2023 62nd IEEE Conference on Decision and Control (CDC)
December 13-15, 2023. Marina Bay Sands, Singapore

# Boolean Internal Structure Reconstruction from Collapsed Small-Scale Networks

Basi Teng[1,†], Yuxuan Zhao[2,†], Ye Yuan[2] and Jorge Goncalves[3,*]

*Abstract*— **Dynamic network reconstruction aims to infer network structure from input-output data. Dynamical structure functions (DSFs) have been introduced to represent structural information between observable nodes of linear time-invariant systems. However, reconstructing large-scale DSFs can be difficult since most existing methods do not scale. Instead of inferring large DSFs directly, an alternative approach is to reconstruct many small-scale DSFs that are easier to infer. Given a sparsity constraint on the network, this paper proposes a necessary and sufficient condition for perfect reconstruction of the Boolean network using collapsed small-scale networks. For sparse networks, such as gene regulatory networks, this method can significantly reduce time and computational costs of Boolean network inference for most links in the network, especially when using parallel computing.**

## I. Introduction

Dynamical network systems have two fundamental properties: their system dynamics and network structure. Both can be captured by dynamical structure functions (DSF) [1], [2]. Reconstruction of DSF has been the focus of considerable research especially in systems biology, such as inference of gene regulatory networks [3], [4] and organic reaction mechanism classification [5].

There are several methods to reconstruct DSFs like BINGO [6], dynGENIE3 [7], and many others [8]–[12], while some focus on reconstructing part of the network [13], [14]. However, reconstructing large-scale DSFs (as those shown in [3], [4]) can be difficult considering the costs of computation, time and accuracy. An alternative approach is to collapse the original network into many small sub-networks, which are easier to infer.

For sparse networks, this paper provides necessary and sufficient conditions that guarantee the inference of the Boolean network from many and small collapsed networks. The paper is organised as follows: the next section provides a motivating example, followed by the main results section. Finally, section IV contains an illustrative example.

## II. Motivating example

**Example 1.** Consider the directed graph in Fig. 1a and assume we are interested in inferring whether the link from node 2 to node 1 exists or not. While keeping nodes 1 and 2, three 3-node collapsed networks can be obtained from the original 5-node network (Figs. 1b-d). All three collapsed networks give a nonzero link from node 2 to 1 via collapsed nodes, given the impression that this link does exist. For example, when keeping nodes 1, 2 and 3 (Figs. 1b), nodes 2 to 1 link via the now hidden node 4. However, we see that the original network does not have this link.

Alternatively, consider collapsing the original network into three 4-node sub-networks (Fig.2b-d). One of the collapsed networks (Fig.2b) does not show a link from node 2 to 1, while the other two (Fig.2c-d) do. By a one-vote veto strategy we obtain the correct answer.

Example 1 shows that the order of the collapsed networks influences the inference results. In particular, in this example collapsing to 3-node networks is not enough for inference, and collapsing to 4-node networks yields the correct Boolean structure. This paper will formalise this result into necessary and sufficient conditions on the number of nodes of the collapsed network to guarantee the correct inference of the original Boolean network.

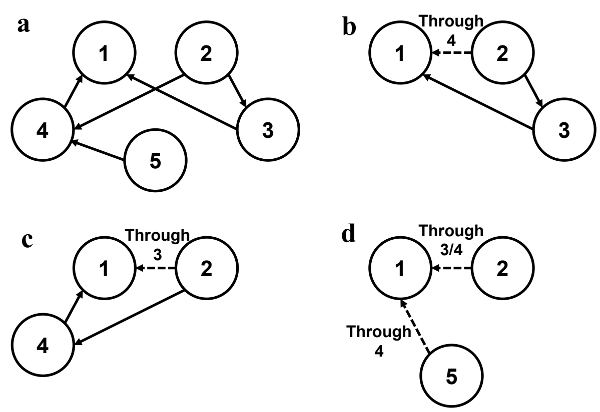


Fig. 1. The case when collapse the original network into three 3-node sub-networks. a. The original network. b-d. The collapsed sub-networks. The directed dash line implies the link that does not exist but is wrongly inferred. Hence, the link from node 2 to 1 is misjudged.

## III. Main results

### A. Dynamical Structure Function

Dynamical structure functions (DSFs) are derived from the signal structure of an LTI system and are characterized by:

$$Y = QY + PU \tag{1}$$

where $Y$ represents measured states, $U$ represents measured inputs, and $(Q, P)$ is the DSF of the system. $Q$ has zero on its

[1]Basi Teng is with the Department of Plant Sciences, University of Cambridge, England bt387@cam.ac.uk.
[2]Yuxuan Zhao and Ye Yuan are with the School of Artificial Intelligence and Automation, Huazhong University of Science and Technology, P.R.China yxz9711@gmail.com, yye@hust.edu.cn.
[3]Jorge Goncalves is with the Luxembourg Centre for Systems Biomedicine, Luxembourg jmg77@cam.ac.uk.
† Co-first authorship.
* Corresponding author.

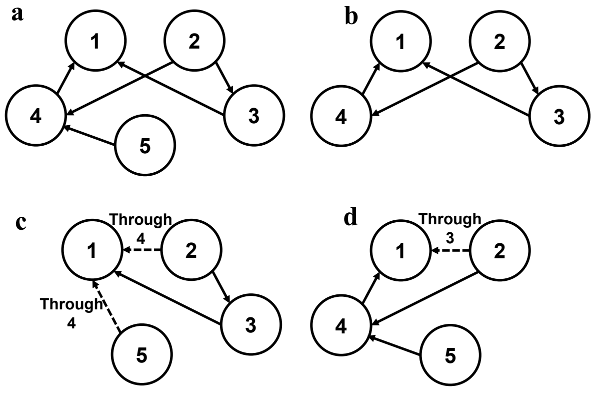


Fig. 2. The case when collapse the original network into three 4-node sub-networks. a. The original network. b-d. The collapsed sub-networks. The directed dash line implies the link that does not exist but is wrongly inferred. With one-vote veto strategy the link can be inferred correctly.

diagonal and describes the transfer functions between measured states. $P$ describes the transfer functions from inputs to outputs without depending on any additional measured states [1], [15], [16].

We call $Q$ and $P$ the internal structure and control structure, respectively, since they can be interpreted as the weighted adjacency matrix of a directed graph indicating the system topology. Note that the DSF in [1] was derived from state space systems where $C = [I, 0]$ and $D = 0$, thus ensuring $(Q, P)$ to be strictly proper. This was extended to general state space system $(A, B, C, D)$ in [15], [16] that makes $(Q, P)$ non-strictly proper. We assume the network is well-posed [17].

### B. Network Collapse

Identification of large-scale systems is typically challenging as it can lead to large dimension computational problems, which are prone to errors. This is also true when reconstructing large-scale DSFs. An alternative to identifying high dimentional DSFs is to reconstruct many small-scale DSFs. Define *network collapse* as the reduction on the number of observed nodes in the network without changing its underlying dynamical system. More precisely, consider an $n$-output dynamical structure function $(Q^{(n)}, P^{(n)})$ from (1). To collapse the DSF to $m$ nodes, first partition the system as follows:

$$\begin{bmatrix} \hat{Y} \\ Z \end{bmatrix} = \begin{bmatrix} Q_{11}^{(n)} & Q_{12}^{(n)} \\ Q_{21}^{(n)} & Q_{22}^{(n)} \end{bmatrix} \begin{bmatrix} \hat{Y} \\ Z \end{bmatrix} + \begin{bmatrix} P_1^{(n)} \\ P_2^{(n)} \end{bmatrix} U \tag{2}$$

where $Y = [\hat{Y}\ Z]^T \in \mathbb{R}^n$ are the observed nodes before collapse, $\hat{Y} \in \mathbb{R}^m$ are the observed nodes after collapse, $Z$ are the $n - m$ nodes to be "hidden" or collapsed. Solving for $Z$ gives

$$Z = (I - Q_{22}^{(n)})^{-1} Q_{21}^{(n)} \hat{Y} + (I - Q_{22}^{(n)})^{-1} P_2^{(n)} U. \tag{3}$$

Substituting (3) into (2) then yields

$$\hat{Y} = W\hat{Y} + VU \tag{4}$$

where

$$W = Q_{11}^{(n)} + Q_{12}^{(n)} (I - Q_{22}^{(n)})^{-1} Q_{21}^{(n)}; \tag{5}$$

$$V = Q_{12}^{(n)} (I - Q_{22}^{(n)})^{-1} P_2^{(n)} + P_1^{(n)}. \tag{6}$$

Subtracting $\text{diag}\{W\}\hat{Y}$ from both sides of (4) yields $(I - \text{diag}\{W\})\hat{Y} = (W - \text{diag}\{W\})\hat{Y} + VU$. Note that $W - \text{diag}\{W\}$ is a matrix with zeros on its diagonal. We then have

$$\hat{Y} = Q^{(m)}\hat{Y} + P^{(m)}U$$

where

$$Q^{(m)} = (I - \text{diag}\{W\})^{-1}(W - \text{diag}\{W\}) \tag{7}$$

Combining (5) and (7), we have the following proposition.

**Proposition 1.** *An $n$-node internal structure $Q^{(n)}$ and its collapsed $m$-node internal structure $Q^{(m)}$ have the following relation:*

$$(I - \textit{diag}\{K\})Q^{(m)} = Q_{11}^{(n)} + K - \textit{diag}\{K\}$$

*where*

$$K = Q_{12}^{(n)} (I - Q_{22}^{(n)})^{-1} Q_{21}^{(n)}$$

In particular, if $m = n - 1$,

$$\begin{aligned} (I - \text{diag}\{K\})Q^{(n-1)} &= Q_{11}^{(n)} + K - \text{diag}\{K\} \\ K &= Q_{12}^{(n)} Q_{21}^{(n)} \end{aligned} \tag{8}$$

Proposition 1 outlines the relationship between the internal structure $Q$ before and after network collapse. Each collapsed internal structure $Q^{(m)}$ provides insights into the original internal structure $Q^{(n)}$, suggesting that the large-scale network could be reconstructed from many collapsed networks that feature different combinations of observed nodes. In the following, we aim to identify a condition that enables the network to be mapped to its multiple small-scale collapsed networks without losing any structural information. This allows for the reconstruction of a large-scale network to be transferred to a small-scale network instead.

This paper focuses on the reconstruction problem of the Boolean structure of $Q^{(n)}$. Henceforth, let $\Theta^{(n)}$ denote the $n$-node Boolean internal structure matrix.

### C. Boolean Internal Structure Reconstruction

There are a total of $\binom{n}{m}$ networks that can be collapsed from an $n$-node to a $m$-node network. Due to their smaller size, each of the smaller $m$-node networks should be easier to reconstruct than the original $n$-node. However, it remains unclear whether the process of network collapse leads to any loss of structural information. We are interested in determining whether the use of a sparsity assumption about internal structure topology can contribute to perfectly converting the reconstruction problem from one large-scale network to many small-scale ones. In contrast to the traditional sparsity requirement that limits the number of regulations (links) in the entire network, we propose a novel sparsity constraint that is defined based on the directed link between any two nodes in the network:

**Definition 1.** *Let $a_i$ ($b_i$) be the indegree (outdegree) of an observable node $i$, $i = 1...n$. Given any two observable nodes $i$ and $j$ in the network, let the sparsity parameter of the directed link from $j$ to $i$ be defined as*

$$r_{ij} = \min\{a_i, b_j\} \tag{9}$$

.

The sparsity parameter of a directed link indicates the number of other regulatory pathways that are similar to this link. When the sparsity parameter $r_{ij}$ is high, there are many other potential indirect pathways that come out from node $j$ and arrive at node $i$, which makes it more challenging to distinguish the direct pathway from a vast number of those indirect. By using the sparsity parameter of a directed link, we can quantify the difficulty of reconstructing every single link without considering the properties of the entire network. Next, is a similar definition of sparsity on the entire network.

**Definition 2.** *Given an $n$-node network, the sparsity parameter $R$ of the network is defined as the maximum sparsity parameter among all pairs of observable nodes in the network:*

$$R = \max_{\substack{i,j\in\{1,2,\ldots,n\}\\ i\neq j}} r_{ij} \tag{10}$$

When reconstructing a single link of Boolean internal structure (we call it Boolean causality of a link), its sparsity parameter represents how "distinguishable" it is. If a link has a low sparsity parameter, which means the number of potential similar pathways is low, it is possible to collapse the network to a lower size without losing any topological information of the target link. Hence, the following theorem explores feasible sizes of collapsed networks based on a link sparsity parameter.

**Theorem 1.** *Given two observable nodes $i, j$ and their sparsity parameter $r_{ij}$, the Boolean causality from node $j$ to node $i$ can be recovered with probability 1 from collapsed $m$-node networks if and only if $m \geq r_{ij} + 2$.*

We have the following corollary to reconstruct the entire Boolean internal structure.

**Corollary 1.** *Consider an $n$-node network with sparsity parameter $R$. Its Boolean internal structure $\Theta^{(n)}$ can be recovered with probability 1 from collapsed $m$-node networks if and only if $m \geq R + 2$*

Theorem 1 states that the dimension of the collapsed networks required to recover a link's Boolean causality depends solely on the sparsity parameter of the link, and not on the network size and other properties. In reality, most of the links in large-scale networks have relatively low sparsity parameters. Theorem 1 offers a practical approach to reconstructing the basic Boolean structure of such a large-scale network by collapsing it to low-dimensional networks. The proof for Theorem 1 is divided in two lemmas.

**Lemma 1.** *Given two observable nodes $i, j$ and their sparsity parameter $r_{ij}$, the Boolean causality from node $j$ to node $i$ can be recovered with probability 1 from all the collapsed $(r_{ij} + 2)$-node networks as*

$$\Theta_{ij}^{(n)} = \prod_{\substack{n_1,\ldots,n_{r_{ij}}\in\{1,2,\ldots,n\}\setminus\{i,j\}\\ n_s\neq n_t,\ for\ s\neq t}} \Theta_{ij}^{\text{node } i,j,n_1,\ldots,n_{r_{ij}} \text{ retained}} \tag{11}$$

*where $\Theta_{ij}^{\text{node } i,j,n_1,\ldots,n_{r_{ij}} \text{ retained}}$ is the Boolean causality of the directed link from node $j$ to node $i$, in the collapsed network that node i, node j and other $r_{ij}$ number of nodes are retained.*

For convenience, we use function $\mathbf{T}$ to describe the multiplication of every link Boolean causality:

$$\mathbf{T}_{ij}^{(n)}(r, K) = \prod_{\substack{n_1,\ldots,n_r\in\{1,2,\ldots,n\}\setminus K\\ n_s\neq n_t,\ for\ s\neq t}} \Theta_{ij}^{\text{node } i,j,n_1,\ldots,n_r \text{ retained}} \tag{12}$$

$r$ is the number of retained nodes apart from node $i$ and $j$, $K$ is a set of nodes that will not be used. Note that node $i$ and $j$ are always retained, and $i, j \in K$.

*Proof.* For simplicity, and without loss of generality, suppose $i = 1, j = 2$ and $r$ stands for $r_{12}$. We prove this by induction.

Base case: Show that Lemma holds for $n = r + 2$. When $n = r + 2$, the collapsed network has the same size as the original network, then $\Theta_{12}^{(n)} = \Theta_{12}^{\text{all nodes retained}}$.

Induction step: Suppose Lemma 1 holds for any $n$ that $n \geq r+2$. Prove that it also holds for $n+1$. We first collapse the $(n+1)$-node network to $n$-node networks. $n-1$ number of networks can be obtained by fixing node $1, 2$ as two observed nodes in the network collapse process. To collapse the $k^{th}$ node ($k \in \{3, 4, \ldots, n + 1\}$), we apply (8) and get

$$(1-q_{1k}^{(n+1)} q_{k1}^{(n+1)}) q_{12}^{(n+1)\setminus\{k\}} = q_{12}^{(n+1)} + q_{1k}^{(n+1)} q_{k2}^{(n+1)} \tag{13}$$

where the superscript '$(n + 1) \setminus \{k\}$' means removing $k^{th}$ node from original $n+1$ nodes network. Since the system is well-posed, $(1-q_{1k}^{(n+1)} q_{k1}^{(n+1)})^{-1}$ exists and is proper. Regard $1 - q_{1k}^{(n+1)} q_{k1}^{(n+1)}$ in (13) as a non-zero term and booleanize it

$$\Theta_{12}^{(n+1)\setminus\{k\}} = \Theta_{12}^{(n+1)} + \Theta_{1k}^{(n+1)} \Theta_{k2}^{(n+1)} \tag{14}$$

Note that in some extreme cases, $q_{12}^{(n+1)}$ and $q_{1k}^{(n+1)} q_{k2}^{(n+1)}$ could be nonzero and perfectly cancel each other. However, for random networks this happens with probability 0. Thus, we assume the booleanization holds from (13) to (14). Then, since Lemma 1 holds for $n$-order systems, we further collapse *LHS* of (14) to $(r + 2)$-node networks:

$$\mathbf{T}_{12}^{(n+1)\setminus\{k\}}(r, \{1, 2, k\}) = \Theta_{12}^{(n+1)} + \Theta_{1k}^{(n+1)} \Theta_{k2}^{(n+1)} \tag{15}$$

Hence, $n - 1$ equations can be obtained from (15) as $k \in \{3, 4, \ldots, n + 1\}$. Since either node 1 has up to $r$ incoming links or node 2 has up to $r$ outgoing link ($r \leq n-2$), there is at least one $\Theta_{1k}^{(n+1)} \Theta_{k2}^{(n+1)}$ equal to zero. Multiply all $n - 1$

equations by each other to obtain

$$\begin{aligned}
&\prod_{k=3}^{n+1} \mathbf{T}_{12}^{(n+1)\setminus\{k\}}(r,\{1,2,k\}) \\
&= \prod_{k=3}^{n+1} \left(\Theta_{12}^{(n+1)} + \Theta_{1k}^{(n+1)}\Theta_{k2}^{(n+1)}\right) \\
&= \Theta_{12}^{(n+1)} + \prod_{k=3}^{n+1} \Theta_{1k}^{(n+1)}\Theta_{k2}^{(n+1)} \\
&= \Theta_{12}^{(n+1)}
\end{aligned} \tag{16}$$

Expanding the *LHS* of (16) and using (12) yields

$$\prod_{k=3}^{n+1} \left[ \prod_{\substack{n_1,\ldots,n_r \in \\ \{3,4,\ldots,n+1\}\setminus\{k\} \\ n_s \neq n_t,\ for\ s\neq t}} \Theta_{12}^{\text{node 1, 2, } n_1,\ldots,n_r \text{ retained}} \right] = \Theta_{12}^{(n+1)} \tag{17}$$

We want to prove $\mathbf{T}_{12}^{(n+1)}(r,\{1,2\}) = \Theta_{12}^{(n+1)}$ from (17). If $\Theta_{12}^{(n+1)} = 0$, then there exists an $k \in \{3,4,\ldots,n+1\}$ and $n_1, n_2, \ldots, n_r \in \{3,4,\ldots,n+1\} \setminus \{k\}$ ($n_s \neq n_t$, if $s \neq t$) such that $\Theta_{12}^{\text{node 1, 2, } n_1,\ldots,n_r \text{ retained}} = 0$. This is one of the factor in the multiplication behind $\mathbf{T}_{12}^{(n+1)}(r,\{1,2\})$. Hence, $\mathbf{T}_{12}^{(n+1)}(r,\{1,2\}) = 0$.

If $\mathbf{T}_{12}^{(n+1)}(r,\{1,2\}) = 0$, there exist $n_1, n_2, \ldots, n_r \in \{3,4,\ldots,n+1\}$ ($n_s \neq n_t$, if $s \neq t$) such that $\Theta_{12}^{\text{node 1, 2, } n_1,\ldots,n_r \text{ retained}} = 0$. Without loss of generality, suppose $n_1 = 3, n_2 = 4, \ldots, n_r = r+2$. Then $\Theta_{12}^{\text{node 1, 2, 3, ..., } r+2 \text{ retained}} = 0$. Suppose $k = r+3$, we find a set of $n_1, \ldots, n_r \in \{3,4,\ldots,n+1\} \setminus \{k\}$ that satisfies $\Theta_{12}^{\text{node 1, 2, } n_1,\ldots,n_r \text{ retained}} = 0$. Then $\Theta_{12}^{(n+1)} = 0$ from (17).

Im summary, we have the following equations proving that $n+1$ also holds

$$\Theta_{12}^{(n+1)} = \mathbf{T}_{12}^{(n+1)}(r,\{1,2\})$$

■

Lemma 1 states that collapsed $(r_{ij}+2)$-node networks can reconstruct the original Boolean structure $\Theta_{ij}$ perfectly. Next, we show that Boolean causality cannot always be recovered from collapsed smaller sized networks than $(r_{ij}+2)$-nodes.

**Lemma 2.** *Given two observable nodes $i, j$ and their sparsity parameter $r_{ij}$, the Boolean causality from node $j$ to node $i$ cannot always be recovered from the collapsed $(r_{ij}+1)$-node networks.*

*Proof.* We still focus on $\Theta_{12}^{(n)}$ for simplicity. Consider the construction of a network with $n \geq r_{ij}+2$ nodes as shown in Fig. 3. We demonstrate that $\Theta_{12}^{(n)}$ in this network cannot be recovered correctly using collapsed $(r_{12}+1)$-node networks. As the number of retained nodes (i.e. $r_{12}-1$) is smaller than the number of nodes connecting node 1 and 2 (i.e. $r_{12}$), for any $r_{12}-1$ retained nodes $n_1, n_2, \ldots, n_{r_{12}-1} \in \{3,4,\ldots,n\}$ ($n_s \neq n_t$, if $s \neq t$), $\Theta_{12}^{\text{node 1, 2, } n_1,\ldots,n_{r_{12}-1} \text{ retained}} = 1$. Denote

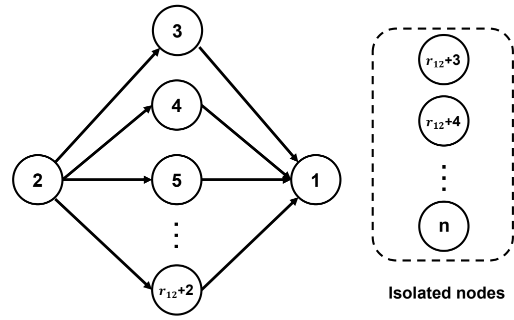


Fig. 3. The special case where node 2 does not directly regulate node 1, but it can regulate node 2 indirectly via separate intermediate nodes, including nodes 3, 4, ..., $r_{12}$+2. Hence, $\Theta_{12}$ cannot be correctly reconstructed by collapsing the network to $r_{12}+1$-node networks.

$\widetilde{\Theta}_{12}^{(n)}$ as the estimate of the real $\Theta_{12}^{(n)}$. Then, we have

$$\widetilde{\Theta}_{12}^{(n)} = \mathbf{T}_{12}^{(n)}(r_{12}-1,\{1,2\}) = 1$$

which is inconsistent with the real $\Theta_{12}^{(n)} = 0$ ■

We further use Example 1 again to illustrate Lemma 2. We can build the Boolean internal structure matrix in terms of the graph topology:

$$\Theta^{(5)} = \begin{bmatrix} 0 & 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix} \tag{18}$$

To infer the Boolean causality from node 2 to node 1 which has the sparsity parameter of $r_{12} = 2$. If collapsing the network to 3-node, three collapsed sub-networks can be obtained

$$\begin{aligned}
\Theta^{(5)\setminus\{4,5\}} &= \begin{bmatrix} 0 & 1 & 1 \\ 0 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix} \\
\Theta^{(5)\setminus\{3,5\}} &= \begin{bmatrix} 0 & 1 & 1 \\ 0 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix} \\
\Theta^{(5)\setminus\{3,4\}} &= \begin{bmatrix} 0 & 1 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}
\end{aligned} \tag{19}$$

Then $\widetilde{\Theta}_{12}^{(5)} = \mathbf{T}_{12}^{(5)}(1,\{1,2\}) = 1$, which is inconsistent with the real $\Theta_{12}^{(5)} = 0$.

The proof of Theorem 1 follows from Lemmas 1 and 2.

*Proof of Theorem 1.* When $m \geq r_{ij}+2$. Since Lemma 1, any $m$-node Boolean network can be recovered by collapsed $(r_{ij}+2)$-node networks (Let $n = m$). Therefore, for any $n$-node network, we can always collapse it to well-recovered $m$-node networks. These $m$-node networks must be able to recover $\Theta_{ij}^{(n)}$ by Lemma 1.

For $m \leq r_{ij}+1$, if it is possible to recover $\Theta_{ij}^{(n)}$ through collapsing to $m$-node networks, then if $n = r_{ij}+1$ we can still get $\Theta_{ij}^{(r_{ij}+1)}$ by collapsing to $m$-node networks.

Therefore, for any $n$-node network, we can always collapse it to well-recovered $(r_{ij}+1)$-node networks, which indicates through $(r_{ij}+1)$-node networks we will always be able to recover $\Theta_{ij}^{(n)}$. This contradicts Lemma 2. Therefore, collapsing the network to the dimension of $m \leq r_{ij}+1$ cannot recover $\Theta_{ij}^{(n)}$ perfectly.

■

From Theorem 1, only those links that have a sparsity parameter of 1 in Example 1 can be reconstructed by collapsed 3-node networks, such as

$$\widetilde{\Theta}_{13}^{(5)} = \mathbf{T}_{13}^{(5)}(1, \{1,3\}) = 1 = \Theta_{13}^{(5)};$$
$$\widetilde{\Theta}_{32}^{(5)} = \mathbf{T}_{32}^{(5)}(1, \{2,3\}) = 0 = \Theta_{32}^{(5)}$$

For some links that have a sparsity parameter of 2, they can be reconstructed by a 4-node networks

$$\widetilde{\Theta}_{12}^{(5)} = \mathbf{T}_{12}^{(5)}(2, \{1,2\}) = 0 = \Theta_{12}^{(5)}$$

When reconstructing the entire network, the sparsity parameters of all the links must be taken into consideration. The size of the collapsed networks is determined by the maximal sparsity parameter of all the links, which is the sparsity parameter of the network defined in Definition 2.

Theorem 1 and Corollary 1 present criteria to collapse the network for reconstructing a single link and the entire network, respectively. However, if the networks consist predominantly of links with low sparsity parameters, Theorem 1 can be used instead of Corollary 1 to choose the dimension of collapsed networks and reconstruct the basic Boolean internal structure. Although this may have slightly lower accuracy due to potential misreconstruction of high sparsity parameter links, it allows for the network to collapse to a smaller dimension. Ultimately, this trade-off between accuracy and dimensionality enables a more effective reconstruction of the Boolean internal structure.

In practice, the formula in Theorem 1 and Corollary 1 can be explained as a "one-vote veto" strategy. When reconstructing a link Boolean causality, if the reconstruction result from any collapsed network is zero (non-existent), then the final result will be zero without the need to consider other collapsed networks. This can greatly speed up the computation as it is unnecessary to compute all possible collapsed networks. The whole reconstruction process combining the one-vote veto strategy is summarized as follows.

**Algorithm 1** Link Boolean causality reconstruction using collapsed networks

**Input:** Experiments on $n$-node network, Dimension of collapsed networks $m$

**Output:** Link Boolean causality $\Theta_{ij}^{(n)}$

```
    Initialisation
    k = 1
    Θ_ij^(n) = 1
    Loop Process
    while k ≤ C(n−2, m−2) and Θ_ij^(n) = 1 do
        Besides node i, j, choose other m − 2 retained nodes:
        i, j, n_1, ... n_{m−2}
        Θ_ij^(n) = Θ_ij^{node 1, 2, n_1,...,n_{m−2} retained}
        k = k + 1
    end while
```

## IV. Example

Here we demonstrate the approach with a large-scale network reconstruction example.

**Example 2.** Consider a 10-node network shown in Fig. 4, where arrows denote direct relations containing all the dynamics of the hidden states. 15 links in the network form typical tree structures and ring structures. The sparsity parameter of the network $R = 3$.

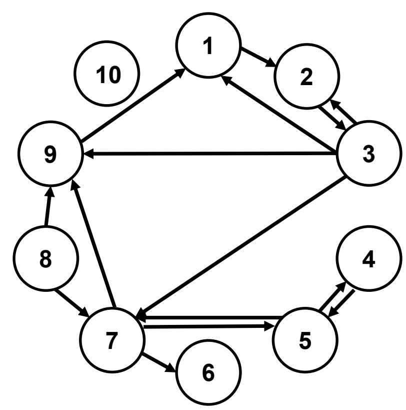


Fig. 4. Network structure in Example 2

The network topology is fully characterised by the Boolean internal structure as

$$\Theta^{(10)} = \begin{bmatrix} 0&0&1&0&0&0&0&0&1&0 \\ 1&0&1&0&0&0&0&0&0&0 \\ 0&1&0&0&0&0&0&0&0&0 \\ 0&0&0&0&1&0&0&0&0&0 \\ 0&0&0&1&0&0&1&0&0&0 \\ 0&0&0&0&0&0&1&0&0&0 \\ 0&0&0&0&1&0&0&1&0&0 \\ 0&0&0&0&0&0&0&0&0&0 \\ 0&0&1&0&0&0&1&1&0&0 \\ 0&0&0&0&0&0&0&0&0&0 \end{bmatrix} \quad (20)$$

We can calculate a matrix of $m$ that describes the minimal dimension of collapsed networks for reconstructing every single link in the network:

$$[m]_{ij}^{(10)} = [r_{ij}^{(10)}+2] = \begin{bmatrix} * & 3 & 4 & 3 & 4 & 2 & 4 & 4 & 3 & 2 \\ 3 & * & 4 & 3 & 4 & 2 & 4 & 4 & 3 & 2 \\ 3 & 3 & * & 3 & 3 & 2 & 3 & 3 & 3 & 2 \\ 3 & 3 & 3 & * & 3 & 2 & 3 & 3 & 3 & 2 \\ 3 & 3 & 4 & 3 & * & 2 & 4 & 4 & 3 & 2 \\ 3 & 3 & 3 & 3 & 3 & * & 3 & 3 & 3 & 2 \\ 3 & 3 & 5 & 3 & 4 & 2 & * & 4 & 3 & 2 \\ 2 & 2 & 2 & 2 & 2 & 2 & 2 & * & 2 & 2 \\ 3 & 3 & 5 & 3 & 4 & 2 & 5 & 4 & * & 2 \\ 2 & 2 & 2 & 2 & 2 & 2 & 2 & 2 & 2 & * \end{bmatrix} \tag{21}$$

The maximal element in matrix $[m]_{ij}^{(10)}$ is 5, corresponding to $R = 3$. To reconstruct the entire network topology perfectly from Corollary 1, the dimension of collapsed networks used must be at least 5. If using 4-dimensional collapsed networks, most of the links can still be correctly reconstructed, except for two links $\Theta_{73}^{(10)}$ and $\Theta_{93}^{(10)}$. If we reduce the dimension of the collapsed networks to 3, 73 out of 90 potential links can still be correctly reconstructed.

## V. Conclusion

Reconstructing large-scale DSFs at one time is sometimes difficult due to the accuracy and the cost of computation and time. To address this problem, this paper defined the sparsity parameter to describe the reconstructability of the network. We proposed a reconstruction method that involves network collapses and a one-vote veto strategy. This lead to necessary and sufficient condition that ensures reconstruction. For sparse networks, this work has the potential to significantly reduce the computational costs and improve accuracy for large-scale networks. Especially for links with low scarcity parameter $r_{ij}$. Moreover, collapsed networks can be computed in parallel.

## References


[1] J. Goncalves and S. Warnick, "Necessary and sufficient conditions for dynamical structure reconstruction of lti networks," *Automatic Control, IEEE Transactions on*, vol. 53, pp. 1670 – 1674, 09 2008.

[2] P. Van den Hof, A. Dankers, P. Heuberger, and X. Bombois, "Identification of dynamic models in complex networks with prediction error methods—basic methods for consistent module estimates," *Automatica*, vol. 49, pp. 2994–3006, 10 2013.

[3] M. Balcerowicz, M. Mahjoub, D. Nguyen, H. Lan, D. Stöckle, S. Conde, K. Jaeger, P. Wigge, and D. Ezer, "An early-morning gene network controlled by phytochromes and cryptochromes regulates photomorphogenesis pathways in arabidopsis," *Molecular plant*, vol. 14, 03 2021.

[4] Y. Kim, M. Yeom, H. Kim, J. Lim, H. Koo, D. Hwang, D. Somers, and H. Nam, "Gigantea and early flowering 4 in arabidopsis exhibit differential phase-specific genetic influences over a diurnal cycle," *Molecular plant*, vol. 5, pp. 678–87, 02 2012.

[5] J. Burés and I. Larrosa, "Organic reaction mechanism classification using machine learning," *Nature*, vol. 613, pp. 689–695, 01 2023.

[6] A. Aalto, L. Viitasaari, P. Ilmonen, L. Mombaerts, and J. Goncalves, "Gene regulatory network inference from sparsely sampled noisy data," *Nature Communications*, vol. 11, 07 2020.

[7] V. A. Huynh-Thu and P. Geurts, "Dyngenie3: Dynamical genie3 for the inference of gene networks from time series expression data," *Scientific Reports*, vol. 8, 02 2018.

[8] A. Aderhold, D. Husmeier, and M. Grzegorczyk, "Approximate bayesian inference in semi-mechanistic models," *Statistics and Computing*, vol. 27, 07 2017.

[9] J. Casadiego, M. Nitzan, S. Hallerberg, and M. Timme, "Model-free inference of direct network interactions from nonlinear collective dynamics," *Nature Communications*, vol. 8, 12 2017.

[10] J. Castro, I. Valdés, L. González, G. Danies, S. Cañas, F. Vischi Winck, C. Ñústez, S. Restrepo, and D. M. Riaño-Pachón, "Gene regulatory networks on transfer entropy (grnte): A novel approach to reconstruct gene regulatory interactions applied to a case study for the plant pathogen phytophthora infestans," *Theoretical Biology and Medical Modelling*, vol. 16, 04 2019.

[11] Y. Yuan, G.-B. Stan, S. Warnick, and J. Goncalves, "Robust dynamical network structure reconstruction," *Automatica*, vol. 47, pp. 1230–1235, 06 2011.

[12] Y. Yuan, G.-B. Stan, S. Warnick, and J. Gonçalves, "Robust dynamical network reconstruction," in *49th IEEE Conference on Decision and Control (CDC)*, 2010, pp. 810–815.

[13] V. C. Rajagopal, K. R. Ramaswamy, and P. M. J. Van Den Hof, "Learning local modules in dynamic networks without prior topology information," in *2021 60th IEEE Conference on Decision and Control (CDC)*, 2021, pp. 840–845.

[14] S. Shi, X. Cheng, and P. Van den Hof, "Single module identifiability in linear dynamic networks with partial excitation and measurement," *IEEE Transactions on Automatic Control*, vol. PP, pp. 1–1, 12 2021.

[15] V. Chetty and S. Warnick, "Network semantics of dynamical systems," in *2015 54th IEEE Conference on Decision and Control (CDC)*, 2015, pp. 1557–1562.

[16] P. Van den Hof, A. Dankers, P. Heuberger, and X. Bombois, "Identification of dynamic models in complex networks with prediction error methods—basic methods for consistent module estimates," *Automatica*, vol. 49, pp. 2994–3006, 10 2013.

[17] M. Araki and M. Saeki, "A quantitative condition for the well-posedness of interconnected dynamical systems," *IEEE Transactions on Automatic Control*, vol. 28, no. 5, pp. 569–577, 1983.