2017 IEEE 56th Annual Conference on Decision and Control (CDC)
December 12-15, 2017, Melbourne, Australia

# Network Topology Identification From Corrupt Data Streams

Venkat Ram Subramanian, Andrew Lamperski, and Murti V. Salapaka

*Abstract*— **The interconnectivity structure of many complex systems can be modeled as a network of dynamically interacting processes. Identification of mutual dependencies amongst the agents is of primary importance in many application domains that include internet-of-things, neuroscience and econometrics. Moreover, in many such systems it is not possible to deliberately affect the system and thus passive methods are of particular relevance. However, for an effective framework that identifies influence pathways from dynamically related data streams originating at different sources it is essential to address the uncertainty of data caused by possibly unknown time-origins of different streams and other corrupting influences including packet drops and noise. In this article, a method of reconstructing the network topology from corrupt data streams is provided with emphasis on the characterization of the effects of data corruption on the reconstructed network. The structure of the network is identified by observing the sparsity pattern in the joint power spectrum of the measurements.**

## I. Introduction

Models of systems as networks of interacting systems are central to many domains such as repeated drug testing [1], automatically assisted anesthesia [2], mesh compression/video segmentation for video streaming [3], gene regulatory networks [4], quantitative finance [5] and neuroscience [6]. In the internet-of-things (IOT) (see [7], [8]), data collected from ubiquitously sensorized devices and/or from many sensors of a single large system are processed to glean important insights into the network of interacting components. Here, data-streams from various sensors can be dynamically related, where the inter-dependency can be caused by the interaction physics of system components. Often, in such large systems, the time-origin of time-series data is not known and the data collection mechanism is plagued by uncertainty in the measurement process with noise and lost data packets.

In [9], the authors have introduced and used the concept of the generative model of linear dynamic graphs and established that optimal multivariate Wiener filters can reconstruct the undirected structure associated with the generative model. However, the effect of measurement uncertainty on the reconstruction of the underlying influences was not investigated. The methods introduced in [9], with ideal measurements do introduce spurious links, however, an attractive feature here is that the spurious links are localized within a hop of a true link. It is desirable to establish conditions under which the combination of spurious links from the method of [9] and the effects of data corruption do not compound so that negative effects remain localized.

In this article, we seek insights into how the network structure influences propagation of the effects of data uncertainty. Time-series data uncertainty sources studied include noise, uncertainty in time-origin of the various data-streams, and lost data packets. Our main result, Theorem 2, shows that data corruption can lead to a spurious addition of links. Furthermore, we identify a set of nodes in the network in which spurious links could potentially appear. The results can be utilized to better inform on what part of the reconstruction can be trusted and for allocation of resources to minimize the affects of data corruption.

Section II-B presents the class of models used for representing the complex network systems. Next, Section III describes the general form of perturbations or corruptions that are of practical relevance. The main results and methods to identify the network structure are discussed in Section IV. Section V deals with the simulation results to demonstrate the theoretical predictions.

*Notation:*

$x_i$ or $\{x\}_i$ means *i*th element of vector $x$.
$M^T$ denotes the transpose of a matrix or vector $M$.
$M_{ij}$ indicates the $(i,j)^{th}$ entry of a matrix $M$.
If $M(z)$ is a transfer function matrix, then $M(z)^\sim = M(z^{-1})^T$ is the conjugate transpose.
$\mathbb{E}[\cdot]$ denotes expectation operator.
$R_{XY}(k) := \mathbb{E}[X[n+k]Y[n]]$ is the cross-correlation function of jointly wide-sense stationary processes $X$ and $Y$. If $Y = X$ then $R_{XX}(k)$ is called the auto-correlation.
$\Phi_{XY}(z) := \mathcal{Z}(R_{XY}(k))$ represents the cross-power spectral density while $\Phi_{XX}(z) := \mathcal{Z}(R_{XX}(k))$ denotes the power spectral density(PSD). $\mathcal{Z}(\cdot)$ is the Z-transform operator.
$b_i$ represents the $i^{th}$ element of the canonical basis of $\mathbb{R}^n$.

## II. Preliminaries

### A. Definitions

*Definition 1 (Directed and Undirected Graphs):* A *directed graph* $\mathcal{G}$ is a pair $(V, A)$ where $V$ is a set of vertices or nodes and $A$ is a set of edges given by ordered pairs $(i, j)$ where $i, j \in V$. If $(i, j) \in A$, then we say that there is an edge from $i$ to $j$. $(V, A)$ forms an *undirected graph* if $V$ is a set of nodes or vertices and $A$ is a set of the ordering of the un-ordered pairs $\{i, j\}$.

*Definition 2 (Children and Parents):* Given a directed graph $\mathcal{G} = (V, A)$ and a node $j \in V$, the children of $j$ are defined as $\mathcal{C}_j := \{i|(j, i) \in A\}$ and the parents of $j$ as $\mathcal{P}_j := \{i|(i, j) \in A\}$.

The authors are with the Department of Electrical and Computer Engineering, University of Minnesota, Minneapolis, MN 55455, USA.subra148@umn.edu, alampers@umn.edu, murtis@umn.edu

978-1-5090-2873-3/17/$31.00 ©2017 IEEE

(a) (b)

Fig. 1: 1a Directed Graph and 1b Kin Graph associated for a Dynamic Influence Model.

*Definition 3 (Kins):* Given a directed graph $\mathcal{G} = (V, A)$ and a node $j \in V$, kins of $j$ are defined as $\mathcal{K}_j := \left\{i | i \neq j \text{ and } i \in \mathcal{C}_j \cup \mathcal{P}_j \cup \mathcal{P}_{\mathcal{C}_j}\right\}$. Kins are formed by parents, children and spouses. A spouse of a node is another node where both nodes have at-least one common child.

*Definition 4 (Kin-Graph):* Given a directed graph $\mathcal{G} = (V, A)$, its kin-graph is the undirected graph $\mathcal{G}^M = (V, A^M)$ where $A^M := \{\{i, j\} | j \in V, i \in \mathcal{K}_j\}$.

Fig. 1 provides an example of a directed graph and its kin graph.

### B. Dynamic Influence Model

Let $\mathcal{G} = (V, A)$ where $V = \{1, \ldots, N\}$. To each node, we associate an agent. For each agent $i$, we associate a discrete time sequence $x_i[\cdot]$ and a sequence $e_i[\cdot]$. The process $e_i[\cdot]$ is considered innate to agent $i$ and thus $e_i$ is independent of $e_j$ if $i \neq j$. The process $x_j$ depends dynamically on the processes of is parents, $x_i$ with $i \in \mathcal{P}_j$:

$$x_j = \sum_{i \in \mathcal{P}_j}^{N} G_{ji}(z) x_i + e_j \quad \text{for } j = 1, \ldots, N. \tag{1}$$

Here $G_{ji}$ is a linear convolution operator.

Let $x = (x_1, x_2, \ldots, x_N)^T$ and $e = (e_1, e_2, \ldots, e_N)^T$. Then (1) is equivalent to

$$x = G(z)x + e. \tag{2}$$

We refer to (2) as the Dynamic Influence Model (DIM). Here, $G$ is termed as the DIM *generative connectivity* matrix. The diagonal entries $G_{jj}(z)$ are assumed to be zero. The DIM will be denoted by $(G, e)$ and the underlying graph, $\mathcal{G}$ is called the *generative directed graph* of $(G, e)$.

We illustrate the notation by an example. Consider a network of five agents whose node dynamics are given by,

$$\begin{aligned} x_1 &= e_1 \\ x_2 &= G_{21}(z)x_1 + e_2 \\ x_3 &= G_{31}(z)x_1 + e_3 \\ x_4 &= G_{42}(z)x_2 + G_{43}(z)x_3 + e_4 \\ x_5 &= G_{54}(z)x_4 + e_5 \end{aligned}$$

with $G = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 \\ G_{21} & 0 & 0 & 0 & 0 \\ G_{31} & 0 & 0 & 0 & 0 \\ 0 & G_{42} & G_{43} & 0 & 0 \\ 0 & 0 & 0 & G_{54} & 0 \end{bmatrix}$.

The DIM is described by $(G, e)$ where the generative directed graph is given by Fig. 1 (a). Note that we do not show the processes $e_i$ in the generative graph.

## III. Uncertainty Description

In this section we provide a description for how uncertainty affects the time-series $x_i$.

### A. General Perturbation Models

The data stream $u_i$ associated with $i^{th}$ node is obtained from the unperturbed time-series $x_i$ as

$$u_i[.] = f_i(x_i[.], \zeta_i[.]), \tag{3}$$

where $u_i$ can depend dynamically on $x_i$ and $\zeta_i$ represents a stochastic description of the corruption. We highlight three such perturbation models that are practically relevant.

*Temporal Uncertainty:* Consider a node $i$ in a DIM $(G, e)$. Suppose $n$ is the true clock index but the node $i$ measures a noisy clock index which is given by a random process $\tau_i[n]$. One such probabilistic model is given by:

$$\tau_i[n] = \begin{cases} n + k_1, & \text{with probability } p_i \\ n + k_2, & \text{with probability } (1 - p_i). \end{cases}$$

The corruption model from (3) for randomized delays in information transmission takes the form:

$$u_i[n] = x_i[\tau_i[n]] \tag{4}$$

*Measurement Noise:* Given a node $i$ in a DIM $(G, e)$ the data stream $x_i$ are corrupted with uncorrelated measurement noise $v_i[\cdot]$, which is another wide-sense stationary process and is independent of $x_i[\cdot]$. Hence the perturbation models is given by:

$$u_i[n] = x_i[n] + v_i[n],$$

*Packet Drops:* Here the corrupted data stream $u_i$ is obtained from $x_i$ as follows:

$$u_i[n] = \begin{cases} x_i[n], & \text{with probability } p_i \\ u_i[n-1], & \text{with probability } (1 - p_i) \end{cases} \tag{5}$$

### B. Relationship of the power spectra of the corrupted data streams to the original power spectra

Reconstruction of the generative graph of the DIM depends intimately on the sparsity of the inverse of the cross-power-spectral density matrix (see [9]). Here, the relationship between the cross-spectral density matrix of the corrupted data streams and the original cross-spectra is described for the corruption models presented earlier.

Under these perturbation models, the signals $u_i$ will have cross-spectra and power spectra of the form:

$$\Phi_{u_i x_i}(z) = h_i(z)\Phi_{x_i x_i}(z) \tag{6a}$$

$$\Phi_{u_i u_i}(z) = h_i(z)h_i(z^{-1})\Phi_{x_i x_i}(z) + d_i(z), \tag{6b}$$

for some transfer functions $h_i$ and $d_i$. If the perturbations were deterministic and time invariant so that $u_i = h_i(z)x_i$, then the power spectrum formulas would hold with $d_i(z) = 0$. However, the randomized perturbations imply that $d_i(z) \neq 0$. The temporal uncertainty case is described in detail, while the other disturbance models are just sketched.

*Time-Origin Uncertainty:* Appendix I briefly sketches the derivation of the cross-spectrum between $u_i$ and $x_i$ which is given by,

$$\Phi_{u_i x_i}(z) = (p_i z^{k_1} + (1-p_i)z^{k_2})\Phi_{x_i x_i}(z) \tag{7}$$

From this equation, it would appear that the effect on the spectrum is equivalent to passing the measurements through the filter $h_i(z) := (p_i z^{k_1} + (1-p_i)z^{k_2})$.

However, the situation is more subtle. If the effect were simply equivalent to applying a filter, the Power Spectral Density (PSD) of $u_i$ is given by: $\Phi_{u_i u_i}(z) = h_i(z)h_i(z^{-1})\Phi_{x_i x_i}(z)$.

As seen from the analysis in Appendix I the PSD is in fact given by,

$$\Phi_{u_i u_i}(z) = h_i(z)\Phi_{x_i x_i}(z)h_i(z^{-1}) + d_i \tag{8}$$

where $d_i = R_{x_i x_i}(0)\left\{1 - p_i{}^2 - (1-p_i)^2\right\} - p_i(1-p_i)[R_{x_i x_i}(k_1 - k_2) + R_{x_i x_i}(k_2 - k_1)]$. Thus, the psd of the process $u_i$ in (4) is equivalent to the PSD of a process obtained by filtering $x_i$ and adding white noise with variance $d_i$.

*Measurement Noise:* If $v_i[n]$ is a zero-mean wide-sense stationary process with with power spectrum $\Phi_{v_i v_i}(z)$, independent of $x_i[n]$, then the perturbation transfer functions are given by:

$$\begin{aligned} h_i(z) &= 1 \\ d_i(z) &= \Phi_{v_i v_i}(z). \end{aligned}$$

This is immediate from the independence of $x_i$ and $v_i$.

*Packet Dropping:* For the packet dropping link, the perturbation transfer functions are given by:

$$h_i(z) = \frac{p_i}{1 - z^{-1}(1-p_i)}$$

$$d_i(z) = R_{x_i x_i}(0) - \frac{p_i^2 R_{x_i x_i}(0) + p_i(1-p_i)(R_{u_i x_i}(-1) + R_{x_i u_i}(1))}{1-(1-p_i)^2} \tag{9}$$

In this case, the cross-correlations $R_{u_i x_i}$ and $R_{x_i u_i}$ can be computed from the relationships $\Phi_{u_i x_i}(z) = h_i(z)\Phi_{x_i x_i}(z)$. The derivation is briefly sketched in Appendix I.

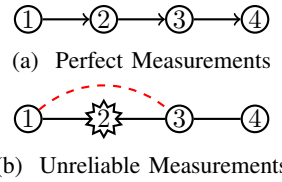
## IV. Network Topology Identification

Before tackling the problem of determining the graph of a DIM for the case with data corruption we discuss the case where the measurements are ideal and thus $u_k = x_k$.

### A. Determining the Generative topology of a DIM: Perfect Measurement Case

The following results are obtained from [9] where the authors have leveraged Wiener filters for determining generative graphs of a DIM.

***Theorem** 1: Consider a DIM $(G(z), e)$ consisting of N nodes with generative graph $\mathcal{G}$. Let the output of the DIM be given by $x = (x_1, \ldots, x_N)'$. Suppose that $S$ is the span of all random variables $x_k[t]$, $t = \ldots -2, -1, 0, 1, 2 \ldots$*




(a) Perfect Measurements

(b) Unreliable Measurements

Fig. 2: When node 2 has corrupt measurements an external observer might wrongly infer that the third node is directly influenced by node 1.

*excluding $x_j$. Define the estimate $\hat{x}_j$ of the time-series $x_j$ via the optimization problem of*

$$\min_{\hat{x}_j \in S} \mathbb{E}\left[(x_j - \hat{x}_j)^T (x_j - \hat{x}_j)\right].$$

*Then a unique optimal solution to the above exists and is given by*

$$\hat{x}_j = \sum_{i \neq j} \mathbf{W}_{ji}(z) x_i \tag{10}$$

*where $\mathbf{W}_{ji}(z) \neq 0$ implies $x_i \in \mathcal{K}_{x_j}$ (equivalently $x_j \in \mathcal{K}_{x_i}$); that is $i$ is a kin of $j$.*

***Corollary** 1: Under the assumptions of Theorem 1, let $\Phi_{xx}$ be the power spectral density matrix of the vector process $x$. Then the $(j,i)$ entry of $\Phi_{xx}^{-1}$ is non zero implies that $i$ is a kin of $j$.*

The solution in (10) is the Wiener Filter solution which is given by $\Phi_{x_j x_{\bar{j}}}\Phi_{x_{\bar{j}} x_{\bar{j}}}^{-1}$ where $x_{\bar{j}}$ denotes the vector of all processes excluding $x_j$ and $\Phi$ denotes the power spectral density.

*Remark:* We emphasize that the PSD $\Phi_{xx}$ can be computed based solely on the measurements $x_i$, $i = 1, \ldots, n$..

### B. Example of Data Corruption

Consider the generative graph of a directed chain in Figure 2a. Suppose the measured data-streams are denoted by $u_i$ for node $i$ where $u_i = x_i$ for $i = 1, 3, 4$ (thus no data uncertainty at nodes 1, 3 and 4) and $u_2$ is related to $x_2$ via the randomized delay model described in (4). In this case, the processes $u_i$ are jointly WSS and the PSD of the vector process $u = (u_1, \cdots, u_4)^\top$ is related to the PSD of the vector process $x$ via:

$$\Phi_{uu}(z) = \underbrace{\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & h_2(z) & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}}_{H(z)} \Phi_{xx}(z) \underbrace{\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & h_2(z^{-1}) & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}}_{H^\sim(z)} + \underbrace{\begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & d_2 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}}_{D},$$

where $h_2$ and $d_2$ were described in Subsection III-B.

Note that $D = b_2 d_2 b_2^T$, where $b_2 = \begin{pmatrix} 0 & 1 & 0 & 0 \end{pmatrix}^T$. Set $\Psi(z) = H(z)\Phi_{xx}(z)H^\sim(z)$. It follows from the Woodbury

matrix identity [10] that

$$\Phi_{uu}^{-1}(z) = \Psi^{-1}(z) - \Psi^{-1}(z)b_2 b_2^T \Psi^{-1}(z)\Delta^{-1}, \qquad (11)$$

where $\Delta = d_2^{-1} + b_2^T \Psi^{-1}(z) b_2$ is a scalar.

Corollary 1 implies that the sparsity pattern of $\Phi_{xx}^{-1}(z)$ is given by:

$$\Phi_{xx}^{-1}(z) = \begin{bmatrix} * & * & 0 & 0 \\ * & * & * & 0 \\ 0 & * & * & * \\ 0 & 0 & * & * \end{bmatrix} \qquad (12)$$

where $*$ indicates a potential non-zero entry.

Since $H(z)$ is diagonal, it follows that $\Psi^{-1}(z)$ and $\Phi_{xx}^{-1}(z)$ have the same sparsity pattern. Thus, the sparsity pattern of $\Psi^{-1}(z)b_2$ and $\Psi^{-1}(z)b_2 b_2^T \Psi^{-1}(z)$ are given by:

$$\Psi^{-1}(z)b_2 = \begin{bmatrix} * \\ * \\ * \\ 0 \end{bmatrix}, \quad \Psi^{-1}(z)b_2 b_2^T \Psi^{-1}(z) = \begin{bmatrix} * & * & * & 0 \\ * & * & * & 0 \\ * & * & * & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} \qquad (13)$$

Combining (11)-(13), it follows that the $\Phi_{uu}^{-1}(z)$ has sparsity pattern given by:

$$\Phi_{uu}^{-1}(z) = \begin{bmatrix} * & * & * & 0 \\ * & * & * & 0 \\ * & * & * & * \\ 0 & 0 & * & * \end{bmatrix}.$$

### C. Determining Generative Topology from Corrupted Data Streams

In this subsection, we will generalize the insights from the preceding subsection to arbitrary DIMs. The following definitions are needed for the development to follow.

*Definition 5 (Neighbors $\mathcal{N}$):* Let $\mathcal{G}^M = (V, A^M)$ be a kin graph. The neighbor set of node $i$ is given by $\mathcal{N} = \{j : \{i,j\} \in A^M\} \cup \{i\}$.

*Definition 6 (Erroneous Links):* Let $\mathcal{G}^M = (V, A^M)$ be a kin graph. An edge or arc $\{i,j\}$ is called an erroneous link when it does not belong to $A^M$ where $i, j \in V$.

*Definition 7 (Perturbed Kin Graph):* Let $\mathcal{G}^M = (V, A^M)$ be a kin graph. Suppose $Y \subset V$ is the set of perturbed nodes. Then the perturbed kin-graph of $\mathcal{G}^M$ with respect to set $Y$ is the graph $\mathcal{G}_Y^M = (V, A_Y^M)$ such that $\{i,j\} \in A_Y^M$ if either $\{i,j\} \in A^M$ or there is a path from $i$ to $j$ in $\mathcal{G}^M$ such that all intermediate nodes are in $Y$.

Note that this graph has the property that the neighbors of any node in $Y$ must form a clique in $\mathcal{G}_Y^M$. Furthermore, by construction, if $Y \subset \hat{Y}$, then $A_Y^M \subset A_{\hat{Y}}^M$.

The following theorem is the main result of the paper.

***Theorem** 2: Consider a DIM $(G(z), e)$ consisting of $N$ nodes with the kin graph $\mathcal{G}^M = (V, A^M)$. Let $Y = \{v_1, v_2, \ldots, v_n\}$ be the set of $n$ perturbed nodes where each perturbation satisfies* (6)*. Then $(\Phi_{uu}^{-1}(z))_{pq} \neq 0$ implies that $p$ and $q$ are neighbors in the perturbed kin graph $\mathcal{G}_Y^M$.*

*Proof:* First, we will describe the structure of $\Phi_{uu}(z)$. For compact notation, we will often drop the $z$ arguments.

For $p = 1, \ldots, N$, if $p$ is not a perturbed node, set $h_p(z) = 1$ and $d_p(z) = 0$. With this notation, (6) implies that the entries of $\Phi_{uu}$ are given by:

$$(\Phi_{uu})_{pq} = \begin{cases} h_p (\Phi_{xx})_{pq} h_q^{\sim} & \text{if } p \neq q \\ h_p (\Phi_{xx})_{pp} h_p^{\sim} + d_p & \text{if } p = q \end{cases}$$

When $p \neq q$, there is no $d$ term because the perturbations were assumed to be independent.

In matrix notation, we have that:

$$\Phi_{uu} = H \Phi_{xx} H^{\sim} + \sum_{k=1}^{n} D_{v_k}$$

where $H$ is the diagonal matrix with entries $h_p$ on the diagonal and $D_{v_k}(z) = b_{v_k} d_{v_k}(z) b_{v_k}^T$ where $b_{v_k}$ is the canonical unit vector with 1 at entry $v_k$.

Set $\Psi_0 = H \Phi_{xx} H^{\sim}$. For $k = 0, \ldots, n-1$, we can inductively define the matrices:

$$\Psi_{k+1} = \Psi_k + b_{v_{k+1}} d_{v_{k+1}} b_{v_{k+1}}^T \qquad (14)$$

For $k = 1, \ldots, n$ let $Y_k = \{v_1, \ldots, v_k\}$ and let $\mathcal{G}_{Y_k}^M$ be the perturbed kin graph constructed by adding edges $\{i,j\}$ to the original kin graph if there is a path from $i$ to $j$ whose intermediate nodes are all in $Y_k$.

We will inductively prove the following claim: For $k = 1, \ldots, n$, if $(\Psi_k^{-1})_{pq} \neq 0$, then $p$ and $q$ are neighbors in $\mathcal{G}_{Y_k}^M$. Proving this claim is sufficient to prove the theorem, since $\Psi_n = \Phi_{uu}$ and $Y_n = Y$.

First we focus on the $k = 1$ case. Using the Woodbury Matrix identity we have, $\Psi_1^{-1} = \Psi_0^{-1} - \Gamma_1$, where $\Gamma_1 := (\Psi_0^{-1} b_{v_1} b_{v_1}^T \Psi_0^{-1}) \Delta_{v_1}^{-1}$ and $\Delta_{v_1} = d_{v_1}^{-1} + b_{v_1}^T \Psi_0^{-1}(z) b_{v_1}$ is a scalar. Therefore, $(\Psi_1^{-1})_{pq} = (\Psi_0^{-1})_{pq} - (\Gamma_1)_{pq}$.

If $(\Psi_1^{-1})_{pq} \neq 0$ then at least one of the conditions (i) $(\Psi_0^{-1})_{pq} \neq 0$ or (ii) $(\Gamma_1)_{pq} \neq 0$ must hold.

Suppose that $(\Psi_0^{-1})_{pq} \neq 0$. Then $(H^{-\sim}(z) \Phi_{xx}^{-1} H^{-1}(z))_{pq} \neq 0$. As $H$ is diagnoal it follows that $(\Phi_{xx}^{-1})_{pq} \neq 0$. From Corollary 1, it follows that $p$ and $q$ are neighbors in $\mathcal{G}^M$. Thus $p$ and $q$ are neighbors in $\mathcal{G}_{Y_1}^M$.

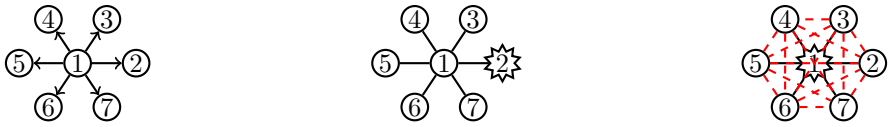
Suppose that $(\Gamma_1)_{pq} \neq 0$. Then it follows that $(\Psi_0^{-1} b_{v_1} b_{v_1}^T \Psi_0^{-1})_{pq} \Delta_{v_1}^{-1} \neq 0$. Thus $(\Psi_0^{-1} b_{v_1})_p \neq 0$ and $(b_{v_1}^T \Psi_0^{-1})_q \neq 0$. Noting that $\Psi_0 = H \Phi_{xx} H^{\sim}$, it follows that , $(\Phi_{xx}^{-1})_{pv_1} \neq 0$ and $(\Phi_{xx}^{-1})_{v_1 q} \neq 0$. From Corollary 1 it follows that $\{v_1, p\}$ and $\{v_1, q\}$ are edges in the kin graph $\mathcal{G}^M$. Thus, there is a path from $p$ to $q$ whose only intermediate node is $v_1 \in Y_1$. Thus, $p, q$ are neighbors in $\mathcal{G}_{Y_1}^M$ and the claim is verified for $k = 1$.

Now assume that the claim holds for some $k > 1$. Combining the Woodbury matrix identity with (14) implies that

$$\Psi_{k+1}^{-1} = \Psi_k^{-1} - \Gamma_{k+1}$$

where $\Gamma_{k+1} = \Psi_k^{-1} b_{v_{k+1}} b_{v_{k+1}}^T \Psi_k^{-1} \Delta_{v_{k+1}}^{-1}$ and $\Delta_{v_{k+1}} = d_{v_{k+1}}^{-1} + b_{v_{k+1}}^T \Psi_k^{-1}(z) b_{v_{k+1}}$ is a scalar.

As before, if $(\Psi_{k+1}^{-1})_{pq} \neq 0$, then either $(\Psi_k^{-1})_{pq} \neq 0$ or $(\Gamma_{k+1})_{pq} \neq 0$.




(a) Broadcast Architecture (b) Corrupted leaf (c) Corrupted Hub

Fig. 3: This figure shows an extreme example of the effect of data corruption of even a single node. 3a shows the original directed graph. 3b shows that even if the leaf is corrupted there are no erroneous links introduced. But if the hub is corrupted as shown in 3c then all the nodes become spuriously correlated.

If $(\Psi_k^{-1})_{pq} \neq 0$, then the induction hypothesis implies that $p$ and $q$ are neighbors in $\mathcal{G}^M_{Y_k}$. Since $Y_k \subset Y_{k+1}$, it follows that $p$ and $q$ are neighbors in $\mathcal{G}^M_{Y_{k+1}}$.

If $(\Gamma_{k+1})_{pq} \neq 0$, then as in the $k = 1$ case, we must have that $(\Psi_k^{-1})_{pv_{k+1}} \neq 0$ and $(\Psi_k^{-1})_{v_{k+1}q} \neq 0$. This implies that $\{p, v_{k+1}\} \in A^M_{Y_k}$ and $\{v_{k+1}, q\} \in A^M_{Y_k}$. Thus, either $p$ and $v_{k+1}$ are kins in the original kin graph, or there is a path from $p$ to $v_{k+1}$ whose intermediate nodes are in $Y_k$. Similarly, for $q$ and $v_{k+1}$. It follows that there is a path from $p$ to $q$ whose nodes are in $Y_{k+1}$, and thus $p$ and $q$ are neighbors in $\mathcal{G}^M_{Y_{k+1}}$. The claim, and thus the theorem, are now proved. ■

## V. Simulation Results

Power spectrum estimates were computed after $10^4$ simulation time steps. The estimated spectra were then averaged over 100 trials. The red boxes indicate the erroneous links introduced as a result of the network perturbation in addition to the the links in the true kin topology as indicated by the black boxes. For both the networks, the sequences $e_i$ are zero mean white Gaussian noise.

### A. Star Topology

The transfer function for each link is $z^{-1}$.

*1) Corrupted Leaf:* The perturbation considered here is

$$\tau_2[n] = \begin{cases} n-3, & \text{with probability } 0.65 \\ n-1, & \text{with probability } 0.35. \end{cases}$$

$\Phi^{-1}_{uu}(z) =$

| | | | | | | |
|---|---|---|---|---|---|---|
| 15.02 | 0.14 | 1.49 | 1.49 | 1.50 | 1.50 | 1.45 |
| 0.14 | 1.74 | 0.05 | 0.05 | 0.05 | 0.05 | 0.04 |
| 1.49 | 0.05 | 2.36 | 0.05 | 0.06 | 0.06 | 0.06 |
| 1.49 | 0.05 | 0.05 | 2.35 | 0.06 | 0.05 | 0.06 |
| 1.50 | 0.05 | 0.06 | 0.06 | 2.36 | 0.05 | 0.05 |
| 1.50 | 0.05 | 0.06 | 0.05 | 0.05 | 2.36 | 0.05 |
| 1.45 | 0.04 | 0.06 | 0.06 | 0.05 | 0.05 | 2.34 |

As predicted by Theorem 2, perturbation of Node 2 for this architecture does not introduce any erroneous links. See Figure 3b.

*2) Corrupted Hub:* The perturbation considered here is

$$\tau_1[n] = \begin{cases} n-2, & \text{with probability } 0.75 \\ n-4, & \text{with probability } 0.25. \end{cases}$$

Theorem 2 predicts that perturbing the central node could introduce erroneous links between all of the nodes. See Figure 3c.

$\Phi^{-1}_{uu}(z) =$

| | | | | | | |
|---|---|---|---|---|---|---|
| 5.08 | 0.40 | 0.40 | 0.40 | 0.39 | 0.39 | 0.38 |
| 0.40 | 2.07 | 0.27 | 0.27 | 0.27 | 0.26 | 0.27 |
| 0.40 | 0.27 | 2.08 | 0.27 | 0.27 | 0.28 | 0.27 |
| 0.40 | 0.27 | 0.27 | 2.07 | 0.27 | 0.27 | 0.27 |
| 0.39 | 0.27 | 0.27 | 0.27 | 2.07 | 0.27 | 0.27 |
| 0.39 | 0.26 | 0.28 | 0.27 | 0.27 | 2.08 | 0.27 |
| 0.38 | 0.27 | 0.27 | 0.27 | 0.27 | 0.27 | 2.08 |

### B. Chain Topology

The chain topology in Figure 4 is considered. The transfer functions are: between nodes 1 and 2, $1.2+0.9z^{-1}$, between nodes 2 and 3, $1 + 0.2z^{-1}$, between nodes 3 and 4, $1 - 0.9z^{-1} + 0.3z^{-2}$ and then for the last link $z^{-1}$. Figure 4 In the simulations, nodes 2 and 3 are simultaneously corrupted with the temporal uncertainty models

$$\tau_2[n] = \begin{cases} n-1, & \text{with probability } 0.83 \\ n-2, & \text{with probability } 0.17. \end{cases}$$

$$\tau_3[n] = \begin{cases} n-2, & \text{with probability } 0.85 \\ n-4, & \text{with probability } 0.15. \end{cases}$$

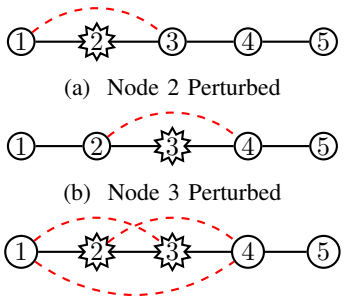
$\Phi^{-1}_{uu}(z) =$

| | | | | |
|---|---|---|---|---|
| 4.23 | 0.54 | 0.12 | 0.25 | 0.05 |
| 0.54 | 1.20 | 0.16 | 0.13 | 0.02 |
| 0.12 | 0.16 | 1.06 | 0.12 | 0.02 |
| 0.25 | 0.13 | 0.12 | 2.22 | 0.90 |
| 0.05 | 0.02 | 0.02 | 0.90 | 1.42 |

Perturbation of 2 adds a false relationship between 1 and 3. In addition, perturbation of 3 introduces erroneous relations between the nodes 1 and 4 as well as between 2 and 4. Thus the erroneous relationships could arise between any nodes that are kins of 3 including the already introduced false kins of 3. Despite this cascaded effect the erroneous links remain local in the sense that the dependency of 5 is unaffected.

## VI. Conclusions and Future Work

We established that network topology reconstruction from corrupt data streams can result in erroneous links/correlations between the nodes. Intuitively, these erroneous links appear with reason that since there is a loss of information occurring due to the corruption of an agent, the nodes dependent on the corrupted agent now has to rely on measurements from other agents. Through the examples considered we also observed that certain architectures and certain nodal corruption has more drastic effects compared to others. Therefore, it is important to rigorously study and come up with topological




(a) Node 2 Perturbed

(b) Node 3 Perturbed

(c) Nodes 2 and 3 Perturbed

Fig. 4: This figure shows how multiple perturbations can lead to a cascade effect as predicted by Theorem 2. Here the original kin graph is a chain. 4a and 4b show the erroneous edges that can arise from perturbing a single node. If nodes 2 and 3 are both perturbed, then another erroneous link between 1 and 4 must be added.

designs that can reduce the spread of corruption. Moreover, motivated with real-world applications such as IOT, it is imperative to devise stronger models for data-uncertainty for networks that can have highly non-linear inter-dependencies.

## Appendix I
## Random Delays

Let us examine the auto-correlation of $u_i$. Let $k = n - m$ and let $k \neq 0$. Then

$$R_{u_iu_i}(n-m) = \mathbb{E}\left[x_i[\tau_i[n]]x_i[\tau_i[m]]\right] = {p_i}^2 R_{x_ix_i}(k) + (1-p_i)^2 R_{x_ix_i}(k) + p_i(1-p_i)R_{x_ix_i}(k+k_2-k_1) + p_i(1-p_i)R_{x_ix_i}(k+k_1-k_2). \quad (15)$$

$$R_{u_iu_i}(0) = \mathbb{E}\left[x_i[\tau_i[n]]x_i[\tau_i[n]]\right] = p_i R_{x_ix_i}(0) + (1-p_i)R_{x_ix_i}(0) = R_{x_ix_i}(0). \quad (16)$$

The PSD is given by,

$$\Phi_{u_iu_i}(z) = h_i(z)\Phi_{x_ix_i}(z)h_i(z^{-1}) + d_i, \quad (17)$$

where $h_i(z) = p_i z^{k_1} + (1-p_i)z^{k_2}$ and $d_i = R_{x_ix_i}(0)\left\{1 - {p_i}^2 - (1-p_i)^2\right\} - p_i(1-p_i)[R_{x_ix_i}(k_1-k_2) + R_{x_ix_i}(k_2-k_1)]$.
The cross correlation between $u_i$ and $x_i$ is given by,

$$R_{u_ix_i}(k) = \mathbb{E}[x_i[\tau_i[n]]x_i[m]] = p_i R_{x_ix_i}(k+k_1) + (1-p_i)R_{x_ix_i}(k+k_2). \quad (18)$$

Perform $\mathcal{Z}(\cdot)$ on both sides of (18) to obtain the cross spectrum of $u_i$ and $x_i$:

$$\Phi_{u_ix_i}(z) = h_i(z)\Phi_{x_ix_i}(z). \quad (19)$$

## Packet Dropping Links

Let $k = n - m$. Substituting (5) into the definition of the cross correlation gives:

$$R_{u_ix_i}(n-m) = \mathbb{E}[u_i[n]x_i[m]] = p_1 R_{x_ix_i}(n-m) + (1-p_i)R_{u_ix_i}(n-m-1).$$

Taking Z-transform of both sides and rearranging gives:

$$\Phi_{u_ix_i}(z) = \frac{p_i}{1 - z^{-1}(1-p_i)}\Phi_{x_ix_i}(z) = h_i(z)\Phi_{x_ix_i}(z).$$

Now we examine the PSD of $u_i$. By a similar argument from the derivation of $R_{u_ix_i}$, for $k \neq 0$, we have that

$$R_{u_iu_i}(k) = \frac{p_i^2}{1-(1-p_i)^2}R_{x_ix_i}(k) + \frac{p_i(1-p_i)}{1-(1-p_i)^2}\left(R_{u_ix_i}(k-1) + R_{x_iu_i}(k+1)\right). \quad (20)$$

Similar to the derivation of (16), we have that $R_{u_iu_i}(0) = R_{x_ix_i}(0)$.

It can be shown that taking Z-transform of $R_{u_iu_i}(k)$ yields:

$$\Phi_{u_iu_i}(z) = h_i(z)\Phi_{x_ix_i}(z)h_i(z^{-1}) + d_i(z)$$

where $d_i(z)$ is same as in (9). We remark that both the random delay uncertainty and the packet drop uncertainty result in $u_i$ with $\phi_{u_iu_i}(z)$ and $\phi_{u_ix_i}(z)$ of the form (6).

Note that we have omitted the proof that $u_i$ and $x_i$ are JWSS due to space constraints.

## References


[1] W. R. Shadish, T. D. Cook, and D. T. Campbell, *Experimental and quasi-experimental designs for generalized causal inference*. Wadsworth Cengage learning, 2002.

[2] K. Soltesz, K. Van Heusden, G. A. Dumont, T. Hägglund, C. Petersen, N. West, and J. M. Ansermino, "Closed-loop anesthesia in children using a pid controller: A pilot study," in *IFAC Conference on Advances in PID Control*, 2012.

[3] Y. Wang, T. Tan, and K.-F. Loe, "Video segmentation based on graphical models," in *Computer Vision and Pattern Recognition, 2003. Proceedings. 2003 IEEE Computer Society Conference on*, vol. 2. IEEE, 2003, pp. II–335.

[4] R. Porreca, S. Drulhe, H. Jong, and G. Ferrari-Trecate, "Structural identification of piecewise-linear models of genetic regulatory networks," *Journal of Computational Biology*, vol. 15, no. 10, pp. 1365–1380, 2008.

[5] S. Ross, "The arbitrage theory of capital asset pricing," *Journal of Economic Theory*, vol. 13, pp. 341–360, 1976.

[6] E. Bullmore and O. Sporns, "Complex brain networks: graph theoretical analysis of structural and functional systems," *Nature Reviews on Neuroscience*, vol. 10, pp. 186–198, 2009.

[7] H. Sundmaeker, P. Guillemin, P. Friess, and S. Woelfflé, "Vision and challenges for realising the internet of things," *Cluster of European Research Projects on the Internet of Things, European Commision*, 2010.

[8] D. Datla, X. Chen, T. Tsou, S. Raghunandan, S. Hasan, S., H. Reed, J., B. Dietrich, C., T. Bose, B. Fette, and H. Kim, J., "Wireless distributed computing: a survey of research challenges," *IEEE Communications Magazine*, vol. 50, no. 1, pp. 144–152, 2012.

[9] D. Materassi and M. V. Salapaka, "On the problem of reconstructing an unknown topology via locality properties of the wiener filter," *IEEE transactions on automatic control*, vol. 57, no. 7, pp. 1765–1777, 2012.

[10] K. B. Petersen and M. S. Pedersen, "The matrix cookbook," 2012. [Online]. Available: http://www2.imm.dtu.dk/pubdb/p.php?3274.