Negative Link Prediction and Its Applications in Online Political Networks

Mert Ozer, Mehmet Yigit Yildirim, Hasan Davulcu
School of Computing, Informatics, and Decision Systems Engineering
Arizona State University
Tempe, US
{mozer,yigityildirim,hdavulcu}@asu.edu

ABSTRACT
Disagreements, oppositions and negative opinions are indispensable parts of online political debates. In social media, people express their beliefs and attitudes not only on issues but also about each other through both their conversations and platform-specific interactions such as like, share in Facebook and retweet in Twitter. While there are explicit “like” features in these platforms, there is no explicit “dislike” feature. Many network analysis tasks, such as detecting communities and monitoring their dynamics (i.e. polarization patterns) require information about both positive and negative linkages. Hence, predicting negative links between users is an important task and a challenging problem. In this study, we propose an unsupervised framework to predict the negative links between users by utilizing explicit positive interactions and sentiment cues in conversations. We show the effectiveness of the proposed framework on a political Twitter dataset annotated through Amazon MTurk crowdsourcing platform. Our experimental results show that the proposed framework outperforms other well-known methods and proposed baselines. To illustrate the contribution of the predicted negative links, we compare the community detection accuracies using signed and unsigned user networks. Experimental results using predicted negative links show superiority on three political datasets where the camps are known a priori. We also present qualitative evaluations related to the polarization patterns (i.e. rivalries and coalitions) between the detected communities which is only possible in the presence of negative links.

KEYWORDS
Negative Link Prediction; Online Political Networks; Social Media Mining; Sentiment Analysis

1 INTRODUCTION
Beyond any doubt, social media has become a prominent platform for people to express their political stances and opinions for more than a decade. It developed into a medium for politicians and political organizations to interact with the public [22]. While 4th President of the United States, Barack Obama makes an appearance on a Reddit Ask Me Anything. 45th President Donald Trump tweets about how hypocrite he thinks the mainstream media is. While many protesters mobilize their political movements, online social networks more and more start to show the characteristics of public sphere in the online world [1].

Many researchers have extensively studied the nature of online political networks [2], [3], [12], [24]. Most of the existing works utilize 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. In [2], Conover et al. presents how platform-specific positive interactions in Twitter shows a polarized 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. For that reason 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 [5], [16], [28]. Certainly, these online platforms are not the hotspots where people participate to express their political views through.

Therefore, we focus on inferring the negative links between users of online political networks. We aim to predict 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 the 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 [25], [20] 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 phenomena such as social balance and status theory are proven to be helpful in predicting negative links in certain network configurations[17].
In this work, we first propose a nonnegative matrix factorization framework SociLS-Fact that combines signals from sentiment lexicon of words, platform-specific positive interactions and social balance theory to predict negative and positive links in online political networks. We 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. Then, we discuss two applications where predicted 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 predicted negative links in community detection tasks. The second application is proposed to show the informativeness of the predicted negative links related to polarization patterns between political groups. The main contributions of the paper are,

- Proposing an unsupervised model for negative link prediction in social media platforms where platform-specific negative interactions or negative links between users are not present.
- Showing the added value of the negative links in community detection tasks for online political networks.
- Presenting the effectiveness of negative links in describing the rivalries, coalitions between groups and its temporal dynamics qualitatively.

2 RELATED WORK

We survey link prediction and sentiment classification methods proposed for similar line of research 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 [8] that became popular in early 20th century. Link prediction studies standing out as most related to our problem definition are [13], [16], [25], [28]. In [16], Leskovec et al. propose a framework that predicts the sign of links in user networks in social media. 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. In [28], Yang et al. 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, in [13], Kunegis et al. 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 Facebook or Twitter, in which no platform-specific negative interaction is available, is not discussed. In [25], 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 labeled links. This method also relies on experiments conducted only on Slashdot and Epinions datasets. On the other hand, our framework stands out as it is proposed for the online settings that does not provide any platform-specific negative interaction capabilities to its users. Moreover, it is experimented by utilizing political Twitter datasets, which does not include any platform-specific negative interactions.

Second line of research related to our work is sentiment classification in social media. Hu et al., in [11], propose a supervised sentiment classification model which takes advantage of connected text messages having similar sentiment labels. Hu et al., in [10], 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., in [7], 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., in [27], develop a model that combinatorially optimizes 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. Our work differentiates itself, from aforementioned others in the literature by using platform-specific positive interactions, and a sentiment lexicon of words to predict the negative link between users.

3 PROPOSED FRAMEWORK

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

Before going into the details of the framework, the notation that is used throughout the paper can be seen in Table 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 1a and 1b. All textual interaction happening between two users are represented as rows of $X$. $X$ encodes how many times each sentiment word occurs in textual interactions of two users. In Figure 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 $S_{u0}$. Initial sentiment lexicon is embedded in $S_{w0}$. Positive and negative polarities are represented as two latent dimensions in matrix $S_{u0}$ and $S_{w0}$. Which user links should have the same polarity following the social balance theory is governed by matrix $M$. Further details of how matrices $S_{u0}$, $S_{w0}$, $M$ are derived is given later in this section.
As referred in Section 1, 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, we propose the following optimization problem; as in $SU$, first column of $SW$ 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 $\alpha$ 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 $Du$ 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. $M$ encodes the information of pair of links that should have the same polarity if they are forming a triangle with another positive link.

### 3.1 Constructing $SW$  
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 1b, initial sentiment lexicon is embedded in $SW$ such that 1st, 3rd, 4th and 8th words as positive sentiment words and 2nd, 5th, 6th and 7th words as negative sentiment words.

### 3.2 Constructing $SU$ and $Du$
Each row of the initial user link polarity matrix $SU0$ encodes the information of the prior inference of the polarity of user link. First

---

$^1$http://www.cs.uic.edu/~liub/FBS/opinion-lexicon-English.rar
column of the polarity matrix $S_{u0}$ 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, we infer the initial polarity of them as positive and embed it as $[0, 1]$ in the corresponding row of $S_{u0}$ and as $1$ in the corresponding diagonal entry of $D_u$. For the links that connect user pairs having no previous platform-specific positive interaction, we do not infer any initial polarity and represent them as $[0, 0, 1]$ in $S_{u0}$ and as $0$ in the corresponding diagonal entry of $D_u$. To illustrate in Figure 1b, the positive interaction between user A and C is inferred as positive.

The objective function proposed in Section 3 is not convex for all variables of $S_u, S_w, H$. We introduce an alternating optimization solution for our problem similar to [18]. We update each variable $S_u, S_w, H$ iteratively while fixing others to find a local minimum in the solution space. The update rules for each variable is given as;

$$S_u \leftarrow S_u \odot \frac{X S_w H^T + \gamma (M + M^T) S_u + \beta D_u S_{u0}^0}{S_w H S_w^T S_u H^T + \gamma S_u S_u^T S_u + \beta D_u S_u} \quad (1)$$

$$H \leftarrow H \odot \frac{S_u^T X S_w}{S_w H S_w^T S_u^T S_u} \quad (2)$$

$$S_w \leftarrow S_w \odot \frac{X^T S_u H + \alpha S_{w0}}{S_w H^T S_u H + \alpha S_{w0}} \quad (3)$$

Derivation of the update rules is presented in Appendix A,B and C. The proposed algorithm employs an iterative scheme of the above rules until convergence. Each step of the algorithm is shown in Algorithm 1.

### Algorithm 1: Proposed Algorithm for the Optimization Problem

**Input:** $X, S_{u0}, S_{w0}, M$.  
**Output:** $S_u, S_w$.  
1. Initialize $S_u \leftarrow S_{u0}, H \leftarrow I_2, S_w \leftarrow S_{w0}$  
2. **while** not convergent **do**  
5. Update $S_w$ using Equation 3.  

Finally, the polarity of the latent dimension with higher numerical value in the $i$th row of $S_u$ is assigned as the polarity output of the link $i$. To illustrate in Figure 1c and 1d, it can be seen that the value in the first column is greater than the second column for the first and the third rows of $S_u$. 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 $S_u$ 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 1 is $O(mn + m^2 + m + n^2)$. Complexity of the update rule in 2 is $O(mn + m + n)$. Complexity of update rule in 3 is $O(mn + m + n^2)$. Therefore, overall time complexity of the Algorithm 1 complexity is $O(i(m^2 + n^2 + m^2 + mn + m + n))$ 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 [4]. The source code for the whole running pipeline presented in this section can be reached at www.public.asu.edu/~mozer/HT2017Code.tar.gz.
4 EXPERIMENTS

In this section, we present three experiments we design to demonstrate our method’s effectiveness and different use-cases. In the first experiment, we investigate the effectiveness of SocLS-Fact for negative link prediction. In the second experiment, we explore how these predicted negative links contribute to community detection performance. In the third experiment, we qualitatively analyze the added value of predicted negative links in revealing polarization patterns of political party members in social media.

4.1 Dataset

We work with politician Twitter networks from United Kingdom, United States and Canada. Each politician account in the dataset either self declares her political party membership in her user profile or has the abbreviation of the political party in her user name as suffix or prefix. Baseline communities are constructed according to each account’s self-identification of political party memberships.

- **UK Dataset** covers 421 prominent political figures’ twitter accounts from 5 major political parties, namely, Conservative Party (Cons), Labour Party (Lab), Scottish Nationalist Party (SNP), Liberal Democrat Party (LibDem), and United Kingdom Independence Party (UKIP).
- **UK-annotated Dataset** covers 1,074 user pairs sampled from aforementioned UK dataset and polarity of each user interaction is annotated using crowdsourcing. Details is explained in Section 4.1.1.
- **US Dataset** covers 603 prominent political figures’ twitter accounts from Republican (Rep) and Democrat (Dem) Party.
- **Canada Dataset** covers twitter accounts of 192 parliament members from 5 major political parties, namely, Liberal Party of Canada (Lib), Green Party (Green) of Canada, Conservative Party of Canada (Cons), New Democratic Party (NDP), and Bloc Quebecois (BLOC).

Latest 3,200 tweets of each identified account are crawled by using Twitter’s REST API. Users who do not participate in any textual user interaction are removed from the dataset. An overview of the preprocessed data can be seen in Table 2.

<table>
<thead>
<tr>
<th></th>
<th>UK</th>
<th>UK-annotated</th>
<th>US</th>
<th>Canada</th>
</tr>
</thead>
<tbody>
<tr>
<td>Textual interactions</td>
<td>4,217</td>
<td>18,903</td>
<td>41,276</td>
<td>5,001</td>
</tr>
<tr>
<td>Users</td>
<td>400</td>
<td>260</td>
<td>596</td>
<td>136</td>
</tr>
<tr>
<td>Interacting pair of users</td>
<td>3,367</td>
<td>1,074</td>
<td>6,114</td>
<td>1,291</td>
</tr>
<tr>
<td>Positive/negative links</td>
<td>N/A</td>
<td>948/126</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>Baseline communities</td>
<td>5</td>
<td>5</td>
<td>2</td>
<td>5</td>
</tr>
</tbody>
</table>

For the second experiment, we directly make use of UK, US and Canada datasets.

For the third experiment, we utilize UK dataset to create 3 datasets as a representation of political environment in different time frames.

4.1.1 Labeling Through Crowdsourcing.

First, we extracted user pairs that interact with each other at least three times. Then, all the textual interactions (i.e., tweets identified as mentions and replies to each other) of these user pairs were aggregated. While aggregating, we filtered the data to include textual interactions which contains a single user mentioned to avoid the confusion as it is ambiguous which user is addressed in the multiple mentions case.

We requested 3 Mechanical Turk Masters (elite workers demonstrates high accuracy in the previous tasks) who had knowledge of UK politics to rate the polarity of given all textual interactions between two politicians. We have also provided users’ political party affiliations and retweet counts between them to help the labelers assess the polarity of the link better.

After retrieving all the answers from 3 labelers, we assigned the polarity labels using majority voting. Then, we analyzed the labelers inter-rater agreement using Cohen’s Kappa [15] and Fleiss’ Kappa [6]. Two-way inter-rater agreement is nearly perfect according to [15] 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 our problem formulation.

4.2 Negative Link Prediction

Our first experiment aims to demonstrate the negative link prediction performance of SocLS-Fact in political Twitter networks.

To assess the performance of our method, we explain and compare with two existing state-of-the-art matrix factorization approaches along with three other baseline predictors we define as follows:

- **Random**: Motivated by [19], this method predicts 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. Note that the predictor can simply be modeled as $XS_{w0}$, thanks to our initialization scheme we provide for $S_{w0}$ and $X$ in Section 3.
- **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[4]**: This predictor is a simple non-negative matrix tri-factorization method without any regularizers of sentiment lexicon, link prior or social balance.
- **SSMFLK[18]**: 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 ten from -6 to 2 and the best performance is reported.

4.2.1 **Evaluation Metrics.** We use three gold-standard metrics, namely; accuracy, precision and F-measure to evaluate our method. Scores are reported in terms of our method’s prediction performance on the negative links. We do not report recall explicitly as we emphasize quality over quantity; retrieving meaningful negative links is the most important task in this work as suggested for many tasks in [26]. The change in recall can be indirectly observed through F-measure. Although we present the accuracy for reader convenience it may be misleading considering the imbalanced nature of our dataset. Hence, we focus mainly on precision and F-measure throughout the discussion of our results.

4.2.2 **Negative Link Prediction Results.** An overview of the negative link prediction performance of the proposed and baseline methods can be found in Table 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-optimizing the link information with sentiment information in LS-Fact framework results in superior performance compared to both only link and only sentiment predictors. It may be reasonable to think that our encoding strategy for starting conditions contributes to this result marginally.
- Finally, our 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.

4.2.3 **Parameter Analysis.** It is essential that our framework performs effectively under different parameter settings. So, we experiment with various values of \( \alpha \), \( \beta \), and \( \gamma \) and then report the performance in terms of F-measure scores. Best performance was obtained using the parameters \( \alpha = 10^{-2}, \beta = 100, \) and \( \gamma = 10^{-1} \).

Figure 3 demonstrates the effect of prior platform-specific positive interaction parameter \( \alpha \) and sentiment lexicon parameter \( \beta \) when the social balance regularizer \( \gamma \) is fixed at optimal value, \( 10^{-1} \). \( \alpha \) and \( \beta \) are tweaked as powers of ten between -6 to 2. Parameters out of this range gives very low F-measure scores thus excluded.

<table>
<thead>
<tr>
<th></th>
<th>Precision</th>
<th>F-Measure</th>
<th>Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>Random</td>
<td>0.1450</td>
<td>0.2344</td>
<td>0.5317</td>
</tr>
<tr>
<td>SSMFLK[18]</td>
<td>0.3143</td>
<td>0.4490</td>
<td>0.7737</td>
</tr>
<tr>
<td>Only Sentiment</td>
<td>0.4010</td>
<td>0.4892</td>
<td>0.8464</td>
</tr>
<tr>
<td>Only Link</td>
<td>0.6032</td>
<td>0.6726</td>
<td>0.9062</td>
</tr>
<tr>
<td>NMTF[4]</td>
<td>0.6741</td>
<td>0.6973</td>
<td>0.9264</td>
</tr>
<tr>
<td>LS-Fact</td>
<td>0.6976</td>
<td>0.7059</td>
<td>0.9302</td>
</tr>
<tr>
<td>SocLS-Fact</td>
<td>0.7236</td>
<td>0.7149</td>
<td>0.9339</td>
</tr>
</tbody>
</table>

- SocLS-Fact is robust to changes of \( \alpha \) and \( \beta \) as F-measure does not differ more than 0.07 ranging from 0.65 to 0.72.
- Lower values of \( \alpha \) yield the lowest F-measure scores. Performance sharply increases when \( \alpha \) is incremented from \( 10^{-6} \) to \( 10^{-2} \). After \( \alpha = 10^{-2} \), a decrease can be observed at \( \alpha = 0 \); then F-measure stays fairly stable until \( \alpha \) becomes 100.
- Change of \( \beta \) creates rather stable results for any given \( \alpha \). When \( \alpha = 10^{-2} \), there is an increasing pattern between \( \beta \) values from 3 and 1. Finally, a very slight rise can be observed between \( \beta \) values 1 and 100, hence maximal F-measure 0.7149 is observed when \( \beta = 100 \).

Figure 4 shows how social balance regularizer \( \gamma \) affects the performance when the other parameters are fixed at optimal values, \( 10^{-2} \) and 100 respectively, \( \gamma \) is supplied incrementally as powers of ten between -5 to 1. As the chart shows, SocLS-Fact is robust also to changes of \( \gamma \) performing in a F-measure margin of 0.025. F-measure is minimally constant around 0.69 for lower values \( \gamma \). There is a significant performance gain between \( \gamma \) values \( 10^{-3} \) and \( 10^{-1} \). After reaching the optimal score, a decreasing pattern is observed for larger \( \gamma \) values. For our UK dataset, maximal F-measure is obtained when \( \gamma = 10^{-1} \).

4.3 **Added Value of Negative Links**

4.3.1 **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. We feed both unsigned links of the given dataset and predicted signed links by our framework SocLS-Fact separately. We employ UK, Canada and US datasets to evaluate the performance of our method. Parameters for SocLS-Fact are set to be the ones which minimizes the residual error of the objective function.

**Spectral Clustering.** As proposed by [14], we define the laplacian matrix \( \mathcal{L} \) of an adjacency matrix \( A \) of signed network as:

\[
\mathcal{L} = \bar{D} - A
\]  \hspace{1cm} (4)

where

\[
\bar{D}_{ii} = \sum_{j \neq i} |A_{ij}|
\]  \hspace{1cm} (5)

The rest of the clustering framework follows the standard spectral clustering as given in Algorithm 2.
To evaluate the contribution of predicted negative links in community detection tasks, we make use of three well-known clustering quality metrics, namely: purity, adjusted rand index and normalized mutual information.

### Community Detection Results

Table 4 shows the community detection results for UK, US and Canada datasets. Inclusion of the predicted negative links of our framework consistently contributes to the performance of community detection tasks. The ground-truth community counts for UK is 5, Canada is 5 and US is 2 as described in 4.1.

#### Algorithm 2: Spectral Clustering Algorithm for Signed and Unsigned Networks

**Input:** \( \tilde{L} \) (signed) or \( L \) (unsigned).

**Output:** Clusters \( C_1, C_2, ..., C_k \).

1. Find the smallest \( k \) eigenvalues of \( \tilde{L} \) (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.

For experiments having matching \( k \)'s with number of ground-truth communities of datasets, following observations is made. Significant improvement in all three metrics can be observed in the results of UK and Canada datasets. US dataset reveals even more intriguing results: purity increases by \( \%25 \), ARI by \( \%208 \), 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. Most increase by percentage in all three metrics is achieved when \( k = 5 \) in UK and Canada, and \( k = 2 \) in US dataset. This further suggests the informativeness of the predicted negative links in implying the underlying communities.

#### 4.3.2 Group Polarization

To show a use-case of our 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 our 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 visualize and qualitatively analyze the predicted polarities of links among groups and their change over time.

We sample UK dataset and create three dataset spanning different time intervals to represent political climate change in an online setting. First dataset covers the whole timespan which we treat as the overall political climate among members. This dataset constitutes our baseline for detecting divergences from conventional behaviours of political party members in the sampled representative data. The second dataset spans the 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. The 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 SocLS-Fact algorithm and predict the polarity of each user link. Links that connect users are aggregated 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,
Table 4: Contribution of negative links in community detection tasks with different k’s.

<table>
<thead>
<tr>
<th>k</th>
<th>United Kingdom</th>
<th>Canada</th>
<th>United States</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Purity</td>
<td>ARI</td>
<td>NMI</td>
</tr>
<tr>
<td>2</td>
<td>Unsigned Links</td>
<td>0.4818</td>
<td>0.1195</td>
</tr>
<tr>
<td></td>
<td>SocLS-Fact Links</td>
<td>0.4844</td>
<td>0.1213</td>
</tr>
<tr>
<td>3</td>
<td>Unsigned Links</td>
<td>0.8333</td>
<td>0.6572</td>
</tr>
<tr>
<td></td>
<td>SocLS-Fact Links</td>
<td>0.9167</td>
<td>0.8074</td>
</tr>
<tr>
<td>4</td>
<td>Unsigned Links</td>
<td>0.9167</td>
<td>0.8070</td>
</tr>
<tr>
<td></td>
<td>SocLS-Fact Links</td>
<td>0.9427</td>
<td>0.8587</td>
</tr>
</tbody>
</table>

thus omitted. The overview of the resulting polarity among and within groups for each of the three datasets is presented in Figure 5.

Table 5: Popular hashtags in the textual interactions of two samples from UK dataset.

<table>
<thead>
<tr>
<th>Sampled Datasets</th>
<th>Popular Hashtags</th>
</tr>
</thead>
<tbody>
<tr>
<td>General Election</td>
<td>#GE2015, #labourdoorstep, #GE15, #VoteSNP, #Labour, #VoteLabour, #bedroomtax, #NHS, #PMQs, #voteSNP</td>
</tr>
<tr>
<td>Brexit</td>
<td>#StrongerIn, #Brexit, #EUref, #VoteLeave, #labourdoorstep, #Remain, #LabourInForBritain, #BackZac2016, #BothVotesSNP, #EU</td>
</tr>
</tbody>
</table>

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 5. 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 UK politics can be obtained from [21].

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 [9] that took 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 Conservative 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.

4.3.3 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. The findings can be summarized as;

- Comparing Figure 5a and Figure 5b shows the increasing positive link ratio in inner-party links. De Nooy et al., in [23], suggests 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. The behavior can be justified with the existence of the general election.

4.3.4 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 5b) to 2016 (Figure 5c) 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 competitors in the general 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.

5 CONCLUSION
In this paper, we propose a negative link prediction framework that performs well on online political networks in which no platform-specific negative interactions or explicit negative links between users are present. We further show two relevant applications of our framework that may help researchers to better make sense with their political social media data. For future work, we plan to experiment with more annotated datasets from different platforms to evaluate the generalizability of the SocLS-Fact framework.

ACKNOWLEDGMENTS
This work is supported by the Office of Naval Research under Grant No.: N00014-16-1-2015, and N00014-15-1-2722.

Appendices

A DERIVATION OF SU’S UPDATE RULE
By rewriting the optimization formulation as:

\[
\min_{S_u, H, S_w} (X-S_uHS_w^T)(X-S_uHS_w^T)^T + \alpha \text{Tr}(S_u - S_{u0})(S_u - S_{u0})^T + \beta \text{Tr}(S_u S_u^T D_u (S_u - S_{u0})) + \gamma \text{Tr}(M - S_u S_u^T)(M - S_u S_u^T)^T
\]

subject to

\[S_u \geq 0, H \geq 0, S_w \geq 0\]

Objective function with respect to \(S_u\) of the rewritten optimization formulation is:

\[
\frac{\partial L_{S_u}}{\partial S_u} = -2XH^T S_u + \text{Tr}(S_u HS_w^T S_u H + 2\beta D_u S_u - 2\beta D_u S_{u0}) + \gamma (M + M^T)S_u - 2\gamma S_u S_u^T S_u - \Gamma
\]

where \(\Gamma\) is the Lagrange multiplier for the constraint of \(S_u \geq 0\). The derivative of the objective function with respect to \(S_u\) is:

\[
\frac{\partial L_{S_u}}{\partial S_u} = -2XH^T S_u + \text{Tr}(S_u HS_w^T S_u H + 2\beta D_u S_u - 2\beta D_u S_{u0}) + \gamma (M + M^T)S_u - 2\gamma S_u S_u^T S_u - \Gamma
\]

By setting the derivative to 0, we get:

\[
\Gamma = -2XH^T S_u + \text{Tr}(S_u HS_w^T S_u H + 2\beta D_u S_u - 2\beta D_u S_{u0}) + \gamma (M + M^T)S_u - 2\gamma S_u S_u^T S_u
\]

Having Karush Kuhn Tucker (KKT) complementary condition of the nonnegativity of \(S_u\) as \(\Theta_{ij}(S_u)ij = 0\) gives:

\[
\begin{align*}
(S_u HS_w^T S_u H + \gamma (M + M^T)S_u - 2\gamma S_u S_u^T S_u) & \geq 0 \\
- \left(\text{Tr}(S_u HS_w^T S_u H + \gamma (M + M^T)S_u - 2\gamma S_u S_u^T S_u)\right)_{ij} & = 0
\end{align*}
\]

which leads to the update rule of \(S_u\):

\[
S_u \leftarrow S_u \odot \frac{XH^T S_u + \gamma (M + M^T)S_u - 2\beta D_u S_{u0}}{\text{Tr}(S_u HS_w^T S_u H + \gamma (M + M^T)S_u - 2\beta D_u S_{u0})}
\]

B DERIVATION OF SW’S UPDATE RULE
Objective function with respect to \(S_w\) of the rewritten optimization formulation in Appendix A is:

\[
\min_{S_w} \quad -2XH^T S_u + \text{Tr}(S_u HS_w^T S_u H + \gamma (M + M^T)S_u - 2\beta D_u S_{u0})
\]

where \(\Theta\) is the Lagrange multiplier for the constraint of \(S_w \geq 0\). The derivative of the objective function with respect to \(S_w\) is:

\[
\frac{\partial L_{S_w}}{\partial S_w} = -2XH^T S_u + \text{Tr}(S_u HS_w^T S_u H + \gamma (M + M^T)S_u - 2\beta D_u S_{u0}) - \Theta
\]

By setting the derivative to 0, we get:

\[
\Theta = -2XH^T S_u + \text{Tr}(S_u HS_w^T S_u H + \gamma (M + M^T)S_u - 2\beta D_u S_{u0})
\]

By employing the KKT complentary condition of the nonnegativity of \(S_w\) as \(\Theta_{ij}(S_w)ij = 0\) it yields:

\[
\left(\text{Tr}(S_u HS_w^T S_u H + \gamma (M + M^T)S_u - 2\beta D_u S_{u0})\right)_{ij} = 0
\]

which leads to the update rule of \(S_w\):

\[
S_w \leftarrow S_w \odot \frac{XH^T S_u + \gamma (M + M^T)S_u - 2\beta D_u S_{u0}}{\text{Tr}(S_u HS_w^T S_u H + \gamma (M + M^T)S_u - 2\beta D_u S_{u0})}
\]

C DERIVATION OF H’S UPDATE RULE
Objective function with respect to \(H\) of the rewritten optimization formulation in Appendix A is:

\[
\min_{H} \quad -2XH^T S_u + \text{Tr}(S_u HS_w^T S_u H + \gamma (M + M^T)S_u - 2\beta D_u S_{u0})
\]

where \(\Phi\) is the Lagrange multiplier for the constraint of \(H \geq 0\). The derivative of the objective function with respect to \(H\) is:

\[
\frac{\partial L_H}{\partial H} = -2XH^T S_u + \text{Tr}(S_u HS_w^T S_u H + \gamma (M + M^T)S_u - 2\beta D_u S_{u0}) - \Phi
\]

By setting the derivative to 0, we get:

\[
\Phi = -2XH^T S_u + \text{Tr}(S_u HS_w^T S_u H + \gamma (M + M^T)S_u - 2\beta D_u S_{u0})
\]
Employing the KKT complementary condition of the nonnegativity of $H$ as $\Phi_{ij}H_{ij} = 0$ yields:

$$S_i^T S_i H S_i^T S_i - S_i^T X S_i H_{ij} = 0$$

leading to the update rule of $H$:

$$H \leftarrow H \odot \frac{S_i^T X S_i}{S_i^T S_i H S_i^T S_i}$$

REFERENCES

[1] Julian Ausserhofer and Axel Maireder. 2013. NATIONAL POLI-
arXiv: http://dx.doi.org/10.1080/1369118X.2012.756050


citation.cfm?id=2392954.2392956


[18] Tao Li, Yi Zhang, and Vikas Sindhwani. 2009. 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-Vol. 1 Association for Computational Linguistics, 244–252.


