Learning Time-Varying Networks

Manuel Llamas, ECS Department, University of Southampton

***Abstract*— Time-varying networks is recent field developed to explain and model the evolution of networks in different subjects aiming to understand how, when and why the networks evolve and even leading to predictive models of the phenomena tracked in the network. In network science distinct elements or actors are represented by nodes (or vertices) and the connections between the elements or actors as links (or edges) allowing us to study from how a single element can affect the whole network to even predict which topological changes indicate coming critical events or understand how a network can reshape to minimize contagion between its elements. All of this provides a powerful tool for studying financial networks, giving us a framework to track the evolution of the stock market and understand its behavior against critical events.**

***Index Terms*— Bayesian Network, Dynamic Networks, Random Graphs, Scale-free Networks, TESLA algorithm.**

I. Introduction

In many problems it is helpful and even necessary to study the relationships between entities of a system since we find them interconnected forming networks in which some small changes on an element can propagate and affect the entire system. To do so we use graph, and more precisely networks, theories which represent each individuals or actors by nodes and the connection or relationships as edges between the nodes. We find a large list of examples in which it is necessary to apply network theory on different and wide subjects as social science, finance, biology or climatology.

Studying such networks can reveal lots of information like how an event can trigger a topological change of the entire network, how entities of the network can depend on each other's states or share similar properties and organize themselves into groups or which individuals are the main agents that lead the evolution of the network.

Classically these analysis have been carried in the simplest way assuming static (non-varying) networks, however this is usually a simplification since in most of the cases the relationships between agents in a network evolve over time as the nodes' states change. The internet, biological processes or the economy are some examples of networks evolving over time, shrinking or creating new links between the nodes.

Although there is wide literature and methods on modelling static networks it has not been until recent years that interest has grown on studying dynamic or time-varying networks ([1], [3], [5], [6], [7], [8], [11]).

Usually, only time series measurements, such as microarray, stock price, etc., of the activity of the nodal entities, but not their linkage status, are available. The goal is to recover the latent time-varying networks with temporal resolution up to every single time point based on time series measurements.

To be able to understand the current state of the research on the field and the strengths, weakness and challenges of the techniques proposed this work is going to analyze several algorithms, focusing mainly on TESLA (from the acronym TESLLOR, which stands for temporally smoothed *l*1-regularized logistic regression), built to recover the structure of time-varying networks over a fixed set of nodes from their time series nodal attributes [1].

II. Network Main Properties

Depending on the problem under consideration, it may be convenient to study the network from different points of view, focusing on relationships between the entities (edge-centric), one or more entities (vertex-centric) or the topological properties of the network as a whole (graph-centric) [5].

When studying edge-centric networks some important measures include transitivity (similar to the mathematical property: if u connected to v and v connected to w, then u connected to w) and reciprocity (when we find loops in a network with 2 edges length, so pairs of nodes that have edges running in both directions).

Meanwhile in vertex-centric networks centrality measures are used to quantify the relevance of the different nodes in a network. The most basic example is the degree of centrality of the nodes: the number of edges attached to them. Nodes with high degree are called hubs and obviously play an important role in the system.

Regarding graph-centric approach we are able to analyze the general shape by studying the clusters in which the vertices organize themselves, many techniques like the well-known hierarchical clustering are useful in this task showing important properties of the network like its tendency to follow an assortative or dissortative mixing pattern.

All these properties [9] give us important information to understand the behavior of the system.

III. Dynamic Networks

Latest years the science of networks developed three important aspects that define this field nowadays: it is used to model real-world problems, it frequently assumes the networks are not static and it aims to understand networks not just as topological objects but also as the framework upon distributed dynamical systems are built [3].

Then studying networks as dynamical systems is the best way to understand evolving underlying processes in complex systems that can trigger critic events reshaping the network and changing its properties, one example of this is how a disease spreads: the structure of a network through which a contagious agent is transmitted can have a dramatic impact on outcomes at the level of entire populations [3].

IV. LEARNING NETWORKS

Looking at their topology we find that networks are organized in two main classes: scale-free networks and random graphs.

Scale-free network term was coined to describe the networks that show a power law degree distribution (distribution of the connections over the whole graph) after Barabasi mapped a portion of the World Wide Web [2] finding that it followed this distribution of the links and that some nodes had many more links than others (hubs). Then the most notable characteristic in scale-free networks is that it is relatively common to find vertices with a much higher degree (links) than the average. They also have other important features like robustness due to its hierarchical organization or clustering coefficient that may help to understand networks following this distribution.

A random graph is obtained by starting with a set of isolated vertices and adding successive edges between them at random. The aim of the study in this field is to determine at what stage a particular property of the graph is likely to arise. Different random graph models produce different probability distributions on graphs [4].

A. Exponential Random Graph Model

Based on this last one, the exponential random graph model (ERGM) has been extensively used to model static networks and also used as a base for two of the first algorithms proposed to study dynamic networks, the temporal exponential random graph model (tERGM) and the hidden tERGM for modelling a sequence of node attribute [7]. When tERGM assume that the sequence of networks is available, htERGM explores the possible dependencies of unobserved rewiring networks and leads to the algorithm that can reconstruct such networks from a snapshots' sequence of nodal attributes.

Although this algorithms overcame approaches that recover single time-invariant networks they are not the most useful ones since they depend on unobserved network variables being unable to compute likelihood ratios, needing inference algorithms specifically for each problem.

B. Hidden Markov Dynamic Bayesian Network

Another classical approach are the Bayesian networks, widely used to describe biological systems, risk analysis and even financial networks.

Bayesian networks are graphs whose edges represent conditional dependencies and nodes are random variables in a Bayesian sense (observable quantities, latent variables, unknown parameters or hypotheses) and are associated with probability functions that takes as input the node's parent variables and gives the probability of the variable represented by the node.

Its standard assumption is 'stationarity', and therefore, several research efforts have been recently proposed to relax this restriction. However, those methods suffer from three challenges: long running time, low accuracy and reliance on parameter settings. [11] propose a non-stationary DBN model by extending each hidden node of Hidden Markov Model into a DBN (called HMDBN), which properly handles the underlying time-evolving networks resulting in a promising experimental evaluation of the method, demonstrating more stably high prediction accuracy and significantly improved computation efficiency (even with no prior knowledge and parameter settings) on both synthetic and real biological data.

Although this probabilistic model is more complex than TESLA algorithm it could be a good method to implement in order to compare results and performance.

C. Bayesian non-parametric model

In this line there was proposed a Bayesian non-parametric model including time-varying predictors in dynamic network inference precisely for financial studies [6].

$$\pi_{ij}(t) = \frac{1}{1 + e^{-s_{ij}(t)}}$$

$$s_{ij}(t) = \mu(t) + z_{ij,t}^{T}\beta(t) + x_i(t)^{T}x_j(t)$$

This model computes edge specific predictors where the link probabilities ($s_{ij}(t)$) are estimated via a logistic regression, with $\mu(t)$ a baseline process quantifying the overall propensity to form links in the network across time, $x_i(t)^T x_j(t)$ are vectors containing the latent coordinates favoring a higher link probability when units *i* and *j* have latent coordinates in the same direction and $z_{ij,t} = [z_{ij1,t}, \ldots, z_{ijP,t}]^T$ is a P-dimensional vector of time-varying edge-specific predictors for units *i* and *j* at time t and $\beta(t)$ are the corresponding dynamic coefficients. This allows the proximity between units *i* and *j* at time *t* to depend on predictors in a manner that varies smoothly with time.

The main issue regarding this method is that it assumes time-constant smoothness while in finance and other network applications we expect smoothness to vary over time, however it had some promising results when analyzing the global stock market network in the 2008 crisis and this makes it a good candidate to test in this project.

D. Temporally smoothed l1-regularized logistic regression

TESLA represents an extension of the lasso-style sparse structure recovery technique and is based on a key assumption that temporally adjacent networks are likely not to be dramatically different from each other in topology and therefore are more likely to share common edges than temporally distant networks.

Building on the *l*1-regularized logistic regression algorithm for estimating single sparse networks [10] it was developed a regression regularization scheme that connects multiple time-specific network inference functions via a first-order edge smoothness function that encourages edge retention between time-adjacent networks. An important property of this idea is that it fully integrates all available samples of the entire time series in a single inference procedure, what means an advantage in contrast of htERGM algorithm.

TESLA estimates $\{\hat{\theta}^t\}_{t=1}^{T}$, the correlation (or dependency strength) matrix between the nodes using a time-series of the observed stated of the nodes $\mathbf{x^t}$.

$$\hat{\theta}_i^1, \dots, \hat{\theta}_i^T = \arg\min_{\hat{\theta}_i^1,\dots,\hat{\theta}_i^T} \sum_{t=1}^{T} \ell(\boldsymbol{x}^t; \boldsymbol{\theta}_i^t) + \lambda_1 \sum_{t=1}^{T} \|\boldsymbol{\theta}_i^t\|_1 + \lambda_2 \sum_{t=2}^{T} \|\boldsymbol{\theta}_i^t - \boldsymbol{\theta}_i^{t-1}\|_1, \quad \forall i$$

where,

$$\ell(\boldsymbol{x}^t; \boldsymbol{\theta}_i^t) = \sum_{d=1}^{N^t} (\log\left(1 + \exp(\boldsymbol{x}_{d,-i}^t \boldsymbol{\theta}_i^t)\right) - \boldsymbol{x}_{d,-i}^t \boldsymbol{\theta}_i^t \boldsymbol{x}_{d,i}^t)$$

Where the graph structure is given by the locations of the nonzero elements of the parameter vectors $\boldsymbol{\theta}_i^t$. Components of the vectors $\boldsymbol{\theta}_i^t$ are indexed by distinct pairs of nodes and a component *j* of the vector $\boldsymbol{\theta}_i^t$ is nonzero if and only if the corresponding edge $(i, j) \in E\tau$. $\boldsymbol{x}_{d,-i}^t$ denotes the observed states of all nodes but node *i* and $\ell()$ is the log conditional likelihood of state $\boldsymbol{x}_{d,i}^t$ under a logistic regression model.

First term of the algorithm are p logistic regressions, one for each node with respect the rest of them ($\boldsymbol{x}_{d,-i}^t$ are the states of all nodes but node I in the dth sample in time epoch t).

The second term is a lasso regularization that introduces sparsity and consistency in neighborhood selection, as showed in [10], converging to the true graph structure. Also a sparse graph effectively limits the degree of freedom of the model, which makes structure recovery possible given a small sample size [8].

Since in this problem we are estimating dynamic networks (so more than one single graph), and, as assumed before, temporary adjacent networks are going to be similar one to each other, we need to introduce the third term in the equation that penalizes the discrepancy between time-adjacent parameters θ.

Summarizing, the first term is the main one while the last two terms are the penalties ($L^{shrink} + L^{smooth}$) that are introduced to enforce sparsity and smoothness.

E. TESLA updated

It is important to note that in this specific method we are assuming that $\boldsymbol{\theta}_i^t$ are a piecewise constant function with abrupt changes in parameters (jumps).

However in some cases the changes in parameters are continuous, so further work [8] was carried out to adapt this algorithm including a weighting term that turns $\boldsymbol{\theta}_i^t$ in continuous functions instead of the penalty in the structural changes (L^{smooth}) that assumes small “jumps” between time-adjacent values.

$$\widehat{\boldsymbol{\theta}}_i^\tau = \min_{\boldsymbol{\theta} \in \mathbb{R}^{p-1}} \{\ell(\boldsymbol{\theta}_i; \mathcal{D}_n) + \lambda_1 \|\boldsymbol{\theta}_i^t\|_1\}$$

$$\ell(\boldsymbol{\theta}_i; \mathcal{D}_n) = -\sum_{t \in \tau_n} \omega_t^\tau \gamma(\boldsymbol{\theta}_i; \boldsymbol{x}^t)$$

Where the weights are defined by

$$\omega_t^\tau = \frac{K_h(t-\tau)}{\sum_{t' \in \tau_n} K_h(t'-\tau)}$$

And $K_h()$ is a symmetric nonnegative kernel function.

V. Financial Application

When studying dynamic networks we aim to solve real world problems, so it is usual to look at meta-networks, this is multi-link (many types of links), multi-mode (many types of nodes) and multi-level (nodes can represent subnetworks) networks.

This added to the fact that not many work has been done to recover dynamic networks on finance and the promising algorithms reviewed on this paper sum up enough motivations to build an interesting topic: testing the performance of these different methods in the financial area to find the model that best defines it and extending the previous work done in this subject (analyze the behavior of financial networks against critic events [6]) by looking at meta-networks instead of simple networks.

VI. Conclusion

Summarizing we find crucial studying data through a network lens since it can add substantial new insights. This work evaluated the state-of-the-art methods to recover dynamic networks in order to prepare further research based on applying (mainly) TESLA algorithm in a finance context to study the changes that may warn or even drive significant global events like the 2008 crisis.

It is important to use scalable computational methods that can fit very large networks (huge number of nodes), so models like TESLA that exploit sparsity are promising in this matter.

The two TESLA-based algorithms provide two different ends of the same algorithm: one is tailored toward estimation of structural changes in the model while the other is able to estimate smoothly changing networks. This gives us a great opportunity to study the same field with these two approaches.

References

[1] Ahmed A and Xing E. (2009) Recovering time-varying networks of dependencies in social and biological studies. *PNAS* 106 (29) 11878-11883

[2] Barabási A, Albert R. (1999). Emergence of scaling in random networks. *Science* 286 (5439): 509–512

[3] Barabasi A, Watts D and Newman M. (2011) Structure and Dynamics of Networks. *Princetown University Press*.

[4] Bollobás B. (2001) Random Graphs. 2nd Edition, *Cambridge University Press*.

[5] Casteigts A, Flocchini P, Quattrociocchi W and Santoro N. (2012) Time-Varying Graphs and Dynamic Networks. *Ad-hoc, Mobile, and Wireless Networks*. Volume 6811 of the series Lecture Notes in Computer Science pp 346-359

[6] Durante D. and Dunson D. (2014) Bayesian dynamic financial networks with time-varying predictors. *Statistics & Probability Letters*, 93, 19-26

[7] Guo F, Hanneke S, Fu W and Xing E (2007) Recovering temporally rewiring networks: A model-based approach. *The 24th International Conference of Machine Learning, 2007* (Association for Computing Machinery, New York).

[8] Kolar M, Song L, Ahmed A and Xing, Eric P. (2010) Estimating time-varying networks. *Annals Applied Statistics Vol. 4*. No. 1, 94-123

[9] Newman, M. (2010) Networks: An Introduction. *Oxford: Oxford University Press.*

[10] Wainwright M, Ravikumar P, Lafferty J (2006) High dimensional graphical model selection using l1-regularized logistic regression. *Advances in Neural Information Processing Systems 19 (MIT Press, Cambridge, MA),* pp 1465–1472

[11] Zhu S. and Wang Y. (2015). Hidden Markov induced Dynamic Bayesian Network for recovering time evolving gene regulatory networks. *Scientific Reports*, *5*, 17841.