



**HAL**  
open science

## Source identification of propagating waves inside a network

Franck Plouraboue, Pierre Uszes, Romain Guibert

► **To cite this version:**

Franck Plouraboue, Pierre Uszes, Romain Guibert. Source identification of propagating waves inside a network. *IEEE Transactions on Network Science and Engineering*, 2022, 9 (3), pp.1437-1450. 10.1109/TNSE.2022.3144647 . hal-03872595v2

**HAL Id: hal-03872595**

**<https://hal.science/hal-03872595v2>**

Submitted on 28 Nov 2022

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Source identification of propagating waves inside a network

F. Plouraboué, P. Uszes & R. Guibert

**Abstract**—The localization of short events arising within a network subsequently leading to wave propagation into it, is of interest in many applications. This work extends [1] which demonstrated the identifiability of a source from two detectors in a  $N$  nodes graph. We show that, rather than a source, a boundary condition identification is also possible and demonstrate a generalized unicity result. Furthermore, we extend the identification algorithm proposed in [1] to an arbitrary number of sensors, and estimate its complexity which depends on sensors number  $N_s$  and time discretisation  $N_t$ . Increasing detectors number increases source identification robustness to noise up to a cut-off number being a small fraction of  $N$ . This cut-off detector density for efficiency in noise reduction is of practical significance. We also analyze and discuss the method sensitivity to total recording time  $T_e$ , sampling frequency, and signal to noise ratio. Finally we propose a pre-sectorisation to improve the systematic exploration algorithm proposed in [1] and we show a  $O(N)$  drop in complexity leading to a  $O(N^3)$  cost for the source identification. Several tests of the method on model and real graphs confirm and support the presented results.



## 1 INTRODUCTION

### 1.1 General context and motivations

In various contexts like water-hammer waves within pipes networks, river tidal waves in river basin, explosives acoustic waves in gas pipes tubing, electrical waves in power-distribution networks, it is interesting to locate the origin of waves propagating inside networks [1], [1], [2], [3], [4], [5], [6], [7]. The simplest example of spatio-temporal source identification is the localization of a sudden breakage or leakage event within a single pipe [8]. In this case, the origin of wave event propagating at velocity  $c$ , in both  $\pm x$  directions, is searched from two distinct sensors recording the signal at some location along the pipe axis  $x$ . Standard signal processing methods can then be applied [8], looking for the highest correlation between temporally translated signals at each sensor location within the two-dimensional half-space  $x - t$ . Nevertheless, even in the most simple networks, wave rebounds, or multiple path-waves produce as many temporal patterns for the signal recorded at the sensor location as possible wave-paths from event to sensor preventing simple signal processing methods to provide an answer to locate sources. Networks are hereby defined by their non-oriented graph  $G(V, E)$ , vertex (node) set  $V$ , and edge set  $E$ , having a symmetric adjacency matrix. The wave propagation along each edge is supposed to arise at constant (possibly non-homogeneous) speed. The wave spreading into the network, i.e its bouncing and propagation at each node are determined by the graph Laplacian matrix  $\Delta_G$  which sets the network's harmonic coupling between vibrating modes [2], [3]. For general networks four important questions are interesting to address :

- (i) For a (non dissipative) wave equation is there a way to find the –possibly unambiguous– source location in time and space from recording the induced signal at some sensor location ?
- (ii) How is the identification method threatened by noise and possibly improved by increasing the number of sensors ?
- (iii) Is there some possibly optimal way of distributing the sensors inside the networks so as to improve both identification and its reliability ?
- (iv) Finally, is it possible to improve the computational efficiency of source identification ?

Here we consider ‘single event’ localization only as in [1], [9], [10], [11]. Question (i) has been addressed in [1] for wave-source identification in the case of two sensors. We provide an extended description of their approach in §2.1. Furthermore, in the context of diffusive processes [9] found an identifiability condition associated with the need for signal not having the same time to reach any detector from two distinct sources. More general source location algorithms for heterogeneous propagation models have been proposed [10], [11].

Issue (ii) has been partly addressed in [1] in the case of two sensors for intern-type event identifications. Also in [9] the influence of random time-delay has been investigated for some discrete value of noise. Question (iii) has been qualitatively addressed in some contributions, using networks concepts such as node's centrality [12], [13], [14]. Concerning (iv), in the context of diffusive processes over graphs source identification complexity has received dedicated attention in the literature [9], [15], [16], [17]. In the case of breadth-first tree structure of first arrival, [17] have provided a  $O(N^3)$  cost method for source identification. Furthermore, in the case of diffusion propagation within a network [9] have proposed a  $O(N^2 \ln(N))$  method using a time-reversal backward spreading algorithm.

- F. Plouraboué, P. Uszes & R. Guibert are at Institut de Mécanique des Fluides de Toulouse, IMFT, Université de Toulouse, CNRS, Toulouse, France.  
E-mail: fplourab@imft.fr
- Toulouse University

Manuscript received 04-Feb-2021

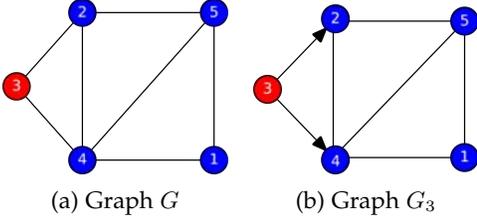


Fig. 1: Illustration of graphs resulting from intern type or boundary condition type sources. (a) Non-oriented unchanged graph  $G$  with inner source node #3 (red) as in [1]. (b) Oriented graph resulting from a boundary condition source (BC-event) node #3 (red). Notation  $G_3$  is introduced to describe this locally oriented graph at node #3 distinct from graph  $G$ .

In some applications, the ‘event’ can be described by the presence of an external source term appearing at given nodes of the network (subsequently called an intern-type event), as in the case examined in [1]. Nevertheless, in other applications, the event results in a structural change of the network itself for which a given node becomes exclusively triggered by an external source impinging its interactions with neighbor nodes of the network. This is found in network pipe disruption, for which the external pressure is suddenly applied at one node, whatever the pressure nearby. Those boundary-condition driven events (subsequently called BC-events) produce a subtle change in the graph topology illustrated in Figure 1a-b. They suddenly locally shift the graph topology from locally un-directed (Cf Figure 1a) to locally directed (Cf Figure 1b). This subtle change has a mostly unfortunate consequence : the graph mathematical structure is changed by BC-events, and the results provided in [1] no more apply in this case. Hence, the aim of this contribution is to

- Extend the identification algorithm proposed in [1] to BC-events identification with an arbitrary number of sensors.
- Analyze and discuss the method sensitivity to sensor’s recording frequency, total recording time  $T_e$  and signal to noise ratio.
- Propose a pre-sectorisation method reducing the numerical cost of source identification.

Section 1.2 provides some useful definitions. Section 2 provides the mathematical analysis of two distinct extensions of [1], the details of which is first presented in §2.1. Then, the source identification proof for an arbitrary number of sensors is detailed in §2.2. Lately, §2.3.1 provides the theory for BC-events identification. Section §3 describes the algorithms for both intern type or BC-type source identification. In §4 practical improvements are discussed, among which a pre-sectorisation method to improve the algorithmic cost is presented in 4.3. §5 provides numerical tests and illustrations of source identification in various real-world graphs. A nomenclature is provided in Appendix A.

## 1.2 Some concepts and definitions

**Definition 1 (Source).** A source is the location  $s \in V$ , within the vertex set of the graph, of a time-varying perturbation, an event, arising at a single node. The event can

either be of intern type or of Boundary-Condition type (i.e BC-type). The type of the event also define the type of the source, i.e intern source or BC-source.

**Definition 2 (Event).** An event is defined by the couple  $(\lambda(t), \mathbf{S})$ , where  $\lambda(t)$  is an applied, compact support, temporal perturbation  $\lambda(t) \in \mathcal{L}^2$  and  $\mathbf{S}$  a source vector, with  $\mathbf{S} \in \{0,1\}^N$ ,  $\mathbf{S} = \hat{\mathbf{s}}$ , i.e, having components  $S_k = \delta_{sk}$ . The event can either be of intern type or of Boundary-Condition type (i.e BC-type).

**Definition 3 (Intern event).** An event is said of intern type if its contribution is only added from the contribution of neighbors of this node.

**Definition 4 (Boundary condition event).** An event is said of Boundary Condition type, i.e BC-type, if the imposed value at the node cancels the contributions of neighbors. A BC-event at vertex  $v$  leads to a locally  $v$ -oriented graph  $G_v$ .

**Definition 5 ( $v$ -oriented graph  $G_v$ ).** A BC-event at vertex  $v$  locally imposes a graph orientation of every edges pointing at  $v$ , outward from  $v$ .

**Definition 6 (Strategic set).** . Given  $v^n$  the  $n$ th eigenvector of the graph  $G$  Laplacian, a set  $K \subset V$  is defined as strategic if  $\forall n \in V = [1, N], \exists k \in K / v_k^n \neq 0$ .

The origin for the need for strategic set is illustrated in figure 2. For the same graph as in figure 1a, figure 2 showcases

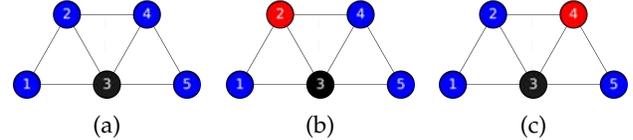


Fig. 2: Illustration of non-detectability of some sources resulting from the graph symmetries. (a) For sensor in node #3 (in black), a source in node #2 (b) or #4 (c) (in red) produce the same signal. Hence node #3 only does not provide a strategic set, i.e  $\xi = \{3\}$  is not strategic.

internal symmetries. From figure 2, it is then easy to find that some sources are not possible to distinguish with limited detector set. The fact that node #3 of figure 2 does not permit to built a strategic set can also be found from computing the eigenvector set of the graph Laplacian given in (1), where we showcase the third line (associated with node #3 highlighted in red) which violates Definition 6 for strategic set.

$$\mathfrak{V} = \begin{pmatrix} 0.4472 & -0.6533 & 0.5 & -0.2706 & -0.2236 \\ 0.4472 & -0.2706 & -0.5 & 0.6533 & -0.2236 \\ \mathbf{0.4472} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0.8944} \\ 0.4472 & 0.2706 & -0.5 & -0.6533 & -0.2236 \\ 0.4472 & 0.6533 & 0.5 & 0.2706 & -0.2236 \end{pmatrix} \quad (1)$$

## 2 THEORETICAL ANALYSIS

This section details Caputo et al. [1] results in 2.1 the extension of which for an arbitrary number of sensors is detailed in 2.2. Furthermore, the extension of Theorem 1 for uniqueness of BC-events identification is given in 2.3.

## 2.1 Detailed presentation of [1]

### 2.1.1 Evaluation of $\mathbf{X}$ .

The wave amplitude at time  $t \in [0, T_e]$  at each vertex of graph  $G(V, E)$ , with  $\text{card}(V) = N$ , is represented by vector  $\mathbf{X}$

$$\mathbf{X}(t) = (X_1(t), X_2(t), \dots, X_N(t))^T \quad (2)$$

With  $T_e$  the final recording time.  $\Delta_G$  is the Laplacian matrix of non-oriented graph  $G$ . The governing equation of  $\mathbf{X}(t)$  is given by the d'Alembert operator

$$\begin{cases} \ddot{\mathbf{X}} - \Delta_G \mathbf{X} = \lambda(t) \mathbf{S} & \text{for } t \in [0, T_e] \\ X(0) = \mathbf{a} \in \mathbb{R}^N & \text{and } \dot{X}(0) = \mathbf{b} \in \mathbb{R}^N \end{cases} \quad (3)$$

where  $\mathbf{a}$  et  $\mathbf{b}$  are the initial state and velocity of the system.  $\lambda$  is the time-variation of a single source localized in single node (vertex)  $s \in V$ , with  $\mathbf{S}_i = \delta_{is}$ . Model (3) represents forced harmonic vibrations into a network resembling coupled vibrating strings, with a time-varying force imposed at node  $s$ .

It is assumed that  $(\lambda(t), \mathbf{S})$  are chosen in admissible event sets defined by

$$\mathcal{A}_s := \begin{cases} \exists T^0 \in [0, T] / \lambda(t) = 0 \forall t \in [T^0, T] \\ \mathbf{S} = (s_1, s_2, \dots, s_N)^T \\ \text{with } s_k \in \{0, 1\} \ k \in V, \ \& \sum_{k=1}^N s_k = 1 \end{cases} \quad (4)$$

This admissible set provides the condition for vector source  $\mathbf{S}$  to represent only one node among all possible ones, and the need for the forcing term to have compact support so that it disappears after finite time  $T^0$ . The observation operator is defined in [1] as an application from the event space to observation space :

$$M(\lambda(t), \mathbf{S}) = \{\mathbf{D}(t) \text{ with } t \in [0, T]\} \quad (5)$$

where the detection vector  $\mathbf{D}$ , is chosen at two sensors locations  $(i, j) \in V$  [1]. This abstract observation operator is useful to consider since it provides the formal link between the event, i.e the source node where some transient forcing occurs, and the later observation of the resulting disturbances within the network at some pre-defined locations provided by  $\mathbf{D}$ . It is assumed that the eigenvalue of  $\Delta_G$  are distinct and ordered

$$0 = -\omega_1^2 > -\omega_2^2 > \dots > -\omega_N^2 \quad (6)$$

Since  $\Delta_G$  is a symmetric matrix of a non-oriented graph, its eigenvectors  $\{\mathbf{v}^1, \mathbf{v}^2, \dots, \mathbf{v}^N\}$  produce an orthogonal base of  $\mathbb{R}^N$ , so that

$$\mathbf{X}(t) = \sum_{n=1}^N y_n(t) \mathbf{v}^n \quad \text{with } y_n(t) = \langle \mathbf{X}(t), \mathbf{v}^n \rangle \forall n \in [1, N] \quad (7)$$

**Lemma 1.** For  $T^* \in (T_0, T)$ , and  $X$  solution of

$$\begin{cases} \ddot{X}(t) - \Delta_G X(t) = 0 \forall t \in (T^*, T) \\ X(T) \in \mathbb{R}^N \text{ et } \dot{X}(T) \in \mathbb{R}^N \end{cases} \quad (8)$$

$\xi = (k_1, k_2, \dots, k_{N_s})$  a strategic set. Then,

$$x_{k_1}(t) = \dots = x_{k_{N_s}}(t) = 0, \forall t \in (T^*, T) \Rightarrow X(T) = \dot{X}(T) = 0 \quad (9)$$

This lemma states that, if the restriction of the wave amplitude at strategic set (Cf definition 6 and figure 2 for

illustration) location  $X|_{\xi}(t)$  is zero at any time after the event to occur, then, the wave amplitude and velocity are zero everywhere at final time. Using lemma 1 [1] have shown that

**Theorem 1.** If  $(\lambda, \mathbf{S})$  is part of admissible events set, i.e  $(\lambda, \mathbf{S}) \in \mathcal{A}_s$ ,  $\mathbf{X}(t)$  solution of (3) given at nodes of strategic set  $\xi$ , i.e  $\mathbf{X}|_{\xi}(t)$ , permit to find a unique  $\mathbf{X}(T)$ .

The symmetric Laplacian matrix  $\Delta_G$  admit an orthogonal base of eigenvectors. Considering the linear homogeneous differential problem  $\ddot{\mathbf{X}} - \Delta_G \mathbf{X} = 0$ , it is possible to find the general solution projected in the  $\Delta_G$  eigenvector base  $\mathcal{B}$

$$\begin{aligned} \mathbf{X}(t) &= (y_1(T) + (t - T)y_1'(T)) \mathbf{v}^1 + \\ &\sum_{n=2}^N \left( y_n(T) \cos(\omega_n(t - T)) + \frac{y_n'(T)}{\omega_n} \sin(\omega_n(t - T)) \right) \mathbf{v}^n \end{aligned} \quad (10)$$

This projection permits to infer the signal everywhere in the network, from recording it in a limited number of sensors located in  $\xi \subset V$ . Components  $y_n(T)$  for  $n \in [1, N]$  permits to build the mode amplitude vector  $\mathbf{y}(T) \in \mathbb{R}^N$  which gives, for each mode, the oscillation amplitude. Sensors at positions  $\xi$  indeed permit to evaluate the mode amplitude vector  $\mathbf{y}(T)$  and its time derivative  $\dot{\mathbf{y}}(T)$ , encapsulated into  $2N$ -vector  $\mathbf{Y}(T) \equiv [\mathbf{y}(T), \dot{\mathbf{y}}(T)]$ — as in [18]— from linear regression with observations. This linear regression uses the detection vector  $\mathbf{D} \equiv X_k(t)$ ,  $k \in \xi$  with  $\text{card}(\xi) = N_s$ , sampled at discrete value  $t_m = m\Delta t$ , with  $m = 0, \dots, M$

$$\begin{aligned} \mathbf{D} &= (X_{k_1}(T_0), \dots, X_{k_1}(t_m), \dots, X_{k_1}(t_M), \\ &X_{k_2}(T_0), \dots, X_{k_2}(t_m), \dots, X_{k_2}(t_M), \\ &X_{k_{N_s}}(T_0), \dots, X_{k_{N_s}}(t_m), \dots, X_{k_{N_s}}(t_M))^T \end{aligned} \quad (11)$$

The regression aims minimizing  $\|\mathbf{X}|_{\xi} - \mathbf{D}\|^2$ , the difference between detection and the wave amplitude prediction at sensor location  $\xi$ , which is a function of unknown wave amplitudes  $\mathbf{Y}(t)$ . Since  $\mathbf{X}$  fulfills a linear dependence with  $\mathbf{Y}(t)$ , a matrix  $\mathbf{R}$  can be built such that,  $\mathbf{X}|_{\xi} = \mathbf{R}\mathbf{Y}$  with regression matrix  $\mathbf{R}$  equals to

$$\begin{bmatrix} \mathbf{v}_1^1 & (T_0 - T) \mathbf{v}_1^1 & \cos(\omega_2(T_0 - T)) \mathbf{v}_1^2 & \frac{\sin(\omega_2(T_0 - T))}{\omega_2} \mathbf{v}_1^2 & \dots & \cos(\omega_N(T_0 - T)) \mathbf{v}_1^N & \frac{\sin(\omega_N(T_0 - T))}{\omega_N} \mathbf{v}_1^N \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \mathbf{v}_1^1 & (t_M - T) \mathbf{v}_1^1 & \cos(\omega_2(t_M - T)) \mathbf{v}_1^2 & \frac{\sin(\omega_2(t_M - T))}{\omega_2} \mathbf{v}_1^2 & \dots & \cos(\omega_N(t_M - T)) \mathbf{v}_1^N & \frac{\sin(\omega_N(t_M - T))}{\omega_N} \mathbf{v}_1^N \\ \mathbf{v}_2^1 & (T_0 - T) \mathbf{v}_2^1 & \cos(\omega_2(T_0 - T)) \mathbf{v}_2^2 & \frac{\sin(\omega_2(T_0 - T))}{\omega_2} \mathbf{v}_2^2 & \dots & \cos(\omega_N(T_0 - T)) \mathbf{v}_2^N & \frac{\sin(\omega_N(T_0 - T))}{\omega_N} \mathbf{v}_2^N \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \mathbf{v}_2^1 & (t_M - T) \mathbf{v}_2^1 & \cos(\omega_2(t_M - T)) \mathbf{v}_2^2 & \frac{\sin(\omega_2(t_M - T))}{\omega_2} \mathbf{v}_2^2 & \dots & \cos(\omega_N(t_M - T)) \mathbf{v}_2^N & \frac{\sin(\omega_N(t_M - T))}{\omega_N} \mathbf{v}_2^N \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \mathbf{v}_{N_s}^1 & (T_0 - T) \mathbf{v}_{N_s}^1 & \cos(\omega_2(T_0 - T)) \mathbf{v}_{N_s}^2 & \frac{\sin(\omega_2(T_0 - T))}{\omega_2} \mathbf{v}_{N_s}^2 & \dots & \cos(\omega_N(T_0 - T)) \mathbf{v}_{N_s}^N & \frac{\sin(\omega_N(T_0 - T))}{\omega_N} \mathbf{v}_{N_s}^N \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \mathbf{v}_{N_s}^1 & (t_M - T) \mathbf{v}_{N_s}^1 & \cos(\omega_2(t_M - T)) \mathbf{v}_{N_s}^2 & \frac{\sin(\omega_2(t_M - T))}{\omega_2} \mathbf{v}_{N_s}^2 & \dots & \cos(\omega_N(t_M - T)) \mathbf{v}_{N_s}^N & \frac{\sin(\omega_N(t_M - T))}{\omega_N} \mathbf{v}_{N_s}^N \end{bmatrix} \quad (12)$$

Then, the linear regression permits to find unknown wave amplitudes  $\mathbf{Y}$  from

$$\mathbf{Y} = [\mathbf{R}^T \cdot \mathbf{R}]^{-1} \mathbf{R}^T \mathbf{D} \quad (13)$$

and then  $\mathbf{X}(t)$  is found from using (13) in (10).

### 2.1.2 Source location detection:

in order to find the source location, the first step consists in simplifying the temporal ordinary differential equation

from projecting it into a suitable base. Since this temporal problem fulfills a Sturm-Liouville problem

$$\begin{cases} \ddot{\varphi}_m(t) = \mu_m \varphi_m(t) \forall t \in [0, T] \\ \varphi_m(0) = \varphi_m(T) = 0 \end{cases} \quad (14)$$

with eigenfunctions  $\varphi_m$  associated with eigenmodes  $\mu_m$

$$\varphi_m(t) = \sqrt{\frac{2}{T}} \sin\left(\frac{m\pi}{T}t\right) \text{ et } \mu_m = \left(\frac{m\pi}{T}\right)^2 \quad (15)$$

Functions  $\{\varphi_m\}$  form a complete orthogonal set of  $\mathcal{L}^2(0, T)$  equipped with scalar product  $\langle f, g \rangle = \int_0^T f(t)g(t)dt$ . This base is then used to simplify the temporal o.d.e problems of the wave amplitude (3). Multiplying (3) by  $\varphi_m$  and integrating twice by part leads to

$$-(\Delta_G + \mu_m I)\overline{\mathbf{X}}_m = \overline{\lambda}_m \mathbf{S} + \mathbf{P}_m, \quad (16)$$

Where  $\overline{\mathbf{X}}_m \equiv \langle \mathbf{X}, \varphi_m \rangle$ , resp.  $\overline{\lambda}_m \equiv \langle \lambda, \varphi_m \rangle$ , are the projection of  $\mathbf{X}$ , resp.  $\lambda$ , over  $\varphi_m$ , and  $\mathbf{P}_m = \dot{\varphi}_m(T)\mathbf{X}(T) - \dot{\varphi}_m(0)\mathbf{X}(0)$ . Defining matrix

$$\mathbf{A}_m \equiv \Delta_G - \mu_m \mathbb{I}, \quad (17)$$

(16) reads

$$-\mathbf{A}_m \overline{\mathbf{X}}_m = \overline{\lambda}_m \mathbf{S} + \mathbf{P}_m. \quad (18)$$

(18) thus represents the second step of the source location strategy for providing a necessary linear algebraic problem for projected mode  $m$ . In this formulation of the amplitude projection dynamic, the right-hand-side term  $\mathbf{P}_m$  takes care of the influence of initial and final conditions of both base-function velocity  $\dot{\varphi}_m$  and amplitude vector  $\mathbf{X}$ . Then [1] search for (18) at every nodes expect those where sensors are located. Denoting  $(i, j)$  the set of the two sensors ([1] consider the case of two sensors only), and removing the corresponding components on the right-hand-side leads to

$$-\mathbf{A}_m^{(i,j)} \overline{\mathbf{X}}_m^{(i,j)} = \overline{\lambda}_m \mathbf{S} + \mathbf{P}_m^{(i,j)} \quad (19)$$

$\mathbf{A}_m^{(i,j)}$  is thus a  $N \times N - 2$  matrix obtained from removing the two columns corresponding to the sensors located at nodes  $(i, j)$ .  $\overline{\mathbf{X}}_m^{(i,j)}$  is vector  $\overline{\mathbf{X}}_m$  without components  $i$  &  $j$ . Finally

$$\mathbf{P}_m^{(i,j)} = \dot{\varphi}_m(T)\mathbf{X}(T) - \dot{\varphi}_m(0)\mathbf{X}(0) + \mathbf{A}_m \cdot \hat{\mathbf{i}} \overline{X}_i + \mathbf{A}_m \cdot \hat{\mathbf{j}} \overline{X}_j \quad (20)$$

where  $\hat{\mathbf{i}}$  &  $\hat{\mathbf{j}}$  are the unit vectors having components  $\hat{\mathbf{i}}_k = \delta_{ki}$  &  $\hat{\mathbf{j}}_k = \delta_{kj}$  with  $\delta_{ki}$  the Kronecker symbol. Theorem 2 established in [1] states the unicity of the source :

**Theorem 2.** For  $m \in \mathbb{N}^*$ ,  $N \times N$  matrix  $\mathbf{A}_m$  defined in (17), and  $\xi = \{i, j\}$ , if

- $\xi$  is strategic
- $\langle \lambda, \varphi_m \rangle \neq 0$
- All  $N - 2 \times N - 2$  sub-matrix obtained from extracting the two columns  $i$  et  $j$  and any two lines  $s_1^*$  &  $s_2^*$  of  $\mathbf{A}_m$  denoted  $\mathbf{A}_{s_1^*, s_2^*}^{(i,j)}$  are invertible.

The data collected in  $\xi = \{i, j\}$  permit to uniquely identify  $\mathbf{S}$  and  $\lambda$ .

Considering (19), let us describe how obtaining  $\mathbf{S}$ . First, it is important to mention that this is an over-determined problem having  $N$  equations for  $N - 2$  unknown. Source identification is obtained from using

$$-\mathbf{A}_m^{(i,j)} \overline{\mathbf{X}}_m^{(i,j)} = \mathbf{P}_m^{(i,j)} \quad (21)$$

This system is tested many times, by each time removing any couple of distinct lines  $s_1^*$  and  $s_2^*$  and solving (from invertibility condition of  $\mathbf{A}_{s_1^*, s_2^*}^{(i,j)}$ )

$$-\mathbf{A}_{s_1^*, s_2^*}^{(i,j)} \overline{\mathbf{X}}_m^{(i,j)} = \mathbf{P}_{s_1^*, s_2^*}^{(i,j)} \quad (22)$$

If two distinct coupled of lines  $(s_1^*, s_2^*)$  and  $(s_1'^*, s_2'^*)$  give the same  $X_m^{(i,j)}$ , then problem (19) is equivalent to (22), so that, it means that the source term is absent from (19). The repeated line in  $(s_1^*, s_2^*)$ ,  $(s_1'^*, s_2'^*)$  gives the location of the source  $s$ , i.e  $s$  is precisely the repeated line. The detailed algorithm of the source localization proposed in [1] is given in Algorithm 1.

Finally, we do not provide further details about the temporal form of  $\lambda(t)$ , since the algorithm provided in [1] is strongly sensitive to added noise on the detector signal. It is interesting to mention the following limits of [1] :

- Strategic set is limited to two sensors, whilst it is possible to exhibit some graph without possible strategic set built from two sensor location only.
- The regression matrix  $R$  can be rank deficient (the condition for regression is  $N_s(T_e - T_0)/\Delta T > 2N$ ).
- No indication are given for choosing  $T_e$  the total recorded time and which mode  $\mu_m$  is best for projection.
- The proposed method is very sensitive to added noise.
- The algorithmic cost is  $O(N^4)$ .

## 2.2 Extension to an arbitrary number of sensors

[1] uses the working hypothesis that every sub-matrix  $\mathbf{A}_{s_1^*, s_2^*}^{(i,j)}$  are invertible for every couple of line  $s_1^*, s_2^* \in V$  suppressed from  $\mathbf{A}_m^{(i,j)}$ . This hypothesis can be weakens so as to provide a broader generality to the source identification results when considering an arbitrary number of sensors located in  $\xi \subset V$  for which the corresponding hypothesis that every sub-matrix  $\mathbf{A}_{s_1^*, s_2^*}^\xi$  is invertible become a strong assumption. In fact some numerical test with  $N_s = 3$  sensors are sufficient to exhibit non-invertible sub-matrix  $\mathbf{A}_{s_1^*, s_2^*}^\xi$ . Hence there is a need for a technical softening of Theorem 2 hypothesis. In fact, the invertibility of sub-matrix  $\mathbf{A}_{s_1^*, s_2^*}^\xi$  is not strictly necessary. Let us define  $\overline{\mathbf{X}}_m|_{\xi^\perp}$  the restriction of  $\overline{\mathbf{X}}_m$  over the nodes where there is no sensors on  $V$ , and  $\mathbf{P}_{s_1^*, s_2^*}^\xi|_{\xi^\perp}$ , the similar restriction of  $\mathbf{P}_{s_1^*, s_2^*}^\xi$ . One needs to find a solution for  $\overline{\mathbf{X}}_m|_{\xi^\perp}$  for the generalization of (22) to an arbitrary number of sensors, i.e,

$$-\mathbf{A}_{s_1^*, s_2^*}^\xi \overline{\mathbf{X}}_m|_{\xi^\perp} = \mathbf{P}_{s_1^*, s_2^*}^\xi|_{\xi^\perp}. \quad (23)$$

The over-constrained solution of (23) is unique when  $Rank(\mathbf{A}_{s_1^*, s_2^*}^\xi) = N - N_s$ . With this condition, one can then compute

$$\overline{\mathbf{X}}_m|_{\xi^\perp} = -[(\mathbf{A}_{s_1^*, s_2^*}^\xi)^\top \cdot \mathbf{A}_{s_1^*, s_2^*}^\xi]^{-1} \mathbf{A}_{s_1^*, s_2^*}^\xi \mathbf{P}_{s_1^*, s_2^*}^\xi|_{\xi^\perp}. \quad (24)$$

Otherwise, if  $Rank(\mathbf{A}_{s_1^*, s_2^*}^\xi) \neq N - N_s$  several possible solutions for  $\overline{\mathbf{X}}_m|_{\xi^\perp}$  could lead to disregard the true source  $s$ , resulting in non-uniqueness of the method. We now precisely state and prove this result.

**Algorithm 1** Caputo et al. [1] algorithm for source identification

**Require:** Single component Graph  $G(V, E)$  and its adjacency matrix  $A_G$

```

1: Chose two sensors location  $i, j \in V$ 
2: if  $(i, j)$  is a strategic set then
3:   Get  $X_i(t)$  and  $X_j(t)$  for  $t \in [0, T]$ 
4:   Construct the Graph Laplacian matrix  $\Delta_G$  from  $A_G$ 
5:   Compute the set of eigenvectors  $\mathbf{v}^n \in \mathfrak{V}$  & eigenvalues  $\omega_n \in \Omega$  for  $n \in [1, N]$  of  $\Delta_G$ 
6:   Compute the 2N-vector  $\mathbf{Y}(T) \equiv [\mathbf{y}(T), \dot{\mathbf{y}}(T)]$ , for  $T > T_0$ , from inversion of the linear regression matrix  $\mathbf{R} : \mathbf{Y} = [\mathbf{R}^\top \cdot \mathbf{R}]^{-1} \mathbf{R}^\top \mathbf{D} \triangleright \text{Cf (12) for } \mathbf{R} \text{ and (60) for } \mathbf{D}$ 
7:   Then, evaluate  $\mathbf{X}(T)$ 
8:   for each  $s^* \in V$  do
9:     for each  $q \in V \setminus \{s^*\}$  do
10:      Evaluate  $\mathbf{A}_{s^*,q}^{(i,j)}$  &  $\mathbf{P}_{s^*,q}^{(i,j)} \triangleright \text{Cf (22) for } \mathbf{A}_{s^*,q}^{(i,j)}$  for  $\mathbf{P}_{s^*,q}^{(i,j)}$ 
11:      Compute  $\bar{\mathbf{X}}_{s^*,q}^{(i,j)} = -[\mathbf{A}_{s^*,q}^{(i,j)\top} \mathbf{A}_{s^*,q}^{(i,j)}]^{-1} \mathbf{A}_{s^*,q}^{(i,j)\top} \mathbf{P}_{s^*,q}^{(i,j)}$ 
12:     end for
13:     for each  $q, q' \in V \setminus \{s^*\}$  do
14:       Compute  $E_{s^*,q,q'}^{(i,j)} = \|\bar{\mathbf{X}}_{s^*,q}^{(i,j)} - \bar{\mathbf{X}}_{s^*,q'}^{(i,j)}\|^2$ 
15:     end for
16:     Estimate  $E_{s^*}^{(i,j)} = \sum_{q,q' \in V \setminus \{s^*\}} E_{s^*,q,q'}^{(i,j)}$ 
17:   end for
18:   for each  $s^* \in V$  do
19:     Find  $s^{*sol}$  such that  $\min_{s^* \in V} E_{s^*}^{(i,j)} = E_{s^{*sol}}^{(i,j)}$ 
20:   end for
21:   return  $s_{sol}^*$ 
22: end if

```

**Lemma 2. (Extension of theorem 2)** Let us define  $S^{*\perp}$  the set of nodes identified as free of source. Noting  $S^* = V \setminus S^{*\perp}$  the set of candidate nodes for source, If  $\forall (s_1^*, s_2^*) \in S^*$ ,  $\text{Rank}(\mathbf{A}_{s_1^*, s_2^*}^\xi) = N - \text{card}(\xi)$ , then a single observation  $M(\lambda, \mathbf{S})$  uniquely identifies the source node  $s$ .

**Proof**

Let  $(\lambda^{(1)}, \mathbf{S}^{(1)})$  &  $(\lambda^{(2)}, \mathbf{S}^{(2)})$  be two possible events associated with two sources candidates  $s_1^*$  &  $s_2^* \in S^*$ . These events leads to solutions  $\mathbf{X}^{(1)}$  &  $\mathbf{X}^{(2)}$ . Let us define  $\mathbf{Z} = \mathbf{X}^{(2)} - \mathbf{X}^{(1)}$ . Using definition (5) of the observation operator,

$$M(\lambda^{(1)}, \mathbf{S}^{(1)}) = M(\lambda^{(2)}, \mathbf{S}^{(2)}) \Rightarrow \forall k \in \xi, \forall t \in (0, T), Z_k(t) = 0. \quad (25)$$

Using lemma 2.2 of [1] leads to  $\mathbf{Z}(T) = \dot{\mathbf{Z}}(T) = 0$ . Furthermore, from a unique observation  $\mathbf{D}$  at  $\xi$  nodes, then the projection of mode decomposition  $\bar{\mathbf{Z}}_m$  of  $\mathbf{Z}$  on  $\xi$  is zero, i.e,  $\forall k \in \xi, \bar{Z}_k = 0$ , so that

$$\mathbf{A}_m^\xi \bar{\mathbf{Z}} | \xi^\perp = \lambda_m^{(2)} S^{(2)} - \lambda_m^{(1)} S^{(1)}. \quad (26)$$

Removing lines  $s_1^*$  &  $s_2^*$  from  $\mathbf{A}_m^\xi$  leads to

$$\mathbf{A}_{s_1^*, s_2^*}^\xi \bar{\mathbf{Z}} | \xi^\perp = 0. \quad (27)$$

Since  $s_1^* \& s_2^* \in S^*$ ,  $\text{Rank}(\mathbf{A}_{s_1^*, s_2^*}^\xi) = N - N_s$ ,  $\text{Rank}((\mathbf{A}_{s_1^*, s_2^*}^\xi)^\top \cdot \mathbf{A}_{s_1^*, s_2^*}^\xi) = N - N_s$  and one gets

$$\bar{\mathbf{Z}} | \xi^\perp = 0. \quad (28)$$

Hence,

$$(\lambda_m^{(2)}, \mathbf{S}^{(2)}) = (\lambda_m^{(1)}, \mathbf{S}^{(1)}), \quad (29)$$

so that, in particular,

$$\mathbf{S}^{(2)} = \mathbf{S}^{(1)}, \quad (30)$$

and uniqueness follows.

## 2.3 Extension to boundary condition sources (BC-events)

### 2.3.1 New definitions

Let us now consider the conditions needed for BC-events to be identified. We need to define the generalization of strategic set for BC-events, that we name **BC-strategic set**. In this case, we need to enlarge the frame so as to consider all possible  $s$ -directed graphs  $G_s \in \mathcal{G}_{BC}$  obtained from undirected graph  $G$ . This first necessitates the direct generalization of the previous notion of strategic set, although, insufficient in this case, that we thus name *Component-discriminant set* for BC-event, and define as

**Definition 7. (Component-discriminant set)** A set  $K_1 \subset V$  is component-discriminant if  $\forall G_v \in \mathcal{G}_{BC}, \forall \omega \in \Omega, \exists k \in K_1 / \mathbf{v}_k^{\omega|G_v} \neq 0$ .

Furthermore, we need to define a *vector-discriminant set*

**Definition 8. (vector-discriminant set)** A set  $K_2 \subset V$  is vector-discriminant if  $\forall G_{v_1} \& G_{v_2} \in \mathcal{G}_{BC}, \forall \omega \in \Omega_{v_1} \cap \Omega_{v_2}$ ,

$$\mathbf{v}^{\omega|G_{v_1}} \neq \mathbf{v}^{\omega|G_{v_2}} \Rightarrow \exists k \neq k' \in K_2 / \frac{\mathbf{v}_k^{\omega|G_{v_1}}}{\mathbf{v}_{k'}^{\omega|G_{v_1}}} \neq \frac{\mathbf{v}_k^{\omega|G_{v_2}}}{\mathbf{v}_{k'}^{\omega|G_{v_2}}}.$$

so that a **BC-strategic set** can be defined

**Definition 9. (BC-strategic set)** A set  $K \subset V$  is BC-strategic if it is both component-discriminant and vector-discriminant.

### 2.3.2 BC-event uniqueness conditions

For BC-events at candidate node  $s^*$ , the corresponding Laplacian matrix  $\Delta_{G_{s^*}}$  associated with  $G_{s^*}$  is distinct from  $\Delta_G$  which is associated with the graph  $G$ . Let us denote  $\Delta_{G \setminus s^*}$  the  $(N-1) \times (N-1)$  sub-matrix built from the  $N \times N$  matrix  $\Delta_G$  by removing its  $s^*$ th column and its  $s^*$ th line. Let us define  $\mathbf{C}_{\setminus s^*}^{\Delta_G}$  as being the  $s^*$ th column of  $\Delta_G$  without its component  $s^*$ . Then, we define  $\Delta_{G_{s^*}}$  built from  $\Delta_{G \setminus s^*}$  as

$$\Delta_{G_{s^*}} = \left( \begin{array}{c|c} \Delta_{G \setminus s^*} & \mathbf{C}_{\setminus s^*}^{\Delta_G} \\ \hline \mathbf{0} & 0 \end{array} \right), \quad (31)$$

where, on the last line,  $\mathbf{0}$  denotes the  $(N-1)$  component zero vector. This choice is made so as to permit the directed-graph forcing of BC-event from dynamic system

$$\begin{pmatrix} \ddot{\mathbf{X}}_{\setminus s^*} \\ \ddot{\mathbf{X}}_{s^*} \end{pmatrix} + \begin{pmatrix} \Delta_{G \setminus s^*} & \mathbf{C}_{\setminus s^*}^{\Delta_G} \\ \hline 0 & 0 \end{pmatrix} \begin{pmatrix} \mathbf{X}_{\setminus s^*} \\ X_{s^*} \end{pmatrix} = \begin{pmatrix} 0 \\ \vdots \\ 0 \\ \lambda(t) \end{pmatrix}, \quad (32)$$

so that the  $s^*$ -neighbor nodes have no influence on the forced node, which, on the contrary, can propagate the forcing wave into the graph. This is why the event forcing  $\lambda(t)$  drives the source node acceleration term  $\ddot{X}_{s^*}$  in (32). Unicity is secured by initial conditions  $\mathbf{X}(0)$  et  $\dot{\mathbf{X}}(0)$ . Let us now consider the eigenvalues and eigenvectors of matrix  $\Delta_{G_{s^*}}$ .

First, let us consider the  $\mathbb{R}^N$  unit-vector  $\mathbf{1}_N = \frac{1}{\sqrt{N}}[1, 1, \dots, 1]^\top$ . From the structure of  $\Delta_{G_{s^*}}$  given in (31), one can easily find that  $\mathbf{1}_N$  is a zero-eigenvector of  $\Delta_{G_{s^*}}$  since the unit-vector  $\mathbf{1}_{N-1}$  in  $\mathbb{R}^{N-1}$  is also the zero-eigenvector of  $\Delta_{G \setminus s^*}$ . So that (the first) zero-eigenvector of  $\Delta_{G_{s^*}}$  is  $\mathbf{v}^{1|G_{s^*}} = \frac{1}{\sqrt{N}}[1, 1, \dots, 1]^\top$ .

Then, let us consider the sub-matrix  $\Delta_{G \setminus s^*}$  of  $\Delta_{G_{s^*}}$  having an orthogonal eigenvectors base (since it is symmetric) denoted  $\mathbf{v}^{n'|G \setminus s^*}$  with  $n' \in [1, N-1]$ . Let us then built a new base for  $\Delta_{G_{s^*}}$  denoted  $\mathbf{v}^{n|G_{s^*}} = [\mathbf{v}^{n'+1|G \setminus s^*}, 0]^\top$  with  $n = n' + 1 \forall n' \in [1, N-1]$ .  $\mathbf{v}^{n|G_{s^*}}$  form an orthogonal base of  $\mathbb{R}^{N-1}$ . They also are eigenvectors of  $\Delta_{G_{s^*}}$ . Moreover  $\mathbf{v}^{1|G_{s^*}}$  is linearly dependant with  $\mathbf{v}^{n|G_{s^*}}$  with  $n \in [2, N]$ , so that  $\text{Im}(\mathbf{v}^{2|G_{s^*}}, \dots, \mathbf{v}^{N|G_{s^*}})$  spans  $\mathbb{R}^{N-1}$ . Hence, the (non-orthogonal) eigenvector base of  $\Delta_{G_{s^*}}$  is known.

**Lemma 3. (Extension of theorem 1)** If  $(\lambda, \mathbf{S})$  is part of admissible BC-events set, i.e  $(\lambda, \mathbf{S}) \in \mathcal{A}_s$ ,  $\mathbf{X}(t)$  solution of (3) given at nodes of BC-strategic set  $\xi$ , i.e  $\mathbf{X}|_\xi(t)$ , permit to find a unique  $\mathbf{X}(T)$ .

### Proof

Let us consider two solutions  $\mathbf{X}^{s_1^*}(t)$  &  $\mathbf{X}^{s_2^*}(t)$  associated with two source candidates  $s_1^*$  &  $s_2^*$ . From (10) the differences between these solutions, reads

(33)

If those solutions provide de same detected signal on  $\xi$ ,  $\forall k \in \xi$ , we have

$$0 = \left( y_1^{s_2^*}(T) + (t-T)y_1^{s_2^*}(T) \right) \mathbf{v}_k^{1|G_{s_2^*}} \quad (34)$$

$$- \left( y_1^{s_1^*}(T) + (t-T)y_1^{s_1^*}(T) \right) \mathbf{v}_k^{1|G_{s_1^*}} \quad (35)$$

$$+ \sum_{n=2}^N y_n^{s_2^*}(T) \cos \left( \omega_n^{G_{s_2^*}}(t-T) \right) \mathbf{v}_k^{n|G_{s_2^*}} \quad (36)$$

$$- \sum_{n=2}^N y_n^{s_1^*}(T) \cos \left( \omega_n^{G_{s_1^*}}(t-T) \right) \mathbf{v}_k^{n|G_{s_1^*}} \quad (37)$$

$$+ \sum_{n=2}^N \frac{y_n^{s_2^*}(T)}{\omega_n^{G_{s_2^*}}} \sin \left( \omega_n^{G_{s_2^*}}(t-T) \right) \mathbf{v}_k^{n|G_{s_2^*}} \quad (38)$$

$$- \sum_{n=2}^N \frac{y_n^{s_1^*}(T)}{\omega_n^{G_{s_1^*}}} \sin \left( \omega_n^{G_{s_1^*}}(t-T) \right) \mathbf{v}_k^{n|G_{s_1^*}} \quad (39)$$

For  $\tau \in \mathbb{R}^+$  integrating (39) over  $[T, T+\tau]$ , and dividing by  $\tau^2$  whilst taking the limit  $\tau \rightarrow \infty$ , one gets

$$\dot{y}_1^{s_1^*}(T) \mathbf{v}_k^{1|G_{s_1^*}} = \dot{y}_1^{s_2^*}(T) \mathbf{v}_k^{1|G_{s_2^*}} \quad (40)$$

But yet, in every case, the eigenvector associated to the nul eigenvalue is  $(1, 1, \dots, 1)^\top / \sqrt{N}$ , so that  $\mathbf{v}^{1|G_{s_1^*}} = \mathbf{v}^{1|G_{s_2^*}} = (1, 1, \dots, 1)^\top / \sqrt{N}$ , and thus  $\dot{y}_1^{s_1^*}(T) = \dot{y}_1^{s_2^*}(T)$ .

Furthermore, integrating (39) in interval  $[T, T+\tau]$ , whilst later on dividing by  $\tau$  in the limit  $\tau \rightarrow \infty$ , one gets

$$y_1^{s_1^*}(T) \mathbf{v}_k^{1|G_{s_1^*}} = y_1^{s_2^*}(T) \mathbf{v}_k^{1|G_{s_2^*}} \quad (41)$$

Thus we have,  $y_1^{s_1^*}(T) = y_1^{s_2^*}(T)$ .

So that, from (39), one is left with  $\forall k \in \xi$

$$0 = \sum_{n=2}^N y_n^{s_2^*}(T) \cos \left( \omega_n^{G_{s_2^*}}(t-T) \right) \mathbf{v}_k^{n|G_{s_2^*}} \quad (42)$$

$$- \sum_{n=2}^N y_n^{s_1^*}(T) \cos \left( \omega_n^{G_{s_1^*}}(t-T) \right) \mathbf{v}_k^{n|G_{s_1^*}} \quad (43)$$

$$+ \sum_{n=2}^N \frac{y_n^{s_2^*}(T)}{\omega_n^{G_{s_2^*}}} \sin \left( \omega_n^{G_{s_2^*}}(t-T) \right) \mathbf{v}_k^{n|G_{s_2^*}} \quad (44)$$

$$- \sum_{n=2}^N \frac{y_n^{s_1^*}(T)}{\omega_n^{G_{s_1^*}}} \sin \left( \omega_n^{G_{s_1^*}}(t-T) \right) \mathbf{v}_k^{n|G_{s_1^*}} \quad (45)$$

So be it  $n' \in [2, N]$ , let us project over  $\omega_{n'}^{G_{s_1^*}}$  mode from multiplying (45) by  $\cos \left( \omega_{n'}^{G_{s_1^*}}(t-T) \right)$ . Then, integrating over  $[T, T+\tau]$ , whilst, later-on dividing by  $\tau$  and taking the limit  $\tau \rightarrow \infty$ , two distinct cases are possible

$$\begin{cases} \forall k \in \xi, y_{n'}^{s_1^*}(T) \mathbf{v}_k^{n'|G_{s_1^*}} = 0 \text{ if } \omega_{n'}^{G_{s_1^*}} \notin \Lambda(G_{s_2^*}) \\ \forall k \in \xi, y_{n'}^{s_1^*}(T) \mathbf{v}_k^{n'|G_{s_1^*}} = y_{n'}^{s_2^*}(T) \mathbf{v}_k^{n'|G_{s_2^*}} \text{ if } \omega_{n'}^{G_{s_1^*}} \in \Lambda(G_{s_2^*}) \end{cases} \quad (46)$$

The first case imply that  $y_{n'}(T)^{s_1^*} = 0$  because  $k \in \xi$  which is composant-discriminant, i.e  $\exists k$  such that  $\mathbf{v}_k^{n'|G_{s_1^*}} \neq 0$ . Furthermore, since  $\omega_{n'}^{G_{s_2^*}} \notin \Lambda(G_{s_1^*})$ , one also finds that  $y_{n'}(T)^{G_{s_2^*}} = 0$ .

The second case implies that  $\exists k \neq k' \in \xi$  such that  $\frac{\omega_{k'}^{G_{s_1^*}}}{\omega_{k'}^{G_{s_1^*}}} = \frac{\omega_{k'}^{G_{s_2^*}}}{\omega_{k'}^{G_{s_2^*}}}$ . But yet, if one supposes that  $\mathbf{v}^{\omega_{k'}^{G_{s_1^*}}} \neq \mathbf{v}^{\omega_{k'}^{G_{s_2^*}}}$  this implies  $\exists k \neq k' \in \xi$ , being vector-discriminant, so that  $\frac{\omega_{k'}^{G_{s_1^*}}}{\omega_{k'}^{G_{s_1^*}}} \neq \frac{\omega_{k'}^{G_{s_2^*}}}{\omega_{k'}^{G_{s_2^*}}}$ . Thus, reductio ad impossibile,  $\mathbf{v}^{\omega_{k'}^{G_{s_1^*}}} = \mathbf{v}^{\omega_{k'}^{G_{s_2^*}}}$ . We thus conclude that  $y_{n'}^{s_1^*}(T) = y_{n'}^{s_2^*}(T)$ . Thus, finally since for all  $n \in [1, N]$  one gets

$$\begin{cases} y_n^{s_1^*}(T) = y_n^{s_2^*}(T) = 0, \\ \text{or} \\ y_n^{s_1^*}(T) = y_n^{s_2^*}(T) \ \& \ \mathbf{v}^{n|G_{s_1^*}} = \mathbf{v}^{n|G_{s_2^*}}. \end{cases} \quad (47)$$

Unicity follows, i.e

$$\mathbf{X}^{s_2^*}(T) - \mathbf{X}^{s_1^*}(T) = \sum_{n=1}^N y_n(T) \left( \mathbf{v}^{n|G_{s_2^*}} - \mathbf{v}^{n|G_{s_1^*}} \right) = \mathbf{0}. \quad (48)$$

## 3 ALGORITHMS FOR GENERAL SOURCE IDENTIFICATION

### 3.1 Building general strategic sets

Here we describe the algorithm for finding a robust strategic set, both including the one defined in [1] for intern-type

events, and BC-strategic sets (Cf Def 9) associated with BC-events. Algorithm 2 details a general method to build this strategic set.

---

**Algorithm 2** Find Strategic set  $\xi$  for general sources and boundary conditions

---

**Require:** Single component Graph  $G(V, E)$  and its adjacency matrix  $A_G$   
 Construct the Graph Laplacian matrix  $\Delta_G$  from  $A_G$

- 2: **for** each  $v \in V$  **do**  
     Construct the matrix  $\Delta_{G_v}$
- 4:   Compute eigenvectors  $\mathbf{v}_v^n \in \mathfrak{V}_v$  & eigenvalues  $\omega_v^n \in \Omega_v$  for  $n \in [1, N]$  of  $\Delta_{G_v}$   
     **end for**
- 6: **for** each  $v_1, v_2 \in V$  **do**  
     **if**  $\Omega_{v_1} \cap \Omega_{v_2} \neq \{\emptyset\}$  **then**
- 8:     **if**  $\nexists (k, k') \in \xi$  such that  $\frac{\mathbf{v}_{v_1, k}^n}{\mathbf{v}_{v_1, k'}^n} \neq \frac{\mathbf{v}_{v_2, k}^n}{\mathbf{v}_{v_2, k'}^n}$  **then**  
        Find  $k \neq k' \in V$  such that  $\frac{\mathbf{v}_{v_1, k}^n}{\mathbf{v}_{v_1, k'}^n} = \frac{\mathbf{v}_{v_2, k}^n}{\mathbf{v}_{v_2, k'}^n}$
- 10:      $\xi = \xi \cup \{k, k'\}$   
     **end if**
- 12: **end if**  
     **end for**

---

### 3.2 General intern-source or BC-source identification

We now describe how the identification of general source of unknown type (intern or BC type) can be done. The identification method proceed from two main steps. A pre-computational one associated with

- Graph properties evaluation (Adjacency matrix, Laplacian eigenvectors)
- Regression computation to find modes amplitudes  $[\mathbf{Y}(T), \dot{\mathbf{Y}}(T)]$  in (7).

Then, a the source term identification step is performed where the prediction over the state  $\mathbf{X}(t)$  from various source candidates is compared with the one observed on detectors locations. Algorithm 3 provides the details of the source identification implementation for unknown type source to be detected. This algorithm branch into the previous algorithm 1 of [1], in the special case of an identified intern type source. It also necessitates the computation of all possible Laplacian  $\Delta_{G_{s^*}}$  of every possible  $s^*$ -directed graphs  $G_{s^*}$ .

## 4 EXTENSIONS AND IMPROVMENT FOR SOURCE IDENTIFICATION EFFICIENCY

### 4.1 Noise sensitivity

This section specifically consider the noise sensitivity of the source detection method.

#### 4.1.1 Noise sensitivity of linear regression

As previously explained in §3.2, the detection method proceed from two main steps. The (pre-computational) regression step, and later on, the source term identification step. Let us first analyse the noise sensitivity of the regression step. We consider that the detector signal decomposes into

---

**Algorithm 3** General internal source or Boundary Condition (BC) identification

---

**Require:** Single component Graph  $G(V, E)$  and its adjacency matrix  $A_G$   
 Construct the Graph Laplacian matrix  $\Delta_G$  from  $A_G$

- 2: Pre-sectorise  $V$  to get candidates set  $S^* \subset V$  ▷ Cf Alg. 4 for pre-sectorisation
- for** each candidate  $s^* \in S^*$  **do**
- 4:   Construct matrix  $\Delta_{G_{s^*}}$  ▷ Cf (4) for  $\Delta_{G_{s^*}}$   
     Compute eigenvectors  $\mathbf{v}_v^n \in \mathfrak{V}_v$  & eigenvalues  $\omega_v^n \in \Omega_v$  for  $n \in [1, N]$  of  $\Delta_{G_v}$
- 6:   Estimate  $\mathbf{R}_{s^*}$   
     Compute  $\mathbf{Y}_{s^*} = [\mathbf{R}_{s^*}^\top \cdot \mathbf{R}_{s^*}]^{-1} \mathbf{R}_{s^*}^\top \mathbf{D}$  with 2N-vector  $\dot{\mathbf{Y}}_{s^*} \equiv [\mathbf{y}_{s^*}(T), \dot{\mathbf{y}}_{s^*}(T)]$
- 8:   Estimate  $\|\mathbf{R}_{s^*} \cdot \mathbf{Y}_{s^*} - \mathbf{D}\|^2 = r_{s^*}$   
     **end for**
- 10: Find  $s_{sol}^*$  such that  $\min_{s^* \in S^*} r_{s^*} = r_{s_{sol}^*}$   
     Compute  $\mathbf{Y} = [\mathbf{R}^\top \cdot \mathbf{R}]^{-1} \mathbf{R}^\top \mathbf{D}$  with 2N-vector  $\dot{\mathbf{Y}} \equiv [\mathbf{y}(T), \dot{\mathbf{y}}(T)]$
- 12: Estimate  $\|\mathbf{R} \cdot \mathbf{Y} - \mathbf{D}\|^2 = r$   
     **if**  $r < r_{s_{sol}^*}$  **then**
- 14:   pertub\_type\_id=intern source type  
     Apply steps 8 to 21 of Alg. 1 for  $s^* \in S^*$
- 16: **else**  
     pertub\_type\_id=BC
- 18:   **return**  $s_{sol}^*$  pertub\_type\_id  
     **end if**

Even if this computation is much more costly than the single graph Laplacian computation  $\Delta_G$  needed for intern-type source, it only represents a pre-treatment cost of the algorithm, which might also possibly be stored, so that it is not needed each time an identification is performed. Furthermore,  $s^*$  source candidates do not need to span the entire set  $V$ , but only a part of it, i.e  $S^* \subset V$ , a pre-sectorisation issue later-on discussed in §4.3.

the true signal  $\mathbf{D}_T$  plus some random noise vector  $\eta$  (associated with the discrete sampling of a continuous random process  $\eta(t)$ )

$$\mathbf{D} = \mathbf{D}_T + \eta \quad (49)$$

Using the linearity of relation (13) leads to

$$\mathbf{Y} = \mathbf{Y}_T + \delta \mathbf{Y} = [\mathbf{R}^\top \cdot \mathbf{R}]^{-1} \mathbf{R}^\top \mathbf{D}_T + [\mathbf{R}^\top \cdot \mathbf{R}]^{-1} \mathbf{R}^\top \eta \quad (50)$$

i.e

$$\delta \mathbf{Y} = [\mathbf{R}^\top \cdot \mathbf{R}]^{-1} \mathbf{R}^\top \eta \quad (51)$$

This extra-term  $\delta \mathbf{Y}$  is then responsible for an extra-prediction  $\delta \mathbf{X}$ , which, from the linearity of  $\mathbf{X}$  with  $\mathbf{Y}$  in (10) reads

$$\delta \mathbf{X}(t) = \left( \delta y_1(T) + (t - T) \delta y_1(T) \right) \mathbf{v}^1 + \quad (52)$$

$$\sum_{n=2}^N \left( \delta y_n(T) \cos(\omega_n(t - T)) + \frac{\delta y_n(T)}{\omega_n} \sin(\omega_n(t - T)) \right) \mathbf{v}^n$$

We now examine how the extra-term  $\delta \mathbf{X}(t)$  affects the second identification step of the algorithm.

#### 4.1.2 Noise impact of source identification

The next step involves the projection of the state vector  $\mathbf{X}(t)$  into the chosen mode  $m$ , denoted  $\bar{\mathbf{X}}_m$ . Considering (19) associated with the restriction of  $\bar{\mathbf{X}}_m$  over non-detector nodes denoted  $\bar{\mathbf{X}}_m|\xi^\perp$ ,

$$-\mathbf{A}_m^\xi \bar{\mathbf{X}}_m|\xi^\perp = \bar{\lambda}_m \mathbf{S} + \mathbf{P}_m|\xi^\perp, \quad (53)$$

One can realize that the right-hand-side term  $\mathbf{P}_m|\xi^\perp$  also fulfills an extra-noise term that we now wish to evaluate. From using the generalization of (20) to an arbitrary number of sensors

$$\mathbf{P}_m|\xi^\perp = \dot{\varphi}_m(T)\mathbf{X}(T) - \dot{\varphi}_m(0)\mathbf{X}(0) + \sum_{k \in \xi} \mathbf{A}_m \cdot \hat{\mathbf{k}} \bar{X}_k, \quad (54)$$

one is able to compute the extra-noise contribution of  $\mathbf{P}_m|\xi^\perp$

$$\delta \mathbf{P}_m|\xi^\perp = \dot{\varphi}_m(T)\delta \mathbf{X}(T) - \dot{\varphi}_m(0)\delta \mathbf{X}(0) + \sum_{k \in \xi} \mathbf{A}_m \cdot \hat{\mathbf{k}} \delta \bar{X}_k, \quad (55)$$

where obviously,  $\delta \bar{\mathbf{X}}_m \equiv \langle \delta \mathbf{X}, \varphi_m \rangle$  with  $\delta \bar{X}_k$  the  $k$ th component of  $\delta \bar{\mathbf{X}}_m$ . Hence during the identification procedure associated with the inversion of the linear problem (53), this extra term in  $\delta \mathbf{P}$  leads, from (24) to an extra-prediction given by :

$$\delta \bar{\mathbf{X}}_m|\xi^\perp = -[(\mathbf{A}_{s_1^*, s_2^*}^\xi)^\top \cdot \mathbf{A}_{s_1^*, s_2^*}^\xi]^{-1} \mathbf{A}_{s_1^*, s_2^*}^\xi \delta \mathbf{P}_{s_1^*, s_2^*}|\xi^\perp \quad (56)$$

This extra-prediction is responsible for the noise sensitivity of the detection algorithm. (56) and (55) combined with (52) and finally (51) provide the linear relationship between this extra-prediction  $\delta \bar{\mathbf{X}}_m|\xi^\perp$  and the noise vector  $\eta$ , i. e. the sensitivity matrix of the state prediction over noise. Since, this sensitivity matrix involves the  $\bar{X}_k$ , we now study how its sensitivity can be attenuated when increasing the time-projection support  $T$ , or conversely increasing the recording frequency.

## 4.2 Sampling frequency and detectors number effect on $\bar{X}_k$ and $\delta X_k(T)$

### 4.2.1 Sampling frequency effect

First, let us note that the noise is weakened by the scalar product for

$$\delta \bar{X}_k = \int_0^T \phi_m(t) \delta X_k(t) dt, \quad (57)$$

with, as mentioned just above,  $\delta X_k(t)$  linearly related to vector  $\eta(t)$ . For a centered, normalized, decorrelated random noise process  $\eta_k(t)$ , i.e

$$\text{Cov}(\eta_k(t)\eta_k(t')) = \text{E}(\eta_k(t)\eta_k(t')) = \sigma_N^2 \delta(t - t'), \quad (58)$$

the central limit theorem provides an asymptotic limit for the behaviour of  $\delta \bar{X}_k$  as  $T \rightarrow \infty$  such as  $E(\delta \bar{X}_k^2) \sim 1/T$ . Similarly, for fixed  $T$  when increasing the recording frequency  $f$ , one expects  $E(\delta \bar{X}_k^2) \sim 1/f$ . We numerically test this prediction using numerical computations on a  $\text{card}(V) = N = 10$  nodes random graph, having maximum vertex degree  $\Delta(G) = 3$ , minimum vertex degree  $\delta(G) = 2$ , with two, randomly located, detectors in figure 3a (for  $k = 1$ ). It is interesting to observed how well the expected

Fig. 3: Root mean square of the variance estimator of  $\bar{X}_1$ , i.e  $\sqrt{E(\delta \bar{X}_1^2)}$  and  $\delta X(T)$ , i.e  $\sqrt{E(\delta X_1^2(T))}$  versus the sampling frequency  $f$  for fixed  $\sigma_N = 10^{-2}$ ,  $T_0 = 80s$ ,  $T = 100s$ ,  $T_e = 120s$ , and  $N = 10$ . Dots are numerical results. Continuous line are the expected  $\sim 1/\sqrt{f}$  asymptotic behavior.

asymptotic rapidly settles for finite sampling rate as small as  $f = 20$  Hz for a finite time-support  $T = 100s$  and a final time detector recording  $T_e = 120s$ , and a fixed noise r.m.s amplitude  $\sigma_N = 10^{-2}$ . We furthermore analyze the sensitivity of  $\delta X_k(T)$  on the sampling frequency which is similarly ruled by a similar central-limit asymptotic behavior as shown in figure 3b.

### 4.2.2 Noise sensitivity versus detector number

For a fixed  $\sigma_N = 10^{-2}$  the increase of detector number leads to an increase of sampling data in the regression steps. This leads to a decrease of the apparent noise of  $\delta X(T)$ . Nevertheless, this decrease falls short after the detector number reaches a (small) fraction of the total node number  $N$ , where after the benefits of increasing their number almost saturates. This effect is illustrated in Figure 4. Our numerical

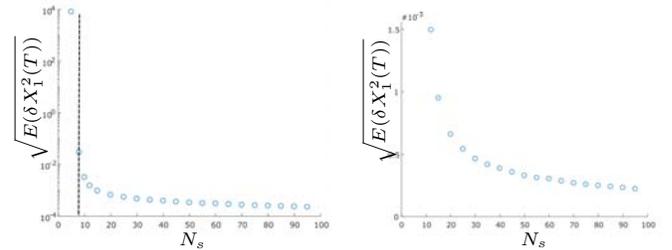


Fig. 4: Impact of sensor number  $N_s$  on the noise of  $X(T)$ , i.e the square root of the variance of  $\delta X(T)$  for fixed  $\sigma_N = 10^{-2}$ ,  $T_0 = 80s$ ,  $T = 100s$ ,  $T_e = 120s$ ,  $f = 100Hz$  and  $N = 100$ . (a) shows the sharp transition nearby  $N_s = 8$  in semi-log representation. (b) Same data as in (a) but in normal plot within  $N_s$  range between 10 to 95.

tests on randomly generated graphs indicates that, after the detector number has reached a small fraction of  $N$ , there is a very modest interest for increasing further their number. More precisely we numerically found a sharp transition observed in Figure 4a nearby  $N_s \simeq N/12$  for which the decrease of the induced noise becomes very weakly sensitive to a further increase in detector number. We believe this can be understood as a statistical effect associated with a partial correlation of observed signals when detector spatial density increases. Since recorded signal are used in the regression process if observations are correlated, their noise becomes weakly sensitive to time averaging.

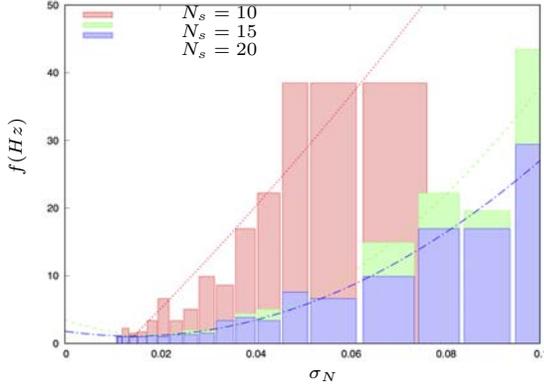


Fig. 5: Frequency sampling  $f$  providing a 95% success for source identification versus r.m.s noise  $\sigma_N$  for randomly generated graphs having maximum vertex degree  $\Delta(G) = 3$ , minimum vertex degree  $\delta(G) = 2$ ,  $N = 100$  nodes and randomly chosen sources. Dotted lines are quadratic fit of the results to help delineate the upper region for which the  $f - \sigma_{SN}$  values offer a 95% confidence for source detection.

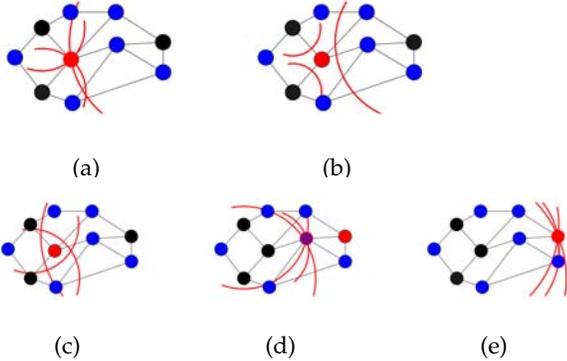


Fig. 6: Principle of pre-sectorisation from a graph (graph nodes are in blue) triangulation analysis from each sensor (black) for finding the source (red). From a guess of the initial time  $t_0$ , one can compute the iso-time sector (represented by arc of a circle in red) from a retrograde wave-front travel along the graph from each sensor in black. (a) The source is capture if  $t_0$  is correctly prescribed. (b) If  $t_0$  is too late, pre-sectorisation does not reach the source, and the intersection between each detector's pre-sectorisation can possibly be null. (c) If  $t_0$  is too soon, pre-sectorisation does reach the source, but also possibly other nodes, resulting in many possible candidates (only one is represented here). (d-e) If the source does not fit inside the –topological– convex hull encapsulating detector's nodes in the graph, depending on the chosen  $t_0$ , the pre-sectorisation can either pick a wrong external candidate (purple circle in (d)), or the good one (red circle in (e)).

This effect is reminiscent of the cut-off phenomenon arising in Markov chains shuffling [19] and might be directly connected with it. If so, the detector cut-off should scale logarithmically with  $N$ . In any case, we believe this cut-off in detector de-noising efficiency to happen in any graph, for any detector distribution. This is an important finding for the design of optimal strategies for detector number choice

in a given network. This now leads us to explore the effect of noise r.m.s on the algorithmic efficiency.

#### 4.2.3 Designing detector number versus noise amplitude

At finite sampling rate  $f$ , for finite time-support  $T = 100s$  and final time detector recording  $T_e = 120s$ , we analyze the effect of the noise r.m.s  $\sigma_N$  on the success rate of the detection algorithm. Indeed, when the noise r.m.s increases, the source detection is not always possible since the signal is degraded. Furthermore, in real applications the noise r.m.s is rarely smaller than  $10^{-2}$  so that it is relevant to analyze how the method performs when increasing noise. Figure 5 provides this analysis for a sampling rate lower than 50Hz (chosen as the relevant maximal sampling rate) for three detector relative number  $N_s/N = 10\%$ ,  $15\%$  and  $20\%$ . The chosen initial condition is  $\lambda(t) = 2\partial_t^2 \tanh((t - 40)/10)$  where 40s is the chosen triggered time for the event, having a typical time variation of 0.1s. This figure showcase an abacus of practical significance, since for a given noise amplitude, providing the number of sensors (and the other sampling and averaging parameters  $T$  &  $T_e$ ) one can design the sampling frequency for which 95% of success identification is reached. It is interesting to note that the progress obtained from having  $N_s = 15$  to  $N_s = 20$  is rather poor. The propagating time is then used backward to evaluate a distance on the graph We believe this observation related to the effect previously illustrated in figure 4, i.e a sharp transition in sampling efficiency when the detector relative number is larger than 8%. Figure 5 confirms that there is a cut-off density efficiency for detector's noise reduction improvement.

### 4.3 Presectorisation

#### 4.3.1 General principle

The general idea of pre-sectorisation is to transpose triangulation methods used in Euclidian space into graphs. This suppose that the propagating wave inside the network indeed spend physical time along each edge traveling from one node to reach its neighbor. This propagating time is used to set a distance to each edge of the graph enforcing a metric onto the graph. as proposed in [9]. From each detected signal at any given detector, and provided the time origin of the unknown event, one is able to evaluate a wave-front of possible source positions. The intersection of these wave-fronts obtained from several detectors able the source pre-sectorisation (Cf Figure 6a). Nevertheless since the time origin  $t_0$  of each event is unknown, one needs to set  $t_0$  as an unknown parameter of the pre-sectorisation method. This result in possible difficulties exemplified in Figure 6b,c,d. Since the traveling time permits to define a metric into the graph, one can built the convex-hull of the detectors set inside the graph. Considering this convex-hull, two possibilities arise. Either, the event appears inside this convex-hull (Cf Figure 6a,b,c), or outside it (Cf Figure 6d,e). In the first case, the retrograde wave-fronts intersection, emergent from each detectors, are spreading along a limited area, resulting in a moderate number of source candidates. In the second case, if the convex hull only represent a moderate fraction of the graph, the source being outside it, almost any node outside the convex hull can provide a

possible source, resulting in a large number of candidates. Hence, for an efficient and robust pre-sectorisation, it is important that the convex-hull of the detector set covers a large portion of the entire graph. This is a supplementary and practical constraint on the detector set, additional to the mathematical requirement that is should also be BC-strategic. This is why considering a detector number larger than two is a relevant practical issue.

We now describe in more details the chosen pre-sectorisation method.

#### 4.3.2 Method and algorithm

As previously mentioned, the event's time origin  $t_0$  is unknown when the first arrival signal is detected. Let us define  $t_{0g}$  for this time origin guess. Let us call  $S^*$  as the set of all possible candidates, and  $S^*(t_{0g})$  the set of all possible candidates obtained from pre-sectorisation whilst supposing the time origin is  $t_{0g}$ . Then, one has

$$S^* = \bigcup_{t_{0g}=0}^{t_{max}} S^*(t_{0g}). \quad (59)$$

To find  $S^*(t_{0g})$ , a retrograde wave-fronts propagation is computed from every detector in  $\xi$ . Let us define  $d_k(t_{0g})$  the distance between the source and each detector  $k \in \xi$  and the distance matrix  $\mathcal{D}$  between any node  $v$  and detector  $k$ . If the graph is embedded in space (such as a pipe network for which the vertex to vertex distance along the graph is a 'natural' feature of the graph), this distance is the "Euclidian/topological" distance: i.e the sum of the Euclidian distances from every successive couple of adjacent nodes along the shortest path into the graph between node  $n$  and detector  $k$ . If the graph is not embedded in space it can be defined as the traveling wave distance define over the (ordered) shortest path  $S_p$  joining every successive vertex pairs  $v', v''$  along itself and associated with a traveling velocity  $c_{v', v''}$  at each edge (in most of the formulation provided in this paper we have supposed an homogeneous traveling wave velocity  $c$ , i.e  $c_{v', v''} = c$ , as well as an homogeneous edges, i.e a non-weighted graphs, but the generalization to weighted graph is direct, this is why, we mention it here) and time traveling  $\Delta T_{v', v''}$  along each edges from vertex  $v'$  to  $v''$

$$\mathcal{D} \equiv \mathcal{D}_{kv} = \sum_{v', v'' \in S_p(k \rightarrow v)} c_{v', v''} \Delta T_{v', v''}. \quad (60)$$

Then one can define  $S^*(t_{0g})$

$$S^*(t_{0g}) = \bigcap_{k \in \xi} \{v \in V \setminus (D_{kv} = d_k(t_{0g}))\} \quad (61)$$

Albeit theoretically sound, this expression does not include some tolerance on the signal detection due to signal noise or detector sensitivity. Including an extra parameter  $\varepsilon$ , one can define a more physically relevant version of (61)

$$S^*(t_{0g}) = \bigcap_{v \in \xi} \{v \in V \setminus (|D_{kv} - d_k(t_{0g})| < \varepsilon)\} \quad (62)$$

So that, (59) with (62) provide a robust and fairly general (independent of specific graph features) formulation so as to locate the source into a pre-sectorised region. Other more dedicated, specific and more effective methods could

possibly be set-up to achieve a similar pre-sectorisation. Nevertheless, the one presented here, already provide a drastic improvement over a systematic exploration algorithm, since it is easy to show that it permits a  $O(N)$  gain in complexity on the source detection algorithm over the systematic exploration proposed in [1]. First let us summary the complexity of a systematic exploration when the pre-sectorisation set is the entire graph, in algorithm 3. In this case in step 2, the subset of source candidates spans the entire vertex set, i.e  $S^* = V$  so that the loop over each candidates is  $O(N)$ . The subsequent steps of algorithm 3, inside this loop, require

- Step 4. Construction of matrix  $\Delta_{G_{s^*}} : O(N^2)$ ,
- Step 5. Eigenvectors  $v_v^n$  computation :  $O(N^3)$ ,
- Step 6. Building matrix  $\mathbf{R}_{s^*} : O(N.N_s)$ ,
- Step 7. Compute  $\mathbf{Y}_{s^*} = [\mathbf{R}_{s^*}^\top \cdot \mathbf{R}_{s^*}]^{-1} \mathbf{R}_{s^*}^\top \mathbf{D} : O(N^3)$ ,
- Step 8. Estimate  $\|\mathbf{R}_{s^*} \cdot \mathbf{Y}_{s^*} - \mathbf{D}\|^2 = r_{s^*} : O(N.N_s)$ ,
- Step 10. Find  $s_{sol}^*$  such that  $\min_{s^* \in S^*} r_{s^*} : O(N)$ ,
- Step 11. Compute  $\mathbf{Y} = [\mathbf{R}^\top \cdot \mathbf{R}]^{-1} \mathbf{R}^\top \mathbf{D} : O(N^3)$ ,
- Step 12. Estimate  $\|\mathbf{R} \cdot \mathbf{Y} - \mathbf{D}\|^2 = r : O(N.N_s)$ .

Thus the overall complexity from Step 4 to Step 12 is  $O(N^3)$ . Including Step 4 to Step 12 in a  $O(N)$  loop results in a  $O(N^4)$  cost associated with a systematic source candidates exploration, as found in [1]. Furthermore, since the pre-sectorisation restrict the source candidate to a finite number of nodes the size of which being independent of  $N$ , the external loop complexity over pre-sectorised candidates is  $O(1)$  leading to an overall  $O(N^3)$  cost of the pre-sectorisation algorithm. To be exhaustive, one should add the complexity of the pre-sectorisation computation being a pre-requisite of algorithm 3. The algorithm for pre-sectorisation is hereby given in algorithm 4.

---

**Algorithm 4** Pre-sectorisation to obtain candidate node set  $S^*$

---

- Require:** Single component Graph  $G(V, E)$  and its adjacency matrix  $A_G$
- Require:** Parameter  $N_s$ -vector `trig_time`, BC-strategic  $N_s$ -vector set  $\xi, \epsilon_{tol}$
- Evaluate  $N_s \times N$  matrix  $\mathcal{D} = \mathcal{D}_{i,j}$  for  $i \in \xi, j \in V \triangleright \text{Cf (60) for definition of } \mathcal{D}$
- Initialization  $S^* = \text{NULL}, S_{t_0}^* = \text{NULL}$
- 3: **for**  $t_0 \in [\min \text{trig\_time} - \max \mathcal{D}/c, \min \text{trig\_time}]$  **by**  $\Delta t_0$  **do**
- $S_{t_0}^* = \bigcap_{k \in \xi} \{v \in V / \mathcal{D}_{kv} - c(\text{trig\_time}[k] - t_0) < \epsilon_{tol}\}$
- $S^* = S^* \cup S_{t_0}^*$
- 6: **end for**
- 

The complexity of algorithm 4 is as follows :

- Step 2. Evaluation of  $N_s \times N$  matrix  $\mathcal{D} : O(N_s.N)$
- Step 3. Back propagation loop over sensors set :  $O(N_s.N)$ , resulting in an overall  $O(N_s.N)$  cost of this pre-computation step. Hence, as expected, the pre-sectorisation has a benign consequence on computational complexity.

#### 4.3.3 Numerical results and illustrations on graph models

This section examines the algorithmic cost and efficiency on various graph models. The effect of pre-sectorisation

on the computational time for source detection is illustrated in figure 7 for poorly connected random graphs. This

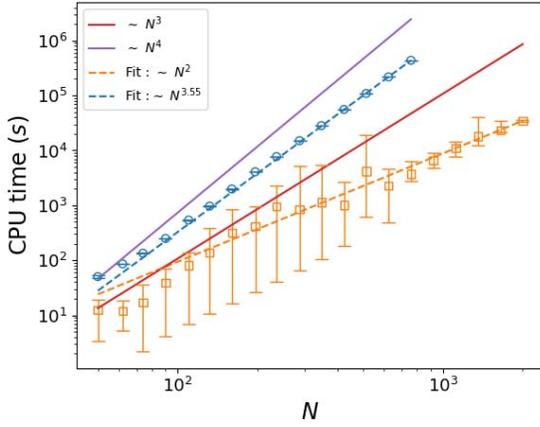


Fig. 7: Log-log plot of computational time versus nodes number  $N$ . Continuous curves display the  $\sim N^4$  (purple) and  $\sim N^3$  (red) behaviors. Square symbols represent the average results over 10 random sources distributed within a random graph, with maximum vertex degree  $\Delta(G) = 3$ , minimum vertex degree  $\delta(G) = 2$ , with  $\sigma_N = 5.10^{-4}$  having  $N$  nodes for the pre-sectorisation method (error bar here indicate the min and max time over ten simulations). Bullets symbols represent the result of the systematic exploration algorithm of [1].

figure compares a systematic exploration of the network for finding the source used in [1] to the pre-sectorisation method proposed in the previous section. A significant number of randomly chosen source location are explored so as to provide an average numerical cost. Figure 7 convincingly illustrates that a  $O(N^3)$  computation cost can be achieved for large  $N$  (here the maximal value of  $N$  was 2000), as opposed to the systematic exploration cost being  $O(N^4)$ . More precisely, we found that, for the chosen weakly connected graphs (having in mind pipe networks applications) the observed efficiency is somehow better than these theoretical predictions. Furthermore the success rate of the method is analyzed when varying the sensor density choosing an homogeneous random distribution of sensors within random graphs tested in [9] : Erdős-Rényi (ER) and a Watts-Strogatz (WS). It is found that the success rate of location events display a very similar behavior in both cases. More precisely, when varying the average node degree  $\langle k \rangle$  (average edge per node) (using a probability  $p = 0.1$  for WS graphs), Figure 8a indicates more than 80% of success for sensor density being larger than 15%. For sensor density smaller than 15% we found a rather sharp degradation of the success rate. However this conclusion is tempered by the observation of the “almost-found” (one degree distant along the graph) sources identification rate. The later is more robust to sensor density, being larger than 90% for sensor density equal to 10%. A similar sharp transition at lower sensor density is also observed for “almost-found” success rate. A similar conclusion can be reached for the Erdős-Rényi graph family in Figure 8b, for which sensor

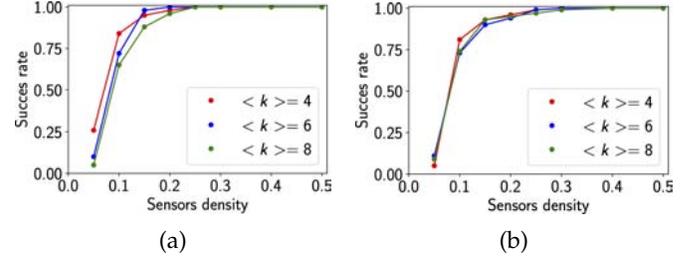


Fig. 8: Success rate of source detection with sensors having r.m.s noise  $\sigma_N = 1.5 \cdot 10^{-2}$  versus sensor density within model networks used in [9]. Continuous curves provide the success for exact source location. Dotted curves the success rate for one degree distance of source location.  $N = 100$  nodes have been chosen as in [9], and  $N = 100$  test events have been considered. (a) Watts-Strogatz (WS) graphs with average node degree  $\langle k \rangle$  (using a probability  $p = 0.1$  for supplementary random connections). (b) Same conventions as (a) for Erdős-Rényi (ER) graphs.

density equal to 10% permits 80% of success in most cases for “almost-found” sources, whereas a 15% sensor density is needed to reach this achievement for exact findings. These performances of our method are close to the ones reported in [9] for the same test cases.

## 5 APPLICATION TO REAL GRAPHS

The ability of the presented method to identify randomly chosen events within real graphs is illustrated in this section.

Concerning real graphs, power networks as well as river basin networks (as in [17]) taken from open data bases [20], have been specifically chosen so as to analyze two complementary class of graphs : a first class with a high density of circuits (power-network graphs) and a second one almost free of circuits having three-like structure (river basin networks). For every case, a sensor r.m.s noise equal to  $\sigma_N = 1.5 \cdot 10^{-2}$  is chosen in order to compare the method’s performances in distinct networks having similar noise level. These two classes of graphs also illustrate two possible classes of applications of the method. The ability to detect BC-events is evaluated here. Figures 9a,b,c,d thus illustrate three power-network distribution and one urban connectivity network (more details on graphs are given table 1) whereby the BC-event detection method with pre-sectorisation has been applied. In each case,  $0.15 \times N$  sensor node number (a sensor density of 15%) have been used. The node number of the chosen real graph varies between  $N = 47$  to  $N = 443$  (as compared with  $N = 287$  examined in [17]).

In almost every cases, the source is exactly recovered, (100% success in Figure 9b,c,d,e,f). Moreover, location failure, if any, are always found very close to the exact source, as found in Figure 9a where only one degree distance is observed between the source and its false prediction. The dependence of success-rate with sensor density is more accurately investigated in Fig. 10, averaging each rate over  $N/2$  randomly seeded sources.

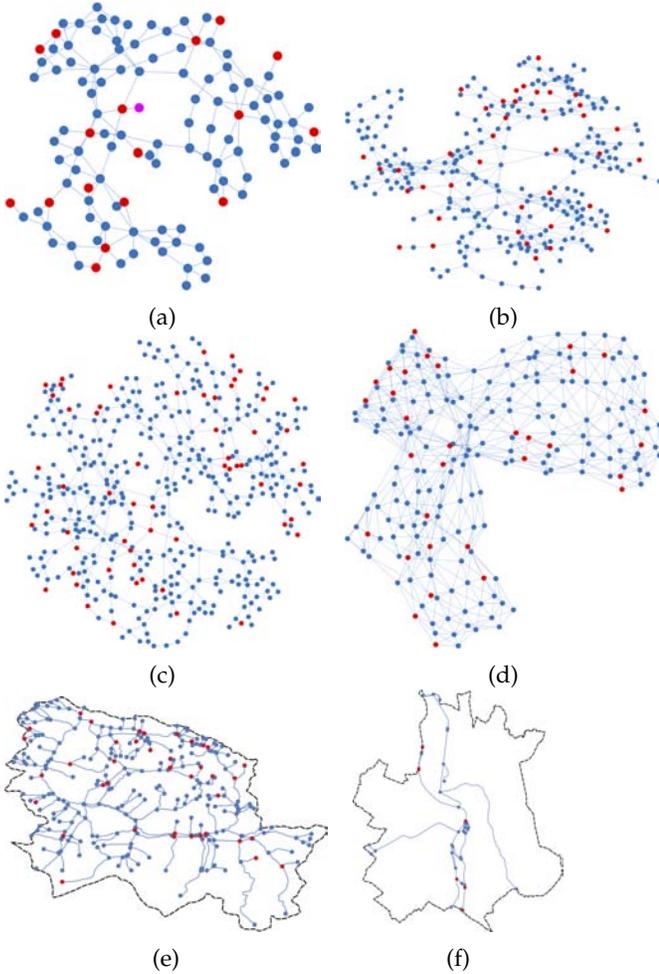


Fig. 9: Illustration of real graphs distribution networks from public data bases. Each exactly detected source is depicted in red whereas wrongly detected ones are represented in purple with a link to the exact (red) source –Cf (a)–. Sensor positions are randomly chosen with a sensor node density equal to 15% and r.m.s noise level  $\sigma_N = 1.5 \cdot 10^{-2}$ . (a,b,c) are three power distribution networks. (d) is a urban connectivity network. Total node number  $N$  and graph origin are indicated in each case. (a) [21]  $N = 118$  (b) [22]  $N = 274$  (c) [23]  $N = 443$  (d) [24]  $N = 209$  (e) Garonne river basin (South of France) [25] mapped with Geo-spatial coordinates [26]  $N = 293$  (f) same as (e), Garonne river, in Toulouse town. [27]  $N = 47$ .

Fig. 10 shows that 10% sensor density already permits more than 90% success rate in every case.

## 6 CONCLUSION

This paper presents various extensions of event identification for non-dissipative wave propagation in networks described by graphs. Having defined intern events and BC-events we found that in both cases, for an arbitrary number of detectors

- 1) Source identification of intern-type events (resp. BC-event) are possible (and unique) if the detector set is strategic (resp. BC-strategic). This result extends the

analysis of [1] dedicated to intern-type events only recorded by two sensors only.

- 2) We found that the noise sensitivity of the source identification detection can be improved in various ways. Increasing the time-averaging of the detectors support, the time-recording of the detectors, or the frequency acquisition are all beneficial, with central-limit asymptotic behavior extending the analysis of [1]. We also estimate the sensitivity matrix response of noisy signals over an arbitrary set of detectors, an analysis not provided in [1]. Another original result obtained in this study has shown that increasing the detector number improves the noise robustness of the detection, up to a cut-off detector density being a small fraction of  $N$  (possibly scaling as  $\ln(N)$ ).
- 3) We analyze a pre-sectorisation strategy so as to improve the computational cost of source identification. We have stressed the interest of a uniform coverage of detectors over the network.
- 4) Finally, using pre-sectorisation we found that it is possible to improve the source detection algorithm within a  $\text{card}(V) = N$  network to an  $O(N^3)$  computation cost, as previously obtained in the context of diffusive processes in graphs [9], [17].

Fig. 10: Detection success rate versus sensor density within real networks using the same conventions as Fig. 8 tested over  $N/2$  events.  $N = 188$  corresponds to power network of Fig. 9a,  $N = 209$  to urban network of Fig. 9d,  $N = 293$  to river basin network Fig. 9e.

The proposed method has been tested in real-world networks for which the applicability of the method on noisy data has been illustrated. Finally it is important to mention that, since the presented method lies upon a time projection upon a discrete base decomposition it necessitates the choice of a particular projection mode. The choice of this mode can affect the efficiency of the method.

## REFERENCES

- [1] J. G. Caputo, A. Hamdi, and A. Knippel, "Inverse source problem in a forced network," *Inverse. Probl.*, vol. 35, no. 5, pp. –, 2019.
- [2] P. B. and G. M., *Traffic Flow on Networks*. American Institute of Mathematical sciences, 2006.
- [3] A. Bressan, Čanić S., G. M., M. H. M., and P. B., "Flows on networks: Recent results and perspectives," *EMS Surv. Math. Sci.*, p. 47–111, 1 2014.
- [4] C. Giudicianni, M. Di Nardo, A. and Di Natale, R. Greco, G. F. Santonastaso, and A. Scala, "Topological Taxonomy of Water Distribution Networks," *Water*, vol. 10, no. 4, 2018.
- [5] K. Prakash, A. Lallu, F. Islam, and K. Mamun, "Review of power system distribution network architecture," 12 2016, pp. 124–130.
- [6] D. S. Wishart, "Advances in metabolite identification," *Bioanalysis*, vol. 3, no. 15, pp. 1769–1782, 2011.
- [7] J.-G. Caputo, A. Knippel, and E. Simo, "Oscillations of networks: The role of soft nodes," *J. Phys. A Math. Theor.*, vol. 46, p. 035101, 12 2012.
- [8] Y. Gao, M. Brennan, P. Joseph, J. Muggleton, and O. Hunaidi, "A model of the correlation function of leak noise in buried plastic pipes," *J. of Sound and Vibration*, vol. 277, pp. 133–148, 10 2004.
- [9] S. Shen, Z. and Cao, W.-X. Wang, Z. Di, and H. E. Stanley, "Locating the source of diffusion in complex networks by time-reversal backward spreading," *Phys. Rev. E*, vol. 93, p. 032301, 2016.

[10] H.-J. Wang, F. Zhang, and K. Sun, "An algorithm for locating propagation source in complex networks," *Phys. Lett., A*, vol. 393, p. 127184, 2021.

[11] H. Wang and K.-J. Sun, vol. 131, no. 4, p. 48001, sep 2020.

[12] G. Scardoni and C. Laudanna, *Centralities Based Analysis of Complex Networks*, 03 2012.

[13] F. Yang, R. Zhang, Y. Yao, and Y. Yuan, "Locating the propagation source on complex networks with Propagation Centrality algorithm," *Knowl Based Syst*, vol. 100, pp. 112–123, 2016.

[14] W. Hongjue, "An universal algorithm for source location in complex networks," *Phys. A: Stat. Mech. Appl.*, vol. 514, pp. 620–630, 2019.

[15] D. Brockmann and D. Helbing, "The hidden geometry of complex, network-driven contagion phenomena," *Science*, vol. 342, no. 6164, pp. 1337–1342, 2013.

[16] S. Peng, S. Yu, and A. Yang, "Smartphone malware and its propagation modeling: A survey," *IEEE Communications Surveys Tutorials*, vol. 16, no. 2, pp. 925–941, 2014.

[17] P. C. Pinto, P. Thiran, and M. Vetterli, "Locating the source of diffusion in large-scale networks," *Phys. Rev. Lett.*, vol. 109, p. 068702, 2012.

[18] F. Flouraboué, E. Thiam, B. Delmotte, and E. Climent, "Identification of internal properties of fibres and micro-swimmers," *Proc. Math. Phys. Eng.*, vol. 473, p. 20160517, 01 2017.

[19] P. Diaconis, "The cutoff phenomenon in finite Markov chains," *Proc. Natl. Acad. Sci. U.S.A.*, vol. 93, no. 4, pp. 1659–1664, 1996.

[20] T. A. Davis and H. Yifan, "The University of Florida Sparse Matrix Collection." *ACM Transactions on Mathematical Software*, vol. 38, no. 1, p. , 2011.

[21] —, "The University of Florida Sparse Matrix Collection." in *ACM Transactions on Mathematical Software*, 2011. [Online]. Available: <https://sparse.tamu.edu/HB/bcspwr03>

[22] —, "The University of Florida Sparse Matrix Collection." in *ACM Transactions on Mathematical Software*, 2011. [Online]. Available: <https://sparse.tamu.edu/HB/bcspwr04>

[23] —, "The University of Florida Sparse Matrix Collection." in *ACM Transactions on Mathematical Software*, 2011. [Online]. Available: <https://sparse.tamu.edu/HB/bcspwr05>

[24] —, "The University of Florida Sparse Matrix Collection." in *ACM Transactions on Mathematical Software*, 2011. [Online]. Available: [https://sparse.tamu.edu/HB/dwt\\_209](https://sparse.tamu.edu/HB/dwt_209)

[25] Sandre Development Team, *Sandre*, Eau France, 2009. [Online]. Available: <https://www.sandre.eaufrance.fr/geo/ZoneHydro/O000>

[26] QGIS Development Team, *QGIS Geographic Information System*, Open Source Geospatial Foundation, 2009. [Online]. Available: <http://qgis.osgeo.org>

[27] Sandre Development Team, *Sandre*, Eau France, 2009. [Online]. Available: <https://www.sandre.eaufrance.fr/geo/ZoneHydro/31555>

## APPENDIX

**Franck Plouraboué** is Senior Scientist at CNRS (DR1). He received his degree from Ecole Normale Supérieure in Paris in 1993, and his PhD in Physics at Paris 7 University in 1996. Since then, he works at the Fluid mechanics Institute of Toulouse (IMFT UMR 5502), being specialist of heterogeneous media, micro-hydrodynamics, multi-physics modeling and complex networks arising in biological micro-circulation and hydraulic systems.

**Pierre Uszes** received his MD Engineering degree at ENSEEIHT National University INPT in 2020.

**Romain Guibert** is research engineer at Toulouse INP. He received his PHD degree Paul Sabatier University toulouse III in 2009. He is working inside MFEED team a dedicated structure to interface research methods and results with industrial problems and needs.

Symbol's	<b>Nomenclature</b>
$\mathbf{A}_m \equiv \Delta_G - \mu_m \mathbb{I}$	Matrix derived from the graph Laplacian matrix acting on $V$ -space so that $\text{card}(\text{Im}(\mathbf{A}_m)) = N$
$\mathbf{A}_m^\xi$	$(N - N_s) \times N$ sub-matrix of $\mathbf{A}_m$ removing all collums in $\xi$
$\mathbf{A}_{s_1^*, s_2^*}^\xi$	$(N - N_s) \times (N - 2)$ sub-matrix of $\mathbf{A}_m^\xi$ formed by suppressing two distinct lines $s_1^*, s_2^* \in S^*$
$\mathbf{A}_m^{(i,j)}$	$N \times (N - 2)$ sub-matrix of $\mathbf{A}_m$ suppressing columns $i, j$
$\mathbf{A}_{s_1^*, s_2^*}^{(i,j)}$	$(N - 2) \times (N - 2)$ sub-matrix of $\mathbf{A}_m^{(i,j)}$ suppressing $s_1^*, s_2^*$
$G(V, E)$	Un-directed Graph defined by Vertex and Edges sets $V$ & $E$
$G_v(V, E)$	$v$ - directed Graph obtained from $G$ by BC-event on $v \in V$
$\Delta_G$	Graph Laplacien matrix of $G$
$\Delta_{G_v}$	Graph Laplacien matrix of $G_v$
$\Delta_v$	symmetric sub-matrix of Graph Laplacien matrix $\Delta_{G_v}$
$\lambda(t)$	Compact support source perturbation temporal variation
$\mathcal{G}_{BC}$	Set of all BC-Graphs $G_v$ , i.e, $\mathcal{G}_{BC} = \cup_{v \in V} G_v$
$E$	Edge set
$\mu_m \equiv (\frac{m\pi}{T})^2$	mode of the temporal adjoint base
$M$	$M + 1$ is the number of time steps
$V$	Vertex set with $\text{card}(V) = N$
$\Omega$	Discrete Spectrum of Graph Laplacien matrix $\Delta_G$
$-\omega^2$	Eigenvalue of $\Omega$
$-\omega_n^2$	$n$ th eigenvalue of ordered set $\Omega = \{-\omega_1^2, -\omega_2^2, \dots, -\omega_N^2\}$
$\mathfrak{V}$	eigenvector set of the Grah Laplacien matrix $\Delta_G$
$\mathfrak{V}_v$	eigenvector set of the Graph Laplacien matrix $\Delta_{G_v}$
$\Omega_v$	Spectrum of the Graph Laplacien matrix $\Delta_{G_v}$
$v$	eigenvector of $\mathfrak{V}$ having $k$ th component $v_k$
$v^n$	$n$ th eigenvector of ordered set $\mathfrak{V} = \{v^1, v^2, \dots, v^N\}$
$\mathbf{R}$	$N_s \times N_t$ regression matrix for $N_s$ sensors & $N_t$ time-steps
$\mathbf{R}_v$	$N_s \times N_t$ regression matrix for $N_s$ sensors in $G_v$
$\mathbf{S}$	$N$ -vector source location in $V$ , i.e, $S_k = \delta_{sk}$
$S^*$	set of source candidates $S^* \subset V$ ,
$s$	source location, $s \in V$
$s^*$	nodes candidates for sources, $s^* \in S^* \subset V$
$T_0$	Final Time of source perturbation $\lambda(t)$ , $\lambda(t) = 0$ for $t > T_0$
$T_e$	Ending Time of the sensor recording
$\Delta T$	Time step increment
$T$	End-time of time-projection support defined in $\langle, \rangle$
$\xi$	Strategic nodes set where signal is detected, $\text{card}(\xi) = N_s$
$\xi^\perp$	$\perp$ of strategic node set, i.e $\xi^\perp \cap \xi = \{\emptyset\}$ , $\xi^\perp \cup \xi = V$
$\mathbf{X}(t)$	$N$ -component vector of network's unknowns $X_k(t)$ , $k \in V$
$\mathbf{X} \xi$	Restriction of $\mathbf{X}(t)$ at sensor's location $\xi \subset V$ , $\dim(\mathbf{X} \xi) = N_s$
$\mathbf{X} \xi^\perp$	Restriction of $\mathbf{X}(t)$ out of sensor's location $\xi^\perp = V \setminus \xi$
$\mathbf{X}^{(i,j)}$	Restriction of $\mathbf{X}(t)$ out of sensor's location $(i, j)$
$y(t)$	Mode projection vector of $\mathbf{X}(t)$ on $\mathfrak{V}$ , $y_n(t) \equiv \langle \mathbf{X}(t), v_n \rangle$
$\phi_m$	Temporal adjoint orthogonal base, i.e, $\langle \phi_{m'}, \phi_m \rangle = \delta_{m,m'}$
$\mathbf{Y}(T)$	$\equiv [y(T), \dot{y}(T)]$ : 2N-vector of mode projection at $t = T$
$\dot{\mathbf{X}}(t) \equiv \frac{d\mathbf{X}(t)}{dt}$	Time derivative of unknown vector $\mathbf{X}$
$\langle f, g \rangle$	$\equiv \int_0^T f(t)g(t)dt$ : Time projection scalar product
$\bar{\mathbf{X}} = \langle \mathbf{X}, \phi_m \rangle$	Projection of $\mathbf{X}$ on $m$ th temporal mode
$\bar{\lambda} = \langle \lambda, \phi_m \rangle$	Projection of $\lambda$ on $m$ th temporal mode

TABLE 1: Real networks structural properties

Ref.	Nodes	Edges	Aver. Deg.	Deg. Std	Deg. max	Deg. min
<b>a</b>	118	179	3.0339	1.5740	9	1
<b>b</b>	274	669	4.8832	3.3052	15	1
<b>c</b>	443	590	2.6637	1.5768	9	1
<b>d</b>	209	767	7.3397	2.2584	16	3
<b>e</b>	293	319	2.1775	1.0898	4	1
<b>f</b>	47	51	2.1702	0.9628	4	1