The Knowledge Engineering Review, Vol. 33, e1, 1–14.
doi:10.1017/S0269888918000024

Relational time series forecasting

RYAN A. ROSSI

Adobe Research, 345 Park Ave, San Jose, CA 95110;
e-mail: rrossi@adobe.com

Abstract

Networks encode dependencies between entities (people, computers, proteins) and allow us to study phenomena across social, technological, and biological domains. These networks naturally evolve over time by the addition, deletion, and changing of links, nodes, and attributes. Despite the importance of modeling these dynamics, existing work in relational machine learning has ignored *relational time series data*. Relational time series learning lies at the intersection of traditional time series analysis and statistical relational learning, and bridges the gap between these two fundamentally important problems. This paper formulates the relational time series learning problem, and a general framework and taxonomy for representation discovery tasks of both *nodes* and *links* including predicting their existence, label, and weight (importance), as well as systematically constructing features. We also reinterpret the prediction task leading to the proposal of two important relational time series forecasting tasks consisting of (i) relational time series classification (predicts a future class or label of an entity), and (ii) relational time series regression (predicts a future real-valued attribute or weight). Relational time series models are designed to leverage both relational and temporal dependencies to minimize forecasting error for both relational time series classification and regression. Finally, we discuss challenges and open problems that remain to be addressed.

1 Introduction

1.1 Motivation

In recent years, relational data have grown at a tremendous rate; it is present in domains such as the Internet and the World Wide Web (Albert *et al.*, 1999; Faloutsos *et al.*, 1999; Broder *et al.*, 2000), scientific citation and collaboration (Newman, 2001; McGovern *et al.*, 2003), epidemiology (Kleczkowski & Grenfell, 1999; Moore & Newman, 2000; May & Lloyd, 2001; Pastor-Satorras & Vespignani, 2001), communication analysis (Rossi & Neville, 2010), metabolism (Jeong *et al.*, 2000; Wagner & Fell, 2001), ecosystems (Camacho *et al.*, 2002; Dunne *et al.*, 2002), bioinformatics (Jeong *et al.*, 2001; Maslov & Sneppen, 2002), fraud and terrorist analysis (Krebs, 2002; Neville *et al.*, 2005), and many others. The links in these data may represent citations, friendships, associations, metabolic functions, communications, co-locations, shared mechanisms, or many other explicit or implicit relationships.

The majority of these real-world relational networks are naturally dynamic—evolving over time with the addition, deletion, and changing of nodes, links, and attributes. Despite the fundamental importance of these dynamics, the majority of work in *relational learning* has ignored the dynamics in relational data (i.e. attributed network data). In particular, dynamic attributed graphs have three main components that vary in time. First, the attribute values (on nodes or links) may change over time (e.g. research area of an author). Next, links might be created and deleted throughout time (e.g. host connections are opened and closed). Finally, nodes might appear and disappear over time (e.g. through activity in an online social network). Figure 1 provides an intuitive view of these underlying dynamics.

Previous research in machine learning (ML) assumes independently and identically distributed data (IID) (Anderson *et al.*, 1986; Bishop *et al.*, 2006); and has ignored relational dependencies (and temporal dependencies). This independence assumption is often *violated* in relational data as (relational) dependencies among data instances are naturally encoded. More specifically, *relational autocorrelation* is a correlation or statistical dependence between the values of the same attribute across linked instances (Jensen *et al.*, 2004) and is a fundamental property of many relational data sets. For instance, people are often linked by business associations, and information about one person can be highly informative for a prediction task involving an associate of that person. Recently, statistical relational learning (SRL) methods (Getoor & Taskar, 2007) were developed to leverage the relational dependencies (i.e. relational autocorrelation (Rossi *et al.*, 2012b), also known as homophily (McPherson *et al.*, 2001); between nodes (Friedman *et al.*, 1999; Macskassy & Provost, 2003; Neville *et al.*, 2003; De Raedt & Kersting, 2008; McDowell *et al.*, 2010). In many cases, these relational learning methods improve predictive performance over traditional IID techniques (Macskassy & Provost, 2003; Neville *et al.*, 2003; McDowell *et al.*, 2009).

Relational learning methods have been shown to improve over traditional ML by modeling relational dependencies, yet they have ignored temporal information (i.e. explicitly assumes the data are independent with respect to time). In that same spirit, our work seeks to make further improvements in predictive performance by incorporating temporal information and designing methods to accurately learn, represent, and model temporal *and* relational dependencies. The temporal information is known to be significantly important to accurately model, predict, and understand relational data (Watts & Strogatz, 1998; Newman *et al.*, 2006). In fact, time plays a key role in many natural laws and is at the heart of our understanding of the universe, that is, the unification of space and time in physics (Einstein, 1906) and how time is related to space and vice versa is fundamentally important to our understanding and interpretation of the universe and its laws (Einstein, 1906; Bock *et al.*, 2008). We make a similar argument here, that ignoring time in attributed networks can only add further uncertainty, as time places a natural ordering on the network components, including the changing of attribute values, links, and nodes.

This work formulates the problem of relational time series learning and proposes a framework that consists of two main components as shown in Figure 2. The first component learns a feature-based

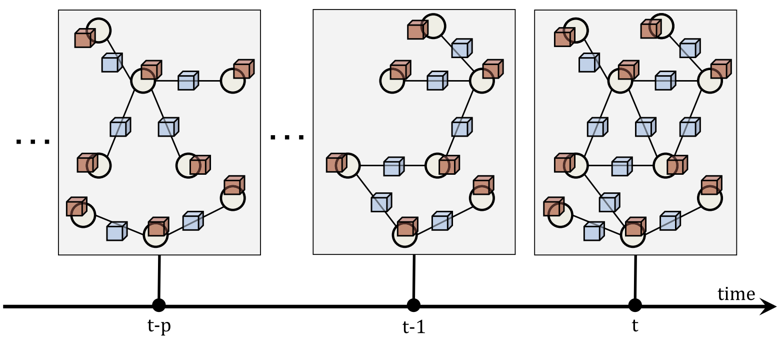

Figure 1 Relational time series data

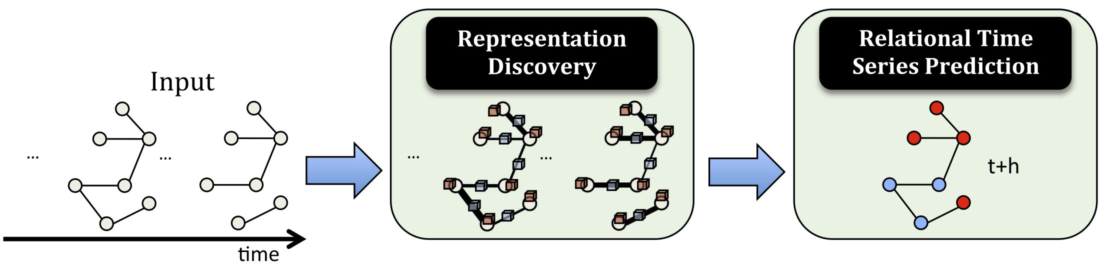

Figure 2 Overview of relational time series learning

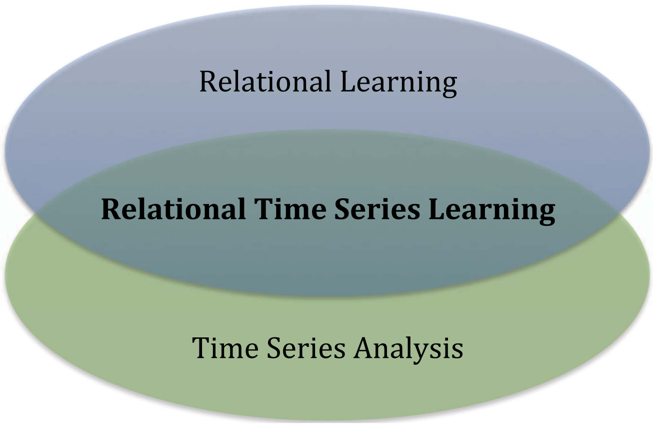

Figure 3 Toward bridging the gap between relational learning and time series analysis. The focus of this work is on the intersection between these two areas which we call *Relational Time Series Learning*. Relational learning has primarily focused on static relational graph data (graphs + attributes), whereas the time series analysis literature avoids graph data and instead focuses on independent time series (i.e. modeling each of time series disjointly) or time series that are assumed to be completely dependent (i.e. clique).

representation from a collection of dynamic relational data (i.e., a time series of graphs and attributes) given as input which incorporates the fundamental temporal dependencies in relational graph data. While the second component leverages the learned feature-based representation for relational time series prediction, which includes both relational time series classification models and regression. In other words, this work proposes techniques for learning an appropriate representation from dynamic attributed graph data for the purpose of improving accuracy of a given relational learning[1] (Getoor & Taskar, 2007) task such as classification (Rossi *et al*., 2012b). We propose a taxonomy shown in Figure 4 for the fundamental representation tasks for nodes in dynamic attributed networks. This work focuses on the three fundamental representation tasks for dynamic attributed networks including dynamic node labeling, dynamic node weighting, and dynamic node prediction. As an aside, let us note that we do not focus on the (symmetric) link-based representation tasks since they have received considerable attention recently (in various other forms) (Hasan *et al*., 2006; Liben-Nowell & Kleinberg, 2007; Lassez *et al*., 2008; Acar *et al*., 2009; Koren *et al*., 2009; Xiang *et al*., 2010a; Al Hasan & Zaki, 2011; Menon & Elkan, 2011; Oyama *et al*., 2011; Schall, 2014) and therefore this work focuses on the novel dynamic representation tasks for nodes.

Ultimately, the goal of this work is to help further bridge the gap between relational learning (Friedman *et al*., 1999; Macskassy & Provost, 2003; Neville *et al*., 2003; Getoor & Taskar, 2007; De Raedt & Kersting, 2008; McDowell *et al*., 2010; Rossi *et al*., 2012b) and traditional time series analysis (Pindyck & Rubinfeld, 1981; Clements & Hendry, 1998; Croushore & Stark, 2001; Brockwell & Davis, 2002; Marcellino *et al*., 2006; Ahmed *et al*., 2010; Box *et al*., 2013) (see Figure 3). This gap between these fundamentally important problems seemingly arose due to the fact that the majority of relational learning techniques from the ML community has ignored temporal or dynamic relational data (Chakrabarti *et al*., 1998; Friedman *et al*., 1999; Domingos & Richardson, 2001; Macskassy & Provost, 2003; Neville *et al*., 2003; Getoor & Taskar, 2007; De Raedt & Kersting, 2008; McDowell *et al*., 2010), whereas the time series work has *ignored graph data* and has mainly focused on (i) modeling independent time series or (ii) multiple time series that are assumed to be completely correlated (Pindyck & Rubinfeld, 1981; Brockwell & Davis, 2002; Ahmed *et al*., 2010; Box *et al*., 2013). The intersection of these two areas is *relational time series learning* and differs significantly from relational learning and time series analysis in the data utilized, model and data assumptions, and their objectives. For instance, the prediction objective of the relational learning problem is mainly for within-network or across-network prediction tasks and does not predict the future node attribute values. Relational learning does not utilize or model temporal information, whereas time series analysis lacks relational information. There are many other differences discussed later.

[1] This area of ML is sometimes referred to as SRL or relational machine learning.

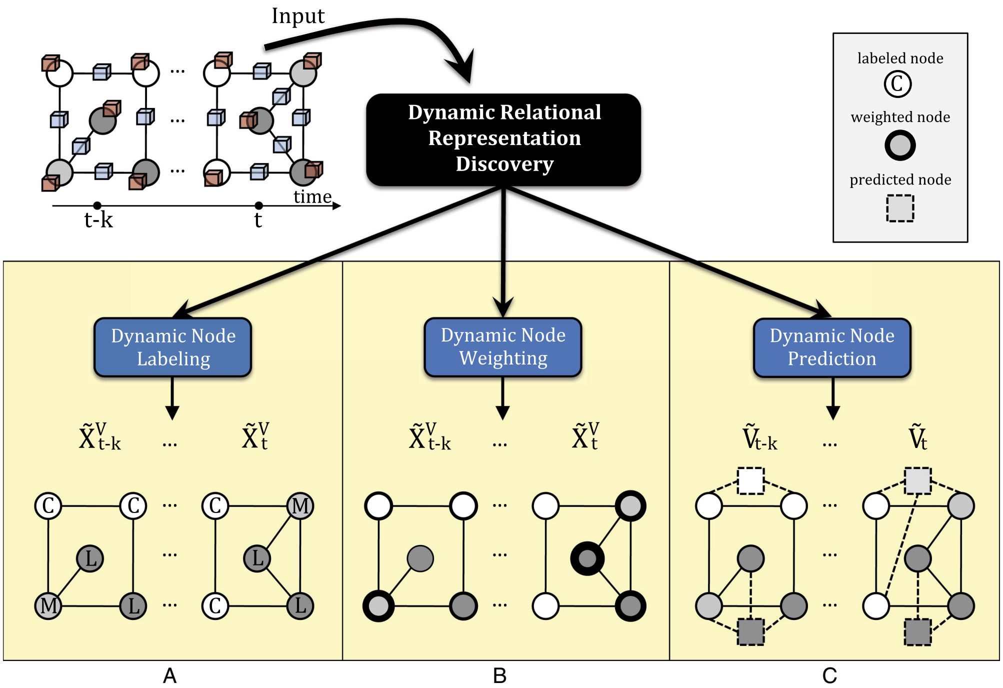

Figure 4 Taxonomy of dynamic relational representation tasks for nodes. For the dynamic node representation tasks, we introduce a dynamic relational representation taxonomy focused on the representation tasks of dynamic node labeling, weighting, and predicting the existence of nodes.

1.2 Scope of this article

This article focuses on the problem of relational time-series forecasting and techniques for learning dynamic relational representations from graph data with timestamps. We do not focus on dynamic network analysis and mining. However, whenever possible, we do reinterpret techniques that were proposed for different problems, and discuss how they could potentially be used for relational time series representation learning for improving relational time series forecasting. As an aside, the relational time series forecasting methods may also be useful for other applications (i.e. besides predicting node attributes) including dynamic network analysis (Tang *et al.*, 2010; Holme & Saramäki, 2012), anomaly detection (Bunke & Kraetzl, 2004; Chandola *et al.*, 2009) in dynamic graphs (Noble & Cook, 2003; Ide & Kashima, 2004; Tong & Lin, 2011; Rossi *et al.*, 2013a), dynamic ranking (O'Madadhain & Smyth, 2005; Das Sarma *et al.*, 2008; Rossi & Gleich, 2012), among many other ML tasks and applications (Rossi *et al.*, 2013b, 2013c; Rossi, 2014). However, this is outside the scope of this article.

1.3 Overview

This article is organized as follows: Section 2 introduces and defines the *relational time series forecasting* problem, which consists of relational time series classification (Section 2.1) and regression (Section 2.2). Next, Section 3 presents the *relational time series representation learning* for relational time series forecasting. Section 4 surveys and reinterprets existing work for use in the relational time series representation learning and forecasting problem. Section 5 presents a few important open and challenging problems. Finally, Section 6 concludes.

2 Relational time series forecasting

Using the learned representation, we demonstrate the effectiveness of these techniques for relational time series *classification* and *regression* of dynamic node attributes. We define relational time series classification (Section 2.1) and regression (Section 2.2) more precisely below.

2.1 Relational time series classification

PROBLEM 1 (RELATIONAL TIME SERIES CLASSIFICATION). *Given a 'known' time series of attributed graph data $\mathcal{G} = \{\mathbf{G}_1, \ldots, \mathbf{G}_t\}$ for learning, the task is to infer the class labels $\mathbf{Y}_{t+h}$ of the nodes at time $t+h$ in the graph where L refers to the set of possible labels.*

As an aside, if $h = 1$ then we call this one-step ahead prediction, whereas multi-step ahead prediction refers to the case when $h > 1$, and thus the prediction is across multiple timesteps. Our relational time series classification methods are also flexible for both binary (i.e. $|L| = 2$) and multi-class problems (i.e. $|L| > 2$), whereas binary classification has been the primary focus of the past relational learning methods for static graphs. Similarly, we also investigate the relational time series regression problem.

2.2 Relational time series regression

PROBLEM 2 (RELATIONAL TIME SERIES REGRESSION). *Given a time series of attributed graphs $\mathcal{G} = \{\mathbf{G}_1, \ldots, \mathbf{G}_t\}$, the task is to estimate the real-valued variable $\mathbf{Y}_{t+h} \in \mathbb{R}^n$ at time $t+h$ for the nodes in the graph.*

The prediction task investigated in this paper is also fundamentally different than the traditional relational learning problems/assumptions. More specifically, we define within-network (e.g. inference) as the task where training instances from a single (static) graph are connected directly to instances whose classification labels are to be predicted (McDowell *et al*., 2009; Xiang *et al*., 2010b). Conversely, the task of across-network inference attempts to learn a model on a (static) network and applying the learned models to a separate network (Craven *et al*., 1998; Lu & Getoor, 2003). For instance, we may learn a model from a static and/or aggregated graph from Facebook and use that learned model for prediction on another social network such as Google+ or Twitter. While both prediction problems for relational learning assume a static network, they also differ fundamentally in their underlying assumptions and goals. On the other hand, we focus on using the past time series of attributed graphs where the training nodes may be connected directly to nodes whose classification labels are to be predicted and similarly the past time series observations of the prediction attribute may also be directly used. The fundamental idea is that both past relational and temporal dependencies and information may be used to predict the future time series values of a given attribute. We also note that we may learn a model using some past data and use it to predict the future value at $t+h$ of an attribute time series, or we could use a technique that does 'lazy learning' in the sense that the past data are determined upon prediction time and used for predicting $t+h$.

3 Representation learning from relational time series data

Recently, relational data representations have become an increasingly important topic due to the recent proliferation of network data (e.g. social, biological, information networks) and a corresponding increase in the application of SRL algorithms to these domains. In particular, appropriate transformations of the nodes, links, and/or features of the data can dramatically affect the capabilities and results of SRL algorithms. See Rossi *et al*. (2012b) for a comprehensive survey on relational representation discovery (for static graph data).

This section first discusses the relational time series data in Section 3.1. The important relational and temporal dependencies are discussed and defined in Section 3.2, whereas Section 3.3 formally defines the key representation discovery tasks for relational time series forecasting.

3.1 Relational time series data

Relational data in the real-world is naturally dynamic—evolving over time with the addition, deletion, and changing of nodes, links, and attributes. Examples of dynamic relational data include social, biological, technological, Web graphs, information networks, among many other types of networks. In particular, dynamic attributed graphs have three main components that vary in time. First, the attribute values (on nodes or links) may change over time (e.g. research area of an author). Next, links might be created and deleted throughout time (e.g. host connections are opened and closed). Finally, nodes might appear and disappear over time (e.g. through activity in an online social network). An intuitive illustration of the underlying dynamics governing relational data is shown in Figure 1.

Definition 1 (Relational Time Series). *Let* $\mathcal{G} = \{\mathbf{G}(t),\ t \in T\}$ denote a relational time series[2], and T denotes the time span of interest. We also define $\mathbf{G}(t) = \langle V(t), E(t), \mathbf{X}^v(t), \mathbf{X}^e(t), \mathbf{Y}(t)\rangle$ *as an attributed network at time* $t \in T$, and $V(t)$ is the set of nodes, $E(t)$ is the set of (possibly directed) edges, each $x_i(t) \in \mathbf{X}^v(t)$ is an attribute vector for node $v_i \in V(t)$, whereas each $x_{ij}(t) \in \mathbf{X}^e(t)$ is an attribute vector for edge $(i, j) \in E(t)$ at time $t \in T$. Further, $\mathbf{Y}(t)$ is the node attribute of interest for prediction where $y_i(t)$ is the prediction attribute value at time t for node $v_i \in V(t)$ and $y_i\,(p{:}t) = \{y_p, \ldots, y_{t-1}, y_t\}$ is the lagged time series attribute vector for node $v_i \in V(t)$.

3.2 Relational and temporal dependencies

In this work, we use relational autocorrelation and along with two temporal dependencies in dynamic attributed networks. More precisely, we observed two fundamental temporal dependencies of dynamic relational network data including the notion of temporal locality and temporal recurrence. We define these temporal dependencies informally below since they apply generally across the full spectrum of temporal-relational information including non-relational attributes, relational node attributes, relational edge attributes, as well as edges and nodes in a graph.

Property 1 (Temporal Locality). *Recent events are more influential to the current state than distant ones.*

This temporal dependency implies that a recent node attribute-value, edge attribute-value, or link is stronger or more predictive of the future than a more distant one. In terms of attribute-values on nodes (or edges) this implies that a recently observed attribute-value (e.g. number of posts) at t is more predictive than past observations at $t-1$ and more distant. Hence, if $x_i(t) = \alpha$ is observed at time t, then at time $t+1$ it is likely that $x_i(t+1) = \alpha$. In the case of edges, this implies that a recently observed edge $(v_i, v_j) \in E_t$ between v_i and v_j at time t implies that there is a high probability of a future edge $(v_i, v_j) \in E_{t+1}$ at $t+1$ will arise.

Property 2 (Temporal Recurrence). *A regular series of observations are more likely to indicate a stronger relationship than an isolated event.*

The notion of temporal recurrence implies that a repeated sequence of observations are more influential or have a higher probability of reappearing in the future than an isolated event. In other words, a repeated or regular sequence of node attribute-values (or edge attribute-values) are more likely to reappear in the future than an isolated node attribute-value. As an example, given a communication network and a node attribute representing the topic of communication for each node in the network, if node v_i has a regular sequence of topics, that is, $x_i(k) = \alpha$, for $k = p, \ldots, t$ across a recent set of timesteps, then there is a higher probability that $x_i(t+1) = \alpha$ is observed than another topic of communication. In terms of edges, temporal recurrence implies that a repeated or recurring series of edges $(v_i, v_j) \in E_k$, for $k = p, \ldots, t$ between v_i and v_j implies a higher probability of a future edge $(v_i, v_j) \in E_{t+1}$ at $t+1$. As an aside, temporal recurrence is based on regular or recurring series of similar observations, whereas temporal locality is based on the notion that the most recent observations are likely to persist in the future.

Learning accurate relational time series representations for nodes in dynamic attributed networks remains a challenge. Just as SRL methods were designed to exploit the relational dependencies in graph data, we instead leverage the relational dependencies *and* the temporal dependencies of the edges, vertices, and attributes to learn more accurate dynamic relational representations.

3.3 Representation tasks

In this paper, we formulate the problem of dynamic relational representation discovery and propose a taxonomy for the dynamic representation tasks shown in Figure 4. More specifically, the dynamic

[2] A time series of relational attributed graph data.

representation tasks for nodes include (i) predicting their label or type, (ii) estimating their weight or importance, and (iii) predicting their existence. We propose methods for each of the dynamic relational node representation tasks in Figure 4 which are defined below.

3.3.1 Dynamic node labeling

Given a time series of attributed graph data, we define the dynamic node labeling problem as the task of learning a time series of node labels $\mathbf{X}_p, \ldots, \mathbf{X}_t$ where for each timestep a given node may be assigned a single label (i.e. class label) or multiple labels (i.e. multiple topics or roles). The time series of labels may represent a known class label previously observed or a latent variable such as roles, topics, among many others.

3.3.2 Dynamic node weighting

Given a time series of attributed graph data, we define the dynamic node weighting representation task as the learning of a time series of weights for the nodes $\mathbf{X}_p, \ldots, \mathbf{X}_t$ that utilize relational and temporal dependencies in the dynamic relational data. The time series of weights may represent the importance or influence of a node in the dynamic attributed network or it may simply represent a latent variable capturing the fundamental dependencies in the dynamic relational data.

3.3.3 Dynamic node prediction

Given a time-series of attributed graph data, we define the dynamic node prediction representation task as the prediction of the existence of a node in a future timestep $t+1$ where the learning leverages past temporal-relational data and more specifically incorporates relational and temporal dependencies in the dynamic relational data. The predicted node may represent a novel type of node, not yet discovered such as a role or topic of communication, or it may be a novel node from an already existing node type such as a Facebook user or group. Most techniques also predict the existence of edges between the predicted node and the set of nodes in the graph.

4 Reinterpreting-related techniques

This section unifies through reinterpretation of a variety of key relational and/or temporal methods for use in the relational time series learning problem. These approaches differ quite drastically in the type of temporal (and/or relational) data used, its characteristics, and key assumptions both within the data and model, as well as the fundamental and underlying task or objective optimized by a particular technique.

4.1 Temporal link representation tasks

While our dynamic relational representation discovery taxonomy shown in Figure 4 focuses on the labeling, weighting, and prediction of nodes, there is also the symmetric dynamic graph representation tasks for links which includes link labeling, link weighting, and link prediction. Our work is not concerned with the link-based dynamic representation tasks as these have been investigated in various contexts (Hasan *et al*., 2006; Liben-Nowell & Kleinberg, 2007; Lassez *et al*., 2008; Acar *et al*., 2009; Koren *et al*., 2009; Xiang *et al*., 2010a; Al Hasan & Zaki, 2011; Menon & Elkan, 2011; Oyama *et al*., 2011; Schall, 2014; Nguyen *et al*., 2018). For instance, link prediction and weighting has been used to improve search engines (Lassez *et al*., 2008), recommendation systems (Koren, 2010) for both products (Koren *et al*., 2009; Xiang *et al*., 2010a) and friends (i.e. social recommendation) (Ma *et al*., 2008), among many others (Chen *et al*., 2013; Liu *et al*., 2013; Li *et al*., 2014). We also note that other work has focused on predicting links in temporal networks using tensor factorizations (Dunlavy *et al*., 2011) and predicting structure in these networks using frequent subgraphs (Lahiri & Berger-Wolf, 2007).

4.2 Temporal centrality and analysis

Recently, there has been a lot of work on analyzing dynamic or temporal graphs which has focused solely on edges that change over time, and has ignored and/or discarded any attributes (both dynamic or static)

(Bhadra & Ferreira, 2003; Xuan *et al.*, 2003; Leskovec *et al.*, 2007a, 2007b; Lahiri & Berger-Wolf, 2008; Tang *et al.*, 2009; Tang *et al.*, 2010; Kovanen *et al.*, 2011; Redmond *et al.*, 2012; Rossi *et al.*, 2012a). Centrality measures have also been extended for temporal networks (Tang *et al.*, 2010; Holme & Saramäki, 2012). While the vast majority of this work has focused only on dynamic edges (i.e. dynamic/temporal/streaming graphs), we instead focus on dynamic relational data and incorporate the full spectrum of dynamics including edges, vertices, and attributes (and their static counterparts as well).

4.3 Time series analysis

Last section discussed temporal graph analysis which lacked attribute data, whereas non-relational attribute-based time series data (Clements & Hendry, 1998; Croushore & Stark, 2001; Marcellino *et al.*, 2006) is the focus of this section. In particular, traditional time series methods ignore graph data all together (Pindyck & Rubinfeld, 1981; Brockwell & Davis, 2002; Ahmed *et al.*, 2010; Box *et al.*, 2013), and focus solely on modeling a time-dependent sequence of real-valued data such as hourly temperatures or economic data such as stock price or gross domestic product (Clements & Hendry, 1998; Croushore & Stark, 2001; Marcellino *et al.*, 2006). In contrast, our proposed methods naturally allow for modeling time series of attributes and graphs (i.e. relational time series data) where each node and edge may have a multi-dimensional time series with arbitrary connectivity or dependencies between them as shown in Figure 5.

At the intersection of time series analysis and ML, Ahmed *et al.* (2010) recently used ML methods such as Neural Networks (Hornik *et al.*, 1989) and support vector machines (Hearst *et al.*, 1998) for time series forecasting. In particular, the authors found that many of these ML methods offered significant improvements over the traditional time series models (Ahmed *et al.*, 2010) such as auto-regressive models and the ilk (Brockwell & Davis, 2002). The main contribution of the work by Ahmed *et al.* (2010) was there use of traditional ML methods for time series forecasting, which has recently attracted numerous follow-up studies (Agami *et al.*, 2009; Ben Taieb *et al.*, 2012; Esling & Agon, 2012). From that perspective, our work makes a similar contribution as we formulate the problem of relational time series learning for dynamic relational graph data, and propose techniques for relational time series classification and regression, which are shown to improve over traditional relational learning and time series methods.

4.4 Relational learning

The majority of research in relational learning has focused on modeling static snapshots or aggregated data (Chakrabarti *et al.*, 1998; Domingos & Richardson, 2001) and has largely ignored the utility of learning and incorporating temporal dynamics into relational representations. Previous work in relational learning on attributed graphs either uses static network snapshots or significantly limits the amount of temporal information incorporated into the models. Sharan and Neville (2008) assumes a strict representation that only uses kernel estimation for link weights, while genetic algorithm enhanced time varying relational

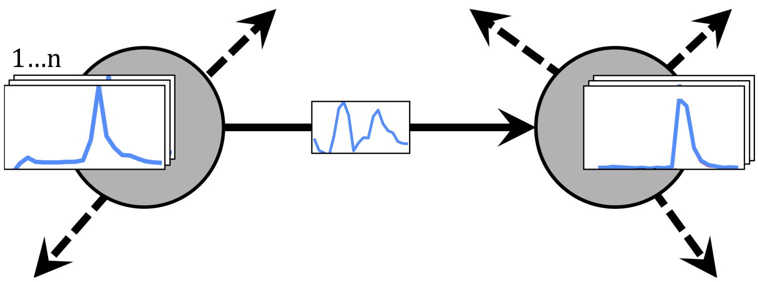

Figure 5 Data model for relational time-series learning. Each node in the network may have an arbitrary number of time series attributes such as blood pressure, number of hourly page views, etc. Further, we also assume that each edge in the network may have an arbitrary number of time series attributes as well (not shown for simplicity). For each edge, we also model the temporal dependencies, resulting in the time series on the edge in the illustration above. As an aside, if there are multiple types of edges between nodes such as friend-edges and email-edges representing friendship between two individuals and email communications, respectively, then we would model the temporal dependencies for each of the edge types resulting in learning multiple time series of edge temporal edge strengths.

classifier (Güneş *et al*., 2011) uses a genetic algorithm to learn the link weights. Spatial-relational probability trees (McGovern *et al*., 2008) incorporate temporal and spatial information in the relational attributes. However, the above approaches focus only on one specific temporal pattern and do not consider different temporal granularities (i.e. they use all available snapshots and lack the notion of a lagged time-series). In contrast, we explore a larger space of temporal-relational representations in a flexible framework that can capture temporal dependencies over *links*, *attributes*, and *nodes*. To the best of our knowledge, we are the first to leverage the full spectrum of dynamic relational data to improve predictions.

We are also the first to propose and investigate *temporal-relational ensemble methods* for time-varying relational classification. However, there has been recent work on relational ensemble methods (Preisach & Schmidt-Thieme, 2006, 2008; Eldardiry & Neville, 2011) and non-relational ensemble methods for evolving streams (Bifet *et al*., 2009). While none of the past work proposes *temporal-relational ensemble methods* for classification, there has been recent work on relational ensemble methods (Preisach & Schmidt-Thieme, 2006, 2008; Eldardiry & Neville, 2011). In particular, Preisach and Schmidt-Thieme (2006) use voting and stacking methods to combine relational data with multiple relations, whereas Eldardiry and Neville (2011) incorporates prediction averaging in the collective inference process to reduce both learning and inference variance.

4.5 Deep learning

Our work is also related to the ML topic of deep learning (LeCun *et al*., 1998; Hinton *et al*., 2006; Marc'Aurelio Ranzato *et al*., 2007; Bengio, 2009; Lee *et al*., 2009; Boureau *et al*., 2010; Deng & Li, 2013), which has recently received a considerable amount of attention from industry due to its success in a variety of real-world applications and systems (Salakhutdinov & Hinton, 2009; Lezama *et al*., 2011; Couprie *et al*., 2013). However, nearly all of this work has focused on images and other similar types of data, whereas we focus on dynamic attributed networks. In view of our work, deep learning for dynamic relational data are informally any method that constructs a representation with varying levels of abstraction or granularity with dependencies between the various layers. For instance, the Dynamic Role Mixed Membership Model (Rossi, 2015) method for node prediction first learns a large set of features, then we discover roles from those features using matrix factorization (i.e. capturing the essence of that set of features), and finally we model the role transitions over time. These representations form a hierarchy of layers each capturing a different level of granularity in the dynamic attributed networks.

5 Challenges and open problems

A discussion of future challenges and directions are discussed below.

5.1 Space and time characterization

Future work should also investigate the tradeoff between space and time. Characterizing thee tradeoffs are challenging, for instance, relational time-series learning methods may learn a model using less data by considering only the most recent observations, whereas relational learning approaches, that ignore temporal information, use all available data. Moreover, modeling temporal dependencies may also lead to simpler/ more accurate models, and more efficient learning and inference algorithms. However, relational time series models typically require an appropriate temporal granularity and kernel function, and learning both of these automatically may be costly.

5.2 Theoretical analysis

There has yet to be any significant theoretical analysis of the existing relational time series forecasting models, despite the fundamental importance of understanding the theoretical limitations of these models. Furthermore, future work should propose simple theoretical models that aid in the theoretical characterization of existing work.

5.3 *Synthetic relational time series graphs*

There have yet to be any synthetic graph models for generating relational time series graph data. However, these models would help evaluate existing approaches and characterize their limitations.

5.4 *Ensemble techniques*

While there has been some work in ensemble techniques for relational time series forecasting (Rossi & Neville, 2012), there has yet to be any systematic investigation into the performance of these across a variety of dynamic networks. More work is needed to identify which types of dynamic networks perform best, among other characteristics.

5.5 *Automatic kernel function learning*

Future work should investigate the related problem of learning the kernel function automatically. While this work investigated a range of kernel functions and found the exponential to work best in most situations, we expect that for certain relational time series data, such an approach is likely to result in a significantly better predictive model. Moreover, it also makes it easier for applying the relational time series learning (for many real-world tasks), without requiring much effort on the part of the user, in terms of knowledge and assumptions about the data. However, techniques proposed in the future must address the challenges associated with the computational cost of such an approach and carefully investigate the benefits (both theoretically and empirically).

5.6 *Robustness to noise*

Another important problem is to investigate the ability of relational time series learning methods to handle varying levels of noise in both the relational and temporal information? Further, does modeling the temporal dependencies reduce the impact of noise, specifically, when the relational data are noisy (e.g. missing or erroneous links)?

5.7 *Model selection*

Selecting the appropriate relational time series model is of fundamental importance. However, there has only been a few techniques and investigations into model learning and more generally hyperparameter optimization for relational time series models (Rossi, 2015). Thus, more work is needed to understand the advantages and disadvantages of each approach, and new techniques developed to overcome problems and/or shortcomings of the existing work.

6 Conclusion

This paper introduces and surveys work related to the problem of *relational time series forecasting*. In particular, this includes the fundamental relational time series prediction tasks: (i) predicting discrete class labels (classification), and (ii) predicting a future real-valued continuous weight (regression). To that end, we introduced a unifying taxonomy that serves as a foundation for studying the main representation tasks that arise in dynamic attributed graph data (for improving relational time series forecasting). This includes the representation tasks of dynamic node and link labeling, weighting, predicting their existence, and discovering important features. Existing work is then surveyed and reinterpreted for the problem of relational time series forecasting. In particular, we discuss how these techniques, which were originally proposed for other ML tasks, can be reinterpreted and used for improving relational time series forecasting. This paper serves as a basis for investigating the important and challenging problem of relational time series forecasting.

Acknowledgements

The authors thank all the reviewers for many helpful suggestions and feedback.

References

Acar, E., Dunlavy, D. & Kolda, T. 2009. Link prediction on evolving data using matrix and tensor factorizations. In *Proceedings of the 9th IEEE International Conference on Data Mining Workshops*, 262–269.

Agami, N., Atiya, A., Saleh, M. & El-Shishiny, H. 2009. A neural network based dynamic forecasting model for trend impact analysis. *Technological Forecasting and Social Change* **76**(7), 952–962.

Ahmed, N., Atiya, A., El Gayar, N. & El-Shishiny, H. 2010. An empirical comparison of machine learning models for time series forecasting. *Econometric Reviews* **29**(5–6), 594–621.

Albert, R., Jeong, H. & Barabási, A. 1999. Internet: diameter of the world-wide web. *Nature* **401**(6749), 130–131.

Al Hasan, M. & Zaki, M. J. 2011. A survey of link prediction in social networks. In *Social Network Data Analytics*, 243–275. Springer.

Anderson, J. R., Michalski, R. S., Michalski, R. S., Carbonell, J. G. & Mitchell, T. M. 1986. *Machine Learning: An Artificial Intelligence Approach*, **2**. Morgan Kaufmann.

Bengio, Y. 2009. Learning deep architectures for AI. *Foundations and Trends in Machine Learning* **2**(1), 1–127.

Ben Taieb, S., Bontempi, G., Atiya, A. F. & Sorjamaa, A. 2012. A review and comparison of strategies for multi-step ahead time series forecasting based on the NN5 forecasting competition. *Expert Systems with Applications* **39**(8), 7067–7083.

Bhadra, S. & Ferreira, A. 2003. Complexity of connected components in evolving graphs and the computation of multicast trees in dynamic networks. In *Ad-Hoc, Mobile, and Wireless Networks*, 259–270.

Bifet, A., Holmes, G., Pfahringer, B., Kirkby, R. & Gavaldà, R. 2009. New ensemble methods for evolving data streams. In *Proceeding of the 15th ACM SIGKDD International Conference on Knowledge Discovery in Data Mining*, 139–148.

Bishop, C. M. 2006. *Pattern Recognition and Machine Learning*. Springer.

Bock, J., Cooray, A., Hanany, S., Keating, B., Lee, A., Matsumura, T., Milligan, M., Ponthieu, N., Renbarger, T. & Tran, H. 2008. The experimental probe of inflationary cosmology (EPIC): a mission concept study for NASA's Einstein inflation probe. *arXiv:0805.4207*.

Boureau, Y.-L., Bach, F., LeCun, Y. & Ponce, J. 2010. Learning mid-level features for recognition. In *IEEE Conference on Computer Vision and Pattern Recognition*, 2559–2566.

Box, G. E., Jenkins, G. M. & Reinsel, G. C. 2013. *Time Series Analysis: Forecasting and Control*. John Wiley & Sons.

Brockwell, P. J. & Davis, R. A. 2002. *Introduction to Time Series and Forecasting*, 1. Taylor & Francis.

Broder, A., Kumar, R., Maghoul, F., Raghavan, P., Rajagopalan, S., Stata, R., Tomkins, A. & Wiener, J. 2000. Graph structure in the web. *Computer Networks* **33**(1–6), 309–320.

Bunke, H. & Kraetzl, M. 2004. Classification and detection of abnormal events in time series of graphs. In Mark Last, Abraham Kandel, Horst Bunke, *Data Mining in Time Series Databases*, Last, M., Kandel, A. & Bunke H (eds). World Scientific, 127–148.

Camacho, J., Guimerà, R. & Nunes Amaral, L. 2002. Robust patterns in food web structure. *Physical Review Letters* **88**(22), 228102: 1–4.

Chakrabarti, S., Dom, B. & Indyk, P. 1998. Enhanced hypertext categorization using hyperlinks. In *Proceedings of the ACM SIGMOD International Conference on Management of Data*, 307–318.

Chandola, V., Banerjee, A. & Kumar, V. 2009. Anomaly detection: a survey. *ACM Computing Surveys* **41**(3), 15.

Chen, C., Yin, H., Yao, J. & Cui, B. 2013. TeRec: a temporal recommender system over tweet stream. *Proceedings of the VLDB Endowment* **6**(12), 1254–1257.

Clements, M. & Hendry, D. 1998. *Forecasting Economic Time Series*. Cambridge University Press.

Couprie, C., Farabet, C. & LeCun, Y. 2013. Causal graph-based video segmentation. *arXiv:1301.1671*.

Couprie, C., Farabet, C., LeCun, Y., & Najman, L. 2013, September. Causal graph-based video segmentation. In Image Processing (ICIP), 2013 20th IEEE International Conference on (pp. 4249–4253). IEEE.

Craven, M., DiPasquo, D., Freitag, D., McCallum, A., Mitchell, T., Nigam, K. & Slattery, S. 1998. Learning to extract symbolic knowledge from the World Wide Web. In *Proceedings of the 15th AAAI Conference on Artificial Intelligence*, 509–516.

Croushore, D. & Stark, T. 2001. A real-time data set for macroeconomists. *Journal of Econometrics* **105**(1), 111–130.

Das Sarma, A., Gollapudi, S. & Panigrahy, R. 2008. Estimating PageRank on graph streams. In *Proceedings of the ACM SIGMOD International Conference on Management of Data*, 69–78.

Deng, L. & Li, X. 2013. Machine learning paradigms for speech recognition: an overview. *Transactions on Audio, Speech and Language Processing* **21**(5), 1060–1089.

De Raedt, L. & Kersting, K. 2008. *Probabilistic Inductive Logic Programming*. Springer-Verlag.

Domingos, P. & Richardson, M. 2001. Mining the network value of customers. In *Proceeding of the 7th ACM SIGKDD International Conference on Knowledge Discovery in Data Mining*, 57–66.

Dunlavy, D. M., Kolda, T. G. & Acar, E. 2011. Temporal link prediction using matrix and tensor factorizations. *Transactions on Knowledge Discovery from Data* **5**(2), 10:1–10:27.

Dunne, J., Williams, R. & Martinez, N. 2002. Food-web structure and network theory: the role of connectance and size. *Proceedings of the National Academy of Sciences of the United States of America* **99**(20), 12917.

Einstein, A. 1906. Zur theorie der brownschen bewegung. *Annalen der Physik* **324**(2), 371–381.

Eldardiry, H. & Neville, J. 2011. Across-model collective ensemble classification. In *Proceedings of the 25th AAAI Conference on Artificial Intelligence*, 343–349.

Esling, P. & Agon, C. 2012. Time-series data mining. *ACM Computing Surveys* **45**(1), 12.

Faloutsos, M., Faloutsos, P. & Faloutsos, C. 1999. On power-law relationships of the internet topology. In *Proceedings of the ACM SIGCOMM International Conference on Applications, Technologies, Architectures, and Protocols for Computer Communication*, 251–262.

Friedman, N., Getoor, L., Koller, D. & Pfeffer, A. 1999. Learning probabilistic relational models. In *Proceedings of the 16th International Joint Conference on Artificial Intelligence*, 1300–1309. Springer-Verlag.

Getoor, L. & Taskar, B. (eds) 2007. *Introduction to Statistical Relational Learning*. MIT Press.

Güneş, İ, Çataltepe, Z. & Öğüdücü, Ş. G. 2011. GA-TVRC: a novel relational time varying classifier to extract temporal information using genetic algorithms. In *Machine Learning and Data Mining in Pattern Recognition*, 568–583. Springer.

Hasan, M. A., Chaoji, V., Salem, S. & Zaki, M. 2006. Link prediction using supervised learning. In *Proceedings of the SDM Workshop on Link Analysis, Counterterrorism and Security*.

Hearst, M. A., Dumais, S., Osman, E., Platt, J. & Scholkopf, B. 1998. Support vector machines. *Intelligent Systems and their Applications* **13**(4), 18–28.

Hinton, G. E., Osindero, S. & Teh, Y.-W. 2006. A fast learning algorithm for deep belief nets. *Neural Computation* **18**(7), 1527–1554.

Holme, P. & Saramäki, J. 2012. Temporal networks. *Physics Reports*.

Hornik, K., Stinchcombe, M. & White, H. 1989. Multilayer feedforward networks are universal approximators. *Neural Networks* **2**(5), 359–366.

Ide, T. & Kashima, H. 2004. Eigenspace-based anomaly detection in computer systems. In *Proceedings of the 10th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 440–449.

Jensen, D., Neville, J. & Gallagher, B. 2004. Why collective inference improves relational classification. In *Proceedings of the 10th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 593–598.

Jeong, H., Mason, S. P., Barabási, A. L. & Oltvai, Z. N. 2001. Lethality and centrality in protein networks. *Nature*, **411** (6833), 41.

Jeong, H., Mason, S., Barabasi, A. & Oltvai, Z. 2001. Lethality and centrality in protein networks. *arXiv preprint cond-mat/0105306*.

Jeong, H., Tombor, B., Albert, R., Oltvai, Z. & Barabási, A. 2000. The large-scale organization of metabolic networks. *Nature* **407**(6804), 651–654.

Kleczkowski, A. & Grenfell, B. 1999. Mean-field-type equations for spread of epidemics: the small world model. *Physica A: Statistical Mechanics and its Applications* **274**(1–2), 355–360.

Koren, Y. 2010. Collaborative filtering with temporal dynamics. *Communications of the ACM* **53**(4), 89–97.

Koren, Y., Bell, R. & Volinsky, C. 2009. Matrix factorization techniques for recommender systems. *IEEE Computer* **42**(8), 30–37.

Kovanen, L., Karsai, M., Kaski, K., Kertész, J. & Saramäki, J. 2011. Temporal motifs in time-dependent networks. *Journal of Statistical Mechanics: Theory and Experiment* **2011**(11), P11005.

Krebs, V. 2002. Mapping networks of terrorist cells. *Connections* **24**(3), 43–52.

Lahiri, M. & Berger-Wolf, T. 2008. Mining periodic behavior in dynamic social networks. In *Proceedings of the 8th IEEE International Conference on Data Mining*, 373–382.

Lahiri, M. & Berger-Wolf, T. Y. 2007. Structure prediction in temporal networks using frequent subgraphs. In *IEEE Symposium on Computational Intelligence and Data Mining*, 35–42.

Lassez, J.-L., Rossi, R. & Jeev, K. 2008. Ranking links on the web: search and surf engines. In *Proceedings of the IEA/AIE International Conference*, 199–208. Springer.

LeCun, Y., Bottou, L., Bengio, Y. & Haffner, P. 1998. Gradient-based learning applied to document recognition. *Proceedings of the IEEE* **86**(11), 2278–2324.

Lee, H., Grosse, R., Ranganath, R. & Ng, A. Y. 2009. Convolutional deep belief networks for scalable unsupervised learning of hierarchical representations. In *Proceedings of the 26th International Conference on Machine Learning*, 609–616.

Leskovec, J., Adamic, L. & Huberman, B. 2007a. The dynamics of viral marketing. *Transactions on the Web* **1**(1), 1–39.

Leskovec, J., Kleinberg, J. & Faloutsos, C. 2007b. Graph evolution: densification and shrinking diameters. *Transactions on Knowledge Discovery from Data* **1**(1), 1–41. `https://dl.acm.org/citation.cfm?id=1217301`.

Lezama, J., Alahari, K., Sivic, J. & Laptev, I. 2011. Track to the future: spatio-temporal video segmentation with long-range motion cues. In *IEEE Conference on Computer Vision and Pattern Recognition*.

Li, L., Zheng, L., Yang, F. & Li, T. 2014. Modeling and broadening temporal user interest in personalized news recommendation. *Expert Systems with Applications* **41**(7), 3168–3177.
Liben-Nowell, D. & Kleinberg, J. 2007. The link-prediction problem for social networks. *Journal of the American Society for Information Science and Technology* **58**(7), 1019–1031.
Liu, N. N., He, L. & Zhao, M. 2013. Social temporal collaborative ranking for context aware movie recommendation. *ACM Transactions on Intelligent Systems and Technology* **4**(1), 15.
Lu, Q. & Getoor, L. 2003. Link-based classification. In *Proceedings of the 20th International Conference on Machine Learning*, 496–503.
Ma, H., Yang, H., Lyu, M. R. & King, I. 2008. SoRec: social recommendation using probabilistic matrix factorization. In *Proceedings of the 17th ACM Conference on Information and Knowledge Management*, 931–940.
Macskassy, S. & Provost, F. 2003. A simple relational classifier. In *Proceedings of the SIGKDD 2nd Workshop on Multi-Relational Data Mining*, 64–76.
Marc'Aurelio Ranzato, Y., Boureau, L. & LeCun, Y. 2007. Sparse feature learning for deep belief networks. *Advances in Neural Information Processing Systems* **20**, 1185–1192.
Marcellino, M., Stock, J. H. & Watson, M. W. 2006. A comparison of direct and iterated multistep are methods for forecasting macroeconomic time series. *Journal of Econometrics* **135**(1), 499–526.
Maslov, S. & Sneppen, K. 2002. Specificity and stability in topology of protein networks. *Science* **296**(5569), 910–913.
May, R. & Lloyd, A. 2001. Infection dynamics on scale-free networks. *Physical Review E* **64**(6), 66112.
McDowell, L., Gupta, K. & Aha, D. 2010. Meta-prediction for collective classification. In *Proceedings of the 23rd International FLAIRS Conference*.
McDowell, L. K., Gupta, K. M. & Aha, D. W. 2009. Cautious collective classification. *Journal of Machine Learning Research* **10**, 2777–2836.
McGovern, A., Collier, N., Matthew Gagne, I., Brown, D. & Rodger, A. 2008. Spatiotemporal relational probability trees: an introduction. In *Proceedings of the 8th IEEE International Conference on Data Mining*, 935–940.
McGovern, A., Friedland, L., Hay, M., Gallagher, B., Fast, A., Neville, J. & Jensen, D. 2003. Exploiting relational structure to understand publication patterns in high-energy physics. *SIGKDD Explorations* **5**(2), 165–172.
McPherson, M., Smith-Lovin, L. & Cook, J. M. 2001. Birds of a feather: homophily in social networks. *Annual Review of Sociology* **27**(1), 415–444.
Menon, A. & Elkan, C. 2011. Link prediction via matrix factorization. In *Proceedings of the European Conference on Machine Learning and Principles and Practice of Knowledge Discovery in Databases*, 437–452.
Moore, C. & Newman, M. 2000. Epidemics and percolation in small-world networks. *Physical Review E* **61**(5), 5678–5682.
Neville, J., Jensen, D. & Gallagher, B. 2003. Simple estimators for relational Bayesian classifers. In *Proceedings of the 3rd IEEE International Conference on Data Mining*, 609–612.
Neville, J., Simsek, O., Jensen, D., Komoroske, J., Palmer, K. & Goldberg, H. 2005. Using relational knowledge discovery to prevent securities fraud. In *Proceedings of the 11th ACM SIGKDD International Conference on Knowledge Discovery in Data Mining*, 449–458.
Newman, M. 2001. The structure of scientific collaboration networks. *Proceedings of the National Academy of Sciences* **98**(2), 404–409.
Newman, M., Barabasi, A.-L. & Watts, D. J. 2006. *The Structure and Dynamics of Networks*. Princeton University Press.
Nguyen, G. H., Lee, J. B., Rossi, R. A., Ahmed, N. K., Koh, E. & Kim, S. 2018. Continuous-time dynamic network embeddings. In *3rd International Workshop on Learning Representations for Big Networks (WWW BigNet)*.
Noble, C. & Cook, D. 2003. Graph-based anomaly detection. In *Proceedings of the 9th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 631–636.
O'Madadhain, J. & Smyth, P. 2005. EventRank: a framework for ranking time-varying networks. In *Proceedings of the LinkKDD Workshop*, 9–16.
Oyama, S., Hayashi, K. & Kashima, H. 2011. Cross-temporal link prediction. In *Proceedings of the 11th International Conference on Data Mining*, 1188–1193.
Pastor-Satorras, R. & Vespignani, A. 2001. Epidemic spreading in scale-free networks. *Physical Review Letters* **86**(14), 3200–3203.
Pindyck, R. S. & Rubinfeld, D. L. 1981. *Econometric Models and Economic Forecasts*, **2**. McGraw-Hill New York.
Preisach, C. & Schmidt-Thieme, L. 2006. Relational ensemble classification. In *Proceedings of the 6th International Conference on Data Mining*, 499–509.
Preisach, C. & Schmidt-Thieme, L. 2008. Ensembles of relational classifiers. *Knowledge and Information Systems* **14**(3), 249–272.
Redmond, U., Harrigan, M. & Cunningham, P. 2012. Identifying time-respecting subgraphs in temporal networks. In *Proceedings of the 3rd International Workshop on Mining Ubiquitous and Social Environments*, 51–63.
Rossi, R. A. 2014. Fast triangle core decomposition for mining large graphs. In *Advances in Knowledge Discovery and Data Mining*, **8443**, 310–322.

Rossi, R. A. 2015. *Improving Relational Machine Learning by Modeling Temporal Dependencies*. PhD thesis, Purdue University.

Rossi, R. A., Gallagher, B., Neville, J. & Henderson, K. 2013a. Modeling dynamic behavior in large evolving graphs. In *Proceedings of the Sixth ACM International Conference on Web Search and Data Mining*, 667–676.

Rossi, R. A. & Gleich, D. 2012. Dynamic PageRank using evolving teleportation. *Algorithms and Models for the Web Graph* **7323**, 126–137.

Rossi, R. A., Gleich, D. & Gebremedhin, A. 2013b. Triangle core decomposition and maximum cliques. In *SIAM Network Science Workshop*, 1–2.

Rossi, R. A., Gleich, D. F., Gebremedhin, A. H. & Patwary, M. A. 2012a. What if clique were fast? Maximum cliques in information networks and strong components in temporal networks. *arXiv:1210.5802*, 1–11.

Rossi, R. A., Gleich, D. F., Gebremedhin, A. H. & Patwary, M. A. 2013c. A fast parallel maximum clique algorithm for large sparse graphs and temporal strong components. *arXiv:1302.6256*, 1–9.

Rossi, R. A., McDowell, L. K., Aha, D. W. & Neville, J. 2012b. Transforming graph data for statistical relational learning. *Journal of Artificial Intelligence Research* **45**, 363–441.

Rossi, R. A. & Neville, J. 2010. Modeling the evolution of discussion topics and communication to improve relational classification. In *Proceedings of the ACM SIGKDD 1st Workshop on Social Media Analytics*, 89–97.

Rossi, R. A. & Neville, J. 2012. Time-evolving relational classification and ensemble methods. In *Advances in Knowledge Discovery and Data Mining* **7301**, 1–13. Springer.

Salakhutdinov, R. & Hinton, G. E. 2009. Deep Boltzmann machines. In *International Conference on Artificial Intelligence and Statistics*, 448–455.

Schall, D. 2014. Link prediction in directed social networks. *Social Network Analysis and Mining* **4**(1), 1–14.

Sharan, U. & Neville, J. 2008. Temporal-relational classifiers for prediction in evolving domains. In *Proceedings of the 8th IEEE International Conference on Data Mining*, 540–549.

Tang, J., Musolesi, M., Mascolo, C. & Latora, V. 2009. Temporal distance metrics for social network analysis. In *Proceedings of the 2nd ACM Workshop on Online Social Networks*, 31–36.

Tang, J., Musolesi, M., Mascolo, C., Latora, V. & Nicosia, V. 2010. Analysing information flows and key mediators through temporal centrality metrics. In *Proceedings of the 3rd Workshop on Social Network Systems*, 1–6.

Tong, H. & Lin, C. 2011. Non-negative residual matrix factorization with application to graph anomaly detection. In *Proceedings of the 7th SIAM International Conference on Data Mining*.

Wagner, A. & Fell, D. 2001. The small world inside large metabolic networks. *Proceedings of the Royal Society of London. Series B: Biological Sciences* **268**(1478), 1803–1810.

Watts, D. J. & Strogatz, S. H. 1998. Collective dynamics of 'small-world' networks. *Nature* **393**(6684), 440–442.

Xiang, L., Yuan, Q., Zhao, S., Chen, L., Zhang, X., Yang, Q. & Sun, J. 2010a. Temporal recommendation on graphs via long-and short-term preference fusion. In *Proceedings of the 16th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 723–732.

Xiang, R., Neville, J. & Rogati, M. 2010b. Modeling relationship strength in online social networks. In *Proceedings of the 19th International World Wide Web Conference*, 981–990.

Xuan, B., Ferreira, A. & Jarry, A. 2003. Computing shortest, fastest, and foremost journeys in dynamic networks. *International Journal of Foundations of Computer Science* **14**(2), 267–285.