Data-Driven Link Prediction Over Graphical Models

Daniele Alpago, Mattia Zorzi, Senior Member, IEEE, and Augusto Ferrante

Abstract—The positive link prediction problem is formulated in a system identification framework: We consider dynamic graphical models for autoregressive moving-average (ARMA) Gaussian random processes. For the identification of the parameters, we model our network on two different time scales: A quicker one, over which we assume that the process representing the dynamics of the agents can be considered to be stationary, and a slower one over which the model parameters may vary. The latter accounts for the possible appearance of new edges. The identification problem is cast into an optimization framework which can be seen as a generalization of the existing methods for the identification of ARMA graphical models. We prove the existence and uniqueness of the solution of such an optimization problem and we propose a procedure to compute numerically this solution. Simulations testing the performances of our method are provided.

Index Terms—Covariance extension, optimization, stochastic systems, system identification.

I. INTRODUCTION

The widespread use of online social networks (e.g., Facebook, Twitter, YouTube, etc.) and the availability of a huge amount of high-quality data related to them, have raised an increasing interest on network analysis research. The nature of big data coming from social networks is rather complex due to the dynamical behavior of the network, responsible of changes in topological and nontopological features of the network over time. In this scenario, the so-called link prediction problem [1], [2]. Given a certain network, link prediction problems can be divided in two categories: Predicting edges that appear or disappear in the future and detecting edges in the current social network that are still unobserved. Link prediction techniques find important applications in recommending systems, in finding business collaborators in e-commerce networks [3], academic social networks (describing co-authors relations) [4], security-related networks [5], and e-mails networks [6]. Particularly interesting for its intersection with the systems and control community, is the work on system biology [7]–[9] employing machine learning and statistical inference tools widely used for several control-related applications. The most part of the research effort has been focused on the positive link prediction (PLP) problem which consists either in detecting newly appeared edges or in predicting unobserved links that are likely to appear, depending on the perspective we are adopting.

Most of the literature on PLP concerns the design of similarity measures, i.e., metrics that quantify the probability that an edge appears between a pair of given nodes: An edge is added between two unconnected nodes if the similarity score (i.e., the value of the similarity measure) between the nodes is sufficiently large with respect to some problem-dependent criterion. Many similarity measures have been proposed in the literature: Common neighbors, Adamic/Adar, Katz, spreading activation, etc. Most of them are designed either by taking into account the topological behavior exhibited in the past by the real network, or on the basis of some properties the real network is expected to have, see [10] and [11]. For example, a property that the real network is often expected to exhibit is the so-called “small world” feature. To enforce this property, a high similarity score is assigned to a pair of unconnected nodes having several common neighbors. These methods essentially hinge on the assumption that the network will tend to exhibit a certain structure and predict the appearance of links that are most compatible with such a structure with little or no attention to the data associated with the network nodes. There are, however, some “unfriendly prediction networks,” namely, networks for which the similarity measures based on the network structure provide unsatisfying prediction accuracy, e.g., Facebook, UC Irvine, see [12]. In this case, it is necessary to rely on the data associated with the network nodes.

Our aim is to propose a new data-driven PLP paradigm leading to a similarity measure which exploits the current model of the network topology and some available noisy data. The PLP problem in the simple scenario in which the nodes are Gaussian random vectors was considered in [13]. In this article, instead, nodes are modeled as autoregressive, moving-average (ARMA) dynamical models and an edge between two nodes means that they are conditionally dependent given the other nodes of the network. Therefore, the only dynamic present in [13] is in the

1This means that most pairs of nodes are related through short chains and the network has an elevated clustering coefficient.
network topology. In this article, instead, we consider the case in which each node of the network represents a stationary Gaussian random process (or, equivalently, a dynamical linear system). Because of this substantial difference the problem requires a much deeper analysis and a completely new approach for its solution. Mathematically speaking, the scenario considered here is obviously much richer than that in [13]. On the other hand, this also corresponds to a much realistic situation in practical applications, where the network nodes model systems that clearly exhibit dynamical behavior such as stock prices.

In view of this approach, the properties on the underlying network are dictated from the data acquired by sampling the random processes associated with the nodes, rather than from properties the network is expected to feature. Thus, the development of such similarity measure is rephrased in the context of system identification with prior: We search the updated graphical model which explains the data and that is as close as possible to the current graphical model (i.e., the prior). The identification paradigm is then cast into an optimization problem for which we prove the existence and uniqueness of the solution. The similarity measure that our method induces exploits not only the topology of the current network (as the classic similarity measures do) but also the model of the current network making it more accurate for prediction.

It is worth noting that our contribution actually considers a detection problem, where the appearance of a new link is detected by considering both the past information (the current model of the network) and a noisy piece of information (data) coming from the nodes of the network where the new link has appeared. The scheme we propose is based on integrating the current network topology with the data and is therefore ready for a recursive implementation.

The present article approaches the link prediction problem from a system identification point of view, employing tools developed in the context of generalized moment problems, widely studied in the systems and control community. More precisely, our approach can be naturally cast in the context of identification of dynamic graphical models [17]–[22]. Particularly appropriate for big data applications are in fact sparse graphical models that are characterized by a reduced number of model’s parameters, thus reducing the risk of overfitting in the estimation scheme. The problem of identifying graphical models is a major topic in statistics and it has been tackled in various ways. Songsiri et al. [23], for instance, proposed to carry out the development of such similarity measure rephrased in the context of system identification with prior: We search the updated graphical model which is as close as possible to the current graphical model (i.e., the prior). The identification paradigm is then cast into an optimization problem for which we prove the existence and uniqueness of the solution. The similarity measure that our method induces exploits not only the topology of the current network (as the classic similarity measures do) but also the model of the current network making it more accurate for prediction.

The article is organized as follows: the notation and the background material on static and dynamic graphical models used throughout the article is recalled in Section II-A together with some references. Section III starts with the introduction of the PLP problem in terms of a regularized optimization problem. An interesting maximum likelihood interpretation is established in Section IV. Section V discusses a recursive version of the optimization problem just mentioned. In Section VI, the existence and the uniqueness of the solution of the optimization problem previously setup is proved. Section VII reports results of some numerical experiments meant to test the performances of the proposed method. In particular, we compare our similarity measure with the common neighbors similarity measure. Finally, in Section VIII we draw the conclusions and analyze some possible further extensions of the present work.

II. NOTATION, BACKGROUND, AND LITERATURE

A. Notation

Given a matrix $F$, $F^\top$ will denote the transpose of $F$, and $F^*$ its transpose-conjugate. If $F$ is square of dimension $p$, $\text{tr}(F)$, $\text{det}(F)$, $F^{-1}$ stand for the trace, the determinant and the inverse of $F$, respectively; moreover, $\text{diag}(F)$ denotes the $p$-dimensional vector whose entries are the diagonal elements of $F$, while $F \geq 0$ and $F > 0$ denote that $F$ is positive semidefinite or positive definite, respectively. $I_p$ denotes the $p \times p$ identity matrix. We denote by $\mathbb{H}_p$ the real vector space of Hermitian matrices of dimension $p \times p$ and by $L^\infty(\mathbb{T}, \mathbb{H}_p)$ the Banach space of essentially bounded functions defined on the unit circle $\mathbb{T} := \{e^{i\theta} : \theta \in [-\pi, \pi]\}$, and taking values in $\mathbb{H}_p$. For any function in $L^\infty(\mathbb{T}, \mathbb{H}_p)$, when it is clear from the context, we will drop the explicit dependence on $\theta$ and we use the shorthand notation $\int F$ for the integral $\int_\pi F(e^{i\theta}) \frac{d\theta}{2\pi}$. In this article we will deal in particular with functions in $S_p^+ := \{F \in L^\infty(\mathbb{T}, \mathbb{H}_p) : F - \alpha I_p \geq 0 \text{ a.e. on } \mathbb{T}, \exists \alpha > 0\}$ namely, we will consider coercive, bounded functions defined on $\mathbb{T}$. The operator $\mathcal{P}_\Omega : L^\infty(\mathbb{T}, \mathbb{H}_p) \to L^\infty(\mathbb{T}, \mathbb{H}_p)$ that maps a function $F$ to its projection $\mathcal{P}_\Omega(F)$ onto a certain support $\Omega \subseteq \{1, \ldots, p\} \times \{1, \ldots, p\}$, is defined as

$$[\mathcal{P}_\Omega(F)]_{ij} = \begin{cases} 0 & \text{if } (i,j) \in \Omega^c \smallskip \mathcal{P}_{ij} & \text{if } (i,j) \in \Omega \end{cases}$$

where $\Omega^c$ denotes the complement of $\Omega$ in $\{1, \ldots, p\} \times \{1, \ldots, p\}$. We define the following finite dimensional vector space of matrix pseudopolynomials

$$\mathcal{P}_{p,n} := \left\{ P \in L^\infty(\mathbb{T}, \mathbb{H}_p) : P(e^{i\theta}) = \sum_{k=-n}^n \mathcal{P}_k e^{ik\theta}, P_{-k} = P_k^* \in \mathbb{R}^{p \times p} \right\}$$

endowed with the norm

$$\|P\|_p := \int_{-\pi}^{\pi} |P(e^{i\theta})| \frac{d\theta}{2\pi} \quad (1)$$

---

\footnote{The reason why we use the term “prediction” is that this is indeed the term consistently used in the literature for these kind of detection problems, see e.g., [11], [14]–[16].}
where \( \nu_\theta(e^{i\theta}) \), \( \theta \in [-\pi, \pi] \), is the eigenvalue of \( P \) having maximum modulus. For a function \( f \) defined in some metric space \( X \) and taking values in \( \mathbb{R} \), we denote by \( \text{epi}(f) \) its epigraph, namely, the subset of \( X \times \mathbb{R} \) defined by
\[
\text{epi}(f) := \{(x, \beta) \in X \times \mathbb{R} : f(x) \leq \beta \}.
\]
The symbol \( \mathbb{E}[\cdot] \) denote the mathematical expectation.

B. Preliminaries on Graphical Models

A static undirected graphical model is a graph \( \mathcal{G} = (V, E) \) associated with a certain \( m \)-dimensional Gaussian random vector \( \mathbf{x} \sim \mathcal{N}(0, \Sigma) \), \( \Sigma = \Sigma^\top > 0 \), having vertexes \( V = \{1, \ldots, m\} \) representing the components \( x_1, \ldots, x_m \) of \( \mathbf{x} \) and edges \( E \subset V \times V \) describing the conditional dependence relations between the components through the following equivalent relations:
\[
(i, j) \notin E \iff x_i \perp_{\mathcal{G}} x_j | \{x_k \}_{k \neq i, j}
\]
\[
(\Sigma^{-1})_{ij} = 0 \quad (2)
\]
see for instance [17], for further details. Following [25], [18], and [19] it is possible to provide a dynamic counterpart of relation (2). In fact, let \( y := \{y(t), t \in \mathbb{Z}\} \) be an \( m \)-dimensional, purely nondeterministic, Gaussian stationary process and let \( \Phi(e^{i\theta}) \) its power spectral density defined for \( \theta \in [-\pi, \pi] \). For any index set \( I \subset V \), define as \( \mathcal{X}_I := \text{span}\{x_j(t) : j \in I, t \in \mathbb{Z}\} \) the closure of the set containing all the finite linear combinations of the variables \( x_j(t), j \in I \), so that for any \( i \neq j \), the notation
\[
\mathcal{X}_{\{i\}} \perp_{\mathcal{X}_{\{j\}} | \mathcal{X}_{\setminus \{i,j\}}} \mathcal{X}_{\setminus \{i,j\}}
\]
means that for all \( t_1, t_2, x_i(t_1) \) and \( x_j(t_2) \) are conditionally independent given the space linearly generated by \( \{x_k(t), k \in V \setminus \{i, j\}, t \in \mathbb{Z}\} \). One can prove that
\[
\mathcal{X}_{\{i\}} \perp_{\mathcal{X}_{\{j\}} | \mathcal{X}_{\setminus \{i,j\}}} \iff [\Phi(e^{i\theta})^{-1}]_{ij} = 0 \quad (3)
\]
for all \( \theta \in [-\pi, \pi] \). The latter generalizes the static relation (2). Accordingly, we can construct the undirected graph \( \mathcal{G} = (V, E) \) representing the conditional dependence relations between the components of the process \( y \) by defining the set of edges as follows:
\[
(i, j) \notin E \iff \mathcal{X}_{\{i\}} \perp_{\mathcal{X}_{\{j\}} | \mathcal{X}_{\setminus \{i,j\}}} \mathcal{X}_{\setminus \{i,j\}}. \quad (4)
\]
In this framework, the graph \( \mathcal{G} \) is completely characterized by the inverse power spectral density of the process; however, the information enclosed on the inverse spectrum \( \Phi^{-1} \) is richer than the one carried by the couple \( \mathcal{G} = (V, E) \).

Modeling each node as linear (Gaussian) process is the simplest way of considering dynamics of the nodes. This is a very general framework with respect to the state of the art, where only static nodes are considered. Of course, there is still an important gap between the theory and the possible applications but this is nevertheless a significative step in the direction of a reasonably realistic dynamical model.

Notice that we can drop the Gaussian assumption and consider only second-order processes. In this case, the conditional independence is replaced by a weaker condition, namely, conditional uncorrelation. Indeed, as shown in [18], for second-order processes, the right-hand side of (3) is equivalent to conditional uncorrelation of the variables in \( \mathcal{X}_{\{i\}} \) and \( \mathcal{X}_{\{j\}} \). To this extent all our results hold for second-order processes, with the proviso of considering the weaker condition of conditional uncorrelation in place of conditional independence.

C. Literature

The identification method proposed in this article can be cast in the framework of generalized moment problem introduced by Byrnes, Georgiou, and Lindquist to address spectral estimation problems, see [26]–[28], and extended to different divergence families in [29]–[32] and to the multidimensional setting in [33] and [34]. An application of this approach to physical layer authentication has been considered in [35]. We will show that our system identification paradigm with prior generalizes the ARMA graphical model identification framework of [19].

III. PARAMETRIC POSITIVE LINK PREDICTION

In this section the PLP problem is introduced in the simplest possible setting. This will serve as a starting point to rephrase the PLP problem for dynamic systems, starting from undirected graphical models for Gaussian time series recalled in the previous section. We start by an example aiming to introduce the PLP problem.

Example 1: The network in Fig. 1 represents five agents (e.g., users of a certain social network, where each user could be modeled—for example—by the number of times per day in which he uses a certain keyword and the observations are noisy because of possible spelling errors) \( a, b, c, d, \) and \( e \), as nodes, and their reciprocal friendship relations at a certain time \( \sigma > 0 \) as solid edges. Suppose that at time \( \tau > \sigma \) agents \( b \) and \( d \) became friends with \( c \), i.e., the dashed-red edges will appear in the network. The PLP problem consists in detecting these red edges based on the information on the network available at the previous time \( \sigma \) and some noisy data at time \( \tau \).

Hereafter we formalize our PLP problem. Let \( y := \{y(t), t \in \mathbb{Z}\} \) be an \( m \)-dimensional, Gaussian stochastic process and suppose that in a sufficiently small time interval \([\sigma - \alpha, \sigma + \alpha]\), for some \( \alpha \in \mathbb{Z}^+, \alpha > 0 \), the process can be considered to be stationary [36] so that it can be approximated by the parametric ARMA representation
\[
\sum_{k=0}^{n} A_j y(t-k) = \sum_{k=0}^{n} B_j e(t-k), \quad e(t) \sim \mathcal{N}(0, I_m) \quad (5)
\]
\footnote{We recall that an ARMA model is the most general linear finite dimensional system. As the order of the model increases, it also provides an arbitrary good approximation for nonrational (infinite dimensional) linear systems.}
where \( A_j, B_j \in \mathbb{R}^{m \times m} \) for \( j = 0, 1, \ldots, m \), and \( A_0 = I_m \). Denote with \( R_k := E[y(t)y(t-k)\dagger] \), \( k \in \mathbb{Z} \), its \( k \)th covariance lag, so that the power spectral density of the process is the Fourier transform of the sequence \((R_k)_{k \in \mathbb{Z}}\)

\[
\Phi_\sigma(e^{i\theta}) = \sum_{k=-\infty}^{\infty} R_k e^{-i\theta k}, \quad R_{-k} = R_k^\dagger, \quad \theta \in [-\pi, \pi]. \tag{6}
\]

Accordingly, in the interval \([\sigma - \alpha, \sigma + \alpha]\) a graphical model \( \mathcal{G}(\sigma) = (V, E_\sigma) \) associated with the process \( y \) is defined. In light of what we have said in the previous section, the edges of \( \mathcal{G}(\sigma) \) are completely characterized by the support \( \Omega_\sigma \subseteq \{(i,j) : i,j = 1, \ldots, m\} \) of the inverse spectrum of the process \( \Phi_\sigma^{-1} \).

\[
\Omega_\sigma = \{(i,j) : [\Phi_\sigma^{-1}(e^{i\theta})]_{ij} \neq 0, \text{ for all } \theta \in [-\pi, \pi]\}.
\]

Notice that, \( \mathcal{G}(\sigma) \) represents the current network so both \( \Phi_\sigma \) and \( \Omega_\sigma \) are assumed to be known. Such a model can be the result of previous studies on the network like, for instance the so-called huge prospective investigations carried out in medical research [37], [38], or it can be the outcome of a previous reliable estimation, more on this second case is developed in Section V. Let \( \mathcal{G}(\tau) = (V, \Omega_\tau) \), \( \tau > \sigma \), be the graphical model associated with the process \( y \) over another time interval \([\tau - \beta, \tau + \beta]\), \( \beta > 0 \), where the process can be considered stationary and can be described by an ARMA representation as (5), of course with different model parameters \( A_j, B_j \in \mathbb{R}^{m \times m} \) for \( j = 0, 1, \ldots, m \), and \( A_0 = I_m \). For our application it is very reasonable to assume that the model parameters do not change very much as the time interval change, see Remark 2 below.

As above, \( \mathcal{G}(\tau) \) corresponds to the support \( \Omega_\tau \) of \( \Phi_\tau^{-1} \): Notice that both \( \Phi_\tau^{-1} \) and its support \( \Omega_\tau \) are unknown. Our aim is to estimate \( \Phi_\tau \) corresponding to a certain support \( \Omega_\tau \). Here \( \mathcal{G}(\tau) \) represents the known current network, while \( \mathcal{G}(\tau) \) plays the role of the new network, whose edges we want to predict. By virtue of relations (3) and (4), the PLP scenario corresponds to the case in which \( \Omega_\sigma \subseteq \Omega_\tau \), i.e., the spectrum \( \Phi_\tau^{-1} \) should have fewer entries identically equal to zeros than the prior spectrum \( \Phi_\sigma^{-1} \), meaning that new edges must have appeared in the corresponding network \( \mathcal{G}(\tau) \).

**Remark 1:** As mentioned in the introduction, our model is based on the assumption that the network’s dynamics is slower than that of the nodes. Therefore, the time instant \( \tau \) is the time subsequent to \( \sigma \) on a time scale much longer than that with which the nodes are sampled. As a consequence, \( \sigma \) and \( \tau \) may be assumed considered to be known a priori.

To the best of our knowledge, this point of view has not yet been considered in the link-prediction community; we think, however, that it could be an interesting way to look at the problem from a different angle. The main motivations come from the following considerations.

1) It seems very reasonable that a similarity measure should take into account the dynamics of nodes and edges and the relation between the two (and thus not only the topology of the network) in predicting if a connection between them will appear;

2) we want to detect pair of nodes that are directly connected, while avoiding to detect a link between nodes that exhibit correlated behaviors due to common connection with other nodes. For this reason, our similarity score is not induced by dependence but rather by conditional dependence. As we have seen, this is equivalent to detect the vanishing entries in the inverse of the (matricial) spectral density. Notice that, by modifying a single entry of the spectral density the whole zero pattern of its inverse may change completely so that our similarity score accounts for the information distributed on the whole networks and has, therefore, a global nature. This contrasts with other data-driven approaches (see, e.g., [5]) where the similarity score is induced by a local property involving only the pair of nodes under scrutiny or, at most, a neighborhood of the target nodes.

Dynamic conditional independence networks cover both aspects; in fact, the presence of a link between two nodes is established on the basis of the whole network’s dynamics. The approach we propose is, therefore, quite different from the classic link prediction techniques: We consider a time-series model for the nodes dynamics, and our similarity measure is computed not only from the information on the current network topology but also from the current network model through the collected data.

Suppose now that a finite number \( N \) of observations \( y(1), \ldots, y(N) \) of the process over the interval \([\tau - \beta, \tau + \beta]\) are given as well. Natural interpretations of this nodes-related information are statistics coming from questionnaires or interviews, very diffuse in the social networks framework. The available observations allow us to compute an estimate of the first \( n + 1 \) covariance lags \( R_0, \ldots, R_n \) of \( y \) as

\[
\hat{R}_k = \frac{1}{N} \sum_{i=k}^{N} y(t)y(t-k)^\dagger, \quad t \in [\tau - \beta, \tau + \beta], \quad k = 0, 1, \ldots, n \tag{7}
\]

so that the truncated periodogram

\[
\hat{\Phi}_{n,\tau}(e^{i\theta}) = \sum_{k=-n}^{n} \hat{R}_k e^{-i\theta k}, \quad \hat{R}_{-k} = \hat{R}_k^\dagger \tag{8}
\]

represents an estimate of the spectrum \( \Phi_\tau \) based on the data. Notice that, \( \Phi_{n,\tau} \) is not positive definite in general and, even if it were, \( \Phi_{n,\tau}^{-1} \) would not be sparse, in other words, it cannot be the estimate we are looking for. Our PLP problem can then be formalized as follows.

**Problem 1:** Given the prior power spectral density \( \Phi_\sigma \) and the observations \( y(1), \ldots, y(N) \), estimate the spectrum \( \Phi_\tau \) and the support \( \Omega_\tau \) of its inverse, such that \( \Omega_\tau \supseteq \Omega_\sigma, \Phi_\tau \) is as close as possible to \( \Phi_\sigma \), and the moments constraints \( \int e^{i\theta k} \Phi_\tau = \hat{R}_k \), are satisfied for \( k = 0, 1, \ldots, n \).

**Remark 2:** Fig. 2 gives an intuitive explanation of the problem setup. We stress the fact that, except for technical details, the framework is precisely the one informally presented in Example 1. It is worth to note that the two intervals cannot be too far away from each other, otherwise the model could be completely different depending on the interval that is selected. Considering the application this assumption is not as restrictive as it seems.
Fig. 2. Pictorial clarification of the PLP set-up presented so far.

indeed, for instance, real-life considerations tell us that it is very rare that edges on a social network change drastically over a relatively short period of time. This could happen after some ground-breaking events that can be safely considered outliers (e.g., social media after the death of Michael Jackson, see for instance [39]).

Having in mind the setup just explained, we propose to reach the solution to Problem 1 through the following steps.

1) Consider a Positive Link Selection Problem: A simplified version of the PLP problem in which the support \( \Omega_r \) is known and only the model dynamics described by \( \Phi^{-1} \) needs to be estimated.

2) Solve Problem 1 by combining step 1) with the estimation of the support \( \Omega_r \), from observations.

In order to ease the presentation, let us call \( \Psi := \Phi_\sigma \) the prior spectral density such that \( P_{\Omega_\tau}(\Psi) = 0 \), \( \Phi := \Phi_\tau \) the spectrum we want to estimate, and \( \Phi_n := \Phi_{n,\tau} \) the corresponding truncated periodogram (8). Moreover, we assume that \( \Phi, \Psi \in S_+^m \) and that \( \Psi \) is rational. It is worth noticing that most of the results of this article could be extended to a nonrational version of the PLP problem in which the support \( \Omega_r \) is known exactly.

In light of the previous observations, we start by assuming the support \( \Omega_r \) to be known. Accordingly, Problem 1 reduces to the estimation of the spectral density \( \Phi \), whose inverse needs to have the specified support \( \Omega_r \), given the prior \( \Psi \) and the observations \( y(1), \ldots, y(N) \). This scenario is appropriate for setting up the following link-selection problem:

\[
\arg\min_{\Phi \in S_+^m} \mathcal{D}(\Phi\|\Psi)
\]

subject to \( P_{\Omega_r} \left( \int e^{i\theta_k} \Phi - \hat{R}_k \right) = 0, \]

\[
k = 0, 1, \ldots, n.
\]

where

\[
\mathcal{D}(\Phi\|\Psi) := \frac{1}{4\pi} \int_{-\pi}^{\pi} - \log \det \Phi + \log \det \Psi + \text{tr}(\Psi^{-1}\Phi) \, d\theta - m
\]

(10)

is the multivariate form of Itakura–Saito pseudodistance among power spectral densities \( \Phi \) and \( \Psi \), employed in signal processing [40]. Computational-complexity arguments and rationality constraints on the spectra motivate the choice of (10) as cost functional to be minimized in these kind of problems, see [41] and [42] for a deeper discussion on that.

Notice that, in our setting there is no loss of generality in assuming that the noise \( \Theta(\cdot) \) in (5) be normalized. Indeed, the only source of uncertainty is due to the fact that the \( \hat{R}_k \) are estimated from a finite number of samples and hence are not known exactly.

Remark 3: It is worth noticing that the support’s constraints on the moments in Problem (9) are the natural way to make the problem accounting the information provided by the observations \( y(1), \ldots, y(N) \), that has to be used in the identification procedure, together with the prior \( \Psi \). In fact, we do not impose that \( \Phi \) matches entirely the estimated moments \( \hat{R}_k, k = 0, 1, \ldots, n \). The condition in (9) has indeed the following interpretation: We believe that only the covariances between the nodes in \( \Omega_r \) are reliable.

Starting from [43] and [44], Dempster problems have been developed, progressively generalized, and adapted to different frameworks in a blooming stream of literature including [26], [27], [29]–[31], [33], [45]–[48]. However, generalized Dempster problems in graphical-models scenarios always involve maximum entropy problems in which some kind of entropic functional is maximized under linear constraints [19]–[21], [23], [49]. In this respect, Problem (9) in which a relative entropy functional is minimized, appears to be an original approach for graphical models applications.

As in Dempster’s setting [43], [50], the natural representation of our problem is given in terms of inverse power spectral densities. Duality theory proves to be the right tool to reparametrize the problem. The Lagrangian for Problem (9) is

\[
\mathcal{L}(\Phi, \Lambda) = \mathcal{D}(\Phi\|\Psi)
\]

\[
- \sum_{k=0}^{n} \left( \int_{-\pi}^{\pi} \int e^{i\theta_k} \Phi - \hat{R}_k \right) \cdot \Lambda_k \]

where \( \Lambda = [\Lambda_0, \Lambda_1, \ldots, \Lambda_n] \), \( \Lambda_k := P_{\Omega_r}(\hat{A}_k) \), \( k = 0, 1, \ldots, n \), are the new multipliers. Recalling that the projection operator is self-adjoint, we can rewrite the Lagrangian as

\[
\mathcal{L}(\Phi, \Lambda) = \mathcal{D}(\Phi\|\Psi)
\]

\[
- \sum_{k=0}^{n} \left( \int_{-\pi}^{\pi} \int e^{i\theta_k} \Phi - \hat{R}_k \right) \cdot \Lambda_k \]

we get

\[
\mathcal{L}(\Phi, \Lambda) = \mathcal{D}(\Phi\|\Psi) - \frac{1}{2\pi} \int_{-\pi}^{\pi} \sum_{k=0}^{n} \text{tr} \left( \Lambda_k^\top (\Phi_n - \Phi) e^{i\theta_k} \right) \, d\theta
\]

\[
= \frac{1}{4\pi} \int_{-\pi}^{\pi} - \log \det \Phi + \log \det \Psi + \text{tr} (\Psi^{-1}\Phi)
\]

\[
+ 2 \sum_{k=0}^{n} \text{tr} (\Lambda_k \Phi) e^{i\theta_k} - 2 \sum_{k=0}^{n} \text{tr} (\Lambda_k^\top \Phi_n) \, d\theta - m.
\]
By defining the pseudopolynomial
\[ Q(e^{i\theta}) := \sum_{k=-n}^{n} Q_k e^{-i\theta k}, \quad \text{with} \quad Q_k := \begin{cases} 2A_0 & \text{if} \quad k = 0 \\ \Lambda_k^T & \text{if} \quad k \neq 0 \end{cases} \]
we can reparametrize the Lagrangian as
\[ \mathcal{L}(\Phi, Q) = \frac{1}{4\pi} \int_{-\pi}^{\pi} \log \det \Phi + \log \det \Psi + \text{tr} \left[ \left( (\Psi^{-1} + Q) \Phi \right) - \text{tr} \left( Q \hat{\Phi}_n \right) \right] \, d\theta - m. \]

Notice that by construction, \( Q = Q(e^{i\theta}) \) has support \( \Omega_r \), i.e., \( P_{\Omega_r} (Q) = 0 \). Thanks to the monotonicity of the integral, it suffices to minimize the integrand (strictly convex in \( \Phi \))
\[ f(\Phi, Q) := -\log \det \Phi + \log \det \Psi + \text{tr} \left[ \left( (\Psi^{-1} + Q) \Phi \right) - \text{tr} \left( Q \hat{\Phi}_n \right) \right]. \]

If \( \delta f(\Phi, Q; \delta \Phi) \) denotes the Gâteaux derivative of \( f \) in a certain direction \( \delta \Phi \in L^\infty(\mathbb{T}, \mathbb{H}_m) \) we have that
\[ \delta f(\Phi, Q; \delta \Phi) = 0, \quad \forall \delta \Phi \in L^\infty(\mathbb{T}, \mathbb{H}_m) \]
if and only if
\[ \Phi = \Phi_o = (\Psi^{-1} + Q)^{-1} \quad (11) \]
provided that \( Q \in P_{m,n} \) is chosen such that \( \Psi^{-1} + Q > 0 \) on \([-\pi, \pi]\). Observe that \( \Phi_o^{-1} \) computed as in \(11 \), has support \( \Omega_s \cup \Omega_r = \Omega_r \) which agrees with the positive link prediction setup. Accordingly, the dual problem readily follows
\[
\arg\max_{Q \in \mathcal{Q}_{\Psi}} \frac{1}{4\pi} \int_{-\pi}^{\pi} \log \det \left( \Psi^{-1} + Q \right) - \text{tr} \left( Q \hat{\Phi}_n \right) \, d\theta \\
\text{subject to} \quad P_{\Omega_r} (Q) = 0 \quad (12)
\]
where
\[ \mathcal{Q}_{\Psi} := \{ Q \in P_{m,n} : \Psi^{-1} + Q(e^{i\theta}) > 0, \forall \theta \in [-\pi, \pi] \} \]
is the domain of optimization, open and unbounded. In view of \(11 \) and \(12 \), the solution to Problem \(9 \) is such that its inverse has support \( \Omega_r \), i.e., we have more (conditional) dependencies between the variables in the new network. We refer the reader to Section VI, Theorem 2, for the discussion of existence and uniqueness of the solution of Problem \(9 \).

**Remark 4:** It is interesting to note that Problem \(9 \) solves the problem of identifying an ARMA graphical model with topology \( \Omega_r \) provided that the prior spectral density \( \Psi \) corresponds to a graphical model having topology \( \Omega_s \subseteq \Omega_r \). It is worth noting that a similar problem has been addressed in \([19]\) with the following formulation:
\[
\arg\max_{\Phi \in \mathcal{S}_m^+} -\text{tr} \int \psi \log(\psi^{-1}) \\
\text{subject to} \quad P_{\Omega_r} \left( \int e^{i\theta k} \Phi - \hat{R}_k \right) = 0, \quad k = 0, 1, \ldots, n \quad (13)
\]
where \( \psi \) is a scalar spectral density *a priori* known. Notice that, \(13 \) is equivalent to minimize
\[ \mathcal{D}_{KL}(\psi I_m \| \Phi) := \text{tr} \left\{ \int \psi \log(\psi^{-1}) - \psi I_m + \Phi \right\} \]
where the last two terms are constant, as \( \psi \) is known and the moments constraint in \(13 \) ensures that \( \text{tr} \int \Phi = \text{tr}(\hat{R}_0) \). Using arguments similar to the ones in \([51]\), it is not difficult to see that \( \mathcal{D}_{KL}(\psi I_m \| \Phi) \) is actually a pseudodistance between \( \psi I_m \) and \( \Phi \) (\( \mathcal{D}_{KL}(\psi I_m \| \Phi) \geq 0 \) with equality if and only if \( \psi I_m = \Phi \)), representing the natural extension of the Kullback–Leibler divergence between multivariate power spectral densities in which the first argument has the particular structure \( \psi I_m \). Therefore, our problem is in the same spirit of Problem \(13 \). In particular, if we take \( \Phi = I_m \) in \(9 \) and \( \psi = 1 \) in \(13 \), then the two problems do coincide, i.e., they maximize the entropy rate of the process. In our setting, however, the scalar prior \( \psi I_m \) would correspond to a graphical model with disconnected nodes. In principle, one could extend Problem \(13 \) to a prior corresponding to a graphical model having topology \( \Omega_s \subseteq \Omega_r \). However, in this case, the variational analysis cannot be carried out.

The second step of our solution consists in combining Problem \(12 \) with the estimation of the support \( \Omega_r \) in order to obtain an optimization problem for the solution of Problem \(1 \). Following \([23]\) and \([21]\), we propose to perform this step by resorting to a regularized version of Problem \(12 \), namely,
\[ \arg\min_{Q \in \mathcal{Q}_{\Psi}} \mathcal{J}(Q) \quad (14) \]
where
\[ \mathcal{J}(Q) := \int \left[ -\log \det \left( \Psi^{-1} + Q \right) + \text{tr} \left( Q \hat{\Phi}_n \right) \right] + \lambda h^\infty(Q) \]
\[ = \int \text{tr} \left[ Q \hat{\Phi}_n - \log \left( \Psi^{-1} + Q \right) \right] + \lambda h^\infty(Q) \]
and
\[ h^\infty(Q) = \sum_{(h, k) \in \Theta} \max_{j=1, \ldots, n} \max_{j=1, \ldots, n} \left( |Q_{0,h,k}|, |Q_{j,k}| \right) \]
with \( \Theta := \{(h, k) \in V \times V \setminus \Omega_s : k > h \} \), plays the role of the \( \ell^1 \)-norm used to induce sparsity on vectors and it has been proposed in \([23]\) for inducing group sparsity to \(Q_0, \ldots, Q_n\). It is worth noting that Problem \(14 \) is a heuristically regularized version of Problem \(12 \). The sparsity-inducing regularization, tuned by the parameter \( \lambda > 0 \), acts only on the elements of the \(Q_k\)s in positions that are not contained in the support \( \Omega_s \), according to the fact that \( \Omega_s \subseteq \Omega_r \). This allows us to reduce the bias introduced by the regularization in the estimation procedure. Indeed, regularization here is used to decide whether an edge is present or not. Since we already know the presence of the edges

---

4Notice that, in order to deal with convex functions, instead of maximizing the objective function [as in Problem \(12 \)] we are now minimizing the opposite function (multiplied by a factor 2).
in $\mathcal{Z}_r$, we do not need regularization for them. As $\lambda$ increases
the corresponding solution tends to be more and more sparse.

The most suitable value of $\lambda$ may be chosen by resorting to a
cross-validation procedure.

The proof of the existence and the uniqueness of the solution
to Problem (14) is devoted to Section VI.

We are aware that at first sight, our setup may seem far apart
with respect to the typical settings adopted in the link prediction
community. However, if we take a closer look at the problem we
readily find out that we are actually just proposing a different
choice of score matrix to decide how much a pair of nodes are
inclined to get connected. Our approach suggests that a suitable
similarity measure should be an indicator of the conditional
dependence between the variables. In order to introduce such
a measure, we define the partial coherence of the predicted
spectrum $\Phi^{-1}_r$ as

$$
\Gamma_r(e^{i\theta}) := \text{diag}[\Phi_r(e^{i\theta})]^{-1/2} \cdot \Phi_r^{-1}(e^{i\theta}) \cdot \text{diag}[\Phi_r(e^{i\theta})]^{-1/2}
$$

(15)

for all $\theta \in [-\pi, \pi]$. More precisely, the element of $\Gamma_r(e^{i\theta})$ in
position $(i,j)$. This is a standard tool the frequency-domain
analysis of time series and it measures the dependence between
two time series after removing the linear time invariant effects
of the other series [18], [23], [25]. The similarity measure that
is naturally induced by our approach in order to rate an edge is,
therefore, the $L_2$ norm of the elements of (15)

$$(G_r)_{ij} := \left\| \Phi_{e_{i,j},e_{i,j}} \right\|_2, \quad (i,j) \in V \times V \setminus \Omega_\sigma$$

and the matrix $G_r := [(G_r)_{ij}]$ represents our score matrix. As
the score matrix $G_r$ will have some small entries but, in general,
will not be exactly sparse, a thresholding procedure is needed
in order to obtain an estimate of the support $\Omega_r$ that defines
a network topology according to relations (3) and (4). More
precisely, we will consider the edge $(i,j) \in V \times V \setminus \Omega_\sigma$ to be
in $\Omega_r$ only if its score is greater than a certain threshold $t_r > 0$
(to be suitably selected), namely,

$$(i,j) \in \Omega_r \iff (G_r)_{ij} > t_r.$$

To conclude, it is interesting to note that given a pair $(i,j) \in V \times
V \setminus \Omega_\sigma$, one can get a straightforward interpretation of $(G_r)_{ij}$
in terms of best (linear) predictors of $y_i$ and $y_j$, namely,

$$\hat{y}_i(t) := \mathbb{E}[y_i | y_{V \setminus \{i,j\}}] \quad \text{and} \quad \hat{y}_j(t) := \mathbb{E}[y_j | y_{V \setminus \{i,j\}},

respectively, where $y_{V \setminus \{i,j\}} := \text{span}\{y_k(t) : k \in V \setminus \{i,j\}\}$. Then $(G_r)_{ij}$ quantifies how much the estimation errors $\epsilon_{i,j}(t) :=

y_i(t) - \hat{y}_i(t)$ and $\epsilon_{j,i}(t) := y_j(t) - \hat{y}_j(t)$ are independent by
considering the $L_2$ norm of the conditional coherence between
$i$ and $j$

$$(G_r)_{ij} = \left\| \Phi_{e_{i,j},e_{i,j}} \right\|_2$$

where $\Phi_{e_{i,j}, e_{i,j}}$ are the spectra of $\epsilon_{i,j}, \epsilon_{j,i}$, and $\Phi_{e_{i,j}, e_{i,j}}$ is the corresponding cross spectrum.

Remark 5: Our estimator of the inverse spectrum is consistent
under quite standard hypotheses. More precisely, it is sufficient
that: 1) the true spectrum, say $\Phi$, belongs to the model class $\mathcal{Q}_\psi$
over which we search our solution; 2) process $y$ is ergodic; 3) the
regularization parameter is such that $\lambda_N \to 0$ as $N \to \infty$
(and we used the subscript $N$ to highlight that $\lambda$ is a function
of $N$). The idea of the proof runs as follows. The dual problem
(14) can be written as

$$
\hat{Q}_N := \underset{Q}{\text{argmax}} \mathcal{J}_\Phi(Q)
= \underset{Q}{\text{argmax}} \mathcal{D}(\{(\Psi^{-1} + Q)^{-1}\|\tilde{\Phi}_N) + \lambda_N h_N^\infty(Q)$$

where $\Phi_N := \hat{\Phi}_n + \sum_{k > N} R_k e^{-itk}$ and $N$ is the number of
observations. Process $y$, being ergodic, implies $\Phi_N \to \tilde{\Phi}$ almost
surely (a.s.) as $N \to \infty$. Using this fact and the dominated
convergence theorem one can further prove that $\mathcal{D}(\{(\Psi^{-1} +
Q)^{-1}\|\tilde{\Phi}_N) \to \mathcal{D}(\{(\Psi^{-1} + Q)^{-1}\|\Phi_{\infty})$ as $N \to \infty$.

Since our model class $\mathcal{Q}_\psi$ contains $\tilde{\Phi}_i$, we have that $\Phi =
(\Psi^{-1} + Q)^{-1}$ where $Q := \underset{Q}{\text{argmax}} \mathcal{D}(\{(\Psi^{-1} + Q)^{-1}\|\Phi_{\infty})$. By
the same arguments detailed in [52], we see that the mapping
$\Phi_N \to \hat{Q}_N$ is continuous. Then, the continuous
mappng theorem ensures that $\hat{Q}_N \to Q$ as $N \to \infty$, a.s. It fol-


dows that the estimator $\Psi^{-1} + Q$ converges to $\tilde{\Phi}_i$ a.s.

Finally, if we take the threshold $t_r$, such that $t_r < \bar{c}$, with
$\bar{c} := \text{argmin}_Q \|\Phi^{-1} - Q\|_2$, the estimator of the support is consistent. Clearly, $\bar{c}$ is not known. On the other
hand, in real applications all the nodes of a network are condi-
tional dependent. In this situation $t_r$ is the upper bound of $L_2$
norm of the conditional coherence between two nodes for which
we consider their conditional dependence negligible.

IV. Maximum Likelihood Interpretation

In this section we show that Problem (14) has a nice inter-
pretation as (regularized) maximum likelihood problem. This
interpretation is based on a frequency approximation of the
likelihood function of a Gaussian sample, first introduced by
Whittle [53], [54] for scalar stationary processes. Over the
years, the so-called Whittle likelihood approximations have been
generalized to multivariate stationary processes [55] and also
extended to the nonstationary case [36].

Consider process (5) having power spectral density $\Phi$ as in (6),
and let $\hat{\Phi}_n$ as in (8) be its truncated periodogram computed on the
basis of $N$ given observations of the process $y(1), \ldots, y(N)$. The
assumptions on model (5) that guarantee the following to hold are
to some extent classical, however, they require a
technical presentation that the interested reader can find in [55, Sec. 2]. Let $y_N = [y(1)\top \cdots y(N)\top]$, then the
$N$-dimensional probability density $p(y_N; \Phi) := p(y(1), \ldots,
y(N); \Phi)$ of the random variables $y(1), \ldots, y(N)$ has the well-
known form

$$p(y_N; \Phi) = \frac{1}{\sqrt{(2\pi)^N n^N}} \exp \left\{ \frac{1}{2} y_N^\top T_N(\Phi)^{-1} y_N \right\}$$

where $T_N(\Phi) = \mathbb{E}[y_N y_N^\top]$ is the $MN \times MN$ Toeplitz matrix
whose $(h, k)$th block is defined as

$$T(\Phi)_{hk} = \int_{-\pi}^{\pi} \Phi(e^{i\theta}) e^{i(h-k)\theta} \frac{d\theta}{2\pi} = R_{h-k}, 1 \leq h, k \leq N + 1.$
The corresponding negative log-likelihood (up to scaling factors and constant terms) is
\[
\hat{\ell}_N(\Phi) = \frac{1}{N} \log \det T_N(\Phi) + \frac{1}{N} y_N^\top T_N(\Phi)^{-1} y_N. \tag{16}
\]
Various frequency approximations of (16) may be introduced \cite{55, 56}. In order for us to define one of those, we introduce the discrete Fourier transform of the data
\[
Y_N(e^{i\theta}) = \frac{1}{\sqrt{N}} \sum_{p=1}^{N} y(p) e^{-i\theta p}
\]
such that
\[
\hat{\Phi}_N(e^{i\theta}) = Y_N(e^{i\theta}) Y_N(e^{i\theta})^* = \sum_{k=-(N-1)}^{N-1} \hat{R}_k e^{-i\theta k}
\]
where \(\hat{R}_k\), such that \(\hat{R}_{-k} = \hat{R}_k^*, k = 0, 1, \ldots, N-1\), computed as in (7), is the periodogram of the process \(y\). With these definitions it can be shown \cite{55} that \((\hat{\ell}_N - \ell_\infty) \to 0\) almost surely as \(N \to \infty\), where
\[
\ell_\infty(\Phi) = \int_{-\pi}^{\pi} - \log \det \Phi(e^{i\theta})^{-1} \text{tr} \left[ \Phi(e^{i\theta})^{-1} \hat{\Phi}_N(e^{i\theta}) \right] \frac{d\theta}{2\pi}.
\tag{17}
\]
In this sense Problem (14) can be interpreted, at least asymptotically, as a (regularized) maximum likelihood problem, in which we have to minimize \(\ell_\infty(\Phi)\) with \(\Phi\) belonging to the parametric family
\[
\Psi_{\Psi, Q} := \left\{ \Phi = (\Psi^{-1} + Q)^{-1} : Q \in Q_0^+ \right\}
\]
for a given \(\Psi \in S_m^+\). Notice that
\[
Q(e^{i\theta}) \hat{\Phi}_N(e^{i\theta}) = \sum_{k, h=-(N-1)}^{N-1} Q_k \hat{R}_h e^{-i\theta(k+h)}
\]
where \(Q_k = 0\) for \(k = n+1, \ldots, N-1\), because \(Q \in \mathcal{P}_{m,n}\).
Hence
\[
Q(e^{i\theta}) \hat{\Phi}_N(e^{i\theta}) = \sum_{k, h=-n}^{n} Q_k \hat{R}_h e^{-i\theta(k+h)} = Q(e^{i\theta}) \hat{\Phi}_n(e^{i\theta})
\]
where \(\hat{\Phi}_n\) is the truncated periodogram (8) of the process. Accordingly, minimizing (17) over \(\Psi_{\Psi, Q}\) is equivalent to minimize
\[
\int_{-\pi}^{\pi} - \log \det \left( \Psi^{-1} + Q \right) + \text{tr} \left( Q \hat{\Phi}_n \right) \frac{d\theta}{2\pi} \tag{18}
\]
over \(Q_0^+\). Save for the regularization term, (18) is precisely the index \(J_\Psi\) of (14).

V. RECURSIVE POSITIVE LINK PREDICTION

Our approach can be easily embedded in a recursive approach that perfects the previous estimation given new available data. Suppose that we want to study a certain network over the time interval \([0, T]\). We refer to the time window \([k - \alpha, k + \alpha]\), \(\alpha > 0, k > 0\), as the time window \(\tau = k\), see the beginning of this section. Moreover, we assume that at time \(\tau = 0\) a prior information concerning the whole network is summarized in the spectrum \(\Phi_0\) whose inverse has support \(\Omega_0\) and observations \(y^{(1)}(1), \ldots, y^{(1)}(N_1)\) related to the present time window \(\tau = 1\) are available. Then we can find \(Q_1\) by solving Problem (14) so that the predicted network at time \(\tau = 1\) is given by \(\Phi_1^{-1} = \Phi_0^{-1} + Q_1\). The same reasoning applied at time \(k + 1\) leads to
\[
Q_{k+1} = \arg\min_{Q \in Q_0^+} J_{\Phi_k}(Q) \tag{19}
\]
from which
\[
\Phi_{k+1}^{-1} = \Phi_k^{-1} + Q_{k+1} = \Phi_0^{-1} + \sum_{l=1}^{k+1} Q_l(e^{i\theta})
\]
where \(Q_l \in \mathcal{P}_{m,n}, l = 1, \ldots, k + 1\), \(\Phi_k\) is the output of the estimation procedure at step \(k\). Therefore, Problem (19) is equivalent to
\[
Q_{k+1} = \arg\min_{Q \in Q_0^+} \tilde{J}_{\Phi_0,k}(Q)
\]
where
\[
\tilde{J}_{\Phi_0,k}(Q) := \frac{1}{4\pi} \int_{-\pi}^{\pi} \log \det (\Phi_0^{-1} + Q) - \text{tr} \left( Q \hat{\Phi}_{n,k} \right) d\theta.
\]
What we just saw implies that the upper bound on Mcmillan degree of \(\Phi_k\) does not depend on \(k\); more precisely
\[
\deg(\Phi_k) \leq \deg(\Phi_0) + n \forall k > 0.
\]
It is worth noticing that, thanks to the latter upper bound, the complexity of the model is guaranteed not to explode even if the number of iterations gets considerably high (potentially infinite). The corresponding iterative scheme is represented in Fig. 3.

1) Initialization. Initial prior \(\Phi_0\) and observations \(y^{(1)}(1), \ldots, y^{(1)}(N_1)\).
2) Iteration. Given \(y^{(k+1)}(1), \ldots, y^{(k+1)}(N_{k+1})\) for \(k = 1, 2, \ldots, T\) compute \(Q_{k+1}\) by solving (19).

VI. EXISTENCE AND UNIQUENESS OF THE SOLUTION

This section is devoted to the proof of existence and uniqueness of the solution for the regularized Problem (14).

Theorem 1: Problem (14) admits a unique solution \(Q_0 \in Q_0^+\).

Proof: We first extend the definition of \(J_{\Phi}(Q)\) to the boundary of \(Q_0^+\) and show that this extended function do admit a unique
minimum there. We then show that this minimum cannot be on the boundary. Let
\[
\partial_\Psi^+ := \left\{ Q \in \mathcal{P}_{m,n} : \Psi^{-1}(e^{i\theta}) + Q(e^{i\theta}) \geq 0 \text{ and singular } \exists \theta \in [-\pi, \pi] \right\}
\]
be the boundary of \( \Omega^+_\Psi \) and introduce the sequence of functions \((\partial_\Psi^+)_n\) over
\[
\tilde{\Omega}^+_\Psi := \Omega^+_\Psi \cup \partial_\Psi^+
\]
\[
= \left\{ Q \in \mathcal{P}_{m,n} : \Psi^{-1}(e^{i\theta}) + Q(e^{i\theta}) \geq 0, \exists \theta \in [-\pi, \pi] \right\}
\]
defined as
\[
\beta_n^+(Q) := \int_{-\pi}^{\pi} \text{tr} \left[ Q \hat{\Phi}_n - \log \left( \Psi^{-1} + Q + \frac{1}{n} I_m \right) \right] \frac{d\theta}{2\pi} + \lambda \hat{h}_n^+(Q)
\]
for \( n \geq 1 \). Notice that \( \beta_n^+ \) is a strictly convex and continuous function on \( \tilde{\Omega}^+_\Psi \), therefore, \( \text{epi}(\beta_n^+) \) is convex and closed on \( \tilde{\Omega}^+_\Psi \times \mathbb{R} \) for any \( n \geq 1 \). Moreover, the sequence \((\beta_n^+)_n\) is monotonically increasing so the pointwise limit
\[
\beta^+_{\Psi}(Q) := \lim_{n \to \infty} \beta_n^+(Q)
\]
exists and it is a strictly convex and continuous function over \( \tilde{\Omega}^+_\Psi \), because \( \text{epi}(\beta^+_{\Psi}) = \cap_{n \geq 1} \text{epi}(\beta_n^+) \). The proof now follows in view of the following three lemmata proven in the Appendix. The first Lemma ensures that \( \beta^+_{\Psi} \) coincides with the original \( \beta_{\Psi} \) on its domain of definition \( \Omega^+_\Psi \).

**Lemma 1:** The following relation holds on \( \Omega^+_\Psi \):
\[
\beta^+_{\Psi} = \beta_{\Psi}.
\]

The next lemma establishes an existence and uniqueness result for the extended function \( \beta^+_{\Psi} \).

**Lemma 2:** The function \( \beta^+_{\Psi} \) admits a unique minimum on \( \tilde{\Omega}^+_\Psi \). The last step consists in showing that the minimum of \( \beta^+_{\Psi} \) cannot be on the boundary.

**Lemma 3:** The minimum of \( \beta^+_{\Psi} \) is attained in \( \Omega^+_\Psi \).

Lemma 3 concludes the proof of Theorem 1.

Before concluding the section, we prove an analogous existence result for problem (9) that, as discussed in Remark 4, solves the problem of identifying an ARMA graphical model with topology \( \Omega_t \), thus extending to a general prior the work [19].

**Theorem 2:** Problem (9) admits a unique solution.

**Proof:** Problem (14) is a regularized (and thus, more complex) version of Problem (12). Thus, by following the same steps of the proof of Theorem 1, we can extend also the objective function of Problem (12) to the closure of the original domain \( \Omega^+_\Psi \). Hence, we can apply Weierstrass theorem and conclude that Problem (12) admits a minimum in the closed domain. Then, by resorting to the same argument of Lemma 3 we can show that the minimum is actually attained on \( \Omega^+_\Psi \), i.e., it cannot be achieved on the boundary. Strong duality occurring between Problem (9) and Problem (12) yields the conclusion.

VII. SIMULATION RESULTS

In this section, we present some numerical examples illustrating the performances of the proposed method for positive link prediction problems in the case of AR and ARMA dynamics. Moreover, we compare our approach with the common neighbors method.

VII. AR Dynamics

Suppose that the dynamic of the agents composing the network is described by model (5) with \( B_k = 0 \) and \( F_k = -A_0^{-1} A_k \) for all \( k > 0 \)
\[
y(t) = \sum_{k=1}^{n} F_k y(t-k) + \epsilon(t)
\]
i.e., by the AR process \( y(t) = F(z)y(t) + \epsilon(t) \) in which \( F(z) = \sum_{k=0}^{n} F_k z^{-k} \), the dimension of the process is \( m = 10 \) and the order of the process is \( n = 2 \). The setup for this test is the one of recursive PLP explained in Section V, for a window of length \( T = 2 \). The initial information of the network is enclosed in a prior spectral density \( \Phi_0 \) whose inverse has support \( \Omega_0 \), and the supports \( \Omega_1 \) and \( \Omega_2 \) of the spectra \( \Phi_1^{-1} \) and \( \Phi_2^{-1} \) need to be estimated according to the iterative scheme
\[
\Phi_0 \sim \Phi_1 \sim \Phi_2
\]
outlined in Section V. Fig. 4 depicts the poles of the shaping filter \( I - F(z) \) of model (21) at times \( \tau = 0, 1, 2 \) from left to right, respectively, while Fig. 5 reports the support of \( \Phi_0^{-1} \) (left) and the supports of the spectra \( \Phi_1^{-1} \) (center) and \( \Phi_2^{-1} \) (right) that have to be estimated. It is worth noting that, we are dealing with an unfriendly prediction network. Indeed, let \( N_i \) be the set of neighbors of agent \( i \) according to the graph \( \mathcal{G}(0) \) and consider the common neighbors similarity measure at
time $\tau = 1$, $(CN_0)_{ij}$ defined as the cardinality of set $N_i \cap N_j$. It is not difficult to see that $(CN_0)_{ij} = 0$ for any edge $(i, j)$ appearing in $\Omega_1$ and $(CN_0)_{ij} = 1$ for any $(i, j) \in \Omega_1^{CN}$. $\Omega_1^{CN}$ contains edges not appearing in $\Omega_1$. A similar scenario happens for the prediction time $\tau = 2$ revealing that the common neighbors measure leads to low prediction accuracy for this kind of network.

In regard to our method, the computation of the estimates $\hat{\Omega}_1$ and $\hat{\Omega}_2$ of the respective supports $\Omega_1$ and $\Omega_2$, exploits the information coming from $N_1 = N_2 = 1000$ data samples together with the previously estimated spectrum and it is performed by solving Problem (19) for each estimation step using the CVX package [57], [58]. As mentioned in Section III, the estimates of the inverse spectra produced by our method are close to be sparse. Fig. 6 displays the score matrices $G_1$, $G_2$ for estimation of the supports $\Omega_1$ and $\Omega_2$, respectively.

![Fig. 6. Score matrices $G_1$ and $G_2$ for estimation of the supports $\Omega_1$ and $\Omega_2$, respectively.](image)

---

**VII. ARMA Dynamics**

This test illustrates the performances of the proposed method when the model used to fit the data is an ARMA model of the type of (5) in which the dimension of the process is $m = 4$, while the order of the polynomial part $Q$ is set to $n = 4$. This experiment generalizes the approach proposed in [23] to the case of a vector MA part. To relate our results with [23], we actually consider the solution of Problem (14) in two different cases: The case in which an ARMA prior spectral density $\Psi^{-1}$ and its support $\Omega_\sigma$ are available, and the case in which $\Psi \equiv I_m$ (and $\Omega_\sigma \equiv I_m$), namely, the regularized maximum likelihood estimator of [23], in which we require our estimation procedure to fit the ARMA model (5) with an AR model of order $n = 6$ (so that to have a comparable number of parameters). Fig. 8 reports the inverse of

![Fig. 8. Inverse of the ARMA prior (above), prior’s support (below left) and true support of $\Phi^{-1}$ (below right).](image)
the ARMA prior power spectral density chosen for the first test (above) with its support $\Omega_\tau$ (below left), and the support $\Omega_\tau$ of $\Phi^{-1}$ that has to be estimated (below right). The results of the simulations corresponding to $\lambda = 0.04$ are illustrated in Fig. 9, where the true $\Phi^{-1}$ (blue) is compared to its estimates obtained from the two different priors. In particular, the red line represents the estimate $\hat{\Phi}_{\text{ARMA}}^{-1}$ of $\Phi^{-1}$ computed from the ARMA prior $\Psi$ depicted in Fig. 8, while the estimate $\hat{\Phi}_{\text{AR}}^{-1}$ computed without prior information is represented with the green line. Also for these set of simulations, the supports of the estimated spectra have been computed on the basis of the partial coherence-based similarity measure followed by the thresholding procedure (with $t_\tau = 0.1$) and we can see from Fig. 9 how the estimation procedure is able to recover the true support $\Omega_\tau$ with both priors. However, it can be noticed that the presence of the prior leads to an enhancement of the estimation capabilities of the proposed procedure. This is particularly highlighted by the comparison between the entries in position (4,1) in Fig. 9.

The ARMA model has been tested also on an AR model of the same order $\Phi \in \mathbb{R}_+^{p \times p}$ and its estimate $\hat{\Phi}$ of the ARMA model used for the estimation. As expected, the estimation procedure is able to recover the true support $\Omega_\tau$ from the two different priors. In particular, the red line represents the entry $\Phi^{-1}$ (above) with its support $\Omega_\tau$, while the estimate $\hat{\Phi}_{\text{ARMA}}^{-1}$ computed with $\lambda = 0.04$ and its estimate $\hat{\Phi}_{\text{AR}}^{-1}$ computed with no prior information.


defined for $\theta \in [-\pi, \pi]$. Clearly $f_n$ is a continuous and, therefore, measurable function of $\theta$, for any $n \geq 1$, and moreover $f_n \uparrow f$ pointwise (recall that $\Psi^{-1}$ and $Q$ are continuous functions of $\theta$) ensuring that the limit $f$ is itself measurable. In addition, $|f_n| \leq g$ pointwise (and, therefore, a.e.) for any $n \geq 1$, where $g = |f| \in L^1(T, \mathbb{R})$ since $Q \in \Omega_\Psi^+$. By Lebesgue’s dominated-convergence theorem, $f \in L^1(T, \mathbb{R})$ and

$$
\lim_{n \to \infty} \int - \log (\Psi^{-1} + Q + \frac{1}{n} I_m) = \lim_{n \to \infty} \int f_n = \int \lim_{n \to \infty} f_n = \int - \log (\Psi^{-1} + Q)
$$

where the last equality follows from continuity of $\log(\cdot)$ and $\log(\cdot)$. The conclusion is now straightforward

$$
\beta_\Psi(Q) = \lim_{n \to \infty} \beta_\Psi(Q) = \int \log (\Psi^{-1} + Q + \frac{1}{n} I_m) + \lambda h_\sigma^*(Q) = \beta_\Psi(Q)
$$

for any $Q \in \Omega_\Psi^+$.

Proof of Lemma 2: Since we have already seen that $\beta_\Psi$ is strictly convex on $\Omega_\Psi^+$, it is sufficient to show that it is also proper on $\Omega_\Psi^+$. First, notice that $\beta_\Psi$ is not identically $+\infty$ on $\Omega_\Psi^+$, and it is easy to see that for any $n \geq 1$

$$
\beta_\Psi(Q) \geq \text{cost.} + \int_{-\pi}^\pi \mu_1 \sum_{i=1}^m \lambda_i - \frac{1}{\mu_1} \log \left( \lambda_i + \frac{1}{n} \right) \frac{d\theta}{2\pi} \geq \text{cost.} + \frac{1}{\mu_1} \frac{1}{n} - \infty
$$

where $\mu_1(e^{i\theta})$ is the minimum eigenvalue of $\Phi$, while $\lambda_1(e^{i\theta}) \leq \cdots \leq \lambda_m(e^{i\theta})$ are the eigenvalues of $\Psi^{-1} + Q$. Taking the limit for $n \to \infty$ on both sides, we conclude that $\beta_\Psi > -\infty$ on $\Omega_\Psi^+$.

VIII. CONCLUSION

In this article, positive link prediction as detection problem is approached as an identification problem for ARMA graphical models when some a priori information is available. The main contribution of this article is the introduction of a similarity measure that instead of relying on properties that the network is expected to fulfill, relies on noisy observations of the network at the current time, setting the proposed link-detection method in the context of (partially) data-driven approaches. The positive link prediction problem was rephrased as a suitable optimization problem whose solution has been formally proved to exist and to be unique. Although this work is theoretically focused, the solution has been computed numerically for different synthetic-data case studies and the method has been compared with the existing methods for the identification of graphical models showing an improvement of the performances with respect to the case in which no a-priori information is available.

APPENDIX

PROOF OF SOME TECHNICAL RESULTS

Proof of Lemma 1: The result can be proved by using the dominated-convergence theorem: Let

$$
\begin{align*}
\frac{1}{n} \cdot \log (\Psi^{-1} + Q) + \frac{1}{n} I_m, & \quad n \geq 1 \\
\frac{1}{n} \cdot \log (\Psi^{-1} + Q)
\end{align*}
$$

for $\theta \in [-\pi, \pi]$. Clearly $f_n$ is a continuous and, therefore, measurable function of $\theta$, for any $n \geq 1$, and moreover $f_n \uparrow f$ pointwise (recall that $\Psi^{-1}$ and $Q$ are continuous functions of $\theta$) ensuring that the limit $f$ is itself measurable. In addition, $|f_n| \leq g$ pointwise (and, therefore, a.e.) for any $n \geq 1$, where $g = |f| \in L^1(T, \mathbb{R})$ since $Q \in \Omega_\Psi^+$. By Lebesgue’s dominated-convergence theorem, $f \in L^1(T, \mathbb{R})$ and

$$
\int \lim_{n \to \infty} \int - \log (\Psi^{-1} + Q + \frac{1}{n} I_m) = \lim_{n \to \infty} \int f_n = \int \lim_{n \to \infty} f_n
$$

where the last equality follows from continuity of $\log(\cdot)$ and $\log(\cdot)$. The conclusion is now straightforward

$$
\beta_\Psi(Q) = \lim_{n \to \infty} \beta_\Psi(Q) = \int \log (\Psi^{-1} + Q + \frac{1}{n} I_m) + \lambda h_\sigma^*(Q) = \beta_\Psi(Q)
$$

for any $Q \in \Omega_\Psi^+$.
\[ \mathcal{J}(Q(k)) \geq \text{cost.} + \int \left[ \mu_1 \sum_{i=1}^{m} (\xi_1 + \nu_i^{(k)}) - \frac{1}{\mu_1} \log \left( \nu_i^{(k)} + \xi_m + \frac{1}{n} \right) \right] \]
\[ \geq \text{cost.} + \mu_1 \int \left[ \xi_m + \nu_m^{(k)} - \frac{m}{\mu_1} \log \left( \nu_m^{(k)} + \xi_m + \frac{1}{n} \right) \right] \]
\[ \geq \text{cost.} + \mu_1 \int \left[ \nu_m^{(k)} - \frac{m}{\mu_1} \log \left( \int \nu_m^{(k)} + \xi_m + \frac{1}{n} \right) \right] \]  
(23)

The last step consists in proving that
\[ \mathcal{J}_{\Phi}^n(Q(k)) \rightarrow +\infty \quad \text{when} \quad (Q(k))_{k \geq 1} \subseteq \overline{Q}_{\Phi}^n : \]
\[ \|Q(k)\|_p \rightarrow \infty. \]

Let \((Q(k))_{k \geq 1} \subseteq \overline{Q}_{\Phi}^n\) be such that \(\|Q(k)\|_p \rightarrow \infty\) as \(k \rightarrow \infty\) and denote by \(\lambda^{(k)}_1, (e^{i\theta}) \leq \cdots \leq \lambda^{(k)}_m, (e^{i\theta})\) the eigenvalues of \(\Psi^{-1} + Q(k)\) for any \(k \geq 1\). With \(\xi_i(e^{i\theta})\) and \(\nu_i(e^{i\theta}), i = 1, \ldots, m\), we denote the eigenvalues of \(\Psi^{-1}\) and \(Q(k)\), respectively, assumed to be ordered as the \(\lambda^{(k)}_i\)’s. For the sequence \((Q(k))_{k \geq 1}\) we have \(\|Q(k)\|_p \rightarrow \infty\) as \(k \rightarrow \infty\). Starting from (22) and applying Weyl’s theorem [59, Sec. 4.3] we get (23) shown at the top of this page, where the last step follows from the Jensen’s inequality. Now observe that
\[ \|Q(k)\|_p = \int |\nu_m^{(k)}| = \int_{\nu_m^{(k)}} \nu_m^{(k)} + \int_{|\nu_m^{(k)}} (-\nu_m^{(k)}) \quad (24) \]
where
\[ \nu_m^{(k)} := \{ \theta \in [-\pi, \pi] : \nu_m^{(k)}(e^{i\theta}) > 0 \}, \]
\[ \nu_m^{(k)} := \{ \theta \in [-\pi, \pi] : \nu_m^{(k)}(e^{i\theta}) < 0 \}. \]

From Weyl’s theorem \(-\nu_m^{(k)} \leq \xi_m\), therefore the second integral in (24) is bounded above and
\[ \|Q(k)\|_p \rightarrow \infty \quad \Rightarrow \quad \int_{\nu_m^{(k)}} \nu_m^{(k)} \rightarrow \infty. \]

From (23) the inequality
\[ \mathcal{J}_{\Phi}^n(Q(k)) \geq \text{cost.} + \mu_1 \int \left[ \nu_m^{(k)}(e^{i\theta}) \right] - \mu_1 \int_{|\nu_m^{(k)}} [-\nu_m^{(k)}(e^{i\theta})] \]
\[ - \frac{m}{\mu_1} \log \left( \int \nu_m^{(k)}(e^{i\theta}) + \xi_m + \frac{1}{n} \right) \]
holds for all \(n \geq 1\). By taking the limit for \(k \rightarrow +\infty\) on both sides we obtain that \(\mathcal{J}_{\Phi}^n(Q(k)) \rightarrow +\infty\) when \(\|Q(k)\|_p \rightarrow \infty\) for any \(n \geq 1\) and, therefore, also when \(n \rightarrow \infty\), i.e., \(\mathcal{J}_{\Phi}^n(Q(k)) \rightarrow +\infty\) when \(\|Q(k)\|_p \rightarrow \infty\). Since the index \(\mathcal{J}_{\Phi}^n\) is strictly convex over \(\overline{Q}_{\Phi}\), a Weierstrass’ theorem argument [60, p. 35] allows to conclude that \(\mathcal{J}_{\Phi}^n\) admits unique minimum point \(\overline{Q}_{\Phi}^n\).

Proof of Lemma 3. Let \(Q_o \in \partial \overline{Q}_{\Phi}^n\). For any \(\epsilon > 0\), \(Q_o + \epsilon I_m \in \overline{Q}_{\Phi}^n\) and \(h^\infty(\Phi_{\epsilon} + \epsilon I_m) = h^\infty(Q_o)\). Bringing the limit inside the integral, we can find an upper bound to the right Gâteaux derivative of \(\mathcal{J}_{\Phi}^\infty\) in direction \(\delta Q = I_m\), i.e.,
\[ \delta \mathcal{J}_{\Phi}^\infty(Q_o; I_m) = \lim_{\epsilon \downarrow 0} \frac{\mathcal{J}_{\Phi}^\infty(Q_o + \epsilon I_m) - \mathcal{J}_{\Phi}^\infty(Q_o)}{\epsilon}. \]

Notice that \(\delta \mathcal{J}_{\Phi}^\infty(Q_o; I_m) = dF/\partial \epsilon\), where
\[ F(\epsilon) := \int_{-\pi}^{\pi} \text{tr} \left[ (Q_o + \epsilon I) \Phi_n - \log (\Psi^{-1} + Q_o + \epsilon I) \right] \frac{d\theta}{2\pi} \]
defined for any \(\epsilon > 0\). In fact, notice that
\[ \text{tr} (\theta, \epsilon) := \left[ (Q_o + \epsilon I) \Phi_n - \log (\Psi^{-1} + Q_o + \epsilon I) \right] \]
is integrable for each \(\epsilon > 0\), and by a similar argument as the one used in (23)
\[ \left| \frac{\partial \text{tr} (\theta, \epsilon)}{\partial \epsilon} \right| \leq \text{tr} \left( \Phi_n \right), \]
with \(g \in L^1(\mathbb{T}, \mathbb{H}_m)\), for any \(\epsilon > 0\) and \(\theta \in [-\pi, \pi]\). Accordingly, \(F\) is differentiable and
\[ \delta \mathcal{J}_{\Phi}^\infty(Q_o; I_m) = \frac{dF(\epsilon)}{d\epsilon} = \int \frac{\partial \text{tr} (\theta, \epsilon)}{\partial \epsilon} \]
\[ = \left[ \text{tr} \left( \Phi_n \right) - \text{tr} \left[ I + (\Psi^{-1} + Q_o)^{-1} \right] \right] \leq \mu_m \frac{d\theta}{2\pi} - \text{tr} \left[ I + (\Psi^{-1} + Q_o)^{-1} \right]. \]

1. I. Ahmad, M. U. Akhtar, S. Noor, and A. Shahnaz, “Missing link prediction using common neighbor and centrality based parameterized algorithm,” Sci. Rep., vol. 10, no. 1, 2020, Art. no. 364.
2. S. Kerrache, R. Alharbi, and H. Benhidou, “A scalable similarity-popularity link prediction method,” Sci. Rep., vol. 10, no. 1, 2020, Art. no. 6594.
3. C. G. Akcora, B. Carminati, and E. Ferrari, “Network and profile based measures for user similarities on social networks,” in Proc. IEEE Int. Conf. Inf. Reuse Integration, Aug. 2011, pp. 292–298.
4. T. Wohlfarth and R. Ichise, “Semantic and event-based approach for link prediction,” in Practical Aspects of Knowledge Management, T. Yamaguchi, Ed. Berlin: Springer, 2008, pp. 50–61.
[1] A. Alipour, A. Ferrante, and M. Pavon, “On the well-posedness of inverse problems in signal processing,” IEEE Trans. Autom. Control, vol. 61, no. 9, pp. 2327–2340, Sep. 2016.

[2] D. Wang, D. Pedreschi, C. Song, A.-L. Barabasi, “Human mobility, social ties, and link prediction,” in Proc. 17th ACM SIGKDD Int. Conf. Knowl. Discov. Data Mining, 2011, pp. 1100–1108.

[3] Y. Yang, R. Lichtenwalter, and N. Chawla, “Evaluating link prediction methods,” Knowl. Inf. Syst., vol. 45, no. 3, pp. 751–782, 2015.

[4] S. L. Lauritzen, Graphical Models, vol. 17, Oxford, U.K.: Clarendon Press, 1996.

[5] R. Dahlhaus, “Graphical interaction models for multivariate time series,” Metrika, vol. 51, no. 2, pp. 157–172, 2000.

[6] E. Arvetti, A. G. Lindquist, and B. Wahlberg, “ARMA identification of graphical models,” IEEE Trans. Autom. Control, vol. 58, no. 5, pp. 1167–1178, May 2013.

[7] J. Songseri, J. Dahl, and L. Vandenberghe, “Graphical models of autoregressive processes;” in Convex Optimization in Signal Processing and Communications, Cambridge, U.K.: Cambridge Univ. Press, pp. 89–116, 2010.

[8] M. Zorzi and R. Sepulchre, “AR identification of latent-variable graphical models,” IEEE Trans. Autom. Control, vol. 61, no. 9, pp. 2327–2340, Sep. 2016.

[9] D. Alpago, M. Zorzi, and A. Ferrante, “A scalable strategy for the identification of latent-variable graphical models,” IEEE Trans. Autom. Control, vol. 67, no. 7, Jul. 2022, doi: 10.1109/TAC.2021.3097558.

[10] J. Songseri and L. Vandenberghe, “Topo-logy selection in graphical models of autoregressive processes,” J. Mach. Learn. Res., vol. 11, pp. 2671–2705, 2010.

[11] M. Zorzi, “Empirical Bayesian learning in AR graphical models,” Automatica, vol. 109, 2019, Art. no. 108516.

[12] D. R. Brillinger, “Remarks concerning graphical models for time series and point processes,” Braz. Rev. Econometrics, vol. 16, no. 1, pp. 1–23, 1996.

[13] C. I. Byrnes, S. V. Gusev, and A. Lindquist, “A convex optimization approach to the rational covariance extension problem,” SIAM J. Control Optim., vol. 37, no. 1, pp. 211–229, 1999.

[14] C. Byrnes, T. T. Georgiou, and A. Lindquist, “A new approach to spectral estimation: A tunable high-resolution spectral estimator,” IEEE Trans. Signal Process., vol. 48, no. 11, pp. 3189–3205, Nov. 2000.

[15] A. Lindquist and G. Picci, “Modeling of stationary periodic time series by arma representations,” in Optimization and Its Applications in Control and Data Sciences (Springer Optimization and Its Applications), vol. 115, B. Goldengorin, Ed. Cham, Switzerland: Springer, 2016, pp. 281–314.

[16] T. T. Georgiou and A. Lindquist, “Kullback–Leibler approximation of spectral density functions,” IEEE Trans. Inf. Theory, vol. 49, no. 11, pp. 2910–2917, Nov. 2003.

[17] T. T. Georgiou, “Relative entropy and the multivariable multidimensional moment problem,” IEEE Trans. Inf. Theory, vol. 52, no. 3, pp. 1052–1066, Mar. 2006.

[18] M. Zorzi, “A new family of high-resolution multivariate spectral estimators,” IEEE Trans. Autom. Control, vol. 59, no. 4, pp. 892–904, Apr. 2014.
Daniele Alpago was born in Belluno, Italy, on September 15, 1993. He received the M.S. degree in automation engineering (cum laude), in 2017 from the University of Padova, Padova, Italy, where he is currently working toward the Ph.D degree in algorithms for graphical models selection and control of linear stochastic systems with the Department of Information Engineering.

His research interests include system identification, spectral estimation, and optimal mass transport.

Mattia Zorzi (Senior Member, IEEE) received the M.S. degree in automation engineering and the Ph.D. degree in information engineering from the University of Padova, Padova, Italy, in 2009 and 2013, respectively.

He held Postdoctoral appointments with the Department of Electrical Engineering and Computer Science, University of Liege, Liege, Belgium, and with the Human Inspired Technology Research Centre, University of Padova, Padova, Italy. He held visiting positions with the Department of Electrical and Computer Engineering, University of California, Davis, CA, USA, and with the Department of Engineering, University of Cambridge, Cambridge, U.K., in 2011, 2013, and 2014, respectively. He is currently an Associate Professor with the Department of Information Engineering, University of Padova. His research interests include linear systems, spectral estimation, optimal control and filtering, quantum control, and stochastic realization.

Dr. Zorzi has been an Associate Editor for IEEE CONTROL SYSTEMS LETTERS, since 2019, and Automatica, since 2021. He serves as an Associate Editor on the IEEE Control System Society Conference Editorial Board, since 2017, and in many other international conferences. He is a member of the IFAC Technical Committee on Modelling, Identification and Signal Processing.

Augusto Ferrante was born in Piove di Sacco, Italy, on August 5, 1967. He received the laurea degree (cum laude) in electrical engineering, and the Ph.D. degree in control systems engineering from the University of Padova, Padova, Italy, in 1991 and 1995, respectively.

He has been a Faculty Member with the College of Engineering, University of Udine, Udine, Italy, and of the Politecnico di Milano, Milan, Italy. He is currently the Professor with the Department of Information Engineering, University of Padova. His research interests include linear systems, spectral estimation, optimal control and filtering, quantum control, and stochastic realization.