ARCHIVIO ISTITUZIONALE
DELLA RICERCA

Alma Mater Studiorum Università di Bologna
Archivio istituzionale della ricerca

A dynamic network model with persistent links and node-specific latent variables, with an application to the interbank market

This is the final peer-reviewed author's accepted manuscript (postprint) of the following publication:

Published Version:
Mazzarisi P., Barucca P., Lillo F., Tantari D. (2020). A dynamic network model with persistent links and node-specific latent variables, with an application to the interbank market. EUROPEAN JOURNAL OF OPERATIONAL RESEARCH, 281(1), 50-65 [10.1016/j.ejor.2019.07.024].

Availability:
This version is available at: https://hdl.handle.net/11585/718078 since: 2021-04-06

Published:
DOI: http://doi.org/10.1016/j.ejor.2019.07.024

Terms of use:

Some rights reserved. The terms and conditions for the reuse of this version of the manuscript are specified in the publishing policy. For all terms of use and more information see the publisher's website.

This item was downloaded from IRIS Università di Bologna (https://cris.unibo.it/).
When citing, please refer to the published version.

(Article begins on next page)

05 September 2026

This is the final peer-reviewed accepted manuscript of:

Mazzarisi, P., Barucca, P., Lillo, F., & Tantari, D. (2020). A dynamic network model with persistent links and node-specific latent variables, with an application to the interbank market. *European Journal of Operational Research*, *281*(1), 50-65.

The final published version is available online at :

https://doi.org/10.1016/j.ejor.2019.07.024

Rights / License:

The terms and conditions for the reuse of this version of the manuscript are specified in the publishing policy. For all terms of use and more information see the publisher's website.

This item was downloaded from IRIS Università di Bologna (https://cris.unibo.it/)

When citing, please refer to the published version.

A DYNAMIC NETWORK MODEL WITH PERSISTENT LINKS AND NODE-SPECIFIC LATENT VARIABLES, WITH AN APPLICATION TO THE INTERBANK MARKET

P. MAZZARISI

Department of Mathematics, University of Bologna, Italy
Scuola Normale Superiore, Pisa, Italy

P. BARUCCA

Department of Computer Science, University College London, London, UK

F. LILLO

Department of Mathematics, University of Bologna, Italy
Scuola Normale Superiore, Pisa, Italy

D. TANTARI

DISEI, University of Florence, Firenze, Italy
Scuola Normale Superiore, Pisa, Italy

E-mail addresses: piero.mazzarisi@sns.it (corresponding author, permanent adress: p.mazzarisi@gmail.com), p.barucca@ucl.ac.uk, fabrizio.lillo@unibo.it, daniele.tantari@unifi.it.

Date: February 16, 2021.

Key words and phrases. Temporal networks, Markov processes, Expectation-Maximization, Interbank market, Link prediction.

Abstract. We propose a dynamic network model where two mechanisms control the probability of a link between two nodes: (i) the existence or absence of this link in the past, and (ii) node-specific latent variables (dynamic fitnesses) describing the propensity of each node to create links. Assuming a Markov dynamics for both mechanisms, we propose an Expectation-Maximization algorithm for model estimation and inference of the latent variables. The estimated parameters and fitnesses can be used to forecast the presence of a link in the future. We apply our methodology to the e-MID interbank network for which the two linkage mechanisms are associated with two different trading behaviors in the process of network formation, namely preferential trading and trading driven by node-specific characteristics. The empirical results allow to recognize preferential lending in the interbank market and indicate how a method that does not account for time-varying network topologies tends to overestimate preferential linkage.

Introduction

In recent years there has been a growing interest in the study of complex networks (Newman et al. (2001), Holme and Saramäki (2012)). One of the reasons is that many natural and artificial systems are characterized by the presence of a sparse structure of interactions, i.e. only a small fraction of the possible pairs of elements mutually interact (at least at each time). Thus the topology of the network of interactions plays an important role in understanding the aggregate behavior of many complex systems. Moreover most of the investigated systems evolve over time and the structure of the network is generically not constant but new links are formed and old ones are destroyed at each time. Understanding and modeling network dynamics is therefore of paramount importance in many disciplines as testified by the recent literature (see below for a review).

From a modeling perspective there are several mechanisms that can lead to link formation and destruction. The presence of a link may depend on present node properties of the system but also on previous network states. Consider as an example a trading network such as the interbank network studied in the empirical application below. The probability of a link between two nodes representing a transaction between the corresponding entities depends generically on the current supply and demand of the two entities, as well as on the existence (or absence thereof) of a link in the past between the two entities. The former driver is associated with node-specific properties (supply and demand) which evolve in time with their own dynamics, possibly dependent also on the network state. The latter driver, instead, is associated with a link specific property, namely the persistence of links which describes the tendency to interact with whom we have interacted with in the past.

The objective of this paper is to introduce a dynamic network model where both mechanisms are present and to propose a statistical estimation technique which allows to disentangle the

importance of the two mechanisms for each link in the network. The estimation method is based on an Expectation-Maximization scheme for maximum likelihood estimation. As we will show, the estimation of the model allows also to forecast the probability of the existence of a specific link in the future when the model parameters and the past network history are known.

More specifically, in our model we define a Markovian process on link dynamics combined with an autoregressive model for the latent variables governing the link probability. There is a latent variable in each node and it is termed the fitness of the node. Thus at each time-step a link can be created - or not - either as a consequence of a copying process of the past link state or as a consequence of a random sampling whose probability depends on the current value of the latent variables of the two considered nodes. Clearly both mechanisms give rise to time correlation of the link state, even if its origin is quite different in the two cases. Being able to disentangle the role in link persistence due to explicit copy from the past or to fitness dynamics allows to identify genuine patterns of preferential linkage.

As a specific application in this paper we study the interbank market, which is an important infrastructure of the financial system. Banks borrow and lend money in the interbank market to meet liquidity shortages or to allocate liquidity surpluses on a daily basis. The decision of whom to trade with is complex but certainly two aspects play an important role: first, the internal state (e.g. balance sheet, liquidity available or needed) of the bank and second the knowledge of the counterpart. Concerning this last aspect, all else being equal a bank will prefer to trade with someone who was a counterpart in the past, since lending money requires some trust on the borrower's solvability. For instance, it has been theoretically shown by Weisbuch et al. (2000) that in an Over-The-Counter market if buyers learn from their own previous experience there is a sharp division between those who learn to be loyal to certain sellers and those who continue to 'shop around'. The relationship of trust between a buyer (borrower) and a seller (lender) is defined as preferential trading, and has been documented in many empirical papers (Cocco et al. (2009), Di Maggio et al. (2017)).

Our model is able to assess the importance of preferential trading between two banks when the (possibly time-varying) internal states of the two banks are taken into account. It is important to stress that the same argument can be made for social networks where the copying mechanism favors extant links due to a minor social cost of entertaining new relationships, see Hellmann and Staudigl (2014) for a review on the evolution of social networks.

Related Literature. The literature on statistical models of temporal networks is rapidly growing and there are several dynamic network models that have been investigated. Each model tries to capture different aspects of spatial and time dependencies in temporal networks and

the main aims are the description of how network topology evolves through time and the prediction of links. From a mathematical point of view, there exist two principal approaches: (i) description of the graph dynamics using (generalized) Markov chains on network observables (ii) description of the graph dynamics using models of latent variables whose dynamics determines the evolution of the network topology. We briefly describe both approaches in the following.

Concerning the first stream of literature, a milestone study is represented by the work of Hanneke et al. (2010) where Exponential Random Graphs (ERG) model has been adapted to temporal networks. The method is called Temporal Exponential Random Graphs (TERG) and aims to model several metrics involving two consecutive network snapshots in a similar fashion to ERG. Krivitsky and Handcock (2014) investigated further TERG by studying a specific parametrization allowing maximum likelihood estimation of the model. More recently, Peixoto and Rosvall (2017) have proposed an extension of the Stochastic Block Model (SBM) to temporal networks by modeling a n-th order Markov chain with suitable transition probabilities which generate the time sequence of links. The aim is to select the most appropriate Markov order and number of communities. Zhang et al. (2017) have proposed generalizations of a number of standard network models, including the classic random graph, the configuration model as well as SBM, to the case of dynamic networks in a similar fashion to Peixoto and Rosvall (2017) . Let us notice that temporal networks can be also interpreted as multiplex networks, each network snapshot representing a layer of the multiplex. Very recently, Papadopoulos and Kleineberg (2019) embed nodes in hyperbolic spaces and introduce a persistence parameter to explain the edge overlap among the layers of real multiplex networks, as coordinate correlations are not found to provide a sufficient explanation. Thus there is some analogy with our approach where temporal (auto)correlations are governed by the persistence parameters. However, they do not consider latent dynamics describing the network topology. More importantly, we further consider link-specific persistence patterns, thus describing temporal autocorrelation of each link sequence, individually. Finally, in this stream of literature several works have focused on the problem of link prediction in time-evolving graphs. One example is the work of Richard et al. (2014) where the authors propose to describe the time series of graph snapshots with a vector autoregressive process. Furthermore, they propose an efficient model estimation based on proximal methods.

However, this first kind of approach has been criticized because it can not fully capture the time-varying patterns of the network structure. This opens to a second stream of literature which aims to describe these patterns with models having time-varying (latent) parameters that capture how network topology changes in time, see Kim et al. (2017) for a review. A milestone work is represented by the study of Sarkar and Moore (2005) that generalized the latent

space model introduced in Hoff et al. (2002) to dynamic networks. The dynamics of the network structure is modelled through random effects in a latent space. Sewell and Chen (2015) proposed a Markov Chain Monte Carlo (MCMC) algorithm to estimate the model parameters and latent positions of the nodes in the network. Durante and Dunson (2014) proposed a further extension of the model in Sarkar and Moore (2005) by describing a latent space model for dynamic networks in which latent node positions evolves in time via stochastic differential equations. They introduced also an efficient MCMC algorithm for Bayesian inference to learn model parameters (Durante and Dunson (2016)). Giraitis et al. (2016) proposed a novel methodology for dynamic modeling of temporal networks with application to interbank networks. They describe the link dynamics with a Tobit model, allowing for deterministic or stochastic time-varying parameters that take into account the possibility of structural changes in network dynamics. Brauning and Koopman (2016) have applied the dynamic factors model to the case of dynamic networks. Depending on the number of factors, the model allows to reduce the dimensionality of the problem and to describe cross-sectional dependencies in network data. Lee et al. (2017) have recently introduced a generalization of TERG, called varying-coefficient exponential random graph model, that characterizes the evolution of network topologies through smoothly time varying parameters whose dynamics can capture temporal heterogeneity of dynamic networks. Finally, in this stream of literature we can also include all the generalizations (Yang et al. (2011), Xu and Hero (2014), Xu (2015), Ghasemian at al. (2016), Barucca et al. (2017), Matias and Miele (2017)) of SBM which account for time evolving community memberships and/or link persistence. However, these works are more focused on the problem of community detection when dynamic effects are considered. The combined effect of latent variable for nodes and Markov processes for edges has been investigated in a set of limited studies, and most of these have neglected the direct effect of link persistence in statistical models. Heaukulani and Ghahramani (2013) focus on the influence of observed relationships, i.e. edges, on the future hidden structure, e.g. community structure or fitness, which effectively creates a conditional dependence of future links on past links. On the other hand, Xu (2015) focuses on a stochastic block transition model in which the probability of future links is influenced both by past links and present communities, but preserving the static SBM marginal distribution for the adjacency matrix of a graph at any given time step. In order to investigate reciprocity in asymmetric networks, Bartolucci at al. (2018) introduce a dyadic sampling of reciprocal pairs of links, based on the dynamic stochastic block model. Recently, Friel et al. (2016) study the persistence in company boards of directors of Irish companies by introducing a bipartite model which utilises two Latent Euclidean spaces for the two types of nodes, i.e. directors and boards, and two persistence parameters for conditioning future probabilities on both absent and present

past links. To the best of our knowledge this is the closest statistical approach to the problem investigated in the present study. In particular, the work by Friel et al. (2016) differs in the Markovian behavior of links. In fact, the mechanism described by Friel et al. (2016) captures the average link persistence of the network, without distinguishing different persistent behaviors of links. In our approach, though, the local behavior captured by a link-specific persistence parameter is of fundamental importance when we aim to disentangle the two persistence patterns associated with both the node fitness and the links incident to the node. At difference with the aforementioned study, we are focusing on a model of dynamic networks with latent fitness variables, where past links are directly influencing the probability of future links, proposing an original Expectation-Maximisation procedure to learn all the parameters of the model, both latent variables and persistence parameters. Further, we test this model on a benchmark set of interbank networks, demonstrating the explanatory power of the persistence variables in a financial case. In particular, we exploit explicitly the link-specific persistence in order to assess properly what are the preferential lending relations between banks, distinguished from random trading mediated by node-specific characteristics which can be persistent, though.

The methodology we propose in this paper exploits both of the aspects that characterize the two streams of literature. From one side, we describe link persistence coming from the mechanism of copying from the past by modeling a Markov chain for link stability, i.e. the tendency of a link that does (or does not) exist at time $t-1$ to continue existing (or not existing) at time t, similarly to Hanneke et al. (2010) . From the other side, we describe the stochastic dynamics of node-specific latent variables that we call fitnesses, with a similar aim of Brauning and Koopman (2016), Lee et al. (2017) . The node fitness describes the tendency of a node in creating links and its evolution determines how the degree of the node changes in time. It is noteworthy that a temporal network ensemble can also be regarded as the set of possible actions of a co-evolutionary N-players game where preferences of players are determining the stochastic dynamics of links within the network, giving rise to a latent fitness dynamics governing the evolution of the network, as in the model studied by Staudigl (2013) .

From the point of view of generative network models, link stability tends to capture preferential linkage mechanism between the nodes of the network while the fitness dynamics accounts for the evolving network topology. The main goal of this work is to disentangle the two temporal patterns generated by the two linkage mechanisms in network data. Hence, we apply our methodology to the financial network of electronic Market of Interbank Deposit (e-MID) where the two linkage mechanisms are associated with two different trading behaviors, i.e. random and preferential trading (Hatzopoulos et al. (2015)), in the process of network formation.

The remainder of this paper is organized as follows. In Section 1 we describe three different models of temporal networks. In Section 2 we present a novel Expectation-Maximization algorithm for model estimation and in Section 3 we run a Monte Carlo exercise to assess the goodness of fit of our estimation method. In Section 4 we apply our methodology to the network of the electronic Market of Interbank Deposit. Finally, we conclude with a discussion of our method and open areas for future research in Section 5.

1. The models

In this Section, we introduce three models of temporal networks: (i) in the first one, the presence or absence of a link can be a copy of the past with a given probability or can be sampled according to a Bernoulli marginal distribution; (ii) in the second model, each graph snapshot does not have an explicit dependence from the past snapshots but the link probability depends on node-specific latent dynamical variables, i.e. the node fitnesses, which evolve stochastically in time with memory of past information; (iii) the third network model combines the copying mechanism of the first model with dynamic node fitnesses of the second model.

We define a temporal network as a time series of graphs, that is the set $(\mathcal{N}, \{\boldsymbol{A}^t\}^{t=0,1,\dots,T})$ with $|\mathcal{N}| = N$ nodes and adjacency matrices $\{\boldsymbol{A}^t\}^{t=0,1,\dots,T}$. A network snapshot is the observed graph at a given time t and is described by the adjacency matrix $\boldsymbol{A}^t$ which has entry $A^t_{ij} = 1$ if the edge from node i to node j is present at time t and zero otherwise. In our models, we exclude graphs with self loops, i.e. the diagonal of $\boldsymbol{A}^t$ is null for all t. The adjacency matrix can be symmetric (undirected graphs) or not (directed ones). In the following, we refer to the undirected case for notational simplicity. The generalization is straightforward and is indeed used in the empirical analysis of the interbank market of Section 4.

In our framework, a temporal network is the observable of the following hidden Markov chain:

$$\begin{cases} \mathbb{P}(\Theta^t|\Theta^{t-1}, \boldsymbol{\Phi}) & = h(\Theta^t, \Theta^{t-1}, \boldsymbol{\Phi}) \\ \mathbb{P}(\boldsymbol{A}^t|\boldsymbol{A}^{t-1}, \Theta^t, \boldsymbol{\beta}) & = g(\boldsymbol{A}^t, \boldsymbol{A}^{t-1}, \Theta^t, \boldsymbol{\beta}) \end{cases} \tag{1.1}$$

where $\{\Theta^t\}^{t=0,1,\dots,T}$ represents the set of dynamic parameters, which are also called latent variables of the Markov chain. Their dynamics is determined by the one-step transition probability h, whereas g represents the likelihood for the network snapshot at time t given the information about the previous network snapshot, as well as the latent variables Θ^t. Finally $\Pi \equiv \{\boldsymbol{\beta}, \boldsymbol{\Phi}\}$ represents the set of static parameters.

Since the Markov chain in Eq. 1.1 has a high-dimensional set of parameters, we reduce the dimensionality by assuming that the node-specific latent variables evolve independently and that there are no explicit spatial correlations among links. Nevertheless, spatial correlations between links are implicitly induced by the latent dynamics.

As mentioned above, in the following, we consider three different specifications of Eq. 1.1.

1.1. **Discrete AutoRegressive Graphs (DAR(1)).** We model link stability with the following discrete autoregressive process,

$$A_{ij}^t = V_{ij}^t A_{ij}^{t-1} + (1 - V_{ij}^t) Y_{ij}^t \quad \forall i, j = 1, ..., N \text{ and } j > i \tag{1.2}$$

where $V_{ij}^t \sim \mathcal{B}(\alpha_{ij})$ with $\alpha_{ij} \in [0, 1]$, $Y_{ij}^t \sim \mathcal{B}(\chi_{ij})$ with $\chi_{ij} \in [0, 1]$ and $\mathcal{B}$ indicates the Bernoulli distribution. In the process of Eq. 1.2, the value of A_{ij}^t is copied from the past value with probability α_{ij} or obtained by tossing a coin according to the marginal distribution $\mathcal{B}(\chi_{ij})$ with probability $1 - \alpha_{ij}$. Highly persistent links (or no-links) are described by high values of α_{ij}. As a consequence, networks characterized by high values of $\boldsymbol{\alpha} \equiv \{\alpha_{ij}\}_{i,j=1,...,N}$ tend to preserve the past structure through time.

The Markov chain described by Eq. 1.2 is the first order process DAR(1), belonging to the more general class of discrete autoregressive processes DAR(p) (Jacobs and Lewis (1978) (A)). Here, we do not consider the hidden dynamics associated with the latent variables.

Hence, the specification of Eq. 1.1 for this model is the following,

$$\mathbb{P}(\boldsymbol{A}^t | \boldsymbol{A}^{t-1}, \boldsymbol{\alpha}, \boldsymbol{\chi}) = \prod_{i,j>i} \mathbb{P}(A_{ij}^t | A_{ij}^{t-1}, \alpha_{ij}, \chi_{ij}) = \prod_{i,j>i} \left(\alpha_{ij} \mathbb{I}_{A_{ij}^t A_{ij}^{t-1}} + (1 - \alpha_{ij}) \chi_{ij}^{A_{ij}^t} (1 - \chi_{ij})^{1 - A_{ij}^t} \right), \tag{1.3}$$

where $\mathbb{I}_{A_{ij}^t A_{ij}^{t-1}}$ is the indicator function taking value equal to 1 if $A_{ij}^t = A_{ij}^{t-1}$ and zero otherwise. Eq. 1.3 describes $\binom{N}{2}$ independent Markov chains for each link. This model of temporal networks is fully determined by the $N(N-1)$ parameters $\{\boldsymbol{\alpha}, \boldsymbol{\chi}\} \equiv \{\alpha_{ij}, \chi_{ij}\}_{i=1,...,N; j>i}$ and we estimate them by maximum likelihood method. The estimation of DAR(1) model is a standard result of time series analysis and Jacobs and Lewis (1978) (B) describe, in a second work, the asymptotic properties of generic DAR(p) processes, thus proving the consistency of maximum likelihood estimator. For completeness, the maximum likelihood estimation of DAR(1) is described in Appendix C.

The persistence pattern of this model can be quantified by the autocorrelation functions (ACF) of the links. It is the one of a standard autoregressive process AR(1) but with non negative autoregressive coefficient α_{ij}, i.e. the DAR(1) graph model is able to describe only non negative ACF. The generalization of this model to directed networks is simply obtained by considering not symmetric adjacency matrices.

The (above) model is highly parametrized, since there are two parameters for each link. We choose this framework, exactly because we want to identify the persistence pattern for each individual link. However, for other purposes (for example pure link prediction or very large graphs) one can consider a parameter restriction where all links shares the same parameters. For example, a simplified version of the DAR network, where all αs are the same (but the model

is Markovian of order $p \geq 1$) has been recently considered in Williams et al. (2019) in the study of disease spreading in temporal networks.

1.2. **Temporally Generalized Random Graphs (TGRG).** The second model is a generalization of the fitness network model (Caldarelli et al. (2002), Garlaschelli and Loffredo (2004)) to a dynamic setting that accounts for time evolving node fitness. Fitness is a node property determining its capability of creating links. We assume that each node i is characterized by the fitness θ_i which evolves in time by following a covariance stationary autoregressive process AR(1),

$$\theta_i^t = \phi_{0,i} + \phi_{1,i}\theta_i^{t-1} + \epsilon_i^t\,, \quad \forall i = 1, ..., N \tag{1.4}$$

where $\phi_{0,i} \in \mathbb{R}$, $|\phi_{1,i}| < 1$ and the i.i.d. variables $\epsilon_i^t \sim \mathcal{N}(0, \sigma_i^2)$. This choice is consistent with the Markovian assumption in Eq. 1.1. Moreover, the hidden node state θ_i^t evolves in $\mathbb{R}$ between timesteps, but large changes are unlikely because of the Gaussian transition probabilities. This is consistent with the idea that the network topology changes smoothly in time. Finally, assuming a Gaussian transition probability represents a simplification for model estimation.

The conditional probability for the network at time t is

$$\mathbb{P}(\boldsymbol{A}^t|\Theta^t) = \prod_{i,j>i} \frac{e^{A_{ij}^t(\theta_i^t+\theta_j^t)}}{1+e^{(\theta_i^t+\theta_j^t)}}, \tag{1.5}$$

where $\Theta^t \equiv \{\theta_i^t\}_{i=1,...,N}$ is the vector of time-varying parameters. In Eq. 1.5 we assume that each link is independently sampled and the probability of a link between node i and node j at time t is determined by the corresponding θ_i^t and θ_j^t. The larger is θ_i^t, the larger is the probability for all links incident to node i.

We refer to this model as Temporally Generalized Random Graphs (TGRG) and the specification of Eq. 1.1 for the TGRG is the following,

$$\begin{cases} \mathbb{P}(\theta_i^t|\theta_i^{t-1}, \boldsymbol{\Phi}_i) & = f(\theta_i^t|\phi_{0,i} + \phi_{1,i}\theta_i^{t-1}, \sigma_i^2) \;\; \forall i = 1, ..., N \\ \mathbb{P}(\boldsymbol{A}^t|\Theta^t) & = \prod_{i,j>i} \mathbb{P}(A_{ij}^t|\theta_i^t, \theta_j^t) = \prod_{i,j>i} \frac{e^{A_{ij}^t(\theta_i^t+\theta_j^t)}}{1+e^{(\theta_i^t+\theta_j^t)}} \end{cases} \tag{1.6}$$

and $\mathbb{P}(\Theta^t|\Theta^{t-1}, \boldsymbol{\Phi}) = \prod_{i=1}^N \mathbb{P}(\theta_i^t|\theta_i^{t-1}, \boldsymbol{\Phi}_i)$ according to the hypothesis of independence, where $f(\theta_i^t|\phi_{0,i} + \phi_{1,i}\theta_i^{t-1}, \sigma_i^2)$ is the density of normal variable with mean $\phi_{0,i} + \phi_{1,i}\theta_i^{t-1}$ and variance σ_i^2. The set of static parameters is $\boldsymbol{\Phi} \equiv \{\boldsymbol{\Phi}_i\}_{i=1,...,N}$ with $\boldsymbol{\Phi}_i \equiv \{\phi_{0,i}, \phi_{1,i}, \sigma_i\}$.

The TGRG model is fully determined by the $3 \times N$ static parameters $\boldsymbol{\Phi}$. In the next Section we propose an Expectation-Maximization scheme to estimate the model parameters and time-varying parameters. It alternates between an Expectation step where we infer the time-varying parameters $\{\Theta^t\}^{t=0,1,...,T}$ and the Maximization step where we maximize the log-likelihood of the static parameters conditional on the expectations $\{\hat{\Theta}^t\}^{t=0,1,...,T}$. Regarding the consistency of the estimator, Chatterjee and Diaconis (2011), Yan and Xu (2013) show that, in the

asymptotic limit of dense (undirected) networks with diverging number of nodes, maximum likelihood estimator of the (static) fitness model is consistent, thus the model is identifiable, see Gabrielsen (1978) . However, identifiability of the dynamic fitness model in the sparse regime is still an open question and is left for future research.

Time autocorrelated node fitnesses may induce link persistence. In fact, the probability of a link between two specific nodes $e^{(\theta_i^t+\theta_j^t)}/(1+e^{(\theta_i^t+\theta_j^t)})$ is persistent if θ_i^t and θ_j^t are autocorrelated. Note that link persistence occurs as a consequence of node properties. For TGRG, the two-point distribution function for lagged links and the ACF of link state can be semi-analytically computed (see Appendix B).

The generalization of the TGRG model to directed networks can be accomplished by distinguishing between the out-degree and the in-degree and by introducing two fitnesses for each node i, i.e. $\theta_i^{t,out}$ and $\theta_i^{t,in}$. The probability of a link from node i to node j at time t given the latent variables $\theta_i^{t,out}$ and $\theta_j^{t,in}$ is $\mathbb{P}(A_{ij}^t|\theta_i^{t,out},\theta_j^{t,in}) = \frac{e^{A_{ij}^t(\theta_i^{t,out}+\theta_j^{t,in})}}{1+e^{(\theta_i^{t,out}+\theta_j^{t,in})}}$. Then, everything follows similarly to the undirected case with the exception that $\mathbb{P}(\boldsymbol{A}^t|\Theta^t)$ is invariant under a linear transformation for the hidden node states: $\theta_i^{t,out} \mapsto \theta_i^{t,out} + c_t \quad \forall i = 1,...,N$, $\theta_j^{t,in} \mapsto \theta_j^{t,in} - c_t \quad \forall j = 1,...,N$, where $\{c_t\}_{t=0,1,...,T} \in \mathbb{R}^{T+1}$. This symmetry arises because the total number of outgoing links has to be equal to the total number of incoming links at each time. It makes the model as non-identifiable. However, when we are interested in the temporal patterns captured by the model, results remain unchanged as long as we preserve the sum of any $\theta_i^{t,out}$ and any $\theta_j^{t,in}$. Hence, we can take one of the fitnesses as constant in time and infer the other ones. Notice that this identifiability problem is well known for the exponential family of probability distributions for directed static graphs since the milestone work of Holland and Leinhardt (1981) and several studies have been focused on the same parameterization we adopt here, *e.g.* see Rinaldo et al. (2013), Luo et al. (2017) . In this case, any comparison between inferred fitnesses and any observed quantity has to be properly considered.

Finally, let us notice that we can interpret TGRG as an extension of Exponential Random Graphs (ERG) by Park and Newman (2004) to the dynamic case. ERG ensembles are probability distributions of networks obtained by maximizing the Shannon entropy under some constraints on the average value of a set of network observables. If this set is the degree sequence, the Lagrange multipliers of the entropy constrained optimization can be directly linked to the latent variables of our model. Differently from other dynamic extension of ERG (see for example Hanneke et al. (2010)) where dynamical (i.e. two-time) observables are used as constraints, here we choose a dynamical model for the latent variable, namely the AR(1) process, and introduce an estimation method for them.

1.3. **Discrete AutoRegressive Temporally Generalized Random Graphs (DAR-TGRG).** The persistence pattern associated with the copying mechanism described by Eq. 1.2 can coexist with the node fitnesses evolving in time according to Eq. 1.4. This can be captured by the following specification of the model in Eq. 1.1,

$$\begin{cases} \mathbb{P}(\theta_i^t|\theta_i^{t-1},\mathbf{\Phi}_i) & = f(\theta_i^t|\phi_{0,i}+\phi_{1,i}\theta_i^{t-1},\sigma_i^2) \;\; \forall i=1,...,N \\ \mathbb{P}(\boldsymbol{A}^t|\boldsymbol{A}^{t-1},\Theta^t,\boldsymbol{\alpha}) & = \prod_{i,j>i}\left(\alpha_{ij}\mathbb{I}_{A_{ij}^t A_{ij}^{t-1}}+(1-\alpha_{ij})\frac{e^{A_{ij}^t(\theta_i^t+\theta_j^t)}}{1+e^{(\theta_i^t+\theta_j^t)}}\right) \end{cases} \tag{1.7}$$

and $\mathbb{P}(\Theta^t|\Theta^{t-1},\mathbf{\Phi})=\prod_{i=1}^N \mathbb{P}(\theta_i^t|\theta_i^{t-1},\mathbf{\Phi}_i)$ according to the hypothesis of independence, where $f(\theta_i^t|\phi_{0,i}+\phi_{1,i}\theta_i^{t-1},\sigma_i^2)$ is the density of a normal variable with mean $\phi_{0,i}+\phi_{1,i}\theta_i^{t-1}$ and variance σ_i^2, $\alpha_{ij}\in[0,1]$ and $\boldsymbol{\alpha}\equiv\{\alpha_{ij}\}$ $\forall i,j=1,...,N$ with $\alpha_{ij}=\alpha_{ji}$ for undirected networks, $\mathbf{\Phi}\equiv\{\phi_{0,i},\phi_{1,i},\sigma_i\}_{i=1,...,N}$ with $\phi_{0,i}\in\mathbb{R}$, $|\phi_{1,i}|<1$ and $\sigma_i>0$ $\forall i$, and $\theta_i^t\in\mathbb{R}$ $\forall i,t$.

This model can be interpreted as a mixture of the two mechanisms, i.e. the one of copying the presence or absence of a link from the past with probability α_{ij} and the one of time evolving marginals described by the TRGR model with probability $1-\alpha_{ij}$. Let us stress that the temporal pattern generated by the fitness dynamics does not concern a specific link but it is a node property. Thus under this mechanism, links incident on the same node tend to have similar persistence properties. On the contrary, the persistence of the copying mechanism is a link property, and links incident on the same node can have very different persistence properties. The parameter α_{ij} disentangles the importance of these two effects in determining the dynamics of the link (i,j).

The model in Eq. 1.7, here referred as DAR-TGRG, is fully determined by the $\binom{N}{2}$ parameters $\boldsymbol{\alpha}$ and the $3\times N$ parameters $\mathbf{\Phi}$ and it can be estimated by the Expectation-Maximization algorithm we propose in the next Section.

DAR-TGRG is a generalization of the TGRG model, then the same distinction between undirected and directed networks needs to be considered and similar conclusions can be obtained for the identifiability of the model. In particular, in the case of undirected networks, if TGRG is identifiable, then this is a strong evidence in favor of the identifiability of the DAR-TGRG model. In fact, a symmetry of the likelihood involving both the parameters associated with the mechanism of copying from the past and the parameters associated with the fitness dynamics, should do not depend on observations $\{A_{ij}^{t-1},A_{ij}^t\}_{i,j=1,...,N}$ $\forall t$, as is the case.

2. Estimation method

We now describe the procedure for the estimation of the DAR-TGRG model. We propose a Marginal Maximum a Posteriori (MMAP) approach together with an Expectation-Maximization like (EM) procedure to estimate the static parameters and track the latent variables dynamics.

The estimation method for TGRG model is simply obtained by setting parameters $\alpha_{ij}\ \forall i,j = 1,...,N$ equal to zero in the following equations.

Let denote $\boldsymbol{\Theta} \equiv \{\Theta^t\}^{t=1,...,T}$, $\boldsymbol{A} \equiv \{\boldsymbol{A}^t\}^{t=0,1,...,T}$ and $\Pi \equiv \{\boldsymbol{\Phi}, \boldsymbol{\alpha}\}$. The marginal posterior distribution of the static parameters reads as

$$\mathbb{P}(\Pi|\boldsymbol{A}) = \frac{\mathbb{P}(\Pi)}{\mathbb{P}(\boldsymbol{A})} \int [d\Gamma] \mathbb{P}(\boldsymbol{A}, \Gamma|\Pi) \propto \mathbb{P}(\Pi) Z_\Pi \tag{2.1}$$

where $[d\Gamma]$ represents the measure over the probability space for $\boldsymbol{\Theta}$ and $Z_\Pi = \int [d\Gamma] \mathbb{P}(\boldsymbol{A}, \Gamma|\Pi)$. It is used to learn the most likely set of parameters $\hat{\Pi}$ given the data. Using smooth priors $\mathbb{P}(\Pi)$, $\hat{\Pi}$ is obtained by extremizing over Π the log-likelihood $l(\Pi) \equiv \log \mathbb{P}(\Pi|\boldsymbol{A})$, i.e. by solving the equations

$$\begin{aligned} \partial_\Pi \log Z_\Pi = \partial_\Pi \log \int [d\boldsymbol{\Gamma}] \mathbb{P}(\boldsymbol{A}, \boldsymbol{\Gamma}|\Pi) = \frac{\int [d\boldsymbol{\Gamma}] \partial_\Pi \mathbb{P}(\boldsymbol{A}, \boldsymbol{\Gamma}|\Pi)}{\int [d\boldsymbol{\Gamma}] \mathbb{P}(\boldsymbol{A}, \boldsymbol{\Gamma}|\Pi)} &= \\ \langle \partial_\Pi \log(\mathbb{P}(\boldsymbol{A}, \Gamma|\Pi)) \rangle_{Z_\Pi^{-1} \mathbb{P}(\boldsymbol{A}, \boldsymbol{\Theta}|\Pi)} &= -\partial_\Pi \log(\mathbb{P}(\Pi)). \end{aligned} \tag{2.2}$$

Since maximizing Eq. 2.1, i.e. solving Eq. 2.2, needs computing expectations with respect to the posterior distribution of the latent variables

$$\mathbb{P}(\boldsymbol{\Theta}|\boldsymbol{A}, \Pi) = \frac{\mathbb{P}(\boldsymbol{A}, \boldsymbol{\Theta}|\Pi)}{\int [d\Gamma] \mathbb{P}(\boldsymbol{A}, \boldsymbol{\Gamma}|\Pi)} = Z_\Pi^{-1} \mathbb{P}(\boldsymbol{A}, \boldsymbol{\Theta}|\Pi), \tag{2.3}$$

the latter requiring in turn an estimate of Π, one typically repeats the two steps until convergence starting from an initial guess, in a standard Expectation-Maximization framework (Friedman et al. (2001)). Once obtained $\hat{\Pi}$, the dynamics of the fitnesses $\hat{\boldsymbol{\Theta}}$ is obtained by maximizing $\mathbb{P}(\boldsymbol{\Theta}|\boldsymbol{A}, \hat{\Pi})$.

Since we have a large number of latent variables this approach is preferred with respect to the maximization of the joint posterior distribution $\mathbb{P}(\boldsymbol{\Theta}, \Pi|\boldsymbol{A}) \propto \mathbb{P}(\boldsymbol{A}|\boldsymbol{\Theta}, \Pi) \mathbb{P}(\boldsymbol{\Theta}|\Pi) \mathbb{P}(\Pi)$ (an approach named Joint Maximum A Posteriori (JMAP) and, for completeness, compared with MMAP in Appendix A) that suffers inherent problems first recognized by Little et al. (1983), Neyman et al. (1948) : they showed that when static parameters are estimated simultaneously with the latent variables, the maximum likelihood estimates of the former need not be consistent as sample size increases. The drawback is that the integral over the posterior distribution of latent variables is typically intractable and cannot be computed exactly, thus the expectation step needs some approximations. In the following we explain in full detail each step of the proposed procedure.

A last remark concerns the choice of the prior $\mathbb{P}(\Pi)$: in the following we always assume flat distributions, i.e. the right hand side of Eq. 2.2 is put to zero. However we find high stability with respect to the choice of the prior, as soon as it is a smooth distribution. This is not surprisingly because, since the parameters are not time-varying, the l.h.s of Eq. 2.2 is typically

T times larger than the r.h.s., meaning that the contribution of the prior asymptotically must disappear as sample size increases.

The remainder of this section is organized as follows.

(1) In Subsection 2.1, we describe how to infer the latent variables of the model(s) by maximizing the log-posterior $\log \mathbb{P}(\boldsymbol{\Theta}|\boldsymbol{A}, \Pi)$ (E-step), see Eq. 2.3;

(2) in Subsection 2.2, we show the procedure for the inference of the static parameters of the model(s) by solving Eq. 2.2 (M-step);

(3) in Subsection 2.3, we hence describe the Expectaion-Maximization (EM) algorithm based on the MMAP approach and consisting in alternating both the E-step and the M-step, until convergence;

(4) in Subsection 2.3, we present also a naive estimation method, namely the Single Snapshot Inference (SSI) of the model(s), which does not exploit the fully information of the time series of network snapshots, but, once inferred the fitness variables snapshot by snapshot, consider them as observations, thus estimating naively the remaining static parameters by fitting both AR(1) and DAR(1) autoregressive processes.

The reader not interested in the details of the inference method can skip the next subsections, moving directly to Section 3.

2.1. **Inference of time-varying parameters.** The Expectation step of the MMAP approach refers to the estimation of the time-varying parameters $\boldsymbol{\Theta}$ by conditioning on the value of the static parameters Π.

We do not solve the inference problem for the time series of dynamic parameters $\boldsymbol{\Theta}$ overall, i.e. by maximizing Eq. 2.3. Instead, we infer step by step the parameters Θ^t by conditioning on the expectations $\hat{\Theta}^{t-1}$, that are the one step backward estimates[1] for Θ^{t-1}.

Let us focus on the inference at the generic time $t \neq 0$ when the previous network snapshot is observed and let $\mathcal{F}^t \equiv \{A^{t-1}, \Pi\}$ be the information set for the considered problem. From the Bayes' theorem, it is

$$\mathbb{P}(\Theta^t|\boldsymbol{A}^t, \Theta^{t-1}, \mathcal{F}^t) = \frac{\mathbb{P}(\boldsymbol{A}^t|\Theta^t, \mathcal{F}^t)\mathbb{P}(\Theta^t|\Theta^{t-1}, \mathcal{F}^t)\mathbb{P}(\Theta^{t-1}|\mathcal{F}^t)}{\mathbb{P}(\boldsymbol{A}^t, \Theta^{t-1}|\mathcal{F}^t)}. \tag{2.4}$$

Hence, by conditioning on the expectation for Θ^{t-1}, i.e. $\hat{\Theta}^{t-1}$, the inference problem can be solved by maximizing

$$\mathbb{P}(\Theta^t|\boldsymbol{A}^t, \boldsymbol{A}^{t-1}, \hat{\Theta}^{t-1}, \Pi) \propto \mathbb{P}(\boldsymbol{A}^t|\boldsymbol{A}^{t-1}, \Theta^t, \boldsymbol{\alpha})\mathbb{P}(\Theta^t|\hat{\Theta}^{t-1}, \boldsymbol{\Phi}) \,, \quad \forall t = 1, ..., T. \tag{2.5}$$

[1]Here, we are assuming to know the expectation for Θ^0, i.e. $\hat{\Theta}^0$. Below, we explain how to infer the initial point for the latent dynamics.

Maximizing Eq. 2.5 is equivalent to solve the following problem

$$\hat{\Theta}^t = \underset{\Theta_t}{\operatorname{argmax}} \left(\log \mathbb{P}(\boldsymbol{A}^t | \boldsymbol{A}^{t-1}, \Theta_t, \boldsymbol{\alpha}) + \log \mathbb{P}(\Theta_t | \hat{\Theta}^{t-1}, \boldsymbol{\Phi}) \right), \tag{2.6}$$

where $\mathbb{P}(\Theta^t | \hat{\Theta}^{t-1}, \boldsymbol{\Phi}) \equiv \prod_{i=1}^{N} f(\theta_i^t | \phi_{0,i} + \phi_{1,i} \hat{\theta}_i^{t-1}, \sigma_i^2)$ is the Gaussian probability density function associated with the transition probability for the latent variables. Eq. 2.6 is equivalent to the following system of non linear equations,

$$\left[\sum_{j \neq i} \left(\frac{(1-\alpha_{ij}) \frac{e^{A_{ij}^t (\theta_i^t + \theta_j^t)}}{1+e^{(\theta_i^t + \theta_j^t)}}}{\alpha_{ij} \mathbb{I}_{A_{ij}^t A_{ij}^{t-1}} + (1-\alpha_{ij}) \frac{e^{A_{ij}^t (\theta_i^t + \theta_j^t)}}{1+e^{(\theta_i^t + \theta_j^t)}}} \right) \left(-A_{ij}^t + \frac{e^{(\theta_i^t + \theta_j^t)}}{1 + e^{(\theta_i^t + \theta_j^t)}} \right) \right] - \frac{\theta_i^t - \phi_{0,i} - \phi_{1,i} \hat{\theta}_i^{t-1}}{\sigma_i^2} = 0, \forall i = 1, ..., N. \tag{2.7}$$

It can be solved by the following iterative proportional fitting procedure:

(i) assume any starting point $\hat{\theta}_i^t \ \forall i = 1, ..., N$ for the node fitness[2];

(ii) then, solve one by one the equations in Eq. 2.7 by conditioning on $\hat{\theta}_j^t \ \forall j \neq i$;

(iii) update the value for $\hat{\theta}_i^t$ with the solution of the corresponding equation;

(iv) repeat until convergence.

2.1.1. *Link prediction.* The proposed method for the inference of the time-varying parameters is a statistical filtering algorithm. Filtering is an operation that involves the extraction of information about a latent quantity of interest at time t by using data measured up to and including t (Chen (2003)), like in Kalman filter and its extensions. However, differently from Kalman filter, we study the case of a Hidden Markov Model with continuous-valued state space, i.e. continuous-valued state vector Θ^t, but binary measurement matrix, i.e. $\boldsymbol{A}^t$. Notice that the proposed method can be used for on-line inference: once the off-line learning of the static parameters is completed, we can solve the filtering problem for Θ^t in Eq. 2.6 whenever the new measurement $\boldsymbol{A}^t$ is available.

On-line inference is particularly useful for link prediction: let $\boldsymbol{A}^t$ be the observation at the current time and we want to construct the one-step-ahead forecast, i.e. $\mathbb{E}[\boldsymbol{A}^{t+1} | \boldsymbol{A}^t]$. Once $\hat{\Theta}^t$ is inferred on-line by solving Eq. 2.6, the one-step-ahead forecast is constructed by averaging over the probability distribution

$$\mathbb{P}(\boldsymbol{A}^{t+1} | \boldsymbol{A}^t, \hat{\Theta}^t, \Pi) = \int [d\Theta^{t+1}] \mathbb{P}(\boldsymbol{A}^{t+1} | \boldsymbol{A}^t, \Theta^{t+1}, \boldsymbol{\alpha}) \mathbb{P}(\Theta^{t+1} | \hat{\Theta}^t, \boldsymbol{\Phi})$$

obtained by projecting the latent state. In Section 4 we show an application of this procedure.

[2] A possible choice is $\hat{\theta}_i^t = \phi_{0,i} + \phi_{1,i} \hat{\theta}_i^{t-1}$. However, it has been proved by Chatterjee and Diaconis (2011) that the iterative proportional fitting procedure for the (static) fitness model converges fast and independently from initial conditions to the Maximum Likelihood Estimator (MLE) of the fitness variables in the regime of dense networks, where the consistency of MLE is guaranteed. We have further noticed numerically the independence from initial conditions for the dynamic generalization, in both dense and sparse regimes.

2.2. **Inference of static parameters.** The Maximization step of the MMAP approach refers to the estimation of the static parameters Π by using the information on time-varying parameters $\boldsymbol{\Theta}$. It requires the maximization of the posterior in Eq. 2.1. However, as claimed before, the posterior distribution of latent variables is intractable because of strict nonlinearity of the distribution function. Hence, we use an approximated posterior, where the transition probabilities for the latent variables (Γ_t) are conditioned on the expectations at the previous step ($\hat{\boldsymbol{\Theta}}_{t-1}$), *i.e.*

$$\mathbb{P}(\Pi|\boldsymbol{A}) \propto \int [d\Gamma]\mathbb{P}(\boldsymbol{A},\Gamma|\Pi) \approx \prod_t \int [d\Gamma_t]\mathbb{P}(\boldsymbol{A}^t,\Gamma_t|\hat{\boldsymbol{\Theta}}_{t-1},\boldsymbol{A}^{t-1},\Pi). \tag{2.8}$$

In the next Section, we verify numerically the consistency of this approximation by showing how MMAP approach leads to unbiased estimation of the parameters.

2.2.1. *Learning* $\boldsymbol{\alpha}$. Let us assume to know the static parameters $\boldsymbol{\Phi}$ and consider the problem of learning $\boldsymbol{\alpha}$. The most likely estimate for $\boldsymbol{\alpha}$ is obtained by maximizing the associated posterior in Eq. 2.1, that is

$$\mathbb{P}(\boldsymbol{\alpha}|\boldsymbol{A}) \propto \int [d\boldsymbol{\Theta}]\, \mathbb{P}(\boldsymbol{A},\boldsymbol{\Theta}|\boldsymbol{\alpha},\boldsymbol{\Phi}) = \int \prod_{t=1}^{T} \left[\, d\Theta^t\right] \mathbb{P}(\boldsymbol{A}^t|\boldsymbol{A}^{t-1},\Theta^t,\boldsymbol{\alpha}) f(\Theta^t|\Theta^{t-1},\boldsymbol{\Phi}). \tag{2.9}$$

The integral in Eq. 2.9 is infeasible because of the nonlinearity of the probability distribution. Hence, we approximate the likelihood of parameters $\boldsymbol{\alpha}$ as

$$l_{\boldsymbol{\alpha}} \equiv \int \prod_{t=1}^{T} \left[\, d\Theta^t\right] \mathbb{P}(\boldsymbol{A}^t|\boldsymbol{A}^{t-1},\Theta^t,\boldsymbol{\alpha}) f(\Theta^t|\Theta^{t-1},\boldsymbol{\Phi}) \approx \prod_{t=1}^{T} \int \left[d\Theta^t\right] \mathbb{P}(\boldsymbol{A}^t|\boldsymbol{A}^{t-1},\Theta^t,\boldsymbol{\alpha}) f(\Theta^t|\hat{\Theta}^{t-1},\boldsymbol{\Phi}) \equiv \tilde{l}_{\boldsymbol{\alpha}}, \tag{2.10}$$

where $\hat{\theta}_i^{t-1}$ is the expectation of θ_i^{t-1} that we obtained by solving Eq. 2.7.

Let us focus on the learning of parameter α_{ij}. When we aim to obtain the solution for α_{ij}, the only time-varying parameters that are involved in the learning are the ones associated with node i and node j, i.e. $\{\theta_i^t\}^{t=1,\dots,T}$ and $\{\theta_j^t\}^{t=1,\dots,T}$. Hence, the most likely estimate for α_{ij} is the value that maximizes

$$\tilde{S}_{\alpha_{ij}} = \log \tilde{l}_{\alpha_{ij}} = \sum_{t=1}^{T} \log \int dxdy \left(\alpha_{ij} \mathbb{I}_{A_{ij}^t A_{ij}^{t-1}} + (1-\alpha_{ij}) \frac{e^{A_{ij}^t (x+y)}}{1+e^{(x+y)}} \right) f(x|\phi_{0,i}+\phi_{1,i}\hat{\theta}_i^{t-1},\sigma_i^2) f(y|\phi_{0,j}+\phi_{1,j}\hat{\theta}_j^{t-1},\sigma_j^2). \tag{2.11}$$

In the learning procedure, the following double integral is involved,

$$\mathcal{I}_{A_{ij}^t}(\hat{\theta}_i^{t-1},\hat{\theta}_j^{t-1},\boldsymbol{\Phi}_i,\boldsymbol{\Phi}_j) = \tag{2.12}$$
$$\int dxdy \frac{e^{A_{ij}^t (x+y)}}{1+e^{(x+y)}} f(x|\phi_{0,i}+\phi_{1,i}\hat{\theta}_i^{t-1},\sigma_i^2) f(y|\phi_{0,j}+\phi_{1,j}\hat{\theta}_j^{t-1},\sigma_j^2),$$

which can be solved numerically. However, we propose to apply the following integral identity proposed by Polson et al. (2013)

$$\frac{(e^{\psi})^a}{(1+e^{\psi})^b} = 2^{-b} e^{(a-\frac{b}{2})\psi} \int_0^{\infty} e^{-\frac{\omega\psi^2}{2}} p_{PG}(\omega) d\omega \tag{2.13}$$

where $b > 0$, $a, \psi \in \mathbb{R}$, and $p_{PG} : [0, \infty) \mapsto [0, 1)$ is the density of the Pólya-Gamma distribution. There is no a closed-form expression for p_{PG} but we evaluate it numerically. The double integral in Eq. 2.12 is equivalent to the following integral,

$$\mathcal{I}_{A_{ij}^t}(\hat{\theta}_i^{t-1}, \hat{\theta}_j^{t-1}, \mathbf{\Phi}_i, \mathbf{\Phi}_j) \equiv \int_0^{\infty} \frac{d\omega}{2} p_{PG}(\omega) \mathcal{K}_{A_{ij}^t}(\omega, \hat{\theta}_i^{t-1}, \hat{\theta}_j^{t-1}, \mathbf{\Phi}_i, \mathbf{\Phi}_j) \tag{2.14}$$

where

$$\mathcal{K}_{A_{ij}^t}(\omega, \hat{\theta}_i^{t-1}, \hat{\theta}_j^{t-1}, \mathbf{\Phi}_i, \mathbf{\Phi}_j) = \frac{exp\left(\frac{\sigma_i^2+\sigma_j^2+4(\phi_{0,i}+\phi_{1,i}\hat{\theta}_i^{t-1}+\phi_{0,j}+\phi_{1,j}\hat{\theta}_j^{t-1})(2A_{ij}^t-1-\omega(\phi_{0,i}+\phi_{1,i}\hat{\theta}_i^{t-1}+\phi_{0,j}+\phi_{1,j}\hat{\theta}_j^{t-1}))}{8(1+\omega(\sigma_i^2+\sigma_j^2))}\right)}{\sqrt{1+\omega(\sigma_i^2+\sigma_j^2)}}.$$

We propose to evaluate numerically the integral in Eq. 2.14. This gives the advantage of computing a single integral.

Then α_{ij} is estimated by solving the equation $\partial_{\alpha_{ij}} \tilde{S}_{\alpha_{ij}} = 0$, which can be explicitly rewritten as

$$\sum_{t=1}^{T} \frac{\mathbb{I}_{A_{ij}^t A_{ij}^{t-1}} - \mathcal{I}_{A_{ij}^t}(\hat{\theta}_i^{t-1}, \hat{\theta}_j^{t-1}, \mathbf{\Phi}_i, \mathbf{\Phi}_j)}{\alpha_{ij}\mathbb{I}_{A_{ij}^t A_{ij}^{t-1}} + (1-\alpha_{ij})\mathcal{I}_{A_{ij}^t}(\hat{\theta}_i^{t-1}, \hat{\theta}_j^{t-1}, \mathbf{\Phi}_i, \mathbf{\Phi}_j)} = 0. \tag{2.15}$$

The solution of Eq. 2.15 represents the most likely estimate $\hat{\alpha}_{ij}$ given the data.

2.2.2. *Learning* $\mathbf{\Phi}$. Let us assume to know the static parameters $\boldsymbol{\alpha}$ and consider the problem of learning $\mathbf{\Phi}$. Similarly to the previous Subsection, we use conditions on the expectations for the latent variables to obtain an approximated log-likelihood for $\mathbf{\Phi}$,

$$\tilde{S}_{\mathbf{\Phi}} = \sum_{t=1}^{T} \log \int \left[\prod_{k=1}^{N} dx_k \, f(x_k|\phi_{0,k} + \phi_{1,k}\hat{\theta}_k^{t-1}, \sigma_k^2)\right] \left[\prod_{i,j>i} \alpha_{ij}\mathbb{I}_{A_{ij}^t A_{ij}^{t-1}} + (1-\alpha_{ij})\frac{e^{A_{ij}^t(x_i+x_j)}}{1+e^{(x_i+x_j)}}\right]. \tag{2.16}$$

Let us focus on the learning of parameters $\mathbf{\Phi}_i \equiv \{\phi_{0,i}, \phi_{1,i}, \sigma_i\}$. Because of the marginal distribution, each time-varying parameter θ_i^t is coupled with all the others and this prevents the valuation of the multiple integral in Eq. 2.16. Hence, we adopt the following approximation for the probability measure,

$$\prod_{k=1}^{N} dx_k f(x_k|\phi_{0,k}+\phi_{1,k}\hat{\theta}_k^{t-1}, \sigma_k^2) \approx dx_i f(x_i|\phi_{0,i}+\phi_{1,i}\hat{\theta}_i^{t-1}, \sigma_i^2) \prod_{k\neq i} dx_k \delta(x_k - \hat{\theta}_k^t) f(x_k|\phi_{0,k}+\phi_{1,k}\hat{\theta}_k^{t-1}, \sigma_k^2),$$

i.e. we condition on the expectations at time t for all the latent variables with the exception of θ_i^t. Then, we maximize the following approximated likelihood,

$$\tilde{S}_{\mathbf{\Phi}_i} = \sum_{t=1}^{T} \log \int_{-\infty}^{\infty} dx_i f(x_i|\phi_{0,i}+\phi_{1,i}\hat{\theta}_i^{t-1}, \sigma_i^2) \left(\prod_{j\neq i} \alpha_{ij}\mathbb{I}_{A_{ij}^t A_{ij}^{t-1}} + (1-\alpha_{ij})\frac{e^{A_{ij}^t(x_i+\hat{\theta}_j^t)}}{1+e^{(x_i+\hat{\theta}_j^t)}}\right), \tag{2.17}$$

or, equivalently, we solve the system of equations $\partial_{\mathbf{\Phi}_i}\tilde{S}_{\mathbf{\Phi}_i}=0$.

Let us define the following partition function $\forall t=1,...,T$,

$$Z^t_{\mathbf{\Phi}_i}\equiv\int_{-\infty}^{+\infty}dx\; f(x_i|\phi_{0,i}+\phi_{1,i}\hat{\theta}_i^{t-1},\sigma_i^2)\left(\prod_{j\neq i}\alpha_{ij}^{\mathbb{I}_{A^t_{ij}A^{t-1}_{ij}}}+(1-\alpha_{ij})\frac{e^{A^t_{ij}(x+\hat{\theta}^t_j)}}{1+e^{(x+\hat{\theta}^t_j)}}\right) \tag{2.18}$$

and let $\mu^t_{\mathbf{\Phi}_i}$ and $\Sigma^t_{\mathbf{\Phi}_i}$ be the first and the second moment of the distribution, respectively.

The system of equations $\partial_{\mathbf{\Phi}_i}\tilde{S}_{\mathbf{\Phi}_i}=0$ reads explicitly as

$$\begin{cases}\langle\boldsymbol{\mu}_{\mathbf{\Phi}_i}\rangle-\phi_{0,i}-\langle L\hat{\boldsymbol{\theta}}_i\rangle\phi_{1,i} & =0\\ \frac{1}{T}(L\hat{\boldsymbol{\theta}}_i^{\intercal}\,\boldsymbol{\mu}_{\mathbf{\Phi}_i})-\langle L\hat{\boldsymbol{\theta}}_i\rangle\phi_{0,i}-\frac{1}{T}(L\hat{\boldsymbol{\theta}}_i^{\intercal}\,L\hat{\boldsymbol{\theta}}_i)\phi_{1,i} & =0\\ \sigma_i^2-\left(\langle\boldsymbol{\Sigma}_{\mathbf{\Phi}_i}\rangle+\phi_{0,i}^2+\frac{1}{T}(L\hat{\boldsymbol{\theta}}_i^{\intercal}\,L\hat{\boldsymbol{\theta}}_i)\phi_{1,i}^2-2\langle\boldsymbol{\mu}_{\mathbf{\Phi}_i}\rangle\phi_{0,i}-2\frac{1}{T}(L\hat{\boldsymbol{\theta}}_i^{\intercal}\,\boldsymbol{\mu}_{\mathbf{\Phi}_i})\phi_{1,i}+2\langle L\hat{\boldsymbol{\theta}}_i\rangle\phi_{0,i}\phi_{1,i}\right) & =0\end{cases} \tag{2.19}$$

where bold symbols represent T-dimensional vectors, *e.g.* $\boldsymbol{x}=(x^1,x^2,...,x^T)'$, angle brackets denote time average, *e.g.* $\langle\boldsymbol{x}\rangle\equiv\frac{1}{T}\sum_{t=1}^T x^t$, and L is the *lag operator*, *i.e.* $Lx^t=x^{t-1}$.[3]

The system of nonlinear equations can be solved with the following iterative proportional fitting procedure:

(i) assume any starting point $\mathbf{\Phi}_i^0$;

(ii) compute $\mu^t_{\mathbf{\Phi}_i^0}$ and $\Sigma^t_{\mathbf{\Phi}_i^0}$ $\forall t=1,...,T$;

(iii) solve the system of equation in Eq. 2.19 by substituting $\mu^t_{\mathbf{\Phi}_i^0}\to\mu^t_{\mathbf{\Phi}_i}$ and $\Sigma^t_{\mathbf{\Phi}_i^0}\to\Sigma^t_{\mathbf{\Phi}_i}$ $\forall t=1,...,T$;

(iv) update the values for $\mathbf{\Phi}_i^0$ and continue until convergence.

2.3. **Expectation-Maximization method.** The estimation procedure consists in alternating the inference of the latent variables (Expectation step) and the learning of the static parameters (Maximization step) up to convergence.

2.3.1. *Single Snapshot Inference (SSI).* As a starting point of the method, the time-varying parameters $\{\Theta^t\}^{t=0,1,...,T}$ can be estimated by single snapshot inference, *i.e.* given the network snapshot at time t and by assuming $\mathbb{P}(\boldsymbol{A}^t|\boldsymbol{\gamma})=\prod_{i,j>i}\frac{e^{A^t_{ij}(\gamma_i+\gamma_j)}}{1+e^{(\gamma_i+\gamma_j)}}$, we solve snapshot by snapshot the following problem,

$$\tilde{\Theta}^t=\underset{\boldsymbol{\gamma}}{\text{argmax}}\;\;\log\mathbb{P}(\boldsymbol{A}^t|\boldsymbol{\gamma})\quad\forall t=0,1,...T \tag{2.20}$$

and we obtain a naive estimation $\tilde{\mathbf{\Theta}}\equiv\{\tilde{\Theta}^t\}^{t=0,1,...,T}$ of the hidden states of the Markov chain in Eq. 1.7. In particular, we infer the latent state at the initial time, i.e. $\hat{\Theta}^0\equiv\tilde{\Theta}^0$.

Then, we estimate the autoregressive process AR(1) in Eq. 1.4 for each naively inferred fitness, *i.e.* $\{\tilde{\theta}_i^t\}^{t=0,1,...,T}$ $\forall i=1,...,N$, to obtain a naive estimation of the static parameters

[3]Notice that $L\hat{\theta}_i^1=\hat{\theta}_i^0$ represents the latent state at the initial time. It is estimated as described in Section 2.3.

$\tilde{\mathbf{\Phi}}$. Finally, a naive estimation of $\tilde{\boldsymbol{\alpha}}$ can be obtained by solving Eq. 2.15 with naively inferred $\{\tilde{\Theta}^t\}^{t=0,1,\dots,T}$ and $\tilde{\mathbf{\Phi}}$. We refer to this naive estimation method as the Single Snapshot Inference (SSI) of the model.

2.3.2. *The algorithm.* The Expectation-Maximization algorithm based on the MMAP approach alternates the following steps up to convergence:

(i) Assume SSI as starting point, *i.e.* $\tilde{\mathbf{\Theta}}$ (see Eq. 2.20) and $\tilde{\Pi} = \{\tilde{\mathbf{\Phi}}, \tilde{\boldsymbol{\alpha}}\}$;
(ii) infer $\hat{\mathbf{\Theta}} \equiv \{\hat{\Theta}^t\}^{t=1,\dots T}$ by solving Eq. 2.7 with $\tilde{\Pi}$;
(iii) learn $\hat{\boldsymbol{\alpha}}$ by solving Eq. 2.15 for each possible couple of nodes with previously inferred $\hat{\mathbf{\Theta}}$ and $\tilde{\mathbf{\Phi}}$;
(iv) learn $\hat{\mathbf{\Phi}}$ by solving Eq. 2.19 for each i with previously inferred $\hat{\mathbf{\Theta}}$ and $\hat{\boldsymbol{\alpha}}$;
(v) update $\tilde{\mathbf{\Theta}} \leftarrow \hat{\mathbf{\Theta}}$;
(vi) update $\tilde{\Pi} \leftarrow \hat{\Pi}$;
(vii) repeat up to convergence.

This is in many aspects an Expectation-Maximization learning algorithm (see Dempster et al. (1977)). In the latent variables inference step (line ii) we also use a generalization of the RAS algorithm (Bacharach (1965)). The RAS algorithm is usually adopted to solve the problem of estimating nonnegative matrices from marginal data[4] and is preferred to other methods due to its computational speed, numerical stability and algebraic simplicity. In Subsection 2.1 we generalize the RAS algorithm to the case of time-varying parameters. The main cycle of the algorithm takes $O(N \times T)$ time. The number of iterations needed for the generalized RAS algorithm to converge is not deterministic, similarly to the original algorithm. However, we observe numerically it takes $O(10^{\log_{10} N-1})$ iterations when $N \in [10^2, 10^3]$. The number of operations needed for learning static paramenters (lines iii and iv) is, in general, a more complicated question. Learning $\boldsymbol{\alpha}$ takes $O(N^2)$ steps, one for each α_{ij}, and each step takes $T+1$ operations, the numerical evaluation of T single integrals and finding the zero of a function. Learning $\mathbf{\Phi}$ takes N steps, one for each $\mathbf{\Phi}_i$, but each step takes a non deterministic number of cycles in order to solve the system of integral equations in Eq. 2.19. In average, each step takes $O(T)$ cycles. Each cycle takes $3 \times T$ operations, i.e. the numerical evaluation of $3 \times T$ single integrals. Finally, the number of iterations for the algorithm to converge is not deterministic but we observe numerically that it is quite constant in the size of the system. Table 1 shows how much time the EM algorithm takes in average to converge.

[4]The problem in Eq. 2.20 can be solved with the RAS algorithm where the generic entry of the matrix is $\frac{e^{\gamma_i+\gamma_j}}{1+e^{\gamma_i+\gamma_j}}$ and the marginal data are represented by the degree sequence.

	$N = 100$	$N = 250$	$N = 500$	$N = 1000$
time (h)	2.8(5)	11(1)	45(4)	151(12)

TABLE 1. The average time of convergence for the algorithm applied to the DAR-TGRG model in the case of undirected networks with $T = 200$. In the model simulations, the parameters are randomly determined as explained in Section 3. The simulations were performed using a Matlab code executed on an ordinary dual-core Intel Core i5, with 8 GB RAM. The number in parenthesis is the uncertainty on the last digit.

3. Monte Carlo simulations

Before applying our methodology to real data, we run Monte Carlo simulations to study the performance of the proposed estimation method when applied both to undirected and to directed networks. Data are generated according to the described models with randomly chosen static parameters. In the case of undirected networks, the DAR(1) model parameters are sampled uniformly in the unit interval. For TGRG we sample $\phi_{1,i} \sim U(-1, 1)$, $\sigma_i \sim U(0, 1)$ and $\phi_{0,i} \sim \mathcal{N}(0, 1)$. For DAR-TGRG, $\alpha_{ij} \sim U(0, 1)$. For both models time-varying parameter θ_i^t follows the stationary AR(1) process of Eq. 1.4. We estimate the models with the proposed Expectation-Maximization (EM) algorithm and compare the results also with the Single Snapshot Inference (SSI). For each simulation, we estimate $\mathbf{\Phi}_i$ for each node i. For DAR-TGRG model we obtain also $\binom{N}{2}$ estimates for α_{ij}, one for each possible couple of nodes (i, j). For both models, we infer the time series of the latent variables $\{\theta_i^t\}_{i=1,\dots,N}^{t=0,1,\dots,T}$. We simulate each model 100 times. In evaluating the goodness of fit of the proposed estimation method, we report the mean absolute relative error for the estimate of parameters. The mean is obtained by averaging over the nodes and the number of simulations. For the time-varying parameters we consider also the time average of the absolute relative errors. A similar study is performed for the case of directed networks, with the exception that for each node we have two fitnesses, $\theta_i^{t,out}$ and $\theta_i^{t,in}$, and as a consequence two sets of static parameters $\mathbf{\Phi} \equiv \{\mathbf{\Phi}^{out}, \mathbf{\Phi}^{in}\}$. For DAR-TGRG model, we obtain $N(N-1)$ estimates for α_{ij}, one for each possible couple of ordered nodes.

The simulation results are summarized in Tables 2, 3, 4, 5 and 6. In Table 2 we show the results for the maximum likelihood estimation of the DAR(1) process in Eq. 1.2. The remaining Tables show that the EM method greatly outperforms the single snapshot inference SSI. Furthermore, we find that the mean absolute relative error for both EM and SSI declines with the number of nodes N since the number of observations increases as N^2, while the number

	$T = 100$	$T = 200$	$T = 500$	$T = 1000$
χ_{ij}	0.28	0.21	0.16	0.10
α_{ij}	0.29	0.19	0.12	0.08

TABLE 2. The mean absolute relative error of the estimates of parameters of the DAR(1) model as a function of the length T of time series. We simulate the DAR(1) model 100 times.

	N	θ_i^t	$\phi_{0,i}$	$\phi_{1,i}$	σ_i
SSI	100	0.30	0.58	0.46	0.69
EM	100	0.22	0.13	0.13	0.06
SSI	200	0.20	0.31	0.27	0.31
EM	200	0.10	0.10	0.10	0.05

TABLE 3. The mean absolute relative error of the estimates of parameters for the TGRG model in the case of undirected networks. We set $T = 200$ and simulate the model 100 times. We compare the proposed Expectation-Maximization algorithm (EM) with the Single Snapshot Inference (SSI).

	N	$\theta_i^{t,out(in)}$	$\phi_{0,i}^{out(in)}$	$\phi_{1,i}^{out(in)}$	$\sigma_i^{out(in)}$
SSI	100	0.31	0.59	0.47	0.71
EM	100	0.23	0.12	0.12	0.06
SSI	200	0.21	0.33	0.29	0.33
EM	200	0.10	0.11	0.10	0.05

TABLE 4. The mean absolute relative error of the estimates of parameters for the TGRG model in the case of directed networks. We set $T = 200$ and simulate the model 100 times. We compare the proposed Expectation-Maximization algorithm (EM) with the Single Snapshot Inference (SSI).

of parameters increases linearly with N. Figure 1 shows the estimated density of the relative errors of the static parameters of the DAR-TGRG model in the case of undirected networks by applying both EM and SSI estimation methods. The SSI leads to biased estimation of the static

	α_{ij}	θ_i^t	$\phi_{0,i}$	$\phi_{1,i}$	σ_i
SSI	0.22	0.29	0.27	0.18	0.22
EM	0.18	0.14	0.15	0.10	0.06

TABLE 5. The mean absolute relative error of the estimates of parameters for the DAR-TGRG model in the case of undirected networks. We compare the EM algorithm with the single snapshot inference SSI. We set $N = 200$, $T = 200$ and simulate the model 100 times.

	α_{ij}	$\theta_i^{t,out(in)}$	$\phi_{0,i}^{out(in)}$	$\phi_{1,i}^{out(in)}$	$\sigma_i^{out(in)}$
SSI	0.22	0.30	0.28	0.18	0.23
EM	0.17	0.14	0.14	0.10	0.05

TABLE 6. The mean absolute relative error of the estimates of parameters for the DAR-TGRG model in the case of directed networks. We compare the EM algorithm with the single snapshot inference SSI. We set $N = 200$, $T = 200$ and simulate the model 100 times.

	$T = 300$	$T = 500$	$T = 1000$
θ_i^t	0.13	0.13	0.12
α_{ij}	0.13	0.10	0.08
$\phi_{0,i}$	0.10	0.09	0.07
$\phi_{1,i}$	0.09	0.08	0.07
σ_i	0.05	0.04	0.04

TABLE 7. Mean absolute relative error of the parameters for the DAR-TGRG model estimated via the EM algorithm. The network is undirected. We set $N = 200$ and the number of simulations is equal to 100.

parameters $\mathbf{\Phi}_i$, while the estimation obtained with the proposed EM method is unbiased. For illustrative purposes, in the right panel of figure 1 we show a typical sample realization and the estimate of time-varying parameter θ_i^t. Clearly, the values inferred with EM track the simulated data fairly closely. Table 7 shows how the mean absolute relative error of the parameters of the DAR-TGRG model of a dynamic undirected graph decreases with the length of the time series.

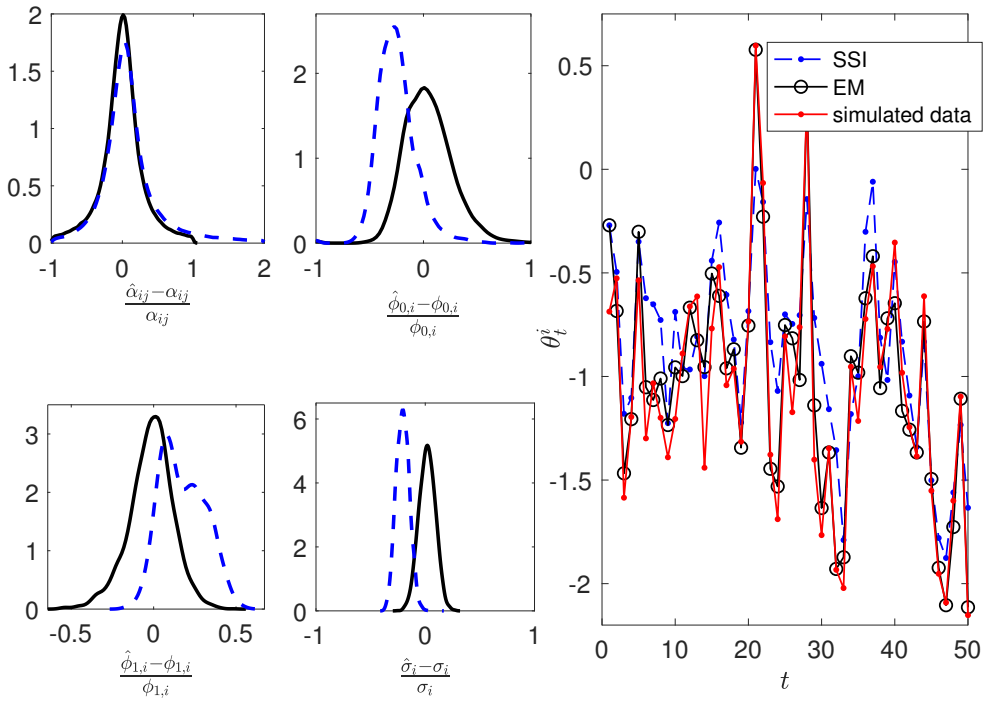

FIGURE 1. Left panels: estimated density of relative errors of α_{ij}, $\phi_{0,i}$, $\phi_{1,i}$ and σ_i. We compare the EM method (solid line) with SSI (dotted line). Right panel: latent dynamics for a generic θ_i^t compared with the inferred one according to EM and SSI.

When the dynamics of the link is affected both by link persistence and by dynamic fitness, neglecting the last one can lead to an overestimation of the importance of the persistence. To show this we simulate a DAR-TGRG model for undirected networks taking $\phi_{1,i}$ equal for all time-varying parameters θ_i^t[5]. Then we estimate α_{ij} according to a DAR(1) model (which neglects dynamic fitness) and to a DAR-TGRG model. Figure 2 shows the mean absolute error of α_{ij} for the two estimations as a function of $\phi_{1,i}$. When this parameter is small both the DAR(1) model and the DAR-TGRG model perform quite equivalently. On the contrary, when the dynamic fitness has a significant persistence due to a high value of $\phi_{1,i}$, the DAR(1) model wrongly imputes this to a link persistence which now has a large bias with respect to the DAR-TGRG model.

[5] $\phi_{1,i}$ determines the autocorrelation of node fitness and as a consequence the link persistence associated with the time-varying marginal.

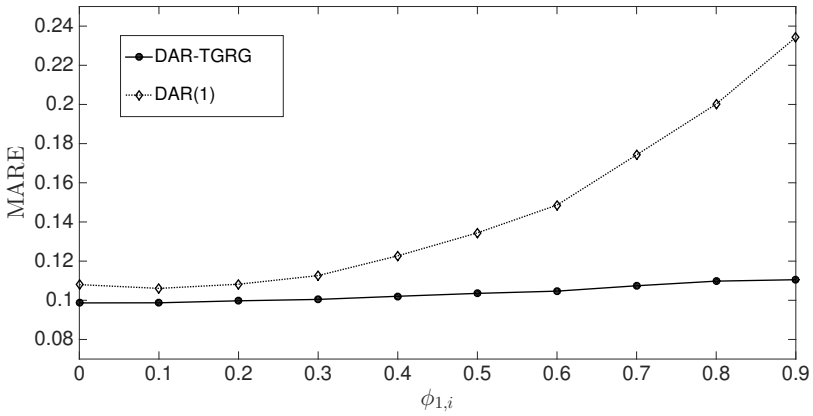

FIGURE 2. Mean Absolute Relative Error (MARE) of the estimates of α_{ij} as a function of the autoregressive coefficient $\phi_{1,i}$ for the time-varying parameters θ_i^t. In the simulation of the DAR-TGRG model for the case of undirected networks, α_{ij}, $\phi_{0,i}$ and σ_i are randomly sampled while $\phi_{1,i}$ are equal for all i. We compare the goodness of fit of the estimates of α_{ij} via the EM method for the DAR-TGRG model (solid line) with the Maximum Likelihood (ML) estimates of α_{ij} according to the DAR(1) model (dotted line). We set $N = 200$ and $T = 400$.

4. Empirical application: understanding link persistence in the interbank market

Trading and credit networks are a natural application case for dynamic networks with persistence, like the one described by our model. Financial institutions lend mutually money on a daily basis and interbank markets are considered an important channel of propagation of systemic risk. While there is a vast literature on the static case, only few papers deal with the dynamic property of interbank networks. The static fitness model has been proved to characterize quantitatively several topological properties of the e-MID network (De Masi et al. (2006), Musmeci et al. (2013)), to outperform other network models in the problem of reconstructing the e-MID network from limited information (Gabrielli et al. (2014), Mazzarisi and Lillo (2017)) and to give useful insights for systemic risk analysis of the interbank market (Cimini et al. (2015)). The ability of the fitness model to describe the static interbank network indicates that the size of two banks correlates with the existence of a credit between them. However it has been documented (Iori et al. (2015)) the presence of memory effects in the process of network formation for interbank markets, according to the idea that a borrower, having asked for a loan many times to a lender in the past, is more likely to borrow from that lender again in

the future than from other lenders, with which the borrower has never (or infrequently) interacted. Finger and Lux (2017) considered a similar problem in a behavioral model of network formation, finding that link persistence is very important in e-MID.

In this section we estimate our dynamic model on data of an interbank market to disentangle the relative importance of fitness and link persistence in determining the future state of the network. This will allow also to perform a forecasting exercise to predict the existence of a credit relation between two banks.

4.1. **Data.** We investigate data from the electronic Market of Interbank Deposit (e-MID), a market where banks extend loans to one another for a specified term and/or collateral. A significant fraction of interbank loans are for maturities of one week or less, the majority being overnight. The e-MID is an electronic market in the Euro Area and it was founded in Italy in 1990 for Italian Lira transactions and denominated in Euros in 1999. According to the "Euro Money Market Study 2006" published by the European Central Bank in February 2007, e-MID accounted for 17% of total turnover in unsecured money market in the Euro Area. More recently the amount of overnight lending in e-MID has significantly declined, especially around the sovereign debt crisis, see Barucca and Lillo (2017) . The e-MID network has been thoroughly studied to understand bank liquidity management, as for instance in Iori et al. (2008), Finger et al. (2013), Barucca and Lillo (2017) .

The dataset contains the edge list of all credit transactions in each day from March 9^{th}, 2012 to February 27^{th}, 2015. In our analysis, we investigate the interbank network aggregated weekly. Each network snapshot of interbank deposits is constructed from the list of transactions where a bank, the lender, extends a loan to another bank, the borrower, that repays the loan in seven days, at most. Hence, we exclude loans with a term larger than a week. However, we account approximately for the 92% of all the traded volume in the market since there are few credit relations with longer maturity. Then, we describe the e-MID weekly network with the unweighted and directed adjacency matrix $\boldsymbol{A}^t$: a generic element A^t_{ij} is 1 if the bank i lends money at least once to bank j during the week t, 0 otherwise. We do not consider banks that interact less than 5% of times in the considered period, i.e. in a period of $T = 156$ weeks a bank has at least a credit relation for more than 7 weeks. Hence, the credit network is formed by $N = 98$ banks.

4.2. **Estimated fitness and link persistence in e-MID.** We estimate the three models on the time series of e-MID networks. Figure 3 shows the estimated density of the α_{ij} link parameters (left panel) and of the $\phi_{1,j}$ node parameters (right panel) for the different model. We see that the DAR(1) model estimates larger α_{ij} parameters, i.e. larger link persistence, than the DAR-TGRG model. Similarly, the TGRG model estimates larger $\phi_{1,j}$, i.e larger

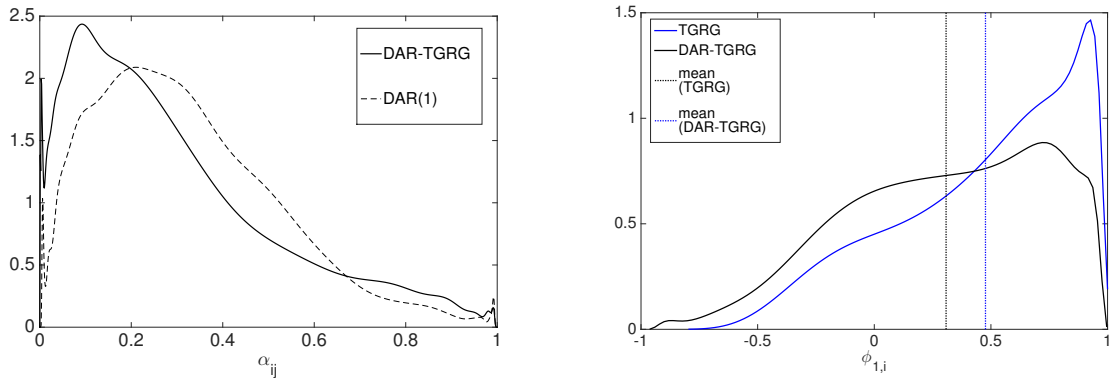

FIGURE 3. Left panel: distributions of α_{ij} estimated via EM on e-MID data. The solid line refers to the DAR-TGRG model while the dotted line refers to the DAR(1) model. Right panel: distribution of parameters $\phi_{1,i}$ estimated via EM. The black line refers to DAR-TGRG while the blue line to TGRG. The dotted lines represent the mean of the two distributions.

fitness persistence, than the complete DAR-TGRG model. Thus the full model balances the relative role of the two persistence mechanisms.

Node fitness is a latent variable whose time evolution is not observed but inferred according to models of temporal networks. However it is interesting to ask if there exists an observable quantity correlated with it. We show that for the considered dataset, node fitness is correlated with the bank exposure in the e-MID interbank market[6].

In fact we observe that the quantity $x_i^{t,out(in)} \equiv e^{\theta_i^{t,out(in)}} \quad \forall i = 1, ..., N$ estimated on data for both TGRG and DAR-TGRG models is strongly correlated with the corresponding bank's exposure in e-MID for the considered week t, see the top left panel in Figure 4. This result suggests that, at a given time, banks with larger exposures are the nodes with larger fitnesses $\theta_i^{t,out(in)}$ or equivalently with larger degrees. Furthermore, the time-varying fitness of a node is correlated significantly with its bank exposure (see the top right panel of Figure 4). Finally, in the bottom panel of Figure 4 we show an example of this behavior for node 3 whose correlation coefficient is $\rho_{x_3^t, s_3^t} \approx 0.90$. Thus the dynamic fitness model can be seen as a procedure allowing to have some insights on bank exposures having only information on the binary network.

[6]Exposure of bank i is defined as the strength of node i in the weighted network. We refer to it as $s_i^{t,out(in)}$ for generic node i. The node out-strength corresponds to the bank asset exposure in e-MID while the node in-strength to the liability.

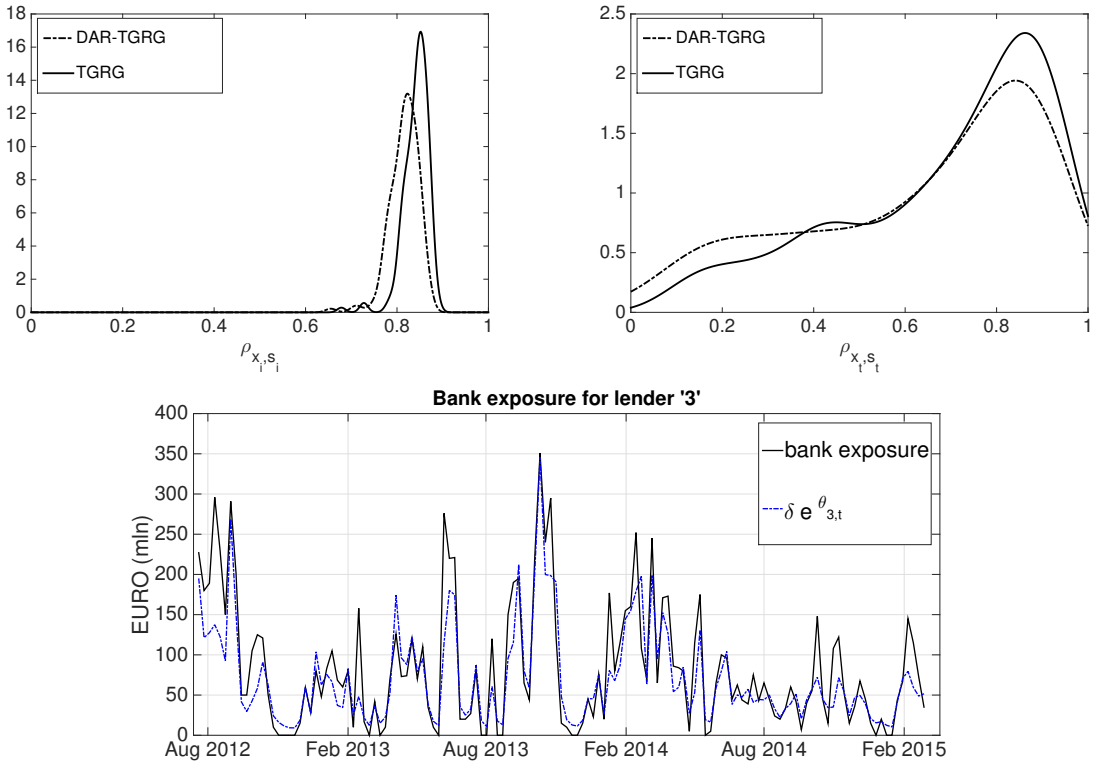

FIGURE 4. Density estimation of cross sectional (top left panel) and temporal (top right panel) Spearman correlation between the inferred $x_i^{t,out(in)} \equiv e^{\theta_i^{t,out(in)}}$ and the corresponded bank exposure s_i^t in e-MID. Bottom panel: an example of time-varying fitness compared with the bank exposure for node '3'. The parameter δ is chosen in such a way that the maxima of the two time series correspond.

4.3. **Link stability and preferential trading in e-MID.** For credit networks like e-MID, the preferential linkage mechanism reflects the presence of banks which trade preferentially each others. Preferential trading between banks can be detected by comparing empirically observed trading relationships with a null hypothesis that assumes random trading. Hatzopoulos et al. (2015) have introduced a statistical test to assess the statistical significance of the observed interbank transactions in order to reveal preferential credit relationships among banks. We apply the same statistical test to show that preferential trading relations in e-MID are associated with link stability.

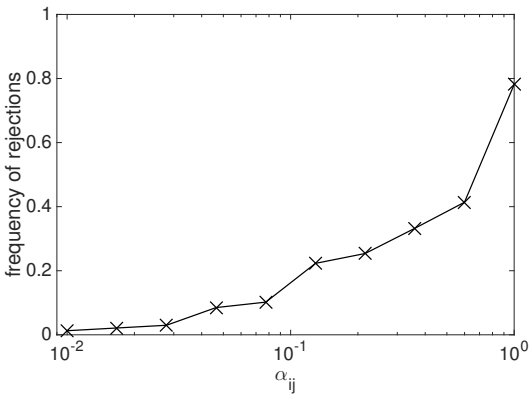

FIGURE 5. Fraction of statistically validated links according to the test in Hatzopoulos et al. (2015) conditional to the value of the estimated α_{ij} parameter measuring the link persistence in the TGRG-DAR model.

Following Hatzopoulos et al. (2015) we apply the test to the weekly aggregated e-MID data split in time windows of 3-maintenance periods[7] In each time window and for each link (i, j) we count the number of times n_{ij}^{lb} bank j borrowed money from bank i. Then, let n_i^l be the number of times bank i lent money to any other banks and let n_j^b be the number of times bank j borrowed money from any other bank. Finally, let us define N_T as the total number of trades among banks in the system for the considered 3-maintenance period. Under the null hypothesis of random trading, n_{ij}^{lb} follows the hypergeometric distribution $H\left(n_{ij}^{lb}|n_i^l, n_j^b, N_T\right)$. Hatzopoulos et al. associate preferential trading with over-expressed number of links with respect to the null hypothesis of random trading, i.e. they use the hypergeometric distribution to associate a p-value with the observed number n_{ij}^{lb}. Preferential trading relations $i \to j$ are the ones rejected according to the statistical test, i.e. with a p-value smaller than the threshold value $\frac{0.05}{a}$ where a is the Bonferroni correction to avoid a large number of false positive validated links because of the multiple hypothesis testing, see Hatzopoulos et al. (2015) for more details.

Figure 5 shows the frequency of rejection for the statistical test by Hatzopoulos et al. (2015) conditional to the estimated α_{ij} parameter measuring the link persistence in the TGRG-DAR model. The clear monotonic behavior indicates that link stability is statistically associated with preferential trading detected according to Hatzopoulos et al. (2015) .

4.4. **Forecasting links.** Finally we compare the proposed network models in their out-of-sample link forecasting performance. We use the first 106 weekly network observations for model estimation and the last 50 as our out-of-sample period. In the training phase we estimate the

[7]The period of time in which credit institutions have to comply with the minimum reserve requirements is called the reserve maintenance period. Each reserve maintenance period is equivalent to one calendar month and we aggregate the maintenance periods in groups of three. Hence, we consider 12 3-maintenance periods ranging from April 2^{nd}, 2012 to February 27^{th}, 2015.

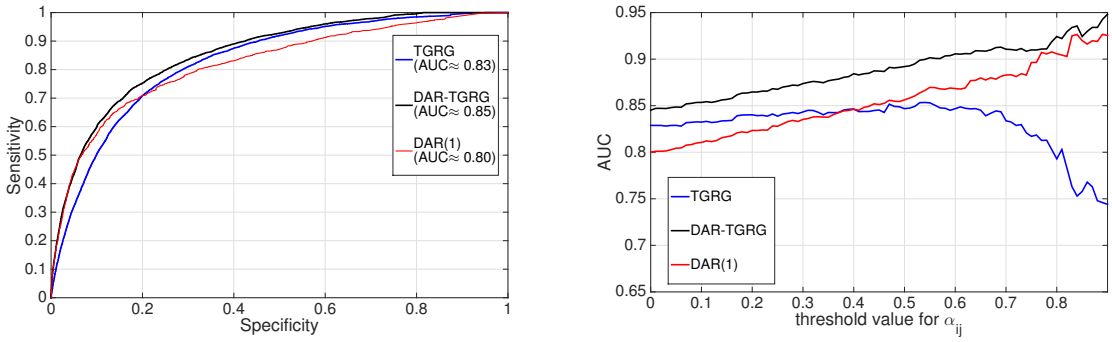

FIGURE 6. Left panel: ROC curve drawn according to the out-of-sample forecasting exercise: TGRG (blue line), DAR-TGRG (black line) and DAR(1) (red line). Right panel: Area Under the Curve (AUC) of the three models as a function of the threshold value for α_{ij} inferred according to DAR-TGRG.

static parameters for the three models and then we adopt the following forecast scheme based on on-line inference. Rolling over the out-of-sample period, at each week t we use the new observed snapshot $\boldsymbol{A}^t$ to infer the expected $\hat{\Theta}^t$ via Eq. 2.6. Then, for DAR-TGRG model we produce the one-step-ahead forecast for each link as

$$\mathbb{E}[A_{ij}^{t+1}|A_{ij}^t,\hat{\theta}_i^t,\hat{\theta}_j^t] = \int d\theta_i^{t+1}d\theta_j^{t+1}\mathbb{P}[A_{ij}^{t+1}=1|A_{ij}^t,\theta_i^{t+1},\theta_j^{t+1}]n(\theta_i^{t+1}|\hat{\theta}_i^t)n(\theta_j^{t+1}|\hat{\theta}_j^t) =$$

$$= \alpha_{ij}A_{ij}^t + (1-\alpha_{ij})\int_0^\infty \frac{d\omega}{2}p_{PG}(\omega)\frac{e^{\frac{-4\omega(\phi_{0,i}+\phi_{1,i}\hat{\theta}_i^t+\phi_{0,j}+\phi_{1,j}\hat{\theta}_j^t)^2+(\sigma_i^2+\sigma_j^2)+4(\phi_{0,i}+\phi_{1,i}\hat{\theta}_i^t+\phi_{0,j}+\phi_{1,j}\hat{\theta}_j^t)}{8(1+\omega(\sigma_i^2+\sigma_j^2))}}}{\sqrt{1+\omega(\sigma_i^2+\sigma_j^2)}}, \tag{4.1}$$

where we have applied the result of Polson et al. (2013) as before. The one-step-ahead forecast for the TGRG model is simply obtained by putting α_{ij} equal to 0 in Eq. 4.1. The one-step-ahead forecast for DAR(1) model is a standard result of time series analysis given by

$$\mathbb{E}[A_{ij}^{t+1}|A_{ij}^t] = \alpha_{ij}A_{ij}^t + (1-\alpha_{ij})\chi_{ij}. \tag{4.2}$$

In comparing the forecasting performances of the models, we restrict to those links which are not always zero in the out-of-sample period. Finally, the forecasting perfomance of each model is assessed by constructing the Receiving Operating Characteristic (ROC) curve (see Friedman et al. (2001) for the definition of ROC curve).

The results are summarized in Figure 6. In the left plot, we compare the three ROC curves and we can notice how the DAR-TGRG model (slightly) outperforms the other models. Furthermore, in the right plot we show the area under the curve (AUC) as a function of a threshold for $\hat{\alpha}_{ij}$ estimated according to DAR-TGRG model. In other words, we compare the AUC considering only the links for which the $\hat{\alpha}_{ij}$ estimated by the the DAR-TGRG model is larger

than a threshold value. We find that taking into account both of fitness dynamics and preferential linkage better forecast links, i.e. DAR-TGRG outperforms always the other models. When we consider links with both high and low persistence, the TGRG model outperforms the DAR(1) network model, that is the evolution of the network topology is more important than preferential linkage in determining the average characteristics of the e-MID network. However, the link copying mechanism associated with the DAR(1) model characterizes better than the fitness dynamics the persistence pattern associated with a smaller set of links representing the preferential relations among banks. In fact, there exists a value of the threshold (around 0.4) after which the AUC associated with the DAR(1) model is larger than the one for TGRG.

5. Conclusions

In this paper we introduce a novel state-of-the-art statistical methodology to describe link persistence and fitness dynamics in temporal networks. We model a Markov dynamics for both observed and unobserved time-varying states which drive the evolution of the network. The analytic tractability of the autoregressive network ensemble we propose allows us to easily calibrate our parameters from the data with a general likelihood maximization iterative procedure. The introduction of the autoregressive dynamics permits link forecasting by taking account of memory properties of the network system. Then, the estimation method we introduce allows online-inference of the time-varying parameters which is particularly useful from a computational point of view to face the problem of link prediction.

The contribution of the paper is twofold. First, the introduction of autoregressive endogenous components displays the clear advantage of describing the network evolution via time-varying states which reproduce the network topology as well as capturing the local property of link persistence, thus going beyond a single snapshot analysis where parameters are chosen for each network snapshot, independently. Second, the analysis on real data from the eMID interbank network from 2012 to 2015 (weekly aggregated) displays the statistical equivalence between link stability, identified by positive value of the persistence parameter, and preferential trading, identified by over-expressed number of trades between counterparties. Hence, our methodology permits to disentangle preferential trading from random trading in dynamic trading networks such as the eMID money market. Finally, the forecasting performance of the model points out both fitness dynamics and link persistence as linkage mechanisms in the process of network formation for the credit market.

As future outlooks, the formalism discussed in the paper could also be applied to more general memory kernel function of the autoregressive model governing the evolution of the system as well as it could permit the introduction of exogenous factors driving the fitness dynamics or

the local link probability. Furthermore, a challenging issue is the introduction of a dependence structure for the dynamic fitnesses. We also note that the estimation method we introduce to obtain our results for dynamic networks is quite general, and could be used to obtain similar results for other types of fitness dynamics. Finally, in temporal networks nodes may become inactive or could disappear at any time, especially in the case of long time intervals. In the case of observed node activity, the proposed model can be generalized to capture this behavior and a generalization of the proposed MMAP approach for the model estimation can be obtained.

Acknowledgment

We acknowledge financial support from the grant SNS16LILLB "Financial networks: statistical models, inference, and shock propagation". FL acknowledges support by the European Community's H2020 Program under the scheme INFRAIA-1- 2014-2015: Research Infrastructures, grant agreement no. 654024 SoBigData: Social Mining & Big Data Ecosystem. PB acknowledges support from the University of Zurich and the London Institute for Mathematical Sciences, and from FET Project DOLFINS nr. 640772. DT acknowledges support from the SNS grant SNS18ATANTARI and from Indam-GNFM.

References

Bacharach (1965). Bacharach, Michael. "Estimating nonnegative matrices from marginal data." International Economic Review 6.3 (1965): 294-310.

Bartolucci at al. (2018). Bartolucci, F., Marino, M. F., Pandolfi, S. (2018). Dealing with reciprocity in dynamic stochastic block models. Computational Statistics & Data Analysis, 123, 86-100.

Barucca and Lillo (2017). Barucca, Paolo, and Fabrizio Lillo. "The organization of the interbank network and how ECB unconventional measures affected the e-MID overnight market." F. Comput Manag Sci, 1-21 (2017).

Barucca et al. (2017). Barucca, P., Lillo, F., Mazzarisi, P., & Tantari, D. "Disentangling group and link persistence in dynamic stochastic block models". arXiv preprint arXiv:1701.05804. (2017, Unpublished results).

Brauning and Koopman (2016). Brauning, Falk, and Siem Jan Koopman. "The Dynamic Factor Network Model with an Application to Global Credit-Risk." (2016, Unpublished results).

Caldarelli et al. (2002). Caldarelli, G., Capocci, A., De Los Rios, P., & Munoz, M. A. "Scale-free networks from varying vertex intrinsic fitness." Physical review letters, 89(25), 258702. (2002).

Chatterjee and Diaconis (2011). Chatterjee, S., Diaconis, P., & Sly, A. (2011). Random graphs with a given degree sequence. *The Annals of Applied Probability*, 21(4), 1400-1435.

Chen (2003). Chen, Zhe. "Bayesian filtering: From Kalman filters to particle filters, and beyond." Statistics 182.1 (2003): 1-69.

Cimini et al. (2015). Cimini, G., Squartini, T., Garlaschelli, D., & Gabrielli, A. "Systemic risk analysis on reconstructed economic and financial networks." Scientific reports, 5. (2015).

Cocco et al. (2009). Cocco, Joao F., Francisco J. Gomes, and Nuno C. Martins. "Lending relationships in the interbank market." Journal of Financial Intermediation 18.1 (2009): 24-48.

De Masi et al. (2006). De Masi, Giulia, Giulia Iori, and Guido Caldarelli. "Fitness model for the Italian interbank money market." Physical Review E 74.6 (2006): 066112.

Dempster et al. (1977). Dempster, Arthur P., Nan M. Laird, and Donald B. Rubin. "Maximum likelihood from incomplete data via the EM algorithm." Journal of the royal statistical society. Series B (methodological) (1977): 1-38.

Di Maggio et al. (2017). Di Maggio, M., Kermani, A., & Song, Z. (2017). The value of trading relations in turbulent times. Journal of Financial Economics, 124(2), 266-284.

Durante and Dunson (2014). Durante, Daniele, and David B. Dunson. "Bayesian dynamic financial networks with time-varying predictors." Statistics & Probability Letters 93 (2014): 19-26.

Durante and Dunson (2016). Durante, Daniele, and David B. Dunson. "Locally adaptive dynamic networks." The Annals of Applied Statistics 10.4 (2016): 2203-2232.

Friel et al. (2016). Friel, N., Rastelli, R., Wyse, J., & Raftery, A. E. (2016). Interlocking directorates in Irish companies using a latent space model for bipartite networks. Proceedings of the National Academy of Sciences, 113(24), 6629-6634.

Finger et al. (2013). Finger, Karl, Daniel Fricke, and Thomas Lux. "Network analysis of the e-MID overnight money market: the informational value of different aggregation levels for intrinsic dynamic processes." Computational Management Science 10.2-3 (2013): 187-211.

Finger and Lux (2017). Finger, Karl and Thomas Lux. (2017) "Network Formation in the Interbank Market: An Application of the Actor-Oriented Model." Social Networks 48 (2017): 237-249

Friedman et al. (2001). Friedman, Jerome, Trevor Hastie, and Robert Tibshirani. "The elements of statistical learning." Vol. 1. Springer, Berlin: Springer series in statistics, (2001).

Gabrielli et al. (2014). Gabrielli, A., Battiston, S., Caldarelli, G., Musmeci, N., & Puliga, M. "Reconstructing topological properties of complex networks from partial information using the Fitness Model." In APS March Meeting Abstracts. (2014).

Gabrielsen (1978). Gabrielsen, A. (1978). Consistency and identifiability. Journal of Econometrics, 8(2), 261-263.

Garlaschelli and Loffredo (2004). Garlaschelli, Diego, and Maria I. Loffredo. "Fitness-dependent topological properties of the world trade web." Physical review letters 93.18 (2004): 188701.

Ghasemian at al. (2016). Ghasemian, A., Zhang, P., Clauset, A., Moore, C., & Peel, L. "Detectability thresholds and optimal algorithms for community structure in dynamic networks." Physical Review X, 6(3), 031005. (2016).

Giraitis et al. (2016). Giraitis, L., Kapetanios, G., Wetherilt, A., & Zikes, F. "Estimating the dynamics and persistence of financial networks, with an application to the Sterling money market". Journal of Applied Econometrics, 31(1), 58-84. (2016)

Hanneke et al. (2010). Hanneke, Steve, Wenjie Fu, and Eric P. Xing. "Discrete temporal models of social networks." Electronic Journal of Statistics 4 (2010): 585-605.

Hatzopoulos et al. (2015). Hatzopoulos, V., Iori, G., Mantegna, R. N., Miccichè, S., & Tumminello, M. "Quantifying preferential trading in the e-MID interbank market". Quantitative Finance, 15(4), 693-710. (2015).

Heaukulani and Ghahramani (2013). Heaukulani, C., & Ghahramani, Z. (2013, February). Dynamic probabilistic models for latent feature propagation in social networks. In *International Conference on Machine Learning* (pp. 275-283).

Hellmann and Staudigl (2014). Hellmann, Tim, and Mathias Staudigl. "Evolution of social networks." European Journal of Operational Research 234.3 (2014): 583-596.

Hoff et al. (2002). Hoff, Peter D., Adrian E. Raftery, and Mark S. Handcock. "Latent space approaches to social network analysis." Journal of the american Statistical association 97.460 (2002): 1090-1098.

Holland and Leinhardt (1981). Holland, P. W., & Leinhardt, S. (1981). An exponential family of probability distributions for directed graphs. Journal of the american Statistical association, 76(373), 33-50.

Holme and Saramäki (2012). Holme, Petter, and Jari Saramäki. "Temporal networks." Physics reports 519.3 (2012): 97-125.

Iori et al. (2008). Iori, G., De Masi, G., Precup, O. V., Gabbi, G., & Caldarelli, G. "A network analysis of the Italian overnight money market." Journal of Economic Dynamics and Control, 32(1), 259-278. (2008).

Iori et al. (2015). Iori, G., Mantegna, R. N., Marotta, L., Miccichè, S., Porter, J., & Tumminello, M. "Networked relationships in the e-MID Interbank market: A trading model with memory." Journal of Economic Dynamics and Control, 50, 98-116. (2015).

Jacobs and Lewis (1978) (A). Jacobs, Patricia A., and Peter AW Lewis. "Discrete Time Series Generated by Mixtures. III. Autoregressive Processes (DAR (p))." No. NPS55-78-022. Naval Postgraduate School Monterey Calif, (1978).

Jacobs and Lewis (1978) (B). Jacobs, P. A., & Lewis, P. A. (1978). Discrete time series generated by mixtures II: asymptotic properties. Journal of the Royal Statistical Society: Series B (Methodological), 40(2), 222-228.

Kim et al. (2017). Kim, B., Lee, K., Xue, L., & Niu, X. "A Review of Dynamic Network Models with Latent Variables." arXiv preprint arXiv:1711.10421 (2017).

Krivitsky and Handcock (2014). Krivitsky, Pavel N., and Mark S. Handcock. "A separable model for dynamic networks." Journal of the Royal Statistical Society: Series B (Statistical Methodology) 76.1 (2014): 29-46.

Little et al. (1983). Little, Roderick JA, and Donald B. Rubin. "On jointly estimating parameters and missing data by maximizing the complete-data likelihood." The American Statistician 37.3 (1983): 218-220.

Lee et al. (2017). Lee, Jihui, Gen Li, and James D. Wilson. "Varying-coefficient models for dynamic networks." arXiv preprint arXiv:1702.03632 (2017, Unpublished results).

Luo et al. (2017). Luo, J., Qin, H., Yan, T., & Zeyneb, L. (2017). A note on asymptotic distributions in directed exponential random graph models with bi-degree sequences. Communications in Statistics-Theory and Methods, 46(18), 8852-8864.

Matias and Miele (2017). Matias, C., & Miele, V. (2017). Statistical clustering of temporal networks through a dynamic stochastic block model. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 79(4), 1119-1141.

Mazzarisi and Lillo (2017). Mazzarisi, Piero, and Fabrizio Lillo. "Methods for Reconstructing Interbank Networks from Limited Information: A Comparison." Econophysics and Sociophysics: Recent Progress and Future Directions. Springer, Cham, (2017). 201-215.

Musmeci et al. (2013). Musumeci, N., Battiston, S., Caldarelli, G., Puliga, M., & Gabrielli, A. "Bootstrapping topological properties and systemic risk of complex networks using the fitness model." Journal of Statistical Physics, 151(3-4), 720-734. (2013).

Neyman et al. (1948). Neyman, Jerzy, and Elizabeth L. Scott. "Consistent estimates based on partially consistent observations." Econometrica 16.1 (1948): 1-32.

Newman et al. (2001). Newman, Mark EJ, Steven H. Strogatz, and Duncan J. Watts. "Random graphs with arbitrary degree distributions and their applications." Physical review E 64.2 (2001): 026118.

Papadopoulos and Kleineberg (2019) . Papadopoulos, F., & Kleineberg, K. K. (2019). Link persistence and conditional distances in multiplex networks. Physical Review E, 99(1), 012322.

Park and Newman (2004). Park, Juyong, and Mark EJ Newman. "Statistical mechanics of networks." Physical Review E 70.6 (2004): 066117.

Peixoto and Rosvall (2017). Peixoto, Tiago P., and Martin Rosvall. "Modeling sequences and temporal networks with dynamic community structures." Nature Communications 8.582 (2017).

Polson et al. (2013). Polson, Nicholas G., James G. Scott, and Jesse Windle. "Bayesian inference for logistic models using Pólya-Gamma latent variables." Journal of the American statistical Association 108.504 (2013): 1339-1349.

Richard et al. (2014). Richard, Emile, Stéphane Gaiffas, and Nicolas Vayatis. "Link prediction in graphs with autoregressive features." Journal of Machine Learning Research 15.1 (2014): 565-593.

Rinaldo et al. (2013). Rinaldo, A., Petrovic, S., & Fienberg, S. E. (2013). Maximum lilkelihood estimation in the β-model. The Annals of Statistics, 41(3), 1085-1110.

Sarkar and Moore (2005). Sarkar, Purnamrita, and Andrew W. Moore. "Dynamic social network analysis using latent space models." ACM SIGKDD Explorations Newsletter 7.2 (2005): 31-40.

Sewell and Chen (2015). Sewell, Daniel K., and Yuguo Chen. "Latent space models for dynamic networks." Journal of the American Statistical Association 110.512 (2015): 1646-1657.

Staudigl (2013). Staudigl, Mathias. "Co-evolutionary dynamics and Bayesian interaction games." *International Journal of Game Theory* 42.1 (2013): 179-210.

Tsay (2005). Tsay, R. S. (2005). Analysis of financial time series. *NJ: Wiley-Interscience.*

Weisbuch et al. (2000). Weisbuch, Gerard, Alan Kirman, and Dorothea Herreiner. "Market organisation and trading relationships." The economic journal 110.463 (2000): 411-436.

Williams et al. (2019). Williams, Oliver E., Fabrizio Lillo and Vito Latora, "Effects of memory on spreading processes in non-Markovian temporal networks." New Journal of Physics 21, (2019), 043028.

Xu and Hero (2014). Xu, Kevin S., and Alfred O. Hero. "Dynamic stochastic blockmodels for time-evolving social networks." IEEE Journal of Selected Topics in Signal Processing 8, no. 4 (2014): 552-562.

Xu (2015). Xu, Kevin. "Stochastic Block Transition Models for Dynamic Networks." In AISTATS. (2015).

Yan and Xu (2013). Yan, T., & Xu, J. (2013). A central limit theorem in the β-model for undirected random graphs with a diverging number of vertices. *Biometrika*, 100(2), 519-524.

Yang et al. (2011). Yang, Tianbao, Yun Chi, Shenghuo Zhu, Yihong Gong, and Rong Jin. "Detecting communities and their evolutions in dynamic social networks: a Bayesian approach." Machine learning 82, no. 2 (2011): 157-189.

Zhang et al. (2017). Zhang, Xiao, Cristopher Moore, and Mark EJ Newman. "Random graph models for dynamic networks." The European Physical Journal B 90.10 (2017): 200.

Appendix A. Comparison between Maximum A Posteriori approaches

A.1. **JMAP approach.** For the sake of completeness, in this appendix section we briefly describe the Joint Maximum A Posteriori (JMAP) approach for the inference of the DAR-TGRG model (the inference method for the TGRG model is obtained by setting parameters α_{ij} $\forall i, j = 1, ..., N$ equal to zero, similarly to the MMAP approach) in order to compare it with the MMAP presented in Section 2.

JMAP inference consists in the maximization of the joint posterior

$$\mathbb{P}(\boldsymbol{\Theta}, \Pi|\boldsymbol{A}) \propto \mathbb{P}(\boldsymbol{A}|\boldsymbol{\Theta}, \Pi)\mathbb{P}(\boldsymbol{\Theta}|\Pi)\mathbb{P}(\Pi) \tag{A.1}$$

for both $\boldsymbol{\Theta}$ and Π, thus treating the latent variables as the static parameters. By assuming uniform prior, the posterior of both static and time-varying parameters in Eq. A.1 for DAR-TGRG reads more explicitly as

$$\mathbb{P}(\boldsymbol{\Theta}, \boldsymbol{\alpha}, \boldsymbol{\Phi}|\boldsymbol{A}) \propto \mathbb{P}(\boldsymbol{A}|\boldsymbol{\Theta}, \boldsymbol{\alpha})\mathbb{P}(\boldsymbol{\Theta}|\boldsymbol{\Phi}). \tag{A.2}$$

In a similar fashion of what we propose in Subsection 2.1, the inference of the time-varying parameters at time t is obtained by maximizing the posterior in Eq. A.2 with respect to Θ^t. By conditioning on the expectation for Θ^{t-1}, *i.e.* $\hat{\Theta}^{t-1}$, and the static parameters, this is in effect equivalent to Eq. 2.6, thus JMAP is equivalent to MMAP regarding the estimation of the latent variables.

Then, the static parameters are estimated by maximizing

$$\max_{\boldsymbol{\alpha}} \log \mathbb{P}(\boldsymbol{A}|\boldsymbol{\Theta}, \boldsymbol{\alpha}), \tag{A.3}$$

and

$$\max_{\boldsymbol{\Phi}} \log \mathbb{P}(\boldsymbol{\Theta}|\boldsymbol{\Phi}). \tag{A.4}$$

By conditioning on the expectations for the latent variables, the maximization problem (A.3) corresponds to solve the following equation for α_{ij},

$$\sum_{t=1}^{T} \frac{\delta_{A_{ij}^t, A_{ij}^{t-1}} - \left(1 + e^{-(\hat{\theta}_i^t + \hat{\theta}_j^t)}\right)^{-1}}{\alpha_{ij}\delta_{A_{ij}^t, A_{ij}^{t-1}} + (1 - \alpha_{ij})\left(1 + e^{-(\hat{\theta}_i^t + \hat{\theta}_j^t)}\right)^{-1}} = 0. \tag{A.5}$$

Finally, the maximization problem in Eq. A.4 is equivalent to the maximum likelihood estimation of a AR(1) process when we use conditions over the latent variables, *i.e.* we consider the inferred $\hat{\boldsymbol{\Theta}}$ as observable variables in the inference of parameters $\boldsymbol{\Phi}$. See Tsay (2005) for details about MLE of the AR(1) model.

Then, we use the following Expectation-Maximization scheme for the JMAP estimation of the DAR-TGRG model:

(1) Assume SSI as starting point, *i.e.* $\tilde{\boldsymbol{\Theta}}$ and $\tilde{\Pi} = \{\tilde{\boldsymbol{\Phi}}, \tilde{\boldsymbol{\alpha}}\}$ obtained as explained in Subsection 2.3;[8]

(2) infer $\hat{\boldsymbol{\Theta}} \equiv \{\hat{\Theta}^t\}^{t=1,\dots T}$ by solving Eq .2.7 with $\tilde{\Pi}$;

(3) learn $\hat{\boldsymbol{\alpha}}$ by solving Eq. A.5 for each possible couple of nodes with previously inferred $\hat{\boldsymbol{\Theta}}$ and $\tilde{\boldsymbol{\Phi}}$;

[8] In the SSI inference of $\boldsymbol{\alpha}$, we use Eq. A.5 instead of Eq. 2.15, but conditioning on $\tilde{\boldsymbol{\Theta}}$ as described in Subsection 2.3.

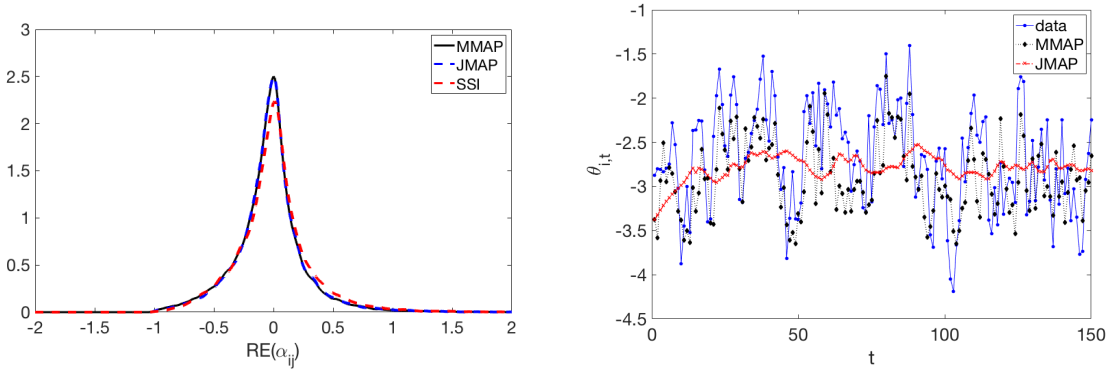

FIGURE 7. Left: distribution of relative errors RE for parameters α_{ij}, inferred according to MMAP (black solid line), JMAP (blue dashed line), and SSI (red dotted line). Data are simulated according to DAR-TGRG model with $N = 300$ nodes, $T = 200$, and model parameters sampled uniformly in the domain of definition. Right: an example of the fitness dynamics following a AR(1) process for a generic fitness θ_i^t generated according to the DAR-TGRG model (blue dots). Black dots represent the fitness inferred according to MMAP, while red crosses represent the fitness inferred according to JMAP.

(4) learn $\hat{\mathbf{\Phi}}$ by MLE of the N independent AR(1) processes, each one describing one time series $\{\hat{\theta}_i^t\}^{t=0,1,\dots,T}$;
(5) update $\tilde{\mathbf{\Theta}} \leftarrow \hat{\mathbf{\Theta}}$;
(6) update $\tilde{\Pi} \leftarrow \hat{\Pi}$;
(7) repeat until convergence.

As already stated previously, this approach can suffer inherent problems with the estimates of static parameters if the number of latent variables increases with the sample size (see below).

A.2. **Comparing SSI, MMAP, and JMAP.** Here, we compare the three different estimation methods, SSI, MMAP, and JMAP, in order to show numerically the consistency of the MMAP approach. According to Monte Carlo simulations of the model DAR-TGRG, the three estimation methods are quite equivalent in inferring the set of parameters $\boldsymbol{\alpha}$ (when sampled uniformly in the unit interaval), see the left panel of Figure 7. However, the three estimation methods differ significantly in the estimation of the latent dynamics. In the right panel of Figure 7, we show a typical example of the inferred fitness according to both JMAP and MMAP, and compared with data simulated according to DAR-TGRG. Notice that the JMAP method underestimates largely the variability of data.

In Figure 8, we show the relative error in the inference of parameters $\Phi \equiv \{\phi_{0,i}, \phi_{1,i}, \sigma_i\}_{i=1,\dots,N}$ of the latent dynamics of both TGRG (left panels) and DAR-TGRG (right panels) models for

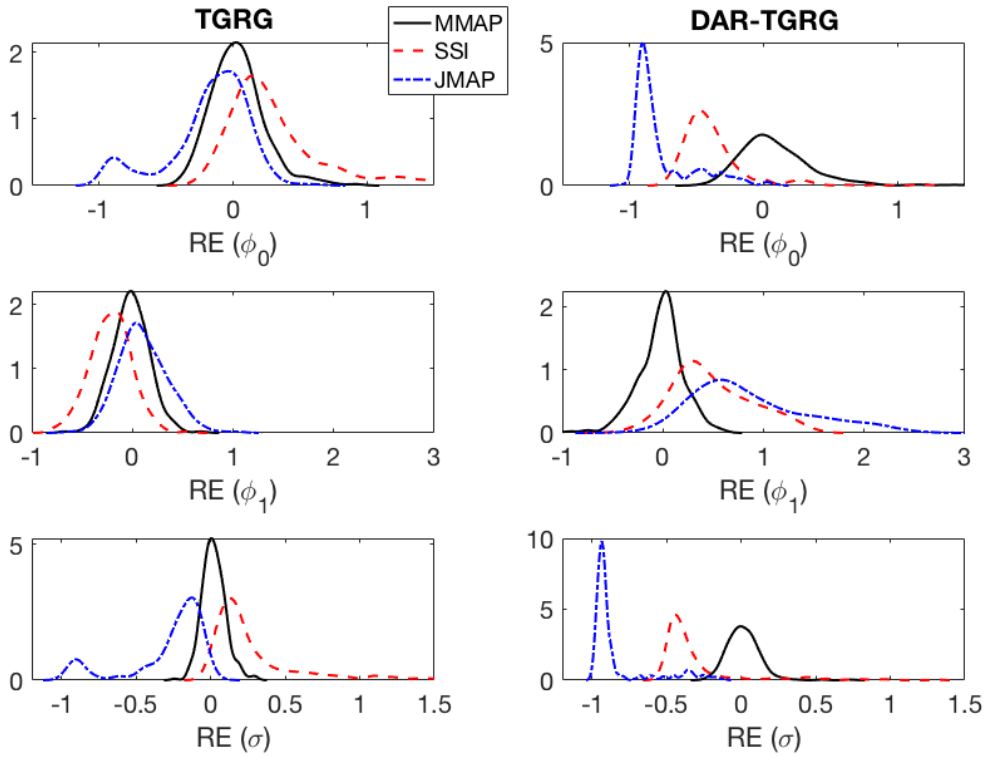

FIGURE 8. Distribution of relative errors RE in the inference of parameters $\Phi_i = \{\phi_{0,i}, \phi_{1,i}, \sigma_i\}$ of the latent AR(1) dynamics for both TGRG (left panels) and DAR-TGRG (right panels) models according to the three estimation methods: MMAP (black), SSI (red), and JMAP (blue). Both TGRG and DAR-TGRG models are simulated by sampling the static parameters uniformly in the domain. The convergence conditions are the same for both MMAP and JMAP expectation-maximization schemes.

the three estimation methods, *i.e.* SSI, MMAP, and JMAP. From numerical simulations, we can notice that both SSI and JMAP lead to a bias in the inference of parameters, whereas MMAP outperforms the others:

(1) in the case of DAR-TGRG, both SSI and JMAP infer $\mathbf{\Phi}$ by solving Eq. A.4, without using explicitly the information about link persistence. On the contrary, MMAP exploits this information, specifically the information carried by each link (i, j) is weighted with its memory parameter α_{ij} , see Eq. 2.17. Persistence of links incident to node i tends to increase the persistence of the node degree. If the estimation method is not able to disentangle the origin of persistence, *i.e.* link persistence of the Markovian behavior for the edges or degree persistence because of the node fitness autocorrelated in time, the result is an overestimation of the autoregressive coefficient $\phi_{1,i}$ associated with the node

fitness θ_i, as it happens for both SSI and JMAP (see the middle right panel of Figure 8). The overestimation of fitness persistence is related to the underestimated variability of the inferred latent dynamics, see the right panel of Figure 7.

As a consequence of the overestimation of $\phi_{1,i}$, the parameter $\phi_{0,i}$ is underestimated in order to preserve the overall mean of the latent autoregressive dynamics AR(1) (*i.e.* $\phi_{0,i}/(1-\phi_{1,i})$), see the top right panel of Figure 8.

Finally, since the variance of a AR(1) process is $\sigma_i^2/(1-\phi_{1,i}^2)$, an overestimation of $\phi_{1,i}$ leads to an underestimation of σ_i (once the observed variance is fixed). This is what we observe for the naive estimation method SSI, see the red dotted line in the bottom right panel of Figure 8. For JMAP, this behavior is even more evident because of a feedback effect in the algorithm. In fact, the maximization problem in Eq. 2.6 for the inference of the latent variables is

$$\frac{\partial \log \mathbb{P}(A^t|A^{t-1},\boldsymbol{\alpha},\boldsymbol{\Theta}^t)}{\partial \theta_i^t} - \frac{\theta_i^t - \phi_{0,i} - \phi_{1,i}\theta_i^{t-1}}{\sigma_i^2} = 0, \;\; \forall t = 1,...,T, \tag{A.6}$$

thus an underestimation of σ_i in the M-step has the consequence of weighting less the information about the network and more the (biased) information about the latent dynamics in the previous step, resulting in a further amplification of the bias.

(2) in the case of TGRG (left panels of Figure 8), the bias of both SSI and JMAP estimators is less evident. Nevertheless, it is still present and by comparing the performances of the three methods, MAP outperforms the others.

Appendix B. Two-point probability mass function for TGRG

In the TGRG model, the node fitness is autocorrelated in time when $\phi_{i,1} \neq 0$. An autocorrelated fitness reflects the autocorrelation of the degree and ultimately of all links incident to the node[9]. A positive autocorrelated fitness is associated with persistence of links. This effect can be characterized by studying the two-point distribution function or equivalently the autocorrelation function. In the TGRG model we can compute semi-analytically the the two-point

[9]This is always true in the case of finite network size.

distribution function,

$$\mathbb{P}(A_{ij}^t=1,A_{ij}^{t-\tau}=1)=\int d\theta_i^t d\theta_j^t d\theta_i^{t-\tau}d\theta_j^{t-\tau}\mathbb{P}(A_{ij}^t=1|\theta_i^t,\theta_j^t)\mathbb{P}(A_{ij}^{t-\tau}=1|\theta_i^{t-\tau},\theta_j^{t-\tau})p(\theta_i^t,\theta_i^{t-\tau})p(\theta_j^t,\theta_j^{t-\tau})=$$
$$=\int d\theta_i^{t-\tau}d\theta_j^{t-\tau}\frac{1}{1+e^{-(\theta_i^{t-\tau}+\theta_j^{t-\tau})}}n(\theta_i^{t-\tau})n(\theta_j^{t-\tau})\times$$
$$\times\int\left[\prod_{q=1}^{\tau-1}\prod_{a=i,j}f(\theta_a^{t-\tau+q}|\theta_a^{t-\tau+(q-1)})d\theta_a^{t-\tau+q}\right]\int d\theta_i^t d\theta_j^t\frac{1}{1+e^{-(\theta_i^t+\theta_j^t)}}f(\theta_i^t|\theta_i^{t-1})f(\theta_j^t|\theta_j^{t-1})=$$
$$=\int d\theta_i^{t-\tau}d\theta_j^{t-\tau}\frac{1}{1+e^{-(\theta_i^{t-\tau}+\theta_j^{t-\tau})}}n(\theta_i^{t-\tau})n(\theta_j^{t-\tau})\int_0^\infty\frac{d\omega}{2}p_{PG}(\omega)K^\tau(\omega|\theta_i^{t-\tau},\theta_j^{t-\tau}) \tag{B.1}$$

where we have applied the result of Polson et al. (2013) as before and

$$K^\tau(\omega|\theta_i^{t-\tau},\theta_j^{t-\tau})=\frac{e^{\frac{-4\omega(\mu_i^\tau+\mu_j^\tau)^2+((\sigma_i^\tau)^2+(\sigma_j^\tau)^2)+4(\mu_i^\tau+\mu_j^\tau)}{8(1+\omega((\sigma_i^\tau)^2+(\sigma_j^\tau)^2))}}}{\sqrt{1+\omega((\sigma_i^\tau)^2+(\sigma_j^\tau)^2)}}$$

with

$$\mu_a^\tau=\phi_{0,a}\left(\sum_{t=0}^{\tau-1}(\phi_{1,a})^t\right)+(\phi_{1,a})^\tau\theta_a^{t-\tau}\quad a=i,j$$

$$(\sigma_a^\tau)^2=\sigma_a^2(\sum_{t=0}^{\tau-1}(\phi_{1,a}^2)^t)\quad a=i,j.$$

The last recursive formulas are obtained by integrating over the Gaussian transition probabilities in Eq. B.1. Let us notice that μ_a^τ and $(\sigma_a^\tau)^2$ converge to the mean and the variance of the marginal distribution for θ_a^t in the limit $\tau\to\infty$ as we can expect for the standard AR(1) process.

Then, the two-point distribution function can be obtained by integrating over the Gaussian marginals, i.e. $n(\theta_i^{t-\tau})$ and $n(\theta_j^{t-\tau})$ with $\tilde{\mu}_a\equiv\frac{\phi_{0,a}}{1-\phi_{1,a}}$ and $\tilde{\sigma}_a^2\equiv\frac{\sigma_a^2}{1-\phi_{1,a}^2}$ $a=i,j$ as the mean and the variance of the Gaussian marginal distribution for $\theta_a^{t-\tau}$, and finally by performing the numerical integration over the probability density function associated with the Polya-Gamma distribution. It is

$$\mathbb{P}(A_{ij}^t=1,A_{ij}^{t-\tau}=1)=\int_0^\infty\frac{d\omega}{2}p_{PG}(\omega)\int_0^\infty\frac{d\zeta}{2}p_{PG}(\zeta)\times$$
$$\times\frac{e^{g(\omega,\zeta,\phi_{0,i},\phi_{0,j},\phi_{1,i},\phi_{1,j},\sigma_i,\sigma_j)}}{\sqrt{1+\zeta(\tilde{\sigma}_i^2+\tilde{\sigma}_j^2)+\omega(C_i^\tau\sigma_i^2+C_j^\tau\sigma_j^2+(B_i^\tau)^2\tilde{\sigma}_i^2+(B_j^\tau)^2\tilde{\sigma}_j^2)+\zeta\omega(\tilde{\sigma}_i^2(C_i^\tau\sigma_i^2+C_j^\tau\sigma_j^2)+\tilde{\sigma}_j^2(C_i^\tau\sigma_i^2+C_j^\tau\sigma_j^2)+\tilde{\sigma}_i^2\tilde{\sigma}_j^2(B_i^\tau-B_j^\tau)^2)}} \tag{B.2}$$

where

$$g(\omega,\zeta,\mathbf{\Phi}_i,\mathbf{\Phi}_j)=\frac{h(\omega,\zeta,\mathbf{\Phi}_i,\mathbf{\Phi}_j)}{8\left(1+\zeta(\tilde{\sigma}_i^2+\tilde{\sigma}_j^2)+\omega(C_i^\tau\sigma_i^2+C_j^\tau\sigma_j^2+(B_i^\tau)^2\tilde{\sigma}_i^2+(B_j^\tau)^2\tilde{\sigma}_j^2)+\zeta\omega(\tilde{\sigma}_i^2(C_i^\tau\sigma_i^2+C_j^\tau\sigma_j^2)+\tilde{\sigma}_j^2(C_i^\tau\sigma_i^2+C_j^\tau\sigma_j^2)+\tilde{\sigma}_i^2\tilde{\sigma}_j^2(B_i^\tau-B_j^\tau)^2)\right)}$$

and

$$
\begin{aligned}
h(\omega,\zeta,\mathbf{\Phi}_i,\mathbf{\Phi}_j) =& 4(T_i^\tau + T_j^\tau) + 4(1+B_i^\tau)\tilde{\mu}_i + 4(1+B_j^\tau)\tilde{\mu}_j + C_i^\tau\sigma_i^2 + C_j^\tau\sigma_j^2 + (1+B_i^\tau)^2\tilde{\sigma}_i^2 + (1+B_j^\tau)^2\tilde{\sigma}_j^2 + \\
&+\zeta\ (\tilde{\sigma}_i^2\tilde{\sigma}_j^2(B_i^\tau - B_j^\tau)^2 - 4(\tilde{\mu}_i+\tilde{\mu}_j)^2 + \\
&\quad + \tilde{\sigma}_i^2\ (4T_i^\tau + 4T_j^\tau + 4\tilde{\mu}_j(B_j^\tau - B_i^\tau) + C_i^\tau\sigma_i^2 + C_j^\tau\sigma_j^2) + \\
&\quad + \tilde{\sigma}_j^2(4T_i^\tau + 4T_j^\tau + 4\tilde{\mu}_i(B_i^\tau - B_j^\tau) + C_i^\tau\sigma_i^2 + C_j^\tau\sigma_j^2)) + \\
&-\omega\ (4(T_i^\tau + T_j^\tau)^2 + 4(B_i^\tau\tilde{\mu}_i + B_j^\tau\tilde{\mu}_j)(2T_i^\tau + 2T_j^\tau + B_i^\tau\tilde{\mu}_i + B_j^\tau\tilde{\mu}_j) + 4(\tilde{\mu}_i+\tilde{\mu}_j)(-1+\zeta(\tilde{\mu}_i+\tilde{\mu}_j))(C_i^\tau\sigma_i^2 + C_j^\tau\sigma_j^2) + \\
&\quad + \tilde{\sigma}_i^2(4(T_i^\tau + T_j^\tau + (B_j^\tau - B_i^\tau)\tilde{\mu}_j)(B_i^\tau(1-\zeta\tilde{\mu}_j) + \zeta(T_i^\tau + T_j^\tau + B_j^\tau\tilde{\mu}_j)) - (C_i^\tau\sigma_i^2 + C_j^\tau\sigma_j^2)) + \\
&\quad + \tilde{\sigma}_j^2(4(T_i^\tau + T_j^\tau + (B_i^\tau - B_j^\tau)\tilde{\mu}_i)(B_j^\tau(1-\zeta\tilde{\mu}_i) + \zeta(T_i^\tau + T_j^\tau + B_i^\tau\tilde{\mu}_i)) - (C_i^\tau\sigma_i^2 + C_j^\tau\sigma_j^2)) + \\
&\quad - \tilde{\sigma}_i^2\tilde{\sigma}_j^2(B_i^\tau - B_j^\tau)^2)
\end{aligned}
$$

where we have defined for notational simplicity

$$
T_a^\tau \equiv \phi_{0,a}\left(\sum_{t=0}^{\tau-1}(\phi_{1,a})^t\right),\quad B_a^\tau \equiv (\phi_{1,a})^\tau,\quad C_a^\tau \equiv (\sum_{t=0}^{\tau-1}(\phi_{1,a}^2)^t)\quad a=i,j.
$$

Finally, the ACF can be obtained by noticing that $\mathbb{E}[A_{ij}^t A_{ij}^{t-\tau}] \equiv \mathbb{P}(A_{ij}^t = 1, A_{ij}^{t-\tau} = 1)$ and the unconditional expectation for A_{ij}^t is

$$
\mathbb{E}[A_{ij}^t] = \int d\theta_{i,t}d\theta_{j,t}\mathbb{P}(A_{ij}^t = 1|\theta_{i,t},\theta_{j,t})n(\theta_{i,t})n(\theta_{j,t}) = \int \frac{d\omega}{2}p_{PG}(\omega)\frac{e^{\frac{-4\omega(\tilde{\mu}_i+\tilde{\mu}_j)^2+(\tilde{\sigma}_i^2+\tilde{\sigma}_j^2)+4(\tilde{\mu}_i+\tilde{\mu}_j)}{8(1+\omega(\tilde{\sigma}_i^2+\tilde{\sigma}_j^2))}}{\sqrt{1+\omega(\tilde{\sigma}_i^2+\tilde{\sigma}_j^2)}}.
$$

Appendix C. Maximum likelihood estimation of DAR(1) model

The standard DAR(1) model introduced by Jacobs and Lewis (1978) (A) describes a binary variable which has order one Markov dependence and Bernoulli marginal distribution. It is

$$
x_t = v_t\, x_{t-1} + (1-v_t)y_t \tag{C.1}
$$

where $x_t \in \{0,1\}\ \forall t$, $v_t \sim \mathcal{B}(\alpha)$ is a Bernoulli random variable with $\alpha \in [0,1]$ being the probability of copying from the past, *i.e.* x_{t-1}, and y_t is a binary random variable sampled according to the Bernoulli marginal distribution $\mathcal{B}(\chi)$ with $\chi \in [0,1]$.

Let us assume to observe a sequence of binary random variables $\boldsymbol{x} \equiv \{x_t\}_{t=0,1,\dots,T}$. We aim to find the maximum likelihood estimators of α and χ by assuming that x_t follows a DAR(1) process. The probability (likelihood) of the observations is

$$
\mathbb{P}(\boldsymbol{x}|\alpha,\chi) = \prod_{t=1}^{T}\mathbb{P}(x_t|x_{t-1},\alpha,\chi) = \prod_{t=1}^{T}\left(\alpha\mathbb{I}_{x_t,x_{t-1}} + (1-\alpha)\chi^{x_t}(1-\chi)^{1-x_t}\right) \tag{C.2}
$$

where $\mathbb{I}_{x_t,x_{t-1}}$ is the indicator function which takes value equal to 1 if $x_t = x_{t-1}$, zero otherwise. The maximum likelihood estimators of α and χ are the values which maximize the posterior

$\mathbb{P}(\alpha, \chi|\boldsymbol{x})$ or, equivalently, $\log \mathbb{P}(\alpha, \chi|\boldsymbol{x})$. Hence, possible candidates are the solution of the following system of equations

$$\partial_{\alpha,\chi} \log \mathbb{P}(\alpha, \chi|\boldsymbol{x}) = \partial_{\alpha,\chi} \sum_{t=1}^{T} \log \left(\alpha \mathbb{I}_{x_t, x_{t-1}} + (1-\alpha)\chi^{x_t}(1-\chi)^{1-x_t} \right) = 0, \tag{C.3}$$

or more explicitly,

$$\begin{cases} \sum_{t=1}^{T} \frac{\mathbb{I}_{x_t,x_{t-1}} - \chi^{x_t}(1-\chi)^{1-x_t}}{\alpha \mathbb{I}_{x_t,x_{t-1}} + (1-\alpha)\chi^{x_t}(1-\chi)^{1-x_t}} = 0 \\ \sum_{t=1}^{T} \frac{\mathbb{I}_{x_t,1} - \mathbb{I}_{x_t,0}}{\alpha \mathbb{I}_{x_t,x_{t-1}} + (1-\alpha)\chi^{x_t}(1-\chi)^{1-x_t}} = 0. \end{cases} \tag{C.4}$$

Given the solution(s) of Eq. C.4 in $(0,1) \times (0,1)$, we check what is the maximum among the possible candidates and the values on the bounds. Hence, we obtain the maximum likelihood estimators for α and χ.