Local Identifiability of Fully-Connected Feed-Forward Networks with Nonlinear Node Dynamics

Martina Vanelli and Julien M. Hendrickx

***Abstract*— We study the identifiability of nonlinear network systems with partial excitation and partial measurement when the network dynamics is linear on the edges and nonlinear on the nodes. We assume that the graph topology and the nonlinear functions at the node level are known, and we aim to identify the weight matrix of the graph. Our main result is to prove that fully-connected layered feed-forward networks are generically locally identifiable by exciting sources and measuring sinks in the class of analytic functions that cross the origin. This holds even when all other nodes remain unexcited and unmeasured and stands in sharp contrast to most findings on network identifiability requiring measurement and/or excitation of each node. The result applies in particular to feed-forward artificial neural networks with no offsets and generalizes previous literature by considering a broader class of functions and topologies.**

I. Introduction

Networks of interconnected dynamical systems are widespread across various domains [1]. Analyzing these systems and developing control strategies require understanding the interconnections, typically modeled as edges of a graph. However, identifying these systems from partial excitations and partial measurements is challenging because measured signals reflect combined dynamics [2].

In this work, we consider network systems where the dynamics is linear on the edges and nonlinear on the nodes. More precisely, we denote with $\mathcal{G} = (\mathcal{N}, \mathcal{E}, W)$ weakly connected directed acyclic graphs with set of nodes $\mathcal{N}$, set of directed links $\mathcal{E}$, and weight matrix W whose entries are such that $w_{ij} \neq 0$ if and only if $(i,j) \in \mathcal{E}$. For a given nonlinear function f, we assume that the output of a node at time k is

$$y_i^k = f\big(\sum_j w_{ij} y_j^{k-1}\big) + u_i^{k-1} \tag{1}$$

where $u_i^{k-1} \in \mathbb{R}$ is an external excitation signal. If a node is not excited, its corresponding excitation signal is set to zero. For example, for the network in Fig. 1 where node 1 and 2 are excited, the output of node 4 at time k is given by $y_4^k = f(w_{43} f(w_{31} u_1^{k-3} + w_{32} u_2^{k-3}))$. In our setting, the topology of the graph $\mathcal{E}$ and the nonlinear function f are known, while the weight matrix W is unknown. We are then interested in determining the identifiability of W with partial excitation and partial measurement, that is, we want to determine if, given the sets of excited and measured nodes, there exist two different weight matrices leading to the same input-output behaviors. In this latter case, we say that the network is not identifiable as different weight matrices cannot be distinguished by experiments in which we excite and measure only such nodes.

M. Vanelli and J. M. Hendrickx are with the ICTEAM Institute, UCLouvain, B-1348 Louvain-la-Neuve, Belgium (email: {martina.vanelli, julien.hendrickx}@uclouvain.be). This work was supported by F.R.S.-FNRS via the KORNET project.

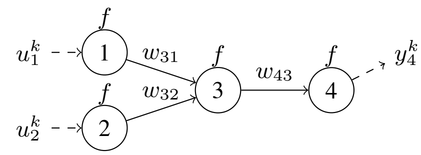

Fig. 1. Nonlinear node dynamics where nodes 1 and 2 are excited and node 4 is measured.

Network identifiability of linear systems (when $f(x) = x$) with partial excitation and partial measurement has been the subject of recent research [3]–[7]. While some graph theoretical conditions exist for full measurement scenarios, a general solution remains elusive. Furthermore, real-world systems are mostly nonlinear. Identifiability of nonlinear systems is even more challenging and far less studied in the literature. The recent work in [8]–[10] addresses nonlinear network identifiability with full/partial excitation and partial measurement when the dynamics is additive on the edges.

In this work, we introduce the notion of genericity in the class of analytic functions and we prove that fully-connected feed-forward networks are generically locally identifiable by *only* exciting sources and measuring sinks for almost all analytic functions that are zero in zero. This result contrasts with findings in identifiability of network systems, where exciting and/or measuring every node is proved to be necessary in both linear [3] and nonlinear [10] systems. In our setting, the presence of the node nonlinearity enables the identification of paths where some nodes remain unexcited and unmeasured, while the linear dynamics on the edges provide sufficient structure for network identification. In the full version of this work, we will generalize the result to more general network structures.

Examples of network systems characterized by nonlinear subsystems with linear interactions are continuous thresholds models [11], that are continuous generalizations of linear threshold models [12], and nonlinear consensus problems [13], [14]. In these two settings, the action (resp. the opinion) of an agent at time k depends in a nonlinear manner on the weighted sum of the actions (resp. the opinions) of his neighbors at time $k-1$. Identifying the weights allows to quantify the strengths and the types of the interconnections among the agents [15]. Networks systems of the form in (1) can be found also in network games, production networks, and models of financial interactions (see [16]).

Our result applies in particular to layered feed-forward neural networks with no offsets [17]. In this context, the objective is to determine if, given a function F and a nonlinearity f, it is possible to determine the architecture, weights, and biases of all feed-forward neural networks that realize F. In [18], Sussmann studied single-hidden-layer hyperbolic tangent networks and showed that two irreducible networks are functionally equivalent (i.e., give the same input-output map) if and only if they are related by simple operations of exchanging and negating the weights of hidden units. This result was later extended to architectures with a broader class of nonlinearities [19], to architectures with multiple hidden layers and hyperbolic tangent as nonlinearity [20] and to certain recurrent architectures [19]. More recently, the same problem has been studied for ReLU networks [21] and for a finite number of samples [22]. In particular, [23], [24] have generalised Sussmann's results to a broader class of sigmoidal nonlinearities. These last works are particularly relevant as the authors derive necessary conditions for the identifiability of deep networks of arbitrary depth, connectivity and nonlinearities and construct a family of nonlinearities for which these genericity conditions both necessary and sufficient. Here, we extend previous work by considering a broader class of nonlinearities and/or networks, though having some limitations such as a given architecture, the absence of biases and local results that holds generically in the class of functions.

II. Problem formulation

We consider the model class in (1), where all nodes have the same activation function. We assume to know the nonlinearity f and the graph topology $\mathcal{E}$ and we want to determine the identifiability of W with partial excitation and partial measurement.

For the identification process, we assume that the relations between excitations and outputs of the nodes have been perfectly identified. Following [9], we denote with $\mathcal{N}^e$ and $\mathcal{N}^m$ in $\mathcal{N}$ the set of excited and measured nodes, respectively. We consider a node $i \in \mathcal{N}^m$ and we let $\mathcal{N}_i^{e,p} \subseteq \mathcal{N}^e$ denote the set of excited nodes that have a path to node i. If we measure a node i at time k, we obtain

$$y_i^k = u_i^{k-1} + F_i(u_1^{k-2}, \dots, u_1^{k-m_1}, \dots, u_{n_i}^{k-2}, \dots, u_{n_i}^{k-m_{n_i}}), \quad 1, \dots, n_i \in \mathcal{N}_i^{e,p}. \tag{2}$$

The function F_i is implicitly defined in (2) and only depends on a finite number of inputs due to the absence of memory on the edges and nodes, and the absence of cycles.

The identifiability problem is related to the possibility of identifying the weight matrix W based on several measurements. Given a set of measured nodes $\mathcal{N}^m \subseteq \mathcal{N}$, the set of measured functions $F(\mathcal{N}^m)$ associated with $\mathcal{N}^m$ is:

$$F(\mathcal{N}^m) := \{F_i \, , \, i \in \mathcal{N}^m\} \, . \tag{3}$$

Example 1: Consider the graph $\mathcal{G}$ in Fig. 1 and assume that $\mathcal{N}^e = \{1, 2\}$, that is, we can only excite the sources. Then, if we measure the node 4 at time k, we obtain

$$F_4(u_1^{k-3}, u_2^{k-3}) = y_4^k = f(w_{43} f(w_{31} u_1^{k-3} + w_{32} u_2^{k-3})) \tag{4}$$

Observe that the function F_4 depends on the inputs of the excited nodes 1 and 2 that have a path to the node 4.

We say that a weight matrix W is *consistent* with the edge set $\mathcal{E}$ if $w_{ij} \neq 0$ if and only if (i,j) in $\mathcal{E}$. We then denote with $F(\mathcal{N}^m)$ the set of measured functions generated by a weight matrix W consistent with $\mathcal{E}$ and with $\tilde{F}(\mathcal{N}^m)$ the set generated by another matrix $\tilde{W}$ consistent with $\mathcal{E}$.

Definition 1 (Identifiability of a network): Consider a graph $\mathcal{G} = (\mathcal{N}, \mathcal{E}, W)$ with sets of excited and measured nodes $\mathcal{N}^e$ and $\mathcal{N}^m$, respectively, and let $\mathcal{F}$ be a class of functions. A *network* $\mathcal{G}$ is *identifiable* in $\mathcal{F}$ if, for any given f in $\mathcal{F}$, $F(\mathcal{N}^m) = \tilde{F}(\mathcal{N}^m)$ implies $W = \tilde{W}$.

If a network is not identifiable, different weight matrices can generate exactly the same behavior, and therefore, recovering the weights from several experiments is impossible. On the other hand, if the network is identifiable, the weights can be distinguished from all others. Then, if the functions in $F(\mathcal{N}^m)$ can be well approximated after sufficiently long experiments, it becomes feasible to approximate the weights through excitation and measurement.

We remark that, if sources are not excited, it is not possible to identify the weights of their outgoing edges, while, if sinks are not measured, the weights of their incoming edges are not identifiable. Therefore, the minimal sets of excited and measured nodes must contain the sets of sources and sinks, respectively. The goal of this work is to determine conditions on the edge set and the function space under which the network is identifiable by *only* exciting sources and measuring sinks.

In this preliminary work, we consider layered feed-forward networks (LFNs) of depth $L > 0$, that is, directed acyclic graphs $\mathcal{G} = (\mathcal{N}, \mathcal{E}, W)$ with $\mathcal{N} = \mathcal{N}^0 \cup \dots \cup \mathcal{N}^L$ where $\mathcal{N}^l \cap \mathcal{N}^k = \emptyset$ for every $l \neq k$ and

$$\mathcal{E} \subseteq \mathcal{E}_{\mathrm{F}} := \{(i,j) \, , i \in \mathcal{N}^{l-1} \, , j \in \mathcal{N}^l \, , \forall l\} \, . \tag{5}$$

If $\mathcal{E} = \mathcal{E}_{\mathrm{F}}$, we say that $\mathcal{G}$ is a *fully-connected* LFN. For LFNs, we can study every pair of source-sink separately by studying an associated function F and the corresponding weight matrix W. More precisely, consider a source i_e in $\mathcal{N}^e$ and sink i_m in $\mathcal{N}^m$ and let and $u_i^k = 0$ for every $i \neq i_e$. Then, we can define the function $F : \mathbb{R} \to \mathbb{R}$, $F(x) := F_{i_m}(x, \mathbf{0})$ where $x = u_{i_e}^{k-L+1}$ is the input of node i_e. For instance, for the graph in Fig. 1, if we select the source 1 and the sink 4, by letting $u_2^{k-3} = 0$ for all k and $x = u_1^{k-3}$, we find

$$F(x) := F_4(x, 0) = f(w_{43} f(w_{31} x)) \tag{6}$$

When $f(0) = 0$, we further have that $F(x)$ depends only on the paths from i_e to i_m. Our goal is then to determine if $F = \tilde{F}$ implies $W = \tilde{W}$, which is the problem commonly addressed in artificial neural networks. If this holds for all pairs i_e, i_m in the network, identifiability of the whole network is guaranteed. From now on, we will assume, without loss of generality, that the LFN has one source and one sink.

A. Some fundamental examples and local identifiability

Let us start by providing a simple example where, depending on the choice of the nonlinearity f, the network can be identifiable by exciting sources and measuring sinks.

Example 2: Consider a path graph with 3 nodes and edge set $\mathcal{E} = \{(2,1),(3,2)\}$ and let $\mathcal{N}^e = \{1\}$ and $\mathcal{N}^m = \{3\}$, i.e., we can only excite the source and measure the sink. If we denote by $x = u_1^{k-3}$ the input of the source 1, we obtain that the measured function of node 3 is given by

$$F(x) = f(w_{32} f(w_{21} x)) \,. \tag{7}$$

Let $f(x) = a_1 x + a_2 x^2$ for some a_1, a_2 in $\mathbb{R}$. Then,

$$\begin{aligned} F(x) =& w_{32} w_{21} a_1^2 x + w_{32} a_1 a_2 (1 + a_1 w_{32}) w_{21}^2 x^2 \\ &+ 2 w_{32}^2 w_{21}^3 a_1 a_2^2 x^3 + w_{32}^2 w_{21}^4 a_2^2 x^4 \,. \end{aligned}$$

Observe that, for $a_1 = 0$ and $a_2 \neq 0$, we obtain that $F = \tilde{F}$ if and only if $w_{32}^2 w_{21}^4 = \tilde{w}_{32}^2 \tilde{w}_{21}^4$. Then, for $\tilde{w}_{21} = \frac{1}{\gamma} w_{21}$ and $\tilde{w}_{32} = \gamma^2 w_{32}$ with $\gamma \notin \{0,1\}$, we obtain $F = \tilde{F}$ and $W \neq \tilde{W}$, which implies that the network is not identifiable for $f(x) = a_2 x^2$ with $a_2 \neq 0$. We obtain the same result for $a_1 \neq 0$ and $a_2 = 0$. If instead $a_i \neq 0$ for i in $\{1,2\}$, by equating $F = \tilde{F}$, we obtain 4 conditions, that is,

$$F = \tilde{F} \Leftrightarrow \begin{cases} w_{32} w_{21} = \tilde{w}_{32} \tilde{w}_{21} \\ w_{32} w_{21} (1 + w_{32}) w_{21} = \tilde{w}_{32} \tilde{w}_{21} (1 + \tilde{w}_{32}) \tilde{w}_{21} \\ w_{32}^2 w_{21}^3 = \tilde{w}_{32}^2 \tilde{w}_{21}^3 \\ w_{32}^2 w_{21}^4 = \tilde{w}_{32}^2 \tilde{w}_{21}^4 \end{cases}$$

If we substitute the third equation in the fourth equation we obtain $w_{21} = \tilde{w}_{21}$ and therefore $w_{32} = \tilde{w}_{32}$. Then, the only solution is $W = \tilde{W}$ and the network is identifiable.

In Ex. 2, we observed that, when $f(x) = a_1 x + a_2 x^2$, $a_i \neq 0$, for i in $\{1,2\}$, a path graph with 3 nodes is identifiable by only exciting the source and measuring the sink. Despite being a very simple example, this observation is fundamental as it is in contrast with the findings in the literature of identifiability of network systems. Indeed, it was proved in [3] for linear systems and in [10] for nonlinear systems that a necessary condition for identifiability is to excite and/or measure every node of the graph. In our setting, the presence of the nonlinearity on the nodes allows to identify paths where some nodes are neither excited or measured, as the superposition principle does not further apply. On the other hand, the linear dynamics on the edges provide enough structure to identify the network.

Example 3: Let us consider a graph with $n = 4$ nodes and edge set as in Fig. 2. Assume that $\mathcal{N}^e = \{1\}$ and $\mathcal{N}^m = \{4\}$. In this case, the measured function of node 4 is given by

$$F(x) = f(w_{42} f(w_{21} x) + w_{43} f(w_{31} x))$$

where $x = u_1^{k-3}$ denotes the input of node 1. Observe that, for every weight matrix W, if we set $\tilde{w}_{21} = w_{31}$, $\tilde{w}_{31} = w_{21}$, $\tilde{w}_{43} = w_{42}$ and $\tilde{w}_{42} = w_{43}$ we have $W \neq \tilde{W}$ and $F = \tilde{F}$. In words, we can at most identify edges up to node permutations. Therefore, in the general setting, the network is not identifiable. Let us now make a second fundamental remark: in the special case when $w_{21} = w_{31}$, the output of node 4 becomes $F(x) = f((w_{42} + w_{43}) f(w_{21} x))$. This implies that, for some particular weight matrices, we can at most identify $w_{42} + w_{43}$ and therefore the network is not identifiable.

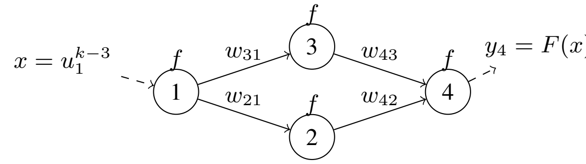

Fig. 2. Graph considered in Example 3

According to Ex. 3, the identifiability of the graph in Fig. 2 cannot be guaranteed unless we introduce more hypotheses and/or some relaxations. This problem has been studied in the context of identifiability of feed-forward neural networks [18], [19]. In their setting, identifiability is studied up to an equivalence class. For instance, one could define that two weight matrices W and $\tilde{W}$ are equivalent if W can be transformed into $\tilde{W}$ by means of a finite number of node-permutations. In this work, instead, we consider the weaker notion of *local* identifiability, which corresponds to identifiability provided that $\tilde{W}$ is sufficiently close to W (i.e., at distance less than ε, for some $\varepsilon > 0$). This relaxation is not novel in the literature (e.g., [5]) and allows us to find results for broader class of functions and topologies.

Definition 2 (Local identifiability): The network $\mathcal{G} = (\mathcal{N}, \mathcal{E}, W)$ is *locally* identifiable in a class $\mathcal{F}$ with excitations $\mathcal{N}^e$ and measurement $\mathcal{N}^m$ if, for any given f in $\mathcal{F}$, there exists $\varepsilon > 0$ such that for any $\tilde{W}$ consistent with the graph satisfying $\|W - \tilde{W}\| < \varepsilon$ there holds

$$F(\mathcal{N}^m) = \tilde{F}(\mathcal{N}^m) \quad \Rightarrow \quad W = \tilde{W} \,.$$

Furthermore, we address the second remark in Ex. 3 by considering a *generic* notion of identifiability. The notion of generic identifiability has been already introduced for transfer functions in several works, including [3], [5]. A property holds generically, or for almost all variables, if it holds for all variables except a zero measure set in $\mathbb{R}^N$ where N is the number of parameters of each variable. Since the edge set is known, the number of parameters of the weight matrix W is $N = |\mathcal{E}|$, i.e., we shall study identifiability of W for almost all choices of the elements w_{ij} that are not known to be zero. We have the following definition.

Definition 3: Given a graph $\mathcal{G}$ with sets $\mathcal{N}^e$ and $\mathcal{N}^m$ of excited and measured nodes and a fixed nonlinearity f, we say that the network is W-generically (locally) identifiable if it is (locally) identifiable for all W consistent with $\mathcal{E}$ except possibly those lying on a zero measure set in $\mathbb{R}^{|\mathcal{E}|}$.

In the remainder of the paper, we consider a second type of genericity: genericity in the class of functions $\mathcal{F}$. We recall that genericity typically refers to *all variables except possibly* *some lying on a lower dimensional space/zero measure set*. In infinite dimension, these notions present some challenges. Here, we restrict the attention to the set of analytic functions, denoted with $\mathcal{C}^{\omega}$, and we make use of the series of the MacLaurin coefficients (i.e., Taylor coefficients in zero) to describe the function. We then define genericity by applying the zero measure set/lower dimensional subspace to a *subset of MacLaurin coefficients*.

Definition 4: Given a class of functions $\mathcal{F} \subseteq C^{\omega}$, a property holds *$f$-generically in $\mathcal{F}$*, or *for almost all functions* f in $\mathcal{F}$, if there exists $M > 0$ such that the property holds for all functions f in $\mathcal{F}$ except possibly a subset of functions whose first M MacLaurin coefficients all lie on a zero measure set in $\mathbb{R}^M$.

Example 4: Consider a path graph with 3 nodes (see Ex. 2) and let f be analytic with MacLaurin coefficients $a = \{a_k\}_{k\in\mathbb{N}}$. We assume $f(0) = 0$, i.e., we let $a_0 = 0$. Since f is analytic, the measurement F is also analytic. If we substitute f in (7), we obtain that, for x sufficiently small,

$$F(x) = a_1^2 w_{21} w_{32} x + a_2 a_1 w_{21}^2 w_{32}(1 + a_1 w_{32})x^2 + o(x^2)\,. \tag{8}$$

Following a similar reasoning as in Ex. 2, we obtain that the network is identifiable for all f whose first two MacLaurin coefficients satisfy $a_1 \neq 0$ and $a_2 \neq 0$. Observe that the set $E = \{(a_1, a_2) \in \mathbb{R}^2 : a_1 a_2 = 0\}$ is a zero measure set of $\mathbb{R}^M$ where $M = 2$. The network is then identifiable for almost all analytic functions satisfying $f(0) = 0$.

Our main result considers both forms of genericity introduced in Section II-A, according to the following definition.

Definition 5: Given a graph $\mathcal{G}$ with sets $\mathcal{N}^e$ and $\mathcal{N}^m$ of excited and measured nodes and a class of functions $\mathcal{F} \subseteq C^{\omega}$, we say that the network is generically (locally) identifiable in the class $\mathcal{F}$ if it is f-generically W-generically (locally) identifiable in $\mathcal{F}$, that is, if it is W-generically (locally) identifiable for almost all functions in $\mathcal{F}$.

III. Main result

We study local identifiability when the class of functions considered is made of all analytic functions that cross the origin, i.e.,

$$\mathcal{F}_Z := \{f : \mathbb{R} \to \mathbb{R} \mid f \text{ analytic in } \mathbb{R}\,,\ f(0) = 0\}\,. \tag{9}$$

The reason for the assumption $f(0) = 0$ will be clarified soon. We now state our main result, which is proved in Section IV.

Theorem 1: Let $\mathcal{G}$ be a fully-connected LFN and let $\mathcal{F}_Z$ be as in (9). Then, the network $\mathcal{G}$ is generically locally identifiable in the class $\mathcal{F}_Z$ by exciting the sources and measuring the sinks.

According to Theorem 1, fully-connected feed-forward networks are generically locally identifiable by only exciting sources and measuring sinks for almost all analytic functions that are zero in zero. This is in contrast with the results on identifiability of network systems where measuring/exciting each node is proved to be a necessary condition [3], [10]. It also generalizes previous literature on artificial neural networks where the problem (with biases) was studied either for shallow networks or for deep networks for a class of sigmoidal functions [23], [24] or ReLU activation function [21]. We discuss the main assumptions of Theorem 1

- In this preliminary work, we consider fully-connected LFNs. The layered structure helps when defining the function $F(x)$ starting from measurements, while the assumption of fully-connectedness is used in the last step of the proof, in the analysis of the exponential example (see Section IV-C). Current work includes extending Theorem 1 to directed acyclic graphs. Also, we aim to extend the result by allowing the activation functions to differ between nodes.
- The result holds for analytic functions. As pointed out in Definition 4, we need to consider analytic functions to properly define genericity in the space of functions. Future work includes extensions to piecewise analytic functions, such as ReLU.
- We assume that $f(0) = 0$ and that there are no biases. These are necessary conditions to prove that every MacLaurin coefficient of the measurement F depends on a finite number of coefficients of f (see Ex. 4 and Remark 3), which is fundamental to prove genericity in f. Relaxing these two assumptions requires the use of different techniques but it is definitely of interest.
- As discussed in Section II-A, we study local identifiability. In Ex. 3, we observed that global identifiability cannot be achieved even in simple network structures. Local identifiability is also a fundamental hypothesis in our results as it can be studied in terms of local injectivity and full-rankness properties as shown in the following example.

Example 4 (continued): Consider the same setting as in Ex. 4. Since f is analytic, we have, for x sufficiently small,

$$F(x) = \sum_k A_k(W, a)x^k$$

where, according to (8),

$$\begin{cases} A_1(W, a) = & a_1^2 w_{21} w_{32} \\ A_2(W, a) = & a_2 a_1 w_{21}^2 w_{32}(1 + a_1 w_{32}) \end{cases}$$

We can study local identifiability in terms of locally injectivity of $A|_I$ in W, where $I = \{1, 2\}$. Observe that the coefficients A_k, for k in $\{1, 2\}$, are polynomial functions of the weights W and *only* the first two coefficients (a_1, a_2) of f. We remark that every coefficient A_k depends on a *finite* number of coefficients and this is true since, by assumption, there are no biases and $f(0) = 0$. Local injectivity of $A|_I$ can be studied by computing the rank of its Jacobian in W, denoted with $\nabla_w A|_I(W, a)$. By imposing

$$\det(\nabla_w A|_I(W, a)) = -a_1^3 a_2 w_{21}^2 w_{32} \neq 0\,.$$

we obtain the same result as before, i.e., the network is generically locally identifiable for all f whose first two MacLaurin coefficients satisfy $a_1 \neq 0$ and $a_2 \neq 0$. The network is indeed identifiable for almost all functions in $\mathcal{F}_Z$.

IV. Proof of the main result

The proof of Theorem 1, is divided in three main steps. In Section IV-A, we prove that, for a fixed analytic f, local identifiability is a generic property in W, that is, it either holds for almost all W or for no W. In Section IV-B we further prove that local identifiability is a generic property for f in $\mathcal{F}_Z$. By combining the two results, we obtain that local identifiability either holds for almost all W for almost all functions f or for no W and no f. It is then sufficient to show local identifiability for one f^* in $\mathcal{F}_Z$ and one W^* consistent with $\mathcal{E}$ to obtain generic local identifiability of the network $\mathcal{E}$ in the class $\mathcal{F}_Z$. Section IV-C is devoted to prove local identifiability in fully-connected FNs for $f^*(x) = e^x - 1$, thus leading to the main result.

A. Genericity in the weight matrix

We start by proving that, for f in $\mathcal{C}^\omega$, local identifiability is a W-generic property. For f analytic with Maclaurin series coefficients $a = \{a_k\}_{k\in\mathbb{N}}$, the measurement F is analytic and for x sufficiently small it must take the form

$$F(x) = \sum_k A_k(W,a)x^k \tag{10}$$

for some coefficients $A_k(W,a)$ that depend on the weights W and the coefficients a. Let us denote with $A(W,a)$ the operator that maps the weights W and the MacLaurin coefficient of f to the MacLaurin coefficients of F. Then, a network $\mathcal{G}$ is locally identifiable for f in $\mathcal{C}^\omega$ if, for all $\tilde{W}$ sufficiently close to W, it holds

$$A(W,a) = A(\tilde{W},a) \quad \Rightarrow \quad W = \tilde{W}\,. \tag{11}$$

We start by formulating the local identifiability question in terms of local injectivity.

Definition 6: A function $g : \mathbb{R}^N \to \mathbb{R}^M$ is *locally injective* at x if there exists $\epsilon > 0$ such that, for all $\tilde{x}$ satisfying $\|x - \tilde{x}\| < \epsilon$, there holds

$$g(x) = g(\tilde{x}) \quad \Rightarrow \quad x = \tilde{x}$$

To use local injectivity, we need to consider finite restrictions of the operator A in (10). For $I \subset \mathbb{N}$ such that $|I| = M > 0$, we shall denote with $A|_I$ the restriction of A on I. Local identifiability of the network and local injectivity of $A|_I$ for I finite are linked in the following way.

Proposition 1: Consider a LFN with one source and one sink and let f in $\mathcal{C}^\omega$. Then, the two facts are equivalent

(i) the network is (locally) identifiable in W by exciting the source and measuring the sink;
(ii) there exists a finite subset $I \subset \mathbb{N}$ such that $A|_I(W,a)$ is (locally) injective in W.

Proof: For f in $\mathcal{C}^\omega$, (i) is equivalent to (ii) if and only if (11) is equivalent to (ii). The implication $(ii) \Rightarrow (i)$ is then trivial since I is just a subset. The opposite implication $(i) \Rightarrow (ii)$ is true for Hilbert's basis theorem [25], from which it follows that a locus-set of a collection of polynomials is the locus of finitely many polynomials, i.e. the intersection of finitely many hypersurfaces. ■

According to Proposition 1, the network is (locally) identifiable if and only if $\exists I$, I finite, such that $A|_I$ is (locally) injective in W. We then focus on $A|_I$, for I finite, and we study under which conditions $A|_I$ is local injective in W. In the following Lemma, we prove that, for every given f in $\mathcal{C}^\omega$ and I finite, local injectivity of $A|_I$ is a generic property in W, that is, it either holds for almost all weights or for no weights. Moreover, we show a direct link between local injectivity of $A|_I$ in W and full-rankness of its Jacobian, that we denote with $\nabla_w A|_I(W,a)$.

Lemma 1: Let $\mathcal{G}$ be a LFN with one source and one sink and let f in $\mathcal{C}^\omega$ with Maclaurin series coefficients $a = \{a_k\}_k$. For every $I \subset \mathbb{N}$, I finite, exactly one of the two following holds:

(i) $A|_I$ is *W-generically* locally injective in W and $\operatorname{rank} \nabla_w A|_I(W,a) = |\mathcal{E}|$ for *almost all* W
(ii) $A|_I$ is locally injective for *no* W and $\operatorname{rank} \nabla_w A|_I(W,a) = |\mathcal{E}|$ for *no* W .

Remark 1: The proof of Lemma 1, which can be found in Appendix V-A, is divided in two parts. First, we link local injectivity of $A|_I$ with full-rankness of its Jacobian. Their relationship is not immediate to show and the proof is based on the results in [5], [7]. The genericity of the property is then derived starting from the genericity of full-rankness of the Jacobian of an analytic function (see Lemma 3.2 in [7]), which follows from the fact that determinants are either nonzero for almost all variables, or for no variables.

If we combine Proposition 1 and Lemma 1, we obtain that local identifiability is a generic property in W, that is, it either holds for almost all weights W or for no W.

Proposition 2: Consider a LFN with one source and one sink and let f in $\mathcal{C}^\omega$. Exactly one of the following holds:

(i) the network is W-generically locally identifiable by exciting the source and measuring the sink;
(ii) the network is locally identifiable by exciting the source and measuring the sink for no W.

Proof: From Lemma 1, it follows that exactly one of the following holds: either A_I is locally injective for no W for all I finite, or there exists I finite such that A_I is locally injective for almost all W. If we combine this with Proposition 1, we obtain the result. ■

Remark 2: According to Proposition 2, for every given f in $\mathcal{C}^\omega$, if the network is locally identifiable for *one* weight matrix W^* consistent with the graph, then, the network is locally identifiable for almost all W.

B. Genericity in the functions

We now restrict the attention to analytic functions that are zero in zero and we prove that local identifiability by exciting sources and measuring sinks is a f-generic property in the class $\mathcal{F}_Z$, that is, it holds for almost all functions (in the sense of Definition 4) or for no functions. Let us start by stating Lemma 2, which is proved in Appendix V-B.

Lemma 2: Consider a LFN with one source and one sink and let f in $\mathcal{F}_Z$ with Maclaurin series coefficients $a = \{a_k\}_{k\in\mathbb{N}_+}$. Then, each coefficient in (10) depends only on a finite number of variables and

$$A_k(W,a) = A_k(W,a_1,\dots,a_k) \tag{12}$$

where, for k in $\mathbb{N}_+$, $A_k : \mathbb{R}^{|\mathcal{E}|}\times\mathbb{R}^k \to \mathbb{R}$ are polynomial functions of the weights W and the first k coefficients $(a_1,\dots,a_k)$.

According to Lemma 2, any restriction $A|_I : \mathbb{R}^{n\times n}\times\mathbb{R}^M \to \mathbb{R}^M$ with $I=\{1,\dots,M\}$, is a polynomial function of W and a *finite* number of coefficients $(a_1,\dots,a_M)$.

Remark 3: The conclusion of Lemma 2 is no longer true when $f(0)\neq 0$. Consider a path graph with 3 nodes as in Ex. 2. and let f in $\mathcal{F}$ have Maclaurin series coefficients $a=\{a_k\}_{k\in\mathbb{N}}$. We then obtain that for x sufficiently small

$$F(x) = \sum_{k'} a_{k'}\big(w_{32}\sum_k a_k w_{21}^k x^k\big)^{k'} =$$

$$\sum_k a_k(a_0 w_{32})^k + \big(a_1 w_{32}w_{21}\sum_{k>0} k a_k (a_0 w_{32})^{k-1}\big)x + o(x)$$

Observe that the first coefficient of F is given by $A_0(W,a) = \sum_k a_k(a_0w_{32})^k$ and depends on all the series of coefficients of f, and so does $A_1(W,a)$.

This finite reduction allows us to prove the following proposition.

Proposition 3: Consider a LFN with one source and one sink that are excited and measured, respectively. Then, exactly one of the following holds:

(i) the network is generically locally identifiable in $\mathcal{F}_Z$;
(ii) the network is locally identifiable for no W and no f in $\mathcal{F}_Z$.

Proof: We remark that (i) and (ii) are mutually exclusive. Furthermore, if (i) holds then (ii) does not hold. We then prove the statement by showing that if (ii) does not hold then (i) holds, i.e., we assume that the network is locally identifiable for some W^* and f^* and we show that this implies that the network is locally identifiabile for almost all W and almost all f.

Let a^* denote the MacLaurin coefficients of f^*. By Proposition 1, if the network is locally identifiable for some W^*, then there exists a finite subset $I\subset\mathbb{N}$ such that $A|_I$ is locally injective in W^*. By Lemma 1, this implies that $A|_I$ is W-generically locally injective in W and $\operatorname{rank}\nabla_w A|_I(W,a^*) = |\mathcal{E}|$ for almost all W. In particular, we have that $\operatorname{rank}\nabla_w A|_I(W^*,a^*) = |\mathcal{E}|$. We now want to study the genericity in both W and a. Without loss of generality, let us assume that $I=\{1,\dots,M\}$, for some $M>0$. By Lemma 2, we have that $A|_I$ depends only on the first M coefficients of f and therefore $\operatorname{rank}\nabla_w A|_I(W,a) = \operatorname{rank}\nabla_w A|_I(W,a|_I)$. Let us denote $v=a|_I$, where v is in $\mathbb{R}^M$. Then, by the genericity of the full-rankness of the Jacobian of an analytic function (see Lemma 3.2 in [7]), we have that either (i) $\operatorname{rank}\nabla_w A|_I(W,v) = |\mathcal{E}|$ for almost all W consistent with $\mathcal{E}$ and v in $\mathbb{R}^M$ or (ii) $\operatorname{rank}\nabla_w A|_I(W,v) = |\mathcal{E}|$ for no W and no v. Recall that we have $\operatorname{rank}\nabla_w A|_I(W^*,v^*) = |\mathcal{E}|$, for $v^*=a^*|_I$. Then, it must be case (i), that is, $\operatorname{rank}\nabla_w A|_I(W,v) = |\mathcal{E}|$ for almost all W for almost all v in $\mathbb{R}^M$. By Lemma 1, we have $A|_I$ is W-generically locally injective in W for almost all $v\in\mathbb{R}^M$, which implies, by Proposition 1 and Definition 4, that the network is W-generically locally identifiable for almost all functions in $\mathcal{F}_Z$. This concludes the proof. ■

Remark 4: According to Proposition 3, if the network is locally identifiable for *one* weight matrix W^* consistent with the graph for *one* f^* in $\mathcal{F}_Z$, then, the network is locally identifiable for almost all W for almost all f.

C. Exponential example

In this preliminary work, we consider fully-connected feed-forward networks of length L with $\mathcal{E}=\mathcal{E}_F$ (see (5)). In order to use Proposition 3 to prove Theorem 1, we need to find f^* in $\mathcal{F}_Z$ and W^* consistent with $\mathcal{E}$ such that $\mathcal{G}$ is locally identifiable by exciting the source and measuring the sink. Here, we consider $f^*(x)=e^x-1$. This function is of particular interest since we can study identifiability in the limit as $x\to+\infty$. As we shall see, when all weights are positive, different and ordered, we can prove identifiability of the edges through the limit of functions associated to F.

We make use of the following notation. We define L submatrices of W, that is, for l in $\{1,\dots,L\}$, we let $W^l\in\mathbb{R}^{|\mathcal{N}^{l-1}|\times|\mathcal{N}^l|}$ be such that $W^l = W[\mathcal{N}^{l-1};\mathcal{N}^l]$. We rename nodes in $\mathcal{N}^l=\{1,\dots,|\mathcal{N}^l|\}$ for $l=0,\dots L$ and weights in W^l accordingly. In words, each submatrix W^l contains all the weights connecting a layer and the following one and the quantity w_{ij}^l denotes the weight of the edge that connects the node i in the layer $l-1$ to node j of layer l. We also denote the output of node i in layer l with F_i^l. Since $f^*(x)=e^x-1$, we have that, according to (1),

$$F_i^l(x) = \exp\big(\sum_j w_{ij}^l F_j^{l-1}(x)\big) - 1 \tag{13}$$

where we recall that $F_1^0(x)=x$. For $l>0$, we denote with $\log_l$ the function that applies the logarithm l times. Finally, we define

$$G_{ij}^l = \frac{\log\big(\log_{L-l}(F) - \sum_{i'<i} w_{1i'}^{l+1}F_{i'}^l\big) - \sum_{j'<j} w_{ij'}^l F_{j'}^{l-1}}{F_j} \tag{14}$$

We assume w.l.g. that W^* has nonnegative entries, that is, $w_{ij}^*>0$ for all i,j such that $(i,j)\in\mathcal{E}$. Furthermore, we assume that the entries W^* are ordered in such a way that, for each layer l and for each node j in layer $l-1$, we have that $w_{1j}^l > w_{2j}^l > \dots > w_{|N_l|j}^l$, that is,

$$W^* \in \mathcal{W}_{\mathrm{o}} := \{W\in\mathbb{R}_+^{n\times n} \mid w_{i'j}^l > w_{ij}^l\,,\ \forall l\in\mathcal{L}\,,\ j\in\mathcal{N}^{l-1}\,,\ i',i\in\mathcal{N}^l, i'<i\}\,. \tag{15}$$

The following lemma, proved in Appendix V-C, states that, for W in $\mathcal{W}_{\mathrm{o}}$, the limit of G_{ij}^l for $x\to+\infty$ is w_{ij}^l.

Lemma 3: For W in $\mathcal{W}_{\mathrm{o}}$, it holds that, for every layer l, node i in $\mathcal{N}^l$ and j in $\mathcal{N}^{l-1}$,

$$\lim_{x\to+\infty} G_{ij}^l(x) = w_{ij}^l\,. \tag{16}$$

Lemma 3 gives us a way to prove identifiability of the weight w_{ij}^l through the quantities in G_{ij}^l. Indeed, according to Lemma 3, it holds that, for W, $\tilde{W}$ in $\mathcal{W}_{\rm o}$,

$$G_{ij}^l = \tilde{G}_{ij}^l \quad \Rightarrow \quad w_{ij}^l = \tilde{w}_{ij}^l \,. \tag{17}$$

If we combine (17) with the definitions of G_{ij} in (14) and F_i in (13), we obtain the following recursive argument. Let $W, \tilde{W}$ in $\mathcal{W}_{\rm o}$ be such that $F = \tilde{F}$. Then, for every layer l, node i in $\mathcal{N}^l$ and j in $\mathcal{N}^{l-1}$, there holds, for $i = 1$,

$$\begin{cases} w_{1j'}^l = \tilde{w}_{1j'}^l \,, & \forall j' < j \\ w_{j'k}^{l-1} = \tilde{w}_{j'k}^{l-1} \,, & \forall j' \leq j \,, \forall k \\ w_{kk'}^{l'} = \tilde{w}_{kk'}^{l'} \,, & \forall l' < l-1 \,, \forall k, k' \end{cases} \quad \Rightarrow \quad w_{1j}^l = \tilde{w}_{1j}^l \tag{18}$$

while, for $i > 1$,

$$\begin{cases} w_{1i'}^{l+1} = \tilde{w}_{1i'}^{l+1} \,, w_{i'j'}^l = \tilde{w}_{i'j'}^l & \forall i' < i \, \forall j' \\ w_{ij'}^l = \tilde{w}_{ij'}^l \,, & \forall j' < j \\ w_{kk'}^{l'} = \tilde{w}_{kk'}^{l'} \,, & \forall l' < l \, \forall k, k' \end{cases} \Rightarrow w_{ij}^l = \tilde{w}_{ij}^l \tag{19}$$

We then have the following.

Proposition 4: Let $\mathcal{G}$ be a fully-connected feed-forward network with $W^* \in \mathcal{W}_{\rm o}$ and let $f^*(x) = e^x - 1$. Then, $\mathcal{G}$ is identifiable by exciting the source and measuring the sink.

Proof: Assume by contradiction that for some $W, \tilde{W} \in \mathcal{W}_{\rm o}$ we have that $F = \tilde{F}$ and $W \neq \tilde{W}$. Then, we can compute

$$\begin{aligned} \bar{l} &= \max\{l : w_{kk'}^{l'} = \tilde{w}_{kk'}^{l'} \,, \ \forall l' < l \,, \forall k, k'\} \\ \bar{i} &= \max\{i : w_{i'j}^{\bar{l}} = \tilde{w}_{i'j}^{\bar{l}} \,, \ \forall i' < i \,, \forall j\} \\ \bar{j} &= \max\{j : w_{\bar{i}j'}^{\bar{l}} = \tilde{w}_{\bar{i}j'}^{\bar{l}} \,, \ \forall j' < j\} \end{aligned}$$

If $\bar{l} < L$, we can further define

$$\bar{i}^{\rm next} = \max\{i \leq \bar{i} : w_{1i'}^{\bar{l}+1} = \tilde{w}_{1i'}^{\bar{l}+1} \,, \ \forall i' < i\}$$

Then, we have the following options,

- $\bar{i}^{\rm next} < \bar{i} \overset{(18)}{\Rightarrow} w_{1\bar{i}^{\rm next}}^{\bar{l}+1} = \tilde{w}_{1\bar{i}^{\rm next}}^{\bar{l}+1}$
- $\bar{i}^{\rm next} = \bar{i} \overset{(19)}{\Rightarrow} w_{\bar{i}\bar{j}}^{\bar{l}} = \tilde{w}_{\bar{i}\bar{j}}^{\bar{l}}$

If $\bar{l} = L$, then $\bar{i} = 1$ and $w_{\bar{i}\bar{j}}^{\bar{l}} = \tilde{w}_{\bar{i}\bar{j}}^{\bar{l}}$. In all cases, we reach an absurd. Therefore, we have $W = \tilde{W}$. ∎

Proposition 4 combined with Proposition 3 leads to Theorem 1. This concludes the proof.

V. Conclusions

We considered local generic identifiability of networks characterized by linear edge dynamics and nonlinear node dynamics. Given known graph topology and nonlinearity, the objective is to identify the weight matrix of the graph. We proved that fully-connected layered feed-forward networks are locally generically identifiable in the class of analytic functions that are zero in zero by exciting only sources and measuring only sinks. In the full version of this work, we aim to generalize the results to more general network structures and heterogeneous nonlinearities. Future research directions include broadening the space of nonlinearities, incorporating offsets and studying conditions for global identifiability and/or identifiability up to an equivalence class.

References

[1] F. Bullo, *Lectures on network systems*, vol. 1. CreateSpace, 2018.

[2] J. Gonçalves and S. Warnick, "Necessary and sufficient conditions for dynamical structure reconstruction of lti networks," *IEEE Transactions on Automatic Control*, vol. 53, no. 7, pp. 1670–1674, 2008.

[3] J. M. Hendrickx, M. Gevers, and A. S. Bazanella, "Identifiability of dynamical networks with partial node measurements," *IEEE Transactions on Automatic Control*, vol. 64, no. 6, pp. 2240–2253, 2018.

[4] H. J. Van Waarde, P. Tesi, and M. K. Camlibel, "Necessary and sufficient topological conditions for identifiability of dynamical networks," *IEEE Transactions on Automatic Control*, vol. 65, no. 11, pp. 4525–4537, 2019.

[5] A. Legat and J. M. Hendrickx, "Local network identifiability with partial excitation and measurement," in *2020 59th IEEE Conference on Decision and Control (CDC)*, pp. 4342–4347, IEEE, 2020.

[6] X. Cheng, S. Shi, I. Lestas, and P. M. Van den Hof, "A necessary condition for network identifiability with partial excitation and measurement," *IEEE Transactions on Automatic Control*, vol. 68, no. 11, pp. 6820–6827, 2023.

[7] A. Legat and J. M. Hendrickx, "Identifiability in networked systems: Algebraic and combinatorial conditions," 2024. Preprint on webpage at `https://perso.uclouvain.be/antoine.legat/_static/TAC24.pdf`.

[8] R. Vizuete and J. M. Hendrickx, "Nonlinear network identifiability: The static case," in *2023 62nd IEEE Conference on Decision and Control (CDC)*, pp. 443–448, IEEE, 2023.

[9] R. Vizuete and J. M. Hendrickx, "Nonlinear network identifiability with full excitations," 2024. Available at https://arxiv.org/pdf/2405.07636.

[10] R. Vizuete and J. M. Hendrickx, "Nonlinear identifiability of directed acyclic graphs with partial excitation and measurement," in *2024 IEEE 63rd Conference on Decision and Control (CDC)*, pp. 3153–3159, 2024.

[11] Y. D. Zhong and N. E. Leonard, "A continuous threshold model of cascade dynamics," in *2019 IEEE 58th Conference on Decision and Control (CDC)*, pp. 1704–1709, IEEE, 2019.

[12] S. Morris, "Contagion," *The Review of Economic Studies*, vol. 67, no. 1, pp. 57–78, 2000.

[13] A. Bizyaeva, A. Franci, and N. E. Leonard, "Nonlinear opinion dynamics with tunable sensitivity," *IEEE Transactions on Automatic Control*, vol. 68, no. 3, pp. 1415–1430, 2022.

[14] R. Gray, A. Franci, V. Srivastava, and N. E. Leonard, "Multiagent decision-making dynamics inspired by honeybees," *IEEE Transactions on Control of Network Systems*, vol. 5, no. 2, pp. 793–806, 2018.

[15] A. S. Lekamalage, "Identifiability of linear threshold decision making dynamics," 2024.

[16] D. Acemoglu, A. Ozdaglar, and A. Tahbaz-Salehi, "Networks, shocks, and systemic risk," tech. rep., National Bureau of Economic Research, 2015.

[17] M. A. Nielsen, *Neural networks and deep learning*, vol. 25. Determination press San Francisco, CA, USA, 2015.

[18] H. J. Sussmann, "Uniqueness of the weights for minimal feedforward nets with a given input-output map," *Neural networks*, vol. 5, no. 4, pp. 589–593, 1992.

[19] F. Albertini, E. D. Sontag, and V. Maillot, "Uniqueness of weights for neural networks," *Artificial neural networks for speech and vision*, vol. 3, 1993.

[20] C. Fefferman *et al.*, "Reconstructing a neural net from its output," *Revista Matemática Iberoamericana*, vol. 10, no. 3, pp. 507–556, 1994.

[21] J. Bona-Pellissier, F. Bachoc, and F. Malgouyres, "Parameter identifiability of a deep feedforward relu neural network," *Machine Learning*, vol. 112, no. 11, pp. 4431–4493, 2023.

[22] C. Fiedler, M. Fornasier, T. Klock, and M. Rauchensteiner, "Stable recovery of entangled weights: Towards robust identification of deep neural networks from minimal samples," *Applied and Computational Harmonic Analysis*, vol. 62, pp. 123–172, 2023.

[23] V. Vlačić and H. Bölcskei, "Affine symmetries and neural network identifiability," *Advances in Mathematics*, vol. 376, p. 107485, 2021.

[24] V. Vlačić and H. Bölcskei, "Neural network identifiability for a family of sigmoidal nonlinearities," *Constructive Approximation*, vol. 55, no. 1, pp. 173–224, 2022.

[25] D. Hilbert, "Ueber die theorie der algebraischen formen," *Mathematische annalen*, vol. 36, no. 4, pp. 473–534, 1890.

A. Proof of Lemma 1

The proof of Lemma 1 is adapted from the proof of Theorem 4.1 in [7] and make use of Proposition 3.1 in [7] (also Proposition 3.1 in [5]) and Lemma 3.2 in [7]. In order to use these results, we need to introduce the notion of coordinate-injectivity.

Definition 7 ([5]): A function $g : \mathbb{R}^N \to \mathbb{R}^M$ is *locally coordinate-injective* for coordinate e at $x \in \mathbb{R}^N$ if there exists $\epsilon > 0$ such that, for all $\tilde{x}$ satisfying $\|x - \tilde{x}\| < \epsilon$, there holds $g(x) = g(\tilde{x})$ implies $\Rightarrow \quad x_e = \tilde{x}_e$.

Proof: [Proof (Lemma 1)] For a and I fixed, let us define $g(W) := A|_I(W, a)$. Following [5], we denote with $\mathbf{e}_e$ the standard basis vector filled with zeros except 1 at the e-th entry. Then, according to Proposition 3.1 in [5], since $g : \mathbb{R}^{n\times n} \to \mathbb{R}^M$ is an analytic function where $\mathbb{R}^{n\times n}$ and $\mathbb{R}^M$ are smooth manifolds with finite dimension, and $\mathbb{R}^{n\times n}$ is open, then exactly one of the two following holds: (i) $\ker \nabla g(W) \perp \mathbf{e}_e$ for almost all W and g is locally coordinate-injective for e at almost all W, or (ii) $\ker\nabla g(W) \perp \mathbf{e}_e$ for almost no W and g is locally coordinate-injective for e at almost no W. The property $\ker \nabla g(W) \perp \mathbf{e}_e$ for each e is equivalent to $\ker \nabla g(W) = \{0\}$, and therefore implies $\operatorname{rank} \nabla g(W) = |\mathcal{E}|$ by the rank-nullity theorem. This implies that either (i) $\operatorname{rank} \nabla g(W) = |\mathcal{E}|$ for almost all W, or (ii) $\operatorname{rank} \nabla g(W) = |\mathcal{E}|$ for almost no W. On the other hand, from Lemma 3.2 in [7], we know that the full rankness of $\nabla g(W)$ is a generic property, that is, it either holds for all W or for *no* W. Therefore, we obtain that either (i) $\operatorname{rank}\nabla g(W) = |\mathcal{E}|$ for almost all W, or (ii) $\operatorname{rank}\nabla g(W) < |\mathcal{E}|$ for all W. In case (ii), this implies that for each W there is a direction e for which $\ker \nabla g(W) \not\perp \mathbf{e}_e$, since the rank is non-full. We apply this argument for each W, and obtain that in case (ii) g is locally injective for no W. This concludes the proof. ∎

B. Proof of Lemma 2

Recall that, in LFNs, the node set is divided in layers, that is, $\mathcal{N} = \mathcal{N}^0 \cup \cdots \cup \mathcal{N}^L$. We shall prove the statement by induction on the layers. Throughout the proof, we denote with F_j the output of node j in $\mathcal{N}$.

For $l = 0$, the subset $\mathcal{N}^0$ contains the single source of the network. Then, $F_1(x) = x$ and it can be rewritten as $F_1(x) = \sum_{k>0} A_k(W, a_1, \ldots, a_k)x^k$ where $A_1 = 1$ and $A_k = 0$ for all $k > 1$. Therefore, the statement holds for $l = 0$.

Consider now a layer l in $\mathcal{L}$ and assume that, for all j in $\mathcal{N}^{l-1}$, $F_j(x) = \sum_{k>0} A^j_k(W, a_1, \ldots, a_k)x^k$ where, for all k, A^j_k are polynomial functions that depend on the first k coefficients of f and the weights W. For a node i in $\mathcal{N}^l$, we have that, according to (1), it holds

$$
\begin{aligned}
F_i(x) &= f\big(\sum_j w_{ij} \sum_{k>0} A^j_k(W, a_1, \ldots, a_k)x^k\big) \\
&= f\Big(\sum_{k>0} \big(\sum_j w_{ij} A^j_k(W, a_1, \ldots, a_k)\big)x^k\Big) = f\big(\sum_{k>0} b_k x^k\big)
\end{aligned}
$$

where $b_k := b_k(W, a_1, \ldots, a_k) = \sum_j w_{ij} A^j_k(W, a_1, \ldots, a_n)$ is introduced to simplify the notation. Observe that, for each k, the coefficient b_k depends on f only through the first k coefficients $a_1, \ldots, a_k$. Let us now compute $f(x) = \sum_{k>0} a_k x^k$, focusing on the first M parameters, for some $M > 0$, thus obtaining

$$
\begin{aligned}
F_i(x) =& \sum_{k=1}^{M} a_k \big(\sum_{k'>0}^{M} b_{k'} x^{k'}\big)^k + \sum_{k>M} a_k \big(\sum_{k'>0}^{M} b_{k'} x^{k'}\big)^k = \\
\overset{(1)}{=}& \sum_{k=1}^{M} a_k \big(\sum_{k'=1}^{M} b_{k'} x^{k'} + \sum_{k'>M+1} b_{k'} x^{k'}\big)^k + \sum_{k>M} c_k x^k = \\
\overset{(2)}{=}& \sum_{k=1}^{M} a_k \big(\sum_{k'=1}^{M} b_{k'} x^{k'}\big)^k + \sum_{k>M} \tilde{c}_k x^k = \\
=& \sum_{k=1}^{M} A^i_k(a_1, \ldots, a_M, b_1, \ldots, b_M)x^k + \sum_{k>M} \bar{c}_k x^k
\end{aligned}
$$

where, for each $k' > M$, c_k, $\tilde{c}_k$ and $\bar{c}_k$ are polynomial functions of the coefficients a and b. We remark that (1) and (2) hold since $a_k = 0$. Indeed, under this assumption, all powers of degree $k' > M$ give rise to monomials of degree $k' > M$. If we combine these observations with the fact that b_k depends on W and a_k, we obtain that, for all $i \in \mathcal{N}^l$, $A^i_M(W, a) = A^i_M(W, a_1, \ldots, a_M)$. Since this reasoning applies for all $M > 0$, this concludes the proof.

C. Proof of Lemma 3

The proof of Lemma 3 requires the following lemma that provides insights on the behavior of F^l_i as $x \to \infty$.

Lemma 4: For every layer l and node i in $\mathcal{N}^l$, it holds $\lim_{x\to+\infty} F^l_i(x) = +\infty$ and, more in general, for every node j in $\mathcal{N}^{l-1}$,

$$
\sum_{k\geq j} w^l_{ik} F^{l-1}_k(x) = w^l_{ij} F^{l-1}_j(x)(1 + o(1)), \quad x \to +\infty . \tag{20}
$$

Proof: We shall prove the argument by induction. Recall that $\mathcal{N}^0 = \{1\}$. Therefore, for $l = 1$ and i in $\mathcal{N}^1$, we have $\sum_{k\geq1} w^1_{ik} F^0_k(x) = w^1_{i1} x$. Furthermore, by definition of $F^l_i(x)$ we have that $F^l_i \to +\infty$ for $x \to +\infty$. Then, the statement holds for $l = 0$. We now prove that, if the argument holds for l, then it holds for $l+1$. Indeed, for $x \to +\infty$, we have

$$
\begin{aligned}
\sum_{k\geq j} w^{l+1}_{ik} F^l_k(x) =& \sum_{k\geq j} w^{l+1}_{ik} (\exp\big(\sum_{k'\geq1} w^l_{kk'} F^{l-1}_{k'}(x)\big) - 1) \\
\overset{(1)}{=}& \sum_{k\geq j} w^{l+1}_{ik} \exp\big(w^l_{k1} F^{l-1}_1(x)(1 + o(1))\big) \\
=& w^{l+1}_{ij} \exp\big(w^l_{j1} F^{l-1}_1(x)(1 + o(1))\big) \\
& \Big(1 + \sum_{k>j} \frac{w^{l+1}_{ik}}{w^{l+1}_{ij}} \exp\big((w^l_{k1} - w^l_{j1} + o(1))F^{l-1}_1(x)\big)\Big) \\
\overset{(2)}{=}& w^{l+1}_{ij} F^l_j(x)(1 + o(1)) ,
\end{aligned}
$$

where (1) holds true for the induction argument, while (2) holds true since $w^l_{j1} > w^l_{k1} > 0$ for all $k > j$ (see (15)) and $F^{l-1}_1 \to +\infty$ for $x \to +\infty$ (induction argument). Since

$F_j^l \to +\infty$ for $x \to +\infty$ and $w_{ij}^{l+1} > 0$, then $F_i^{l+1} \to +\infty$ for $x \to +\infty$. This concludes the proof. ■

Then, according to (13) and Lemma 4, for $x \to +\infty$, there holds

$$\begin{aligned}
&\log_{L-l}(F(x)) = \log_{L-l}\big(\exp(\sum_j w_{ij}^L F_j^{L-1}(x)) - 1\big)\\
&= \log_{L-l-1}\big(w_{11}^L F_1^{L-1}(x)(1+o(1))\big)\\
&= \log_{L-l-2}\big(\log(w_{11}^L) + \log(F_1^{L-1}(x)) + \log(1+o(1))\big)\\
&\sim \log_{L-l-2}\big(w_{11}^{L-1} F_1^{L-2}(x)(1+o(1))\big) = \cdots =\\
&\sim \log(w_{11}^{l+2}) + \sum_j w_{1j}^{l+1} F_j(x) + \log(1+o(1))\,.
\end{aligned}$$

Therefore, $G_{ij}^l(x) \sim \frac{\log\left(\sum_{j\geq i} w_{1j}^{l+1} F_j(x)\right) - \sum_{k<j} w_{ik}^l F_k^{l-1}(x)}{F_j(x)}$ for $x \to +\infty$. The result then holds since

$$\begin{aligned}
&\lim_{x\to+\infty} \frac{\log\left(\sum_{j\geq i} w_{1j}^{l+1} F_j(x)\right) - \sum_{k<j} w_{ik}^l F_k^{l-1}(x)}{F_j(x)}\\
&= \lim_{x\to+\infty} \frac{\log(w_{1i}^{l+1}) + \sum_{k\geq j} w_{ik}^l F_k^{l-1}(x) + \log(1+o(1))}{F_j(x)}\\
&= \lim_{x\to+\infty} \frac{w_{ij}^l F_j(x)(1+o(1))}{F_j(x)} = w_{ij}^l\,.
\end{aligned}$$