Proceedings of Machine Learning Research vol xxx:1–13, 2021

Learning local modules in dynamic networks

Paul M.J. Van den Hof P.M.J.VANDENHOF@TUE.NL
Eindhoven University of Technology

Karthik R. Ramaswamy K.R.RAMASWAMY@TUE.NL
Eindhoven University of Technology

Abstract

Over the last decade, the problem of data-driven modeling in linear dynamic networks has been introduced in the literature, and has shown to contain many different challenging research questions. The structural and topological properties of networks become a central ingredient in the data-driven modeling problem, as well as the selection of locations for signals to be sensed and for excitation signals to be added. In this survey-type paper we will present an overview of recent results that are obtained for the problem of learning the dynamics of a single link/module in a dynamic network of which the topology is given. The surveyed methods include extensions of classical identification methods, combined with Bayesian kernel-based methods. Particular attention will be given to the selection of signals that need to be available for measurement/excitation, and accuracy properties of the estimated models in terms of consistency and minimum variance properties.

Keywords: System identification, identifiability, dynamic networks, kernel-based methods, systems over graphs.

1. Introduction

Linear dynamic networks are structured systems that are composed of interconnected linear time-invariant systems. Typically a dynamic network induces a graph, with vertices and edges, that represents the topology of the network. Often a network is represented in a state-space form with states as node signals represented by the vertices in the graph, and the state transitions as links or edges in the graph. However in a data-driven modeling setting, where not all states of a system are typically measured, it has appeared to be attractive to represent the network in a graph that has (measured) node signals as vertices, and dynamic transfer functions on the links/edges. The basic setting of *Dynamic Structure Functions* that was introduced in Gonçalves and Warnick (2008), was generalized to a stochastic estimation and identification setting in Van den Hof et al. (2013), and has been adopted by several different authors.

In this setting a dynamic network is built up out of L scalar *internal variables* or *nodes* w_j, $j = 1, \ldots, L$, and K *external variables* r_k, $k = 1, \ldots K$. Each internal variable is described as:

$$w_j(t) = \sum_{\substack{l=1 \\ l \neq j}}^{L} G_{jl}(q) w_l(t) + u_j(t) + v_j(t) \tag{1}$$

where q^{-1} is the delay operator, i.e. $q^{-1}w_j(t) = w_j(t-1)$;

- G_{jl} are proper rational transfer functions, referred to as *modules*.

- u_j is an input signal, $u_j(t) = \sum_{k=1}^{K} R_{jk}(q) r_k(t)$ with r_k *external variables* that can directly be manipulated by the user.
- v_j is *process noise*, where the vector process $v = [v_1 \cdots v_L]^T$ is modelled as a stationary stochastic process with rational spectral density $\Phi_v(\omega)$, such that there exists a white noise process $e := [e_1 \cdots e_L]^T$, with covariance matrix $\Lambda > 0$ such that $v(t) = H(q)e(t)$, where $H \in \mathbb{R}^{L \times p}(q)$ is stable, monic and minimum-phase.

When combining the L node signals we arrive at the full network expression

$$\begin{bmatrix} w_1 \\ w_2 \\ \vdots \\ w_L \end{bmatrix} = \begin{bmatrix} 0 & G_{12} & \cdots & G_{1L} \\ G_{21} & 0 & \ddots & \vdots \\ \vdots & \ddots & \ddots & G_{L-1\,L} \\ G_{L1} & \cdots & G_{L\,L-1} & 0 \end{bmatrix} \begin{bmatrix} w_1 \\ w_2 \\ \vdots \\ w_L \end{bmatrix} + R \begin{bmatrix} r_1 \\ r_2 \\ \vdots \\ r_K \end{bmatrix} + H \begin{bmatrix} e_1 \\ e_2 \\ \vdots \\ e_p \end{bmatrix}$$

which results in the matrix equation:

$$w = G(q)w + R(q)r + H(q)e, \tag{2}$$

where by construction the matrix G is hollow, i.e. it has diagonal entries 0, while it encodes the topology of the network, i.e. $G_{j\ell}(q) \neq 0$ if and only if there is a connection from node w_ℓ to node w_j in the network.
The single module identification problem to be considered is the problem of identifying one particular module $G_{ji}(q)$ on the basis of measured time-series of a subset of variables in w, and possibly r, for the situation that the network topology is known. This is illustrated in the network depicted in Figure 1.

It may be clear that simply measuring the input and output of the target module and estimating a model on the basis of these signals, will generally not lead to accurate results, because of the signal correlations that are induced by the remaining part of the network. For example, in the situation of Figure 1, with G_{21} being the target module for estimation, estimating the dynamics on the basis of input w_1 and output w_2 only will provide an estimated model that includes the dynamics of the "parallel path" $G_{31}G_{43}G_{42}$.

Non-uniqueness

The network representation (2) will in general be non-unique. E.g. for the situation $R(q) = 0$, the dynamic properties of the network are reflected by the spectral density $\Phi_w(\omega)$, while this spectrum can be generated by different combinations of G, H and Λ. In this situation uniqueness is achieved e.g. if the noise model H is restricted to be diagonal (Bottegal et al. (2018)). However, in situations that the network and its topology result from structured first principle modelling, it is still relevant to consider situations of non-diagonal H, as disturbances in different locations of the network can very well be correlated. For the general case of $R(q) \neq 0$, the freedom of transforming the network to equivalent representations is analyzed in Weerts et al. (2020).

2. Main approaches

We can distinguish several different approaches for addressing the single module identification problem, where the target module is indicated by G_{ji}.

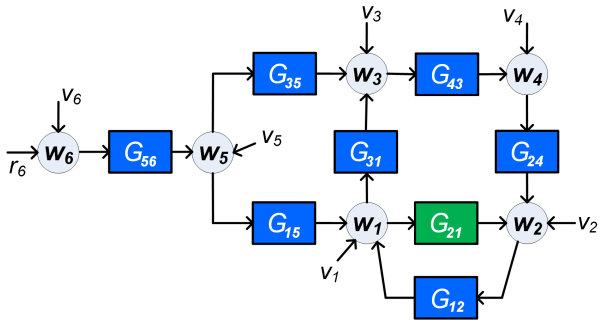

Figure 1: Example network with the green module G_{21} being the target module for identification (Ramaswamy and Van den Hof (2021)).

1. An *indirect method*, that is based on selecting a particular set of predictor input signals r_k, $k \in \mathcal{D}$, and a set of predicted outputs w_ℓ, $\ell \in \mathcal{Y}$, that are used in a predictor model, leading to

$$\varepsilon(t,\theta) = w_{\mathcal{Y}}(t) - \bar{T}(q,\theta) r_{\mathcal{D}}(t). \tag{3}$$

   The matrix $\bar{T}$ refers to a submatrix of the network transfer matrix $T := (I - G)^{-1}R$, which maps external signals r into internal node signals w. In order to extract the dynamics of a particular module G_{ji} from an estimated $\bar{T}$, a postprocessing step is necessary. Consistency of the target module estimate is the typical objective. Different variations of indirect methods exist, including two-stage and instrumental variable (IV) methods.

2. A *direct method*, that is based on selecting a particular set of predictor input signals w_k, $k \in \mathcal{D}$, and a set of predicted output signals w_ℓ, $\ell \in \mathcal{Y}$, with $i \in \mathcal{D}$, $j \in \mathcal{Y}$, and estimating a dynamic model based on a prediction error:

$$\varepsilon(t,\theta) = \bar{H}(q,\theta)^{-1}[w_{\mathcal{Y}}(t) - \bar{G}(q,\theta) w_{\mathcal{D}}(t) - \bar{R} r_{\mathcal{P}}(t)], \tag{4}$$

   where $\bar{G}(q,\theta)$ and $\bar{H}(q,\theta)$ are parametrized transfer function matrices, $\bar{R}$ is a constant selection matrix, and $r_{\mathcal{P}}$ a subset of excitation signals r. The target module is then embedded in the model $\bar{G}(q,\theta)$, and the objective is to estimate the target module consistently and possibly with minimum variance.

3. Recently a generalization of the direct and indirect method was introduced based on a predictor model with prediction error

$$\varepsilon(t,\theta) = \bar{H}(q,\theta)^{-1}[w_{\mathcal{Y}}(t) - \bar{G}(q,\theta) w_{\mathcal{D}_w}(t) - \bar{T}(q,\theta) r_{\mathcal{D}_r}(t)],$$

   and where the target module estimate is obtained after post-processing the estimated $\bar{G}$ and $\bar{T}$. This method allows for more flexibility in selecting the node signals to be measured for identification, see Ramaswamy et al. (2019).

In all of the three approaches, models can be estimated by minimizing a scalar cost function over θ, as e.g., the quadratic cost function

$$\hat{\theta}_N := \arg\min_\theta \frac{1}{N} \sum_{t=1}^{N} \varepsilon^T(t,\theta) P \varepsilon(t,\theta) \tag{5}$$

with P a positive definite weighting function. Specific estimation algorithms are further addressed in Section 5.

Besides the above methods, which are based on selecting an appropriate predictor model, there are alternative methods including non-parametric approaches, where relations between particularly estimated spectral densities of internal signals are used as a basis for module estimation, see Materassi and Salapaka (2015, 2020). Recently also a subspace method has been presented that can handle non-measured interacting signals between the modules, see Yu and Verhaegen (2018). In this paper we will focus on the prediction error approaches listed above.

In order to arrive at accurate, consistent (and possibly minimum variance) estimates of our target module, there are two prime requirements for the estimation setup:

1. An appropriate *predictor model* needs to be chosen. This choice of predictor model includes a selection of node signals to be included as measured signals, and to select inputs and outputs in the predictor model. The predictor model determines where sensors should be available in the network. The predictor model needs to satisfy particular properties in order to guarantee that the target module indeed can be estimated and no uncontrolled bias occurs in the estimated model.

2. The measured data that is taken from the network needs to satisfy condition of *data-informativity*, in other words it needs to be sufficiently rich in order to provide accurate estimates.

The three identification approaches listed above are distinguished by the choice of predictor model. Whereas indirect methods use predictor models having only external excitation signals r as predictor input, the direct method uses node signals w as predictor input, and the generalized method combines both. Different choices of predictor models will lead to different conditions for data-informativity. The direct method (4) has node signals $w_{\mathcal{D}}$ as predictor inputs, and therefore utilizes both external signals r and e for creating data-informativity. On the other hand, indirect methods rely on external excitation signals r only for data informativity, and therefore will typically require more "expensive" external excitations. The direct method provides asymptotically efficient estimates (i.e. consistency and minimum variance for the identification setup) at the cost of the need to include noise models $\bar{H}(q,\theta)$. The indirect method and its variations provides consistent estimates but not with minimum variance. When the node signals are measured with sensor noise (errors-in-variables (EIV) situation), the direct method becomes biased and the indirect method provides consistent estimates of the target module.

3. Indirect method

3.1. General approach

The network model (2) can be rewritten as $w = Tr + \bar{v}$ where $\bar{v} = (I-G)^{-1}He$. A consistent estimate $\hat{T}(q)$ of $T(q)$ can be obtained using open loop MIMO identification method as in (3). On the basis of $\hat{T}(q)$, a consistent estimate $\hat{G}$ of G can be obtained by solving $(I-\hat{G})\hat{T}(q) = R$. By identifiying only a submatrix of T and solving only a subset of the above equations, a target module embedded in the dynamic network can be identified, see Gevers et al. (2018); Hendrickx et al. (2019); Bazanella et al. (2019).

This latter situation leads to a predictor model setup as depicted in Figure 2. The output $w_{\mathcal{Y}}$ of the predictor model is selected to be composed of w_j and all node signals $w_{\mathcal{N}}$ that are in-neighbours

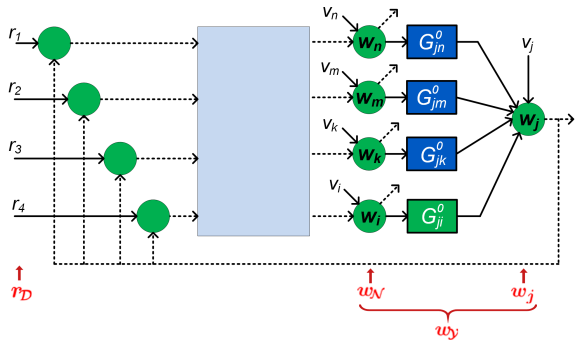

Figure 2: Predictor model for the indirect method.

of w_j, i.e. $\mathcal{N} = \mathcal{N}_j^-$. Using the predictor model

$$\varepsilon(t,\theta) = \bar{H}(q,\theta)^{-1}\left[w_\mathcal{Y}(t) - \bar{T}(q,\theta)r_\mathcal{D}\right]$$

where $\bar{T}$ is decomposed as

$$\bar{T} = \begin{bmatrix} \bar{T}_{jr_\mathcal{D}} \\ \bar{T}_{\mathcal{N}r_\mathcal{D}} \end{bmatrix}. \tag{6}$$

If a consistent estimate $\hat{T}$ of $\bar{T}$ is made, then a consistent estimate of $G^0_{j\mathcal{N}}$ is obtained according to

$$\hat{G}_{j\mathcal{N}} = \hat{T}_{jr_\mathcal{D}}\hat{T}^\dagger_{\mathcal{N}r_\mathcal{D}} \tag{7}$$

where $\hat{T}^\dagger_{\mathcal{N}r_\mathcal{D}}$ is the right inverse of $\hat{T}_{\mathcal{N}r_\mathcal{D}}$.

The fact that the right inverse of $\hat{T}_{\mathcal{N}r_\mathcal{D}}$ needs to exist, requires the presence of a sufficient number of external excitation signals r in the network, i.e. $dim(r_\mathcal{D}) \geq dim(w_\mathcal{N})$. So we need at least as many external excitation signals r to be present as there are in-neighbours of w_j.

3.2. Reducing the number of node signals

In general it is not necessary to use the full set of in-neighbours $w_\mathcal{N}$ with $\mathcal{N} = \mathcal{N}_j^-$ in the output $w_\mathcal{Y}$. It is sufficient to select a subset $w_\mathcal{N}$ that satisfies the property that upon removal (*immersion*) of the remaining unmeasured nodes from the network, while keeping the remaining node signals invariant, the target module remains invariant too. This is achieved if the *parallel path and loop condition* is satisfied (Dankers et al. (2016)):

Condition 1 (Parallel path and loop condition) *In the graph of the network model, every path from (w_i, w_j) to w_j with length ≥ 2 passes through a node that is included in the predictor model.*

With the parallel path and loop condition satisfied, we can perform the identification steps as presented in Section 3.1, with a subset of the in-neighbour nodes of w_j, i.e., $w_\mathcal{N}$ with $\mathcal{N} \subset \mathcal{N}_j^-$. At the same time, this reduces the required number of excitation signals present in the network, so as to guarantee the existence of the right inverse of $\hat{T}_{\mathcal{N}r_\mathcal{D}}$ in (7).

In Shi et al. (2020) the above situation has been generalized and the handling of Condition 1 has been reformulated in terms of finding a *disconnecting set* between node signals w_i and $\mathcal{N}_j^-$

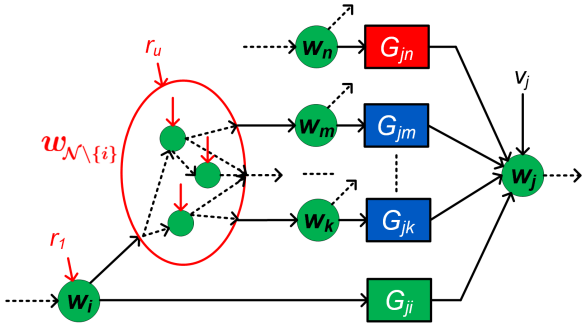

Figure 3: Node signal selection through disconnecting set.

while excluding node signals that are input to modules of which the dynamics is known a priori. This situation is depicted in Figure 3. In this Figure the red module G_{jn} is a module of which the dynamics is a priori known, and the set of nodes in $w_{\mathcal{N}\backslash\{i\}}$ is a (minimum) disconnecting set between the nodes w_i and $\{w_k \cdots w_m\}$. With $r_{\mathcal{D}} = (r_1, r_u^T)^T$ and $dim(r_u) = dim(w_{\mathcal{N}\backslash\{i\}})$, a consistent estimate of G_{ji}^0 can then be obtained according to

$$\hat{G}_{ji} = \hat{T}_{jr_{\mathcal{D}}} \hat{T}^{\dagger}_{\mathcal{N}r_{\mathcal{D}}} \pi_i \tag{8}$$

with π_i the $i-th$ unit vector. Excitation requirements on $r_{\mathcal{D}}$ can now again be formulated in terms of the existence of the right inverse of $\hat{T}_{\mathcal{N}r_{\mathcal{D}}}$, implying that $dim(r_{\mathcal{D}}) \geq dim(w_{\mathcal{N}})$. The r-signals can be added to the nodes in $w_{\mathcal{N}}$ directly, or can also be added elsewhere in the network, as in Figure 2. If $\mathcal{N}\backslash\{i\}$ is a *minimum disconnecting set* then automatically the number of required external excitation signals is minimized, thus providing the "cheapest" possible experimental setup. Note that the node signals that need to be available for estimating the dynamics of G_{ji}^0 now become: $w_j, w_{\mathcal{N}}$ and w_n. The in-neighbours $\{w_k, \cdots w_m\}$ are not necessarily included now. A further generalization of the signal selection problem can be achieved by replacing the parallel path and loop condition by a result from network abstractions (Weerts et al. (2020)), allowing in stead of measuring nodes in the parallel paths and loops, measuring particular descendants of those nodes.

The considerations above show the results for the choice of an appropriate *predictor model*. For the indirect method the requirements for *data-informativity* are rather straighforward: a sufficient condition for data-informativity is that $\Phi_r(\omega) > 0$ for almost all ω, with $\Phi_r(\omega)$ the spectral density of r, see Van den Hof and Ramaswamy (2020).

The indirect method has been studied in different settings. The above reasoning follows from the starting point of *full measurement*, i.e. all node signals are assumed available for measurement, although in the final result only a subset of node signals needs to be measured. In a dual setting of *full excitation*, i.e., having excitation signals on all node signals, it has been analyzed which node signals to measure for consistent identification of the target module, see Bazanella et al. (2017); Hendrickx et al. (2019). This situation is further relaxed in Bazanella et al. (2019).

Related indirect methods, such as the *two-stage method* and the *Instrumental Variable* (IV) method have been presented in Van den Hof et al. (2013) and Dankers et al. (2015) respectively. A semi-parametric approach has been introduced in Galrinho et al. (2017) where a parametric model of the target module is consistently identified using a multi-step approach, while avoiding non-

convex optimization. All the indirect methods can handle an EIV setting as well as networks with correlated process noise.

4. Direct method

4.1. General setup

In the situation that it is known that the process disturbances are uncorrelated, i.e. Φ_v is diagonal, the direct identification method can typically be reduced to a MISO problem, where (4) reduces to

$$\varepsilon(t,\theta) = \bar{H}(q,\theta)^{-1}\left[w_j(t) - \bar{G}(q,\theta)w_{\mathcal{D}}(t) - \bar{R}r_{\mathcal{P}}(t)\right] \tag{9}$$

where $w_{\mathcal{D}}$ is a vector signal, and $w_{\mathcal{Y}} = w_j$ is scalar. The simplest situation is when all w-in-neighbors of w_j (denoted by $w_{\mathcal{N}_j^-}$) are included, i.e. $w_{\mathcal{D}} = w_{\mathcal{N}_j^-}$.

The target module can then directly be parametrized and estimated as part of a MISO model, see Van den Hof et al. (2013). For consistency, conditions on informativity of the data have to be satisfied, implying that sufficient excitation should be present in the predictor inputs. A typical, but conservative, condition is that $\Phi_{w_m}(\omega)$ is positive definite for a sufficient number of frequencies where w_m is the vector of stacked predicted output and predictor input signals. For a particular situation of limited order models a less conservative condition is formulated in Gevers and Bazanella (2015). Excitation signals r are not explicitly included in the predictor model, but they play an obvious role in the realization of the data-informativity conditions. Note that besides the external excitation signals r, also the disturbance signals e (and their filtered versions v) can contribute to a sufficient excitation of the predictor input signals. This is schematically depicted in Figure 4.

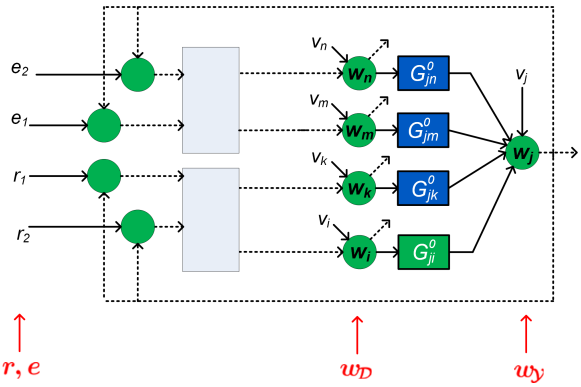

Figure 4: Predictor model for direct method.

Like in the situation of the indirect method, the number of node signals that needs to be included in $w_{\mathcal{D}}$ can be further reduced by satisfying the parallel path and loop condition, or similarly, by constructing a (minimum) disconnecting set between the nodes i and $\mathcal{D}\backslash\{i\}$. However one of the possible consequences of removing an w-in-neighbour of w_j from $w_{\mathcal{D}}$ is that in the immersed network, in which the considered node is removed, the disturbance signals that are acting on w_j and those acting on $w_{\mathcal{D}}$ can get correlated. The same situation occurs if in the original network the disturbance signals v are correlated, i.e., the spectral density $\Phi_v(\omega)$ is non-diagonal. In those situations we have to deal with the presence of confounding variables as discussed next.

4.2. Handling confounding variables in the direct method

In one of the following situations

a. $\Phi_v(\omega)$ is non-diagonal such that v_j and $v_k, k \in \mathcal{D}$ are correlated, or

b. Some in-neighbours of w_j are not included in $w_\mathcal{D}$, after immersion,

correlations between inputs and output in the predictor model are possibly introduced, that are referred to as confounding variables.

Definition 1 (Confounding variable) *A confounding variable is an unmeasured variable that affects both the input and output of an estimation problem.*

When not properly accounted for, confounding variables typically destroy the consistency properties of the direct method for estimating G_{ji}^0, as they introduce correlation between the measured node signals w_i and w_j that is not induced by the module dynamics G_{ji}^0. Phrased in identification-terms, confounding variables are correlated disturbances. We distinguish two types of confounding variables, depending on their origin of appearance. Next to the *indirect confounding variables* that can occur due to non-measured input signals, we distinguish *direct confounding variables* that appear as a result of disturbance correlations in Φ_v. There are different solutions for mitigating the effect of confounding variables that affect the construction of the predictor model.

(a) Direct confounding variables can only be handled by including the predictor input node with a disturbance that correlates with the output, in the predictor output too. In this way the correlated disturbance between the two node signals can be modelled through a multivariate noise model $\bar{H}$;

(b). Indirect confounding variables can be handled by either following the previous strategy (a), or by adding a measured node signal to $w_\mathcal{D}$ that blocks the path from the confounding variable to the input signal.

Handling confounding variables by adding predictor inputs has been addressed in Dankers et al. (2016, 2017), while the solution through increasing the number of predicted outputs has been introduced in Ramaswamy and Van den Hof (2021). The handling of confounding variables is illustrated in the following example.

Example 1 *Consider the network as sketched in Figure 5 (left) where node signal w_7 is unmeasured, node signals w_1, w_2, w_3, w_6 are measured, and the target is to estimate module G_{21}^0. Becasue of the path from node w_7 to both input w_1 and output w_2, the unmeasured node signal w_7 acts as an indirect confounding variable for the estimation problem $(w_1, w_3, w_6) \to w_2$. The effect of this (indirect) confounding variable can be mitigated in two different ways. If the path from w_7 to w_1 can be "blocked" by an additional node signal that can be measured, like node w_4 in Figure 5 (right), then w_4 is included as predictor input, leading to the predictor model $(w_1, w_3, w_4, w_6) \to w_2$. Alternatively node w_1 can be added to the output, where the correlated disturbances are modelled through a 2×2 noise model in the predictor model $(w_1, w_3, w_6) \to (w_1, w_2)$.*

The construction of an appropriate predictor model can now be performed according to the following steps:

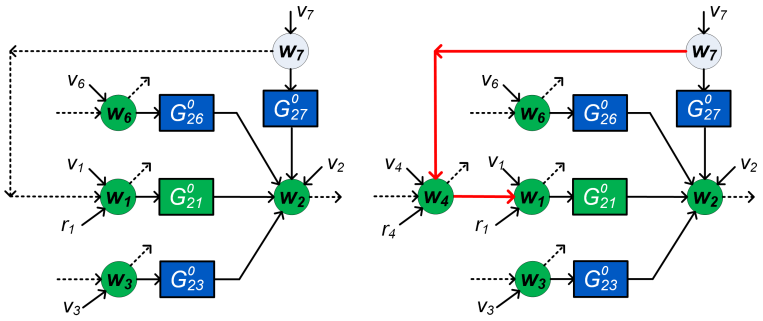

Figure 5: Handling of an indirect confounding variable. Left Figure: w_7 is an indirect confounding variable for the estimation problem $(w_1, w_3, w_6) \to w_2$; Right Figure: an additional node signal w_4 is included in $w_{\mathcal{D}}$ that blocks the path from w_7 to w_1.

1. Select $w_{\mathcal{D}} = w_i$ and $w_{\mathcal{Y}} = w_j$;
2. Add predictor inputs in $w_{\mathcal{D}}$ to satisfy the parallel path and loop condition;
3. Mitigate the effect of indirect and direct confounding variables, by adding predictor inputs and/or adding predicted outputs;

This leads to a description of the relation between the node variables that are incorporated in the predictor model, as:

$$w_{\mathcal{Y}}(t) = \bar{G} w_{\mathcal{D}}(t) + \bar{R} r_{\mathcal{P}}(t) + \bar{H} \xi_{\mathcal{Y}}(t) \tag{10}$$

where only those r signals are included in $r_{\mathcal{P}}$ that warrant matrix $\bar{R}$ to be a known selection matrix, and so non-dynamic. The effect of all remaining r-signals is incorporated in the noise term $\bar{H}\xi_{\mathcal{Y}}$. The prediction error that forms the basis for the estimation algorithm then becomes

$$\varepsilon(t, \theta) = \bar{H}(q, \theta)^{-1} \left[w_{\mathcal{Y}}(t) - \bar{G}(q, \theta) w_{\mathcal{D}}(t) - \bar{R} r_{\mathcal{P}}(t) \right]. \tag{11}$$

It has to be noted that there can be a set of node signals, denoted by $w_{\mathcal{Q}}$ that appear both in $w_{\mathcal{D}}$ and in $w_{\mathcal{Y}}$, as also illustrated in Example 1.

Path-based conditions for the construction of an appropriate predictor model have been formulated in Ramaswamy and Van den Hof (2021). The results provide multiple options for the user to choose from, in terms of how many and which node signals to include in the estimation setup, dependent on the availability of measured node signals.

Data-informativity conditions are typically formulated in the form of a positive definite signal spectrum of a specified signal vector, and thus requiring sufficient (external) excitation, either through excitation or through disturbance signals. Also for this property path-based conditions are formulated that guarantee data-informativity generically, and that are based on a sufficient number of paths that connect particular externals signals r, e to the predictor inputs, see Van den Hof and Ramaswamy (2020). One of the specific results on data-informativity is that $dim(r) \geq dim(w_{\mathcal{Q}})$, i.e. there should be at least as many external excitation signals as there are node signals in $w_{\mathcal{Q}}$.

5. Algorithmic aspects - kernel based methods

The (non-convex) optimization problems (5) that need to be solved for the different identification approaches will typically grow in dimension of the number of unknown parameters, to a large extent because the number of modules that needs to be estimated in the resulting MISO or MIMO predictor models will grow. Rather than solving for the large scale non-convex optimization problems, algorithms inspired by machine learning that more effectively handle the complexity of the estimation problem are favoured.

By incorporating kernel-based method, the impulse response(s) of the modules are modeled as zero-mean Gaussian processes whose covariance(s) are described by a kernel that ensures smoothness and stability of the model. A probabilistic description of the model is obtained and the coefficients of the impulse response(s) are obtained by estimating the hyperparameters of the kernel by maximizing the marginal likelihood of the data. In this way, the impulse response of each module is obtained through estimating only a few hyperparameters (eg. 2 hyperparamters per module for stable spline kernel).
In Chiuso and Pillonetto (2012), this has been applied in a time domain non-parametric approach for estimating models in a MISO setup with white output noise. Following a semi-parametric approach to a dynamic network with only sensor noise (no process noise), the increase in variance due to high order modeling in a two-stage method is reduced in Everitt et al. (2018) by applying a kernel-based method.
Similarly, the direct method demands a model order selection step for all modules in the MISO setup, which increases the complexity and induces estimation of large number of nuisance parameters. In Ramaswamy et al. (2018), a kernel-based method has been presented to tackle these problems by modeling the target module as a parametric model and the remaining modules in the MISO setup as Gaussian processes, thus avoiding the model order selection step and decreasing the number of parameters. This offers a substantially reduced variance of the target module estimate. In Rajagopal et al. (2020) this approach has been extended to the situation of MIMO predictor models.

6. Identifiability

When considering network identifiability (Weerts et al. (2018a)), conditions can be formulated for identifiability of a single module in a network model set. This typically leads to rank conditions on particular transfer functions from external signals to particular node signals. In a generic sense, this can be translated to path-based conditions on the graph of the network model set (Hendrickx et al. (2019) and followed up by Weerts et al. (2018b)). A synthesis procedure to assign and locate the minimum number of external excitation signals for guaranteeing local module generic identifiability, is provided in Shi et al. (2020). These results are typically independent of the particular identification method considered.

7. Conclusions

We have surveyed a class of data-driven modeling methods for learning the dynamics of a single module that is embedded in a dynamic network of which the topology is a priori given.

Acknowledgments

This project has received funding from the European Research Council (ERC), Advanced Research Grant SYSDYNET, under the European Union's Horizon 2020 research and innovation programme (grant agreement No 694504).
Extended version of an extended abstract presented as survey at the 2020 IFAC World Congress, submitted for presentation at the 3rd Annual Learning for Dynamics and Control Conference, June 7-8, 2021, ETH Zürich, Switzerland.

References

A. S. Bazanella, M. Gevers, J. M. Hendrickx, and A. Parraga. Identifiability of dynamical networks: which nodes need to be measured? In *Proc. 56th IEEE Conference on Decision and Control (CDC)*, pages 5870–5875, 2017.

A. S. Bazanella, M. Gevers, and J. M. Hendrickx. Network identification with partial excitation and measurement. In *Proc. 58th IEEE Conference on Decision and Control (CDC)*, pages 5500–5506, 2019.

G. Bottegal, A. Chiuso, and P. M. J. Van den Hof. On dynamic network modeling of stationary multivariate processes. *IFAC-PapersOnLine*, 51-15:850–861, 2018. Proc. 18th IFAC Symp. System Identification.

A. Chiuso and G. Pillonetto. A Bayesian approach to sparse dynamic network identification. *Automatica*, 48(8):1553—-1565, 2012.

A. G. Dankers, P. M. J. Van den Hof, X. Bombois, and P. S. C. Heuberger. Errors-in-variables identification in dynamic networks – consistency results for an instrumental variable approach. *Automatica*, 62:39–50, 2015.

A. G. Dankers, P. M. J. Van den Hof, P. S. C. Heuberger, and X. Bombois. Identification of dynamic models in complex networks with prediction error methods: Predictor input selection. *IEEE Trans. on Automatic Control*, 61(4):937–952, 2016.

A. G. Dankers, P. M. J. Van den Hof, D. Materassi, and H. H. M. Weerts. Relaxed predictor input selection rules for handling confounding variables in dynamic networks. *IFAC-PapersOnLine*, 50-1:3983–3988, 2017. Proc. 20th IFAC World Congress.

N. Everitt, G. Bottegal, and H. Hjalmarsson. An empirical Bayes approach to identification of modules in dynamic networks. *Automatica*, 91:144–151, May 2018.

M. Galrinho, N. Everitt, and H. Hjalmarsson. Incorporating noise modeling in dynamic networks using non-parametric models. *IFAC-PapersOnLine*, 50(1):10568–10573, 2017. Proc. 20th IFAC World Congress.

M. Gevers and A. S. Bazanella. Identification in dynamic networks: identifiability and experiment design. In *Proc. 2015 IEEE 54th Conf. Decision and Control, Osaka, Japan*, pages 4006–4011, 2015.

M. Gevers, A.S. Bazanella, and G. Vian da Silva. A practical method for the consistent identification of a module in a dynamical network. *IFAC-PapersOnLine*, 51-15:862–867, 2018. Proc. 18th IFAC Symp. System Identif. (SYSID2018).

J. Gonçalves and S. Warnick. Necessary and sufficient conditions for dynamical structure reconstruction of LTI networks. *IEEE Trans. Automatic Control*, 53(7):1670–1674, August 2008.

J.M. Hendrickx, M. Gevers, and A.S. Bazanella. Identifiability of dynamical networks with partial node measurements. *IEEE Trans. Autom. Control*, 64(6):2240–2253, 2019.

D. Materassi and M. Salapaka. Identification of network components in presence of unobserved nodes. In *Proc. 2015 IEEE 54th Conf. Decision and Control, Osaka, Japan*, pages 1563–1568, 2015.

D. Materassi and M. V. Salapaka. Signal selection for estimation and identification in networks of dynamic systems: a graphical model approach. *IEEE Trans. Automatic Control*, 65(10):4138–4153, october 2020.

V. C. Rajagopal, K. R. Ramaswamy, and P. M. J. Van den Hof. A regularized kernel-based method for learning a module in a dynamic network with correlated noise. In *59th IEEE Conf. Decision and Control*, Jeju Island, Republic of Korea, 2020.

K. R. Ramaswamy and P. M. J. Van den Hof. A local direct method for module identification in dynamic networks with correlated noise. *IEEE Trans. Automatic Control*, 66(11), November 2021. ArXiv: 1908.00976.

K. R. Ramaswamy, G. Bottegal, and P. M. J. Van den Hof. Local module identification in dynamic networks using regularized kernel-based methods. In *Proc. 57th IEEE Conf. on Decision and Control (CDC)*, pages 4713–4718, Miami Beach, FL, 2018.

K. R. Ramaswamy, P. M. J. Van den Hof, and A. G. Dankers. Generalized sensing and actuation schemes for local module identification in dynamic networks. In *Proc. 58th IEEE Conf. on Decision and Control (CDC)*, pages 5519–5524, Nice, France, 2019.

S. Shi, X. Cheng, and P. M. J. Van den Hof. Excitation allocation for generic identifiability of a single module in dynamic networks: A graphic approach. In *Proc. 21st IFAC World Congress*, Berlin, Germany, 2020.

P. M. J. Van den Hof and K. R. Ramaswamy. Path-based data-informativity conditions for single module identification in dynamic networks. In *59th IEEE Conf. Decision and Control*, Jeju Island, Republic of Korea, 2020.

P. M. J. Van den Hof, A. G. Dankers, P. S. C. Heuberger, and X. Bombois. Identification of dynamic models in complex networks with prediction error methods - basic methods for consistent module estimates. *Automatica*, 49(10):2994–3006, 2013.

H. H. M. Weerts, P. M. J. Van den Hof, and A. G. Dankers. Identifiability of linear dynamic networks. *Automatica*, 89:247–258, March 2018a.

H. H. M. Weerts, P. M. J. Van den Hof, and A. G. Dankers. Single module identifiability in linear dynamic networks. In *Proc. 57th IEEE Conf. on Decision and Control (CDC)*, pages 4725–4730, Miami Beach, FL, 2018b. IEEE.

H. H. M. Weerts, J. Linder, M. Enqvist, and P. M. J. Van den Hof. Abstractions of linear dynamic networks for input selection in local module identification. *Automatica*, 117(108975), July 2020.

C. Yu and M. Verhaegen. Subspace identification of individual systems operating in a network (SI^{2}ON). *IEEE Trans. Automatic Control*, 63(4):1120–1125, 2018.