Hierarchical Overlapping Belief Estimation by Structured Matrix Factorization

Chaoqi Yang, Jinyang Li, Ruijie Wang, Shuochao Yao, Huajie Shao, Dongxin Liu,
Shengzhong Liu, Tianshi Wang, Tarek F. Abdelzaher

University of Illinois at Urbana-Champaign, Urbana, Illinois, 61801

{chaoqiy2, jinyang7, ruijiew2, syao9, hshao5, dongxin3, sl29, tianshi3, zaher}@illinois.edu

Abstract—Much work on social media opinion polarization focuses on a *flat categorization* of stances (or orthogonal beliefs) of different communities from media traces. We extend in this work in two important respects. First, we detect not only points of *disagreement* between communities, but also points of *agreement*. In other words, we estimate community beliefs in the presence of *overlap*. Second, in lieu of flat categorization, we consider *hierarchical* belief estimation, where communities might be hierarchically divided. For example, two opposing parties might disagree on core issues, but within a party, despite agreement on fundamentals, disagreement might occur on further details. We call the resulting combined problem a *hierarchical overlapping belief estimation* problem. To solve it, this paper develops a new class of unsupervised Non-negative Matrix Factorization (NMF) algorithms, we call *Belief Structured Matrix Factorization* (BSMF). Our proposed unsupervised algorithm captures both the latent belief intersections and dissimilarities, as well as hierarchical structure. We discuss properties of the algorithm and evaluate it on both synthetic and real-world datasets. In the synthetic dataset, our model reduces error by 40%. In real Twitter traces, it improves accuracy by around 10%. The model also achieves 96.08% self-consistency in a sanity check.

I. Introduction

This paper introduces and solves the novel problem of *unsupervised hierarchical overlapping belief estimation* that uncovers points of agreement and disagreement among communities (as well as their sub-communities), given their social media posts on a polarizing topic. Most prior work clusters sources or beliefs into flat classes or stances [1]. Instead, we focus on scenarios where the underlying social groups disagree on *some* issues but agree on others (i.e., their beliefs *overlap*). Moreover, we consider a (shallow) hierarchical structure, where communities can be further subdivided into subsets with their own agreement and disagreement points.

Our work is motivated, in part, by the increasing polarization on social media [2]. Individuals tend to connect with like-minded sources [3], ultimately producing echo-chambers [3] and filter bubbles [4]. Tools that could automatically extract social beliefs, and distinguish points of agreement and disagreement among them, might help generate future technologies (e.g., less biased search engines) that summarize information for consumption in a manner that gives individuals more control over (and better visibility into) the degree of bias in the information they consume.

A key advantage of the solutions described in this paper is that they are *unsupervised* and mostly *language-agnostic*. By

unsupervised, we mean that our approach does not need prior training, labeling, or remote supervision (in contrast, for example, to deep-learning solutions [5]–[7] that usually require labeled data). By (mostly) *language-agnostic*, we mean that the approach does not use language-specific prior knowledge [2], [8], distant-supervision [9], [10], or prior embedding [5], [6]. Rather, it relies only on tokenization (the ability to separate individual words). While we test the solution only with English text, we conjecture that the unsupervised nature of the work will facilitate its application to other languages (with the exception of those that do not have spaces between words, such as Chinese and Japanese, because we expect spaces as token separators). An advantage of unsupervised techniques is that they do not need to be retrained for new domains, jargon, or hash-tags (as opposed to techniques that rely, for example, on language embedding or neural networks). To the authors' knowledge, ours is the first unsupervised solution to the problem of hierarchical overlapping belief separation.

The work is a significant generalization of approaches for polarization detection (e.g., [3], [11]–[13]), that identifies opposing positions in a debate but do not explicitly search for points of agreement. The unsupervised problem addressed in this paper is also different from unsupervised techniques for topic modeling [14], [15] and polarity detection [11], [16]. Prior solutions to these problems aim to find *orthogonal* topic components [16] or *conflicting* stances [12]. In contrast, we aim to find components that adhere to a given (generic) *overlap* structure. Moreover, unlike solutions for hierarchical topic decomposition [10], [17], we consider not only message content but also user attitudes towards it (e.g., who forwards it), thus allowing for better separation, because posts that share a specific stance are more likely to overlap in the target community (who end up spreading them).

The work is evaluated using both synthetic data as well as real-life data sets, where it is compared to approaches that detect polarity by only considering who posted which claim [11], approaches that separate messages by content or sentiment analysis [18], [19], approaches that identifies different communities in social hypergraphs [20], and approaches that detects user stance by density-based feature clustering [21]. The results of this comparison show that our algorithm significantly outperforms the state of the art. An ablation study further illustrates the impact of different design decisions on accomplishing this improvement.

The rest of the paper is organized as follows. Section II formulates the problem and summarizes the solution approach. Section III proposes our new belief structured matrix factorization model, and analyzes some model properties. Section IV presents an experimental evaluation. We review the related literature on belief mining and matrix factorization in Section V. The paper concludes with key observations and a statement on future directions in Section VI. Note that for the purpose of readability, we have simplified model derivation in the main text, and readers could find technical details in the Appendix.

II. PROBLEM FORMULATION

Consider an observed data set of posts collected from a social medium, such as Twitter, where each post is associated with a *source* and with semantic content, called a *claim*. Let $\mathcal{S}$ be the set of sources in our data set, and $\mathcal{C}$ be the set of claims made by those sources. While, in this paper, a claim is the content of a tweet (or retweet), the analytical treatment does not depend on this interpretation. Let matrix $\mathbf{X}$, of dimension $|\mathcal{S}| \times |\mathcal{C}|$, be a matrix of binary entries denoting who claimed what. If source S_i posted claim C_j, then $x_{ij} = 1$, otherwise $x_{ij} = 0$. In general, a claim can be made by multiple sources.

A. Problem Statement

We assume that the set of sources, $\mathcal{S}$, is divided into a small number, K, of latent social groups, denoted by the subsets $\mathcal{G}_1$, $\mathcal{G}_2$, ..., $\mathcal{G}_K$, that form a tree. In this tree, group $\mathcal{G}_i$ is a child (i.e., a subgroup) of $\mathcal{G}_k$ if $\mathcal{G}_i \subset \mathcal{G}_k$. Children of the same parent are disjoint groups. Members of group $\mathcal{G}_k$ that do not belong to any of its children are denoted by the residual set $\mathcal{G}_k^-$. Within each group, $\mathcal{G}_k$, $1 \le k \le K$, individuals have *shared beliefs* expressed by a set of claims. A shared belief of a group is a belief espoused by *all members* of the group. By definition, therefore, a child group inherits the shared beliefs of its parent. The child group may have additional shared beliefs within its group (not shared by the remaining members of the parent). Thus, we define the *incremental belief set*, $\mathcal{B}_k$, of group $\mathcal{G}_k$ to be the beliefs held by group $\mathcal{G}_k$, beyond what's inherited from its parent. The overall belief set of group $\mathcal{G}_k$ is thus the union of incremental beliefs of its ancestors and itself. The problem addressed in this paper is to simultaneously (i) uncover all latent groups, $\mathcal{G}_k$, in the tree, and (ii) uncover their incremental belief sets, $\mathcal{B}_k$.
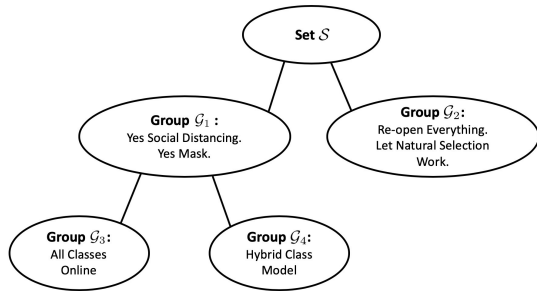

Fig. 1. A Notional Example of Hierarchical Overlapping Beliefs.

Figure 1 illustrates an example, inspired by the first wave of the COVID-19 pandemic in 2020. In this figure, a hypothetical community is divided on whether to maintain social distancing ($\mathcal{G}_1$) or reopen everything and let natural selection take place ($\mathcal{G}_2$). Furthermore, while $\mathcal{G}_1$ agree on social distancing, they disagree on some implementation detail, such as whether classes should be entirely online ($\mathcal{G}_3$) or hybrid ($\mathcal{G}_4$).

B. Solution Approach

We develop a novel non-negative matrix factorization algorithm that decomposes the "who said what" matrix, $\mathbf{X}$, into (i) a matrix, $\mathbf{U}$, that maps sources to latent groups, (ii) a matrix, $\mathbf{B}$, that maps latent groups to latent incremental belief sets (called the *belief structure matrix*), and (iii) a matrix, $\mathbf{M}$, that maps latent incremental belief sets to claims. Importantly, since groups and belief sets are *latent*, the belief structure matrix, $\mathbf{B}$, in essence, specifies the *latent structure* of the solution space that the algorithm needs to populate with specific sources and claims, thereby guiding factorization.

III. STRUCTURED MATRIX FACTORIZATION

The novel aspect of our structured matrix factorization is the existence of the structure matrix, $\mathbf{B}$, that represents the relation between the latent groups, $\mathcal{G}_k$ (that we wish to discover), and their incremental belief sets, $\mathcal{B}_k$ (that we wish to discover as well). An element, b_{ij} of the matrix, $\mathbf{B}$, is 1 if group $\mathcal{G}_i$ adopts the belief set $\mathcal{B}_j$. Otherwise, it is zero. In a typical (non-overlapping) clustering or matrix factorization framework, there is an one-to-one correspondence between groups and belief sets, reducing $\mathbf{B}$ to an identity matrix. Structured matrix factorization extends that structure to an arbitrary relation. Matrix $\mathbf{B}$ can be thought of as a template relating latent groups (to be discovered) and belief sets (to be identified). It is a way to describe the structure that one wants the factorization algorithm to populate with appropriate group members and claims. While it might seem confusing to presuppose that one knows the latent structure, $\mathbf{B}$, before the groups and belief sets in question are populated, below we show why this problem formulation is very useful.

A. An Illustrative Example

Consider a conflict involving two opposing groups, say, a minority group $\mathcal{G}_1$ and a majority $\mathcal{G}_2$. Their incremental belief sets are denoted by $\mathcal{B}_1$ and $\mathcal{B}_2$, respectively. The two groups disagree on everything. Thus, sets $\mathcal{B}_1$ and $\mathcal{B}_2$ do not overlap. An agent wants to weaken the majority and conjectures that the majority group might disagree on something internally. Thus, they postulate that group $\mathcal{G}_2$ is predominantly made of subgroups $\mathcal{G}_{2a}$ and $\mathcal{G}_{2b}$. While both subgroups agree on the shared beliefs, $\mathcal{B}_2$, each subgroup has its own incremental belief sets, $\mathcal{B}_{2a}$ and $\mathcal{B}_{2b}$, respectively. The structure matrix in Figure 2 represents the belief structure postulated above.

$$
\begin{array}{c}
\begin{array}{cccc} \mathcal{B}_1 & \mathcal{B}_2 & \mathcal{B}_{2a} & \mathcal{B}_{2b} \end{array} \\
\begin{array}{c} \mathcal{G}_1 \\ \mathcal{G}_2^- \\ \mathcal{G}_{2a} \\ \mathcal{G}_{2b} \end{array}
\left[\begin{array}{cccc}
1 & 0 & 0 & 0 \\
0 & 1 & 0 & 0 \\
0 & 1 & 1 & 0 \\
0 & 1 & 0 & 1
\end{array} \right]
\end{array}
$$

Fig. 2. The Belief Structure Matrix, $\mathbf{B}$.

For example, the second column indicates that the belief set $\mathcal{B}_2$ is shared by all members of group $\mathcal{G}_2$ (hence, there is a "1" in rows of subgroups $\mathcal{G}_{2a}$, $\mathcal{G}_{2b}$, and the residual $\mathcal{G}_2^-$), but that belief sets $\mathcal{B}_{2a}$ and $\mathcal{B}_{2b}$ are unique to subgroups $\mathcal{G}_{2a}$ and $\mathcal{G}_{2b}$, respectively. It is also evident that the beliefs espoused by different groups overlap. For example, from the third and fourth rows, we see that groups $\mathcal{G}_{2a}$ and $\mathcal{G}_{2b}$ overlap in the set of beliefs, $\mathcal{B}_2$.

An interesting question might be: so, which sources belong to which group/subgroup? What are the incremental belief sets $\mathcal{B}_{2a}$ and $\mathcal{B}_{2b}$ that divide group $\mathcal{G}_2$ (i.e., are shared only by the individual respective subgroups)? What are the shared beliefs $\mathcal{B}_2$ that unite it? What are the beliefs, $\mathcal{B}_1$, of group $\mathcal{G}_1$? These are the questions answered by our structured matrix factorization algorithm whose input is (only) matrix, $\mathbf{X}$, and matrix, $\mathbf{B}$ (Fig. 2).

B. Mathematical Formulation

To formulate the hierarchical overlapping belief estimation problem, we introduce the notion of claim *endorsement*. A source is said to *endorse* a claim if the source finds the claim agreeable with their belief. Endorsement, in this paper, represents a state of *belief*, not a physical act. A source might find a claim agreeable with their belief, even if the source did not explicitly post it. Let the probability that source S_i endorses claim C_j be denoted by $Pr(S_iC_j)$. We further denote the proposition $S_i \in \mathcal{G}_p$ by S_i^p, and the proposition $C_j \in \mathcal{B}_q$ by C_j^q. Thus, $Pr(S_i^p)$ denotes the probability that source $S_i \in \mathcal{G}_p$. Similarly, $Pr(C_j^q)$ denotes the probability that claim $C_j \in \mathcal{B}_q$. Following the law of total probability:

$$Pr(S_iC_j) = \sum_{p,q} Pr(S_iC_j|S_i^pC_j^q)Pr(S_i^pC_j^q)$$
$$= \sum_{p,q} Pr(S_iC_j|S_i^pC_j^q)Pr(S_i^p)Pr(C_j^q). \quad (1)$$

By definition of the belief structure matrix, $\mathbf{B}$, we say that $Pr(S_iC_j|S_i^pC_j^q) = 1$ if $b_{pq} = 1$ in the belief structure matrix. Otherwise, $Pr(S_iC_j|S_i^pC_j^q) = 0$.

Let $u_{ip} = Pr(S_i^p)$ and $m_{jq} = Pr(C_j^q)$. Let u_i and m_j be the corresponding vectors, with elements ranging over values of p and q respectively. Thus, we get: $Pr(S_iC_j) = u_i^\top \mathbf{B} m_j$. Let the matrix $\mathbf{X}^G$ be the matrix of probabilities, $Pr(S_iC_j)$, such that element $x_{ij}^G = Pr(S_iC_j)$. Thus:

$$\mathbf{X}^G = \mathbf{UBM}^\top \quad (2)$$

where $\mathbf{U}$ is a matrix whose elements are u_{ip} and $\mathbf{M}$ is a matrix whose elements are m_{jq}. Factorizing $\mathbf{X}^G$, given $\mathbf{B}$, would directly yield $\mathbf{U}$ and $\mathbf{M}$, whose elements are the probabilities we want: elements of matrix $\mathbf{U}$ yield the probabilities that a given source S_i belongs to a given group $\mathcal{G}_p$, whereas elements of matrix $\mathbf{M}$ yield the probabilities that a claim C_j belongs to a belief $\mathcal{B}_q$. Each source is then assigned to a group and each claim to a belief set, based on the highest probability entry in their row of matrix $\mathbf{U}$ and $\mathbf{M}$, respectively. In practice, $\mathbf{B}$ could be customized up to certain tree depth to meet different granularity of belief estimation.

C. Estimating $\mathbf{X}^G$

Unfortunately, we do not really have matrix $\mathbf{X}^G$ to do the above factorization. Instead, we have the *observed* source-claim matrix $\mathbf{X}$ that is merely a sampling of what the sources endorse. (It is a sampling because a silent source might be in agreement with a claim even if they did not say so.) Using $\mathbf{X}$ directly is suboptimal because it is very sparse. It is desired to estimate that a source endorses a claim even if the source remains silent. We do so on two steps.

Message interpolation (the M-module): First, while source S_i might have not posted a specific claim, C_j, it may have posted similar ones. Thus, even when $x_{ij} = 0$, in reality the source may endorse claim C_j. Let matrix $\mathbf{D}$, of dimension $|\mathcal{C}| \times |\mathcal{C}|$, denote the *similarity* structure between claims, by some similarity measure. Hence, each element, d_{ij}, denotes how similar claims C_i and C_j are. To keep our approach simple and language-agnostic, as we detail later, we use radial basis functions (RBF) that consider only lexical (i.e., bag of words) overlap between the respective claims. We then update x_{ij} if source S_i posted claims similar to C_j, according to matrix, $\mathbf{D}$. This is called *message interpolation* (the M-module). The exact expression is derived in the Appendix. The output of this module is a matrix $\mathbf{X}^M$.

Social graph convolution (S-module): To further improve our estimation of matrix $\mathbf{X}^G$, we assume that sources generally hold similar beliefs to those in their immediate social neighborhood. Thus, we perform a smoothing of matrix $\mathbf{X}^M$ by replacing each cell, x_{ij}, by a weighted average of itself and the entries pertaining to neighbors of its source, S_i, in the social graph. Let matrix $\mathbf{A}$, of dimension $|\mathcal{S}| \times |\mathcal{S}|$, denote the social graph. Each entry, a_{ij}, denotes the influence of user S_i on user S_j. $\mathbf{A}$ is thus the adjacency matrix of a social network graph. In this paper, we construct $\mathbf{A}$ by calculating the frequency of each source S_i retweeting posts of source S_j. We call it the *retweet graph*. The update of entries of $\mathbf{X}^M$ by smoothing over entries of neighboring sources is called *social graph convolution* (S-module). The mathematical details are also in the Appendix. It results in an improved estimate, $\mathbf{X}^{MS}$.

We can now take, $\mathbf{X}^G \approx \mathbf{X}^{MS}$, and decompose it as presented in Section III-B, Equation (2), with L_1 and L_2 regularizations to enforce sparsity and prevent overfitting. We specify the iterative estimation rule in the Appendix.

IV. Experiments

In the section, we evaluate *Belief Structured Matrix Factorization (BSMF)* using both a synthetic dataset and real-world Twitter datasets with hierarchical overlapping beliefs. Our model is compared to six baselines and three model variants. We elaborate the experimental settings and results below. The code base and additional analytical and experimental results could be found in github [1].

---

[1] https://github.com/ycq091044/narrative-detection

A. Synthetic Data

1) Dataset Construction: In order to understand the behavior of our algorithm in a simplified and controllable setting, we build a synthetic dataset where two groups of users are created, a minority, $\mathcal{G}_1$ (100 users) of belief set $\mathcal{B}_1$, and a majority, $\mathcal{G}_2$ (300 users) of belief set $\mathcal{B}_2$. The majority includes two subgroups, $\mathcal{G}_{2a}$ and $\mathcal{G}_{2b}$ (100 users each) of incremental belief sets $\mathcal{B}_{2a}$ and $\mathcal{B}_{2b}$, respectively. Essentially, the groups follow the hierarchical structure illustrated in Fig. 2. For each of the group, we built disjoint claim corpus, called c_1, c_2, c_{2a} and c_{2b}, respectively. Users were simulated who chose claims to emit based on their corpus or their parent's corpus (we randomly generate 20 claims for each user). Thus, for example, users in group $\mathcal{G}_{2a}$ could emit claims generated from c_{2a} or from the parent corpus c_2, but users in group $\mathcal{G}_1$ only emit claims from corpus c_1. In sum, 400 users and 8,000 claims were created. In this experiment, we do not impose social relations. Instead, we use the identity matrix for the adjacency $\mathbf{A}$.

2) Method Comparison: For this experiment, the factorization uses the belief structure matrix in Fig. 2. Two simpler variants are introduced: (i) the first variant substitutes $\mathbf{B}$ with an identity matrix, and takes a standard NMF formulation $\mathbf{X}^G = \mathbf{U}\mathbf{M}^\top$; (ii) the second variant substitutes $\mathbf{B}$ with an learnable matrix $\tilde{\mathbf{B}}$, which takes a standard non-negative matrix tri-factorization (NMTF) form, $\mathbf{X}^G = \mathbf{U}\tilde{\mathbf{B}}\mathbf{M}^\top$. Obviously, NMTF offers more freedom. However, the need to learn parameters of matrix $\tilde{\mathbf{B}}$ can cause overfitting. We use the same regularization settings for NMF, NMTF and our BSMF to make sure the comparison is fair. Empirically, after $150 \sim 200$ iterations, all three methods converge. The predicted belief set label for each claim is given by the index of the maximum value in this final representation from matrix, $\mathbf{M}$.[2]

3) Results of 200 Rounds: We run each model for 200 times and BSMF consistently outperforms NMF and NMTF. The average accuracy for BSMF, NMF and NMTF are 97.34%, 93.78%, 95.54%, respectively. As might be expected, specifying matrix, $\mathbf{B}$, guides subsequent factorization to a better result compared to both NMF and NMTF.

B. Real-world Twitter Trace: Eurovision2016

To illustrate versatility of our structured matrix factorization approach, we start with a simplified example where we break sources into two subgroups of overlapping beliefs. The example allows easy comparison with the prior state of the art on polarization that explicitly addresses this common case.

1) Dataset: We use the Eurovision2016 dataset, borrowed from [11]. Eurovision2016 contains tweets about the Ukrainian singer, Jamala, who won the Eurovision (annual) song contest in 2016. Her success was a surprise to many as the expected winner had been from Russia according to pre-competition polls. The song was on a controversial political topic, about story of deportation of Crimean Tatars by Soviet Union forces in the 1940s. Tweets related to Jamala were collected within five days of the contest. Basic statistics are reported in Table I.

---

[2]In practice, we permute the labels and pick the best matching as a result, since our approach does clustering not classification.

TABLE I
BASIC STATISTICS FOR TWO TWITTER DATASETS

Dataset	# sources	# claims	# all tweets	# retweets
Eurovision2016	3,514	5,812	9,868	6,001
Global Warming	14,752	7,030	16,418	9,341

As pre-processed in [11], the most popular 1,000 claims were manually annotated. They were separated into 600 pro-Jamala, 239 anti-Jamala, and 161 neutral claims.

In the context of the dataset, the entire set of sources is regarded as a big group, $\mathcal{G}_1$, with belief set, $\mathcal{B}_1$, agreed among all users. Upon the disagreements, the group is further divided into three disjoint groups, group $\mathcal{G}_{1a}$ (with inherited belief $\mathcal{B}_1$ and incremental belief $\mathcal{B}_{1a}$), group $\mathcal{G}_{1b}$ (with inherited belief $\mathcal{B}_1$ and incremental belief $\mathcal{B}_{1b}$), and the residual group $\mathcal{G}_1^-$ (with belief $\mathcal{B}_1$). In this case, the belief structure matrix is:

$$\mathbf{B} = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ 1 & 0 & 1 \end{bmatrix} \tag{3}$$

where rows correspond to groups $\mathcal{G}_1^-$, $\mathcal{G}_{1a}$, and $\mathcal{G}_{1b}$, respectively, whereas columns correspond to belief sets, $\mathcal{B}_1$, incremental sets, $\mathcal{B}_{1a}$ and $\mathcal{B}_{1b}$, respectively.

2) Baselines: We carefully select seven baseline methods that encompass different perspectives on belief separation:

- *Random*: A trivial baseline that annotates posts randomly, giving equal probability to each label.
- *DBSCAN* [21]: A density-based clustering technique that extracts user-level features and performs DBSCAN for unsupervised user stance detection in Twitter. In this paper, we use DBSCAN and then map the user stance to claim stances with majority voting, as a baseline.
- *Sentiment140* [18] and SANN [19]: Content-aware solutions based on language or sentiment models. In the implementation, each of the claims is a query through Sentiment140 API, which responds with a polarity score. SANN similarly outputs three polarity labels.
- *H-NCut* [20]: The method views the bipartite structure of the source-claim network as a hypergraph, where claims are nodes and sources are hyperedges. The problem is thus seen as a hypergraph community detection problem, where community nodes represent posts. We implement *H-NCut*, a hypergraph normalized cut algorithm.
- *Polarization* [11]: A baseline that uses an NMF-based solution for social network belief extraction to separate biased and neutral claims.
- *NMTF*: A baseline with a learnable mixture matrix. We compare our model with it to demonstrate that pure learning without a prior is not enough to unveil the true belief overlap structure in real-world applications.

Different variants of BSMF are also evaluated to verify the effectiveness of message similarity interpolation (the M-module) and social graph convolution (the S-module). BSMF incorporates both modules. Models without the M-module or the S-module are named BSMF_{M-} and BSMF_{S-}, respectively, while BSMF_{MS-} denotes the model without either module.

TABLE II
MACRO- AND WEIGHTED- METRICS COMPARISON (EUROVISION2016)

| Models | Eurovision 2016 | | | | | |
| | Macro | | | Weighted | | |
	Prec.	Recall	F1	Prec.	Recall	F1
Random	0.340	0.339	0.319	0.436	0.352	0.368
Sentiment140	0.425	0.377	0.339	0.413	0.384	0.354
SANN	0.410	0.393	0.334	0.453	0.370	0.365
H-NCut	0.462	0.481	0.413	0.518	0.487	0.474
DBSCAN	0.513	0.514	0.488	0.556	0.528	0.520
Polarization	0.531	0.497	0.443	0.588	0.491	0.463
NMTF	0.531	0.527	0.471	0.585	0.526	0.491
BSMF_{MS-}	0.565	0.516	0.460	0.601	0.549	0.492
BSMF_{M-}	0.656	0.623	0.578	0.741	0.582	0.605
BSMF_{S-}	0.586	0.592	0.574	0.664	0.660	0.645
BSMF	0.678	0.643	0.596	0.768	0.602	0.623

*Prec. stands for precision.

TABLE III
TOP 3 TWEETS FROM SEPARATED BELIEFS (EUROVISION2016)

(Incremental) Beliefs	Sample Tweets
$\mathcal{B}_1$: Agreement	BBC News - Eurovision Song Contest: Ukraine's Jamala wins competition https://t.co/kL8SYOPOYL
	Parents of "#Ukrainian" Susana #Jamaludinova - @Jamala are #Russian citizens and prosper in the Russian #Crimea
	A politically charged ballad by the Ukrainian singer Jamala won the @Eurovision Song Contest http://nyti.ms/1qlmmNs
$\mathcal{B}_{1a}$: Pro-Jamala	@jamala congratulations! FORZA UKRAINE!
	@DKAMBinUkraine: Congratulations @jamala and #Ukraine!!! You deserved all the 12 points from #Denmark and the victory, #workingforDK
	@NickyByrne: Well done to Ukraine and @jamala
$\mathcal{B}_{1b}$: Anti-Jamala	jamala The song was political and agaisnt The song contest rules shows NATO had influence on jury decision.
	@VictoriaLIVE @BBCNews @jamala Before voting we rated it worst song in the contest. Not changed my mind.
	@JohnDelacour So @jamala has violated TWO ESC rules - the song is not new, and it includes political content. Result MUST be annulled

3) Evaluation Metrics: We evaluate claim separation, since only claim labels are accessible. We use the Python *scikit-learn* package to help with the evaluation. Multiple metrics are employed. *Macro-evaluation* simply calculates the mean of the metrics, giving equal weight to each class. It is used to highlight model performance of infrequent classes. *Weighted-metrics* account for class imbalance by computing the average of metrics in which each class score is weighted by its presence in the true data sample. Standard *precision, recall and f1-score* are considered in both scenarios. Note that weighted averaging may produce an f1 that is not between precision and recall.

4) Result of Eurovision2016: The comparison results are shown in Table II. It is not surprising that all baselines beat Random. Overall, matrix factorization methods work well for this problem. With both the *M-module* and *S-module*, our BSMF algorithm ranks the top in terms of all metrics. Among other baselines, Sentiment140 and SANN work poorly for this task, because (i) they use background language models that are pre-trained on another corpus; and (ii) they do not user dependency information, which matters in real-world data. H-NCut and DBSCAN also yield weak performance, sicne they do not consider the underlying overlapping and hierarchical structure. We also notice that the NMF-based algorithm actually outperforms NMTF. The reason might be that, for real-world data, the latent belief structure is harder to capture, and NMTF could be trapped in poor local optima.

Table III shows the top 3 tweets from each belief set ($\mathcal{B}_1$, $\mathcal{B}_{1a}$, $\mathcal{B}_{1b}$) estimated by our model. Note that, due to an update of the Twitter API, the crawled text field is truncated to 140 characters. Our algorithm runs on the text within that range only. For human readability and interpretability, we manually fill in the rest of the tweet, showing the additional text in yellow (the same for Table IV). Note that, the labels shown in the first column, called *Beliefs*, are inserted manually after the fact (and not by our algorithm). The algorithm merely does the separation/clustering.

C. A More Complicated Scenario: Global Warming

We consider a more complicated scenario in this section, with a majority group, $\mathcal{G}_1$, and a minority group, $\mathcal{G}_2$, of beliefs

$\mathcal{B}_1$ and $\mathcal{B}_2$ that do not overlap. Since more data is expected on $\mathcal{G}_1$ (by definition of majority), we opt to further divide it into subgroups $\mathcal{G}_{1a}$ and $\mathcal{G}_{1b}$, who (besides believing in $\mathcal{B}_1$) hold the incremental belief sets $\mathcal{B}_{1a}$ and $\mathcal{B}_{1b}$, respectively. The corresponding belief structure is reflected by the belief matrix:

$$\mathbf{B} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \tag{4}$$

where rows represent groups $\mathcal{G}_1^-$, $\mathcal{G}_{1a}$, $\mathcal{G}_{1b}$, and $\mathcal{G}_2$ and columns represent the belief sets $\mathcal{B}_1$, $\mathcal{B}_{1a}$, $\mathcal{B}_{1b}$, and $\mathcal{B}_2$. The matrix is not an identity matrix because beliefs overlap (e.g, groups $\mathcal{G}_1^-$, $\mathcal{G}_{1a}$, and $\mathcal{G}_{1b}$ share belief $\mathcal{B}_1$). It also features a hierarchical subdivision of $\mathcal{G}_1$, into $\mathcal{G}_1^-$, $\mathcal{G}_{1a}$, and $\mathcal{G}_{1b}$.

1) Datasets: We apply this belief structure to an unlabeled dataset, Global Warming, which is crawled in real time with the *Apollo Social Sensing Toolkit*[3]. This dataset is about a twitter discussion of global warming in the wake of Australia wildfires that ravaged the continent, in September 2019, where at least 17.9 million acres of forest have burned in the fire. Our goal is to identify and separate posts according to the above abstract belief structure.

Table IV shows the algorithm's assignment of claims to belief groups (only the top 3 claims are shown for space limitations). The first column shows the abstract belief categories $\mathcal{B}_1$, $\mathcal{B}_{1a}$, $\mathcal{B}_{1b}$, and $\mathcal{B}_2$. While the algorithm allocates posts to categories based on the structure of matrix **B**, for readability, we manually inspect posts assigned to each category in the matrix, and give that category a human-readable name, also shown in the first column. For each belief category, the table also shows the top ranked statements. The table reveals that sources in our data set are polarized between a group, $\mathcal{G}_1$, that believes in global warming (offering statements that urge a serious response) and a group, $\mathcal{G}_2$, that does not (offering statements that oppose the thesis of man-made global warming). Within group, $\mathcal{G}_1$ (apart from $\mathcal{G}_1^-$), there are two subgroups, $\mathcal{G}_{1a}$ and $\mathcal{G}_{1b}$. The former blames the fossil fuel industry, whereas the latter is concerned with rising sea levels.

[3]http://apollo2.cs.illinois.edu/

TABLE IV
TOP 3 TWEETS FROM SEPARATED BELIEFS (GLOBAL WARMING)

(Incremental) Beliefs	Sample Tweets
$\mathcal{B}_1$: **Global Warming / urge response**	Australia's top scientists urge government to do more on global warming https://t.co/NclFqGKXE1
	Australia's most prestigious scientific organisation has added to growing pressure on Prime Minister Scott Morrison over climate change policy, calling on the government to "take stronger action" in response to the bushfire crisis
	"Have we now reached the point where at last our response to global warming will be driven by engineering and economics rather than ideology and idiocy?" #auspol
$\mathcal{B}_{1a}$: **Global Warming / fossil fuel**	As long as the ALP keep accepting 'donations' (bribes) from the climate change deniers the fossil fuel industry, who spent millions and millions spreading lies about global warming, they have zero credibility when they talk about phasing out fossil fuel
	To mitigate the effects of climate change, we must do away with fossil fuel burning as they are the major contributors of global warming.
	Turnbull: "The world must, and I believe will, stop burning coal if we are to avoid the worst consequences of global warming. And the sooner the better." Malcom Turnbull, The Guardian 12 January #ScottyfromMarketing
$\mathcal{B}_{1b}$: **Global Warming / sea level**	That time when The Australian misrepresented @JohnChurchOcean to say sea level rise wasn't linked to global warming. After I wrote about it, they pulled the story.
	Brave global warming researchers are studying sea level rise in the Maldives this morning. https://t.co/aqGtgXAj2t
	CO2 is a magical gas which causes Lake Michigan water levels to both rise and fall https://t.co/8FrC1Cx2Rm
$\mathcal{B}_2$: **No Global Warming**	CLIMATE'S FATAL FLAW : 'Greenhouse Gases Simply Do Not Absorb Enough Heat To Cause Global Warming' – "New data and improved understanding now show that there is a fatal flaw in greenhouse-warming theory."
	"Three new research studies confirm that geothermal heat flow, not man-made global warming, is the dominant cause of West Antarctic Ice Sheet (WAIS) melting," writes geologist James Edward Kamis.
	Left Media talks about.. Climate Change, Global Warming, But... Jihad is reason for recent Forest Fires in Australia !

While we do not claim to have reached conclusions on global warming, the table shows how structured matrix factorization can fit data sets automatically to useful belief structures, thereby offering visibility into what individuals are concerned with, what actions they agree on, and what they disagree about.

2) Quantitative Measurements: Next, we do a sanity check by measuring user grouping consistency. Specifically, we first identify the belief sets (by claim separation) and then assign belief labels to users by having a user inherit the assigned belief set label for each claim they made. The inherited labels are inconsistent if they belong to *different* groups according to matrix, **B**. For example, if the same user has been assigned belief labels $\mathcal{B}_1$ and $\mathcal{B}_{1a}$, then the labeling is coherent because both represent beliefs of $\mathcal{G}_1$ (remember that a group inherits the beliefs of its parent). If another user is labeled with both $\mathcal{B}_1$ and $\mathcal{B}_2$, then it is apparently wrong, since belief sets $\mathcal{B}_1$ and $\mathcal{B}_2$ belong to different groups.

The percentage of coherently labeled users was 96.08%. Note that, we do not conduct comparison in this dataset, since most baselines do not uncover hierarchical group/belief structures, whereas those that do generally break up the hierarchy differently (e.g., by hierarchical topic, not hierarchical stance) thus not offering an apples to apples comparison. In future work, we shall explore more comparison options.

V. RELATED WORK

The problem of belief mining has been a subject of study for decades [2]. Solutions include such diverse approaches as detecting social polarization [11], [12], opinion extraction [5], [6], [9], stance detection [21] and sentiment analysis [8], [22], to name a few.

Pioneers, like Leman *at el.* [23] and Bishan *at el.* [24], had used Bayesian models and other basic classifiers to separate social beliefs. On the linguistic side, many efforts extracted user opinions based on domain-specific phrase chunks [25], and temporal expressions [26]. With the help of pre-trained embedding, like Glove [6] or word2vec [7], deep neural networks (e.g., variants of RNN [5], [6]) emerged as powerful tools (usually with attention modules [7]) for understanding the polarity or sentiment of user messages. In contrast to the above supervised or language-specific solutions, we consider the challenge of developing an *unsupervised* approach.

In the domain of unsupervised algorithms, our problem is different from the related problems of unsupervised topic detection [14], [15], sentiment analysis [8], [22], and unsupervised community detection [27]. Topic modeling assigns posts to polarities or topic mixtures [28], independently of actions of users on this content. Hence, they often miss content nuances or context that helps better interpret the stance of the source. Community detection [29], on the other hand, groups nodes by their general interactions, maximizing intra-class links while minimizing inter-class links [27], [29], or partitioning (hyper)graphs [20]. While different communities may adopt different beliefs, this formulation fails to distinguish regions of belief overlap from regions of disagreement.

The above suggests that belief mining must consider both sources (and forwarding patterns) and content. Prior solutions used a source-claim bipartite graph, and determined disjoint polarities by iterative factorization [11], [23]. Our work is novel by postulating a more generic and realistic view: social beliefs could overlap and can be hierarchically structured. In this context, we developed a new matrix factorization scheme that considers (i) the source-claim graph [11]; (ii) message word similarity [10] and (iii) user social dependency [30] in a new class of non-negative matrix factorization techniques to solve the hierarchical overlapping belief estimation problem.

The work also contributes to non-negative matrix factorization. NMF was first introduced by Paatero and Tapper [31] as the concept of positive matrix factorization and was popularized by the work of Lee and Seung [32], who gave an interesting interpretation based on parts-based representation. Since then, NMF has been widely used in various applications, such as pattern recognition [33], signal processing [34].

Two main issues of NMF have been intensively discussed during the development of its theoretical properties: solution uniqueness [35], [36] and decomposition sparsity [37], [38]. By only considering the standard formula $\mathbf{X} \approx \mathbf{UM}^\top$, it is usually not difficult to find a non-negative and non-singular matrix $\mathbf{V}$, such that $\mathbf{UV}$ and $\mathbf{V}^{-1}\mathbf{M}^\top$ could also be a valid solution. Uniqueness will be achieved if $\mathbf{U}$ and $\mathbf{M}$ are sufficiently sparse or if additional constraints are included [39]. Special constraints have been proposed in [40], [41] to improve the sparseness of the final representation.

Non-negative matrix tri-factorization (NMTF) is an extension of conventional NMF (i.e., $\mathbf{X} \approx \mathbf{UBM}^\top$ [42]). Unconstrained NMTF is theoretically identical to unconstrained NMF. However, when constrained, NMTF possesses more degrees of freedom [39]. NMF on a manifold emerges when

the data lies in a nonlinear low-dimensional submanifold [43]. Manifold Regularized Discriminative NMF [44], [45] were proposed with special constraints to preserve local invariance, so as to reflect the multilateral characteristics.

In this work, instead of including constraints to impose structural properties, we adopt a novel belief structured matrix factorization by introducing the mixture matrix $\mathbf{B}$. The structure of $\mathbf{B}$ can well reflect the latent belief structure and thus narrows the search space to a good enough region.

VI. Conclusion

In this paper, we proposed a new class of NMF, where the structure of parts is already known (or assumed to follow some generic form). Specifically, we introduced a belief structure matrix $\mathbf{B}$, and proposed a novel *Belief Structured Matrix Factorization* algorithm, called BSMF, to separate overlapping, hierarchically structured beliefs from large volumes of user-generated messages. The factorization could be briefly formulated as $\mathbf{X}^{MS} \approx \mathbf{UBM}^\top$, where $\mathbf{B}$ is known. The model is tested on a synthetic dataset. Further evaluations are conducted on real-world Twitter events. The results show that our algorithm consistently outperform baselines by a great margin. We believe this paper could seed a research direction on automatically separating data sets according to arbitrary belief structures to enable more in-depth understanding of social groups, attitudes, and narratives on social media.

References

[1] D. Küçük and F. Can, "Stance detection: A survey," *ACM Computing Surveys (CSUR)*, vol. 53, no. 1, pp. 1–37, 2020.

[2] B. Liu, "Sentiment analysis and opinion mining," *Synthesis lectures on human language technologies*, 2012.

[3] A. Bessi, F. Zollo, M. Del Vicario, M. Puliga, A. Scala, G. Caldarelli, B. Uzzi, and W. Quattrociocchi, "Users polarization on facebook and youtube," *PloS one*, 2016.

[4] E. Bakshy, S. Messing, and L. A. Adamic, "Exposure to ideologically diverse news and opinion on facebook," *Science*, 2015.

[5] O. Irsoy and C. Cardie, "Opinion mining with deep recurrent neural networks," in *EMNLP*, 2014.

[6] P. Liu, S. Joty, and H. Meng, "Fine-grained opinion mining with recurrent neural networks and word embeddings," in *EMNLP*, 2015.

[7] W. Wang, S. J. Pan, D. Dahlmeier, and X. Xiao, "Coupled multi-layer attentions for co-extraction of aspect and opinion terms," in *AAAI*, 2017.

[8] Y. Hu, F. Wang, and S. Kamb, "Listening to the crowd: automated analysis of events via aggregated twitter sentiment," in *IJCAI*, 2013.

[9] M. Srivatsa, S. Lee, and T. Abdelzaher, "Mining diverse opinions," in *MILCOM 2012-2012 IEEE Military Communications Conference*, 2012.

[10] T. Weninger, Y. Bisk, and J. Han, "Document-topic hierarchies from document graphs," in *CIKM*, 2012.

[11] M. T. Al Amin, C. Aggarwal, S. Yao, T. Abdelzaher, and L. Kaplan, "Unveiling polarization in social networks: A matrix factorization approach," in *INFOCOM*, 2017.

[12] M. D. Conover, J. Ratkiewicz, M. Francisco, B. Gonçalves, F. Menczer, and A. Flammini, "Political polarization on twitter," in *ICWSM*, 2011.

[13] G. Demartini, S. Siersdorfer, S. Chelaru, and W. Nejdl, "Analyzing political trends in the blogosphere," in *Fifth International AAAI Conference on Weblogs and Social Media*, 2011.

[14] R. Ibrahim, A. Elbagoury, M. S. Kamel, and F. Karray, "Tools and approaches for topic detection from twitter streams: survey," *Knowledge and Information Systems*, 2018.

[15] I. Litou and V. Kalogeraki, "Pythia: A system for online topic discovery of social media posts," in *ICDCS*, 2017.

[16] K. Cheng, J. Li, J. Tang, and H. Liu, "Unsupervised sentiment analysis with signed social networks," in *AAAI*, 2017.

[17] C. Zhang, F. Tao, X. Chen, J. Shen, M. Jiang, B. Sadler, M. Vanni, and J. Han, "Taxogen: Unsupervised topic taxonomy construction by adaptive term embedding and clustering," in *SIGKDD*, 2018.

[18] *Sentiment140 - A Twitter Sentiment Analysis Tool*, 2016, Jul. [Online]. Available: http://www.sentiment140.com

[19] Y. Zhang, G. Lai, M. Zhang, Y. Zhang, Y. Liu, and S. Ma, "Explicit factor models for explainable recommendation based on phrase-level sentiment analysis," in *SIGIR*, 2014.

[20] D. Zhou, J. Huang, and B. Schölkopf, "Learning with hypergraphs: Clustering, classification, and embedding," in *NIPS*, 2007.

[21] K. Darwish, P. Stefanov, M. Aupetit, and P. Nakov, "Unsupervised user stance detection on twitter," in *Proceedings of the International AAAI Conference on Web and Social Media*, vol. 14, 2020, pp. 141–152.

[22] X. Hu, J. Tang, H. Gao, and H. Liu, "Unsupervised sentiment analysis with emotional signals," in *WWW*, 2013.

[23] L. Akoglu, "Quantifying political polarity based on bipartite opinion networks," in *ICWSM*, 2014.

[24] B. Yang and C. Cardie, "Extracting opinion expressions with semi-markov conditional random fields," in *EMNLP*, 2012.

[25] C. Wu, F. Wu, S. Wu, Z. Yuan, and Y. Huang, "A hybrid unsupervised method for aspect term and opinion target extraction," *Knowledge-Based Systems*, 2018.

[26] A. Schulz, B. Schmidt, and T. Strufe, "Small-scale incident detection based on microposts," in *Proceedings of the 26th ACM Conference on Hypertext & Social Media*, 2015.

[27] S. Fortunato and D. Hric, "Community detection in networks: A user guide," *Physics reports*, vol. 659, 2016.

[28] J. Han, H. Cheng, D. Xin, and X. Yan, "Frequent pattern mining: current status and future directions," *TKDD*, 2007.

[29] J. Yang and J. Leskovec, "Overlapping community detection at scale: a nonnegative matrix factorization approach," in *WSDM*, 2013.

[30] H. Zhang, T. N. Dinh, and M. T. Thai, "Maximizing the spread of positive influence in online social networks," in *ICDCS*, 2013.

[31] P. Paatero and U. Tapper, "Positive matrix factorization: A non-negative factor model with optimal utilization of error estimates of data values," *Environmetrics*, 1994.

[32] D. D. Lee and H. S. Seung, "Algorithms for non-negative matrix factorization," in *NIPS*, 2001.

[33] A. Cichocki, R. Zdunek, A. H. Phan, and S.-i. Amari, *Nonnegative matrix and tensor factorizations: applications to exploratory multi-way data analysis and blind source separation*. John Wiley & Sons, 2009.

[34] I. Buciu, "Non-negative matrix factorization, a new tool for feature extraction: theory and applications," *International Journal of Computers, Communications and Control*, 2008.

[35] D. Donoho and V. Stodden, "When does non-negative matrix factorization give a correct decomposition into parts?" in *NIPS*, 2004.

[36] B. Klingenberg, J. Curry, and A. Dougherty, "Non-negative matrix factorization: Ill-posedness and a geometric algorithm," *Pattern Recognition*, 2009.

[37] S. Moussaoui, D. Brie, and J. Idier, "Non-negative source separation: range of admissible solutions and conditions for the uniqueness of the solution," in *ICASSP*, 2005.

[38] H. Laurberg, M. G. Christensen, M. D. Plumbley, L. K. Hansen, and S. H. Jensen, "Theorems on positive data: On the uniqueness of nmf," *Computational intelligence and neuroscience*, 2008.

[39] Y.-X. Wang and Y.-J. Zhang, "Nonnegative matrix factorization: A comprehensive review," *TKDE*, 2012.

[40] N. Mohammadiha and A. Leijon, "Nonnegative matrix factorization using projected gradient algorithms with sparseness constraints," in *ISSPIT*, 2009.

[41] P. O. Hoyer, "Non-negative matrix factorization with sparseness constraints," *JMLR*, 2004.

[42] J. Yoo and S. Choi, "Orthogonal nonnegative matrix tri-factorization for co-clustering: Multiplicative updates on stiefel manifolds," *Information processing & management*, 2010.

[43] D. Cai, X. He, X. Wu, and J. Han, "Non-negative matrix factorization on manifold," in *ICDM*, 2008.

[44] N. Guan, D. Tao, Z. Luo, and B. Yuan, "Manifold regularized discriminative nonnegative matrix factorization with fast gradient descent," *IEEE Transactions on Image Processing*, 2011.

[45] S. Ana, J. Yoob, and S. Choi, "Manifold-respecting discriminant non-negative matrix factorization [j]," *Pattern Recognition Letters*, 2011.

APPENDIX

A. Message Interpolation (M-module)

The approximation, $\mathbf{X}^M \approx \mathbf{X}^G$, is developed as follows. First, if a source $\mathcal{S}_i$ posted, retweeted, or liked claim $\mathcal{C}_j$ in our data set (i.e., $x_{ij} = 1$ in matrix $\mathbf{X}$), then we know that the source endorses that claim (i.e., $x_{ij}^M = 1$ in matrix $\mathbf{X}^M$). The question is, what to do when $x_{ij} = 0$? In other words, we need to estimate the likelihood that the source endorses a claim, when no explicit observations of such endorsement were made. We do so by considering the claim similarity matrix $\mathbf{D}$. If source $\mathcal{S}_i$ was observed to endorse claims $\mathcal{C}_k$ similar to $\mathcal{C}_j$, then it will likely endorse $\mathcal{C}_j$ with a probability that depends on the degree of similarity between $\mathcal{C}_j$ and $\mathcal{C}_k$. Thus, when $x_{ij} = 0$, we can estimate x_{ij}^M by weighted sum interpolation:

$$x_{ij}^M = \sum_{k:\ x_{ik} \neq 0} d_{kj} \cdot x_{ik} \tag{5}$$

To compute matrix $\mathbf{D}$, in this work, we first compute a bag-of-words (BOW) vector w_j for each claim j. We then normalize it using vector L_2 norm, $\bar{w}_j = w_j / \|w_j\|_2$. We select non-zero entries x_{ij} in each row i of $\mathbf{X}$ as *medoids* $\{\bar{w}_j \mid x_{ij} \neq 0\}$. We assume that claims close to any of the *medoids* could also be endorsed by $\mathcal{S}_i$ as well. Based on that, we use:

$$d_{kj} = \phi(\|\bar{w}_j - \bar{w}_k\|) \tag{6}$$

in Equation (5). A Gaussian radial basis function (RBF) is used for $\phi(r) = e^{-(\epsilon r)^2}$. If the resulting value of x_{ij}^M is less than 0.2, we regard that it is far from all of the *medoids* and set it back to 0. In the experiment, ϵ is set 0.5 for sythetic datset and 0.05 for both Eurovision 2016 and Global Warming.

B. Social Graph Convolution (S-module)

To further improve our estimation of matrix $\mathbf{X}^G$, denoted by $\mathbf{X}^{MS}$, we consider the social dependency matrix $\mathbf{A}$.

The fundamental insight we would like to leverage is that users that are close in the social graph, $\mathbf{A}$, are likely to endorse the same claims, even if an explicit endorsement was not observed in the data set. Thus, we consider the social dependency matrix $\mathbf{A}$ (user-user retweet frequency) and compute the a degree matrix $\mathbf{F}$ by summing each row of $\mathbf{A}$ and the random walk normalized adjacency is denoted as $\tilde{\mathbf{A}}_{rw} = \mathbf{F}^{-1}\mathbf{A}$. We define our propagation operator based on $\mathbf{A}_{rw}$ with self-loop re-normalization, $\bar{\mathbf{A}}_{rw} \leftarrow \frac{1}{2}\tilde{\mathbf{F}}_{rw}^{-1}(\tilde{\mathbf{A}}_{rw} + I)$. Thus, the new source-claim network is given by,

$$\mathbf{X}^{MS} = \bar{\mathbf{A}}_{rw} X^M, \tag{7}$$

where each row of $\bar{\mathbf{A}}_{rw}$ adds up to 1. The effect of the propagation operator is to convolve the information from 1-hop neighbors, while preserving half of the information from itself. Note that, we deem dependency beyond 1-hop too weak to capture, so we do not consider $\mathbf{A}^n$, where $n > 1$. From a macroscopic perspective, this social graph convolution recovers some of the possible source-claim connections and also enforces the smoothness of matrix $\mathbf{X}^{MS}$.

C. Overall Loss and Optimization

Given a belief mixture matrix, $\mathbf{B}$, we now factorize $\mathbf{X}^{MS}$ to estimate matrices $\mathbf{U}$ and $\mathbf{M}$ that decide the belief regions associated with sources and claims, respectively. (e.g., the estimated belief for claim $\mathcal{C}_j$ is given by the index of maximum entry in the j_{th} row of $\mathbf{M}$).

Regularization. To avoid model overfitting, we include widely used L_2 regularization. Also, we enforce the sparsity of $\mathbf{U}$ and $\mathbf{M}$ by introducing L_1 norm. The overall objective function becomes (defined by the Forbenious-norm),

$$J = \|\mathbf{X}^{MS} - \mathbf{UBM}^\top\|_F^2 + \lambda_1\|\mathbf{U}\|_F^2 + \lambda_1\|\mathbf{M}\|_F^2$$
$$+ \lambda_2\|\mathbf{U}\|_1 + \lambda_2\|\mathbf{M}\|_1. \tag{8}$$

We rewrite J using matrix trace function $tr(\cdot)$,

$$J = tr(\mathbf{X}^{MS^\top}\mathbf{X}^{MS}) - 2tr(\mathbf{X}^{MS^\top}\mathbf{UBM}^\top)$$
$$+ tr(\mathbf{UBM}^\top\mathbf{MB}^\top\mathbf{U}^\top) + \lambda_1 tr(\mathbf{U}^\top\mathbf{U}) + \lambda_1 tr(\mathbf{M}^\top\mathbf{M})$$
$$+ \lambda_2\|\mathbf{U}\|_1 + \lambda_2\|\mathbf{M}\|_1. \tag{9}$$

We minimize J by gradient descent. Since only the non-negative region is of our interests, derivatives of L_1 norm is differentiable in this setting. By referring to gradient of traces of product with constant matrix A, $\nabla_{\mathbf{X}} tr(\mathbf{AX}) = \mathbf{A}^\top$ and $\nabla_{\mathbf{X}} tr(\mathbf{XAX}^\top) = \mathbf{X}(\mathbf{A} + \mathbf{A}^\top)$, the partial derivative of J w.r.t. $\mathbf{U}$ and $\mathbf{M}$ are calculated as,

$$\nabla_{\mathbf{U}} = -2\mathbf{X}^{MS}\mathbf{MB}^\top + 2\mathbf{UBM}^\top\mathbf{MB}^\top + 2\lambda_1\mathbf{U} + \lambda_2\mathbf{1},$$
$$\nabla_{\mathbf{M}} = -2\mathbf{X}^{MS^\top}\mathbf{UB} + 2\mathbf{MB}^\top\mathbf{U}^\top\mathbf{UB} + 2\lambda_1\mathbf{M} + \lambda_2\mathbf{1}.$$

The gradient matrix $\nabla_{\mathbf{U}}$ is of dimension $|\mathcal{S}| \times K$, and $\nabla_{\mathbf{M}}$ is of dimension $|\mathcal{C}| \times K$. Estimation step begins by updating $\mathbf{U} \leftarrow \mathbf{U} - \eta\nabla_{\mathbf{U}}$ and $\mathbf{M} \leftarrow \mathbf{M} - \eta\nabla_{\mathbf{M}}$, and η is the step size. Negative values might appear in the learning process, which are physically meaningless in this problem. Thus, we impose the non-negative constraints for $\mathbf{U}$ and $\mathbf{M}$ during the update. A *ReLU*-like strategy is utilized: when any entry of $\mathbf{U}$ or $\mathbf{M}$ becomes negative, it is set to be ξ. In the experiment, we set $\xi = 10^{-8}$, $\lambda_1 = \lambda_2 = 10^{-3}$. Note that the initial entry values of $\mathbf{U}$ and $\mathbf{M}$ are randomized uniformly from $(0, 1)$.

D. Complexity

After *M-module* and *S-module*, non-zero entries in the estimated matrix, $\mathbf{X}^{MS} \approx \mathbf{X}^G$, are still far fewer than $|\mathcal{S}| \times |\mathcal{C}|$. We consider to use sparse matrix multiplications and avoid dense intermediate matrices, which makes the computation efficient. Note that, K (number of beliefs) is picked according to the dataset, and it typically satisfies $K \ll min(|\mathcal{S}|, |\mathcal{C}|)$. During the estimation, we generalize standard NMF multiplicative update rules [32] for our tri-factorization,

$$\eta_{\mathbf{U}} = \frac{1}{2}\frac{\mathbf{U}}{\mathbf{UBM}^\top\mathbf{MB}^\top}, \quad \eta_{\mathbf{M}} = \frac{1}{2}\frac{\mathbf{M}}{\mathbf{MB}^\top\mathbf{U}^\top\mathbf{UB}}. \tag{10}$$

Algorithmically, updating $\mathbf{U}$ and $\mathbf{M}$ takes $O(K|\mathcal{S}||\mathcal{C}|)$ per iteration. We could also take the advantages of the structure of $\mathbf{B}$, and reduce the complexity to $O(|\mathcal{S}||\mathcal{C}|)$, identical to typical NMF. The number of iterations before the empirical convergence is usually no more than 200 for random initialization, and thus we claim that our model is scalable and efficient.