(Lloyd, 2006) and conventional non-negative matrix factorization algorithm (Lee and Seung, 2000).

For approaches employing both connectivity and content information of users, I test GNMF (Cai *et al.*, 2011) and NMTF (Pei *et al.*, 2015) algorithms besides proposed methods. GNMF algorithm is introduced by Cai et al. to incorporate intrinsic geometric similarity of users. I feed previously defined two types of user connectivity graphs' adjacency matrices as graph regularization terms to the GNMF algorithm.

Pei et al. work in (Pei *et al.*, 2015) applies non-negative matrix tri-factorization with regularization to Twitter data. It makes use of user similarity, [tweet x word] and [user x word] matrices and regularize the objective function with tweet similarity and user connectivity matrices. Complexity of the algorithm is $O(rk(mn + mw + nw + m^3 + n^2))$ where r is the iteration times. m, n, k, and w denote the number of users, messages, features and communities.

2.3.4 Experimental Design

First set of experiments test the performance of using connectivity-only information for community detection, labeled as the Experiment Set 1. I test Louvain and CNM algorithms on three different types of connectivity graphs. Second set of experiments test the performance of content-only methods, labeled as Experiment Set 2. I test k-means and NMF methods. Third set of experiments test the performance of methods utilizing both connectivity and content information, labeled as Experiment Set 3. I test GNMF and NMTF frameworks proposed by (Pei *et al.*, 2015) as baseline algorithms, alongside my proposed MultiNMF, TriNMF and DualNMF methods. In user content dimension, I use DualNMF method to test the experiment design that only uses user-word content. I use TriNMF method to test the experiment design that uses user-hashtag or user-domain information in combination with the user-word information. I use MultiNMF method to test the experiment design that uses

all of user-word, user-hashtag and user-domain contents. I label these experiments as Experiment Set 1, 2 and 3 respectively.

2.3.5 Experimental Results

First, I present statistics of retweets without edits and user mentions on the full and endorsement filtered user connectivity graphs. Table 2.3 shows that retweeting without edits indeed occurs mostly inside like-minded political camps, rather than cross-camps. Roughly 97% of retweets in the UK data, and 88% of retweets in the Ireland data occur inside like-minded groups, while these percentages are much lower for users mentions. My endorsement filtered connectivity network boosts the percentage of inner group user mentions from 83% to 97% in the UK data and from 59% to 87% in the Ireland data evidencing that TSB rule in fact identifies positive user mentions and retweets with edits with high accuracy.

Table 2.3: Effect of Endorsement Filtered Mention Links

	UK	Ireland
Inner Group Retweet Links	962	1,652
Inter Group Retweet Links	28	216
Inner Group Retweet + Mention Links	1,986	3,056
Inter Group Retweet + Mention Links	398	2,092
Inner Group Retweet + ΔMention Links	1,456	2,820
Inter Group Retweet + ΔMention Links	40	432

I run each experiment 20 times for every method and pick the maximum score achieved for reporting. Each regularizer parameter (α, γ, θ, β) are experimented with

values 1, 10, 100 and 1000. Best accuracies are usually reached with experiments in which α and β equal to 10 or 100 while γ and θ equal to 1. This shows the contribution of user connectivity and word similarity regularizers, and considerably lower contributions of hashtag and domain name regularizers towards overall performance of the algorithms.

Table 2.4: UK & Ireland Experiment Set 1 Results

		UK Dataset				Ireland Dataset			
Algorithm	User Graph	k	Purity	ARI	NMI	k	Purity	ARI	NMI
Louvain	$R + M$	20	.9313	.4661	.5854	13	.8720	.7277	.6849
	$R + \Delta M_w$	42	.9484	.4291	.5916	31	.9224	.7536	.7518
CNM	$R + M$	17	.8498	.5656	.5257	10	.7016	.4509	.4720
	$R + \Delta M_w$	41	.9700	.6150	.6496	29	.8333	.6426	.6381

Major findings for Experiment Set 1 can be summarized as follows:

- Relatively larger clustering scores occur due to artificially large number of clusters that are found. Considering the number of users in both datasets, the number of clusters identified in Experiment Set 1 are not practical for use (e.g. 29 clusters in Ireland data for 7 political parties).

- Using endorsement filtered user connectivity graph usually gives better clustering performance compared to using full user connectivity graph. There is a pattern of weighted graph approach outperforming the others.

Experiment Set 2 indicates that word usage-only based clustering yields considerably lower accuracies compared to user connectivity-only based clustering.

Table 2.5: UK & Ireland Experiment Set 2 Results

		UK Dataset			Ireland Dataset		
Algorithm	User Content	Purity	ARI	NMI	Purity	ARI	NMI
k-Means	user x word	.6738	.2378	.2018	.4651	.0488	.1672
NMF	user x word	.6395	.1541	.1709	.4186	.0434	.1139

Major findings from Experiment Set 3 can be summarized as follows;

- Regardless of the experiment set and algorithms used, endorsement filtered user connectivity graph yields higher accuracy clustering performance compared to using the full connectivity graph. Usually weighted graph approach outperforms the others.

- DualNMF method which factorizes user-word matrix alongside user connectivity and word similarity regularizers yields the highest accuracy clustering performance.

- I get much higher scores of clustering accuracy in Experiment Set 3 compared to Experiment Set 2. Regularizing content-only methods with user connectivity graphs(GNMF (Cai *et al.*, 2011)), dramatically increases the quality of the clustering. DualNMF which incorporates keyword similarity regularization to GNMF further boosts the quality of clustering.

- Compared to DualNMF method, including tweet messages for NMTF method proposed in (Pei *et al.*, 2015) does not help to further improve the clustering quality, while it increases complexity dramatically. DualNMF provides 9% additional purity, 46% additional ARI score while doubling the NMI score compared

Table 2.6: UK & Ireland Experiment Set 3 Results.

Algorithm	User Graph	User Content	Purity	ARI	NMI
GNMF*	$R + M$	word	.7854	.4955	.4120
	$R + \Delta M_w$		.8326	.6469	.5461
NMTF*	$R + M$	word, tweet	.8197	.6448	.2593
	$R + \Delta M_w$		.8412	.5331	.3751
TriNMF	$R + M$	word, domain	.7597	.3707	.3158
	$R + \Delta M_w$		.8283	.6375	.5006
TriNMF	$R + M$	word, #tag	.7897	.5232	.4320
	$R + \Delta M_w$		.7768	.5001	.3837
MultiNMF	$R + M$	word, domain, #tag	.7554	.4025	.3343
	$R + \Delta M_w$		.8112	.6108	.4978
DualNMF	$R + M$	word	.8326	.5674	.5146
	$R + \Delta M_w$		**.8970**	**.7616**	**.6380**

to the baseline NMTF method of Pei et al. in (Pei *et al.*, 2015).

- Compared to DualNMF method, utilizing hashtag and/or domain usage information (i.e. TriNMF and MultiNMF) do not contribute to the overall clustering quality.

Table 2.7: Ireland Experiment Set 3 Results

Algorithm	User Graph	Content	Purity	ARI	NMI
GNMF*	$R + M$	word	.5543	.2447	.2881
	$R + \Delta M_w$		.8178	.6978	.6399
NMTF*	$R + M$	word, tweet	.5969	.3119	.2144
	$R + \Delta M_w$		.7597	.5198	.4469
TriNMF	$R + M$	word, domain	.7209	.5051	.5237
	$R + \Delta M_w$		.8101	.6807	.6372
TriNMF	$R + M$	word, #tag	.6938	.4202	.4431
	$R + \Delta M_w$		.8062	.6784	.6885
MultiNMF	$R + M$	word, domain, #tag	.7481	.4777	.4938
	$R + \Delta M_w$		.8178	.6953	.6411
DualNMF	$R + M$	word	.7364	.5561	.5397
	$R + \Delta M_w$		**.8721**	**.7536**	**.7096**

Table 2.8: Comparison of NMF Methods for Experiment Set 3

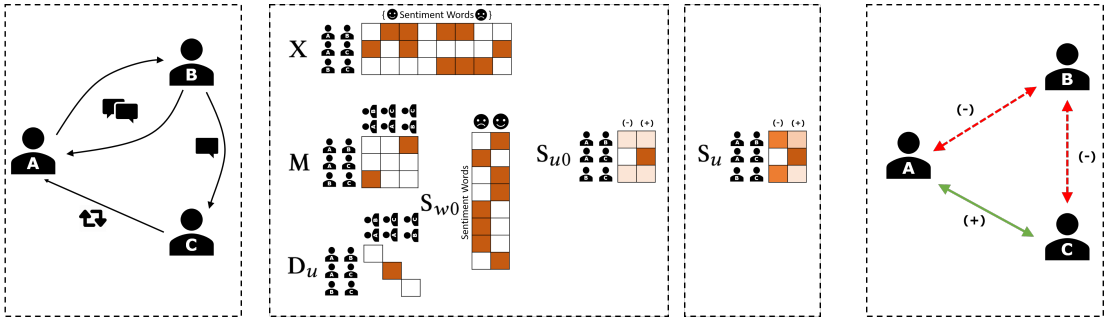
		Connectivity	
		$R + M$	✓$\mathbf{R + \Delta M_w}$
Content	✓**word**	DualNMF	✓✓**DualNMF**
	word, {#tag or domain}	TriNMF	✓**TriNMF**
	word, #tag, domain	MultiNMF	✓**MultiNMF**

Chapter 3

IMPLICIT NEGATIVE LINK DETECTION

3.1 Proposed Frameworks

3.1.1 Offline Framework



(a) Data Input (b) Algorithm Input (c) Alg. Output (d) Output

Figure 3.1: Input Representation of Social Media Data and Interpretation of Algorithm Output.

In this section, I first present the notation used throughout the chapter, formally define the problem and then propose the SocLS-Fact optimization solution. Finally, I provide the details of how to build the prior knowledge that the SocLS-Fact requires.

Before going into the details of the framework, the notation that is used throughout the chapter can be seen in Table 3.1. Let m be the number of interacting user pairs, and n be the number of unique sentiment words. An example with 3 interacting user pairs and 8 unique sentiment words can be seen in Figure 3.1a and 3.1b. All textual interaction happening between two users are represented as rows of $\mathbf{X}$. $\mathbf{X}$ encodes how many times each sentiment word occurs in textual interactions of two

Table 3.1: Notation

Symbol	Size	Explanation
m		Number of interacting user pairs
n		Number of sentiment words
I_k	$k \times k$	Identity matrix of size k
$\mathbf{X}$	$[m \times n]$	Matrix of occurrences of sentiment words in textual interactions of user pairs
$\mathbf{S}_u$	$[m \times 2]$	User link polarity
$\mathbf{S}_{u0}$	$[m \times 2]$	Initial user link polarity
$\mathbf{D}_u$	$[m \times m]$	Binary diagonal matrix of user pairs with positive interaction
$\mathbf{S}_w$	$[n \times 2]$	Sentiment word polarity
$\mathbf{S}_{w0}$	$[n \times 2]$	Initial sentiment lexicon
$\mathbf{M}$	$[m \times m]$	Social balance matrix

users. In Figure 3.1b, when user a and b interacts they use 2nd, 3rd, 5th and 6th words while user b and c interacts they use 1st, 3rd and 8th and so on. Initial user link polarities are embedded in matrix $\mathbf{S}_{u0}$. Initial sentiment lexicon is embedded in $\mathbf{S}_{w0}$. Positive and negative polarities are represented as two latent dimensions in matrix $\mathbf{S}_{u0}$, and $\mathbf{S}_{w0}$. Which user links should have the same polarity following the social balance theory is governed by matrix $\mathbf{M}$. Further details of how matrices $\mathbf{S}_{u0}$, $\mathbf{S}_{w0}$, $\mathbf{M}$ are derived is given later in this section.

As I discuss earlier, sentiment of words used in user interactions are proven to be good predictors of the polarity of user links. Moreover, built-in positive interactions

(i.e. retweet, like, share) are good predictors of positive user links by their nature. As referred in Section 1.1.2, how user links form triangles with each other is also a decisive factor of their polarities since they tend to follow social balance theory. To factorize all textual interactions between users into two latent dimensions as positive and negative and enjoy aforementioned three predictors of polarity of user links at the same time, I propose the following optimization problem;

$$
\begin{aligned}
\min_{\mathbf{S_u},\mathbf{H},\mathbf{S_w}} \quad & ||\mathbf{X} - \mathbf{S}_u\mathbf{H}\mathbf{S}_w^T||_F^2 && (0)\\
& + \alpha||\mathbf{S}_w - \mathbf{S}_{w0}||_F^2 && (1)\\
& + \beta Tr\Big((\mathbf{S}_u - \mathbf{S}_{u0})^T\mathbf{D}_u(\mathbf{S}_u - \mathbf{S}_{u0})\Big) && (2)\\
& + \gamma||\mathbf{M} - \mathbf{S}_u\mathbf{S}_u^T||_F^2 && (3)\\
\text{subject to} \quad & \mathbf{S_u} > 0, \mathbf{S_w} > 0, \mathbf{H} > 0
\end{aligned}
$$

Optimization formulation consists of 4 terms. $(0)^{th}$ term factorizes user pair textual interactions into three matrices. $\mathbf{S}_u \in \mathbb{R}_+^{m\times 2}$ is the lower-rank projection of matrix $\mathbf{X}$. The first column of $\mathbf{S}_u$ is the latent negative and second column is the latent positive dimension. $\mathbf{S}_w$ is the lower-rank projection of columns of matrix $\mathbf{X}$. Note that each column of $\mathbf{X}$ represents a sentiment word. Projection matrix $\mathbf{S}_w$ corresponds to distributed polarity representation of each sentiment word. As in $\mathbf{S}_u$, first column of $\mathbf{S}_w$ is the latent negative and the second column is the latent positive dimension.

$(1)^{st}$ term in the optimization formulation penalizes the meaning change of the sentiment words compared their initial lexicon meaning. Parameter α governs the relaxation on the penalty.

$(2)^{nd}$ term governs how much the polarity prediction of links diverges from their initial inferred labels. Initial labels are inferred as positive if there is any platform-specific positive interaction between users that the link connecting to. Diagonal

matrix D_u helps to penalize divergences of links which have platform-specific positive interactions only.

$(3)^{rd}$ term in the optimization formulation penalizes the triangles in the user network that do not follow social balance theory. $\mathbf{M}$ encodes the information of pair of links that should have the same polarity if they are forming a triangle with another positive link.

Constructing $\mathbf{S}_{w0}$

A well-known off-the-shelf sentiment word lexicon is utilized [1] to populate the initial sentiment polarities of words. A word is represented as $[1, 0]$ if it has negative sentiment meaning. It is represented as $[0, 1]$ if it has positive sentiment meaning. In Figure 3.1b, initial sentiment lexicon is embedded in $\mathbf{S}_{w0}$ such that 1st, 3rd, 4th and 8th words as positive sentiment words and 2nd, 5th, 6th and 7th words as negative sentiment words.

Constructing $\mathbf{S}_{u0}$ and $\mathbf{D}_u$

Each row of the initial user link polarity matrix $\mathbf{S}_{u0}$ encodes the information of the prior inference of the polarity of user link. First column of the polarity matrix $\mathbf{S}_{u0}$ is the latent negative dimension, while the second column is the latent positive dimension. For the links that connect user pairs having previous platform-specific positive interaction, I infer the initial polarity of them as positive and embed it as $[0, 1]$ in the corresponding row of $\mathbf{S}_{u0}$ and as 1 in the corresponding diagonal entry of $\mathbf{D}_u$. For the links that connect user pairs having no previous platform-specific positive interaction, I do not infer any initial polarity and represent them as $[0.5, 0, 5]$ in $\mathbf{S}_{u0}$ and as 0 in the corresponding diagonal entry of $\mathbf{D}_u$. To illustrate in Figure 3.1b, the

[1] http://www.cs.uic.edu/∼liub/FBS/opinion-lexicon-English.rar

positive interaction between user A and C is represented as $[0, 1]$ in the second row of $\mathbf{S}_{u0}$ and as 1 in the second diagonal entry of $\mathbf{D}_u$.

Incorporating Social Balance Theory

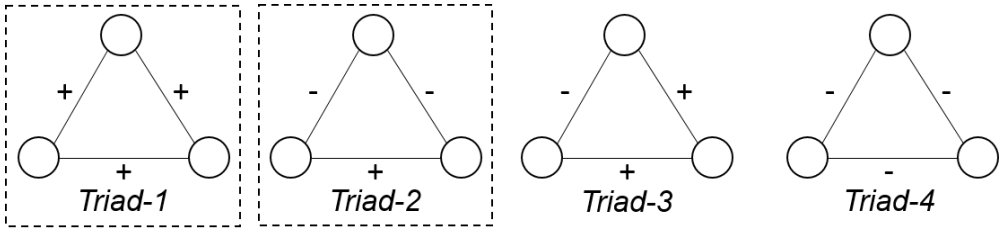


Figure 3.2: Possible Configurations of Undirected Signed Links in a Triad. Balanced Ones in Dashed Rectangles.

The theory of social balance of signed links in triads is extensively studied since its introduction by Heider et al. in (Heider, 1958) as structural balance of signed links. It suggests that for a signed triad to be balanced, it has to have an odd number of positive links (i.e. one or three positive links), otherwise it is not balanced. The balanced configurations among all possible configurations are presented with dashed frames in Figure 3.2. The definition of structural balance is analogous to common daily phrase of "enemy of my enemy is my friend" and "friend of my friend is my friend" in social settings.

To encode the social balance theory, I utilize the prior knowledge of positive links inferred from platform-specific positive interactions. My intuition is that if two users have any prior platform-specific positive interaction, the polarity of their interaction with any other third user should be similar. They can connect to third user either with both negative or positive links (i.e. Triad-1 and Triad-2 in Figure 3.2). The cases which they connect to a third user with different polarities are not socially balanced configurations (i.e. Triad-3 in Figure 3.2).

The matrix $\mathbf{M} \in \{0,1\}^{m \times m}$ encodes the link pairs that are needed to have the same polarity to follow social balance theory by having 1 in the related row and column of $\mathbf{M}$ and 0 for the rest. In Figure 3.1a, link between user A and B should have the same polarity with link between user B and C. It is because they are forming a triad with link between user A and C which has prior platform-specific positive interaction. In Figure 3.1b, it is encoded as 1 in the $\mathbf{M}(1,3)$ and $\mathbf{M}(3,1)$. Eventually, minimizing the squared frobenious norm of the difference between $\mathbf{M}$ and $\mathbf{S}_u\mathbf{S}_u^T$ forces triads to have odd number of positive links in the whole network.

Algorithm

The objective function proposed in Section 3.1.1 is not convex for all variables of $\mathbf{S}_u$,$\mathbf{S}_w$,$\mathbf{H}$. I introduce an alternating optimization solution for my problem similar to (Li *et al.*, 2009). I update each variable $\mathbf{S}_u$,$\mathbf{S}_w$,$\mathbf{H}$ iteratively while fixing others to find a local minimum in the solution space. The update rules for each variable is given as;

$$\mathbf{S}_u \leftarrow \mathbf{S}_u \odot \sqrt{\frac{\mathbf{X}\mathbf{S}_w\mathbf{H}^T + \gamma(\mathbf{M}+\mathbf{M}^T)\mathbf{S}_u + \beta\mathbf{D}_u\mathbf{S}_{u0}}{\mathbf{S}_u\mathbf{H}\mathbf{S}_w^T\mathbf{S}_w\mathbf{H}^T + \gamma\mathbf{S}_u\mathbf{S}_u^T\mathbf{S}_u + \beta\mathbf{D}_u\mathbf{S}_u}} \tag{3.1}$$

$$\mathbf{H} \leftarrow \mathbf{H} \odot \sqrt{\frac{\mathbf{S}_u^T\mathbf{X}\mathbf{S}_w}{\mathbf{S}_u^T\mathbf{S}_u\mathbf{H}\mathbf{S}_w^T\mathbf{S}_w}} \tag{3.2}$$

$$\mathbf{S}_w \leftarrow \mathbf{S}_w \odot \sqrt{\frac{\mathbf{X}^T\mathbf{S}_u\mathbf{H} + \alpha\mathbf{S}_{w0}}{\mathbf{S}_w\mathbf{H}^T\mathbf{S}_u^T\mathbf{S}_u\mathbf{H} + \alpha\mathbf{S}_w}} \tag{3.3}$$

Derivation of the update rules are presented in the Appendix B. The proposed algorithm employs an iterative scheme of the above rules until convergence. Each step of the algorithm is shown in Algorithm 2.

Finally, the polarity of the latent dimension with higher numerical value in the i^{th} row of $\mathbf{S}_u$ is assigned as the polarity output of the link i. To illustrate in Figure 3.1c and 3.1d, it can be seen that the value in the first column is greater than the second

Algorithm 2 Proposed Algorithm for the Optimization Problem

Input:{$\mathbf{X}$,$\mathbf{S}_{u0}$,$\mathbf{S}_{w0}$, $\mathbf{M}$}

Output: $\mathbf{S}_u$,$\mathbf{S}_w$

Initialize $\mathbf{S}_u \leftarrow \mathbf{S}_{u0}$,$\mathbf{H} \leftarrow I_2$, $\mathbf{S}_w \leftarrow \mathbf{S}_{w0}$.

while not convergent **do**

    Update $\mathbf{S}_u$ using Equation 3.1.

    Update $\mathbf{H}$ using Equation 3.2.

    Update $\mathbf{S}_w$ using Equation 3.3.

end while

column for the first and the third rows of $\mathbf{S}_u$. Therefore, the polarity of the link between user A and B and the link between user B and C are inferred as negative. Since the value in the second column is greater than the first column for the second row of $\mathbf{S}_u$ the polarity of the link between user A and C is inferred as positive.

The most computationally costly operations of the update rules are matrix multiplications since matrix summation, matrix hadamard product and element-wise division can be handled in linear time. Complexity of the update rule in Equation 3.1 is $\mathcal{O}(mn + m^2 + m + n^2m)$. Complexity of the update rule in 3.2 is $\mathcal{O}(mn + m + n)$. Complexity of update rule in 3.3 is $\mathcal{O}(mn + m^2n)$. Therefore, overall time complexity of the Algorithm 2 complexity is $\mathcal{O}(i(m^2n + n^2m + m^2 + mn + m + n))$ where i is the iteration count that algorithm takes until update rules converges to a local minima. Experiments empirically show that convergence takes usually less than 20 iterations.

The proof of the convergence of the algorithm is omitted here due to space constraints which can be followed in similar works using the auxiliary function approach, such as presented in (Ding *et al.*, 2006). The source code for the whole running pipeline presented in this section can be reached at `www.public.asu.edu/~mozer/HT2017Code.tar.gz`.

3.1.2 Online Framework

Given the dynamic nature of online political networks, it is necessary to handle streaming data in an online fashion. It usually is computationally expensive to re-run offline methods from scratch each time a new piece of arrives. One naive solution is running the offline method only on the new data. It is a faster solution, yet, it ignores the rich historical information.

To alleviate the aforementioned problems, we introduce an online framework. It follows similar principles as the offline framework besides the modelling of temporal dimension. It uses sentiment words, prior positive interactions, and socially balanced triads to infer implicit negative links. To take previous snapshots' detected implicit links into account, we propose using a temporal smoothing term. This smoothing term penalises abrupt changes of the signs of the links in consecutive snapshots. Thus, we propose solving the following optimisation problem for online settings;

$$
\begin{aligned}
\min_{\mathbf{S_u}^{(t)},\mathbf{H}^{(t)},\mathbf{S}_w^{(t)}} \quad & ||\mathbf{X}^{(t)} - \mathbf{S}_u^{(t)}\mathbf{H}^{(t)}{\mathbf{S}_w^{(t)}}^T||_F^2 && (0)\\
& + \alpha||\mathbf{S}_w^{(t)} - \mathbf{S}_{w0}||_F^2 && (1)\\
& + \beta Tr\Big((\mathbf{S}_u^{(t)} - \mathbf{S}_{u0}^{(t)})^T\mathbf{D}_u^{(t)}(\mathbf{S}_u^{(t)} - \mathbf{S}_{u0}^{(t)})\Big) && (2)\\
& + \gamma||\mathbf{M}^{(t)} - \mathbf{S}_u^{(t)}{\mathbf{S}_u^{(t)}}^T||_F^2 && (3)\\
& + \tau\sum_{i=1}^{t} e^{-(t-i)}||\mathbf{S}_u^{(t)} - \mathbf{S}_u^{(i)}||_F^2 && (4)\\
\text{subject to} \quad & \mathbf{S}_u^{(t)} > 0, \mathbf{S}_w^{(t)} > 0, \mathbf{H}^{(t)} > 0
\end{aligned}
$$

In above formulation, t stands for the current time snapshot and any matrix superscripted by parameter t (e.g. $\mathbf{X}^{(t)}$) spans the data of tth snapshot from time $(t-1)$ to t. First four terms $(0, 1, 2, 3)$ in the objective function are inherited from the offline framework. $(4)^{th}$ term controls the divergence of current snapshot's signs of

links from previous time snapshots'. An inverse exponential decay function ($e^{-(t-i)}$) is employed to weight previous snapshots' importance in temporal order. One can simply plug another decay function based on their application's constraints when necessary. Parameter τ controls the importance of temporal smoothing.

Algorithm

To optimise the online framework objective function, we follow similar iterative multiplicative update rules as in the offline framework. While updating $\mathbf{S}_u^{(t)}$, we treat emerging links and continuing links exclusively, since they are subject to different temporal smoothing constraints. For emerging links as there is no precedent of them in previous snapshot, we employ the update rule of the offline framework. We denote rows of $\mathbf{S}_u$ corresponding to emerging links as $\mathbf{S}_{ue}$,

$$\mathbf{S}_{ue}^{(t)} \leftarrow \mathbf{S}_{ue}^{(t)} \odot \sqrt{\frac{\mathbf{X}_e^{(t)}\mathbf{S}_w^{(t)}\mathbf{H}^{(t)^T} + \gamma(\mathbf{M}_e^{(t)} + \mathbf{M}_e^{(t)^T})\mathbf{S}_{ue}^{(t)} + \beta\mathbf{D}_{ue}^{(t)}\mathbf{S}_{ue0}^{(t)}}{\mathbf{S}_{ue}^{(t)}\mathbf{H}^{(t)}\mathbf{S}_w^{(t)^T}\mathbf{S}_w^{(t)}\mathbf{H}^{(t)^T} + \gamma\mathbf{S}_{ue}^{(t)}\mathbf{S}_{ue}^{(t)^T}\mathbf{S}_{ue}^{(t)} + \beta\mathbf{D}_{ue}^{(t)}\mathbf{S}_{ue}^{(t)}}} \tag{3.4}$$

For continuing links, we incorporate the temporal smoothing term, so, the update rule for continuing links become,

$$\mathbf{S}_{uc}^{(t)} \leftarrow \mathbf{S}_{uc}^{(t)} \odot \sqrt{\frac{\mathbf{X}_c^{(t)}\mathbf{S}_w^{(t)}\mathbf{H}^{(t)^T} + \gamma(\mathbf{M}_c^{(t)} + \mathbf{M}_c^{(t)^T})\mathbf{S}_{uc}^{(t)} + \beta\mathbf{D}_{uc}^{(t)}\mathbf{S}_{uc0}^{(t)} + \tau\sum_{i=1}^{t} e^{-(t-i)}\mathbf{S}_{uc}^{(i)}}{\mathbf{S}_{uc}^{(t)}\mathbf{H}^{(t)}\mathbf{S}_w^{(t)^T}\mathbf{S}_w^{(t)}\mathbf{H}^{(t)^T} + \gamma\mathbf{S}_{uc}^{(t)}\mathbf{S}_{uc}^{(t)^T}\mathbf{S}_{uc}^{(t)} + \beta\mathbf{D}_{uc}^{(t)}\mathbf{S}_{uc}^{(t)} + \tau t\mathbf{S}_{uc}^{(t)}}} \tag{3.5}$$

From the perspective of matrices $\mathbf{H}^{(t)}$ and $\mathbf{S}_w^{(t)}$, there is no temporal smoothing involved. So, same update rules can be employed as in the offline framework in a snapshot-based fashion;

$$\mathbf{H}^{(t)} \leftarrow \mathbf{H}^{(t)} \odot \sqrt{\frac{\mathbf{S}_u^{(t)^T}\mathbf{X}^{(t)}\mathbf{S}_w^{(t)}}{\mathbf{S}_u^{(t)^T}\mathbf{S}_u^{(t)}\mathbf{H}^{(t)}\mathbf{S}_w^{(t)^T}\mathbf{S}_w^{(t)}}} \tag{3.6}$$

$$\mathbf{S}_w^{(t)} \leftarrow \mathbf{S}_w^{(t)} \odot \sqrt{\frac{\mathbf{X}^{(t)^T}\mathbf{S}_u^{(t)}\mathbf{H}^{(t)} + \alpha\mathbf{S}_{w0}^{(t)}}{\mathbf{S}_w^{(t)}\mathbf{H}^{(t)^T}\mathbf{S}_u^{(t)^T}\mathbf{S}_u^{(t)}\mathbf{H}^{(t)} + \alpha\mathbf{S}_w^{(t)}}} \tag{3.7}$$

Derivation of the update rule of $\mathbf{S}_{uc}^{(t)}$ is given in Appendix B.4. Derivation of the update rules of $\mathbf{S}_{ec}^{(t)}$, $\mathbf{H}^{(t)}$, and $\mathbf{S}_w^{(t)}$ follow the exact offline framework calculations in (Ozer *et al.*, 2017). Given the update rules, algorithm for finding the optimal $\mathbf{S}_{ue}^{(t)}$ and $\mathbf{S}_{uc}^{(t)}$ becomes straightforward, and presented in Algorithm 3

Algorithm 3 Proposed Algorithm for the Online Framework's Optimisation Problem

Input:$\{\mathbf{X}^{(t)}$,$\mathbf{S}_{u0}^{(t)}$,$\mathbf{S}_{w0}$, $\mathbf{M}^{(t)}$, $\mathbf{S}_u^{(i)}$, $\quad i = 1, 2..., t-1\}$

Output: $\{\mathbf{S}_u^{(t)}$,$\mathbf{S}_w^{(t)}\}$

1: Initialise $\mathbf{S}_u^{(t)} \leftarrow \mathbf{S}_{u0}^{(t)}$,$\mathbf{H}^{(t)} \leftarrow I_2$, $\mathbf{S}_w^{(t)} \leftarrow \mathbf{S}_{w0}$.

2: **while** not convergent **do**

3: Update emerging links $\mathbf{S}_{ue}^{(t)}$ using Equation 3.4.

4: Update continuing links $\mathbf{S}_{uc}^{(t)}$ using Equation 3.5.

5: Update $\mathbf{H}^{(t)}$ using Equation 3.6.

6: Update $\mathbf{S}_w^{(t)}$ using Equation 3.7.

7: **end while**

As in the offline framework, the polarity of a link is assigned based on the values in the corresponding row of the link in $\mathbf{S}_u^{(t)}$. If the first value is larger, the link is inferred as negative, and otherwise, as positive.

Complexity of the algorithm can be formulated as follows. Update rule for $\mathbf{S}_{ue}^{(t)}$(Eq. 3.4), $\mathbf{H}^{(t)}$(Eq. 3.6), and $\mathbf{S}_w^{(t)}$ (Eq. 3.7) are same as in the offline framework; Equation 3.1, 3.2 and 3.3. They are $\mathcal{O}(mn+m^2+m+n^2m)$, $\mathcal{O}(mn+m+n)$ and $\mathcal{O}(mn+m^2n)$, respectively. Complexity of updating $\mathbf{S}_{uc}^{(t)}$ is $\mathcal{O}(mn+m^2+m+n^2m+tm)$. Therefore, whole complexity of the online framework becomes $\mathcal{O}(i(m^2n+n^2m+m^2+mn+tm+n))$ where i is the number of iterations of applying multiplicative update rules. The added

time complexity the online framework introduces is due to the term tm. The source code for the whole running pipeline for both offline and online frameworks can be reached at `www.public.asu.edu/~mozer/NRHMcode.tar.gz`.

In this section, we present experiments to evaluate the performance of my offline and online frameworks. In the first experiment, we investigate the effectiveness of the offline framework and in the second experiment, we compare online framework's performance with variants of offline framework in implicit negative link detection task.

Dataset

We crawl tweets by members of the 56th and 57th Parliament of United Kingdom using GET user_timeline function of Twitter API. Each parliament member usually self-describes when the account is associated with their parliament identity in their user profile. All of the accounts in the dataset are verified Twitter accounts.

- **56th Parliament Dataset** covers 1,074 user pairs sampled from 400 members of the 56th Parliament of United Kingdom on Twitter. Polarity of each user link is annotated using three human annotators.

- **57th Parliament Dataset** covers 1,349 user pairs sampled from 561 members of the 57th Parliament of United Kingdom on Twitter. Polarity of each user pair is annotated by yearly snapshots. It spans three snapshots, namely, "2016 $\rightarrow$ 2017", "2017 $\rightarrow$ 2018", and "2018 $\rightarrow$". "2018 $\rightarrow$" snapshot spans the first two months of 2018. The task of annotation involves three human annotators. Details of the annotation are explained in Section 3.1.2.

Users who do not participate in any textual user interaction are removed from the dataset. For implicit negative link detection task, it is essential to obtain labels

Table 3.2: Dataset Statistics

	56th Parliament	57th Parliament		
		2016 → 2017	2017 → 2018	2018 →
Textual interactions	4,217	1,297	3,947	1,459
Interacting user pairs	1,074	460	1,099	602
+/− links	948/126	433/27	977/122	526/76
$(+,+,+)$ triads	732	150	1257	294
$(+,+,-)$ triads	61	0	72	15
$(+,-,-)$ triads	68	12	126	30
$(-,-,-)$ triads	11	0	3	0
Sentiment Tokens	1,225	543	1,064	615

for the links between users to (1) test the effectiveness of my algorithm, (2) have a grasp on the effect of the parameters. Thus, we hired three graduate students for my annotation task. An overview of the annotated datasets can be seen in Table 3.2. Tweet ids, user ids and annotated user links of both datasets used in my experiments can be retrieved from `www.public.asu.edu/~mozer/NRHMdata.tar.gz`.

To evaluate the performance of the offline algorithm, we experiment with the 56th Parliament dataset and an aggregated single-view of the 57th Parliament dataset over three snapshots. To aggregate human annotated labels of the 57th Parliament dataset, we use the latest available label in three snapshots for each link. For online algorithm's experiments, we use the 57th Parliament dataset as it is.

Annotation Task

For 56th Parliament dataset, we aggregated all the textual interactions (i.e. tweets identified as mentions and reply to's) of user pairs. For 57th Parliament we aggregated interactions into three snapshots("2016 $\rightarrow$ 2017", "2017 $\rightarrow$ 2018", "2018 $\rightarrow$"). We filtered the data to include textual interactions which contains a single user mention to avoid the confusion as it is ambiguous which user is addressed in the multiple mentions case.

We requested 3 graduate students who had knowledge of UK politics to rate the polarity of the interactions between two politician accounts. For a pair of users, we have provided all textual interactions, political party affiliations, and retweet counts between the users to help annotators assess the polarity of the link better. After retrieving all the answers from three annotators, we assigned the polarity labels using majority voting.

We analyzed the labelers inter-rater agreement using Cohen's Kappa (Landis and Koch, 1977) and Fleiss' Kappa (Fleiss, 1971) to ensure annotation quality. Two-way inter-rater agreement is nearly perfect according to (Landis and Koch, 1977) with Cohen's Kappa scores calculated as 0.810, 0.898 and 0.911. Fleiss' kappa is reported as 0.731.

Finally, we remove the neutral user links as they are not covered by my problem formulation.

Offline Framework Performance

My first experiment aims to demonstrate the implicit negative link detection performance of SocLS-Fact in offline settings. To assess the performance of my method, we explain and compare with two existing state-of-the-art matrix factorization ap-

proaches along with three other baseline predictors we define as follows:

- **Random**: Motivated by (Liben-Nowell and Kleinberg, 2003), this method predicts signs of user links randomly.

- **Only Sentiment**: This predictor infers the polarity of user pairs' links using only textual interaction. Sum of the inverse distance weighted sentiment values (+1, -1) of words in textual interactions is given as the polarity of the link between user pairs.

- **Only Link**: This predictor infers user pairs' links as positive if there is any historical platform-specific positive interaction between them and negative otherwise.

- **NMTF[(Ding *et al.*, 2006)]**: This predictor is a simple non-negative matrix tri-factorization method without any regularizers of sentiment lexicon, link prior or social balance.

- **SSMFLK[(Li *et al.*, 2009)]**: Proposed as sentiment classification method, it is a semi-supervised matrix factorization framework utilizing prior sentiment lexicon knowledge. This method is similar to SocLS-Fact method, however, it does not encode platform-specific positive interaction between users or social balance theory.

- **LS-Fact**: This predictor is a variant of the proposed method but it does not embed social balance theory. It is introduced as a baseline to show the effect of social balance regularizer.

Methods using regularizer coefficients (i.e. SSMFLK, LS-Fact, SocLS-Fact) are experimented with all powers of 10 from -6 to 2 and the best performance is reported.

Table 3.3: Offline Implicit Negative Link Detection Performance on the 56th and 57th Parliament Datasets

	56th Parliament Dataset			57th Parliament Dataset		
	Prec.	F-meas.	Acc.	Prec.	F-meas.	Acc.
Random	0.1450	0.2344	0.5317	0.1707	0.2709	0.5664
SSMFLK	0.3143	0.4490	0.7737	0.3708	0.4599	0.8426
Only Sentiment	0.4010	0.4892	0.8464	0.3364	0.4207	0.8333
Only Link	0.6032	0.6726	0.9062	0.5312	0.6733	0.9021
NMTF	0.6741	0.6973	0.9264	**0.8243**	0.5622	0.9271
LS-Fact	0.6976	0.7059	0.9302	0.7091	0.7548	0.9434
SocLS-Fact	**0.7236**	**0.7149**	**0.9339**	0.7742	**0.8**	**0.9553**

Evaluation Metrics

We use three gold-standard metrics, namely; accuracy, precision, and F-measure to evaluate my method. Scores are reported in terms of my method's detection performance on the negative links. We do not report recall explicitly as we emphasise quality over quantity; retrieving meaningful negative links is the most important task in this work as suggested for many tasks in (Wang *et al.*, 2011). The change in recall can be indirectly observed through F-measure. Although we present the accuracy for reader convenience solely focusing on accuracy may be misleading considering the imbalanced nature of my dataset. Hence, we focus mainly on precision and F-measure throughout the discussion of my results.

Results

An overview of the implicit negative link detection performance of the proposed and baseline methods can be found in Table 3.3. As can be clearly observed through the table, performance increase is consistent among all three metrics: precision, F-measure and accuracy. Important findings are reported below:

- Encoding the sentiment information using SSMFLK improves the performance over the random classifier.
- An interesting finding can be observed when "only sentiment" predictor is used. It yields better results than SSMFLK due to its deterministic nature; whereas SSMFLK may be highly affected by the random starting conditions.
- Only link predictor gives much better results than using just the sentiment information. A steep increase in all three metrics is evident that prior platform specific positive interaction is a very strong signal that the link between users is not negative.
- Co-optimising the link information with sentiment information in LS-Fact framework results in superior performance compared to both only link and only sentiment predictors.
- Finally, my framework, SocLS-Fact obtains the best results by incorporating the social balance theory into the framework. SocLS-Fact performs slightly better than LS-Fact thanks to the user link triads following social balance theory in formation. F-measure performance contribution of socially balanced triangles is higher for 57th Parliament dataset than 56th, as higher ratio of socially balanced triangles can be observed in Table 3.2.

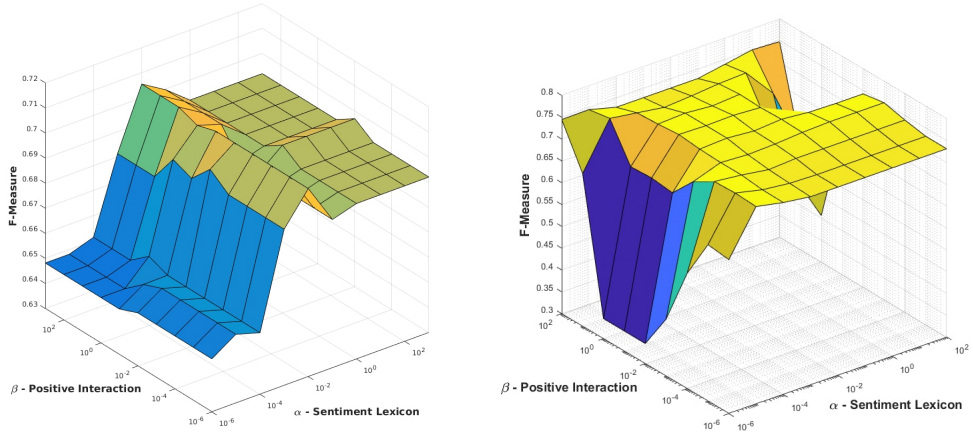
Parameter Analysis

It is essential that my framework performs effectively under different parameter settings. So, we experiment with various values of α, β, and γ then report the performance in terms of F-measure scores. Best performance was obtained using the parameters $\alpha = 10^{-2}$, $\beta = 100$, and $\gamma = 10^{-1}$ for 56th Parliament Dataset, $\alpha = 10$, $\beta = 10^{-5}$, and $\gamma = 10^{-5}$.

Figure 3.3 demonstrates the effect of sentiment lexicon parameter α and prior platform-specific positive interaction parameter β when the social balance regularizer γ is fixed at 0. α and β are tweaked as powers of 10 between -6 to 2. Parameters out of this range gives very low F-measure scores thus excluded.

- SocLS-Fact is robust to changes of α and β when α is in the range of 10^{-5} and 1 as F-measure does not differ more than 0.07 for both datasets.
- Lower values of α yield the lowest F-measure scores. Performance sharply increases when α is incremented from 10^{-6} towards 10^{-2}.
- Change of β creates rather stable results for any given α in 56th Parliament dataset and αs between 10^{-5} and 1 in 57th Parliament dataset.

Figure 3.4 shows how social balance regularizer γ affects the performance when the other parameters are fixed at optimal values, 10^{-2} and 100 for 56th Parliament and 10 and 10^{-5} for 57th Parliament dataset, respectively. γ is supplied incrementally as powers of 10 between -5 to 1. As the chart shows, SocLS-Fact is robust also to changes of γ performing in a F-measure margin of 0.025 for 56th Parliament dataset. The margin for 57th Parliament dataset is 0.1. Both chart shows that with the optimal setting of γ, social balance theory can contribute to achieve a superior performance



(a) 56th Parliament Dataset (b) 57th Parliament Dataset

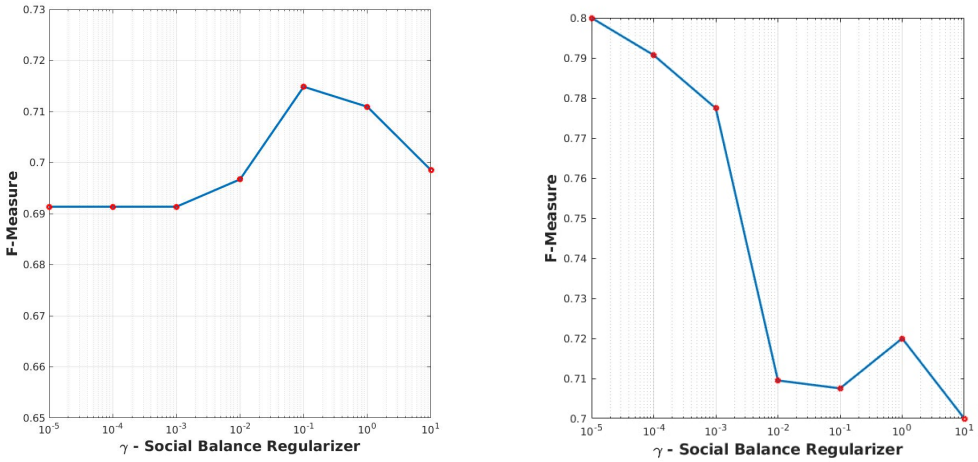
Figure 3.3: Effect of Regularizer Coefficients

in implicit negative link detection task. The optimal γ parameters are 10^{-1} for the 56th Parliament dataset and 10^{-5} for the 57th Parliament dataset.

Online Framework Performance

In this section, we discuss the performance of the online framework by presenting comparisons with variants of the offline framework. As mentioned in the previous sections, conventional ways of dealing with streaming data using an offline methodology usually involves either computing everything from scratch, or ignoring the historical data. Both extremes have their disadvantages. To show the trade-offs between these two approaches, we propose experimenting with the following two baselines and my online method;

- **SocLSFact** detects signs of the links only based on the current snapshot data.
- **SocLSFact** [**A**] detects signs of the links based on aggregation of current and all previous snapshots.



(a) 56th Parliament Dataset (b) 57th Parliament Dataset

Figure 3.4: Effect of Social Balance Regularizer Under Optimal Positive Prior and Sentiment Lexicon Regularizers

- **SocLSFact (Online)** detects signs of the links based on signs of the detected implicit links in the previous snapshots and factorise only the current snapshot data.

In this experimental setup, we utilise 57th Parliament dataset which is labelled in three snapshots. Results are reported based on the last snapshot (2018 $\rightarrow$) data. SocLSFact works only on the last (2018 $\rightarrow$) snapshot data to detect implicit links in it. SocLSFact [A] aggregates all three snapshots into single view and detect implicit links, accordingly. SocLSFact (Online) model uses SocLSFact outputs of two previous snapshots (2016 $\rightarrow$ 2017 and 2017 $\rightarrow$ 2018) for temporal smoothing and factorises only the last snapshot (2018 $\rightarrow$) data to detect implicit links. Parameters $(\alpha, \beta, \gamma, \tau)$ are explored with all powers of 10 from -6 to 2.

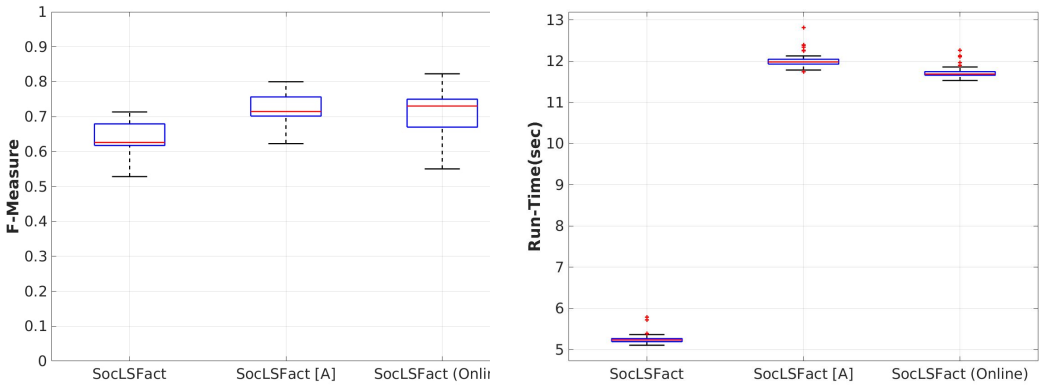
Table 3.4: Online Implicit Negative Link Detection Maximum Performances on the Last Snapshot of the 57th Parliament Dataset

	2018→		
	Prec.	F-Meas.	Acc.
SocLSFact	0.6588	0.7134	0.9233
SocLSFact [A]	0.7742	0.8	0.9553
SocLSFact (Online)	**0.8406**	**0.8227**	**0.9574**

Results

An overview of the implicit negative link detection in online settings can be seen in Table 3.4 and in Figure 3.5a. In terms of maximum performance, SocLSFact (Online) model performs better than both SocLSFact and SocLSFact [A] baseline models. In other words, while modelling SocLSFact for online settings, temporal smoothing among other two options increases the performance in all three metrics. SocLSFact [A], which aggregates previous snapshots' data as they are, is still the better choice than running factorization only on the last snapshot. SocLSFact (Online) achieves 15% higher F-measure, 28% higher precision, and 4% higher accuracy than SocLSFact and achieves 3% higher F-measure, 9% higher precision, and 0.2% higher accuracy than SocLSFact [A] in implicit negative link detection.

To better evaluate the trade-off between run-time and effectiveness of these three methods, we run each method 100 times with parameters α,β,γ, and τ set to 0.01, arbitrarily. We report their run-times in Figure 3.5b. The online framework runs approximately %3 faster on average than SocLSFact [A]. Furthermore, it shows 1% higher F-measure performance on average than SocLSFact [A] method and 11% higher than SocLSFact on average. Much shorter run-time of SocLSFact method should be



(a) Performance Comparison (b) Run-Time Comparison

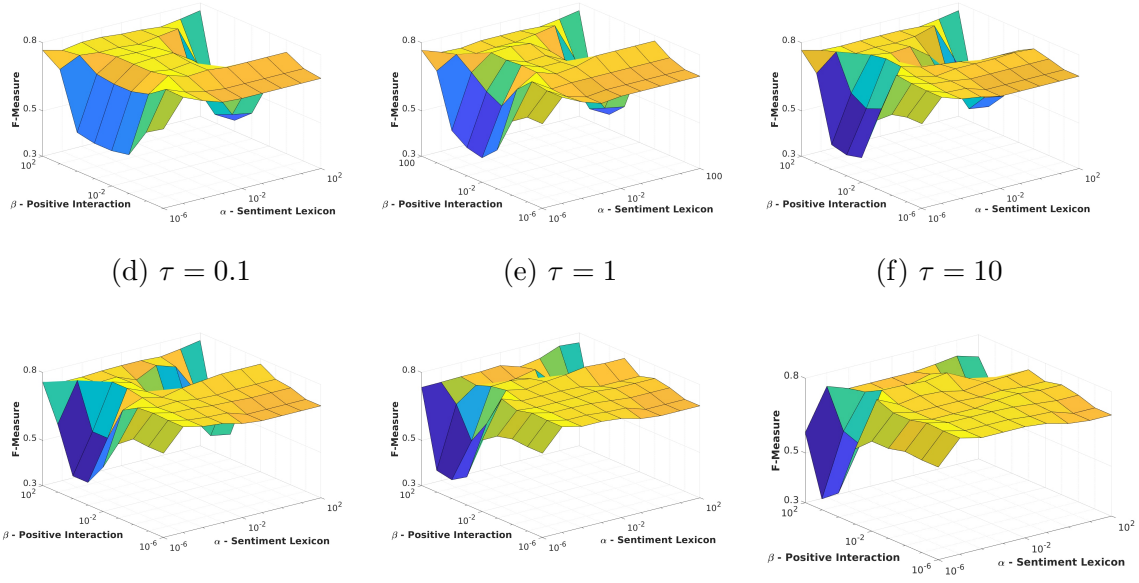
Figure 3.5: Offline & Online Algorithms' Performance Comparison for 57th Parliament Dataset. Online SocLSFact Achieves Competetitive Performances While Having Shorter Run-times.

noted. However, it is not significant as it factorises a much smaller size of data, and shorter run-time is expected.

Temporal Smoothing Parameter Analysis

In this section, we discuss the effect of the temporal smoothing parameter τ in the online framework. We introduce the parameter τ to weight the importance of previous snapshots' detected implicit links. We expect it to behave as a temporal regularizer in the case of data sparsity and any other type of abrupt changes. To evaluate the behaviour of online framework under different τ's, we present the F-measure performances of different parameter settings in Figure 3.6. When the value of τ gets larger, variation in the performance due to the positive prior and sentiment lexicon regularizer parameters decrease. Tweaking τ does not improve the F-measure performance under optimal α and β choices, significantly.

(a) $\tau = 0.0001$ (b) $\tau = 0.001$ (c) $\tau = 0.01$



(d) $\tau = 0.1$ (e) $\tau = 1$ (f) $\tau = 10$

Figure 3.6: Effect of Temporal Smoothing Parameter τ. Deviation in F-measure Decreases with Increasing τs.

3.2 Applications

In this section, we present two applications for online political networks which would not be possible or as effective without detecting implicit negative links, first. As the first application, we demonstrate the added value of implicit negative links in community detection task. Second, we qualitatively analyze the key role of implicit negative links in disclosing group polarization dynamics.

3.2.1 Dataset

Before going into the details of the applications, first, we introduce the datasets we utilise in the application settings. We crawl three datasets from politician accounts of

Twitter from United Kingdom, United States, and Canada using GET user_timeline function of Twitter API. They consist of either parliament member accounts of their country or prominent political figure accounts. Each politician account in the dataset either self declares their political party membership in their user profile or has the abbreviation of the political party in their user name as suffix or prefix. Baseline communities are constructed according to each account's self-identification of political party memberships.

- **United Kingdom Dataset** covers Twitter accounts of 421 prominent members of 56th United Kingdom Parliament from 5 major political parties, namely, Conservative Party (Cons), Labour Party (Lab), Scottish National Party(SNP), Liberal Democrats (LibDem), and United Kingdom Independence Party (UKIP).

- **United States Dataset** covers 596 prominent political figures' Twitter accounts from Republican and Democrat Party.

- **Canada Dataset** covers Twitter accounts of 136 members of 41st Parliament of Canada from 5 major political parties, namely, Liberal Party of Canada, Green Party of Canada, Conservative Party of Canada, New Democratic Party, and Bloc Quebecois (BLOC).

Further statistics about the datasets can be found in Table 3.5, and tweet ids and user ids can be downloaded from `www.public.asu.edu/~mozer/NRHMdata.tar.gz`.

3.2.2 Community Detection

To evaluate the added value of negative links we test the contribution of negative links in detecting the underlying political communities in the dataset. To that end, we employ a simple spectral clustering algorithm for signed networks. We feed both

Table 3.5: Dataset Statistics

	United Kingdom	United States	Canada
Textual interactions	18,903	31,276	5,001
Users	400	596	136
Interacting user pairs	3,367	6,114	1,291
Sentiment tokens	1,685	1,987	1,078
# of communities	5	2	5

unsigned links of the given dataset and predicted signed links by my framework SocLS-Fact separately. We employ United Kingdom, Canada and United States datasets to evaluate the performance of my method. Parameters for SocLS-Fact are set to be the ones which minimises the residual error of the objective function.

Spectral Clustering on Signed Networks

As proposed by (Kunegis *et al.*, 2010), we define the laplacian matrix $\overline{L}$ of an adjacency matrix A of signed network as;

$$\overline{L} = \overline{D} - A \tag{3.8}$$

where

$$\overline{D}_{ii} = \sum_{j \sim i} |A_{ij}| \tag{3.9}$$

The rest of the clustering framework follows the standard spectral clustering as given in Algorithm 4.

Algorithm 4 Spectral Clustering Algorithm for Signed and Unsigned Networks

Input:{$\overline{L}$ (signed) or L (unsigned)}

Output:{Clusters $C_1, C_2, ..., C_k$}

1: Find the smallest k eigenvalues of $\overline{L}$ (or L).

2: Form matrix U as $[v_1, v_2, ..., v_k]$ with corresponding k eigenvectors as columns.

3: Cluster the rows of U into $C_1, C_2, ..., C_k$ by applying k-means.

Evaluation Metrics

To evaluate the contribution of predicted negative links in community detection tasks, we make use of two well known clustering quality metrics, namely; purity and normalised mutual information(NMI).

Community Detection Results

Table 3.6 shows the community detection results for United Kingdom, United States and Canada datasets. Inclusion of the predicted negative links of my framework consistently contributes to the performance of community detection tasks.

For experiments having matching k's with number of ground-truth communities of datasets, following observations are made. Significant improvement in all three metrics can be observed in the results of United Kingdom and Canada datasets. United States dataset reveals even more intriguing results: purity increases by %25, and NMI by %241. This finding suggests that addition of negative links does not only boost the performance but can be of very critical importance for community detection.

Another observation we make is the higher contribution of the predicted negative links in community detection tasks when the number of clusters k given to spectral clustering algorithm is equal to the ground-truth community count of the datasets.

Table 3.6: Contribution of the Detected Implicit Negative Links in Community Detection Tasks with Varying k's.

k		United Kingdom		Canada		United States	
		Purity	NMI	Purity	NMI	Purity	NMI
2	Unsigned Links	0.4818	0.3829	**0.8013**	**0.5485**	0.7445	0.1863
	SocLS-Fact Links	**0.4844**	**0.4052**	0.7947	0.5057	**0.9294**	**0.6364**
3	Unsigned Links	0.8333	0.6770	**0.9338**	**0.7481**	0.8622	0.3962
	SocLS-Fact Links	**0.8411**	**0.6854**	**0.9338**	0.7473	**0.8807**	**0.4709**
4	Unsigned Links	**0.9167**	0.7838	0.9338	0.7026	0.8605	0.3770
	SocLS-Fact Links	**0.9167**	**0.7859**	**0.9470**	**0.7424**	**0.8773**	**0.4268**
5	Unsigned Links	0.9167	0.7794	0.9272	0.6803	0.8706	0.3935
	SocLS-Fact Links	**0.9427**	**0.8041**	**0.9536**	**0.7456**	**0.8790**	**0.4304**

The ground-truth community counts for United Kingdom is 5, Canada is 5 and United States is 2 as described in 3.1.2. Most increase by percentage in all three metrics is achieved when k = 5 in United Kingdom and Canada, and k = 2 in United States dataset. This further suggests the informativeness of the predicted negative links in implying the exact number of underlying communities.

3.2.3 Group Polarization

To show another powerful use-case of my framework SocLS-Fact, we set up an experiment that quantifies the group polarization patterns over time among UK politicians who interact with each other in Twitter. We demonstrate how my method and

predicted negative links can be used to represent political dynamics such as emerging and diminishing rivalries or coalitions among political party members. We visualise and qualitatively analyze the detected polarities of links among groups and their change over time.

We sample United Kingdom dataset and create three datasets spanning different time intervals to represent political climate change on social media. First dataset covers the whole timespan which we treat as the overall political climate among members. This dataset constitutes my baseline for detecting divergences from conventional behaviours of political party members in the sampled representative data. Second dataset spans all tweets in 2015. General election held on May, 5 2015 is considered to be the major political event of the year. We refer to the second dataset as general election dataset for future references. Third dataset spans the time interval of first 6 months of the year 2016. Brexit unequivocally being the major political event of that time interval, we refer to the third dataset as Brexit sample for future references.

After sampling these three datasets, we run offline SocLS-Fact algorithm and detect the polarity of each user link. Links that connect users are aggragated with users' affiliated political parties. Aggregation yields the polarization scores among and within political parties. Positive scores are mapped to hues of greens while negative scores are mapped to reds. Darker color means higher polarity. White color stands for the non-existence or very few links between groups, thus omitted. The overview of the resulting polarity among and within groups for each of the three datasets is presented in Figure 3.7.

Table 3.7: Popular Hashtags in Textual Interactions of Two Samples from the United Kingdom Dataset

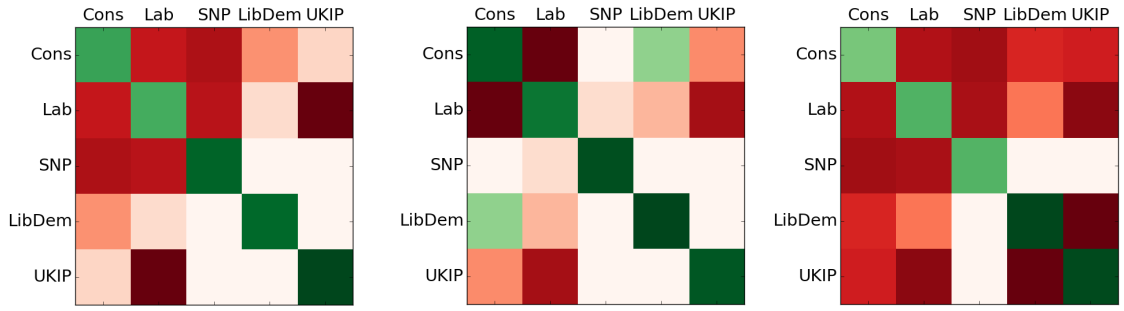
Sampled Datasets	Popular Hashtags
General Election	#GE2015, #labourdoorstep, #GE15, #VoteSNP, #Labour, #VoteLabour, #bedroomtax, #NHS, #PMQs, #voteSNP
Brexit	#StrongerIn, #Brexit, #EUref, #VoteLeave, #labourdoorstep, #Remain, #LabourInForBritain, #BackZac2016, #BothVotesSNP, #EU

General Election Dataset

Major event of the 2015 which this dataset covers is the United Kingdom general election 2015 as implied by the popular hashtags presented in Table 3.7. It took place on May, 5 2015. Conservative Party and Labour Party was the prominent candidates of winning the election. Government before the election was a coalition between Conservative Party and Liberal Democrat Party. Further background information about United Kingdom political parties can be obtained from (Moran, 2015).

Brexit Dataset

The biggest political event of the first 6 months of the year 2016 that Brexit Dataset covers, is clearly the European Union (EU) Referandum [(Hobolt, 2016)] that took



(a) Overall Political Climate (b) General Election Dataset (c) Brexit Dataset

Figure 3.7: United Kingdom Link Prediction Results for Political Parties for Various Time Frames. The Darker the Color is the Higher the Positive or Negative Polarity is among Two Parties.

place on June, 23 2016. UKIP and some politicians from Conservative party supported leaving the EU. On the opposite side of leave campaign, SNP, Labour Party, Liberal Democrats and part of the Consertavite Party were for staying in the EU. UKIP was a prominent political actor in the campaign. As implied by the popular hashtags used in the textual interactions between users, the dataset also covers London mayoral election (i.e. #BackZac2016) and Scottish Parliament Election (#BothVotesSNP). The election in Scotland resulted as a victory for SNP.

Tracking the Divergence of Political Parties From Overall Behaviour

In this section, we elaborate on how much polarization between groups deviate from their overall representation in the full dataset. Findings can be summarised as;

- Comparing Figure 3.7a and Figure 3.7b shows the increasing positive link ratio in inner-party links. (Nooy and Kleinnijenhuis, 2013) suggest that if two politicians belong to the same political party, they are more likely to support each other in an election season as the partisanship increases.

Tracking the Temporal Dynamics of Polarization among Political Parties

To evaluate the performance of the tracking the temporal dynamics of polarization between groups, we qualitatively analyze the polarity shifts from 2015 to 2016 between groups.

- Inner group positive link ratio of Conservative Party members decrease from 2015 (Figure 3.7b) to 2016 (Figure 3.7c) which can be explained by the members of the party diverging apart by having different point of views for EU Referandum.

- The rivalry between Conservative Party and Labour Party members dissolves slightly in 2016, because they were the two most prominent competiters in the general election and considerable amount of two parties' members campaigned for the same voting stance on Brexit election.

- The coalition in 2015 between Conservative Party and Liberal Democrats shifts to rivalry in 2016. It may be due to the coalition government that still existed in 2015 but were not formed again after the election.

- Rivalry increases between UKIP and other parties in Brexit dataset compared to General Election dataset. It can be explained by the EU Referandum in which UKIP was a leading figure.

Chapter 4

MEASURING THE POLARIZATION IMPACT OF AUTOMATED ACCOUNTS

4.1 Methodology

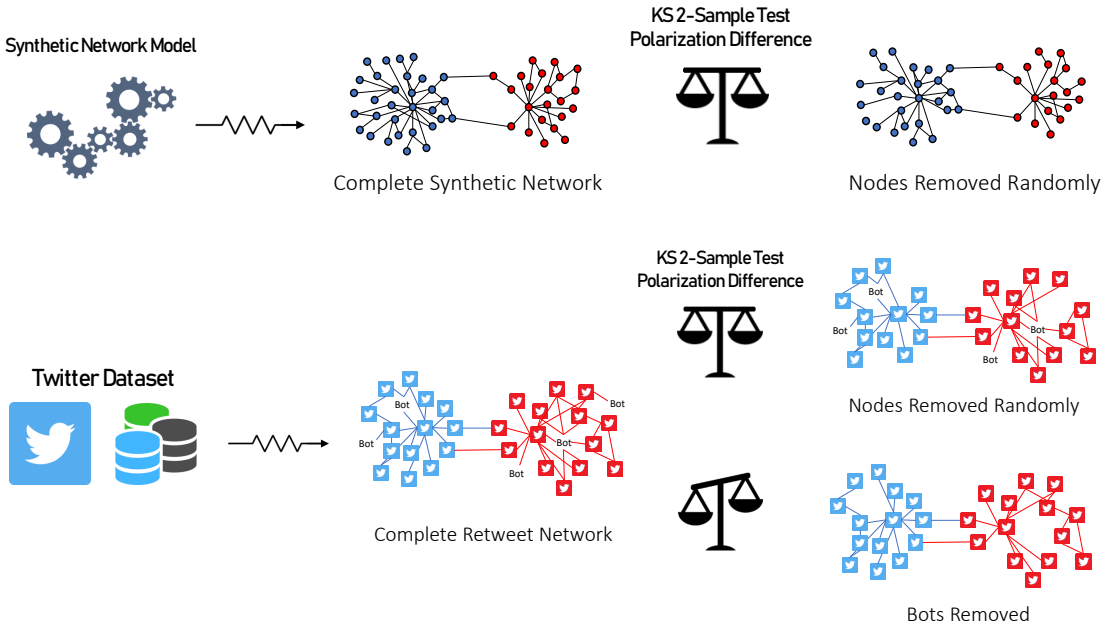


Figure 4.1: Proposed Methodology for Measuring the Effect of Automated Accounts

My methodology consists of several essential pieces to ensure the robustness of my measurements on the polarization impact of the automated accounts. Despite my main experiments are conducted on Twitter data, (1) I extend a directed scale-free graph model to generate polarized networks and test the stability of the polarization metric I utilize on the generated synthetic graph. Once the stability of the polarization metric is established, (2) I crawl a month of Twitter data associated with the Parkland school shooting incident to construct my main dataset. (3) I assign political labels to Twitter users with the help of a supervised classification task, then (4) identify

the automated accounts using a third-party state-of-the-art bot detection tool. From this labeled network, (5) I measure the polarization of the whole network, and the polarization of the network with random accounts removed, and the polarization of the network with bot accounts removed. A comparative analysis between the polarization of these three networks quantifies the effect of the automated accounts on the ecosystem.

To provide a more granular context and understanding, I conduct content analysis and focus on observational impact differences of varying types of automated accounts (e.g. self-identifying automated accounts).

4.1.1 Generating Synthetic Polarized Networks

I initialize a synthetic network with two separate Erdős-Rényi random network models. These initial two sub-networks correspond to the initial stages of two political sides. Then, I adopt a directed scale free graph model (Bollobás *et al.*, 2003) and modify it to be able to generate polarized networks. Details of the algorithm can be seen in Algorithm 5. Notice that my contribution to (Bollobás *et al.*, 2003) is the polarization parameter ρ. When the model connects two nodes to each other (new or old), it depreciates the effect of the indegrees and outdegrees of the nodes that are at the opposite side of political spectrum by ρ. Thus, nodes show political homophily in their connections besides their preferential attachment to higher degree nodes. With the help of the polarization parameter ρ, I are able to generate scale free directed networks in different levels of polarization. Later, I use this model to test my hypothesis on synthetically generated polarized networks.

Algorithm 5 Synthetic Polarized Network Generation

Input: $\alpha, \beta, \gamma, \delta_{in}, \delta_{out}, N, \rho$, where $\alpha + \beta + \gamma = 1$ and $0 \leq \rho \leq 1$

$G_L, G_R \leftarrow Erdős–Rényi()$

$G \leftarrow G_L \cup G_R$

while $|G| < N$ **do**

    with probability α:

    Draw a side from $\{L, R\}$ for a new node v.

    Add the node v and an edge to an existing node w from v, where w is chosen according to $d_{in} + \delta_{in}$ for w's in the same side with v, according to $\rho * (d_{in} + \delta_{in})$, otherwise.

    with probability β:

    Add an edge from an existing node v to an existing node w, where v and w are chosen independently, v according to $d_{out} + \delta_{out}$ and w according to $d_{in} + \delta_{in}$ if v and w are in the same side, according to $\rho * (d_{in} + \delta_{in})$, otherwise.

    with probability γ:

    Draw a side from $\{L, R\}$ for a new node v.

    Add a node v and an edge from an existing node w to v, where w is chosen according to $d_{out} + \delta_{out}$ for w's in the same side with v, according to $\rho * (d_{out} + \delta_{out})$, otherwise.

end while

4.1.2 Quantifying Polarization

Measuring the impact of bot accounts on network polarization requires us to quantify the polarization of a given network precisely. To this end, I refer to a recent study(Garimella *et al.*, 2018) and adopt their random walk controversy score.

$$RWC = P_{LL^{(+)}} P_{RR^{(+)}} - P_{LR^{(+)}} P_{RL^{(+)}} \tag{4.1}$$

where $P_{LL^{(+)}}$ is probability of a random walk starting from any left node (L) ending up at a central left node ($L^{(+)}$). Similarly $P_{RR^{(+)}}$ is probability of starting on any right node and ending on a central right node. $P_{LR^{(+)}}$, $P_{RL^{(+)}}$ follow the same definition and quantifies the probability of a walk crossing sides. To compute the aforementioned probabilities, Garimella et al. (Garimella *et al.*, 2018) suggest a simple Monte Carlo sampling of random walks over network. After having samples of walks they quantify the probabilities $P_{LL^{(+)}}$,$P_{RR^{(+)}}$,$P_{LR^{(+)}}$,$P_{RL^{(+)}}$ as follows;

$$P_{LL^{(+)}} = \frac{C_{LL^{(+)}}}{C_{LL^{(+)}}C_{LR^{(+)}}} \qquad P_{RR^{(+)}} = \frac{C_{RR^{(+)}}}{C_{RR^{(+)}}C_{RL^{(+)}}}$$

$$P_{LR^{(+)}} = \frac{C_{LR^{(+)}}}{C_{LR^{(+)}}C_{LL^{(+)}}} \qquad P_{RL^{(+)}} = \frac{C_{RL^{(+)}}}{C_{RL^{(+)}}C_{RR^{(+)}}}$$

where C stands for the count of walks falling into certain previously defined types. The RWC polarization metric returns values between +1(perfect polarization) and -1(no polarization).

4.1.3 User Classification

To decide the left and right side of the RWC algorithm, I develop a political classification task. To be able to set up a classification task, I first acquire third party intelligence from a crowd-sourcing platform that indicates the political leanings of news domains. I crawl news domains' political scale (left, center-left, center-right, right) from *mediabiasfactcheck.org*. This procedure equips us with 1,241 news domains and their political labels. [1]

Various studies have shown that social media users' political news diet is highly clustered according to their political leaning (pew, 2014). I also adopt a similar heuristic and label social media users based on domains of news articles they share. I execute a simple majority voting for each user based on what they share in their social

[1] `www.public.asu.edu/~mozer/bot_polarization/media_scales.zip`

Table 4.1: Bag-of-words Based and Network Based Classification Performances

		F1-Macro	Accuracy
Text	Random Forest	0.4438	0.7964
	GBM	0.6433	0.8403
	Logistic Regression	0.9101	0.9441
Network	Label Propagation	0.9552	0.9715

media posts. I use -2, for the left domains, -1 for the center-left domains, +1 for the center-right domains, and +2 for the right domains. I keep users having cumulative values greater than +2 and less than -2 as my training dataset. It provides us around 80K social media accounts and their 7M tweets labeled as left or right.

After garnering labeled social media accounts, I develop two separate classification tasks for classifying the rest of the users. Note that these users have not shared enough news articles for us to assess their political ideology. First, I use a label propagation algorithm on the retweet network informed by (Conover *et al.*, 2011). Second, I develop several text-based classification tasks and report each classifier's accuracy with five-fold cross validation in Table 4.1. Given the superior performance of label propagation algorithm, for the rest of the chapter I build my analysis upon its results.

4.1.4 Automated Account Detection

To detect the automated accounts in my dataset I register to the Botometer API provided by Indiana University(Davis *et al.*, 2016). [2] I query a random sample of 260K accounts from my dataset. I tag accounts who have a score over 0.5 as automated and the rest as not automated. The API returns 25K accounts

[2]https://botometer.iuni.iu.edu

flagged as automated (%10) which agrees with the previous literature's findings on the prevalence of automated accounts on social media(Varol *et al.*, 2017).

4.1.5 Measuring the Impact

To measure the impact of automated accounts on network polarization, I set up an experiment as follows. First, I compute the polarization of complete retweet network. I run the RWC algorithm 1,000 times and report the distribution of polarization scores. Second, I compute the polarization of the sub-network without any automated accounts. I also run the RWC algorithm 1,000 times and report the distribution. Finally, I compute the polarization of the sub-network which is acquired by removing number of nodes equal to the number of automated accounts randomly. I run the RWC algorithm 1,000 times also and report the distribution. Then, I compare these three distributions pairwise and report the significance results of Kolmogorov-Smirnov 2-sample test. In my application, Kolmogorov-Smirnov d test(Massey Jr, 1951) assesses if two measured polarization score distributions come from different means and variances of underlying polarization distributions.

4.2 Experimental Results

I branch my analysis into two distinct sets of experiments. First, I set up experiments with artificially generated polarized networks. Second, I set up experiments on my focus study; Twitter dataset regarding the unfolding and aftermath of Parkland school shooting event.

4.2.1 Validating the Experimental Setup

In this set of experiments, I generate synthetic polarized networks emulating the retweet network of Twitter. First, I report the interplay of random walk contro-

versy(RWC) score with the polarization parameter ρ of previously introduced variant of directed scale free network model. I also check if RWC score is robust to the network size changes. My search space spans polarization parameter ρ values between 0.01 and 0.1 with increments of 0.01. It also spans number of nodes between 10,000 and 200,000 with increments of 10,000. I generate the synthetic networks based on the given ρ and number of nodes and with Erdős–Rényi random network having 100 nodes on each side with 0.33 edge probability.

I also experiment with parameters beyond what I report here, but for the sake of brevity of my chapter, I only report results which spans the neighborhood of mean polarization score of my Twitter dataset ($\mu_{RWC} = 0.9067$).

RWC Score on Synthetic Polarized Networks

By tweaking the polarization parameter ρ and number of nodes parameters of the synthetic polarized network generation model, I generate 200 networks in various sizes and polarization levels. I compute the polarization score distribution for each by running the RWC algorithm 1,000 times on them. I observe two main patterns in my experiments. (1) Suggested polarization scoring algorithm (RWC) is in strong linear correlation with the ρ (MSE= 3.7702e−5, R2= 0.9889, pval$<$ 0.001). (2) Change in number of nodes of networks do not provide strong evidence for the polarization score of the underlying configuration (MSE= 0.0034 , R2= 8.0824e−6, pval= 0.2105). Results can further be investigated visually in Figure 4.2.

RWC Score after Random Node Removals

Assessing the impact of node removal on network polarization is crucial in my study. In this subsection, I evaluate the impact of node removals on the polarization of various synthetically generated networks. To emulate the observational data, I ex-

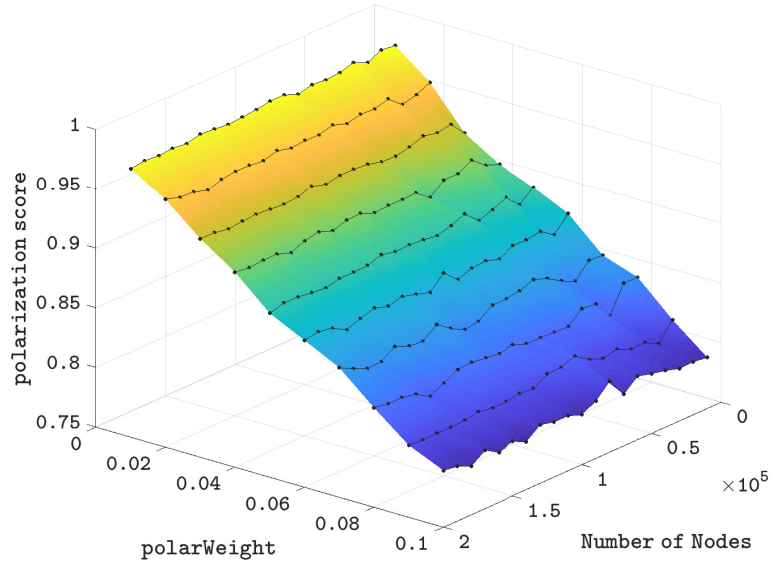


Figure 4.2: Polarization Score Measurements for Varying Synthetic Network Generation Parameters

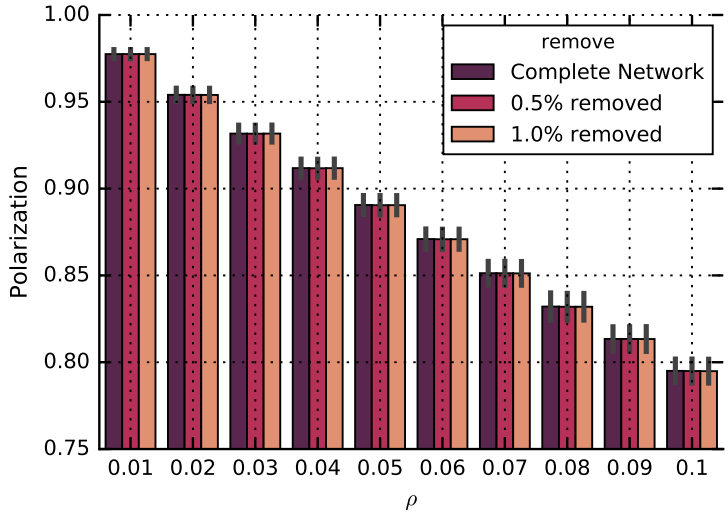


Figure 4.3: The Insignificance of Removals When Nodes are Randomly Removed from Synthetic Networks at Various Polarization(ρ) Levels

periment with removing 0.5% and 1% of nodes from networks. Note that the impact I measure on my observational Twitter dataset is already incurred by less than 0.1% of the accounts of the complete retweet network. My search space spans 400 experiment configurations derived from 20 network sizes, 10 polarization levels, and 2 removal rates.

Figure 4.3 shows the results of my experiments that networks with various levels of polarization do not experience significant polarization change when nodes are removed randomly from them. Out of 400 removal experiments, 388 of my experiments fail to reject the null hypothesis that polarization(RWC score) changes when nodes are removed randomly (Kolmogorov-Smirnov 2-sample test pval$>$ 0.005). Only seven of the experiments present decreased polarization while five of them exhibit increased polarization. I report the average polarization change as 0.0006 and the standard deviation of it as 0.0002 among 400 random node removal experiments.

4.2.2 Measuring the Impact of Automated Accounts

Dataset & Preprocessing

My dataset collection includes 3.7M users and their 25M tweets posted between February 1, 2018, and March 6, 2018. I purchased the dataset from GNIP Twitter by requesting tweets that contain any of the 140 words, subwords, and bigrams listed at `www.public.asu.edu/~mozer/bot_polarization/GNIP_query_list.txt`. I build retweet without edit network by compiling a network of 3.3M nodes and 16M edges.

I detect 25K automated accounts through Botometer API(Davis *et al.*, 2016). I note that this is not the comprehensive list of automated accounts in my dataset as I am constrained by the Twitter API. So, I query 260K accounts due to these resource

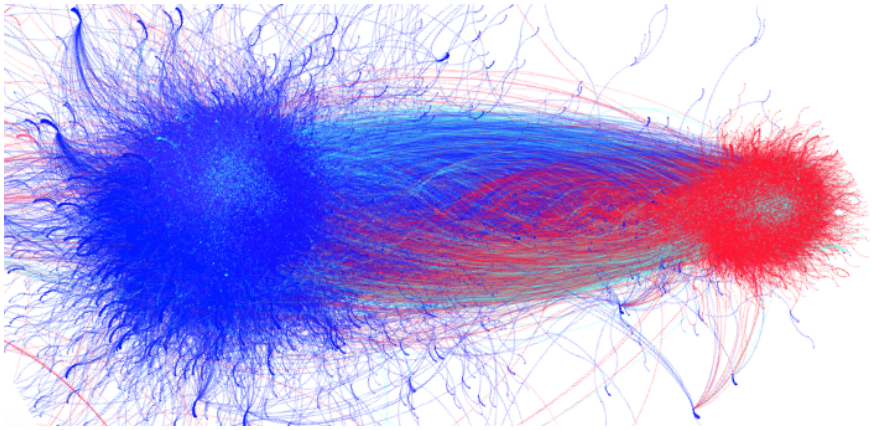


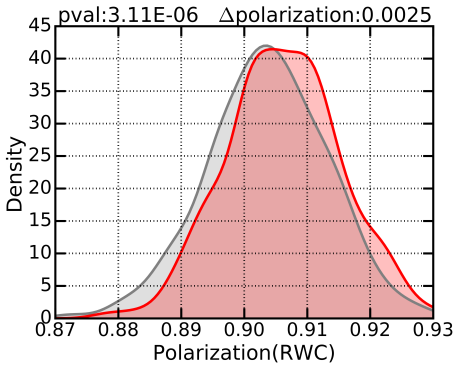
Figure 4.4: Retweet Network During and Aftermath of Parkland School Shooting. Light Blue Represents Automated Activity, Dark Blue Represents Left-leaning and Red Represents Right-leaning.

limitations. The effects I measure represents only a portion of the automated activity in my dataset.

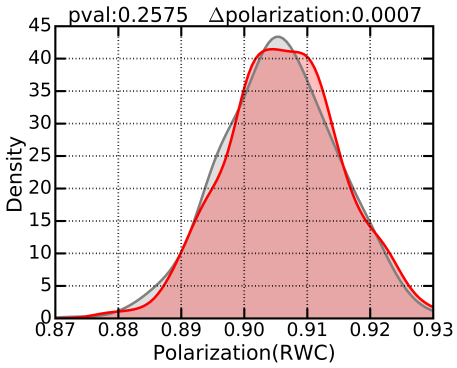
I use a label propagation approach to classify users' political leanings as discussed in the classification section under methodology and assess 3M left-leaning and 300K right-leaning Twitter users.

Overall Network Polarization Change

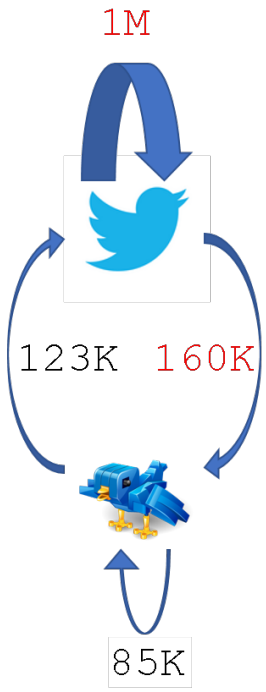
The major finding of this study is as follows; when automated accounts are removed from the retweet network of Twitter activity relating to the unfolding and aftermath of Parkland shooting event, polarization between left-leaning and right-leaning accounts decrease. When the same number of accounts removed randomly from the network the overall polarization score does not get affected significantly. More precisely, the difference between polarization measurements of complete retweet network and the network from which automated accounts removed is 0.0025. The same analysis yields 0.0007 polarization difference when done with random removals; in other



(a) Automated Accounts Removed



(b) Accounts Removed Randomly



(c) Retweet transitions between types of accounts.

Figure 4.5: Difference in Polarization between Complete Retweet Network (red) and When Automated Accounts Removed (grey) from it 4.5a. Indifference in Polarization When Same Amount of Nodes Removed Randomly 4.5b.

words, approximately 3.5 times less difference. The finding can be observed from Figures 4.5a and 4.5b. For possible explanations of this phenomena, I investigate my observational data further in the following sections.

Figure 4.5 presents the overall retweeting interaction between automated and not automated accounts. 160K retweets are initiated by 23K not-automated accounts towards 1.5K automated accounts, while 123K retweets initiated by only 7K automated accounts towards 5K not-automated accounts. This signals a hyper-active automated account activity to promote not-automated accounts' tweets through retweeting. On

the other hand, retweets acquired by automated accounts from not-automated accounts is greater in volume than the other way around (160K>123K). Indeed, if the automated activity was not getting any traction, it would not affect the RWC score, and the impact would not be at measurable levels.

Hashtag-Level Network Polarization Change

Hashtags are popular semantic atomic units that serve as topical hubs on Twitter. In this section, I extend my analysis to a lower granularity level and report hashtag level polarization impact of automated activity. First, I build 100 retweet networks of most participated hashtags from both political leanings. To quantify the participation from left and right sides, I use harmonic mean of the counts of users from both political sides $\frac{2|L||R|}{|L|+|R|}$. These retweet networks of most participated 100 hashtags span 80% of the total retweet activity containing at least a single hashtag in it, and 26% of the complete dataset.

Second, I measure the RWC score distribution of each hashtag network. Similar to the previous analysis, I remove automated accounts from the network and measure the RWC score distribution again. I find that among most participated 100 hashtags majority of them are less polarized retweet networks when automated accounts are removed. In particular, 65 of the hashtags presents a decrease in polarization when automated accounts are removed. Seven of them do not experience any statistically significant change and 28 experience increase in polarization.

Even though the majority of the hashtags (65%) experience a decrease in polarization when automated accounts removed, I observe a heterogeneity in the impact of automated account activity under different hashtags. This opens up a future direction for us and other researchers to study if there is any correlation between the impact and the properties of the hashtag (e.g. semantic, political leaning, emotion). Fig-

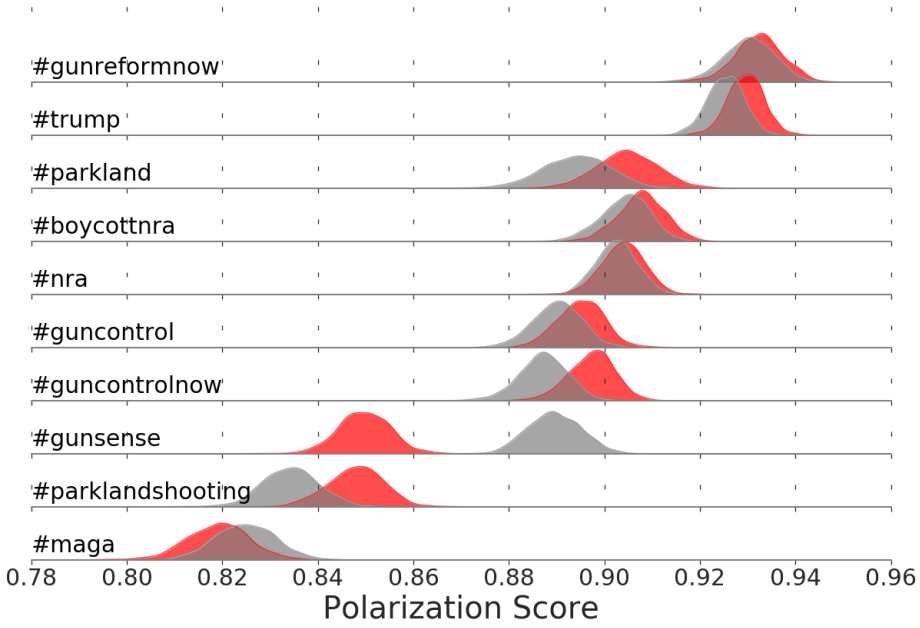


Figure 4.6: The Effect of Automated Accounts on the Hashtags that Attracted the Highest Participation from Two Sides. Red Distribution Represents the Polarization of Complete Retweet Network, and Gray Distribution Represents the Network's Without Automated Accounts.

ure 4.6 presents a brief summary of my findings (the most participated 10 hashtags' change) as a ridge plot.

Content Analysis

So far, I have presented the network polarization impact of automated accounts. This impact is measurable mainly due to the retweets that automated accounts are able to collect. In this section, I focus on the predictors of retweet count of automated accounts' tweets. In other words, I focus on the predictors of the impact. In particular, I study the content generated by these accounts and certain features of their user profiles. My interest is to investigate how predictors of retweet count in automated account case differ from non-automated ones. A natural experiment design would be a

direct comparison of two classes in my dataset, one being automated accounts' tweets and the other non-automated accounts' tweets. I refrain from setting up experiments which differentiate the automated accounts from non-automated accounts' features in my dataset since my classification of account type is collected through a supervised classifier already (Botometer API).

Instead, I present the predictors of retweet count of automated accounts' tweets and report how they differ from or align with previous literature on properties of general engaging content on social media(Bhattacharya *et al.*, 2014; Brady *et al.*, 2017; Suh *et al.*, 2010). In my dataset, I have 102,393 tweets posted by automated accounts. I design a negative binomial regression task with zero inflation to address the over-dispersion in my dataset. My regression's target variable is retweet count, and my predictor variables are as follows;

- *media_count* quantifies how many images or videos are embedded in the tweet,
- *mention_count* quantifies how many user handles are in the tweet,
- *followers_count* quantifies how many followers the automated account has
- *we* quantifies how many times a tweet contains the word I or variants defined by LIWC dictionary,
- *they* quantifies how many times a tweet contains the word they or variants defined by LIWC dictionary,
- *Moral-Emotional* quantifies how many times a moral-emotional word appears in the tweet. The word list is comprised of the intersection of moral words and emotional words dictionaries (Brady *et al.*, 2017).
- *Emotional-Only* quantifies how many times an emotional word appears in the

Table 4.2: Incidence Rate Ratios (IRR) Derived from Zero Inflated Negative Binomial Regression

	IRR	Lower 95%	Upper 95%
media_count***	1.5963	1.5388	1.6558
they***	1.1833	1.1431	1.2249
Moral-Emotional***	1.1105	1.0741	1.1480
Emotional Only***	1.0762	1.0589	1.0938
followers_count***	1.0001	1.0001	1.0001
we	1.0030	0.9726	1.0339
mention_count***	0.9831	0.9777	0.9885
Moral Only***	0.9512	0.9313	0.9715
url_count***	0.5522	0.5350	0.5699

$***p < 0.0001$

tweet. The word list is comprised of the distinctive emotional words that are not in the moral words dictionary at the same time.

- *Moral-Only* quantifies how many times a moral word appears in the tweet. The word list is comprised of the distinctive moral words that are not in the emotional words dictionary at the same time.

For my implementation, I use JMP Pro's Generalized Regression tool with zero inflated negative binomial regression task. My regression analysis yields *media_count* to be the most prominent predictor. It has the highest estimate (0.4675 ± 0.0187) among my predictors aligning with previous studies which also report the importance of visual media in the virality of tweets (Suh *et al.*, 2010). Second most prominent predictor is *they*. Use of the word and its variants in LIWC dictionary increase the

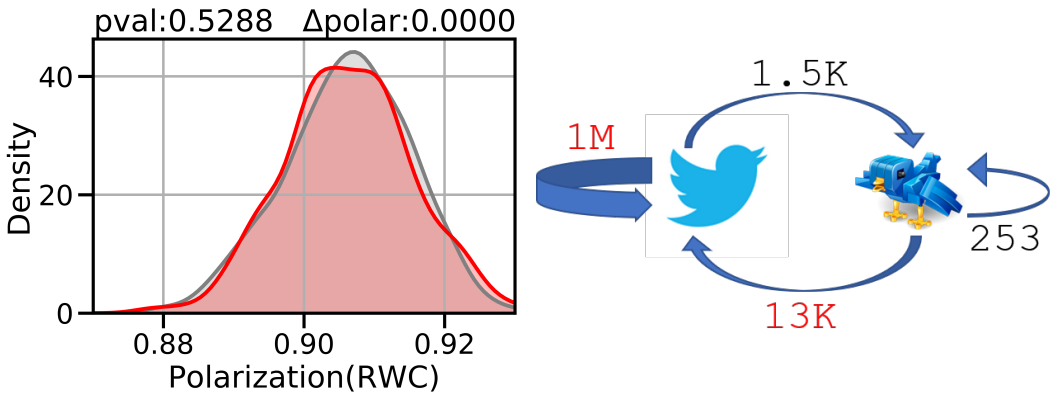
ratio of retweet count by 20%. Following previous literature on persuasive political communication studies(Hameleers *et al.*, 2017; Hameleers and Schmuck, 2017), success of blaming the other is a prominent phenomenon in the age of political populism coupled with social media. Furthermore, in alignment with previous works(Brady *et al.*, 2017; Valenzuela *et al.*, 2017), I find that Moral-Emotional words contribute (0.1048+-0.0170) more to retweet count than Emotional-Only words (0.0734±0.0083).

I report that moral-only words have a small negative predictor coefficient ($-0.0501\pm$ 0.0108) alongside *mention_count* (-0.0171 ± 0.0028). However, change of one unit in url count decreases the chance of a tweet to be retweet almost by half. For more detailed information about the incidence rate ratios of the independent variables, readers can refer to Table 4.2. In the light of these observations, I argue that the impact automated activity incurs are mostly in alignment with previous findings on the characteristics of engaging social media content.

When Automation is Self-Identified

In the previous experiments, I demonstrate the significant effect of the automated accounts on polarization, which necessitates the users to exercise more discretion to comprehend the whole context surrounding a message. Treating an automated message as a human-curated one may have a vast impact in the reception of the message. Thus, it is essential for the users to be aware if the message was indeed produced by an automated system.

I hypothesize that the utilization of a simple, explicit indicator for an account being automated may be an effective way to prevent any confusion. To test my hypothesis on observational data, I distinguish the automated accounts that explicitly uses the words "bot", "robot", or "chatbot" in their screen names or profile names (publicly visible account attributes) from the rest using the following regular expres-



(a) (b)

Figure 4.7: Retweeting Transitions between not Automated and Self-describing Automated Accounts. Insignificance of the Change in Polarization Change when Self-identifying Automated Accounts are Removed from the Retweet Network $pval > 0.05$.

sions:

- $*\_chatbot$
- $*\ chatbot$
- $chatbot\_*$
- $chatbot\ *$
- $[\hat{}\,] + Chatbot$
- $*\_robot$
- $*\ robot$
- $robot\_*$
- $robot\ *$
- $[\hat{}\,] + Robot$
- $*\_bot$
- $*\ bot$
- bot_*
- $bot\ *$
- $[\hat{}\,] + Bot$

I determine 1,802 self-identifying automated accounts matching these regular expressions. Figure 4.7b shows the retweet interactions within and among human-controlled accounts and self-identifying automated accounts. While human-controlled accounts retweeted self-identifying automated generated content 1.5K times, the opposite transition happened 13K times indicating 12 percent relative engagement from humans. This discrepancy is notable especially when compared to all human-automated interactions in Figure 4.5, where the relative engagement from human-

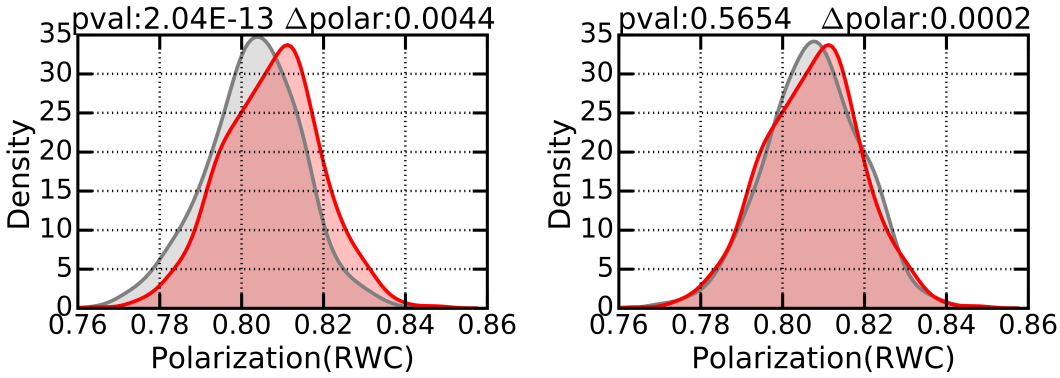
controlled accounts is 130 percent compared to all automated accounts. Thus, self identification clearly changes the dynamics in terms of human-automated account interactions.

Following this observation, I repeat the node-removal experiment described in previous section this time with 1,802 self-identifying automated accounts instead of all automated accounts. As can be observed in Figure 4.7a, removal of self-identifying automated accounts from the network do not result in a statistically significant change in polarization. This finding is the essential proof that the usage of a simple self-identification phrase can reverse the polarization effects created by the current ecosystem. Thus, I strongly recommend the adoption of a "automation identifier" such as a small robot symbol near the account name by the platforms as a simple and elegant way to diminish the unintended polarization effects.

Results with Text-based Political Leaning Classification

In this short section, I present my findings on the impact of automated accounts when political leaning classification is executed with text-only features. I use the labels acquired through the best performing text-based classifier in Table 4.1; logistic regression. In the complete retweet network, automated activity has 44 times more impact in polarization than the random effect (Figure 4.8). Notice that this impact of increase in polarization is much higher than what we report in the main text (three times). Furthermore, I find that 84% of the most popular debate related hashtags experience an increase in polarization with automated activity (Figure 4.9). Overall, the measured polarization impact of automated accounts among left and right leaning accounts is robustly evident approved by two fundamentally different political leaning classification approaches. I also note that when I repeat my analysis on self-disclosing automated accounts with text-based labels, I again find no evidence

that they contribute to polarization.



(a) Automated accounts removed (b) Accounts removed randomly

Figure 4.8: Polarization Impact When Political Leaning of an Account is Classified Using a Text Based Classifier.

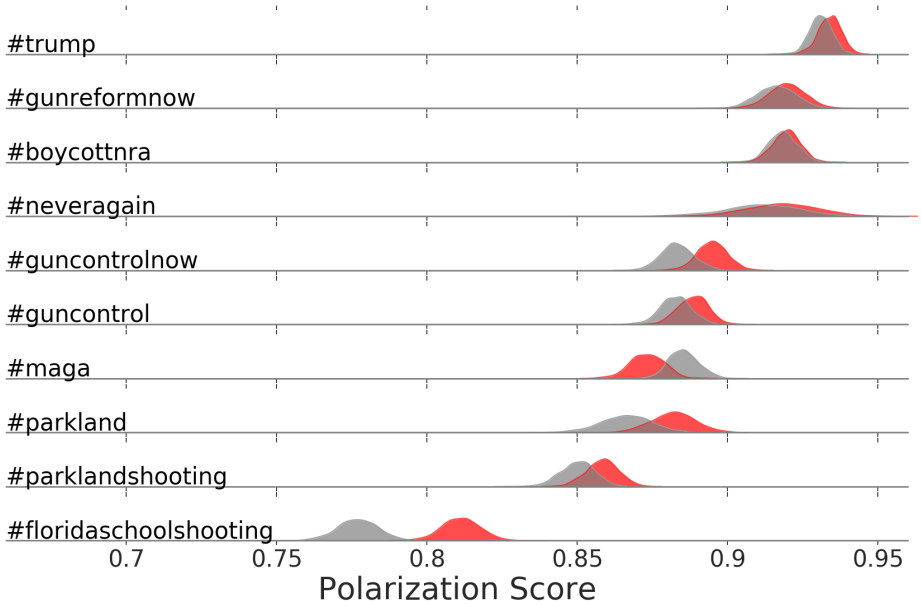


Figure 4.9: The Effect of Automated Accounts on the Hashtags that Attracted the Highest Participation from both Political Leanings when Political Leaning of Accounts are Classified through a Text-based Classifier.

Chapter 5

CONCLUSION

5.1 Summary of Contributions

In this dissertation, I proposed three computational solutions to the problems concerning three major aspects of online political networks. In the following three subsections I summarize my efforts.

5.1.1 Community Detection

First, I investigated the problem of detecting politically aligned communities of users on political Twitter networks. By jointly leveraging endorsement networks, social balance theory, and several facets of textual content generated by users, I show that sparsity problem of only-connectivity approaches can be overcome and I find that best complimentary textual feature to the social network among others(words, hashtags, url domains) is usage of words. By proposing a three non-negative matrix factorization based framework, I present a superior performance in politically aligned community detection task when augmented endorsement network and word usage are utilized together.

5.1.2 Implicit Negative Link Detection

Second, I focused on implicit negative links on social media platforms. Given that every political analytics task require identifying enmities, antagonisms, or any other form of adversarial relationship, I introduce a non-negative matrix factorization framework to detect the implicit negative linkages on political Twitter networks.

Since major social media platforms do not provide their users to connect with peers in a negative fashion, to detect implicit negative linkages, I turned to the social balance theory, sentiment analysis, and prior positive connections between users. By utilizing the aforementioned three pieces of information, I propose two frameworks for detecting implicit negative links in offline and online settings. I show the contribution of detecting implicit negative links on Twitter in two tasks; community detection, and tracking the opposition of political parties to each other on the Brexit referendum.

5.1.3 Impact of Automated Accounts on Network Polarization

Third, I propose a set of experiments to measure the polarization impact of bot accounts on Twitter. I find that on the issue of U.S. gun control debate, automated accounts contribute to increase in polarization three times more than the random effect. Furthermore, they have an impact of increased polarization in the 65% of the most popular debate related 100 hashtags. When I analyze the predictors of their tweets' endorsement (retweet count) levels, I find that usage of memes/videos, moral emotional words, they category words in LIWC dictionary, and the number of followers are the four major predictors of higher endorsement. However, when I conduct a similar analysis on the bot accounts where automated nature of them are self-disclosed in their profile (e.g. having a keyword "chatbot" in their profile name), polarization impact vanishes. Moreover, I find that when automated nature of the bot account is revealed in their profile, retweet counts of bot tweets by human controlled accounts decrease in 10-fold.

5.2 Future Directions

As a future work, I plan to investigate whether measured polarization increase impact of bot accounts on U.S. gun debate Twitter dataset holds for datasets collected

from other countries, other languages, and other time frames. Particularly, I am interested in investigating the fact that polarization impact vanishes when automation is self-identified. This finding can further motivate the ongoing research in automated account detection on social media and suggest a way to alleviate their unintended impact. I also plan to suggest a dynamic framework for detecting politically aligned communities when the data is in streaming nature. Finally, I work on the interplay of communities and word vector representations. I aim to investigate if word embeddings that are formed by the concatenation of community-level word representations help achieve better performance in NLP downstream tasks such as sentiment or political orientation classification.

REFERENCES

"Political polarization and media habits", Pew Research Center URL https://www.pewresearch.org/wp-content/uploads/sites/8/2014/10/Political-Polarization-and-Media-Habits-FINAL-REPORT-7-27-15.pdf (2014).

Adamic, L. A. and N. Glance, "The political blogosphere and the 2004 u.s. election: Divided they blog", in "Proceedings of the 3rd International Workshop on Link Discovery", LinkKDD '05, pp. 36–43 (ACM, New York, NY, USA, 2005), URL http://doi.acm.org/10.1145/1134271.1134277.

Agarwal, S., W. L. Bennett, C. Johnson and S. Walker, "A model of crowd enabled organization: Theory and methods for understanding the role of twitter in the occupy protests", International Journal of Communication **8**, 0, URL https://ijoc.org/index.php/ijoc/article/view/2068 (2014).

Aktunc, R., I. H. Toroslu, M. Ozer and H. Davulcu, "A dynamic modularity based community detection algorithm for large-scale networks: Dslm", in "2015 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM)", pp. 1177–1183 (2015).

Ausserhofer, J. and A. Maireder, "National politics on twitter", Information, Communication & Society **16**, 3, 291–314 (2013).

Barberá, P., "Birds of the same feather tweet together: Bayesian ideal point estimation using twitter data", Political Analysis **23**, 01, 76–91 (2015).

Barber, P., J. Jost, J. Nagler, J. Tucker and R. Bonneau, "Tweeting from left to right: Is online political communication more than an echo chamber?", Psychological science **26** (2015).

Bennett, W. L. and A. Segerberg, "The logic of connective action", Information, Communication & Society **15**, 5, 739–768, URL https://doi.org/10.1080/1369118X.2012.670661 (2012).

Bhattacharya, S., P. Srinivasan and P. Polgreen, "Engagement with health agencies on twitter", PLOS ONE **9**, 11, 1–12, URL https://doi.org/10.1371/journal.pone.0112235 (2014).

Bimber, B., "Information and political engagement in america: The search for effects of information technology at the individual level", Political Research Quarterly **54**, 1, 53–67, URL https://doi.org/10.1177/106591290105400103 (2001).

Blondel, V. D., J.-L. Guillaume, R. Lambiotte and E. Lefebvre, "Fast unfolding of communities in large networks", Journal of Statistical Mechanics: Theory and Experiment **10**, 10008 (2008).

Bollobás, B., C. Borgs, J. Chayes and O. Riordan, "Directed scale-free graphs", in "Proceedings of the Fourteenth Annual ACM-SIAM Symposium on Discrete Algorithms", SODA '03, pp. 132–139 (Society for Industrial and Applied Mathematics, Philadelphia, PA, USA, 2003), URL http://dl.acm.org/citation.cfm?id=644108.644133.

Boyd, D., S. Golder and G. Lotan, "Tweet, tweet, retweet: Conversational aspects of retweeting on twitter", in "2010 43rd Hawaii International Conference on System Sciences", pp. 1–10 (2010).

Boyd, D., S. Golder and G. Lotan, "Tweet, tweet, retweet: Conversational aspects of retweeting on twitter", in "2010 43rd Hawaii International Conference on System Sciences", pp. 1–10 (2010).

Boyd, S. and L. Vandenberghe, *Convex Optimization* (Cambridge University Press, New York, NY, USA, 2004).

Brady, W. J., J. A. Wills, J. T. Jost, J. A. Tucker and J. J. Van Bavel, "Emotion shapes the diffusion of moralized content in social networks", Proceedings of the National Academy of Sciences **114**, 28, 7313–7318, URL https://www.pnas.org/content/114/28/7313 (2017).

Broniatowski, D. A., A. M. Jamison, S. Qi, L. AlKulaib, T. Chen, A. Benton, S. C. Quinn and M. Dredze, "Weaponized health communication: Twitter bots and russian trolls amplify the vaccine debate", American Journal of Public Health **108**, 10, 1378–1384, URL https://doi.org/10.2105/AJPH.2018.304567, pMID: 30138075 (2018).

Cai, D., X. He, J. Han and T. S. Huang, "Graph regularized nonnegative matrix factorization for data representation", IEEE Trans. Pattern Anal. Mach. Intell. **33**, 8, 1548–1560 (2011).

Carney, N., "All lives matter, but so does race: Black lives matter and the evolving role of social media", Humanity & Society **40**, 2, 180–199, URL https://doi.org/10.1177/0160597616643868 (2016).

Castells, M., *Rise of the Network Society: The Information Age: Economy, Society and Culture* (Blackwell Publishers, Inc., Cambridge, MA, USA, 1996).

Clark, M., *To Tweet Our Own Cause: A Mixed-Methods Analysis of the Online Phenomena Known as Black Twitter: a thesis presented in partial fulfilment of the requirements for the degree of Doctor of Philosophy in School of Media and Journalist at University of North Carolina at Chapel Hill*, Ph.D. thesis, University of North Carolina at Chapel Hill (2014).

Clauset, A., M. E. J. Newman and C. Moore, "Finding community structure in very large networks", Phys. Rev. E **70**, 066111 (2004).

Cleaver, H. M., "The zapatista effect: The internet and the rise of an alternative political fabric", Journal of International Affairs **51**, 2, 621–640, URL http://www.jstor.org/stable/24357524 (1998).

Colleoni, E., A. Rozza and A. Arvidsson, "Echo chamber or public sphere? predicting political orientation and measuring political homophily in twitter using big data", Journal of Communication **64** (2014).

Conover, M., J. Ratkiewicz, M. Francisco, B. Goncalves, F. Menczer and A. Flammini, "Political polarization on twitter", (2011a).

Conover, M., J. Ratkiewicz, M. Francisco, B. Gonçalves, A. Flammini and F. Menczer, "Political polarization on twitter", in "Proc. 5th International AAAI Conference on Weblogs and Social Media (ICWSM)", (2011b), URL `http://www.aaai.org/ocs/index.php/ICWSM/ICWSM11/paper/view/2847`.

Conover, M. D., B. Goncalves, J. Ratkiewicz, A. Flammini and F. Menczer, "Predicting the political alignment of twitter users", in "2011 IEEE Third International Conference on Privacy, Security, Risk and Trust and 2011 IEEE Third International Conference on Social Computing", pp. 192–199 (2011c).

Conover, M. D., B. Goncalves, J. Ratkiewicz, A. Flammini and F. Menczer, "Predicting the political alignment of twitter users", in "2011 IEEE Third International Conference on Privacy, Security, Risk and Trust and 2011 IEEE Third International Conference on Social Computing", pp. 192–199 (2011).

Davis, C. A., O. Varol, E. Ferrara, A. Flammini and F. Menczer, "Botornot: A system to evaluate social bots", in "Proceedings of the 25th International Conference Companion on World Wide Web", WWW '16 Companion, pp. 273–274 (International World Wide Web Conferences Steering Committee, Republic and Canton of Geneva, Switzerland, 2016), URL `https://doi.org/10.1145/2872518.2889302`.

Ding, C., T. Li, W. Peng and H. Park, "Orthogonal nonnegative matrix t-factorizations for clustering", in "Proceedings of the 12th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining", KDD '06, pp. 126–135 (ACM, New York, NY, USA, 2006).

DuBois, T., J. Golbeck and A. Srinivasan, "Predicting trust and distrust in social networks", in "2011 IEEE Third International Conference on Privacy, Security, Risk and Trust and 2011 IEEE Third International Conference on Social Computing", pp. 418–424 (2011).

Epstein, R. and R. E. Robertson, "The search engine manipulation effect (seme) and its possible impact on the outcomes of elections", Proceedings of the National Academy of Sciences **112**, 33, E4512–E4521, URL `https://www.pnas.org/content/112/33/E4512` (2015).

Farwell, J. P., "The media strategy of isis", Survival **56**, 6, 49–55, URL `https://doi.org/10.1080/00396338.2014.985436` (2014).

Ferrara, E., "Disinformation and social bot operations in the run up to the 2017 french presidential election", First Monday **22** (2017).

Fleiss, J. L., "Measuring nominal scale agreement among many raters", Psychological Bulletin **76**, 5, 378–382 (1971).

Fortunato, S., "Community detection in graphs", Physics Reports **486**, 3, 75 – 174 (2010).

Garimella, K., G. D. F. Morales, A. Gionis and M. Mathioudakis, "Quantifying controversy on social media", Trans. Soc. Comput. **1**, 1, 3:1–3:27, URL `http://doi.acm.org/10.1145/3140565` (2018).

Garimella, V. and I. Weber, "A long-term analysis of polarization on twitter", in "Proceedings of the 11th International Conference on Web and Social Media, ICWSM 2017", pp. 528–531 (AAAI press, 2017).

Girvan, M. and M. E. J. Newman, "Community structure in social and biological networks", Proceedings of the National Academy of Sciences **99**, 12, 7821–7826 (2002).

Gu, Q. and J. Zhou, "Local learning regularized nonnegative matrix factorization", in "Proceedings of the 21st International Jont Conference on Artifical Intelligence", IJCAI'09, pp. 1046–1051 (Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, 2009).

Halavais, A. and M. Garrido, "Mapping networks of support for the zapatista movement", Cyberactivism: Online activism in theory and practice. London: Routledge pp. 165–184 (2003).

Hameleers, M., L. Bos and C. H. de Vreese, "they did it: The effects of emotionalized blame attribution in populist communication", Communication Research **44**, 6, 870–900, URL `https://doi.org/10.1177/0093650216644026` (2017).

Hameleers, M. and D. Schmuck, "Its us against them: a comparative experiment on the effects of populist messages communicated via social media", Information, Communication & Society **20**, 9, 1425–1444, URL `https://doi.org/10.1080/1369118X.2017.1328523` (2017).

Hassan, A., A. Abu-Jbara and D. Radev, "Extracting signed social networks from text", in "Workshop Proceedings of TextGraphs-7 on Graph-based Methods for Natural Language Processing", TextGraphs-7 '12, pp. 6–14 (Association for Computational Linguistics, Stroudsburg, PA, USA, 2012).

Heider, F., *The psychology of interpersonal relations* (Wiley, New York, 1958).

Hobolt, S. B., "The brexit vote: a divided nation, a divided continent", Journal of European Public Policy **23**, 9, 1259–1277 (2016).

Hosseinmardi, H., S. A. Mattson, R. Ibn Rafiq, R. Han, Q. Lv and S. Mishra, "Analyzing labeled cyberbullying incidents on the instagram social network", in "Social Informatics", edited by T.-Y. Liu, C. N. Scollon and W. Zhu, pp. 49–66 (Springer International Publishing, Cham, 2015).

Hu, X., J. Tang, H. Gao and H. Liu, "Unsupervised sentiment analysis with emotional signals", in "Proceedings of the 22Nd International Conference on World Wide Web", WWW '13, pp. 607–618 (ACM, New York, NY, USA, 2013a).

Hu, X., L. Tang, J. Tang and H. Liu, "Exploiting social relations for sentiment analysis in microblogging", in "Proceedings of the Sixth ACM International Conference on Web Search and Data Mining", WSDM '13, pp. 537–546 (ACM, New York, NY, USA, 2013b).

Hubert, L. and P. Arabie, "Comparing partitions", Journal of Classification **2**, 1, 193–218 (1985).

Johnson, K. and D. Goldwasser, ""all i know about politics is what i read in twitter": Weakly supervised models for extracting politicians' stances from twitter", in "COLING", (2016).

Kunegis, J., J. Preusse and F. Schwagereit, "What is the added value of negative links in online social networks?", in "Proceedings of the 22Nd International Conference on World Wide Web", WWW '13, pp. 727–736 (ACM, New York, NY, USA, 2013).

Kunegis, J., S. Schmidt, A. Lommatzsch, J. Lerner, E. W. De Luca and S. Albayrak, "Spectral analysis of signed graphs for clustering, prediction and visualization", in "Proceedings of the 2010 SIAM International Conference on Data Mining", pp. 559–570 (SIAM, 2010).

Landis, J. R. and G. G. Koch, "The measurement of observer agreement for categorical data", Biometrics **33**, 1, 159–174 (1977).

Lee, D. D. and H. S. Seung, "Algorithms for non-negative matrix factorization", in "Proceedings of the 13th International Conference on Neural Information Processing Systems", NIPS'00, pp. 535–541 (MIT Press, Cambridge, MA, USA, 2000).

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", WWW '10, pp. 641–650 (ACM, New York, NY, USA, 2010a).

Leskovec, J., D. Huttenlocher and J. Kleinberg, "Signed networks in social media", in "Proceedings of the SIGCHI Conference on Human Factors in Computing Systems", CHI '10, pp. 1361–1370 (ACM, New York, NY, USA, 2010b).

Li, T., Y. Zhang and V. Sindhwani, "A non-negative matrix tri-factorization approach to sentiment classification with lexical prior knowledge", in "Proceedings of the Joint Conference of the 47th Annual Meeting of the ACL and the 4th International Joint Conference on Natural Language Processing of the AFNLP: Volume 1-Volume 1", pp. 244–252 (Association for Computational Linguistics, 2009).

Liben-Nowell, D. and J. Kleinberg, "The link prediction problem for social networks", in "Proceedings of the Twelfth International Conference on Information and Knowledge Management", CIKM '03, pp. 556–559 (ACM, New York, NY, USA, 2003).

Lin, C.-J., "On the convergence of multiplicative update algorithms for nonnegative matrix factorization", Trans. Neur. Netw. **18**, 6, 1589–1596 (2007).

Liu, H., F. Morstatter, J. Tang and R. Zafarani, "The good, the bad, and the ugly: uncovering novel research opportunities in social media mining", International Journal of Data Science and Analytics **1**, 3-4, 137–143 (2016).

Lloyd, S., "Least squares quantization in pcm", IEEE Trans. Inf. Theor. **28**, 2, 129–137 (2006).

Lou, X., A. Flammini and F. Menczer, "Information pollution by social bots", (2019).

M Bond, R., C. J Fariss, J. J Jones, A. D I Kramer, C. Marlow, J. Settle and J. H Fowler, "A 61-million-person experiment in social influence and political mobilization", Nature **489**, 295–8 (2012).

Manikonda, L., G. Beigi, S. Kambhampati and H. Liu, "metoo through the lens of social media", in "SBP-BRiMS", (2018).

Mankad, S. and G. Michailidis, "Structural and functional discovery in dynamic networks with non-negative matrix factorization", Phys. Rev. E **88**, 042812, URL `https://link.aps.org/doi/10.1103/PhysRevE.88.042812` (2013).

Massey Jr, F. J., "The kolmogorov-smirnov test for goodness of fit", Journal of the American statistical Association **46**, 253, 68–78 (1951).

Mitra, T., S. Counts and J. Pennebaker, "Understanding anti-vaccination attitudes in social media", URL `https://www.aaai.org/ocs/index.php/ICWSM/ICWSM16/paper/view/13073` (2016).

Moran, M., *Politics and Governance in the UK* (Palgrave Macmillan, 2015).

Myers, S. A., A. Sharma, P. Gupta and J. Lin, "Information network or social network?: The structure of the twitter follow graph", in "Proceedings of the 23rd International Conference on World Wide Web", WWW '14 Companion, pp. 493–498 (ACM, New York, NY, USA, 2014).

NACOS, B. L., "Politics and the twitter revolution: How tweets influence the relationship between political leaders and the public by john h. pamelee and shannon l. bichard. lanham, md, lexington books, 2011. 256 pp. $75.00.", Political Science Quarterly **128**, 1, 178–179 (2013).

Newman, M. E. J., "Modularity and community structure in networks", Proceedings of the National Academy of Sciences **103**, 23, 8577–8582 (2006).

Nooy, W. D. and J. Kleinnijenhuis, "Polarization in the media during an election campaign: A dynamic network model predicting support and attack among political actors", Political Communication **30**, 1, 117–138 (2013).

O'Callaghan, D., D. Greene, M. Conway, J. Carthy and P. Cunningham, "An analysis of interactions within and between extreme right communities in social media", in "Ubiquitous Social Media Analysis", edited by M. Atzmueller, A. Chin, D. Helic and A. Hotho, pp. 88–107 (Springer Berlin Heidelberg, Berlin, Heidelberg, 2013).

Ozer, M., N. Kim and H. Davulcu, "Community detection in political twitter networks using nonnegative matrix factorization methods", in "Proceedings of the 2016 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining", ASONAM '16, pp. 81–88 (IEEE Press, Piscataway, NJ, USA, 2016), URL http://dl.acm.org/citation.cfm?id=3192424.3192440.

Ozer, M., M. Y. Yildirim and H. Davulcu, "Negative link prediction and its applications in online political networks", in "Proceedings of the 28th ACM Conference on Hypertext and Social Media", HT '17, pp. 125–134 (ACM, New York, NY, USA, 2017).

Ozer, M., M. Y. Yildirim and H. Davulcu, "Implicit negative link detection on online political networks via matrix tri-factorizations", New Review of Hypermedia and Multimedia **24**, 2, 63–87 (2018).

Pei, Y., N. Chakraborty and K. Sycara, "Nonnegative matrix tri-factorization with graph regularization for community detection in social networks", in "Proceedings of the 24th International Conference on Artificial Intelligence", IJCAI'15, pp. 2083–2089 (AAAI Press, 2015).

Ruan, Y., D. Fuhry and S. Parthasarathy, "Efficient community detection in large networks using content and links", in "Proceedings of the 22Nd International Conference on World Wide Web", WWW '13, pp. 1089–1098 (ACM, New York, NY, USA, 2013).

Sachan, M., D. Contractor, T. A. Faruquie and L. V. Subramaniam, "Using content and interactions for discovering communities in social networks", in "Proceedings of the 21st International Conference on World Wide Web", WWW '12, pp. 331–340 (ACM, New York, NY, USA, 2012).

Shang, F., L. Jiao and F. Wang, "Graph dual regularization non-negative matrix factorization for co-clustering", Pattern Recognition **45**, 6, 2237 – 2250, brain Decoding (2012).

Shao, C., G. L. Ciampaglia, O. Varol, K.-C. Yang, A. Flammini and F. Menczer, "The spread of low-credibility content by social bots", in "Nature Communications", (2018).

Stella, M., E. Ferrara and M. De Domenico, "Bots increase exposure to negative and inflammatory content in online social systems", Proceedings of the National Academy of Sciences **115**, 49, 12435–12440, URL https://www.pnas.org/content/115/49/12435 (2018).

Strehl, A. and J. Ghosh, "Cluster ensembles—a knowledge reuse framework for combining multiple partitions", Journal of machine learning research **3**, Dec, 583–617 (2002).

Suh, B., L. Hong, P. Pirolli and E. H. Chi, "Want to be retweeted? large scale analytics on factors impacting retweet in twitter network", in "2010 IEEE Second International Conference on Social Computing", pp. 177–184 (2010).

Tang, J., S. Chang, C. Aggarwal and H. Liu, "Negative link prediction in social media", in "Proceedings of the Eighth ACM International Conference on Web Search and Data Mining", WSDM '15, pp. 87–96 (ACM, New York, NY, USA, 2015).

Tang, J., X. Wang and H. Liu, "Integrating social media data for community detection", in "Modeling and Mining Ubiquitous Social Media", edited by M. Atzmueller, A. Chin, D. Helic and A. Hotho, pp. 1–20 (Springer Berlin Heidelberg, Berlin, Heidelberg, 2012).

Theocharis, Y., W. Lowe, J. W. van Deth and G. Garca-Albacete, "Using twitter to mobilize protest action: online mobilization patterns and action repertoires in the occupy wall street, indignados, and aganaktismenoi movements", Information, Communication & Society **18**, 2, 202–220, URL `https://doi.org/10.1080/1369118X.2014.948035` (2015).

Tufekci, Z., "Big questions for social media big data: Representativeness, validity and other methodological pitfalls", pp. 505–514 (2014).

Tufekci, Z., *Twitter and Tear Gas: The Power and Fragility of Networked Protest* (Yale University Press, New Haven, CT, USA, 2017).

Tufekci, Z. and C. Wilson, "Social Media and the Decision to Participate in Political Protest: Observations From Tahrir Square", Journal of Communication **62**, 2, 363–379, URL `https://doi.org/10.1111/j.1460-2466.2012.01629.x` (2012).

Valenzuela, S., M. Pia and J. Ramrez, "Behavioral Effects of Framing on Social Media Users: How Conflict, Economic, Human Interest, and Morality Frames Drive News Sharing", Journal of Communication **67**, 5, 803–826, URL `https://doi.org/10.1111/jcom.12325` (2017).

Varol, O., E. Ferrara, C. Davis, F. Menczer and A. Flammini, "Online human-bot interactions: Detection, estimation, and characterization", (2017).

Varol, O., E. Ferrara, C. L. Ogan, F. Menczer and A. Flammini, "Evolution of online user behavior during a social upheaval", in "Proceedings of the 2014 ACM Conference on Web Science", WebSci '14, pp. 81–90 (ACM, New York, NY, USA, 2014), URL `http://doi.acm.org/10.1145/2615569.2615699`.

Victor, J. N., A. H. Montgomery and M. Lubell, *The Oxford Handbook of Political Networks* (Oxford University Press, 2016).

Vigil-Hayes, M., M. Duarte, N. D. Parkhurst and E. Belding, "#indigenous: Tracking the connective actions of native american advocates on twitter", in "Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing", CSCW '17, pp. 1387–1399 (ACM, New York, NY, USA, 2017), URL `http://doi.acm.org/10.1145/2998181.2998194`.

Waitoa, J. H., *E-whanaungatanga: the role of social media in Māori political engagement: a thesis presented in partial fulfilment of the requirements for the degree of Master of Philosophy in Development Studies at Te Kunenga ki Pūrehuroa, Massey University, Palmerston North, New Zealand*, Ph.D. thesis, Massey University (2013).

Waltman, L. and N. J. van Eck, "A smart local moving algorithm for large-scale modularity-based community detection", The European Physical Journal B **86**, 11, 471 (2013).

Wang, D., D. Pedreschi, C. Song, F. Giannotti and A.-L. Barabasi, "Human mobility, social ties, and link prediction", in "Proceedings of the 17th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining", KDD '11, pp. 1100–1108 (ACM, New York, NY, USA, 2011).

Weber, I., V. R. K. Garimella and A. Batayneh, "Secular vs. islamist polarization in egypt on twitter", in "2013 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM 2013)", pp. 290–297 (2013).

West, R., H. Paskov, J. Leskovec and C. Potts, "Exploiting social network structure for person-to-person sentiment analysis", Transactions of the Association for Computational Linguistics **2**, 297–310 (2014).

Wong, F. M. F., C. W. Tan, S. Sen and M. Chiang, "Quantifying political leaning from tweets, retweets, and retweeters", IEEE Transactions on Knowledge and Data Engineering **28**, 8, 2158–2172 (2016).

Yang, S.-H., A. J. Smola, B. Long, H. Zha and Y. Chang, "Friend or frenemy?: Predicting signed ties in social networks", in "Proceedings of the 35th International ACM SIGIR Conference on Research and Development in Information Retrieval", SIGIR '12, pp. 555–564 (ACM, New York, NY, USA, 2012).

Yao, Y., H. Tong, G. Yan, F. Xu, X. Zhang, B. K. Szymanski and J. Lu, "Dual-regularized one-class collaborative filtering", in "Proceedings of the 23rd ACM International Conference on Conference on Information and Knowledge Management", CIKM '14, pp. 759–768 (ACM, New York, NY, USA, 2014).

Yu, W., C. C. Aggarwal and W. Wang, "Temporally factorized network modeling for evolutionary network analysis", in "Proceedings of the Tenth ACM International Conference on Web Search and Data Mining", WSDM '17, pp. 455–464 (ACM, New York, NY, USA, 2017), URL `http://doi.acm.org/10.1145/3018661.3018669`.

Zhu, L., D. Guo, J. Yin, G. V. Steeg and A. Galstyan, "Scalable temporal latent space inference for link prediction in dynamic social networks", IEEE Transactions on Knowledge and Data Engineering **28**, 10, 2765–2777 (2016).

APPENDIX A

DERIVATION OF EQUATIONS IN MULTINMF

To follow the conventional theory of constrained optimization we rewrite objective function 2.1 as;

$$\begin{aligned}
\mathbf{J}_{\mathbf{U},\mathbf{H},\mathbf{D},\mathbf{W}} &= Tr((\mathbf{X}_{uw} - \mathbf{U}\mathbf{W}^T)(\mathbf{X}^{uw} - \mathbf{U}\mathbf{W}^T)^T) \\
&+ Tr((\mathbf{X}_{uh} - \mathbf{U}\mathbf{H}^T)(\mathbf{X}_{uh} - \mathbf{U}\mathbf{H}^T)^T) \\
&+ Tr((\mathbf{X}_{ud} - \mathbf{U}\mathbf{D}^T)(\mathbf{X}_{ud} - \mathbf{U}\mathbf{D}^T)^T) \\
&+ \alpha Tr(\mathbf{U}^T L_{\mathbf{C}}\mathbf{U}) + \gamma Tr(\mathbf{H}^T L_{\mathbf{H}_{sim}}\mathbf{H}) \\
&+ \theta Tr(\mathbf{D}^T L_{\mathbf{D}_{sim}}\mathbf{D}) + \beta Tr(\mathbf{W}^T L_{\mathbf{W}_{sim}}\mathbf{W}) \\
\mathbf{J}_{\mathbf{U},\mathbf{H},\mathbf{D},\mathbf{W}} &= Tr(\mathbf{X}_{uw}\mathbf{X}_{uw}^T) - 2Tr(\mathbf{X}_{uw}\mathbf{W}\mathbf{U}^T) \\
&+ Tr(\mathbf{U}\mathbf{W}^T\mathbf{W}\mathbf{U}^T) + Tr(\mathbf{X}_{uh}\mathbf{X}_{uh}^T) \\
&- 2Tr(\mathbf{X}_{uh}\mathbf{H}\mathbf{U}^T) + Tr(\mathbf{U}\mathbf{H}^T\mathbf{H}\mathbf{U}^T) \\
&+ Tr(\mathbf{X}_{ud}\mathbf{X}_{ud}^T) - 2Tr(\mathbf{X}_{ud}\mathbf{D}\mathbf{U}^T) + Tr(\mathbf{U}\mathbf{D}^T\mathbf{D}\mathbf{U}^T) \\
&+ \alpha Tr(\mathbf{U}^T L_{\mathbf{C}}\mathbf{U}) + \gamma Tr(\mathbf{H}^T L_{\mathbf{H}_{sim}}\mathbf{H}) \\
&+ \theta Tr(\mathbf{D}^T L_{\mathbf{D}_{sim}}\mathbf{D}) + \beta Tr(\mathbf{W}^T L_{\mathbf{W}_{sim}}\mathbf{W})
\end{aligned}$$

Let Φ, η, Ω and Ψ be the Lagrangian multipliers for constraints $\mathbf{U}, \mathbf{H}, \mathbf{D}, \mathbf{W} > 0$ respectively. So the Lagrangian function $\mathcal{L}$ becomes;

$$\begin{aligned}
\mathcal{L} &= Tr(\mathbf{X}_{uw}\mathbf{X}_{uw}^T) - 2Tr(\mathbf{X}_{uw}\mathbf{W}\mathbf{U}^T) + Tr(\mathbf{U}\mathbf{W}^T\mathbf{W}\mathbf{U}^T) \\
&+ Tr(\mathbf{X}_{uh}\mathbf{X}_{uh}^T) - 2Tr(\mathbf{X}_{uh}\mathbf{H}\mathbf{U}^T) + Tr(\mathbf{U}\mathbf{H}^T\mathbf{H}\mathbf{U}^T) \\
&+ Tr(\mathbf{X}_{ud}\mathbf{X}_{ud}^T) - 2Tr(\mathbf{X}_{ud}\mathbf{D}\mathbf{U}^T) + Tr(\mathbf{U}\mathbf{D}^T\mathbf{D}\mathbf{U}^T) \\
&+ \alpha Tr(\mathbf{U}^T L_{\mathbf{C}}\mathbf{U}) + \gamma Tr(\mathbf{H}^T L_{\mathbf{H}_{sim}}\mathbf{H}) + \theta Tr(\mathbf{D}^T L_{\mathbf{D}_{sim}}\mathbf{D}) \\
&+ \beta Tr(\mathbf{W}^T L_{\mathbf{W}_{sim}}\mathbf{W}) + Tr(\Phi\mathbf{U}^T) + Tr(\eta\mathbf{H}^T) \\
&+ Tr(\Omega\mathbf{D}^T) + Tr(\Psi\mathbf{W}^T)
\end{aligned}$$

The partial derivatives of Lagrangian function $\mathcal{L}$ with respect to $\mathbf{U}, \mathbf{H}, \mathbf{D}, \mathbf{W}$ are as follows;

$$\begin{aligned}
\frac{\partial \mathcal{L}}{\partial \mathbf{U}} &= -2\mathbf{X}_{uw}\mathbf{W} + 2\mathbf{U}\mathbf{W}^T\mathbf{W} - 2\mathbf{X}_{uh}\mathbf{H} + 2\mathbf{U}\mathbf{H}^T\mathbf{H} - \\
&\quad 2\mathbf{X}_{ud}\mathbf{D} + 2\mathbf{U}\mathbf{D}^T\mathbf{D} + 2\alpha L_{\mathbf{C}}\mathbf{U} + \Phi \\
\frac{\partial \mathcal{L}}{\partial \mathbf{H}} &= -2\mathbf{X}_{uh}^T\mathbf{H} + 2\mathbf{U}\mathbf{H}^T\mathbf{H} + 2\gamma L_{\mathbf{H}_{sim}}\mathbf{H} + \eta \\
\frac{\partial \mathcal{L}}{\partial \mathbf{D}} &= -2\mathbf{X}_{ud}^T\mathbf{H} + 2\mathbf{U}\mathbf{D}^T\mathbf{D} + 2\theta L_{\mathbf{D}_{sim}}\mathbf{D} + \Omega \\
\frac{\partial \mathcal{L}}{\partial \mathbf{W}} &= -2\mathbf{X}_{uw}^T\mathbf{U} + 2\mathbf{W}\mathbf{U}^T\mathbf{U} + 2\beta L_{\mathbf{W}_{sim}}\mathbf{W} + \Psi
\end{aligned}$$

Setting derivatives equal to zero and using KKT complementarity conditions (Boyd and Vandenberghe, 2004) of nonnegativity of matrices $\mathbf{U}, \mathbf{H}, \mathbf{D}, \mathbf{W}$, $\Phi\mathbf{U} = 0$, $\eta\mathbf{H} = 0$,

$\Omega\mathbf{D} = 0$ and $\Psi\mathbf{W} = 0$, we get the update rules given in Equations 2.2, 2.3, 2.4, 2.5.

$$\mathbf{U} \leftarrow \mathbf{U} \odot \sqrt{\frac{\mathbf{X}_{uw}\mathbf{W} + \mathbf{X}_{uh}\mathbf{H} + \mathbf{X}_{ud}\mathbf{D} + \alpha L^{-}_{\mathbf{C}}\mathbf{U}}{\mathbf{U}\mathbf{W}^T\mathbf{W} + \mathbf{U}\mathbf{H}^T\mathbf{H} + \mathbf{U}\mathbf{D}^T\mathbf{D} + \alpha L^{+}_{\mathbf{C}}\mathbf{U}}}$$

$$\mathbf{H} \leftarrow \mathbf{H} \odot \sqrt{\frac{\mathbf{X}^T_{uh}\mathbf{H} + \gamma L^{-}_{\mathbf{H}_{sim}}\mathbf{H}}{\mathbf{H}\mathbf{U}^T\mathbf{U} + \gamma L^{+}_{\mathbf{H}_{sim}}\mathbf{H}}}$$

$$\mathbf{D} \leftarrow \mathbf{D} \odot \sqrt{\frac{\mathbf{X}^T_{ud}\mathbf{D} + \theta L^{-}_{\mathbf{D}_{sim}}\mathbf{D}}{\mathbf{D}\mathbf{U}^T\mathbf{U} + \theta L^{+}_{\mathbf{D}_{sim}}\mathbf{D}}}$$

$$\mathbf{W} \leftarrow \mathbf{W} \odot \sqrt{\frac{\mathbf{X}^T_{uw}\mathbf{U} + \beta L^{-}_{\mathbf{W}_{sim}}\mathbf{W}}{\mathbf{W}\mathbf{U}^T\mathbf{U} + \beta L^{+}_{\mathbf{W}_{sim}}\mathbf{W}}}$$

APPENDIX B

DERIVATION OF EQUATIONS IN SOCLS-FACT

B.1 DERIVATION OF $\mathbf{S}_u$'S UPDATE RULE

By rewriting the optimization formulation as;

$$\begin{aligned}\min_{\mathbf{S_u},\mathbf{H},\mathbf{S}_w} \quad & Tr((\mathbf{X}-\mathbf{S}_u\mathbf{H}\mathbf{S}_w^T)(\mathbf{X}-\mathbf{S}_u\mathbf{H}\mathbf{S}_w^T)^T) \\ & + \alpha Tr((\mathbf{S}_w-\mathbf{S}_{w0})(\mathbf{S}_w-\mathbf{S}_{w0})^T) \\ & + \beta Tr\Big((\mathbf{S}_u-\mathbf{S}_{u0})^T\mathbf{D}_u(\mathbf{S}_u-\mathbf{S}_{u0})\Big) \\ & + \gamma Tr((\mathbf{M}-\mathbf{S}_u\mathbf{S}_u^T)(\mathbf{M}-\mathbf{S}_u\mathbf{S}_u^T)^T) \\ \text{subject to} \quad & \mathbf{S}_u \geq 0, \mathbf{H} \geq 0, \mathbf{S}_w \geq 0\end{aligned}$$

Objective function with respect to $\mathbf{S}_u$ of the rewritten optimization formulation is;

$$\begin{aligned}\min_{\mathbf{S}_u} \quad & -2Tr(\mathbf{X}\mathbf{S}_w\mathbf{H}^T\mathbf{S}_u^T) + Tr(\mathbf{S}_u\mathbf{H}\mathbf{S_w}^T\mathbf{S}_w\mathbf{H}\mathbf{S}_u^T) \\ & + \beta Tr(\mathbf{S}_u^T\mathbf{D}_u\mathbf{S}_u) - 2\beta Tr(\mathbf{S}_u^T\mathbf{D}_u\mathbf{S}_{u0}) - \gamma Tr(\mathbf{M}\mathbf{S}_u\mathbf{S}_u^T) \\ & - \gamma Tr(\mathbf{M}^T\mathbf{S}_u\mathbf{S}_u^T) + \gamma Tr(\mathbf{S}_u\mathbf{S}_u^T\mathbf{S}_u\mathbf{S}_u^T) - Tr(\Gamma\mathbf{S}_u^T)\end{aligned}$$

where Γ is the Lagrange multiplier for the constraint of $\mathbf{S}_u \geq 0$. The derivative of the objective function with respect to $\mathbf{S}_u$ is;

$$\begin{aligned}\frac{\partial \mathcal{L}_{\mathbf{S}_u}}{\partial \mathbf{S}_u} = & -2\mathbf{X}\mathbf{S}_w\mathbf{H}^T + 2\mathbf{S}_u\mathbf{H}\mathbf{S_w}^T\mathbf{S}_w\mathbf{H} + 2\beta\mathbf{D}_u\mathbf{S}_u - 2\beta\mathbf{D}_u\mathbf{S}_{u0} \\ & + \gamma(\mathbf{M}+\mathbf{M}^T)\mathbf{S}_u - 2\gamma\mathbf{S}_u\mathbf{S}_u^T\mathbf{S}_u - \Gamma\end{aligned}$$

By setting the derivative to 0, we get;

$$\begin{aligned}\Gamma = & -2\mathbf{X}\mathbf{S}_w\mathbf{H}^T + 2\mathbf{S}_u\mathbf{H}\mathbf{S_w}^T\mathbf{S}_w\mathbf{H} + 2\beta\mathbf{D}_u\mathbf{S}_u - 2\beta\mathbf{D}_u\mathbf{S}_{u0} \\ & + \gamma(\mathbf{M}+\mathbf{M}^T)\mathbf{S}_u - 2\gamma\mathbf{S}_u\mathbf{S}_u^T\mathbf{S}_u\end{aligned}$$

Having Karush Kuhn Tucker (KKT) complementary condition of the nonnegativity of $\mathbf{S}_u$ as $\Gamma_{ij}(S_u)_{ij} = 0$ gives;

$$\begin{aligned}& \Big(\mathbf{S}_u\mathbf{H}\mathbf{S_w}^T\mathbf{S}_w\mathbf{H} + \beta\mathbf{D}_u\mathbf{S}_u + \gamma(\mathbf{M}+\mathbf{M}^T\Big)_{ij}(\mathbf{S}_u)_{ij} \\ & - \Big(\mathbf{X}\mathbf{S}_w\mathbf{H}^T + \beta\mathbf{D}_u\mathbf{S}_{u0} + \gamma\mathbf{S}_u\mathbf{S}_u^T\mathbf{S}_u\Big)_{ij}(\mathbf{S}_u)_{ij} = 0\end{aligned}$$

which leads to the update rule of $\mathbf{S}_u$;

$$\mathbf{S}_u \leftarrow \mathbf{S}_u \odot \sqrt{\frac{\mathbf{X}\mathbf{S}_w\mathbf{H}^T + \gamma(\mathbf{M}+\mathbf{M}^T)\mathbf{S}_u + \beta\mathbf{D}_u\mathbf{S}_{u0}}{\mathbf{S}_u\mathbf{H}\mathbf{S}_w^T\mathbf{S}_w\mathbf{H}^T + \gamma\mathbf{S}_u\mathbf{S}_u^T\mathbf{S}_u + \beta\mathbf{D}_u\mathbf{S}_u}}$$

B.2 DERIVATION OF $\mathbf{S}_w$'S UPDATE RULE

Objective function with respect to $\mathbf{S}_w$ of the rewritten optimization formulation in Appendix B.1 is;

$$\begin{aligned}\min_{\mathbf{S}_w} \quad & -2Tr(\mathbf{X}\mathbf{S}_w\mathbf{H}^T\mathbf{S}_u^T) + Tr(\mathbf{S}_u\mathbf{H}\mathbf{S_w}^T\mathbf{S}_w\mathbf{H}\mathbf{S}_u^T) \\ & + \alpha Tr(\mathbf{S}_w\mathbf{S}_w^T) - 2\alpha Tr(\mathbf{S}_w\mathbf{S}_{w0}^T) - Tr(\Theta\mathbf{S}_w^T)\end{aligned}$$

where Θ is the Lagrange multiplier for the constraint of $\mathbf{S}_w \geq 0$. The derivative of the objective function with respect to $\mathbf{S}_w$ is;

$$\frac{\partial \mathcal{L}_{\mathbf{S}_w}}{\partial \mathbf{S}_w} = -2\mathbf{X}^T\mathbf{S}_u\mathbf{H} + 2\mathbf{S}_w\mathbf{H}^T\mathbf{S}_u^T\mathbf{S}_u\mathbf{H} + 2\alpha\mathbf{S}_w - 2\alpha\mathbf{S}_{w0} - \Theta$$

By setting the derivative to 0, we get;

$$\Theta = -2\mathbf{X}^T\mathbf{S}_u\mathbf{H} + 2\mathbf{S}_w\mathbf{H}^T\mathbf{S}_u^T\mathbf{S}_u\mathbf{H} + 2\alpha\mathbf{S}_w - 2\alpha\mathbf{S}_{w0}$$

By employing the KKT complentary condition of the nonnegativity of $\mathbf{S}_w$ as $\Theta_{ij}(S_w)_{ij} = 0$ it yields;

$$\Big((\mathbf{S}_w\mathbf{H}^T\mathbf{S}_u^T\mathbf{S}_u\mathbf{H} + \alpha\mathbf{S}_w) - (\mathbf{X}^T\mathbf{S}_u\mathbf{H} + \alpha\mathbf{S}_{w0})\Big)_{ij}(\mathbf{S}_w)_{ij} = 0$$

which leads to the update rule of $\mathbf{S}_w$;

$$\mathbf{S}_w \leftarrow \mathbf{S}_w \odot \sqrt{\frac{\mathbf{X}^T\mathbf{S}_u\mathbf{H} + \alpha\mathbf{S}_{w0}}{\mathbf{S}_w\mathbf{H}^T\mathbf{S}_u^T\mathbf{S}_u\mathbf{H} + \alpha\mathbf{S}_w}}$$

B.3 DERIVATION OF $\mathbf{H}$'S UPDATE RULE

Objective function with respect to $\mathbf{H}$ of the rewritten optimization formulation in Appendix B.1 is;

$$\min_{\mathbf{H}} \quad -2Tr(\mathbf{X}\mathbf{S}_w\mathbf{H}^T\mathbf{S}_u^T) + Tr(\mathbf{S}_u\mathbf{H}\mathbf{S_w}^T\mathbf{S}_w\mathbf{H}\mathbf{S}_u^T) + Tr(\Phi H^T)$$

where Φ is the Lagrange multiplier for the constraint of $\mathbf{H} \geq 0$. The derivative of the objective function with respect to $\mathbf{H}$ is;

$$\frac{\partial \mathcal{L}_{\mathbf{H}}}{\partial \mathbf{H}} = -2\mathbf{S}_u^T\mathbf{X}\mathbf{S}_w + 2\mathbf{S}_u^T\mathbf{S}_u\mathbf{H}\mathbf{S}_w^T\mathbf{S}_w - \Phi$$

By setting the derivative to 0, we get;

$$\Phi = -2\mathbf{S}_u^T\mathbf{X}\mathbf{S}_w + 2\mathbf{S}_u^T\mathbf{S}_u\mathbf{H}\mathbf{S}_w^T\mathbf{S}_w$$

Employing the KKT complentary condition of the nonnegativity of $\mathbf{H}$ as $\Phi_{ij}\mathbf{H}_{ij} = 0$ yields;

$$\Big(\mathbf{S}_u^T\mathbf{S}_u\mathbf{H}\mathbf{S}_w^T\mathbf{S}_w - \mathbf{S}_u^T\mathbf{X}\mathbf{S}_w\Big)_{ij}\mathbf{H}_{ij} = 0$$

leading to the update rule of $\mathbf{H}$;

$$\mathbf{H} \leftarrow \mathbf{H} \odot \sqrt{\frac{\mathbf{S}_u^T\mathbf{X}\mathbf{S}_w}{\mathbf{S}_u^T\mathbf{S}_u\mathbf{H}\mathbf{S}_w^T\mathbf{S}_w}}$$

B.4 DERIVATION OF $\mathbf{S}_{uc}^{(t)}$'S UPDATE RULE

Objective function with respect to $\mathbf{S}_{uc}^{(t)}$ of the rewritten optimization formulation of online framework is;

$$\begin{aligned}\min_{\mathbf{S}_{uc}^{(t)}} \quad & -2Tr(\mathbf{X}^{(t)}\mathbf{S}_w^{(t)}\mathbf{H}^{(t)^T}\mathbf{S}_{uc}^{(t)^T}) + Tr(\mathbf{S}_{uc}^{(t)}\mathbf{H}^{(t)}\mathbf{S_w}^{(t)^T}\mathbf{S}_w^{(t)}\mathbf{H}^{(t)}\mathbf{S}_{uc}^{(t)^T}) \\ & + \beta Tr(\mathbf{S}_{uc}^{(t)^T}\mathbf{D}_u^{(t)}\mathbf{S}_{uc}^{(t)}) - 2\beta Tr(\mathbf{S}_{uc}^{(t)^T}\mathbf{D}_u^{(t)}\mathbf{S}_{u0}^{(t)}) - \gamma Tr(\mathbf{M}^{(t)}\mathbf{S}_{uc}^{(t)}\mathbf{S}_{uc}^{(t)^T}) \\ & - \gamma Tr(\mathbf{M}^{(t)^T}\mathbf{S}_{uc}^{(t)}\mathbf{S}_{uc}^{(t)^T}) + \gamma Tr(\mathbf{S}_{uc}^{(t)}\mathbf{S}_{uc}^{(t)^T}\mathbf{S}_{uc}^{(t)}\mathbf{S}_{uc}^{(t)^T}) \\ & + \tau\sum_{i=1}^{t}(e^{-(t-i)}\big(-2Tr(\mathbf{S}_{uc}^{(t)}\mathbf{S}_{uc}^{(i)^T}) + Tr(\mathbf{S}_{uc}^{(t)}\mathbf{S}_{uc}^{(t)})\big) - Tr(\Gamma\mathbf{S}_{uc}^{(t)^T})\end{aligned}$$

where Γ is the Lagrange multiplier for the constraint of $\mathbf{S}_u \geq 0$. The derivative of the objective function with respect to $\mathbf{S}_u$ is;

$$\begin{aligned}\frac{\partial\mathcal{L}_{\mathbf{S}_{uc}^{(t)}}}{\partial\mathbf{S}_{uc}^{(t)}} = & -2\mathbf{X}_c^{(t)}\mathbf{S}_w^{(t)}\mathbf{H}^T + 2\mathbf{S}_{uc}^{(t)}\mathbf{H}^{(t)}\mathbf{S_w}^T\mathbf{S}_w^{(t)}\mathbf{H}^{(t)} + 2\beta\mathbf{D}_{uc}^{(t)}\mathbf{S}_{uc}^{(t)} - 2\beta\mathbf{D}_{uc}^{(t)}\mathbf{S}_{uc0}^{(t)} \\ & + \gamma(\mathbf{M}_c^{(t)} + \mathbf{M}_c^{(t)^T})\mathbf{S}_{uc}^{(t)} - 2\gamma\mathbf{S}_{uc}^{(t)}\mathbf{S}_{uc}^{(t)^T}\mathbf{S}_{uc}^{(t)} - 2\tau\sum_{i=1}^{t}e^{-(t-i)}\mathbf{S}_{uc}^{(i)} \\ & + 2\tau\sum_{i=1}^{t}e^{-(t-i)}\mathbf{S}_{uc}^{(t)} - \Gamma\end{aligned}$$

By setting the derivative to 0, we get;

$$\begin{aligned}\Gamma = & -2\mathbf{X}_c^{(t)}\mathbf{S}_w^{(t)}\mathbf{H}^{(t)^T} + 2\mathbf{S}_{uc}^{(t)}\mathbf{H}^{(t)}\mathbf{S_w}^{(t)^T}\mathbf{S}_w\mathbf{H}^{(t)} + 2\beta\mathbf{D}_{uc}^{(t)}\mathbf{S}_{uc}^{(t)} - 2\beta\mathbf{D}_{uc}^{(t)}\mathbf{S}_{uc0}^{(t)} \\ & + \gamma(\mathbf{M}_c^{(t)} + \mathbf{M}_c^{(t)^T})\mathbf{S}_{uc} - 2\gamma\mathbf{S}_{uc}^{(t)}\mathbf{S}_{uc}^{(t)^T}\mathbf{S}_{uc} - 2\tau\sum_{i=1}^{t}e^{-(t-i)}\mathbf{S}_{uc}^{(i)} \\ & + 2\tau\sum_{i=1}^{t}e^{-(t-i)}\mathbf{S}_{uc}^{(t)}\end{aligned}$$

Having Karush Kuhn Tucker (KKT) complementary condition of the nonnegativity of $\mathbf{S}_{uc}^{(t)}$ as $\Gamma_{ij}(\mathbf{S}_{uc}^{(t)})_{ij} = 0$ gives;

$$\begin{aligned}&\Big(\mathbf{S}_{uc}^{(t)}\mathbf{H}^{(t)}\mathbf{S_w}^T\mathbf{S}_w\mathbf{H}^{(t)} + \beta\mathbf{D}_{uc}^{(t)}\mathbf{S}_{uc}^{(t)} + \gamma(\mathbf{M}_c^{(t)} + \mathbf{M}_c^{(t)^T}) + \tau\sum_{i=1}^{t}e^{-(t-i)}\mathbf{S}_{uc}^{(t)}\Big)_{ij}(\mathbf{S}_{uc}^{(t)})_{ij} \\ &- \Big(\mathbf{X}_c^{(t)}\mathbf{S}_w\mathbf{H}^{(t)^T} + \beta\mathbf{D}_{uc}^{(t)}\mathbf{S}_{uc0}^{(t)} + \gamma\mathbf{S}_{uc}^{(t)}\mathbf{S}_{uc}^{(t)^T}\mathbf{S}_{uc}^{(t)} + \tau\sum_{i=1}^{t}e^{-(t-i)}\mathbf{S}_{uc}^{(i)}\Big)_{ij}(\mathbf{S}_{uc}^{(t)})_{ij} = 0\end{aligned}$$

which leads to the update rule of $\mathbf{S}_{uc}^{(t)}$;

$$\mathbf{S}_{uc}^{(t)} \leftarrow \mathbf{S}_{uc}^{(t)} \odot \sqrt{\frac{\mathbf{X}_c\mathbf{S}_w\mathbf{H}^T + \gamma(\mathbf{M}_c + \mathbf{M}_c^T)\mathbf{S}_{uc} + \beta\mathbf{D}_{uc}\mathbf{S}_{uc0} + \tau\sum_{i=1}^{t} e^{-(t-i)}\mathbf{S}_{uc}^{(i)}}{\mathbf{S}_{uc}\mathbf{H}\mathbf{S}_w^T\mathbf{S}_w\mathbf{H}^T + \gamma\mathbf{S}_{uc}\mathbf{S}_{uc}^T\mathbf{S}_{uc} + \beta\mathbf{D}_{uc}\mathbf{S}_{uc} + \tau te^{-(t-i)}\mathbf{S}_{uc}^{(t)}}}$$