# CSL$^+$: Scalable Collective Subjective Logic under Multidimensional Uncertainty

ADIL ALIM, University at Albany - SUNY
JIN-HEE CHO, Virginia Tech
FENG CHEN, The University of Texas at Dallas

Using unreliable information sources generating conflicting evidence may lead to a large uncertainty which significantly hurts decision making process. Recently many approaches have been taken to integrate conflicting data from multiple sources and/or fusing conflicting opinions from different entities. To explicitly deal with uncertainty, a belief model called *Subjective Logic* (SL), as a variant of Dumpster-Shafer Theory (DST), has been proposed to represent subjective opinions and to merge multiple opinions by offering a rich volume of fusing operators, which have been used to solve many opinion inference problems in trust networks. However, the operators of SL are known to be lack of scalability in inferring unknown opinions from large network data as a result of the sequential procedures of merging multiple opinions. In addition, SL does not consider deriving opinions in the presence of conflicting evidence. In this work, we propose a hybrid inference method that combines SL and Probabilistic Soft Logic (PSL), namely Collective Subjective Plus, CSL$^+$, which is resistible to highly conflicting evidence or a lack of evidence. PSL can reason a belief in a collective manner to deal with large-scale network data, allowing high scalability based on relationships between opinions. However, PSL does not consider an uncertainty dimension in a subjective opinion. To take benefits from both SL and PSL, we proposed a hybrid approach called CSL$^+$ for achieving high scalability and high prediction accuracy for unknown opinions with uncertainty derived from a lack of evidence and/or conflicting evidence. Through the extensive experiments on four semi-synthetic and two real-world datasets, we showed that the CSL$^+$ outperforms the state-of-the-art belief model (i.e., SL), probabilistic inference models (i.e., PSL, CSL) and deep learning model (i.e., GCN-VAE-opinion) in terms of prediction accuracy, computational complexity, and real running time.

CCS Concepts: • **Computing methodologies** → **Learning in probabilistic graphical models**; **Reasoning about belief and knowledge**; *Logical and relational learning*; *Learning latent representations*;

Additional Key Words and Phrases: uncertainty, subjective opinion, vacuity, conflicting evidence, opinion inference

## 1 INTRODUCTION

Effective decision making is closely related to how to manage uncertainty. Uncertainty is one of key causes generating biases which may mislead a decision making. Uncertainty in data and/or information is often originated from different aspects of its quality such as incomplete, missing, corrupted, or modified/forged information caused by either unreliable medium (e.g., wireless medium) or malicious actions (e.g., attacks) by adversarial entities. In addition, inherent cognitive, computational limitations by entities (either machines or humans) can introduce uncertainty being a serious hurdle in proper information processing. In addition, machine learning and/or data mining research has studied how to deal with uncertain data in the areas of collective classification [10], ontology alignment [9], personalized medicine [6], opinion diffusion [1], trust inference in social networks [12], and graph summarization [20].

Since the 1960's, belief models have been studied to solve dynamic decision making problems, such as Dempster-Shafer Theory (DST) [24], Fuzzy Logic (FL) [28], Transferable Belief Model (TBM) [26], Dezert-Smarandache Theory (DSmT) [25] and Subjective Logic (SL) [13]. In particular, in 1990's, SL is proposed to explicitly consider a uncertainty dimension in a subjective opinion.

Uncertainty in SL mainly refers to vacuity (or ignorance) caused by a lack of evidence. SL offers a rich set of operators for binomial, multinomial, and hyper opinions to combine different opinions based on structural relationships between them. However, combining multiple opinions in SL is limited based on dyadic relationships and needs to be performed sequentially, there is a scalability issue for a large-scale network data. In addition, although SL's capability in predicting unknown opinions is highly effective with a small-scale network data, SL has its inherent issue in scalability due to its nature in computing two opinions at a time to derive a fused opinion. Further, the uncertainty in SL is only limited to dealing with vacuity, not uncertainty introduced by conflicting evidence. For example, in a binomial opinion in SL, even if there exists a same amount of evidence supporting two extremes (e.g., pro. vs. con.), uncertainty can be close to 0. In this situation, the information a decision maker has is not really useful for his/her decision making. The scalability issue for large-scale network data can be relaxed by a probabilistic model, called Probabilistic soft logic (PSL). PSL is proposed to resolve intractable, complexity issues of previous opinion inference problems (e.g., Markov logic networks, or MLNs) based on logic rules which allows collective inference of unknown opinions for high scalability. However, PSL does not deal with uncertainty in subjective opinions.

To consider uncertainty in an opinion based on SL as well as scalability via collective inference process in PSL, we previously proposed an opinion inference algorithm, called *Collective Subjective Logic* (CSL) [31], by combining PSL and SL. However, uncertainty considered in CSL is still limited to vacuity, not considering uncertainty caused by conflicting evidence. Hence, in this work, we further enhanced CSL whose opinion inference performance is highly resilient against vacuity and conflicting evidence. We name it Collective Subjective Logic Plus, $\texttt{CSL}^+$. $\texttt{CSL}^+$ is an enhanced version of CSL [31] that achieves high resilience against multidimensional uncertainty in vacuity and conflicting evidence, leading to high prediction accuracy in opinion inference with highly uncertain opinions.

We made the following **key contributions** in this work:

(1) **Providing a scalable solution for opinion inference under uncertainty**: The proposed $\texttt{CSL}^+$ is a hybrid approach taking the merits of both PSL and SL as an enhanced version of CSL [31] that deals with uncertain opinions based on SL while achieving high scalability for a large-scale network data using the collective process of opinions in PSL. We leverage a technique called *posterior regularization* (PR) providing the ability to collectively predict opinions of multiple variables based on known opinions of other variables and their structural relationships following the logic rules in PSL.

(2) **Maximizing prediction accuracy of opinion inference under multidimensional uncertainty**: $\texttt{CSL}^+$ deals with multiple root causes of uncertainty (i.e., lack of evidence and/or conflicting evidence) that can significantly hinders effective decision making. To this end, the proposed $\texttt{CSL}^+$ leverages the learning from known opinions (i.e., training nodes) for maximizing prediction accuracy.

(3) **Validation of $\texttt{CSL}^+$ via extensive experiments based on six datasets**: $\texttt{CSL}^+$ is validated based on four semi-synthetic dataset and two real world datasets with three different tasks: (1) the trust inference task (i.e., trusted vs. not trusted to determine who trusts who) using Epinions (a semi-synthetic) datasets; (2) the traffic congestion inference task (i.e., congested vs. non-congested in road traffic) using District of Columbia (D.C.) and Pennsylvania (PA) road traffic datasets; and (3) the Sybils attack inference task (e.g., Sybil vs. benign node in a network) using Facebook, Enron and Slashdot social network datasets.

(4) **Proving the outperformance of $\texttt{CSL}^+$ over other competitive counterparts**: We conducted extensive experiments to ensure the outperformance of $\texttt{CSL}^+$ over other state-of-the-art counterparts, including SL, PSL, CSL, and a GCN-VAE-opinion model (i.e., a deep learning model

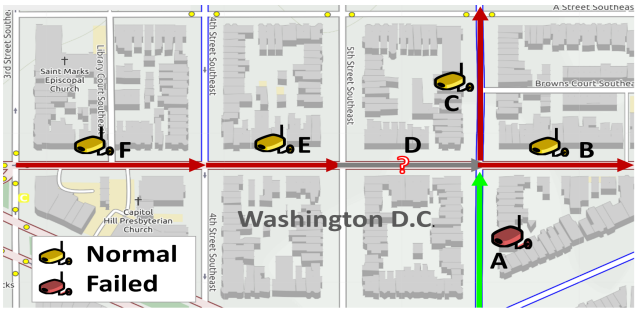



Fig. 1. A scenario of conflicting evidence and the prediction of congestion on Washington D.C. road network. A~F are road sections, 'red' indicates congested road, 'green' is for non-congested. A, B, C, E, and F have observations collected from traffic sensors. The status of D is unknown, the sensor of A is broken and the observed status is non-congested which conflicts with the observed status of A's upstream where road sections B and C are congested.

using graphical convolutional neural networks with variational autoencoder) based on all six datasets (i.e., four semi-synthetic dataset and two real world datasets) in terms of prediction accuracy of unknown opinions, computational algorithmic complexity, and real running time.

The rest of this paper is structured as follows. Section 2 provides the overview of related work including probability models, belief models, and other advanced inference models. Section 3 gives background on SL and PSL, which are mainly leveraged to develop CSL$^+$. Section 4 describes an example scenario and the problem statement. Section 5 provides the details of the proposed CSL$^+$. Section 6 describes the experimental setting and datasets and discusses the observed trends of experimental results. Section 7 concludes the paper and suggests the future research directions.

## 2 RELATED WORK

In this section, we provide the overview of the related approaches for opinion inference under uncertainty, including probability models, belief models, and machine/deep learning models.

### 2.1 Probabilistic Models

Inference problems with lack of information in structured network data have been studied based on a joint probability distribution where each node represents a random variable in the network. To be specific, Markov random fields (MRFs) [21] uses potential functions of cliques to obtain structural relationships that models the joint distribution over a set of variables. A Markov logic network (MLN) [23] models Boolean MRFs via first-order logic. The MLN can be obtained by attaching weights to the clauses (or formulas) in a first-order knowledge base, and can be seen as a constructing template of a simple Markov network. In spite of its high interpretability, inferring opinions from MLNs is computationally intractable. On the other hand, existing approximation counterparts generate high computational complexity with low accuracy. To resolve the high complexity of MLNs, a probabilistic soft logic, namely PSL [12, 18], is introduced to define how truth probabilities are related to each other in binary variables. PSL uses 'Hinge-Loss Markov Random Fields' (HL-MRFs) [5] to define logic rules, on the basis of graphical models with 'log-concave density functions,' which enables better efficiency and scalability than MLNs. Nevertheless, PSL does not deal with relations of truth probabilities when opinions are uncertain.

### 2.2 Belief Theory

Since the 1960's, belief theories have been studied to solve decision making problems under uncertainty. The examples include Dempster-Shafer Theory (DST) [24], Fuzzy Logic (FL) [28], Transferable Belief Model (TBM) [26], Dezert-Smarandache Theory (DSmT)[25] and Subjective

Logic (SL) [13]. Fuzzy Logic [28] evaluates imprecise information considering its vagueness which can be one of the root causes introducing uncertainty. DST is the most well-known belief model but is well-known for its counter-intuitive outcome with input of conflicting evidence. To resolve this issue in DST [24], TBM [26] and DSmT [25] are proposed; but they failed to properly consider conflicting evidence. SL dealing with vacuity as the uncertainty dimension considered in a subjective opinion is substantially used to develop trust networks or security mechanisms by leveraging various types of operators to solve fusion problems [15]. However, most SL operators combine two opinions sequentially which hinders scalability and requires high complexity for the opinion inference of a large-scale network data. In this work, by fully leveraging the merits of PSL and SL, we propose a scalable opinion inference algorithm considering the multidimensional causes of uncertainty.

### 2.3 Other Advanced Inference Models

Collective Subjective Logic (CSL) [31] is an uncertainty opinion reasoning method used for the cases where all the node level subjective opinions in a network have different levels of beliefs (i.e., belief and disbelief) with a same level of uncertainty. Our proposed $\texttt{CSL}^+$ combines SL and PSL as CSL [31], but considers conflicting evidence for the opinion inference. Graph convolutional neural networks (GCNs) are used to infer opinions along with variational autoencoder (VAE) [34]. GCN-VAE-opinion (GCN-VAE) method is proposed by formulating an opinion based on SL to explicitly deal with uncertainty (vacuity) while GCN and VAE are used to achieve low complexity of opinion inference for large-scale network data.

Beta ProbLog [40] was recently proposed, a probabilistic logical programming approach that reasons in presence of uncertain probabilities represented as Beta-distributed random variables. [41] estimates approximate credible intervals or "Bayesian error bars" around the model outputs. [42] generalized belief propagation to infer opinions over binary propositions in a singly connected graph.

Inference and/or prediction with conflicting information is highly challenging and has been studied in various domains including the data mining, machine learning, and database systems, especially the data integration in the database system. Resolving conflict data from multiple data sources are studied in data mining or web mining, including logic-based methods on data integration tasks [37] and investigation of dependencies between data sources to find the true values from conflicting information [36]. These works [36, 37] above mainly focus on resolving conflict values from different data sources (i.e. data from different database tables, or websites) via probabilistic or logic-based methods to choose the truth value from one of the data source; however, they do not resolve uncertainty caused by conflicting evidence upon all evidence considered from multiple sources. DSmT [25] is more efficient in combining conflicting evidence than the DST [24], but it has shortcomings when it comes to combining the rules, has high computational complexity [39]. SL offers a number of operators that combine two subjective opinions based on a transitive trust or consensus where the opinions are independent to each other [15, 35].

## 3 BACKGROUND

In this section, we provide the brief overview of SL and PSL which are mainly leveraged to develop $\texttt{CSL}^+$ in this work.

### 3.1 Subjective Logic

SL provides ways to define three types of subjective opinions: binomial, multinomial, and hyper opinions [15]. In this work, we consider the binomial opinion to help decision making in given proposition. A binomial opinion $\omega$ is defined by [15]:

$$\omega = (b, d, u, a) \tag{1}$$

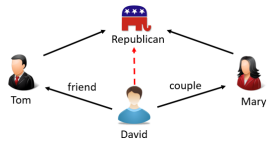



Fig. 2. Prediction of voters intuition based on the opinions of twitter users.

where $b$ indicates belief (e.g., agree), $d$ indicates to disbelief (e.g., disagree), $u$ refers uncertainty (i.e., vacuity or ignorance), and $a$ is a base rate supporting $b$ (or $(1-a)$ supporting $d$), representing a prior knowledge upon no commitment (e.g., neither true nor false), where $b+d+u=1$ for $(b,d,u,a)\in[0,1]^4$. When a belief is multinomial, multiple beliefs exist with a *belief mass distribution* $\mathbf{b}$ over the states of the variable and an *uncertainty mass*. A *base rate* $(a)$ can be seen as a domain prior knowledge related to the given proposition or a human entity's judgmental capability or bias, where $a$ can be interpreted in either way. When an agent's prior domain knowledge is not used, we will omit $a$ and denote an opinion as $\omega=(b,d,u)$.

A binomial subjective opinion corresponds to a Beta PDF (probability density function) via a bijective mapping, denoted by $\text{Beta}(p;\alpha,\beta)$, and is given by:

$$\text{Beta}(p;\alpha,\beta)=\frac{\Gamma(\alpha+\beta)p^{\alpha-1}(1-p)^{\beta-1}}{\Gamma(\alpha)\Gamma(\beta)}. \tag{2}$$

where $0\le p\le 1, \alpha>0, \beta>0$, and the $\alpha$ and $\beta$ is obtained based on the base rate $a$, the amount of positive evidence $r$ and the amount of negative evidence $s$: $\alpha=r+aW$ and $\beta=s+(1-\alpha)W$. Based on the mapping rule in SL [14], the vector of an opinion, $\omega=(b,d,u)$, is given by:

$$b=\frac{r}{r+s+W},\ d=\frac{s}{r+s+W},\ u=\frac{W}{r+s+W}. \tag{3}$$

where $W$ is the default non-informative prior weight, and usually set $W=2$ supposing that there is a complete uncertainty when $\alpha=1$ and $\beta=1$ at the very beginning of opinion update, resulting in $u=1$. The base rate $a$ is set to the default value 0.5 referring to a neutral position. As time elapses, more positive/negative evidence is gathered with more new evidence $\alpha$ and $\beta$, leading to decreasing uncertainty $u$. An subjective opinion $\omega$ can be expressed based on the amount of new evidence received as:

$$\omega=(\alpha,\beta) \tag{4}$$

This can be mapped to $\omega'=(b,d,u,a)$ as in Eq. (3) where the base rate $a$ is given as a constant derived from a historical record. In SL, the transitive trust [17] is estimated based on a discounting operator, $\otimes$. We denote $i$'s trust in $j$ by $\omega_j^i=(b_j^i,d_j^i,u_j^i,a_j^i)$. $j$'s trust in $k$ is expressed by $\omega_k^j=(b_k^j,d_k^j,u_k^j,a_k^j)$. Thus, $i$'s trust in $k$, denoted by $\omega_k^{i:j}:=(b_k^{i:j},d_k^{i:j},u_k^{i:j},a_k^{i:j})=\omega_j^i\otimes\omega_k^j$, is:

$$b_k^{i:j}=b_j^ib_k^j,\ \ d_k^{i:j}=b_j^id_k^j,\ \ u_k^{i:j}=d_j^i+u_j^i+b_j^iu_k^j,\ \ a_k^{i:j}=a_k^j. \tag{5}$$

To combine two opinions, a consensus operator, $\oplus$, is used as follows [17]. After two trust opinions towards a same entity are combined, the combined opinion is represented by $\omega_k^{i\diamond j}:=(b_k^{i\diamond j},d_k^{i\diamond j},u_k^{i\diamond j},a_k^{i\diamond j})=\omega_k^i\oplus\omega_k^j$ with:

$$b_k^{i\diamond j}=\frac{b_k^iu_k^j+b_k^ju_k^i}{\zeta},\ d_k^{i\diamond j}=\frac{d_k^iu_k^j+d_k^ju_k^i}{\zeta} \tag{6}$$

$$u_k^{i\diamond j}=\frac{u_k^iu_k^j}{\zeta},\ a_k^{i\diamond j}=a_k^i.$$

Table 1. An illustration of SL-based Unknown Relation Inference. The relationships are Friend, Spouse, and VoteFor where related entities include David ($D$), Tom ($T$), Mary ($M$), and Republican party ($Rep$).

| **Observed opinions** | **Given/Inferred opinions** |
|---|---|
| $y_1 \equiv \text{Friend}(D, T)$ | $\omega_{y_1} = (1, 0, 0, 0.5)$ |
| $y_2 \equiv \text{Spouse}(D, M)$ | $\omega_{y_2} = (1, 0, 0, 0.5)$ |
| $y_3 \equiv \text{VoteFor}(T, Rep)$ | $\omega_{y_3} = (0.6, 0.1, 0.3, 0.5)$ |
| $y_4 \equiv \text{VoteFor}(M, Rep)$ | $\omega_{y_4} = (0.8, 0.1, 0.1, 0.5)$ |
| $y_5 \equiv y_1 : y_3$ | $\omega_{y_5} = \omega_{y_1} \otimes \omega_{y_3} = (0.6, 0.1, 0.3, 0.5)$ |
| $y_6 \equiv y_2 : y_4$ | $\omega_{y_6} = \omega_{y_2} \otimes \omega_{y_4} = (0.8, 0.1, 0.1, 0.5)$ |
| **Unknown opinion** | **Inferred opinion** |
| $x \equiv \text{VoteFor}(D, Rep)$ | $\omega_x = \omega_{y_5} \oplus \omega_{y_6} = (0.81, 0.11, 0.08, 0.5)$ |

where $\zeta = u_j^i + u_k^j - u_j^i u_k^j > 0$. When $\zeta = 0$, $\omega_k^i \oplus \omega_k^j$ is defined by:

$$b_k^{i\diamond j} = \frac{\psi b_k^i + b_k^j}{\psi + 1},\ d_k^{i\diamond j} = \frac{\psi d_k^i + d_k^j}{\psi + 1},\quad u_k^{i\diamond j} = 0,\ a_k^{i\diamond j} = a_k.$$

where $\psi = \lim(u_k^j / u_k^i)$. SL provides discounting " $\otimes$", and consensus "$\oplus$", operators to estimate trust relationships following a trust chain of subjective trust opinions [17].

In Fig. 2, we show an example for predicting people's voting intentions via a social network with two kinds of relations, "Friend" and "Spouse", when a set of observed/given opinions $\mathbf{y} = \{y_1, \cdots, y_M\}$ and a set of unknown opinions $\mathbf{x} = \{x_1, \cdots, x_N\}$ are given. These opinions are represented by a set of random opinion variables where the format of each opinion follows Eq. (1) (i.e., $\omega_{x_i}$ for $i = 1 \cdots n$ and $\omega_{y_j}$ for $j = 1 \cdots m$).

Given the following rule holds,

$$\text{Friend}(D, T) \wedge \text{VoteFor}(T, Rep) \rightarrow \text{VoteFor}(D, Rep), \tag{7}$$

This rule implies that if David ($D$) and Tom ($T$) are friends and $T$ voted for the Republican ($Rep$) party, $T$ may vote for $Rep$. Similarly, we assume holding the following rule,

$$\text{Spouse}(D, M) \wedge \text{VoteFor}(M, Rep) \rightarrow \text{VoteFor}(D, Rep). \tag{8}$$

In this rule, if $D$'s spouse is Mary ($M$) and $M$ voted for $Rep$, then $D$ may vote for $Rep$. Heads of rule (8) and (9) imply the same opinion for VoteFor(D,Rep), $D$ may vote for Rep. In order to infer the the target opinion for VoteFor(D,Rep), we first identify two independent transitive paths in the network. We can use discounting operator to compute these two transitive paths (different relationships and voting intention opinions) $y_5$ and $y_6$ separately, then and combine these two opinions for VoteFor(D,Rep) using consensus operator. In Table 1, $y_1 \sim y_6$ denotes the trust relations of entities with subjective opinions in a certain proposition. We infer a unknown relation, $x$, based on known relationships, $y_1 \sim y_6$, by using the consensus ($\oplus$) and discounting ($\otimes$) operators in SL. The opinions $y_1 - y_4$ are given while $y_5$ and $y_6$ are derived separately using $\otimes$ and $\oplus$: $\omega_{y_5} = \omega_{y_1} \otimes \omega_{y_3}$ and $\omega_{y_6} = \omega_{y_2} \otimes \omega_{y_4}$. Lastly, the unknown opinion, $x$, is evaluated by $\omega_x = \omega_{y_5} \oplus \omega_{y_6}$.

### 3.2 Probabilistic Soft Logic

PSL is a 'statistical relational learning' (SRL) [38] technique applying logics to formulated SRL problems. PSL offers a formulation tool providing a user friendly language interface for collective, probabilistic reasoning point-valued unknown probabilities under known probabilities. In PSL, the key instances are 'truth probabilities of binary random variables' (i.e., Boolean). PSL employs 'weighted first-order logic rules with conjunctive bodies and single literal heads' that considers the

structural relationships between the random variables. Each logic rule is assigned with non-negative weights with a real value in $[0, 1]$, where the weight indicates the importance of a logic rule. To compare PSL with SL, we present the same example in Section 3.1 showing the prediction of a voter intuition based on the two types of relations, *Friend* in Rule (7) and *Spouse* in Rule (8).

PSL defines a non-negative *rule weight* per rule representing the confidence (i.e., importance) on the logic rule. For instance, given the weight of Rule (7) is 0.3 and the weight of Rule (8) is 0.8, we interpret that the spouse relation is stronger than the friend relation, implying that a same vote can be observed in a stronger relation such as a spouse relation than a friend relation. PSL has its syntax rule basis on the first order logic and uses a real number representing *soft truth probabilities* in $[0, 1]$, instead of a binary decision 0 or 1. We call probability $p_{x_i}$ an *atom* for a random variable $x_i$ that refers to a certain relation. $\mathbf{p_x} = [p_{x_1}, \cdots, p_{x_N}]^T \in [0, 1]^n$ means a vector of unknown atoms. In the above voter example, Rule (8) is weighed more than than Rule (7), which means in a spouse relationship, one more tends to vote for what his/her voted for than what his/her friend voted for. To assess how well the ground logic rule works, PSL adopts the *Lukasiewicz t-norm and co-norm* as the relaxation of the logical operators $\wedge$ (conjunction or logical AND), $\vee$ (disjunction or logical OR), and $\neg$ (negation or logical NOT) [19], respectively, by:

$$p_{x_1} \wedge p_{x_2} = \max[p_{x_1} + p_{x_2} - 1, 0],\ p_{x_1} \vee p_{x_2} = \min[p_{x_1} + p_{x_2}, 1],\ \neg p_{x_1} = 1 - p_{x_1}. \tag{9}$$

These operators coincide with the Boolean logic operators for integer inputs; but it also offers a consistent mapping for intermediate values between integers. A grounding PSL logic rule, $r_k$, is represented by:

$$\bigwedge_{i \in I^-} p_{x_i} \rightarrow \bigvee_{i \in I^+} p_{x_i}, \tag{10}$$

where $I^- \subset \{1, \cdots, N\}$ refers to a set of indices in the body of the rule $r_k$ and $I^+ \subset \{1, \cdots, N\} \setminus I^-$ indicates a set of indices in the head of the rule $r_k$. $r_k(\mathbf{p_x})$ captures the degree of rule k-$th$ being satisfied and is given by:

$$r(\mathbf{p_x}) = \min\Big[\sum_{i \in I^+} p_{x_i} + \sum_{i \in I^-} (1 - p_{x_i}), 1\Big]. \tag{11}$$

For example, considering the realization of the atom,

$$p_{x_1} \equiv \text{Spouse}(B, A) = 1.0,\ p_{x_2} \equiv \text{VotesFor}(A, P) = 0.8, \text{and } p_{x_3} \equiv \text{VotesFor}(B, P) = 0.1.$$

Let $r_k$ be the corresponding grounding instance of Rule (8). We get that $r(\mathbf{p_x}) = \min[0.1 + ((1 - 1) + (1 - 0.8)), 1] = 0.3$. Instead of soft values in $[0, 1]$, if all the atoms are Boolean (binary) variables (True '1' or False '0'), Eq. (11) is similar to:

$$\Big(\bigvee_{i \in I^+} p_{x_i}\Big) \bigvee \Big(\bigvee_{i \in I^-} \neg p_{x_i}\Big), \tag{12}$$

where the satisfaction of this rule (i.e., 1 or 0) is same as that in a standard first-order logic.

Overall PSL uses first-order logic syntax to define constraints and potential functions in a graphical model over the truth values of logical atoms to infer unknown truth values. More detailed information is provided in [5, 6].

## 4 EXAMPLE SCENARIO AND PROBLEM STATEMENT

In this section, we describe an example realistic scenario and a problem formulation.

### 4.1 Example Scenario

Let us take an example real-world application scenario aiming to predict the traffic congestion in a road network where a node refers to an intersection of road links and an edge is a road link. On the commuting roads, we may encounter traffic congestion. To check the road conditions, we often listen to the radio and/or may use GPS or online maps. The prediction of road congestion is expected to be presented with a certain probability. Although a rich, diverse traffic related applications are available, unexpected traffic congestion caused by the latest accidents may occur without any chance of knowing it. Hence, it is highly challenging to predict unknown traffic conditions when some partial updates and/or observations are available. Note that the road traffic prediction is one of examples we can apply our CSL$^+$ and we discuss our experimental results using six different semi-synthetic and real datasets to prove the outperformance of the CSL$^+$ in Section 6.

### 4.2 Problem Statement

Given a structured network, denoted by $\mathcal{G} = (\mathbb{V}, \mathbb{E}, f)$ where $\mathbb{V} = \{1, 2, \cdots, l\}$ is a set of vertices representing intersections of road links and $\mathbb{E} \subseteq \mathbb{V} \times \mathbb{V}$ is a set of edges indicating road links, we define the mapping function $f : \mathbb{E} \rightarrow \{0, 1\}$ with a Boolean variable $f(e_i)$ for each edge $e_i$, with '0' indicating a non-congested road and '1' meaning a congested road at a current time. Suppose there is a subset of edges $\mathbb{E}_y = \{e_1, \cdots, e_M\} \subseteq \mathbb{E}$ with traffic sensors (e.g., cameras, speed radar) installed at these edges representing known opinions. For the edges without traffic sensors, we treat them unknown opinions, denoted by $\mathbb{E}_x = \{o_1, \cdots, o_N\}$, and $\mathbb{E} = \mathbb{E}_y \cup \mathbb{E}_x$. In addition, we consider a set of unknown traffic sensors that report unreliable observations providing conflicting observations to the true traffic observations (labels) of their corresponding edges.

Suppose we have a vector of observations on road congestion status at a current time on $\mathbb{E}_y$ which shows known opinions $\mathbf{y} = [f(e_1), \cdots, f(e_M)] \in \{0, 1\}^M$ based on observations, and the beliefs towards the states of these variables derived based on their historical data on the observations. Given $\mathbb{E}_y$, our goal is to predict the beliefs of the states on the congestion variables at the edges without sensors (i.e., intersections without any camera or speed sensors), denoted as $\mathbf{x} = [f(o_1), \cdots, f(o_N)] \in \{0, 1\}^N$. A belief over the states of a road congestion variable $x_i$ (or $y_j$) can be represented as a subjective opinion, defined as $\omega_{x_i} = (b_{x_i}, d_{x_i}, u_{x_i}, a_{x_i})$, $\omega_{y_j} = (b_{y_j}, d_{y_j}, u_{y_j}, a_{y_j})$ in Eq. (1), or following a Beta distribution with given parameter evidence $(\alpha_{x_i}, \beta_{x_i})$, Beta$(p_{x_i}; \alpha_{x_i}, \beta_{x_i})$ (see Eq. (2)). $\omega_{x_i}$ is also obtained by SL's mapping rule in Eq. (3). For edge $e_i \in \mathbb{E}_y$ with historical observations $\sum_{t=1}^{T} f(e_i^t)$ where $r_y$ and $s_y$ are the counts of 0's (i.e., no traffic congestion) and 1's (i.e., have traffic congestion ) in these observations, respectively. Beta$(p_{y_j}; \alpha_{y_j}, \beta_{y_j})$ is given by:

$$\alpha_{y_j} = r_{y_j} + a_{y_j} W, \quad \beta_{y_j} = s_{y_j} + (1 - a_{y_j}) W. \tag{13}$$

where $r_{y_j}$ and $s_{y_j}$ are the volumes of positive and negative evidence, $W$ is a predefined non-informative prior weight representing an amount of uncertain evidence, and $a_{y_j}$ is the base rate on proposition $y_j$, predefined probability of prior general knowledge on $y_j$ used to interpret $W$. Because the edges in $\mathbb{E}_x$ do not have historical observations, we cannot directly infer their beliefs. However, we can infer the unknown beliefs towards the edges in $\mathbb{E}_x$ based on the structural relationships between the known beliefs on the edges in $\mathbb{E}_y$. We also consider untrustworthy known sources (i.e., edge sensors) providing conflicting evidence.

As the identification of conflicting sources and the prediction of unknown opinions are equally important because of their inter-dependency, we formulate a statistical model for predicting unknown opinions and learning the conflicting sources simultaneously as follows:

**Problem 1 (Uncertainty-based opinion prediction against the conflicting evidence in network data):** *The key notations are:*

- *Let* $\mathbf{y} = (y_1, \cdots, y_M)$ *be a vector of given input binary variables whose opinions are denoted by* $\boldsymbol{\omega}_\mathbf{y} = (\omega_{y_1}, \cdots, \omega_{y_M})$, $\text{Beta}(p_{y_i}; \omega_{y_i})$ *refers to the PDF of the truth probability* $p_{y_i}$ *of the variable* $y_i$.
- *Let* $\mathbf{x} = (x_1, \cdots, x_N)$ *be a vector of target binary variables whose opinions are represented by* $\boldsymbol{\omega}_\mathbf{x} = (\omega_{x_1}, \cdots, \omega_{x_N})$ *to be predicted.* $\text{Beta}(p_{x_i}; \omega_{x_i})$ *is the PDF of the truth probability* $p_{x_i}$ *of the variable* $x_i$.
- *Let* $\mathbf{p}_\mathbf{x} = (p_{x_1}, \cdots, p_{x_N})$, $\mathbf{p}_\mathbf{y} = (p_{y_1}, \cdots, p_{y_M})$, *and* $\mathbf{p}_{\mathbf{x},\mathbf{y}} = (\mathbf{p}_\mathbf{x}, \mathbf{p}_\mathbf{y})$.
- $\mathbf{b}_\mathbf{y} = (b_{y_1}, \cdots, b_{y_M})$ *is a vector of binary variables, in which each* $b_{y_i} \in \{0, 1\}$ *indicates whether* $y_i$ *is a source providing conflicting evidence or not.*

**Given**

- $\mathbf{y} \in \{0, 1\}^M$, $\omega_{\mathbf{y}_i}$ *is the subjective opinion on* $\mathbf{y}_i$, *and*
- $\mathcal{R} = \{r_k, \rho_k\}_{k=1}^K$, *a set of PSL logic rules, in which* $r_k$ *is the* $k$*-th rule over* $\mathbf{p}_{\mathbf{x},\mathbf{y}}$ *and* $\mathbf{b}_\mathbf{y}$ , *and* $\rho_k$ *is the weight of* $r_k$. *A logic rule* $r_k$ *is defined as:*

$$r_k := \bigwedge_{i \in I^-_{\mathbf{x},k}} p_{x_i} \bigwedge_{i \in I^-_{\mathbf{y},k}} \Big(p_{y_i} \wedge (1 - b_{y_i})\Big) \rightarrow \bigvee_{i \in I^+_{\mathbf{x},k}} p_{x_i} \bigvee_{i \in I^+_{\mathbf{y},k}} \Big(p_{y_i} \vee b_{y_i}\Big), \tag{14}$$

*where* $I^-_{\mathbf{x},k}$ *and* $I^-_{\mathbf{y},k}$ *indicate the indices of variables in* $\mathbf{x}$ *and* $\mathbf{y}$ *occurring in the body of the logic rule* $r_k$, *separately; and* $I^+_{\mathbf{x},k}$ *and* $I^+_{\mathbf{y},k}$ *are the indices of variables in* $\mathbf{x}$ *and* $\mathbf{y}$ *occurring in the head of the logic rule* $r_k$.

**Predict** $\boldsymbol{\omega}_\mathbf{x}$, *the unknown opinion on a vector of target variables* $\mathbf{x}$ *and A conflicting evidence indicator vector,* $\mathbf{b}_\mathbf{y}$.

Without concerning the constraints based on PSL logic rules, the joint PDF of the input and output variables has the following form:

$$\begin{aligned} \text{Prob}(\mathbf{p}_{\mathbf{x},\mathbf{y}}, \mathbf{b}_\mathbf{y}, \mathbf{x}, \mathbf{y}; \boldsymbol{\omega}_\mathbf{x}, \boldsymbol{\omega}_\mathbf{y}, \boldsymbol{\omega}_0) \quad = \quad & \prod_{i=1}^{M} \Big\{ \big(\text{Beta}(p_{y_i}; \omega_{y_i})\big)^{1-b_{y_i}} \big(\text{Beta}(p_{y_i}; \omega_0)\big)^{b_{y_i}} \\ & \cdot \text{Bin}(y_i; p_{y_i}) \text{Bin}(b_{y_i}; p_0) \Big\} \prod_{l=1}^{N} \text{Bin}(x_l; p_{x_l}) \text{Beta}(p_{x_l}; \omega_{x_l}), \end{aligned} \tag{15}$$

where $\text{Bin}(\cdot)$ and $\text{Beta}(\cdot)$ refer to the PDF functions of a Binomial distribution and a Beta distribution, respectively. If $y_i$ is a source providing conflicting evidence, then its opinion $\omega_{y_i}$ set to the base rate opinion $\omega_0 = (1, 1)$, and the probability $p_{y_i} = 0.5$, which indicate the totally uncertain opinion and probability. The aim is to find the target unknown opinion vector $\boldsymbol{\omega}_\mathbf{x}$ and the conflicting-source indicator vector $b_\mathbf{y}$, which maximize the likelihood $\text{Prob}(\mathbf{y}; \boldsymbol{\omega}_\mathbf{x}, \boldsymbol{\omega}_\mathbf{y}, \boldsymbol{\omega}_0)$ subject to the constraints defined by the set of PSL rules $\mathcal{R}$. We adopt a generally employed strategy that enforces the PSL logic rule constraints on $\text{Prob}(\mathbf{p}_{\mathbf{x},\mathbf{y}}, \mathbf{b}_\mathbf{y} | \mathbf{y}; \boldsymbol{\omega}_\mathbf{x}, \boldsymbol{\omega}_\mathbf{y}, \boldsymbol{\omega}_0)$ through an expectation operator. In particular, for each rule $r_k$ and each of its related variables $\mathbf{p}_{\mathbf{x},\mathbf{y}}, \mathbf{b}_\mathbf{y}$, we expect

$$\mathbb{E}_{\text{Prob}(\mathbf{p}_{\mathbf{x},\mathbf{y}}, \mathbf{b}_\mathbf{y} | \mathbf{y}; \boldsymbol{\omega}_\mathbf{x}, \boldsymbol{\omega}_\mathbf{y}, \boldsymbol{\omega}_0)}[r_k(\mathbf{p}_{\mathbf{x},\mathbf{y}}, \mathbf{b}_\mathbf{y})] = 1,$$

with a confidence evaluated by the non-negative weight $\rho_k$. By imposing the constraints based on PSL logic rules, our key problem of predicting unknown opinions in the presence of conflicting evidence is defined as a maximization problem based on a constrained log-likelihood, $\mathcal{L}(\boldsymbol{\omega}_\mathbf{x})$ by:

$$\begin{aligned} \arg\max_{\boldsymbol{\omega}_\mathbf{x}, \boldsymbol{\xi} \geq 0} \mathcal{L}(\boldsymbol{\omega}_\mathbf{x}) = \arg\max_{\boldsymbol{\omega}_\mathbf{x}, \boldsymbol{\xi} \geq 0} \log \text{Prob}(\mathbf{y}; \boldsymbol{\omega}_\mathbf{x}, \boldsymbol{\omega}_\mathbf{y}, \boldsymbol{\omega}_0) \\ s.t.\ \rho_k \mathbb{E}_{\text{Prob}(\mathbf{p}_{\mathbf{x},\mathbf{y}}, \mathbf{b}_\mathbf{y} | \mathbf{y}; \boldsymbol{\omega}_\mathbf{x}, \boldsymbol{\omega}_\mathbf{y}, \boldsymbol{\omega}_0)} \Big[1 - r_k\ (\mathbf{p}_{\mathbf{x},\mathbf{y}}, \mathbf{b}_\mathbf{y})\Big] \leq \xi_k, \\ \|\boldsymbol{\xi}\|_\beta \leq \epsilon, k = [1 : K], \end{aligned} \tag{16}$$

where $\boldsymbol{\xi}_k$ indicates a slack variable, $\|\cdot\|_\beta$ represents a norm. We basically allow slight violations with slack variables $\xi_k$ on the PSL logic rules, the norm of $\xi$ is bounded by $\epsilon \geq 0$. Since our

Table 2. Key notations and their meanings.

| Notations | Description |
|---|---|
| $\mathbf{y} = [y_1, \cdots, y_M]$, $\mathbf{p}_\mathbf{y} = [p_{y_1}, \cdots, p_{y_M}]$, $\boldsymbol{\omega}_\mathbf{y} = [\omega_{y_1}, \cdots, \omega_{y_M}]$, $\mathbf{b}_\mathbf{y} = [b_{y_1}, \cdots, b_{y_M}]$ | $\mathbf{y}$ is a vector of $M$ input (observed) binary random variables whose opinions are given. $\mathbf{p}_\mathbf{y}$ and $\boldsymbol{\omega}_\mathbf{y}$ are the corresponding vectors of truth probabilities and opinions of $\mathbf{y}$, respectively. $\mathbf{b}_\mathbf{y}$ is a vector of $M$ binary random variables indicating the conflicting sources. |
| $\mathbf{x} = [x_1, \cdots, x_N]$, $\mathbf{p}_\mathbf{x} = [p_{x_1}, \cdots, p_{x_N}]$, $\boldsymbol{\omega}_\mathbf{x} = [\omega_{x_1}, \cdots, \omega_{x_N}]$ | $\mathbf{x}$ is a vector of $N$ output binary random variables whose opinions are predicted. $\mathbf{p}_\mathbf{x}$ and $\boldsymbol{\omega}_\mathbf{x}$ are the corresponding vectors of truth probabilities and opinions of $\mathbf{x}$, respectively. |
| $\omega_x = (b_x, d_x, u_x, a_x)$ | A binomial opinion of a binary random variable $x$ as defined in Eq. (1) |
| $[1:K]$ | $1, \cdots, K$ |
| $\omega'_x = (\alpha_x, \beta_x)$ | Evidence parameters of Beta$(p_x \vert \alpha_x, \beta_x)$ that corresponds to an opinion $\omega_x$ in Eq. (3). |
| $\mathbf{p}_{\mathbf{x},\mathbf{y}}$, $\boldsymbol{\omega}_{\mathbf{x},\mathbf{y}}$ | $\mathbf{p}_{\mathbf{x},\mathbf{y}} = (\mathbf{p}_\mathbf{x}, \mathbf{p}_\mathbf{y})$, $\boldsymbol{\omega}_{\mathbf{x},\mathbf{y}} = (\boldsymbol{\omega}_\mathbf{x}, \boldsymbol{\omega}_\mathbf{y})$ |
| $\mathcal{R} = \{r_k, \rho_k\}_{k=1}^K$ | A set of PSL logic rules as defined in Eq. (14), in which $r_k$ is the $k$-th rule over $\mathbf{p}_\mathbf{x}$, $\mathbf{p}_\mathbf{y}$, and $\mathbf{b}_\mathbf{y}$, and $\rho_k$ is the non-negative weight of $r_k$. |
| $q(\mathbf{p}_{\mathbf{x},\mathbf{y}}, \mathbf{b}_\mathbf{y})$ | A simpler new PDF function that fits the PSL rules as in $\mathcal{R}$ as well as meets the minimal Kullback-Leibler (KL)-divergence distance to the posterior Prob$(\mathbf{p}_{\mathbf{x},\mathbf{y}}, \mathbf{b}_\mathbf{y} \vert \mathbf{y}; \boldsymbol{\omega}_\mathbf{x})$ (See Eq. (17)). |

maximization problem is analytically intractable, we introduce an efficient robust approximation inference algorithm, CSL$^+$, to address this problem in Section 5.

## 5 CSL$^+$ FOR OPINION PREDICTION

Predicting unknown opinions in the presence of conflicting evidence is defined as a problem in Eq. (16). This problem has two challenging issues, in terms of high computational complexity and low scalability: (1) The expectation operator $\mathbb{E}_{\text{Prob}(\mathbf{p}_{\mathbf{x},\mathbf{y}}, \mathbf{b}_\mathbf{y} \vert \mathbf{y}; \boldsymbol{\omega}_\mathbf{x}, \boldsymbol{\omega}_\mathbf{y})}[1 - r_k(\mathbf{p}_{\mathbf{x},\mathbf{y}}, \mathbf{b}_\mathbf{y})]$ is computationally intractable; and (2) the dimensionality of unknown opinions $\boldsymbol{\omega}_\mathbf{x}$ is often high, aligned with network density (i.e., a number of edges) in the network data. We describe CSL$^+$ in two-fold: (1) prediction of unknown opinions in the presence of conflicting evidence (Eq. (16)) in Section 5.1; and (2) estimation of the expectation components as the part of CSL$^+$ and conflicting evidence inference in Section 5.2.

### 5.1 Prediction of Unknown Opinions in CSL$^+$

To solve the maximization problem in Eq. (16) and obtain a computationally tractable solution, we adopt *posterior regularization* (PR) [11], a probabilistic approach for the structural relational learning. PR is a technique for regularizing relational learning models by encoding prior knowledge in constraints on model posteriors. Via applying PR, CSL$^+$ learns a new simpler density function $q(\mathbf{p}_{\mathbf{x},\mathbf{y}}, \mathbf{b}_\mathbf{y})$ that fits the PSL logic rules while staying close to the posterior PDF (i.e. Prob$(\mathbf{p}_{\mathbf{x},\mathbf{y}}, \mathbf{b}_\mathbf{y} \vert \mathbf{y}; \boldsymbol{\omega}_\mathbf{x})$) concurrently, where we do not explicitly show the parameters $\boldsymbol{\omega}_\mathbf{y}$, $\boldsymbol{\omega}_0$ and $p_0$ that are assumed known as input. For each weighted PSL logic rule, $r_k$, we expect that $\mathbb{E}_{q(\mathbf{p}_{\mathbf{x},\mathbf{y}}, \mathbf{b}_\mathbf{y})}[r_k(\mathbf{p}_{\mathbf{x},\mathbf{y}}, \mathbf{b}_\mathbf{y})] = 1$, with a weight $\rho_k$, where $r_k(\mathbf{p}_{\mathbf{x},\mathbf{y}}, \mathbf{b}_\mathbf{y})$ is the level of the satisfaction of the PSL logic rule $r_k$ as defined in Eq. (11). All the constraints by the PSL logic rules in $\mathcal{R}$ construct a rule-restraint space of all valid distributions. To guarantee the closeness between the posterior PDF and $q(\mathbf{p}_{\mathbf{x},\mathbf{y}}, \mathbf{b}_\mathbf{y})$ and Prob$(\mathbf{p}_{\mathbf{x},\mathbf{y}}, \mathbf{b}_\mathbf{y} \vert \mathbf{y}; \boldsymbol{\omega}_\mathbf{x})$, we minimize the closeness using KL-divergence [8], a measure of the closeness of two distributions. Summing up the above two main considerations and further allowing the slight slackness for each

constraint, we formalize the following equivalent optimization problem:

$$\min_{q \in Q} \mathrm{KL}\Big(q(\mathbf{p}_{\mathbf{x},\mathbf{y}}, \mathbf{b}_\mathbf{y}), \mathrm{Prob}(\mathbf{p}_{\mathbf{x},\mathbf{y}}, \mathbf{b}_\mathbf{y} | \mathbf{y}; \boldsymbol{\omega}_\mathbf{x})\Big), \tag{17}$$

where $Q$, the domain of $q$, denotes the constrained posterior (with slacks) space of PDF as defined by the PSL logic rules, $\mathcal{R}$, and is defined as:

$$Q := \Big\{ q(\mathbf{p}_{\mathbf{x},\mathbf{y}}, \mathbf{b}_\mathbf{y}) : \exists \boldsymbol{\xi} > 0, \rho_k \mathbb{E}_{q(\mathbf{p}_{\mathbf{x},\mathbf{y}}, \mathbf{b}_\mathbf{y})}[1 - r_k(\mathbf{p}_{\mathbf{x},\mathbf{y}}, \mathbf{b}_\mathbf{y})] \le \xi_k; \|\boldsymbol{\xi}\|_\beta \le \varepsilon \Big\}. \tag{18}$$

The intent of the above optimization problem is to project the posterior PDF into the structural constrained posterior space, $Q$. This optimization can be effectively and efficiently solved in its dual problem with analytical closed-form solutions that are provided by the PR framework. Due to the space limit, we show the detailed derivation steps in **Appendix A** in our supplement material and directly show the following final solution:

$$q(\mathbf{p}_{\mathbf{x},\mathbf{y}}, \mathbf{b}_\mathbf{y}) \propto \mathrm{Prob}(\mathbf{p}_{\mathbf{x},\mathbf{y}}, \mathbf{b}_\mathbf{y} | \mathbf{y}; \boldsymbol{\omega}_\mathbf{x}, \boldsymbol{\omega}_\mathbf{y}) \cdot \exp\left\{ - \sum\nolimits_{k=1}^{K} \rho_k \Big(1 - r_k(\mathbf{p}_{\mathbf{x},\mathbf{y}}, \mathbf{b}_\mathbf{y})\Big) \right\}. \tag{19}$$

Apparently, from above Eq. (19), we can observe that a stronger PSL logic rule with large weight $\rho_k$ leads to lower probabilities, $\mathbf{p}_{\mathbf{x},\mathbf{y}}$, that result in failing to meet the constraints.

For a given $\boldsymbol{\omega}_\mathbf{x}$, a new desired density function, $q(\mathbf{p}_{\mathbf{x},\mathbf{y}}, \mathbf{b}_\mathbf{y})$, that fits the structural constraints of PSL logic rules while concurrently being close to our model posterior, can be estimated via given analytical form in Eq. (19). For a given approximated $q(\mathbf{p}_{\mathbf{x},\mathbf{y}}, \mathbf{b}_\mathbf{y})$, by Jensen's inequality [8], we construct an evidence lower bound $F(q, \boldsymbol{\omega}_\mathbf{x})$ of the log likelihood function $\mathcal{L}(\boldsymbol{\omega}_\mathbf{x}) = \log \mathrm{Prob}(\mathbf{y}; \boldsymbol{\omega}_\mathbf{x})$ by:

$$\mathcal{L}(\boldsymbol{\omega}_\mathbf{x}) = \log \sum_{\mathbf{p}_{\mathbf{x},\mathbf{y}}, \mathbf{b}_\mathbf{y}} q(\mathbf{p}_{\mathbf{x},\mathbf{y}}, \mathbf{b}_\mathbf{y}) \frac{\mathrm{Prob}(\mathbf{p}_{\mathbf{x},\mathbf{y}}, \mathbf{b}_\mathbf{y}, \mathbf{y}; \boldsymbol{\omega}_\mathbf{x})}{q(\mathbf{p}_{\mathbf{x},\mathbf{y}}, \mathbf{b}_\mathbf{y})} \tag{20}$$

$$\ge \sum_{\mathbf{p}_{\mathbf{x},\mathbf{y}}, \mathbf{b}_\mathbf{y}} q(\mathbf{p}_{\mathbf{x},\mathbf{y}}, \mathbf{b}_\mathbf{y}) \log \frac{\mathrm{Prob}(\mathbf{p}_{\mathbf{x},\mathbf{y}}, \mathbf{b}_\mathbf{y}, \mathbf{y}; \boldsymbol{\omega}_\mathbf{x})}{q(\mathbf{p}_{\mathbf{x},\mathbf{y}}, \mathbf{b}_\mathbf{y})} = F(q, \boldsymbol{\omega}_\mathbf{x}). \tag{21}$$

$F(q, \boldsymbol{\omega}_\mathbf{x})$ can be reformulated as:

$$\begin{aligned} F(q, \boldsymbol{\omega}_\mathbf{x}) &= \sum_{\mathbf{p}_{\mathbf{x},\mathbf{y}}, \mathbf{b}_\mathbf{y}} q(\mathbf{p}_{\mathbf{x},\mathbf{y}}, \mathbf{b}_\mathbf{y}) \log \Big( \mathrm{Prob}(\mathbf{p}_{\mathbf{x},\mathbf{y}}, \mathbf{b}_\mathbf{y} | \mathbf{y}; \boldsymbol{\omega}_\mathbf{x}) \mathrm{Prob}(\mathbf{y}; \boldsymbol{\omega}_\mathbf{x}) \Big) - \sum_{\mathbf{p}_{\mathbf{x},\mathbf{y}}, \mathbf{b}_\mathbf{y}} q(\mathbf{p}_{\mathbf{x},\mathbf{y}}, \mathbf{b}_\mathbf{y}) \log q(\mathbf{p}_{\mathbf{x},\mathbf{y}}, \mathbf{b}_\mathbf{y}) \\ &= \mathcal{L}(\boldsymbol{\omega}_\mathbf{x}) - \mathrm{KL}\Big(q(\mathbf{p}_{\mathbf{x},\mathbf{y}}, \mathbf{b}_\mathbf{y}), \mathrm{Prob}(\mathbf{p}_{\mathbf{x},\mathbf{y}}, \mathbf{b}_\mathbf{y} | \mathbf{y}; \boldsymbol{\omega}_\mathbf{x})\Big), \end{aligned} \tag{22}$$

where $\sum_{\mathbf{p}_{\mathbf{x},\mathbf{y}}, \mathbf{b}_\mathbf{y}} q(\mathbf{p}_{\mathbf{x},\mathbf{y}}, \mathbf{b}_\mathbf{y}) = 1$ and $\mathcal{L}(\boldsymbol{\omega}_\mathbf{x}) = \log \mathrm{Prob}(\mathbf{y}; \boldsymbol{\omega}_\mathbf{x}) = \log \sum_{\mathbf{p}_{\mathbf{x},\mathbf{y}}, \mathbf{b}_\mathbf{y}} \mathrm{Prob}(\mathbf{p}_{\mathbf{x},\mathbf{y}}, \mathbf{b}_\mathbf{y}, \mathbf{y}; \boldsymbol{\omega}_\mathbf{x}))$. According to the above interpretation, we design a modified Expectation Maximization (EM) inference algorithm to solve the unknown opinion prediction under conflicting evidence problem in Eq. (16). As shown in Fig. 3, beginning from an initial parameter estimate $\boldsymbol{\omega}_\mathbf{x}^l$ (or $\theta^l$) at step $l = 0$ in the first starting iteration, our algorithm iterates two block-coordinate ascent, and has the following $\mathbf{E}'$ and $\mathbf{M}$ steps:

- $\mathbf{E}' - \mathbf{Step}$, a **modified E-step** that includes the constraints defined by the PSL rules $\mathcal{R}$:

$$q^{l+1} = \arg\max\nolimits_{q \in Q} F(q, \boldsymbol{\omega}_\mathbf{x}^l) = \arg\min_{q \in Q} \mathrm{KL}\Big(q(\mathbf{p}_{\mathbf{x},\mathbf{y}}, \mathbf{b}_\mathbf{y}), \mathrm{Prob}(\mathbf{p}_{\mathbf{x},\mathbf{y}}, \mathbf{b}_\mathbf{y} | \mathbf{y}; \boldsymbol{\omega}_\mathbf{x})\Big), \tag{23}$$

  where $Q$ is the PSL logic rule constrained space of the PDF as defined in Eq. (18), and the analytical solution of $q^{l+1}$ is given in Eq. (19).
- $\mathbf{M} - \mathbf{Step}$:

$$\boldsymbol{\omega}_\mathbf{x}^{l+1} = \arg\max\nolimits_{\boldsymbol{\omega}_\mathbf{x}} F(q^{l+1}, \boldsymbol{\omega}_\mathbf{x}) = \arg\max_{\boldsymbol{\omega}_\mathbf{x}} \mathbb{E}_{q^{l+1}}[\log \mathrm{Prob}(\mathbf{p}_{\mathbf{x},\mathbf{y}}, \mathbf{b}_\mathbf{y}, \mathbf{y}; \boldsymbol{\omega}_\mathbf{x}, \boldsymbol{\omega}_\mathbf{y}, \boldsymbol{\omega}_0)]. \tag{24}$$

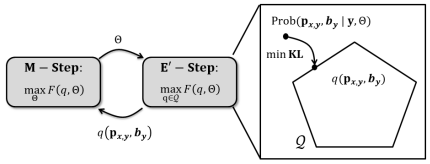



Fig. 3. Modified EM for the unknown opinion prediction under conflicting evidence problem in Eq. (16) in CSL$^+$.

From the joint PDF in Eq. (15), the objective function of E-Step in Eq. (24) can be rewritten as:

$$\begin{aligned}
\boldsymbol{\omega}_{\mathbf{x}}^{l+1} &= \arg\max_{\boldsymbol{\omega}_{\mathbf{x}}} \Big\{ \sum_{i=1}^{N} \mathbb{E}_{q^{l+1}}[\log \text{Beta}(p_{x_i}; \omega_{x_i})] + \sum_{i=1}^{M} \mathbb{E}_{q^{l+1}} \Big[ (1 - b_{y_i}) \log \text{Beta}(p_{y_i}; \omega_{y_i}) \\
&+ b_{y_i} \log \text{Beta}(p_{y_i}; \omega_0) \Big] + \sum_{i=1}^{M} \mathbb{E}_{q^{l+1}} \log \text{Bin}(b_{y_i}; p_0) + \sum_{i=1}^{M} \mathbb{E}_{q^{l+1}}[\log \text{Bin}(y_i; p_{y_i})] \Big\} \\
&= \arg\max_{\boldsymbol{\omega}_{\mathbf{x}}} \sum_{i=1}^{N} \mathbb{E}_{q^{l+1}}[\log \text{Beta}(p_{x_i}; \omega_{x_i})] + const.
\end{aligned}$$

The above point-wise summation equation implies that the opinions in $\boldsymbol{\omega}_{\mathbf{x}}$, including $\{\omega_{x_1}, \cdots, \omega_{x_N}\}$, can be optimized separately. The opinion $\omega_{x_i}$ can be computed by treating the other opinions in Eq. (24) as constants, where $\omega'_{x_i} = (\alpha_{x_i}, \beta_{x_i})$ as we defined in Eq. (4). Ultimately, the main goal of **M-step** is solving the following optimization point-wise sub-problem:

$$\max_{\alpha_{x_i}>0, \beta_{x_i}>0} \quad \mathbb{E}_{q^{l+1}} \Big[ \log \text{Beta}(p_{x_i} | \alpha_{x_i}, \beta_{x_i}) \Big] + const. = \tag{25}$$

$$\max_{\alpha_{x_i}>0, \beta_{x_i}>0} \quad \log \Gamma(\alpha_{x_i} + \beta_{x_i}) - \log \Gamma(\alpha_{x_i}) - \Gamma(\beta_{x_i}) + (\alpha_{x_i} - 1) \mathbb{E}_{q^{l+1}}[\log p_{x_i}] \tag{26}$$

$$+ \quad (\beta_{x_i} - 1) \mathbb{E}_{q^{l+1}}[\log(1 - p_{x_i})],$$

where $const.$ represents the additive terms $\Big( \sum_{j \neq i} \mathbb{E}_{q^{l+1}}[\log \text{Beta}(p_{x_j}; \omega_{x_j})] + \sum_{i=1}^{M} \log \text{Beta}(p_{y_i}; \omega_{y_i}) + \sum_{i=1}^{M} \log \text{Bin}(y_i; p_{y_i}) \Big)$ that are constants w.r.t. $\alpha_{x_i}$ and $\beta_{x_i}$; and the equality can be achieved by replacing the analytical form of $\text{Beta}(p_{x_i} | \alpha_{x_i}, \beta_{x_i})$ in Eq. (2); and $\Gamma(x) = (x-1)!$, $x \in \mathbb{R}$. The above problem formulation is similar to a Maximum Likelihood Estimation (MLE) problem of a standard Beta distribution. Only differences are that the negative constants "$\log p_{x_i}$" and "$\log(1 - p_{x_i})$" are substituted by their expectation on $q^{l+1}$: $\mathbb{E}_{q^{l+1}}[\log p_{x_i}]$ and $\mathbb{E}_{q^{l+1}}[\log(1 - p_{x_i})]$, respectively. Thus, we can directly apply the numerical methods in [4, 7] for the MLE estimation of a Beta distribution to solve Problem (25).

### 5.2 Approximate Expectation Estimation

In this section, we present an efficient approximate expectation estimation algorithm to reduce the computational complexity of $\{\mathbb{E}_{q^{l+1}}[\log p_{x_i}], \mathbb{E}_{q^{l+1}}[\log(1 - p_{x_i})] \mid i = 1, \cdots, N\}$. Because the computation of these expectation terms is impossible and leads to intractable inference, we adopt a commonly used approximation approach: $\mathbf{p}^*_{\mathbf{x},\mathrm{y}}$ and $\mathbf{b}^*_{\mathrm{y}}$ represent the values at the "most probable" setting of $\mathbf{p}_{\mathbf{x},\mathrm{y}}$ and $\mathbf{b}_{\mathrm{y}}$ with the current inferred opinion $\boldsymbol{\omega}^t_{\mathbf{x}}$. The expectation terms $\mathbb{E}_{q^{l+1}}[\log p_{x_i}]$ and $\mathbb{E}_{q^{l+1}}[\log(1 - p_{x_i})]$ can be approximated as $\log \mathbf{p}^*_{x_i}$ and $\log(1 - \mathbf{p}^*_{x_i})$, respectively. We can obtain the "most probable" values $\mathbf{p}^*_{\mathbf{x},\mathrm{y}}$ and $\mathbf{b}^*_{\mathrm{y}}$ by solving the following optimization problem (by replacing

the analytical solution in Eq. (19)):

$$\begin{aligned}\mathbf{p}^*_{\mathbf{x},\mathbf{y}}, \mathbf{b}^*_{\mathbf{y}} &= \arg\min_{\mathbf{p}_{\mathbf{x},\mathbf{y}}, \mathbf{b}_{\mathbf{y}}} -\log q(\mathbf{p}_{\mathbf{x},\mathbf{y}}, \mathbf{b}_{\mathbf{y}}) \\ &= \arg\min_{\mathbf{p}_{\mathbf{x},\mathbf{y}}, \mathbf{b}_{\mathbf{y}}} -\log \text{Prob}(\mathbf{p}_{\mathbf{x},\mathbf{y}}, \mathbf{b}_{\mathbf{y}}|\mathbf{y}) + \sum\nolimits_{k=1}^{K} \rho_k \Big(1 - r_k(\mathbf{p}_{\mathbf{x},\mathbf{y}}, \mathbf{b}_{\mathbf{y}})\Big), \end{aligned} \tag{27}$$

where the parameters $\boldsymbol{\omega}_0, p_0, \boldsymbol{\omega}_{\mathbf{x}}$ and $\boldsymbol{\omega}_{\mathbf{y}}$ are not shown in the Prob$(\cdot)$ function for simplicity. The definition of PSL logic rule $r_k(\mathbf{p}_{\mathbf{x},\mathbf{y}}, \mathbf{b}_{\mathbf{y}}) \in \mathcal{R}$ is given in Eq. (14), and according to the definition of the distance to satisfaction of a given rule $r_k$ in Eq. (11), $r_k(\mathbf{p}_{\mathbf{x},\mathbf{y}}, \mathbf{b}_{\mathbf{y}})$ is defined by :

$$r_k(\mathbf{p}_{\mathbf{x},\mathbf{y}}, \mathbf{b}_{\mathbf{y}}) = \min\Big\{1, \sum_{i \in I^+_{\mathbf{x},k}} p_{x_i} + \sum_{i \in I^+_{\mathbf{y},k}} (p_{y_i} + b_{y_i}) + \sum_{i \in I^-_{\mathbf{x},k}} (1 - p_{x_i}) + \sum_{i \in I^-_{\mathbf{y},k}} (1 - p_{y_i} + b_{y_i})\Big\}, \tag{28}$$

where $I^+_k, I^-_k \subseteq \mathbb{V}$ and $I^+_k \wedge I^-_k = \emptyset$. Let $\mathbf{p}_{\mathbf{x},\mathbf{y},k} \equiv (\mathbf{p}_{\mathbf{x},k}, \mathbf{p}_{\mathbf{y},k})$. Then, we can reformulate the optimization problem in Eq. (27) as follows:

$$\arg\min_{\mathbf{p}_{\mathbf{x},\mathbf{y}}, \mathbf{b}_{\mathbf{y}}} -\log \text{Prob}(\mathbf{p}_{\mathbf{x},\mathbf{y}}, \mathbf{b}_{\mathbf{y}}|\mathbf{y}) + \sum_{k=1}^{K} \rho_k \max\Big\{\ell_k(\mathbf{p}_{\mathbf{x},\mathbf{y},k}, \mathbf{b}_{\mathbf{y},k}), 0\Big\},$$

$$\text{where } \ell_k(\mathbf{p}_{\mathbf{x},\mathbf{y},k}, \mathbf{b}_{\mathbf{y},k}) = 1 - \sum_{i \in I^+_{\mathbf{x},k}} p_{x_i} - \sum_{i \in I^+_{\mathbf{y},k}} (p_{y_i} + b_{y_i}) - \sum_{i \in I^-_{\mathbf{x},k}} (1 - p_{x_i}) - \sum_{i \in I^-_{\mathbf{y},k}} (1 - p_{y_i} + b_{y_i}).$$

In the above objective function, the first term (joint probability of input and output variables) is a convex function and the second term (logical rule constraints) is a hinge-loss function that is convex but non-smooth. Thus, the optimization problem in Eq. (27) is a non-smooth convex optimization problem.

To tackle the various of convex/non-smooth convex optimization problems, many state-of-the-art and off-the-shelf methods are proposed, such as gradient descent (GD) based methods and interior-point methods (IPMs). But these methods are inefficient to solve the optimization problem in Eq. (27) with a large number of variables $N + 2M$, where $N$ is the total number of target unknown variables and $M$ is the total number of given variables. In our paper, we propose a robust and efficient algorithm, uses consensus optimization via adopting Alternating Direction Method of Multipliers (ADMM) [27], to solve this problem. The adopted ADMM based consensus maximum-a-posteriori (MAP) inference process has the following three main steps: (1) *forming and initialing* local copies of the variables in each PSL logic rule by constraining the local copies to be equal to the original global variables; (2) *decomposing* the problem into independent sub-problems; and (3) *block-wise updating* until converging to a consensus on the optimum. Let $\hat{\mathbf{p}}_{\mathbf{x},k}$, $\hat{\mathbf{p}}_{\mathbf{y},k}$, and $\hat{\mathbf{b}}_{\mathbf{y},k}$ be the local copies of the global variables $\mathbf{p}_{\mathbf{x},k}$, $\mathbf{p}_{\mathbf{y},k}$, and $\mathbf{b}_{\mathbf{y},k}$ in the PSL logic rule $(r_k, \rho_k) \in \mathcal{R}$, separately. Finally, our main problem based on the ADMM framework is formulated as follows:

$$\begin{aligned}\min_{\{\hat{\mathbf{p}}_{\mathbf{x},\mathbf{y},k}, \hat{\mathbf{b}}_{\mathbf{y},k}\}_{k=1}^K, \mathbf{p}_{\mathbf{x},\mathbf{y}}, \mathbf{b}_{\mathbf{y}}} &\Big\{ -\log \text{Prob}(\mathbf{p}_{\mathbf{x},\mathbf{y}}, \mathbf{b}_{\mathbf{y}}|\mathbf{y}) + \sum_{k=1}^{K} \rho_k \max\{\ell_k(\hat{\mathbf{p}}_{\mathbf{x},\mathbf{y},k}, \hat{\mathbf{b}}_{\mathbf{y},k}), 0\}\Big\}, \\ & s.t.\ \hat{\mathbf{p}}_{\mathbf{x},\mathbf{y},k} = \mathbf{p}_{\mathbf{x},\mathbf{y},k},\ \hat{\mathbf{b}}_{\mathbf{y},k} = \mathbf{b}_{\mathbf{y},k}\ , k = [1:K]. \end{aligned} \tag{29}$$

Then these constraints are transformed into and augmented Lagrangian with penalty $\kappa$ and Lagrange multipliers $\lambda$ and $\gamma$:

$$\begin{aligned}&\mathcal{L}(\{\hat{\mathbf{p}}_{\mathbf{x},\mathbf{y},k}, \hat{\mathbf{b}}_{\mathbf{y},k}, \boldsymbol{\lambda}_{\mathbf{x},\mathbf{y},k}, \boldsymbol{\gamma}_{\mathbf{y},k}\}_{k=1}^K, \mathbf{p}_{\mathbf{x},\mathbf{y}}, \mathbf{b}_{\mathbf{y}}) = -\log \text{Prob}(\mathbf{p}_{\mathbf{x},\mathbf{y}}, \mathbf{b}_{\mathbf{y}}|\mathbf{y}) \\ &+ \sum_{k=1}^{K} \Big(\rho_k \max\{\ell_k(\hat{\mathbf{p}}_{\mathbf{x},\mathbf{y},k}, \hat{\mathbf{b}}_{\mathbf{y},k}), 0\} + \frac{1}{2\kappa}\|\hat{\mathbf{p}}_{\mathbf{x},\mathbf{y},k} - \mathbf{p}_{\mathbf{x},\mathbf{y},k} + \kappa\boldsymbol{\lambda}_{\mathbf{x},\mathbf{y},k}\|_2^2 + \frac{1}{2\kappa}\|\hat{\mathbf{b}}_{\mathbf{y},k} - \mathbf{b}_{\mathbf{y},k} + \kappa\boldsymbol{\gamma}_{\mathbf{y},k}\|_2^2\Big), \end{aligned} \tag{30}$$

where $\kappa > 0$ represents the step-size (penalty) of ADMM. ADMM aims to find a saddle point of $\mathcal{L}(\mathbf{p}_{\mathrm{x,y}}, \hat{\mathbf{p}}_{\mathrm{x,y}}, \boldsymbol{\lambda}_{\mathrm{x,y}}, \boldsymbol{\gamma}_{\mathrm{y}})$ via updating the four blocks of variables at each iteration $t$:

For $k = [1 : K]$ :

$$\boldsymbol{\lambda}^{t}_{\mathrm{x,y},k} = \boldsymbol{\lambda}^{t-1}_{\mathrm{x,y},k} + \frac{1}{\kappa}(\hat{\mathbf{p}}^{t-1}_{\mathrm{x,y},k} - \mathbf{p}^{t-1}_{\mathrm{x,y},k}) \tag{31}$$

$$\boldsymbol{\gamma}^{t}_{\mathrm{y},k} = \boldsymbol{\gamma}^{t-1}_{\mathrm{y},k} + \frac{1}{\kappa}(\hat{\mathbf{b}}^{t-1}_{\mathrm{y},k} - \mathbf{b}^{t-1}_{\mathrm{y},k}) \tag{32}$$

$$\begin{aligned}\hat{\mathbf{p}}^{t}_{\mathrm{x,y},k}, \hat{\mathbf{b}}^{t}_{\mathrm{y},k} = \arg \min_{\hat{\mathbf{p}}_{\mathrm{x,y},k}, \hat{\mathbf{b}}_{\mathrm{y},k}} \rho_k \max\left\{\ell_k(\hat{\mathbf{p}}^{t-1}_{\mathrm{x,y},k}, \hat{\mathbf{b}}^{t-1}_{\mathrm{y},k}), 0\right\} \\ + \frac{1}{2\kappa}\|\hat{\mathbf{p}}^{t-1}_{\mathrm{x,y},k} - \mathbf{p}^{t-1}_{\mathrm{x,y},k} th + \kappa\boldsymbol{\lambda}^{t}_{\mathrm{x,y},k}\|_2^2 + \frac{1}{2\kappa}\|\hat{\mathbf{b}}^{t-1}_{\mathrm{y},k} - \mathbf{b}^{t-1}_{\mathrm{y},k} + \kappa\boldsymbol{\gamma}^{t}_{\mathrm{y},k}\|_2^2\end{aligned} \tag{33}$$

$$\mathbf{p}^{t}_{\mathrm{x,y}}, \mathbf{b}^{t}_{\mathrm{y}} = \arg \min_{\mathbf{p}_{\mathrm{x,y}}, \mathbf{b}_{\mathrm{y}}} \mathcal{L}(\hat{\mathbf{p}}^{t}_{\mathrm{x,y}}, \mathbf{p}^{t-1}_{\mathrm{x,y}}, \hat{\mathbf{b}}^{t}_{\mathrm{y}}, \mathbf{b}^{t-1}_{\mathrm{y}}, \boldsymbol{\lambda}^{t}_{\mathrm{x,y}}, \boldsymbol{\gamma}^{t}_{\mathrm{y}}) \tag{34}$$

where the rule indices are $k = 1, \cdots, K$. The block-wise ADMM updates make sure that $\mathbf{p}_{\mathrm{x,y}}$ and $\mathbf{b}_{\mathrm{y}}$ converge to the global optimums $\mathbf{p}^{\star}_{\mathrm{x,y}}$ and $\mathbf{b}^{\star}_{\mathrm{y}}$, assuming that there exists a feasible assignment to $\mathbf{p}_{\mathrm{x,y}}$ and $\mathbf{b}_{\mathrm{y}}$. Updating of the Lagrange multipliers $\boldsymbol{\lambda}_{\mathbf{x},\mathrm{y},k}, \boldsymbol{\gamma}_{\mathrm{y},k}$ is a basic step in the gradient direction in Eq. (31) and (32). The local variables related problem in Eq. (33) can be efficiently solved via a customized algorithm proposed in [5]. After solving the sub-problems in Eq. (31) ~ (33), we can treat the local variables $\hat{\mathbf{p}}_{\mathbf{x},k}$, $\hat{\mathbf{p}}_{\mathrm{y},k}$, and $\hat{\mathbf{b}}_{\mathrm{y},k}$ as constants and substitute the local variables in Eq. (30). We can use the same technique in Eq. (25), via grouping the constants and variable terms. The problem related to global variables in Eq. (34) has an analytical solution ensuring that the gradient of the objective function is 0.

Fig. 4 shows the main architecture of the consensus MAP inference algorithm. We decompose the problem into independent subproblems and optimize each potential $\ell_k$ functions independently (can be done in parallel). We let each subproblem to vote to the optimal solution until all the independent distances to satisfaction. Our problem is a convex problem; so it guarantees to converge global optimum. Auxiliary variables ensure consensus reached across subproblems.

### 5.3 Complexity Analysis

Algorithm 1 is pseudo-code of our purposed CSL$^+$ and it summarizes the key steps of CSL$^+$. CSL$^+$ has two main loops: The **outer loop** (Lines 3 to 19) is related to the modified EM inference. The modified E′−Step is implemented in Line 4. The M−Step is implemented by Lines 5 through 17. Especially, Lines 5 through 15 (including the **inner loop**) show the ADMM steps for estimating the "most probable" values $\mathbf{p}^{*}_{\mathrm{x,y}}$ and $\mathbf{b}^{*}_{\mathrm{y}}$ by solving the optimization problem in Eq. (27). The calculated

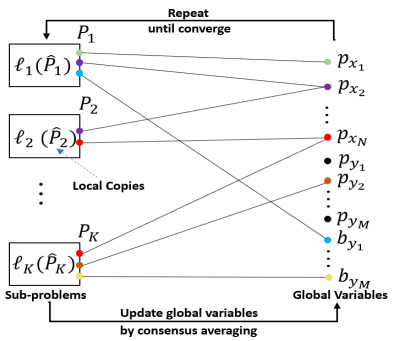



Fig. 4. An illustration of the consensus algorithm.

**Algorithm 1:** CSL$^+$ on the Opinion Prediction

**Input:** $\boldsymbol{\omega}_{\mathbf{y}}, \mathbf{y}, \mathcal{R}, \omega_0, p_0$
**Output:** $\boldsymbol{\omega}_{\mathbf{x}}$
1 Initialize $l = 1$ , $\boldsymbol{\omega}_{\mathbf{x}}^{l}$;
2 **repeat**
3  Update $q^{l}(\mathbf{p}_{\mathbf{x},\mathbf{y}}, \mathbf{b}_{\mathbf{y}})$ via Eq. (19);
4  Initialize $t = 1$ , $\mathbf{p}_{\mathbf{x},\mathbf{y}}, \mathbf{b}_{\mathbf{y}}$;
5  Initialize $\hat{\mathbf{p}}^{t}_{\mathbf{x},\mathbf{y},k}$ and $\hat{\mathbf{b}}^{t}_{\mathbf{y},k}$ as copies of the probability variables, $\mathbf{p}^{t}_{\mathbf{x},\mathbf{y},k}$ and $\mathbf{b}^{t}_{\mathbf{y},k}$, that occur in the $k$-th rule in $\mathcal{R}$, respectively, $k = [1 : K]$;
6  Initialize Lagrange multipliers $\boldsymbol{\lambda}_{\mathbf{x},k}$, $\boldsymbol{\lambda}_{\mathbf{y},k}$ and $\boldsymbol{\gamma}_{\mathbf{y},k}$ corresponding to variable copies $\hat{\mathbf{p}}_{\mathbf{x},k}$ and $\hat{\mathbf{p}}_{\mathbf{y},k}$, respectively, $k = [1 : K]$;
7  **repeat**
8  $t = t + 1$
9  Update Lagrange multiplier $\boldsymbol{\lambda}^{t}_{\mathbf{x},\mathbf{y},k}$ via Eq. (31), $k = [1 : K]$;
10  Update Lagrange multiplier $\boldsymbol{\gamma}^{t}_{\mathbf{x},\mathbf{y},k}$ via Eq. (32), $k = [1 : K]$;
11  Update local copies $\hat{\mathbf{p}}^{t}_{\mathbf{x},\mathbf{y},k}$ and $\hat{\mathbf{b}}^{t}_{\mathbf{y},k}$ via solving the problem in Eq. (33), $k = [1 : K]$;
12  Update global variables $\mathbf{p}^{t}_{\mathbf{x},\mathbf{y}}, \mathbf{b}^{t}_{\mathbf{y}}$ via solving the problem in Eq. (34);
13  **until** *convergence*
14  $l = l + 1$;
15  **for** $i = [1 : N]$ **do**
16  Update opinion $\omega^{l}_{x_i}$ via solving the problem in Eq. (25);
17 **until** *convergence*
18 **return** $\boldsymbol{\omega}_{\mathbf{x}}^{l}$

most probable values $\mathbf{p}^{*}_{\mathbf{x},\mathbf{y}}$ are used to approximate $\mathbb{E}_{q^{l+1}}[\log p_{x_i}]$ and $\mathbb{E}_{q^{l+1}}[\log(1 - p_{x_i})$ as $\log p^{*}_{x_i}$ and $\log(1 - p^{*}_{x_i})$, separately, which are then applied to implement the M−Step in Lines 17-18. The computational complexity and time consumption of Algorithm 1 are shown in Lines 13 , 14, and 18.

Line 13 requires to solve $K$ sub-problems in Eq. (33) that can be solved by applying the algorithm in [5] with $O(KP)$, where $P$ is the maximum number of variables occurred in the PSL logic rules, $\mathcal{R}$. Line 14 requires to solve the optimization of global variables in Eq. (34), and needs to update all global variables and the analytical solutions can be obtained in $O(N + 2M)$. Lines 17-18 require to solve the optimization problem in Eq. (25) which is similar to the MLE problem of a Beta distribution and can be solved applying the method of moments [4] in $O(1)$. $T_1$ and $T_2$ represents the numbers of iterations of the outer and inner loops, separately. Summing up above, the overall computation complexity of CSL$^+$ (Algorithm 1) is $O(T_1 \cdot T_2 \cdot (K + 2M + N + KP))$. As $K$ sub-problems in Line 13 can be calculated via parallel processing, if we have enough processors with count $C$ such that $O(K/C) \approx O(1)$. Finally, the computational complexity is $O(T_1 \cdot T_2 \cdot (K + 2M + N + P))$, which is linear w.r.t. to $M$, $N$, and $K$. This analysis indicates that our proposed CSL$^+$ is scalable to large-scale network data. Compared to CSL, CSL$^+$ via doubling the variable size and not increasing the time complexity too much, achieves better performance on different scale of datasets. We further prove the scalability via extensive empirical evaluations in Section 6.

## 6 EXPERIMENTAL RESULTS & ANALYSIS

### 6.1 Datasets and Experimental Set-Up

In the experiments, we validate our proposed CSL$^+$ method on four semi-synthetic dataset and two real-world datasets on three different tasks under the conflicting evidence scenarios: **Epinions**

Table 3. Dataset statistics

| Dataset | # Nodes | # Edges | # Weeks | # Snapshots in total (hours) |
|---|---|---|---|---|
| **Philadelphia (PA)** | 603 | 708 | 43 | 3440 |
| **Washington, D.C.** | 1,383 | 1,878 | 43 | 3440 |
| **Facebook** | 8,078 | 372,936 | - | - |
| **Epinions** | 47,676 | 477,468 | - | - |
| **Enron** | 67,392 | 743,244 | - | - |
| **Slashdot** | 164,336 | 2,018,920 | - | - |

(semi-synthetic) datasets on the trust inference task, **DC and PA road traffic datasets** on the congestion inference task and **Facebook, Enron and Slashdot social network datasets** on the Sybils attack inference task (also semi-synthetic). Dataset statistics are summarized in Table 3.

*6.1.1* ***Epinions***. *Epinions* [1] is who-trust-who network data that were crawled in 2003 [22]. The used Epinions dataset is a directed network, and has 47,676 users (i.e., nodes) and 467,468 trust relations (i.e., edges). As there are lack of ground truth trust information for Epinions dataset, we infer the actual trust relations between the users according to the trust inference method applied in [17, 31] by proceeding the following main steps:
**Initialization Step**: randomly select 20% of the edges (relationships) and set the trust of the edges to '1's indicates $\text{User}_i$ trusts $\text{User}_j$ (but not necessarily $\text{User}_j$ trusts $\text{User}_i$, and '0' indicates distrust) where $\text{User}_i$ and $\text{User}_j$ are users in the given Epinions network.
**Exploration Step**: 10,000 exploration steps are taken to update trust relationships based on the following trust PSL logic rule:

$$\text{Trusts}(A, B) \wedge \text{Trusts}(B, C) \rightarrow \text{Trusts}(A, C). \tag{35}$$

To generate synthetic trust observation, we select an edge $e_i$ randomly, identify the rule instances (neighborhood edges of $e_i$) related to $e_i$, and generate an observation for $e_i$ (trust '1' or distrust '0') based on the probability of the rule instances. Through the above steps, we obtain 1$st$ realization of trust relationships on the edges in the Epinions network. Now each edge has a single trust observation. We obtain the 2$nd$ realization based on the previous one by randomly choosing 5% of the edges and flip their observations from '1' to '0' or '0' to '1', and then iteratively repeating 10,000 exploration steps to make the generated observations consistent with the above trust rule. Following this process, we obtain $3rd, \cdots$, and $T$-th realizations.
**Performance Evaluation Step**: After generating $T$ realizations, then each edge has $T$ trust relationship observations in total to estimate the opinion of this link. To validate the performances of proposed $\text{CSL}^+$ and the baselines on the networks of different sizes, we randomly generate induced sub-networks size with $N_G \in \{1000, 5000, 10000\}$ from the original Epinions trust network. From all the edges we randomly select the *testing edges* with the percentages (or test ratios (TR):= $\frac{N}{N+M} \times 100\%$) $\in \{10\%, 20\%, 30\%, 40\%, 50\%\}$, and from the rest of edges we select the *conflict edges* randomly with percentage (or conflict ratios (CR)) $\in \{0\%, 10\%, 20\%, 30\%, 40\%\}$; and we flip a half of the observations of the conflict edges. The opinions of testing edges (target edges) are predicted based on the given opinions of the other edges (maybe have conflicting evidences) which are *training edges*.

*6.1.2* ***Road traffic datasets***. We crawled live traffic data from June 1, 2013 to March 31, 2014 across two major cities from INRIX [2] website, Washington D.C. and Philadelphia (PA), as given the summarization in Table 3. The original raw INRIX data provides traffic speed and reference

[1] http://www.trustlet.org/downloaded_epinions.html

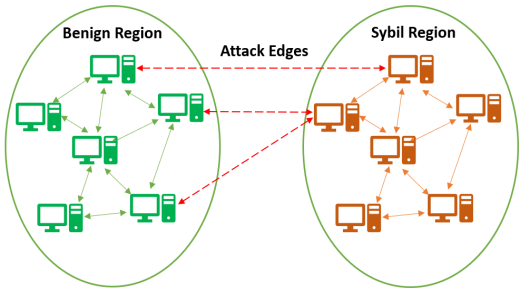



Fig. 5. Illustration of Benign and Sybil region, and the attack edges.

speed information for each road section (link) per hour interval. The reference speed of each road link indicates the "non-congested free flow speed" [3], and is set based on the 60-th percentile of the recorded speed for all time periods over several years, where the reference speed work as a golden threshold labeling into two traffic states, *congested* vs. *non-congested*. We collect for each of these two cities 43 weeks of traffic information in total. An hour is denoted by a tuple $(h, d, w)$, i.e., (hour,day,week), where the hour ($h \in \{8, \cdots, 22\}$), the day (weekday) ($d \in \{1, 2, 3, 4, 5\}$), and the week ($w \in \{1, 2, \cdots, 43\}$): $(h, d, w)$. We only consider weekdays from Monday ($d = 1$) to Friday ($d = 5$) and hours from 8 AM ($h = 8$) to 9 PM ($h = 21$).

**Ground truth opinions of the training and testing edges in the traffic datasets**. For DC and PA traffic datasets, the opinion of a specific (training/testing) road link $i$ at an hour $(h, d, w)$ is calculated based on the historical observations of the same hour in previous $T$ weeks $\{x_{i,h,d,w}, x_{i,h,d,w-1}, ..., x_{i,h,d,w-T+1}\}$ as the evidence, where $x_{i,h,d,w}$ refers to the congestion observation ('0' or '1') of the link $i$ at hour $(h, d, w)$ and $T$ indicates a predefined time window size. Especially, the belief, disbelief, and uncertainty mass variables $b_{x_i}$, $d_{x_i}$, and $u_{x_i}$ of a specific road link $i$ are estimated as:

$$\begin{aligned} b_{x_i} &= \sum_{t=0}^{T-1} x_{i,h,d,w-t}/(T+W) \\ d_{x_i} &= \left(T - \sum_{t=0}^{T-1} x_{i,h,d,w-t}\right)/(T+W) \\ u_{x_i} &= W/(T+W), \end{aligned} \tag{36}$$

where we set the amount of uncertain evidence $W = 2$ and the specific base rate (i.e., prior knowledge) $a_{x_i} = 0.5$ for all road links. For the semi-synthetic dataset, the opinion of each training / testing edge is estimated based on the $T$ observations similar to the above. Similar to the Epinions, we try the same test ratios (TR) $\in \{10\%, 20\%, 30\%, 40\%, 50\%\}$, and conflict ratios (CR) $\in \{0\%, 10\%, 20\%, 30\%, 40\%\}$.

*6.1.3* ***Social networks dataset with synthesized Sybils attack***. We utilize three social networks used in [29, 30], i.e., Facebook, Enron, and Slashdot (see Table 3 for basic statistics) to represent vary application scenarios. These datasets are publicly available at SNAP [2]. (1) **Facebook:** a node represents a user on Facebook and an edge between two nodes indicates they are friends. (2) **Enron:** an email address is denoted by a node, and an edge between two nodes implies at least one email was exchanged between these two. (3) **Slashdot:** a technology-related news website where a node represents a user and an edge between two users indicates a friend relationship. We also follow the method in [29, 30] to synthesize the Sybil attack in different scenarios. That is, in the above social networks, a single user (i.e., a node in a network) can pretend to have multiple

[2]SNAP: http://snap.stanford.edu/data/index.html

identities, performing Sybil attack, with its unknown identity. Our main goal is to infer the identity of the unknown users performing Sybil attack. We set a real social network graph as the Benign region while synthesizing the Sybil region and between the Benign and Sybil regions uniformly at random add attack edges (see Fig. 5). For each social network graph, we use it as the Benign region and replicate it as a Sybil region. We labeled the observation of the nodes in the Sybil region to "1" at time stamp $t = 1$, "0" to the nodes in the Benign region. In Exploration step, we duplicate the observations of each node and process $T$ realizations, and then we randomly swap observations of 1% of nodes each realization. We randomly select % of nodes (or test ratio (TR)) $\in \{10\%, 20\%, 30\%, 40\%, 50\%\}$ (where we randomly select the same amount of nodes from the Benign and Sybil region) as the test nodes. Except for the test nodes, we select from the remaining nodes with the percentage (or conflict ratio (CR)) $\in \{0\%, 10\%, 20\%, 30\%, 40\%\}$ as the conflict nodes, for each conflict node we flip (set '0' to '1', or '1' to '0') the observation of the half of the realization. We also try different numbers of attacking edges between the Benign region and Sybil region, $\{1000, 5000, 10000, 15000, 20000\}$ which make our conflicting inference and prediction task more challenging.

*6.1.4* ***Parameter settings.*** In our experiment, the main parameters of our datasets are $CR$ (Conflict Ratio), $TR$ (the percentage of testing edges/nodes or Test Ratio) and $T$ (time window size). The values of $CR$ are set as $\{0\%, 10\%, 20\%, 30\%, 40\%\}$ and $TR$ are set as $\{10\%, 20\%, 30\%, 40\%, 50\%\}$. The values of $T$ are set as $T \in \{8, 9, 10, 11\}$. The corresponding uncertainty mass values, $u$ of these $T$ values can be obtained based on Eq. (36), $\{20\%, 18\%, 16\%, 15\%\}$. Epinions dataset has three parameters: The parameter $N_G$ (network size) $\in \{1000, 5000, 10000\}$, the positive trust ratio ( the percentage of randomly selected edges) is 20% in the initialization phase and the percentage of edges in the exploration phase (5%) whose observations were flipped between '0' and '1.' In the experimental results, the patterns observed are consistent or similar with the results of other parameter settings. The numbers of attack edges between the Benign and Sybil region are set as $\{1000, 5000, 10000, 15000, 20000\}$. We set $\omega_0 = (1, 1)$ and $p_0 = 0.5$, indicating complete uncertainty. $p_0$ is a hyperparmeter that represents the expected percentage of conflicting observed opinions. If $p_0$ matches the conflict ratio (CR), the CSL+ would exhibit the best performance. However in our experiments, any prior information about the CR is assumed unknown, leading to setting $p_0 = 0.5$.

*6.1.5* ***Performance metrics.*** The uncertainty mass $u_{x_i}$ of each edge (training and testing edges) can be calculated based on Eq. (36). For a predefined the window size $T$, $u_{x_i}$ is a known and constant value, without the actual observations. Because of $u_{x_i}$ is fixed, our empirical analysis of the experiments on all the datasets is focused on the comparison between the proposed $\mathsf{CSL}^+$ and other baselines based on the two metrics: (1) ***Expected truth probability MAE*** (denoted as Probability MAE or EP-MAE) and running time complexity (sec.). Based on the definition of expected belief probability $\mathbb{E}_b = b_x + au_x$ and Table 2, EP-MAE is defined as:

$$\text{EP-MAE}(\boldsymbol{\omega}_{\mathbf{x}}) = \frac{1}{N}\sum_{i=1}^{N}\left|\frac{\alpha_{x_i}}{\alpha_{x_i}+\beta_{x_i}} - \frac{\alpha^{\star}_{x_i}}{\alpha^{\star}_{x_i}+\beta^{\star}_{x_i}}\right| \tag{37}$$

where $\omega_{x_i} = (\alpha_{x_i}, \beta_{x_i})$ and $\omega^{\star}_{x_i} = (\alpha^{\star}_{x_i}, \beta^{\star}_{x_i})$ represent the predicted and true opinions of a target variable $x_i$, separately, and $\frac{\alpha_{x_i}}{\alpha_{x_i}+\beta_{x_i}}$ refers to the predicted expected truth probability (or the expected belief) of the opinion $\omega_{x_i}$. Expected probability MAE is calculated as the mean absolute difference between the estimated expected belief and the true expected belief on all testing links; and (2) ***Average running time*** comparison between $\mathsf{CSL}^+$ and other methods on the real-world dataset experiments.

*6.1.6* ***Baseline methods.*** We compare the proposed $\mathsf{CSL}^+$ (Section 5) with the comparable counterpart methods, including SL [13] (Section 3.1), PSL [5] (Section 3.2), CSL [31] and the deep

learning based GCN-VAE-opinion (for short GCN-VAE) method [34]. GCN-VAE is a DL-based opinion inference model while node-level opinions are still formalized based on SL. Followed the authors recommendation, we use the returned result that give the minimum belief and uncertainty MAE among all epochs as the final results of GCN-VAE. As PSL is proposed to predict the truth probability of the testing edges, but not their subjective opinions. We extend PSL[3] as follows: suppose the truth probability of a testing edge $x_i$ predicted by PSL is denoted as $p_{x_i}$. For a given uncertainty mass $u$, an subjective opinion of $w_{x_i}$ based on the probability $p_{x_i}$ can be calculated as:

$$\omega_{x_i} = (p_{x_i} \cdot (1-u), (1-p_{x_i})(1-u), u) \tag{38}$$

*6.1.7* ***Parameter tuning.*** SL has only one parameter that is the maximum length of its independent paths. We try different settings $\{3, 4, \cdots, 20\}$ where for each dataset, we keep the settings returning the best result. Thus, we set the maximum length to 10 for Epinions, PA and DC, set 3 to Facebook, Enron and Slashdot datasets. For GCN-VAE [34], we used the recommended settings from original paper in our experiments: $\lambda = 0.01$ (the trade-off parameter), $\eta = 0.001$ (the learning rate), $K = 16$ (the mini-batch size), and $P = 16$ (the dimensionality of the latent encoded vectors), and dropout rate = 0.1. CSL$^+$, CSL and PSL require additional input as ***logic rules*** for reasoning. In our experiments, we used the trust rules from [32], for Epinions dataset:

$$\begin{aligned}
\textsc{Trusts}(A,B) \wedge \textsc{Trusts}(B,C) &\rightarrow \textsc{Trusts}(A,C)\\
\neg\textsc{Trusts}(A,B) \wedge \textsc{Trusts}(B,C) &\rightarrow \neg\textsc{Trusts}(A,C)\\
\textsc{Trusts}(A,B) \wedge \neg\textsc{Trusts}(B,C) &\rightarrow \neg\textsc{Trusts}(A,C),
\end{aligned} \tag{39}$$

where A, B and C are users, Trust $(\cdot,\cdot)$ indicates their trust relationship. The logical rules from [29], for Sybil attack dataset:

$$\begin{aligned}
\textsc{Homogeneous}(U,V) \wedge \textsc{Benign}(U) &\rightarrow \textsc{Benign}(V)\\
\textsc{Homogeneous}(U,V) \wedge \neg\textsc{Benign}(U) &\rightarrow \neg\textsc{Benign}(V)\\
\textsc{Heterogeneous}(U,V) \wedge \textsc{Benign}(U) &\rightarrow \neg\textsc{Benign}(V)\\
\textsc{Heterogeneous}(U,V) \wedge \neg\textsc{Benign}(U) &\rightarrow \textsc{Benign}(V),
\end{aligned} \tag{40}$$

where these rules indicate that two linked network entities share the same label with a high probability. A single rule from [33], for the traffic datasets:

$$\textsc{Neighbor}(E_1,E_2) \wedge \textsc{Congested}(E_1) \rightarrow \textsc{Congested}(E_2), \tag{41}$$

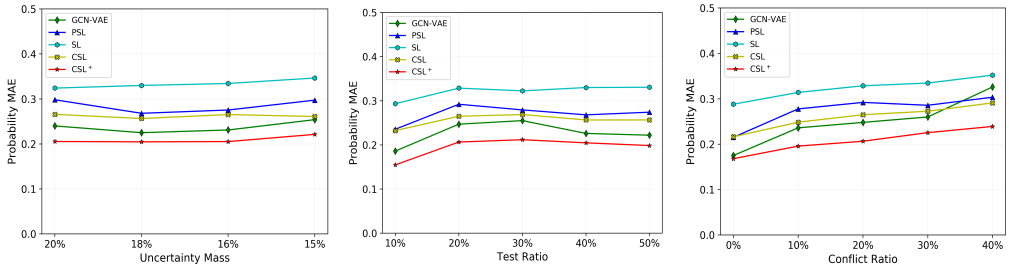
where $E_1$ is a congested road section and $E_2$ is its upper stream neighbor, then $E_2$ is likely congested. As rule weights are applied to model the relative importance of different rules. In our experiments, we set the rule weights to 1.0, where we see all the rules equally important.

## 6.2 Experimental Results on Semi-Synthetic Dataset

*6.2.1* ***Parameter sensitivity study.*** In this section, we vary Uncertainty Mass ($u$), Test Ratio ($TR$), and Conflict Ratio ($CR$) to investigate their impact on the performance of CSL$^+$ and its counterparts. To evaluate how changes to the parameterization of CSL$^+$ and the synthetic dataset affects its performance on the uncertainty learning, we conducted experiments on the Epinions semi-synthetic dataset and compared with the baseline methods. Fig. 6 demonstrates the performance of our CSL$^+$ method and the four baseline methods on the probability MAE of the semi-synthetic dataset based on Epinions.

Fig. 6 (a) shows that the performance of CSL$^+$ exceeds all the baselines on the truth probability MAE with respect to different uncertainty masses, $u$. With fixed conflict ratios and test ratios, on

[3]PSL Code:https://github.com/linqs/psl-examples




(a) Effect of $u$ under $TR = 40\%$ and $CR = 20\%$. (b) Effect of $TR$ under $CR = 20\%$ and $u = 18\%$. (c) Effect of $CR$ under $TR = 20\%$ and $u = 18\%$.

Fig. 6. Probability MAE results on the semi-synthetic network based on Epinions dataset ($N_G = 5000$, Test Ratio (TR), Conflict Ratio (CR), and Uncertainty Mass ($u$))

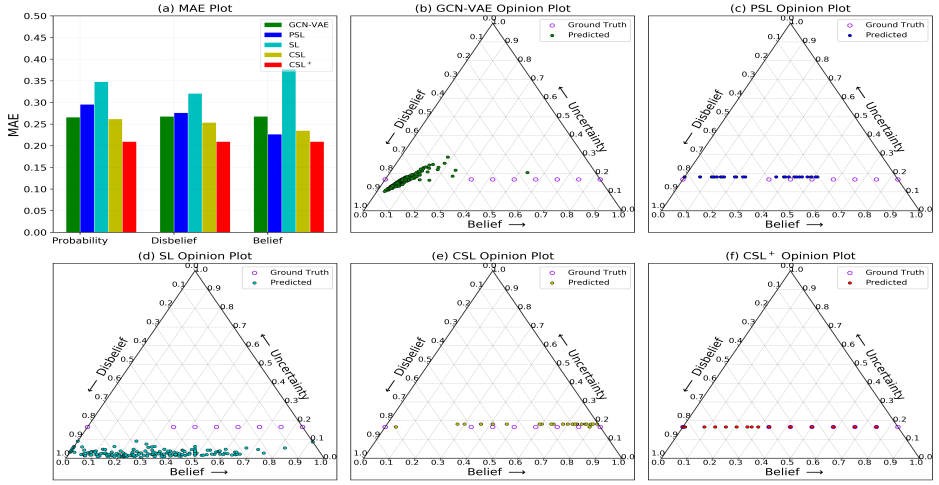



Fig. 7. Uncertainty probability, Belief, Disbelief and Uncertainty MAE and the corresponding simplex plot of Epinions dataset with $N = 5000$ ,$TR = 10\%, CR = 20\%$

varying uncertainty masses $u$, CSL$^+$ shows significantly better performance among the baselines. Compared with the best baseline GCN-VAE, CSL$^+$ decreases the probability MAE 13% $\sim$ 15%.

Fig. 6 (b) demonstrates the sensitivity of testing (or training) ratio on the probability MAE of all the baseline methods. Apparently, CSL$^+$ achieves best performance among all the baseline methods while GCN-VAE demonstrates the second best performance in terms of probability MAE. More or less, all the baselines and CSL$^+$ show sensitivity with respect to the test ratio (i.e., a smaller test ratio implies more training data used while a large test ratio implies to less training data used). Compared with the best baseline GCN-VAE, CSL$^+$ decreases the probability MAE 10% $\sim$ 17%.

Fig. 6 (c) shows the effect of conflicting evidence ratios on the probability MAE of all methods. All the methods show higher sensitivity with respect to the ratio of conflicting evidence; when the conflicting evidence ratio increases, the probability MAE also increases. It is obvious that CSL$^+$ outperforms among all the methods. Compared with the other baselines, CSL$^+$ decreases the probability MAE up to 26%.

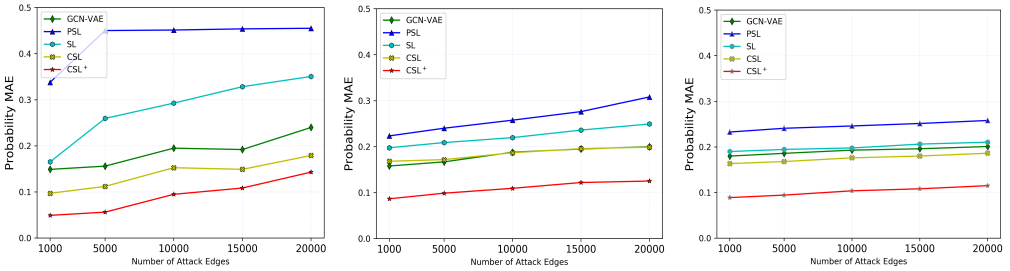
Overall, for all these different settings of parameters and the variety of datasets with conflicting evidence, CSL$^+$ outperforms all the other baseline methods even with high uncertainty and conflicting evidence.

*6.2.2* ***Opinion prediction.*** In Fig. 7, we show the simplex opinion plots, where the binomial opinion point is represented by a subjective opinion tuples (belief, disbelief, uncertainty). We visualize the ground truth and the inferred (or predicted) opinion points on the simplex plots to study the opinion prediction performance of CSL$^+$ and other counterparts on the Epinions testing data. Fig. 7 (a) shows the expected belief probability, belief, disbelief, and uncertainty MAE of the predicted opinions with conflicting evidence. Fig. 7 (b)~(f) are the corresponding opinion simplex plots of the results showed in Fig. 7 (a) for CSL$^+$ and the baseline methods, respectively.

From Fig. 7 (a), we can observe that CSL$^+$ gives the best performance separately on the prediction of the truth probability, belief and disbelief MAE compared to other baselines. In the simplex plots, the bottom, left and right axes are corresponding to the belief, disbelief, and uncertainty, respectively. In Fig. 7 (b)~(f), the purple points represent the ground truth opinion points, and the other points represent the predicted opinion points by CSL$^+$ and the baselines. Most of the predicted opinions of CSL$^+$ are more likely overlapped or close to the ground truth opinions (see Fig. 7 (f)). The predicted opinions of GCN-VAE, PSL and CSL are close to the ground truth opinions but do not overlap each other (see Fig. 7 (b), (c) and (e)). The predicted opinion results of SL are more scattered (see Fig. 7 (d)). These visualization plots demonstrate the low opinion prediction error of CSL$^+$ on opinion prediction against the baselines.

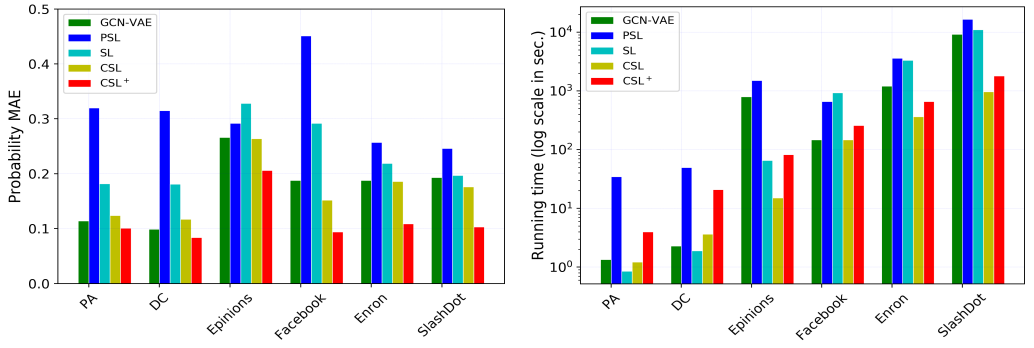
### 6.3 Experimental Results on Real-World Datasets

*6.3.1* ***Parameter sensitivity study.*** In this section, we examine the effect of varying the number of Sybil attack edges on the performance of CSL$^+$ and its counterparts while we fixed the other parameters. Fig. 8 (a) is one of the results on Facebook dataset and it shows the probability MAE of CSL$^+$ and the compared methods on the real-world Facebook dataset. And in this experiment the number of attack edges increasing from 1,000 to 20,000 while $TR = 20\%, CR = 30\%$ and $u = 16\%$. We can observe from the results that CSL$^+$ and the other compared methods achieve low truth probability MAE when a social network has strong homophily, i.e., the number of attack edges is small. One of the main reason is the Benign and Sybil users have fewer ties and are totally separated, and it is easy to detect them. When the conflicting evidence ratio, test ratio and uncertainty mass are fixed, if we increase the number of attack edges, the probability MAE of CSL$^+$ and the baseline methods will also increase. However, CSL$^+$ outperforms (decreases the probability MAE up to 50%) among all the others for varying number of attacking edges. Fig. 8 (b) and (c) shows the experiment results of Enron and Slashdot datasets,respectively. On the other parameter settings, we observe the same patterns.




(a) Facebook (b) Enron (c) Slashdot

Fig. 8. Probability MAE of compared methods as the number of attack edges becomes large,$TR = 20\%, CR = 30\%$ and $u = 16\%$




(a) Probability MAE (b) Running time log scale (sec.)

Fig. 9. Probability MAE and Running time log scale (sec.) for real-world datasets, $TR = 20\%, CR = 30\%$ and $u = 16\%$ (for DC and PA, $u = 4\%$)

*6.3.2* ***Experimental Results.*** We conduct experiments on four semi-synthetic dataset with $N_G = 5,000$ based on the Epinions and other two real world datasets, such as two traffic datasets for PA and DC cities and three social network datasets for Facebook, Enron and Slashdot, with different parameter settings. We validate the **conflicting evidence inference** performance of the proposed CSL$^+$ method on the following tasks: (1) ***Trust inference on Epinions network:*** learning of unknown trust relationships between users from the given trust relationship opinions with highly conflicting evidence; (2) ***Congestion inference on the road traffic networks:*** prediction of the traffic congestion status of unknown road sections (testing links) by using the given opinions of other road sections with conflicting evidence; and (3) ***Sybil attack detection on the social networks:*** detection of the Sybil and Benign network nodes from the given known node information with conflicting evidence.

We observe similar performance patterns from all the datasets experiments with different parameter settings. Fig. 9 (a) is one of the result plot where $TR = 20\%, CR = 30\%$ for all six datasets, $u = 4\%$ for DC and Philly and $u = 16\%$ for other datasets . For Facebook, Enron and Slashdot datasets the number of attack edges is 10000. Fig. 9 (a) shows the probability MAE results of six datasets. We can observe that CSL$^+$ performs the best among all in terms of probability MAE, and CSL$^+$ improves the results up to 44.6%. The overall performance order of the compared methods is CSL$^+$ > GCN-VAE ~ CSL > SL > PSL.

*6.3.3* ***Scalability.*** In our experiments, the network size is varied from Philly traffic network with 603 nodes (users) and 708 edges to Slashdot network with 164,336 nodes and 2,018,920 edges. Also for our proposed CSL$^+$, the inference rule instances are varied from 980 to 2 million. As discussed in Section 5.3, CSL$^+$ shows low complexity, with linear increase with respect to $N$, $M$, and $K$, that are the number of testing and training variables and logic rules, respectively. Fig. 9 (b) shows the average running time of CSL$^+$ and other counterpart methods on real-world datasets. GCN-VAE is a little bit slower for large-scale graphs, due to handling the large size of adjacency and feature matrices. The running time of SL increases exponentially when the network size increases and SL is not scalable for large networks. Even PSL using a consensus algorithm to speed up, still it is slower when it infers large number of rule instances. CSL$^+$ and CSL scale nearly linearly with respect to the network size. The reason for CSL$^+$ being a little slower than CSL is that CSL$^+$ predicts opinions and infers the conflicting evidence simultaneously.

*Trade-off between performance and running time.* As discussed before, CSL$^+$ via doubling the variable size and not increasing the time complexity too much, achieves better performance on

the different scales of datasets. Compared to CSL, CSL+ decreases the MAE error 23~50% for all different datasets. To achieve a lower error rate, we recommend using robust CSL$^+$ method.

## 7 CONCLUSION & FUTURE WORK

In this work, we proposed a method so called Collective Subjective Logic Plus, namely CSL$^+$, that infers unknown opinion in the presence of uncertainty caused by both vacuity and conflicting evidence. CSL$^+$ keeps an opinion format based on SL to consider the degree of uncertainty (i.e., vacuity) while performing opinion reasoning operations based on PSL in order to collectively derive unknown opinions, and also resolves the issue of conflicting evidence which has not been considered by SL or CSL. CSL$^+$ can infer conflicting evidence from the given opinions during derivation of unknown opinions. Through the extensive experiments, the **key findings** are summarized as:

(1) We formulate the unknown opinion prediction problem as **an uncertainty minimization** problem , so that CSL$^+$ can effectively predict unknown opinions with linear complexity. CSL$^+$ achieves high performance under a lack of evidence and/or conflicting evidence, because CSL$^+$ learns conflicting representation of known opinions while inferring the unknown opinions.
(2) In the parameter sensitivity experiment, CSL$^+$ demonstrates less sensitivity over a wide range of test ratios, implying high resilience, compared to SL, PSL, CSL and GCN-VAE. The performance order in the expected belief probability MAE on the parameter sensitivity study experiments of semi-synthetic Epinions datasets follows: CSL$^+$>GCN-VAE > CSL > SL > PSL.
(3) The performance order in expected belief probability MAE on the experiment of real world datasets follows: CSL$^+$>GCN-VAE ~ CSL > SL > PSL. When varying the number of Sybil attack edges on the experiment of Facebook, Enron and Slashdot datasets, the performance follows: CSL$^+$>GCN-VAE > CSL > SL > PSL.
(4) Overall CSL$^+$ outperformed in the **opinion prediction performance** among all baselines and counterparts in both the expected truth belief probability MAE and opinion prediction. In addition, it provided a scalable solution for large-scale network datasets (e.g., Slashdot dataset with 164,000 nodes and 2 million edges), and scaled almost linearly in proportion to the network size.

As the future work, we plan to: (1) explore the impact of changes to hyperparameter $p_0$ on the CSL+ performance; (2) extend our proposed framework to address uncertainty-based online opinion inference problems; and (3) study the possibilities of the adversarial attacks on rule-based graph datasets and propose a new method that has resistance to strong adversarial attacks.

## REFERENCES

[1] Bach, Stephen, Matthias Broecheler, Lise Getoor, and Dianne O'leary. Scaling mpe inference for constrained continuous markov random fields with consensus optimization. In *NIPS*, pages 2654–2662, 2012.
[2] Inrix. http://inrix.com/publicsector.asp.
[3] Reference speed for congestion evaluation. http://www.inrix.com/scorecard/methodology.asp/.
[4] AbouRizk, Simaan M., Daniel W. Halpin, and James R. Wilson. Fitting beta distributions based on sample data. *JCEM*, 120(2):288–305, 1994.
[5] Bach, Stephen H., Matthias Broecheler, Bert Huang, and Lise Getoor. Hinge-loss markov random fields and probabilistic soft logic. *arXiv preprint arXiv:1505.04406*, 2015.
[6] Bach, Stephen H., Matthias Broecheler, Stanley Kok, and Lise Getoor. Decision-driven models with probabilistic soft logic. In *NIPS Workshop on Predictive Models in Personalized Medicine*, 2010.
[7] Beckman RJ, Tiet jen GL. Maximum likelihood estimation for the beta distribution. *JSCS*, 7(3-4):253–258, 1978.
[8] Christopher M. Bishop. *Pattern recognition and machine learning, 5th Edition.* Information science and statistics. Springer, 2007.
[9] Brocheler, Matthias, Lilyana Mihalkova, and Lise Getoor. Probabilistic similarity logic. In *UAI*, pages 73–82, 2010.
[10] Broecheler, Matthias, and Lise Getoor. Computing marginal distributions over continuous markov networks for statistical relational learning. In *NIPS*, pages 316–324, 2010.

[11] Ganchev, Kuzman, Jennifer Gillenwater, and Ben Taskar. Posterior regularization for structured latent variable models. *JMLR*, 11(Jul):2001–2049, 2010.
[12] Huang, Bert, Angelika Kimmig, Lise Getoor, and Jennifer Golbeck. Probabilistic soft logic for trust analysis in social networks. In *International Workshop on Statistical Relational AI*, pages 1–8, 2012.
[13] Jøsang, Audun A logic for uncertain probabilities. *J. of Uncertainty, Fuzziness and Knowledge-Based Systems*, 9(03):279–311, 2001.
[14] Jøsang, Audun Decision making under vagueness and uncertainty. In *Subjective Logic*, pages 51–82. Springer, 2016.
[15] Jøsang, Audun *Subjective Logic: A Formalism for Reasoning Under Uncertainty*. Springer, 2016.
[16] Jøsang,Audun and Touhid Bhuiyan. Optimal trust network analysis with subjective logic. In *IEEE SECURWARE*, pages 179–184, 2008.
[17] Jøsang,Audun , Ross Hayward, and Simon Pope. Trust network analysis with subjective logic. In *ACSCV*, pages 85–94, 2006.
[18] Kimmig, Angelika, Stephen Bach, Matthias Broecheler, Bert Huang, and Lise Getoor. A short introduction to probabilistic soft logic. In *NIPS*, pages 1–4, 2012.
[19] Lukasiewicz, Thomas, and Umberto Straccia. Managing uncertainty and vagueness in description logics for the semantic web. *Web Semantics: Science, Services and Agents on the World Wide Web*, 6(4):291–308, 2008.
[20] Memory, Alex, Angelika Kimmig, Stephen Bach, Louiqa Raschid, and Lise Getoor. Graph summarization in annotated data using probabilistic soft logic. In *Proceedings of the 8th International Conference on Uncertainty Reasoning for the Semantic Web-Volume 900*, pages 75–86, 2012.
[21] Pearl, Judea. *Probabilistic reasoning in intelligent systems: networks of plausible inference*. Morgan Kaufmann, 2014.
[22] Richardson, Matthew, Rakesh Agrawal, and Pedro Domingos. Trust management for the semantic web, 2003.
[23] Richardson, Matthew, and Pedro Domingos. Markov logic networks. *Machine learning*, 62(1):107–136, 2006.
[24] Shafer, Glenn. *A mathematical theory of evidence*, volume 1. Princeton university press Princeton, 1976.
[25] Smarandache, Florentin, and Jean Dezert, eds. *Advances and Applications of DSmT for Information Fusion (Collected works), second volume: Collected Works*, volume 2. Infinite Study, 2006.
[26] Smets, Philippe, and Robert Kennes. The transferable belief model. *Artificial intelligence*, 66(2):191–234, 1994.
[27] Timotheou, Stelios, Christos G. Panayiotou, and Marios M. Polycarpou. Distributed traffic signal control using the cell transmission model via the alternating direction method of multipliers. *TITS*, 16(2):919–933, 2015.
[28] Lotfi A. Zadeh. Fuzzy sets. *Information and control*, 8(3):338–353, 1965.
[29] Wang, Binghui, Jinyuan Jia, Le Zhang, and Neil Zhenqiang Gong. Structure-based sybil detection in social networks via local rule-based propagation. In *IEEE TNSE*, 2018.
[30] Gong, Neil Zhenqiang, Mario Frank, and Prateek Mittal. Sybilbelief: A semi-supervised learning approach for structure-based sybil detection. *IEEE TIFS*,9(6):976–987, 2014.
[31] Chen, Feng, Chunpai Wang, and Jin-Hee Cho. Collective subjective logic: Scalable uncertainty-based opinion inference. *In IEEE Big Data* , 7-16, 2017.
[32] Huang, Bert, Angelika Kimmig, Lise Getoor, and Jennifer Golbeck. A flexible framework for probabilistic models of social trust. *In SBP 2013,Springer.*, 265–273.
[33] Chen, Po-Ta, Feng Chen, and Zhen Qian. Road traffic congestion monitoring in social media with hinge-loss Markov random fields. *ICDM 2014,IEEE*, 80–89.
[34] Zhao, Xujiang, Feng Chen, and Jin-Hee Cho. Deep Learning based Scalable Inference of Uncertain Opinions. *ICDM 2018,IEEE.*
[35] Jøsang, Audun. The consensus operator for combining beliefs. *Artificial Intelligence 141, no. 1-2 (2002): 157-170.*
[36] Dong, Xin Luna, Laure Berti-Equille, and Divesh Srivastava. Integrating conflicting data: the role of source dependence. *VLDB Endowment 2, no. 1 (2009): 550-561.*
[37] Levy, Alon Y. Logic-based techniques in data integration. *In Logic-based artificial intelligence, pp. 575-595. Springer, Boston, MA, 2000.*
[38] Brocheler, Matthias, Lilyana Mihalkova, and Lise Getoor. Probabilistic similarity logic. *arXiv preprint arXiv:1203.3469 (2012)*
[39] Lifang, Hu, Guan Xin, and He You. Efficient combination rule of Dezert-Smarandache theory. *JSEE 19, no. 6 (2008): 1139-1144*
[40] Cerutti, Federico, Lance Kaplan, Angelika Kimmig, and Murat Åđensoy. Probabilistic Logic Programming with Beta-Distributed Random Variables. *In AAAI, vol. 33, pp. 7769-7776. 2019.*
[41] Van Allen, Tim, Ajit Singh, Russell Greiner, and Peter Hooper. Quantifying the uncertainty of a belief net response: Bayesian error-bars for belief net inference. *AI 172, no. 4-5 pp. 483-513.2008.*
[42] Kaplan, Lance, and Magdalena Ivanovska. Efficient belief propagation in second-order Bayesian networks for singly-connected graphs. *IJAR 93.pp. 132-152.2018*