Forecasting Location-based Events with Spatio-temporal Storytelling

Raimundo Dos Santos[1], Sumit Shah[2], Feng Chen[3], Arnold Boedihardjo[1], Chang-Tien Lu[2], Naren Ramakrishnan[2]
[1] U.S. Army Corps of Engineers [2] Virginia Tech [3] SUNY at Albany
{raimundo.f.dos.santos, arnold.p.boedihardjo}@usace.army.mil
{sshah, ctlu, naren}@vt.edu
fchen5@albany.edu

ABSTRACT

Storytelling, the act of connecting entities through relationships, provides an intuitive platform for exploratory analysis. This paper combines storytelling and Spatio-logical Inference (SLI) to generate rules of interaction among entities and measure how well they forecast a real-world event. The proposed algorithm first takes as input the probability of prior occurrences of events along with their spatial distances. It calculates their *soft truths*, i.e., the belief they have indeed been observed with certainty. Subsequently, the algorithm applies a relaxed form of logical conjunction and disjunction to compute a *distance to satisfaction* for each rule. The rules of lowest distances represent the best forecasts. Extensive experiments with social unrest in *Afghanistan* show that storytelling and SLI can outperform common probabilistic approaches by as much as 30% in terms of precision and 13% in terms of recall.

1. INTRODUCTION

Social events are often the byproducts of complex factors of various natures, such as financial, political, and religious. For an event to take place, the right mix of signals must come together in order to elicit reaction. Take as an example Fig. 1, which depicts some of the locations of the *Poll Tax Riots* of Great Britain in 1990. Social unrest broke out after the government enacted a flat-rate tax on each adult. But before those acts of violence occurred, other developments led up to them: activists organized protests at *Trafalgar Square*, police closed a few of London's *Underground* stations, transit was rerouted in some streets, and shops closed in certain areas. The key idea here is that social events, especially the violent ones, tend to be associated to other spatially and temporally-related nearby processes. These processes are composed of any number of constituent parts that, when identified properly, can help uncover the final event. While the above example is not surprising (after all,

ACM SIGSPATIAL LBSN'14, November 4, 2014, Dallas, TX, USA
(c) 2014 ACM ISBN 978-1-4503-3140-1 ...$15.00.

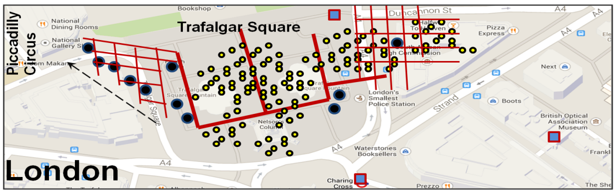


Figure 1: Approximate spread of the *Poll Tax Riots* of London in 1990. Red lines represent street closures around *Trafalgar Square*. Yellow dots denote concentration of protesters while black dots represent the locations of riots propagating north towards *Piccadilly Circus*.

protests can frequently lead to riots), acts of violence are not always transparent. The *Montreal Stanley Cup Riot* of 1993, for instance, developed quickly as the crowd celebrated a win, and had no apparent reason to engage in violence, when in fact it did. Social events can take on many characteristics, four of which are observed in the above example:

1. **event cascading:** single developments provide little insight into the overall event. On their own, the *street closures* of the above example are not alarming. But when combined with other developments, such as *gathering of protesters* and *closed shops*, a much bleaker picture begins to delineate;
2. **event propagation:** developments evolve in spatial regions through nearby areas, fading into distance. Shops, for instance, are closed near the event, but not far away from it;
3. **event sequencing:** the temporal sequence in which developments occur is essential to explain facts. *Disruption in transportation*, for example, commonly takes place after protesters have gathered, but less frequently before;
4. **event interaction:** developments represent interactions among entities: police try to contain protesters, rioters throw stones, looters attack shops, etc. Some interactions provoke strong reactions, while others do not.

Given the spatio-temporal sequence of developments as described above, one interesting question is whether an event, violent or not, can be predicted based on previous knowledge of seemingly related developments. In other words, would it be possible to foresee looting at London's *Piccadilly Circus* knowing that major protests took place earlier at *Trafalgar Square*? Making such determinations has proven elusive even with the most advanced reasoning systems available today.

While forecasting has been an art as much as a science, we can measure the feasibility that an event will occur by expanding the four characteristics mentioned above to the following hypothesis: an event can be identified by the **constituent parts** that lead to it, observing their **spatial propagation** and **time coherence**, and taking into account their **semantic interactions**. The goal of this study is to reason over spatio-temporal sequences of developments that can lead to other events, and provide a probabilistic view of their occurrence. For a focused discussion, we base our case study on violent events and apply *Probabilistic Soft Logic* (PSL) [9] to reason over them. To be more specific, Fig. 2 gives an example of what we work towards. The figure shows two event sequences **A** and **B** across a short timeline (31 March, 1990). **Sequence A** is composed of three developments taking place along the day: *Whitcomb St.* is blocked, protests occur at *Trafalgar Square*, and *Charing Cross* station is closed. They culminate in looting in the vicinity of *Piccadilly Circus* at 7 pm. **Sequence B** has three different events in different locations, but also lead to the same looting at *Piccadilly Circus.* Given the two sequences (and possibly others), the goal is to give each sequence a numerical quantification of its ability to forecast the looting. One would like to say that **Sequence A** forecasts looting at *Piccadilly Circus* with a certain value, while **Sequence B** forecasts the same looting with a lower value than **Sequence A** . Thus, **Sequence A** is a better predictor than **B**. These values represent a probability, which is explored later in this paper. The above sequences represent streams of information that *tells a story* about several things happening, and thus, we begin by framing this problem as one of *storytelling*, which we explain below.

Broadly speaking, *storytelling* is the process of connecting entities through their characteristics, actions, and events [22] in order to create meaningful streams of information. In the *Poll Tax Riots* example above, a possible storyline would be the sequence $\boxed{\textit{\textbf{activists}}} \stackrel{organize}{\rightarrow} \boxed{\textit{\textbf{protest}}} \stackrel{containedby}{\rightarrow} \boxed{\textit{\textbf{police}}} \stackrel{closed}{\rightarrow} \boxed{\textit{\textbf{streets}}}$, where entities {activists, protest, police, streets} are connected through semantic relationships {organize, containedby, closed}, and tagged with a location and timestamp. *Information retrieval* and web research have studied this problem, i.e., modeling storylines from documents and search results, and linking documents into stories [11][6][7] (the terms *stories* and *storylines* are used interchangeably). A violent event can be viewed as a vector of three important dimensions: the **spatial regions** where entities interact; **temporal coherence** which dictates the proper ordering of developments; and the **interactions** that lead to social outcomes. In this study, we enforce all of these three dimensions and focus on spatio-temporal *storytelling* related to violent events, presenting the following contributions:

1. **Reasoning with spatio-logical inference**: Key to understanding violent events is to differentiate their relevant circumstances while filtering out the unimportant ones. We introduce *spatio-logical inference* as an extension of *Probabilistic Soft Logic* to determine the likelihood that parts of an event will occur, and by extension, if the final event is probable or not. In this manner, the analyst can focus on hundreds of important happenings rather than thousands (or millions) of uninformative developments.

2. **Designing spatio-temporal methods to analyze events**: Because the dynamics of violent events are too complex for simple modeling, treating them in short spans of space and time is more conducive to human understanding. We show how such applicability can be achieved by proposing a candidate rule-generating algorithm, which becomes part of our forecasting strategy, and has not been explored in spatio-temporal analysis.

3. **Performing extensive experiments over location-based real-world events**: Because violent events are reported within and across many spatial regions, we perform several experiments using both structured and unstructured data sources. Analysis of violent events is done on the *Global Database of Events, Language, and Tone* (*GDELT*) [12], a well-established dataset of conflicts and social unrest, from which we target events in the Middle East and other parts of Asia.

In this paper, we briefly show how storylines are generated. For full details, however, we refer the reader to the spatio-temporal framework described in [20] and its originating work in [11], which we use as the basis for this research. Our focus is on spatio-temporal techniques of storyline usage to demonstrate how they can be helpful in real-world applications, using violent events as our domain. This paper is organized as follows. In Section 2, we describe related works and point their differences to our approach. Section 3 details the proposed forecasting strategies. Extensive experiments are presented in Section 4. A conclusion is finally given in Section 5.

2. RELATED WORKS

Storytelling is not a single analytical tool with predefined tasks. It can be better described as a platform of knowledge exploration for fact finding, association discovery, and inferencing. Moreover, its goals can range widely according to the domain of application: law enforcement may want to connect criminal behavior; health officials may be interested in drug interactions; and marketers may benefit from repercussion of their products in social media. As such, storytelling depends on a combination of social analysis and the technical quantitative fields. The work proposed in this paper, therefore, spans many areas of expertise, from graph analysis to geographic networks. Our research best lines up with the approaches described below.

Inferencing and **Forecasting**: While the goal of this study is not to compare the best forecasting strategies, we discuss forecasting as a case study. Some authors prefer the terms 'event prediction' while others speak of 'causality' in relation to forecasting. One such work proposed by Radinsky *et al.* reasons over the causes of events described in news articles [19]. They present an algorithm that takes as input a causality pair to find a causality predictor. Objects are defined to be similar if they relate to a third object in

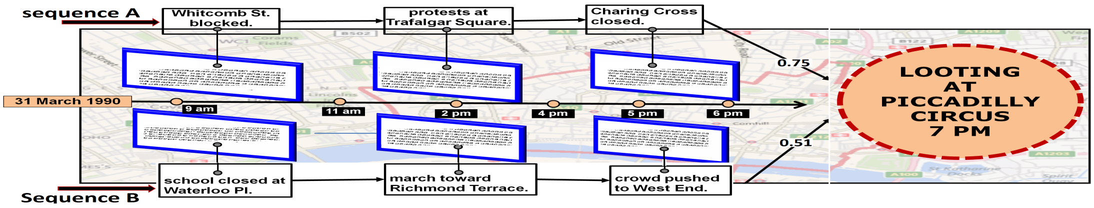


Figure 2: Example of forecasting from spatio-temporal storytelling on two event sequences. Sequence A explains the looting at *Piccadilly Circus* as an implication of the blocking of *Whitcomb St.*, protests at *Trafalgar Square*, and closing of *Charing Cross*. Sequence B, alternatively, relates the same looting with a school closing at *Waterloo Pl*, a march to *Richmond Terrace*, and the pushing of the crowd to *West End*. The 0.75 and 0.51 values indicate the beliefs with which sequences A and B respectively forecast the looting. For its higher value, sequence A is deemed a better predictor than sequence B.

the same way. This departs from our approach, which does not compare entities, but rather investigates if *behavior* is similar when entities are in close spatial proximity. Further, their work utilizes external knowledge databases, such as *LinkedData* to obtain information about well-known objects and entities. Our approach is mostly unsupervised in that all knowledge is self-contained in the targeted datasets.

Another work worth mentioning is prediction from textual data described in [18, 17]. The authors propose to capture the effects of an event by propagating it through a hierarchical model, namely an *abstraction tree*, that contains events and rules. It then finds matching nodes that can produce possible effects. In our work, we also propose a rule-based method, but do not rely on a trained model that stores rules for subsequent use. Our idea is to compare events, which may be viewed as nodes, where each event has a weight based on spatial distance. We favor this methodology as it does not depend on the availability of entity attributes or physical characteristics.

In our discussion, we note the importance of *Bayesian Inference* in forecasting. Among classical methods, it is one of the strongest foundations for *cause-effect* relationships that one can use to justify forecasting. Determining that A happens because B and C also happen is a powerful statement in many areas of knowledge, although it must be taken carefully. *Bayesian Inference* in its traditional form, however, is challenging for a few reasons: (1) it needs many instances of the same events to occur in like sequences to establish certainty; (2) without modification, it does not consider subjective criteria, such as behavioral knowledge or entity characteristics. Things "are" or "are not"; (3) it does not take into account spatial reasoning. Every element, no matter where they reside, are regarded equally. In terms of violent events, these three aspects represent challenges that must be dealt with. For this reason, we do utilize *Bayesian Inference* as a forecasting method, but do not solely rely on it.

Link Analysis: Often relying on graphs as a modeling abstraction, this class of work observes the evolution of entities in space and time ([16], [4]) and the identification of patterns ([5], [3]). The goal of link analysis, however, is not to explore stories or do forecasting. Rather, it is an attempt to quantify changes in entities and manage relationships, which leads to the notion of ranking.

Ranking in terms of link analysis has been popularly applied to web pages since the seminal works of Brin and Page [1] and Kleinberg [10]. The former computes the value of the importance of a web page based on its links and an initial damping factor. The latter also consider the page's links, but is dependent on an initial query that generates a *root set*, and is augmented by other pages that point to the *root set*.

Within the same family of the above approaches, there have been other proposed methods. The *Indegree Algorithm* is a simple heuristic that considers the *popularity* factor as a ranking measure [15]. For social media, *popularity* is a gray area: works well for high-visibility events, but may fail miserably for events that are important, but that do not get much exposure. For *storytelling*, this type of applicability is possible, but too subjective in terms of ranking. The *HITS Algorithm* [10] introduced the notion of *hub and authority*, where authorities are the pages that hold "legitimate" information, and hubs are the pages directing the user to the authorities. In terms of *storytelling*, this type of ranking would be challenging since there is no clear-cut way to determine which entities would be authorities and which would be hubs. It represents an open line of research, but outside the scope of this document.

Differences: Each of the above research fields provides solutions to the various tasks involved in storytelling. Challenges and requirements come in different flavors as a result of application demands or data characteristics. Our work, for instance, requires geolocation of entities as it relies on a spatio-temporal model where both geographical proximity and time ordering are favored. In this sense, our focus is on methods for which spatial influence and time sequencing can be intuitively justified by semantic analysis. Given the many differences in what each technique can contribute, we do not show competing approaches. Rather, we present complementary techniques that demonstrate how storylines can be a valuable analysis tool, covering a spatio-temporal niche which remains largely untapped.

3. FORECASTING VIOLENT EVENTS WITH SPATIO-LOGICAL INFERENCE

The list below provides definitions for various terms used throughout the remainder of this paper:

1. **entity**: a real-world object, whether physical or conceptual. In the context of this paper, an entity is commonly an individual or an organization.
2. **event**: an outcome, development, or happening as in the dictionary sense. A protest, an election, or a sports

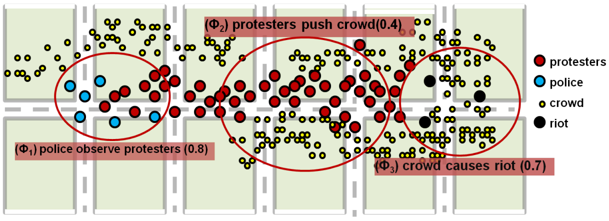


Figure 3: A spatial diagram of entity interactions enclosed in ovals. The left and center ovals represent *trigger events*, and the rightmost one is the *final event*. Each has a text description, is denoted by ϕ_1, ϕ_2, and ϕ_3, and has a *soft truth* value in the range [0,1]. The sequence conveys a storyline in which as police observe protesters, and protesters push against the crowd, a riot ensues.

match are examples. In this study, events are often of a violent nature, or components that can lead to violence.

3. **storyline**: a sequence of entities connected by relationships that describe an event. For example, "protesters –> blocked –> Trafalgar Square –> surrounded by –> police. Loosely speaking, a storyline is a stream of information that does not always obey grammatical rules.

4. **soft truth**: a numerical belief in the range [0,1] related to an event. For example, "protesters –> blocked –> Trafalgar Square [0.75]" denotes that someone has reported the blocking of Trafalgar Square, and that person is 75% certain that this event indeed happened.

5. **rule**: a storyline that has been annotated with the following: a weight in the range [0,1] (to determine the rule's importance); and soft truths of each interaction between adjacent pairs of entities. An example rule would be "[0.45] protesters –> blocked –> Trafalgar Square [0.35] –> surrounded by –> police[0.95] –> caused –> riot [0.55]". Here, [0.45] is the overall weight of the rule and [0.35],[0.95], and [0.55] are the respective soft truths for the three interactions ("blocked","sorrounded", and "caused"). The first two interactions ("blocked" and"sorrounded") are called *trigger events*, whereas the last one ("caused") is the *final event*. Note that the soft truths do not have to add up to the overall weight of the rule.

As explained earlier, violent events can be viewed as the end result of larger processes composed of one or more *trigger events*. In the *Poll Tax Riots*, for example, we identified some of those *trigger events*, two of which were that activists organized protests and police closed some streets. Intuitively, each of these *trigger events* contribute a certain amount of momentum to the riots, with some weighing in more heavily than others. Our goal then is to make use of these weights, which we will call "*soft truths*", such that, when put together, the final violent event can be deemed probable or not. A *soft truth* is simply a *numerical belief* in the range [0,1] that two entities will interact in a particular way. Thus, one person may have seen police observing protesters with a *soft truth* of 0.75, while another person is not sure the police was involved, lowering the *soft truth* to 0.25. The combination of event sequences and *soft truths* allows us to generate rules and determine how well they lead to the violent event (i.e., their *distance to satisfaction*), which we explain below.

3.0.1 Rule Inference

Informally, our problem can be expressed as follows: given a storyline composed of several interacting entities, we seek a method to combine the individual *soft truths* of each interaction and make a decision of whether the consolidated interactions are compatible with the violent event or not, i.e., if they can generate the violent event. Consider Fig. 3 which depicts different sets of entities (*police*, *protesters*, *crowd*) interacting among themselves in the streets. There are three interactions, denoted ϕ_1, ϕ_2, and ϕ_3, each described in text with an associated *soft truth* value. The *soft truths* can be obtained from various sources: historical frequencies, input of domain experts, and random sampling, among others. We wish to find an algorithmic way to answer the following question: is the combination of "*police observe protesters*" (ϕ_1) and "*protesters push against crowd*" (ϕ_2) enough for the crowd to "*cause a riot*" (ϕ_3)? Formally, this problem can be modeled in *First Order Logic* with the following statement:

$$\textbf{observe(police,protesters)} \wedge \textbf{push(protesters,crowd)} \implies \textbf{cause(crowd,riot)} \qquad \text{(Rule r}_1\text{)}$$

The above statement establishes a logical rule (r_1) that relates two *trigger events* via an "and" relationship ($\wedge$) to the *final event*, which is the riot. All of these events are in the format ***predicate*(entity$_x$,entity$_z$)**. It should read that *entity$_x$* performs the *predicate* on *entity$_z$*, meaning that when police observe protesters and protesters push against the crowd, it implies that a riot will break out. This type of statement represents hard logic, *i.e.*, it determines whether developments will or will not happen, such as in a binary fashion. In terms of violent events, hard logic in many instances is not applicable because one can seldom state with certainty that a riot will or will not occur. For this reason, instead of hard logic, a more appropriate direction is to relax the binary restriction, and permit interactions to have a *soft truth* in a continuous fashion. Relaxing these restrictions allows us to rewrite Rule r_1 as in the two examples below:

$$0.25\text{: } \textbf{observe(police,protesters)}(0.8) \wedge \textbf{push(protesters,crowd)}(0.4) \implies \textbf{cause(crowd,riot)} (0.7) \qquad \text{(Rule r}_2\text{)}$$

$$0.44\text{: } \textbf{observe(police,protesters)}(0.9) \wedge \textbf{push(protesters,crowd)}(0.3) \implies \textbf{cause(crowd,riot)} (0.1) \qquad \text{(Rule r}_3\text{)}$$

Generalizing them, we have:

$$\text{RW: } \phi_1(\mathbf{e}_a,\mathbf{e}_b)(\mathrm{w}_1) \wedge \ldots \wedge \phi_n(\mathbf{e}_u,\mathbf{e}_v)\ (\mathrm{w}_n) \implies \phi_{n+1}(\mathbf{e}_w,\mathbf{e}_z)\ (\mathrm{w}_{n+1})$$

where RW is the rule weight, ϕ_i is either a *trigger event* or the *final event*, e_i represents an entity (or set of) and w_i is a *soft truth* value. Note that *trigger events* always appear in the antecedent of the rule (*i.e.*, before the $\implies$ sign), and the *final event* always appear in the consequent of the rule (*i.e.*, after the $\implies$ sign). Subsection 3.0.2 shows a method on how to select *trigger events* and *final events* in order to generate rules. In Rules r_2 and r_3 respectively, the *trigger events* have *soft truths* (0.8, 0.4, 0.9, 0.3) and the *final events* have *soft truths* (0.7, 0.1). The rules themselves have weights 0.25 and 0.44. In practice, the rules put in formal notation statements about what "people think" or "may have seen" or "has happened" given uncertainty. There could be different rules that also lead to the same riot, such as:

0.65: **seen_with(weapons,protesters)**(0.8) ∧ **push(protesters,crowd)**(0.4) $\implies$ **cause(crowd,riot)** (0.7) (Rule r_4)

Given its higher rule weight (0.65), Rule r_4 is preferable to r_2 (0.25) and r_3 (0.44) (possibly because it involves weapons!). In a real application, thousands of such rules can be generated, which requires a numerical method to determine how good each rule actually is. In practice, we must find out whether the *trigger events* satisfy the riot, and if not, their *distance from satisfaction*. What we have described so far is derived from *Probabilistic Soft Logic* (PSL) [9]. PSL allows us to find if a rule's *trigger events* satisfy the *final event*, in which case we can then state that the rule forecasts the *final event*.

Given a set of *trigger events* $\phi=\{\phi_1, ..., \phi_n\}$, the assignment of $\phi_i \rightarrow [0,1]^n$ represents the allocation of a *soft truth* value to an interaction between two entities. This allocation is called an *interpretation* $I(\phi_i)$. PSL uses the *Lukasiewicz t-norm* and *co-norm* to relax the traditional logical conjunction (∧) and disjunction (∨) into continuous values as follows:

$$I = \begin{cases} \phi_1 \;\tilde{\wedge}\; \phi_2 = max\{0, I(\phi_1) + I(\phi_2) - 1\} \\ \phi_1 \;\tilde{\vee}\; \phi_2 = min\{I(\phi_1) + I(\phi_2), 1\} \\ \tilde{\neg} = 1 - I(l_1) \end{cases} \tag{1}$$

The ˜ symbol is applied to denote the relaxed version of the normal logical operators, which allows us to assert the following:

Definition Given a rule r, composed of a set of *trigger events* Φ={ϕ_1,...,ϕ_n} and a *final event* ϕ_{final} where each ϕ_i and ϕ_{final} have an interpretation in [0,1], r is satisfied if and only if $I(\phi_1,...,\phi_n) \leq I(\phi_{final})$.

Simply put, the above definition states that for a rule to be satisfied, its *trigger events* must combine to an overall soft truth that is less than or equal to the soft truth of the *final event*. Or conversely, the interaction established by the entities in the *final event* (ϕ_{final}) must have at least the same *soft truths* as the interactions of its constituent *trigger events* (ϕ_1,...,ϕ_n). The rule's distance to satisfaction for interpretation I is given by:

$$d_r(I) = max\{0, I(\phi_1, ..., \phi_n) - I(\phi_{final})\} \tag{2}$$

As an example, take Rule r_2, for which we wish to compute its distance to satisfaction $d_r(I)$. I(ϕ_1,ϕ_2)=max{0, 0.8+0.4-1}=0.2. Since 0.2≤0.7, we say that the rule is satisfied and $d_r(I)$=0. This contrasts with Rule r_3. where I(ϕ_1,ϕ_2)=max{0, 0.9+0.3-1}=0.2, and $d_r(I)$ = max{0,0.2-0.1} = 0.1. Rule 3 is more distant to satisfaction than Rule 2.

Interpretations can be challenging to deal with because different people have different opinions and different perceptions of facts. From an algorithmic perspective, however, all interpretations are equally valid until some are shown to be more feasible than others. For that purpose, we can calculate a distribution over all interpretations and identify the most probable ones. Given a set of Rules R = {r_1,...,r_n}, each composed of one or more *trigger events* and one *final event* in ϕ={ϕ_1,.., ϕ_n}, the probability density function over all interpretations of the rules in R is given by:

$$f(I) = \frac{1}{Z} e^{[-\sum_{r \in R} \; WR \; (d_r(I))^p]} \tag{3}$$

where WR is the rule's weight, Z is a normalization constant so that interpretations sum up to 1, and p is a *loss function* that affects the rule's distance from satisfaction, in {1,2}. When p=1, interpretations that completely satisfy one rule are preferred over others that contribute a positive distance from satisfaction. When p=2, distance from satisfaction is squared, which favors all rules to some degree. The value of Z, which is derived from *Markov Random Fields*, can be obtained from:

$$Z = \sum_{I=1}^{n} e^{[-\sum_{r \in R} \; WR \; (d_r(I))^p]} \tag{4}$$

Deriving the most probable interpretation is mathematically equivalent to maximizing $f(I)$. Knowing this distribution allows one to pick the maximal interpretations, which can be subsequently used for many purposes, such as ranking, filtering, or even classification of violent events. In the experiments section, we apply them in the context of storylines to forecast probable violent events.

3.0.2 Generation of Candidate Rules

The discussion in Subsection 3.0.1 explains how to find the "goodness" of a rule for comparison purposes. Obviously, it requires a set of rules as input, and therefore, rules must be available as a pre-condition. Rule generation is an open research field, ranging from pattern mining [21, 2] to distributed processing [13, 8]. We do not endorse any specific methods, but would rather show an effective spatial approach for that purpose. This approach injects a spatial distance component into the rules, and thus the name *spatio-logical inference*. Fast forward to Table 1 and the discussion in Subsection 4 for a brief visual example. More formally, this process obeys the steps of Alg. 1, explained below.

The algorithm takes as input a set of storylines composed of many events, where each event is associated to a location, such as latitude and longitude. The user must also input the following items: the number of desired rules to be generated (n), a matrix of probabilities where each cell contains the likelihood of observing the corresponding events (event pair *Probability Matrix*), and the desired size s of each rule. Rule size is defined as the number of *trigger events* that composes the rule, *i.e.*, the number of events concatenated by the ∧ relationship. In the previous example, the size of Rule 4, for instance, is 2. The algorithm first initializes two items: *RULES*, a data structure to hold the final rules, as empty; and the user-selected *final event* ϕ_k (line 1).

In the pre-processing stage, we wish to compute the distance between all events in the area of study, shown in line 3, to be used later. The results are stored in a *Distance-Matrix* (line 4). Rule generation is accomplished in the main stage. First, using the *Distance-Matrix*, a query finds a number s of events (*i.e.*, a number that matches the rule size) within

Algorithm 1: Candidate Rule Generation

```
inputs : set of STORYLINES = {s_1,...,s_n} where each s_i is
         composed of events φ_1, ..., φ_m tagged by locations and
         timestamps in an area of study, number of desired rules n,
         size of rule s, distance d, event-pair Probability-Matrix
output: set of weight-based rules RULES
Initialize
1: |Rules| = 0; φ_final ← φ_k ; // select one event in the dataset to be
   the final event
Pre-processing
2: while STORYLINES exist do
3:   foreach pair (φ_i,φ_j) ∈ {s_i} where φ_i ≠ φ_j do
4:     Distance-Matrix ← store(normalizedDistance(φ_i,φ_j)) ;
       // calculate the distance between each pair of entities.
5:   end
6: end
Main Stage
7: while |Rules| ≤ n do
8:   List{Trigger-Events} = query(Distance-Matrix,φ_final,s,d) ;
     // perform a query for the s closest events within distance d of the
     final event.
9:   rule ← concatenate(List{Trigger-Events},"∧",φ_final) ; // combine
     all trigger events to the final event with an "and" relationship.
10:  foreach (φ_i,φ_j) ∈ rule, φ_i ≠ φ_j, do
11:    set soft-truth(φ_i,φ_j) = Probability-Matrix[(φ_i,φ_j)] ; // set the
       soft truth for each interaction in the rule by looking up the
       probability of its composing events in the probability matrix.
12:  end
13:  set rule_RW = 1 / avgDistEvents(rule,Distance−Matrix) ;
     // set the rule's weight as the inverse of the average normalized
     distance among all its composing events
14:  RULES ← rule ; // store the formed rule.
15:  increment d ; // increase the search distance and perform another
     query.
16: end
17: output RULES;
```

a user-specified spatial distance d of the *final event*. The results are stored in List{Trigger-Events} (line 8). The rule is then formed by concatenating the found *trigger events* in the list to the *final event* ϕ_{final} via the "and " ($\wedge$) operator (line 9). What remains to be done is to set the *soft truths* for each event in the rule. This is represented in lines 10 and 11 by doing a lookup in the probability matrix already provided. The overall rule weight is obtained by averaging the distances of all events for that rule, which can be obtained from the Distance-Matrix (line 13). The formed rule is then stored in the output data structure *RULES* (line 14) and the distance is incremented for a new search for more *trigger events* (line 15). The process continues until the desired number of rules has been reached, at which point the *RULES* are output in line 17.

4. EXPERIMENTS

In this section, we apply *spatio-logical inference* to transform storylines into weight-based rules, which we then use to do forecasting.

We begin with a set of 250,000 *GDELT* events of category type ASSAULT (broken down into three subcategories) that took place in *Afghanistan*. Out of those records, we use 150,000 to extract rules, find events of high probability of occurrence using *spatio-logical inference*, and use those to find the number of similar events that exist in the remaining 100,000. The measures utilized are the following: **recall**=$\frac{similar\ identified}{similar\ identified+similar\ missed}$ as the number of similar events that were identified in the initial 150,000 records over the total number of similar events among the remaining 100,000; **precision**=$\frac{similar\ identified}{all\ retrieved}$ as the number of similar events that were identified over all retrieved records. By similar events, we denote events of the same category as specified by *GDELT* [12] (e.g., bombing and explosion) located within *distance to satisfaction* $\leq t$ of one another, where t is a threshold.

To extract rules from our dataset, we use Alg. 1, for which we give a brief example. Consider the three *GDELT* event types shown in Table 1 and geolocated in the corresponding image, which is *Afghanistan*, our region of study. The frequency for each event type is shown in parenthesis. Because the two closest events are A and B, at a distance of 115 km, these two events make up the body of the rule. The remaining one, event C, becomes the implication:

carryout-vehicular-bombing(AFGMOS,AFGREB) $\wedge$
use-as-human-shield(AFGREB,AFGCVL) $\implies$
attempt-to-assassinate(AFGCVL,AFGMIL)

To add the *soft truths*, we look at Table 1 and see that the probability of event A $= \frac{15}{45} = 0.33$, B $= \frac{5}{45} = 0.11$, and C $= \frac{25}{45} = 0.55$. The overall weight of the rule is the average distance between the three events, normalized in the range [0,1], which can be calculated as 0.76, assuming a min distance of 0 km, and a max distance of 278 km. Thus the final rule looks like:

$$0.76: \overbrace{carryout\ vehicular\ bombing(AFGMOS, AFGREB)}^{0.33} \wedge$$
$$\overbrace{use\ as\ human\ shield(AFGREB, AFGCVL)}^{0.11} \implies$$
$$\overbrace{attempt\ to\ assassinate(AFGCVL, AFGMIL)}^{0.55}$$

We then use the above rule to find its *distance to satisfaction* as described in Subsection 3. In the experiments we set the overall weight of every rule as 1.0 (every rule is equally important) and focus on the soft truths instead. Our forecasts are the rules with the least distance to satisfaction. Based on that, we use *precision* and *recall* as evaluation measures. It should be clear that the above example is a simple scenario with only three events. Given vast numbers of events, the number of rules can easily explode. Optimizations should be done, such as shortening distances or filtering out specific event types in order to alleviate computation costs. Our weights are based on frequencies and spatial distances, but it is possible that different approaches may be better suited for different domains of knowledge.

Discussion: In the context of violent events, a key consideration is whether relevant events can be forecast, knowing that relevance is a highly subjective matter. For measurement purposes, we define relevance in a comparative scale based on either *Euclidean* distance or *distance to satisfaction*: lower values are always more relevant than higher ones. Association among events can be investigated in three configurations: (1) all events are the same, such as when instances of fights result in other fights; (2) all events are different, such as when a fight and police crackdown result in a riot; (3) otherwise, events are mixed. Assume that there exists a set of *trigger events* (ϕ_1 to ϕ_n) that lead to a *final violent event* (ϕ_{final}) with a d *Euclidean* distance or *distance to satisfaction*. Then one can assert a successful forecast for other unseen *final violent events* provided that the *trigger events* lead to the same *final violent events* with the same or lower distance d. In other words, comparing the association between two sets of events, if the events match (or partially match) on at least one *trigger event* and distance is just as low, then a forecast is made. If no events match or distance d is off, then the forecast is a miss.

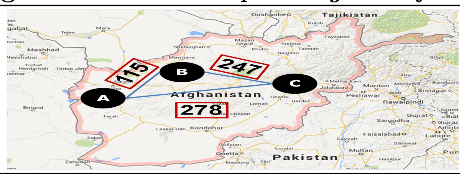
Using the above ideas, we retrieve all events from our dataset that fit those conditions, and count how many we

[!ht]

Table 1: Example of three *GDELT* events in *Afghanistan*. The number in parenthesis is the total number of events of that kind reported in *Afghanistan* in 2011. The latitude and longitude values represent the locations where one instance of that event took place between the source and target groups. The image shows the distance in km between the different events, which are used to generate rules for *spatio-logical inference*.

	Event Description (instances)	Source†	Target†	Lat	Lng
A	Carry out vehicular bombing (15)	AFGMOS	AFGREB	34.3333	70.4167
B	Use as human shield (5)	AFGREB	AFGCVL	34.5167	69.1833
C	Attempt to assassinate (25)	AFGCVL	AFGMIL	32.3472	68.5932

†AFG=Afghanistan, MOS=Muslim group, REB=Rebel group, CVL=Civilians, MIL=Military



were able to forecast, and how many were missed. For simplicity, we limit the views to a range where the max distance between any two events is 100 km. Fig. 4 shows six plots with different measurements for discussion. High recall values indicate that previously-unseen events are being found without going over a distance limit. This is shown in Fig. 4(A), in which recall values range from 30% to 66%. In the range where distance between events lies between 0 to 50 km, recall remains fairly constant at around 62% for mixed events. It indicates that, for many of the generated rules, their constituent events lead to the same *final violent event* with a distance of 50 km or less. For events of the same category, recall trends upward up to 50 km, but only gets worse thereafter. More intriguing are events of unique type, in which recall is good with short distances (0 - 20 km) or long distances (80 - 100 km), but often worse between (21 - 79 km). The lesson learned from this example is the following: the *soft truth* values established in the rules seem to be appropriate for the initial part of the graph (shorter distances) and the late stages (longer distances), but may not be ideal for mid distances. Similar trends as the above is also seen in Fig. 4(B), which shows precision by *Euclidean* distance. In general, one would expect high recall for short distances and vice-versa. Intuitively, government in Kabul experiences many bombings over time, but ones which are not necessarily related to other bombings in far-away cities, such as Charikar. However, our data indicate that, in many instances, longer distances between events display higher precision than shorter ones. This is the case in Fig. 4(B) where the highest precision for mixed events is approximately 60% with a distance of 60 km. For unique events, this fact is even more pronounced, since the highest precision (57%) lines up with the highest distance of 100 km. The above discussion points to the importance of relating event types, locations, distances, and frequencies in the discussion of violent events. We use these components to generate event-based rules. Fig. 4(E) summarizes recall in terms of the number of generated rules according to event type. This time, distance is disregarded, which has a different effect on the results. When distance is not considered, recall is consistently high when events have unique types, but suffer considerably for mixed ones, with a higher variation for same event types. In practice, it seems to relay the message that "a forecast is safe when **a and b lead to c**, but not when **a and a lead to c** or **b and b lead to c**". The closest that the three lines come together is at approximately 33k generated rules, where recall ranges from 44% to 54%. This is a significant difference from the distance approach, which underscores the importance of spatial analysis. Fig. 4(E) also shows that fewer rules is not necessarily better than more rules (as one might expect). In fact, some of the best recall values can be seen exactly at the end of the graph when the number of generated rules hits 100k. The not-all-clear message here is that violent events are better explainable with different types of events, and not with the reoccurrence of the same event types. Fig. 4(C) and Fig. 4(D) show variations of precision and recall in terms of *distance to satisfaction*. Interestingly, recall (Fig. 4(C)) is observed to be higher with lower *distance to satisfaction*, with few exceptions. In the case of precision (Fig. 4(D)), however, the trend is inconsistent. No clear pattern can be established of where higher or lower recall can be equated with higher or lower *distance to satisfaction*.

For illustrative purposes, Fig. 4(F) depicts the distribution of three events for the 10 largest cities in *Afghanistan*, which were used in this dataset. It shows, for instance, that (for this partial dataset) *vehicular bombings* are mostly frequent in *Kandahar*, *Kabul*, and *Jalalabad* (in this order), while *Kabul* itself sees most of the *human shield* events. While this graph is not the complete dataset used in the experiments, it gives the reader a sense of the spatial locations being investigated and the event types we were looking for.

Finally, we display some of the events our approach is able to forecast in Table 2. Starting from a sample generated rule ($G1$), whose *final violent event* relates to *destruction of property*, having a *distance to satisfaction* $= 0.25$, the table first shows a set of four rules that were correctly forecast ($F1$, $F2$, $F3$, $F4$). $F1$, for example, tells about some sort of "negotiation" that involves an action of "release", which eventually ended up as "confiscation of property". Without the benefit of external knowledge, we do not know the details of this case. However, we can affirm with confidence that this event is very close in concept to the original rule $G1$, which also has a "release" component, involves "destruction of property", and has lower *distance to satisfaction* than the original rule $G1$ (0.13 as opposed to 0.25). These events took place in 2010 in Afghanistan at a distance of 34 km from each other. The same is true for $F3$, which also deals with "destruction of property", though coming from totally separate *trigger events* related to "military cooperation" and a "curfew". $F2$ and $F4$ have slightly higher *distance to satisfaction*, albeit still below the limit of 0.25 established by $G1$.

Further down the table, we show four other rules that were not considered valid forecasts based on our pre-established conditions. The first one, $M1$, does not have a similar *final violent event* to $G1$, and thus we do not have a basis to compare distances given that the rules share little in common (only one *trigger event*). $M2$ shares no *trigger events* at all

Table 2: Examples of events that were forecast correctly or missed based on the generated rule shown across the top row. The final violent event is destruction of property as shown in the implication of the generated rule $G1$. For each missed forecast, a reason explains why the forecast was not considered valid. The number in parenthesis at the end of each rule denotes *distance to satisfaction*.

	Generated Rule	Distance to Satisfaction: 0.25
$G1$	engage(AFGGOV,RADMOS) $\wedge$ demand-release(AFG,COP) $\implies$ destroy-property(AFGREB,RADMOS)	
	Events Correctly Forecast by Rule $G1$	
$F1$	halt-negotiation(AFGCOP,UAF) $\wedge$ demand-release(AFG,COP) $\implies$ confiscate-property(AFGGOV,RADMOS) (0.13)	
$F2$	engage(AFGGOV,RADMOS) $\wedge$ impose-embargo(AFGSPY,AFGCRM) $\implies$ seize(AFGGOV,AFGINSTALUAF) (0.17)	
$F3$	cooperate-militarily(AFGCOP,AFG) $\wedge$ impose-curfew(AFGSPY,AFGCVL) $\implies$ destroy-property(AFGGOV,RADMOS) (0.12)	
$F4$	ban parties(AFGCOP,UAF) $\wedge$ demand-material-coop(AFGGOVBUS,AFGCVL) $\implies$ destroy-property(AFGREL,RADMOS) (0.24)	
	Missed Forecasts	**Reason for Miss**
$M1$	engage(AFGGOV,RADMOS) $\wedge$ reject(AFG,AFG) $\implies$ mobilize-armed-forces(AFG,RADMOS) (0.20)	wrong *final violent event*
$M2$	halt-negotiation(AFGCOP,UAF) $\wedge$ use-tactics-violent(AFG,COP) $\implies$ destroy-property(AFGGOV,RADMOS) (0.20)	no match on *trigger events*
$M3$	expel(AFGMIL,AFGELI) $\wedge$ rally-opposition(AFGGOV,AFGREF) $\implies$ demand-release(AFGGOV,RADMOS) (0.38))	high *distance to satisfaction*
$M4$	engage(AFGEDU,AFGMIL) $\wedge$ reduce-econ-aid(UAF,AFGREF) $\implies$ destroy-property(AFGGOV,RADMOS) (0.31)	high *distance to satisfaction*
	AFG=*Afghanistan*, BUS=*business*, COP=*police force*, CRM=*criminal*, CVL=*civilian*, ELI=*elites*, GOV=*government*, MIL=*military*, MOS=*muslim*, RAD=*radical*, REB=*rebels*, REF=*refugee* , SPY=*spy*, UAF=*unidentified armed force*	

with $G1$, and thus is not valid because our approach needs at least one element in common. $M3$ and $M4$ are both too distant in terms of *distance to satisfaction* from 0.25, and thus are rejected as well.

4.1 Comparison of the Different Forecasting Strategies

In this subsection, we put in perspective two forecasting strategies. Our goal is not to find the best forecasting approach, but rather to contrast them. One line of research complimentary to our work, but which often does not include spatial *storytelling*, is event detection, to which we point the reader for further reading [14, 23]. We frame our discussion in terms of *precision* and *recall*, as done before.

Table 3 lists a set of 5 event types, labeled $E1$ through $E5$, from the *GDELT* dataset that we target as *final violent events*. We use 250,000 records: 150,000 as input and 100,000 for validation. For each event type, the table shows precision and recall values, calculated as explained earlier, using two technical approaches. *Bayes* denotes traditional *Bayesian Inference*, where we consider combinations of events whose probability of occurrence is 10% or less (higher than 10% was less significant in our dataset). The other two methods (*sli-1* and *sli-2*) are considered according to the nature of their distances. Again, for *Simple Bayes* (*sB*), precision and recall values are provided for all events whose probabilities are less than or equal to 0.1. Similarly, *sli-1* and *sli-2* are considered for rules where the *distance to satisfaction* is less than or equal to 1.5. These values are selected because they are approximate midpoint distances between the events in our dataset, and thus a reasonable breakoff point for investigative purposes.

The way to interpret the table, exemplified for row 1, is as follows. Upon running *Bayesian Inference* for event $E1$ (*attempt to assassinate*) in the initial set of 150,000 events, the results indicated 5,101 combinations (not shown in table) of *trigger events* that led to $E1$ with a probability $\leq$ 10%. However, when validating against the remaining 100,000 records, those combinations only contained 985 out of 2662 events with a probability$\leq$ 10% (and that shared at least one event with the generating combination), yielding a precision of 0.37. For recall, 985 combinations were found, but 1539 should have been identified, resulting in a recall of 0.64.

Discussion: The first noticeable point is the fairly low levels of precision for *simple Bayes* for all event types, except for $E2$ (*carry out vehicular bombing*). Especially for $E3$ (*engage in violence*), very few combinations of events lead to that type of event, making it hard to identify. Overall, the reason for the low precision values is that traditional *Bayes* requires the same events in the same sequence for the probabilities to be high. For violent events, however, sequence can seldom be guaranteed, rendering Bayes less than ideal. The situation is more favorable in terms of recall, as relevant items are often retrieved with greater success.

For *simple Bayes*, precision displays mixed signals. It is good for $E2$ (0.61) and $E5$ (0.51) , but decreases for $E1$ (0.35). The reason has to do with the distance, which impacts $E1$ negatively. Verifying the dataset, it can be seen that events within short distances of $E1$ are not commonly observed, thus impacting precision. In general, recall is consistently high, not seeing much impact from either side of the 10% threshold.

sli-1 demonstrates the highest precision of any of the approaches for event $E4$ (0.78). It is never lower than 0.51. However, the most robust numbers come from *sli-2*. Its positive aspect is consistency even when it does not account for the highest precision. Spatial distance appears to be a factor here. This is most likely due to many instances of the same pairs of events that lead to $E4$. Its recall values are reasonably high across all events.

Two observations can be made about *spatio-logical inference* (*sli*): first, precision shows good consistency for low *distances to satisfaction*, which is desirable in terms of forecasting. However, one should also expect low precision for high *distances to satisfaction* (*df*), which in general does not occur when *df* >1.0. While high precision is normally a good thing, we would prefer *df* to oppose precision hand-in-hand (low to high, and high to low). *sli* is very stable in terms of recall, and interestingly, especially when distances are long. While Table 3 only shows a limited number of results, our overall experience points to *sli-2* as having most of the highest recall and precision scores.

5. CONCLUSION

This paper combines storytelling with Spatio-Logical Inference to perform event forecasting based on the interactions among real-world entities. The proposed algorithm generates interaction rules, calculates their soft truths, and uses relaxed logical conjunction and disjunction to compute a rule's *distance to satisfaction*. The lower the *distance to satisfaction* the stronger the forecast for the events in that rule. Experiments on Afghanistan data demonstrated that this proposed approach provides significantly higher preci-

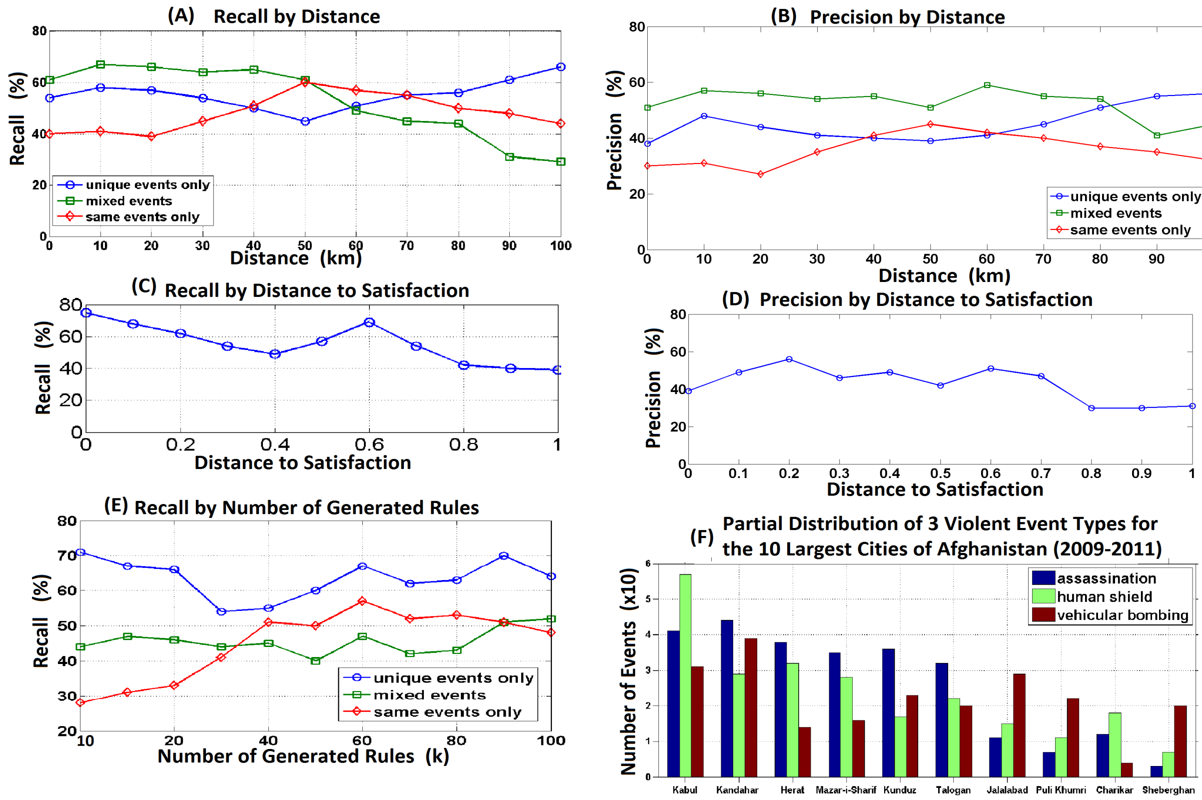


Figure 4: Results from *spatio-logical inference*. (A) Effect of distance between events on recall. (B) Effect of distance between events on precision. (C) Effect of the *distance to satisfaction* on recall. (D) Effect of the *distance to satisfaction* on precision. (E) Effect of the number of generated rules on recall. (F) Distribution of violent events in cities of Afghanistan.

Table 3: Comparison of precision and recall for Simple Bayesian Inference and Spatio-logical Inference. Precision and recall are measured based on thresholds of 0.10 for Bayes and distance to satisfaction of 0.5 for Spatio-Logical Inference. The highest values are shown in red letters.

	simple Bayes (sB)			spatio-logical inference (sli-1)			spatio-logical inference (sl1-2)		
Event	precision	recall	f-measure	precision	recall	f-measure	precision	recall	f-measure
E1-attempt to assassinate	0.35	0.81	0.48	0.51	0.58	0.54	0.55	0.77	0.64
E2-carry out vehicular bombing	0.61	0.75	0.67	0.57	0.72	0.63	0.66	0.80	0.72
E3-engage in violence	0.48	0.71	0.57	0.54	0.72	0.61	0.66	0.79	0.71
E4-conduct strike or boycott for rights	0.45	0.65	0.53	0.78	0.61	0.68	0.67	0.79	0.72
E5-destroy property	0.51	0.70	0.59	0.61	0.77	0.68	0.45	0.52	0.48

sion and recall scores than a typical traditional probabilistic method. For future work, our objective is to establish storytelling as a robust tool for entity reasoning in a wide range of application domains.

6. REFERENCES

[1] S. Brin and L. Page. The anatomy of a large-scale hypertextual web search engine. *Computer Networks and ISDN Systems*, 30:107–117, 1998.

[2] A. Ceglar and J. F. Roddick. Association mining. *ACM Computing Surveys*, 38(2), July 2006.

[3] J. Chan, J. Bailey, and C. Leckie. Discovering correlated spatio-temporal changes in evolving graphs. *Knowledge and Information Systems*, 16:53–96, 2008.

[4] J. Chan, J. Bailey, and C. Leckie. Using graph partitioning to discover regions of correlated spatio-temporal change in evolving graphs. *Intelligent Data Analysis (IDA)*, 13:755–793, 2009.

[5] B. George, J. Kang, and S. Shekhar. Spatio-temporal sensor graphs (stsg): A data model for the discovery of spatio-temporal patterns. *Intelligengt Data Analysis (IDA)*, 13:457–475, 2009.

[6] M. S. Hossain, C. Andrews, N. Ramakrishnan, and C. North. Helping intelligence analysts make connections. In *Workshop on Scalable Integration of Analytics and Visualization*, AAAI '11, pages 22–31, 2011.

[7] M. S. Hossain, J. Gresock, Y. Edmonds, R. Helm, M. Potts, and N. Ramakrishnan. Connecting the dots between pubmed abstracts. *Public Library of Science (PLoS ONE)*, 7(1), 2012.

[8] Y. Huang, Z. Lu, and H. Hu. A new permutation approach for distributed association rule mining. In

ACM International Conference on Information and Knowledge Management, CIKM '05, pages 351–352, 2005.

[9] A. Kimmig, S. H. Bach, M. Broecheler, B. Huang, and L. Getoor. A short introduction to probabilistic soft logic. In *NIPS Workshop on Probabilistic Programming: Foundations and Applications*, 2012.

[10] J. Kleinberg. Authoritative sources in a hyperlinked environment. In *Society of industrial and applied mathematics (SIAM)*, pages 668–677, 1998.

[11] D. Kumar, N. Ramakrishnan, R. F. Helm, and M. Potts. Algorithms for storytelling. *IEEE TKDE*, 20(6):736–751, 2 2008.

[12] K. Leetaru and P. Schrodt. Gdelt: Global database of events, language, and tone, 1979-2012. In *Proc. Intl. Studies Assoc. Annual Conference (ISA)*, 2013.

[13] S. Li, T. Wu, and W. M. Pottenger. Distributed higher order association rule mining using information extracted from textual data. *ACM SIGKDD Explorations Newsletter*, 7(1):26–35, June 2005.

[14] Z. Li, B. Wang, M. Li, and W.-Y. Ma. A probabilistic model for retrospective news event detection. In *ACM SIGIR Conference on Research and Development in Information Retrieval*, SIGIR '05, pages 106–113, 2005.

[15] M. Marchiori. The quest for correct information on web: Hyper search engines. In *World Wide Web Conference (WWW)*, pages 1225–1235, 1997.

[16] G. D. Mondo, M. Rodriguez, C. Claramunt, L. Bravo, and R. Thibaud. Modeling consistency of spatio-temporal graphs. *Data and Knowledge Engineering*, 84:59–80, 2013.

[17] K. Radinsky, S. Davidovich, and S. Markovitch. Learning causality for news events prediction. In *World Wide Web Conference (WWW)*, pages 909–918, 2012.

[18] K. Radinsky, S. Davidovich, and S. Markovitch. Learning to predict from textual data. *Journal of Artificial Intelligence Research (JAIR)*, 45:641–684, 2012.

[19] K. Radinsky and E. Horvitz. Mining the web to predict future events. In *Conference on Web Search and Data Mining*, WSDM '13, pages 255–264, 2013.

[20] R. D. Santos, S. Shah, F. Chen, A. Boedihardjo, P. Butler, C.-T. Lu, and N. Ramakrhishnan. Spatio-temporal storytelling on Twitter. Technical report, Virginia Tech, 10 2013. http://vtechworks.lib.vt.edu/handle/10919/24701.

[21] M.-C. Tseng and W.-Y. Lin. Efficient mining of generalized association rules with non-uniform minimum support. *Data and Knowledge Engineering*, 62(1):41–64, July 2007.

[22] S. Turner. *The creative process: A computer model of storytelling and creativity*, pages 122–123. Psychology Press, 1994.

[23] K. N. Vavliakis, A. L. Symeonidis, and P. A. Mitkas. Event identification in web social media through named entity recognition and topic modeling. *Data and Knowledge Engineering*, 88:1–24, Nov. 2013.