# Subgraph detection for average detectability: application to SIS epidemics

Nicolas Martin, Paolo Frasca, Carlos Canudas de Wit

# Subgraph detection for average detectability: application to SIS epidemics

Nicolas Martin, Paolo Frasca, *Senior Member, IEEE,* Carlos Canudas-de-Wit, *Fellow, IEEE*

**Abstract**—Observation and detection of networked systems aim to reconstruct the evolution of the system based on the measurement of few nodes. In large-scale networks, reconstructing the exact state of each node becomes more complex and in practice it is often superfluous. Reconstructing an aggregated version of the system is often sufficient. In the light of this observation, we consider the notion of average detectability: A system is said to be average detectable if it is possible to reconstruct the average of the subset of its unmeasured nodes. We show here that for a particular type of system, that is negative uniform networks, the average detectability property is reached when the subgraph induced by the unmeasured nodes is regular. Thus, we study the detection of such regular induced subgraph and we propose an algorithm to complete this task. We introduce also the relaxed notion of quasi-regularity ensuring an approximate reconstruction of the average. This paper presents algorithms to detect regular induced subgraphs (*RIS*) and quasi-regular induced subgraph (*q-RIS*). We propose an extension to detect multiple quasi-regular induced subgraphs (*mq-RIS*) in order to reconstruct the average of several subgraphs of the system. Finally we apply our method to the evolution of an epidemic spreading over a simulated contact network over the largest cities in France based on a SIS model.

**Index Terms**—Network theory, Regular induced subgraph, Observability, Detectability, Algorithmic, Network epidemics

## 1 Introduction

In network theory, observation problems aims to reconstruct the state of the whole system by knowing only a fraction of the states [1]. This subject has been widely studied and has raised the question, among others, of the choice of the measured nodes to improve the reconstruction [2], [3], [4]. While this issue has been solved in some cases, in large-scale networks we face the issue of complexity and limited number of sensors. Taking into account this complexity and the cost of a sensor, it is often difficult to reconstruct the state of a large-scale network. In the past few decades, several graph-theoretical approaches for controllability and observability of network systems have been proposed [5], [6], [7], [8]. Most of theses works aims to reconstruct or to control the whole state of the systems. However in numerous cases, there is no need to reconstruct the state of each node, but only an aggregation of these states. Recently, Niazi et al. [9] took advantage of this observation and proposed the notions of average observability and average detectability. These notions refer to the reconstruction of the average state of the unmeasured nodes respectively in closed-loop and in open-loop. In particular, in an average detectable system, it is possible to design an open-loop observer estimating the average of the unmeasured nodes such that the error converges to zero. In the following, we will only focus on this latter notion. In [9] the authors propose a sufficient condition for average detectability. This condition allows to test if a given network and a given subset of measured nodes form an average detectable system. If it is possible to choose the placement of the measured nodes, a question emerges: given a network system what is the smallest subset of measured nodes fulfilling the average detectability condition ? Said otherwise, what is the smallest subset of nodes to observe in order to be able to reconstruct the average of the unmeasured nodes ? This is the question we address through this article.

In order to transform the condition for average detectability into a structural condition, we will restrict ourselves to a particular type of system: negative uniform graph. With such system, when a subgraph of unmeasured nodes forms a regular graph[1], then the system is average detectable. Hence, our problem becomes detecting the largest regular induced subgraph (RIS) out of a given graph. The question of finding induced subgraphs with particular properties has been addressed in several works. For example, we can cite the maximum clique problem [10] which has implications, in particular, in social networks; Frequent subtree mining [11] which is applied to data analysis; Induced subgraph isomorphism problem [12] or its variant as Snake-in-the-box problem or the maximum independent set problem. In most cases, these problems are either oriented to data analysis or are graph-theoretic problems with no direct application. To our knowledge, the present work is the first to use an induced subgraph problem for a reconstruction concern. As said, here the objective is also to detect subgraph with a particular property which is regularity. The regular subgraph detection have been studied in different contexts and after introducing this property, we propose a brief review of the works in this domain in Section 2.

However, the regular induced subgraph detection raises some difficulties: first, the hypothesis on the system are very restrictive and concern few real systems. Moreover the problem is known to be NP-hard and so it is not scalable. Finally the solution found

- *The authors are with Univ. Grenoble Alpes, CNRS, Inria, Grenoble INP, GIPSA-Lab, 38000 Grenoble, France.*
  *E-mail:* {*nicolas.martin;paolo.frasca;carlos.canudas-de-wit*}*@gipsa-lab.fr*

*This project has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement N. 694209)*
www.scale-freeback.eu

1. Here, we call regular a graph in which all the nodes have the same out-degree. The term *out-regular* would be more accurate, but we prefer *regular* for readability concerns.

implies often to measure a large proportion of nodes. Therefore, in Section 3, we relax the problem by introducing the notion of quasi-regularity, which qualifies a graph which is *close* to be regular. We derive then a result linking the error of regularity and the error of reconstruction: the more regular the unmeasured subgraph is, the better is the reconstruction. On these grounds, we treat the problem of quasi-regular induced subgraph detection (q-RIS). We also extend the results, and the algorithm, to multiple quasi-regular induced subgraph (mq-RIS). Finally, in Section 4 we present an application of the mq-RIS approach on an epidemic spreading case.

### 1.1 Preliminaries

In this article we will consider a directed graph $G = (\mathcal{V}, \mathcal{E})$, where $\mathcal{V} = \{v_1, v_2, \dots, v_n\}$ is the set of nodes and $\mathcal{E} \in \mathcal{V} \times \mathcal{V}$ is the set of directed edges. The graph can be represented by its adjacency matrix $A \in \mathbb{R}^{n\times n}$, whose $ij$-th entry is given by

$$[A]_{ij} = \begin{cases} a_{ij} & \text{if } (v_i, v_j) \in \mathcal{E} \\ 0 & \text{otherwise} \end{cases} \tag{1}$$

where $a_{ij}$ is the weight of the edge $(v_i, v_j) \in \mathcal{E}$. We follow the convention that the edge $(v_i, v_j)$ goes from $v_i$ to $v_j : v_i \longrightarrow v_j$. We associate to this graph a linear time-invariant (LTI) network system:

$$\Sigma : \begin{cases} \dot{x}(t) &= Ax(t) + Bu(t) \\ y(t) &= Cx(t) \end{cases}, \tag{2}$$

where $x(t) = [\, x_1(t), \dots, x_n(t)\,]^T$ is the network state vector and $u(t) = [\, u_1(t), \dots, u_p(t)\,]^T$ is the input vector.
We will consider that the output vector $y$ contains a sample of $k$ components of the state vector $x$. This means that $y = [\, x_1(t), \dots, x_k(t)\,]^T$ and so $C = \begin{bmatrix} I_k & 0 \end{bmatrix}$. The nodes $\mathcal{V}_1 := \{v_1, \dots, v_k\}$ are called the measured nodes while $\mathcal{V}_2 := \{v_{k+1}, \dots, v_n\}$ are the unmeasured nodes. We denote $m$ the number of unmeasured nodes: $m = n - k$. We also denote $\mathbf{x}_2(t) = [\, x_{k+1}(t), \dots, x_m(t)\,]^T$, the state of unmeasured nodes and $x_2^{av}$ the average value of the unmeasured nodes:

$$x_2^{av} = \frac{1}{m}\mathbf{1}^T\mathbf{x}_2 \tag{3}$$

We decompose the matrices $A$ and $B$ as follows:

$$A = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix},\ B = \begin{bmatrix} B_1 \\ B_2 \end{bmatrix}. \tag{4}$$

With $A_{11} \in \mathbb{R}^{k\times k}$, $A_{22} \in \mathbb{R}^{m\times m}$ and all the other block matrices of corresponding dimensions. We denote by $\sigma$ the *deviation vector* defined as:

$$\sigma = \mathbf{x}_2 - \mathbf{1}x_2^{av} = \begin{pmatrix} x_{k+1} - x_2^{av} \\ x_{k+2} - x_2^{av} \\ \vdots \\ x_n - x_2^{av} \end{pmatrix} \tag{5}$$

This vector contains the difference between the value of each unmeasured nodes and the average value of the unmeasured nodes. We have $\mathbf{1}^T\sigma = 0$. The evolution of $x_2^{av}$ is described by the following equation:

$$\dot{x}_2^{av} = \alpha x_2^{av} + gy(t) + bu(t) + \eta\sigma(t) \tag{6}$$

with $\alpha = \frac{1}{m}\mathbf{1}^T A_{22}\mathbf{1}$, $g = \frac{1}{m}\mathbf{1}^T A_{21}$, $b = \frac{1}{m}\mathbf{1}^T B_2$ and $\eta = \frac{1}{m}\mathbf{1}^T A_{22}$.
Let $\hat{x}_2^{av}$ be an open-loop observer for $x_2^{av}$ described as:

$$\dot{\hat{x}}_2^{av} = \alpha\hat{x}_2^{av} + gy(t) + bu(t) \tag{7}$$

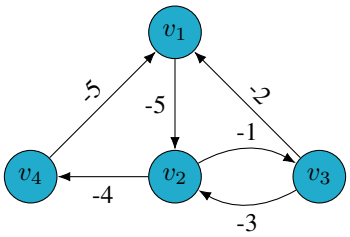



Fig. 1. Outflow balanced graph: the sum of the weights going out is the same for each node.

We will see in the next section the condition on the system to ensure that the observer $\hat{x}_2^{av}$ converges to $x_2^{av}$.

## 2 Regular Induced Subgraph detection for average detectability

### 2.1 Average detectable system

**Definition 1** (Average Detectability: AD)**.** *A system $\Sigma$ with a subset $\mathcal{V}_1$ of measured nodes is said to be average detectable if zero output implies that the average value of $\mathcal{V}_2$ converges to zero:*

$$\forall\, x_2^{av}(0) \in \mathbb{R}^m, \quad y(t) = 0 \implies \lim_{t\to\infty} x_2^{av}(t) = 0 \tag{8}$$

The average detectability of the system $\Sigma$ is equivalent to the convergence of the observer to the average value of the unmeasured nodes [9], which is:

$$\Sigma \text{ AD} \iff \hat{x}_2^{av}(t) \underset{t\to\infty}{\longrightarrow} x_2^{av}(t) \tag{9}$$

In the following, we first exhibit a sufficient condition for average detectability in the general case. We will introduce then a special class of systems: negative uniform graphs. With such a system, it appears that finding a subgraph $\mathcal{V}_1$ ensuring average detectability amounts to find a regular induced subgraph. We investigate then the problem of detecting regular subgraph in a graph.

### 2.2 Sufficient condition for average detectability

In order to introduce the condition for average detectability we first define a graph-based notion.

**Definition 2** (Outflow balanced graph)**.** *Let $G$ be a directed weighted graph represented by its adjacency matrix $A$. $G$ is said to be outflow balanced if the sum of the weights of the outgoing edges is the same for every nodes, which is :*

$$\exists\, \gamma \in \mathbb{R}, \qquad \boldsymbol{1}^T A = \gamma \boldsymbol{1}^T \tag{10}$$

*Moreover if $\gamma < 0$, then the graph is said to be negatively outflow balanced.*

*Example* 1. Consider the graph $G$ represented in fig. 1. Its adjacency matrix $A$ is as follows:

$$A = \begin{pmatrix} 0 & 0 & -2 & -5 \\ -5 & 0 & -3 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & -4 & 0 & 0 \end{pmatrix} \tag{11}$$

We verify that $\mathbf{1}^T A = -5 \times \mathbf{1}^T$. Hence the graph $G$ is negatively outflow balanced.

**Proposition 1** (Sufficient condition for average dectability - Theorem 3 in [9])**.** *Consider a system $\Sigma$ associated to the graph $G$. We*

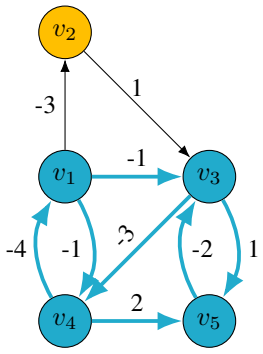



Fig. 2. By measuring the node $v_2$, the unmeasured nodes $\mathcal{V}_2 = \{v_1, v_3, v_4, v_5\}$ induces a negatively outflow balanced $G_{\mathcal{V}_2}$ with $\gamma = -2$. Thus, the system is average detectable.

*note $G_{\mathcal{V}_2}$ the subgraph of $G$ induced by the subset of unmeasured nodes.*
*The system $\Sigma$ is average detectable if $G_{\mathcal{V}_2}$ is negatively outflow balanced.*

Therefore if the set of nodes can be partitioned into two subsets such that one induced subgraph is negatively outflow balanced, then measuring the nodes outside this subgraph makes the system average detectable. Figure 2 gives an example of such a system.

### 2.3 Formal Problem statement

The problem we want to tackle is formulated as follows:

**Problem 1** (Negatively outflow balanced induced subgraph detection)**.** *In a given graph $G$ find the largest induced subgraph of $G$ which is negatively outflow balanced.*

However, in an arbitrary graph with independent weights there is, *almost surely*, no outflow balanced subgraph. That being said, two solutions can be considered: Either we restrict ourselves to particular systems for which we know that negatively outflow balanced subgraphs can be found, either we relax the notion of average detectability and hence the notion of regularity. The second solution will be explored in Section 3 where we introduce the notion of quasi-regularity. In the following of this section, we focus on the first solution: we treat the problem for a particular type of systems: the negative uniform graphs defined hereafter.

**Definition 3** (Negative uniform graph: NUG)**.** *A graph $G$, represented by the adjacency matrix $A$, is said to be negative uniform if all its weights are equal and negative, which is $A_{ij} \in \{0; a\}$ with $a < 0$.*

**Proposition 2** (Sufficient condition for average detectability of negative uniform graph)**.** *Let $\Sigma$ be a system associated to a negative uniform graph $G$. $\Sigma$ is average detectable if $G_{\mathcal{V}_2}$ is regular.*

*Proof.* In a negative uniform graph we remark that

$$\mathbf{1}^T A = a \times \deg_{\text{out}}(G), \tag{12}$$

Therefore, with (10) and (12), a negative uniform graph is negatively outflow balanced if there exists $\gamma < 0$:

$$a \times \deg_{\text{out}}(G) = \gamma \mathbf{1}^T$$
$$\deg_{\text{out}}(G) = \frac{\gamma}{a}\mathbf{1}^T$$

This mean that all the out-degree of $G$ must be the same, which is $G$ has to be regular □

Finally we have:

$$\text{NUG} + \text{RIS} \implies \text{AD} \iff \hat{x}_2^{av}(t) \underset{t\to\infty}{\longrightarrow} x_2^{av}(t) \tag{13}$$

*Remark* 1. While negative uniform graph is a quite restrictive case, we can also consider positive uniform graph with a same negative self-loop $\eta$. Indeed in this case, we have $\mathbf{1}^T A = a \times \deg_{\text{out}}(A) - \eta\mathbf{1}$, and even with $a > 0$ the right side can be negative if the self-loop $\eta$ is large enough. Proposition 2 remains true if the system is associated with such a graph. This type of graphs includes for example some heat systems with high dissipation [13]. The model presented in section 4 falls also in this sclope. In the following, we only consider negative uniform graph for the simplicity of the development.

We can now formulate the problem arising from Problem 1 restricted to the negative uniform graphs case.

**Problem 2** (Regular induced subgraph detection)**.** *Let $G$ be a negative uniform graph. We look for the largest regular induced subgraph of $G$, which is:*

$$\begin{aligned} &\max_{I\subset\mathcal{V}} |I|, \\ &\textit{s.t. } G_I \textit{ is regular.} \end{aligned} \tag{14}$$

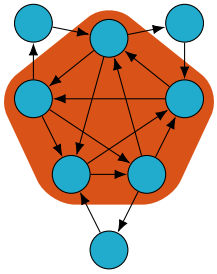
*where $G_I$ is the subgraph of $G$ induced by the subset of nodes $I$.*

By measuring the nodes outside the subgraph solution of Problem 2, we obtain an average detectable system. Figure 3 illustrates this problem.

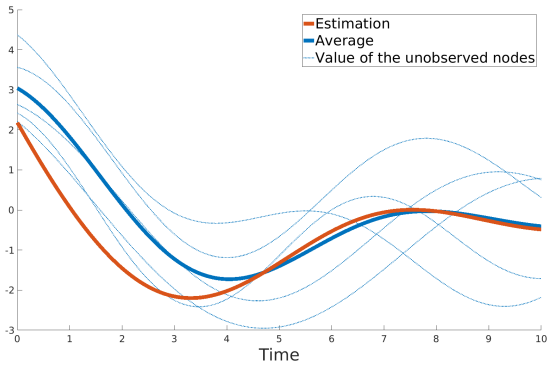
### 2.4 Methods for RIS detection

In this section, we briefly present the existing literature on the regular induced subgraph detection problem. We will see to wich extent previous works can help to tackle Problem 2.
The problem we want to tackle is known as regular induced subgraph (or RIS) detection. In the literature the problem, is treated for a given degree $k$ of the regularity and is called $k$-regular induced subgraph (or k-RIS) problem. The works in this domain can be classified as follows:

- **Complexity of the k-RIS problem**: A first work [14] showed that the problem in the case $k = 0$ is NP-hard. Then several works [15], [16] generalize the result for any $k$ and different type of graphs. Some studies [17], [18] exhibit a polynomial complexity for some particular types of graph.
- **Algorithms to detect the largest k-RIS**: Despite the complexity of the problem, some algorithms have been proposed to solve it. In the case $k = 0$, [19], [20] propose *fast-exponential* algorithms (i.e. in $\mathcal{O}(c^n)$ with $c \leq 2$). For any $k$, a fast-exponential algorithm based on a branch-and-bound approach is proposed in [21] and a polynomial algorithm for a particular type of graphs in [22]. However these results consider undirected graph and standard regularity (not *out-regularity* as us). Still, the branch-and-bound approach proposed in [21] can be extended to our case and will be presented later.
- **Upper-bound on the size of the largest** k-RIS: Facing the complexity of the problem it is interesting to obtain an upper-bound on the size of the optimal solution. An



(a) We detect a regular induced subgraph. Here the five nodes in the center have a same out-degree equal to 2.




(b) The system is excited with an arbitrary sinusoidal input in some nodes. The dotted blue lines represent the evolution of the unmeasured node in the initial system (2) and the solid blue line their average. The solid red line represents the evolution of the observer described in (7) which converges towards the averaged value.

Fig. 3. Representation of the approach proposed here: (a) from an initial negative uniform graph, a regular subgraph is detected; (b) as the system is average detectable, the observer (7) allows to estimate the averaged value of the unmeasured nodes.

intuitive upper-bound on the size of the largest k-RIS is the size of the k-Core of the graph. The k-Core is a subgraph obtained by removing iteratively nodes with a degree smaller than $k$. More elaborate upper-bounds have been proposed for any $k$ [15], [23] or for particular values of $k$ [14], [24]. As above, these results are valid only for undirected graph and. From our knowledge, only the k-Core approach can be extended to our case.

- **Approximating algorithm to find a sub-optimal solution**: While it would be interesting to find an approximate solution of our problem, it is shown that this problem is hard to approximate [25].

### 2.5 The meta-algorithm and the sub-algorithms

As discussed in the previous section, the literature proposes only methods to find the largest $k$-regular induced subgraph but nothing to solve the problem for every $k$. A simple approach consists in solving the problem for each $k$ and then keeping the best solution: We denote $\mathrm{RIS}(G)$ and $\text{k-RIS}(G)$ respectively, the largest regular and $k$-regular induced subgraphs of $G$. We have then:

$$\mathrm{RIS}(G) = \max_{k\in\mathbb{N}} \text{k-RIS}(G) \tag{15}$$

Some tricks can be used to optimize the approach:

- As seen in the previous section, there are some methods allowing to find an upper-bound on $\text{k-RIS}(G)$. We denote this upper-bound by $\theta_k(G)$. While testing the value of $\text{k-RIS}(G)$ for every $k$, if $\theta_{k_0}(G)$ is smaller than the largest $\text{k-RIS}(G)$ so far, it is useless to compute $\text{k}_0\text{-RIS}(G)$. Noticing that the computation of this upper-bound is faster by far than the computation of the $\text{k-RIS}(G)$, this helps the computation of the regular induced subgraph.
- The cases with $k = 0$ and $k = 1$ are particular and a specific algorithm can be applied.
- Using an approximate algorithm for the k-RIS would imply a sub-optimal solution for the RIS which can be interesting if the approximation is good.

Based on these remarks, we propose Algorithm 1 to detect the largest regular induced subgraph within a given graph.

**Algorithm 1** Regular induced subgraph detection

**Require:** $G$
$\mathrm{RIS} = []$
**for** $k = 0 : \max(\deg_{\text{out}}(G))$ **do**
  $\theta_k = \text{UpperBound\_k-RIS}(k, G)$
  **if** $\theta_k > |\,\mathrm{RIS}\,|$ **then**
    $\text{k-RIS} = \text{Find\_k-RIS}(k, G)$
    $\mathrm{RIS} = \max(\mathrm{RIS}, \text{k-RIS})$
  **end if**
**end for**
**Ensure:** RIS

This is actually a *meta-algorithm* as we only give the skeleton of the method and not the *sub-algorithms* UpperBound_k-RIS and Find_k-RIS. The choice of these sub-algorithm is discussed hereafter.

*Remark* 2. The sub-algorithms $\text{Find\_k-RIS}(k, G)$ and $\text{UpperBound\_k-RIS}(k, G)$ can be different algorithms for $k = 0$ and $k = 1$.

We first discuss the implementation of the $\text{UpperBound\_k-RIS}(k, G)$ algorithm. This algorithm is detailed in Algorithm 2 and is based on the k-Core. As said before, the k-Core of a graph is the subgraph obtained after removing repetitively the nodes with a degree smaller than $k$. It is clear that

$$\text{k-RIS}(G) \subset \text{k-Core}(G) \tag{16}$$

and so, the size of the k-Core is an upper-bound on the size of k-RIS.

**Algorithm 2** k-Core

**Require:** $G, k$
$I \leftarrow$ Nodes in $G$ with out-degree $< k$
**while** $I$ is not empty **do**
  Remove $I$ from $G$
  $I \leftarrow$ Nodes in $G$ with out-degree $< k$
**end while**
$\theta_k \leftarrow$ number of nodes in $G$
**Ensure:** $\theta_k$ an upper bound on the size of the k-RIS

The implementation of the second sub-algorithm $\text{Find\_k-RIS}(k, G)$ is described in Algorithm 3. It is an extension of the branch-and-bound approach proposed in [21]. It is a recursive algorithm designed as follows: Given a graph $G$

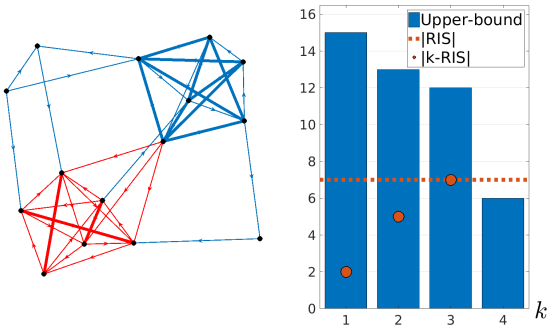



Fig. 4. Left: The largest regular induced subgraph found is highlighted in red. Right: The upper-bound (k-Core) is displayed in blue for $k \in \{1;2;3;4\}$. The real size of k-RIS are in dot red for $k \in \{1;2;3\}$. The 4-RIS is not computed since $\sigma(4) < |3\text{-RIS}|$. The red dotted line represents the size of the RIS.

and a degree $k$, we first compute the k-Core of $G$ (line 1). If the graph obtained is $k-$regular then the k-RIS is found (line 3) and the algorithm terminates. Otherwise there is at least one node, denoted $r$, with a degree larger than $k$ (line 5). It is clear that either $r$ is not in the k-RIS either (at least) one of its successors[2] is not (line 6). Thus, we consider the subgraphs obtained by removing $r$ or a successor of $r$ (line 8). Finally we compute the k-RIS for each of them (line 9) and select the largest one (line 11).

**Algorithm 3** k-RIS

**Require:** $G, k$
1: $G \leftarrow$ k-Core$(G,k)$
2: **if** $G$ is $k-$regular **then**
3: k-RIS $\leftarrow G$
4: **else**
5: $r \leftarrow$ a node of $G$ with degree $> k$
6: $I \leftarrow \mathcal{N}_{out}(r) \cup r$
7: **for** $i \in I$ **do**
8: $G_{tmp} \leftarrow$ remove $i$ in $G$
9: k-RIS$_{tmp}\{i\}$ = k-RIS$(G_{tmp}, k)$
10: **end for**
11: k-RIS $= \max($k-RIS$_{tmp})$
12: **end if**
**Ensure:** k-RIS

*Remark* 3. It is possible to use a technical trick to speed-up the algorithm. The size of the current best solution can be stored and passed to the recursive call to k-RIS (line 9). If the size of the current subgraph is smaller than the size of the best solution, the recursion stops. This shortcut is not in algorithm 3 for the sake of readability.

*Simulation* 1. Figure 4 presents a result obtained with Algorithm 1 (and Algorithms 2-3). Considering that the graph is negative uniform, the system is average detectable by measuring the 8 nodes outside of the red subgraph.

### 2.6 Discussions

The method proposed above suffers some limitations that we present hereafter and which motivate the relaxed problem presented in next section:

2. $\mathcal{N}_{out}(r)$ is the set of successors of $r$

- Negative uniform systems are rare: It seems that there is no practical example in the literature. Even when we consider positive uniform systems with large negative self-loop (as mentioned in Remark 1), the applications are scarce.
- RIS is a fragile notion: For example, a grid graph which is very close to regular (only the nodes on the border have a smaller degree) does not fulfill the condition of the theorem. Moreover the largest RIS in a graph is generally relatively small.
- The RIS problem is hard to solve: there is no specific method for the RIS problem and the k-RIS problem has to be solved several times. Moreover, as said before, the k-RIS problem is a NP-hard problem and there is no approximation algorithm to solve it.

Based on these statements, the RIS detection approach introduced in this section is difficult to apply to real-world problems. We will see that it is possible to accept some errors on the regularity while preserving the possibility to reconstruct efficiently the average. We introduce, in the next section, the notion of quasi-regularity leading to more flexibility.

## 3 Quasi-Regular Induced Subgraph Detection

### 3.1 Preliminaries

In the previous problem, to reach average detectability, the condition to satisfy was $\mathbf{1}^T A_{22} = \gamma \mathbf{1}^T$ with $\gamma < 0$. To relax the problem we consider now that this equality is no more verified and we introduce a perturbation vector $s$ defined as:

$$s = \mathbf{1}^T A_{22} - \gamma \mathbf{1}^T \tag{17}$$

and the regularity error $\epsilon$ defined as:

$$\epsilon = \frac{\|s\|_1}{m|\gamma|} \tag{18}$$

We denote $e_{ss}$ the reconstruction error defined as:

$$e_{ss} = \lim_{t\to\infty} |x_2^{av} - \hat{x}_2^{av}| \tag{19}$$

The following proposition gives the link between the regularity error and the reconstruction error.

**Proposition 3** (Relation between regularity error and reconstruction error)**.** *The reconstruction error* (19) *is linked with the regularity error* (18) *by:*

$$e_{ss} \leq \bar{\sigma} \frac{\epsilon}{1-\epsilon}, \tag{20}$$

*where* $\bar{\sigma} = \limsup_{t\to\infty} |\sigma(t)|$ *is a constant of the system.*

*Proof.* We consider that condition (10) is not fulfilled and so we introduce $s$ a perturbation vector:

$$\mathbf{1}^T A_{22} = \gamma \mathbf{1}^T + s^T \tag{21}$$

We denote $\tilde{x}_2^{av} := x_2^{av} - \hat{x}_2^{av}$ the reconstruction error. From equations (6) and (7) we have then:

$$\dot{\tilde{x}}_2^{av}(t) = \alpha \tilde{x}_2^{av}(t) + \eta\sigma(t) \tag{22}$$

where $\sigma$ defined in (5) is the deviation from average. From (21) and the definition of $\alpha$, we have:

$$\alpha = \frac{1}{m}\mathbf{1}^T A_{22}\mathbf{1} = \frac{1}{m}(\gamma \mathbf{1}^T + s^T)\mathbf{1} = \gamma + \frac{1}{m} s^T \mathbf{1} \tag{23}$$

From (21), the definition of $\eta$ and since $\mathbf{1}^T\sigma = 0$, we have:

$$\eta\sigma = \frac{1}{m}\mathbf{1}^T A_{22}\sigma = \frac{1}{m}(\gamma\mathbf{1}^T + s^T)\sigma = \frac{1}{m}s^T\sigma \quad (24)$$

From these two results, (22) becomes:

$$\dot{\tilde{x}}_2^{av}(t) = (\gamma + \frac{1}{m}s^T\mathbf{1})\tilde{x}_2^{av}(t) + \frac{1}{m}s^T\sigma \quad (25)$$

We denote $e_{ss} := \lim_{t\to\infty}\tilde{x}_2^{av}(t)$. By convergence of (25) we have:

$$e_{ss} = -\frac{\frac{1}{m}s^T\sigma}{\gamma + \frac{1}{m}s^T\mathbf{1}} \quad (26)$$

and therefore

$$|e_{ss}| = \frac{|s^T\sigma|}{|m\gamma + s^T\mathbf{1}|} \leq \frac{|s^T\sigma|}{|m\gamma| - |s^T\mathbf{1}|} \leq \frac{\|s\|_1\bar{\sigma}}{m|\gamma| - \|s\|_1} \quad (27)$$

with $\bar{\sigma} := \limsup_{t\to\infty}\sigma(t)$ and with the definition of $\epsilon$ given in (18) we proved (20). □

The constant $\bar{\sigma}$ depends on the physical system of interest and corresponds to the maximal difference between the value of an unmeasured node and the average value of the unmeasured nodes. According to Proposition 3, in order to minimize the reconstruction error, it is interesting to find a subgraph having a reasonable regularity error $\epsilon$. Relation (20) is emphasized through simulations in the following section.

*Remark* 4. As the reconstruction error $e_{ss}$ grows with the regularity error $\epsilon$, it actually grows with the absolute error of regularity $\|s\|_1$ and decreases with the degree of regularity $\gamma$ and the size of the unobserved subgraph $m$. This means that the reconstruction will be better if the subgraph of unmeasured nodes is *large*, *close to be regular* and with a *large degree of regularity*.

### 3.2 Emphasis of the link between error of regularity and error of reconstruction

Before investigating the q-RIS detection problem, we present some simulations enlightening the relation described in Proposition 3. To this end, we first introduce a family of graphs for which we can control the regularity.

**Definition 4** ($p-$reg graph)**.** *Given a graph $G$, we denote $\mathcal{N}_{out}(i) = \{j,\ (j,i) \in \mathcal{E}\}$, the set of successors of $i$.*
*A graph is said $p-$reg if it verifies:*

$$\begin{cases} |\mathcal{N}_{out}(i)| = 1 & \text{if } i \text{ is odd} \\ |\mathcal{N}_{out}(i)| = p & \text{if } i \text{ is even} \end{cases} \quad (28)$$

Graphs of this family have the particularity to have one half of their nodes with out-degree 1 and the other half with out-degree $p$. In particular if $p = 1$, the graph is 1-regular (it is a cycle). By increasing the value of $p$, the regularity worsens as shown in fig. 5. In the following numerical simulations, we consider a series of graphs composed of a p-reg graph and one additional measured node as in fig. 6. Therefore, with such graph, by tuning the value of $p$ we can modify the regularity of $G_{\mathcal{V}_2}$ while preserving the shape of the system. The graph used in the experiment contains 101 nodes (100 nodes in the p-reg graph plus one extra node to measure). We add some arbitrary inputs to the system and we compare the average value of the p-reg subgraph and the reconstructed value of this average for different value of $p$. The results are displayed in fig. 7. We notice that as predicted by Proposition 3 the error of reconstruction grows with the error

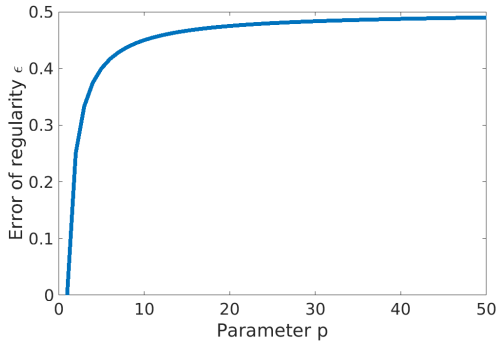



Fig. 5. The family of $p-$reg graphs allows to control the regularity of a graph.

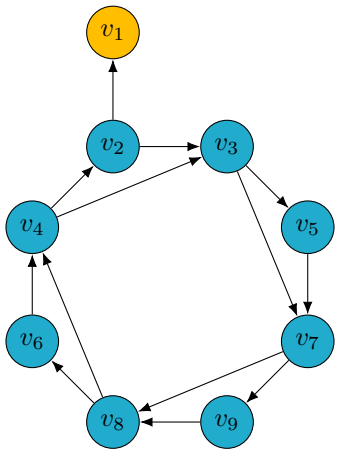



Fig. 6. A 2-reg graph as defined in Definition 4 with an additional node to observe. For the experiments, a similar graph is used with 100 nodes and $p$ varying from 1 to 100.

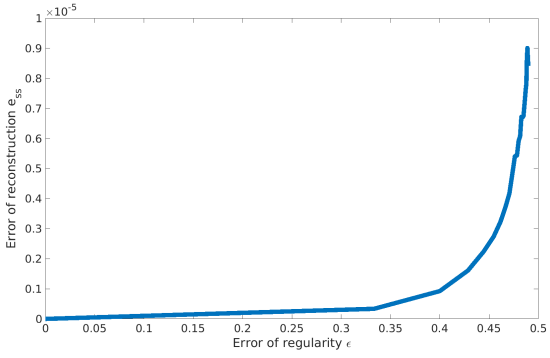



Fig. 7. Error of reconstruction in function of the error regularity for the family of p-reg graph. For each $p \in [1, ..., 100]$, 25 p-reg graph are generated with different inputs. The error of reconstruction is then computed as the bias between the signal and the reconstruction at $t = 1000$ (which is a good approximation of $e_{ss}$). This emphasizes the link between the regularity of a subgraph and the ability to reconstruct its average.

of regularity. This simulation enlights the fact that by minimizing the error of regularity $\epsilon$ we can control the error of reconstruction $e_{ss}$.

*Remark* 5. While detecting quasi-regular, if a small error is imposed on the regularity, the subgraph found might be small (and so the number of nodes to measure would be high). In the other hand, Proposition 3 ensures that the reconstruction would be better. Consequently, a compromise between the number of measure and the quality of the reconstruction has to be found. An interesting way to implement this compromise is by fixing a threshold for the error of regularity and then find the largest subgraph whose regularity error is lower than this threshold. This leads to the q-RIS detection problem defined hereafter.

**Problem 3** (Quasi-regular induced subgraph detection)**.** *Let $G$ be a graph and $\epsilon_0 > 0$ a threshold for the quasi-regularity. We look for the largest quasi-regular subgraph of $G$, which is:*

$$\begin{aligned} &\max_{I\subset\mathcal{V}} |I|, \\ &s.t.\ \epsilon(G_I) < \epsilon_0 \end{aligned} \tag{29}$$

*where $\epsilon(G)$ is the regularity error associated to $G$ as defined in* (18).

Therefore, by measuring the nodes outside the subgraph solution of Problem 3 we can reconstruct the average value of the subgraph with an error of the order of $\bar{\sigma}\frac{\epsilon_0}{1-\epsilon_0}$.

### 3.3 Algorithm for q-RIS detection

We present here an algorithm providing a sub-optimal solution to Problem 3. A combinatorial algorithm exploring every possible subgraph would find the exact solution but it is uncomputable in practice even for graphs with a relatively small size. We present here a beam-search algorithm allowing to find an approximating solution.

The principle of beam-search algorithms is as follows: A set of candidate solutions is considered as a seed, a set of solution deriving from these candidate is considered and the $\beta$ most *promising* are memorized ($\beta$ is called the beam width) and form the new set of candidate. The algorithm stops when a candidate solution is satisfying enough or when the new candidates are no more satisfying. In this latter case, the final solution is chosen among all the previous candidates.

For the quasi-RIS detection the algorithm is described in Algorithm 4: We initialize the set of candidates with the singletons of each node (line 1). Then we iterate while one of the candidate has a regularity error smaller than the minimum accepted $\epsilon_0$ (line 2) (see Remark 6 for a discussion on this point). At each iteration, new candidates are derived from the current candidates (line 4-5). These new candidates are all the subsets composed by one current candidate $c$ and any other nodes of the graph $\mathcal{S}_i$. Finally all these new candidates are united (line 7), the $\beta$ best form the new candidates (line 8) and the best one is stored as quasi$-$RIS (line 9). By repeating this operation several times, the size of the candidates grows until none of the candidate have a regularity error small enough. At the end, the candidate with the smallest regularity error is chosen. Figure 8 illustrates this algorithm. Beam-search algorithms, like this one, are a type of greedy algorithm and hence do not provide an optimal solution. However the computation is relatively fast and the solutions found are rather good, as we will see.

**Algorithm 4** quasi-RIS detection

**Require:** $G$: graph with $n$ nodes, $\epsilon_0$ maximum acceptable error, $\beta$ beam width

1: Cand $\leftarrow \{\{1\};\{2\};\ldots;\{n\}\}$
2: **while** $\min_{I\in\text{Cand}} \epsilon(G_I) < \epsilon_0$ **do**
3:  **for** $c \in$ Cand **do**
4:  $\mathcal{S} \leftarrow [1,\ldots,n]\backslash c$
5:  $J_c = \{c\cup\mathcal{S}_1; c\cup\mathcal{S}_2;\ldots; c\cup\mathcal{S}_{end}\}$
6:  **end for**
7:  $\Omega \leftarrow \bigcup_c J_c$
8:  Cand $\leftarrow \beta$ smallest $\epsilon(G_c)$ for $c \in \Omega$
9:  quasi-RIS $\leftarrow$ smallest $\epsilon(G_c)$ for $c \in \Omega$
10: **end while**

**Ensure:** quasi$-$RIS

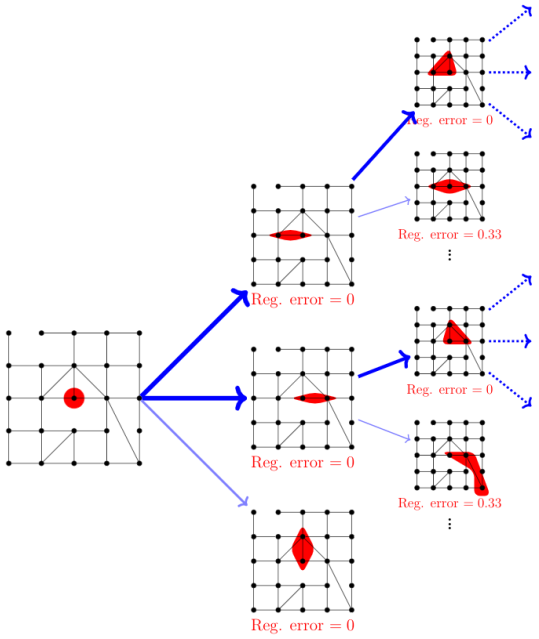
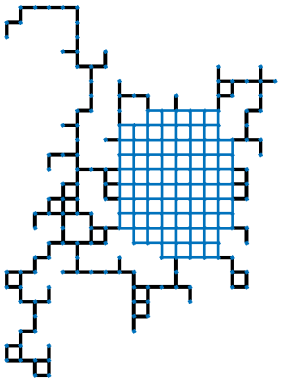


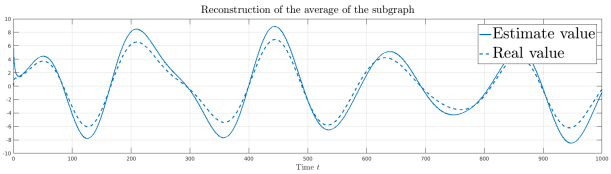

Fig. 8. Illustration of the algorithm 4 (on an undirected for readability). Here the beam width $\beta = 2$, so at each step the 2 best candidate are kept (solid line). Two light modifications are brought to make the example more readable: the seed is a single candidate while in the algorithm the seed is composed of the singletons of each node. Secondly, here the new candidates are the subsets composed of a previous candidate and a node *neighboring* this candidate while it is any node in the algorithm.

*Remark* 6. In Algorithm 4 the *while* loop will stop if none of the candidate have a regularity error small enough. However, this condition is relatively strict as it is possible that it is not verified at some iteration but it will be verified in the future. Thus, a relaxed condition would be to stop if the the condition is disrespected several iterations in a row. Then the solution would be the last candidate verifying the condition.

*Remark* 7. The new candidates are chosen in a way that may induce a disconnected subgraph. This allows to explore more potential solutions. However, for some applications the connectedness of the unobserved subgraph may be required or wished. For example when the network has a geographical nature (as urban traffic network or electrical grid), it is interesting to estimate the average of a geographical area. In this case, the algorithm can be adapted by changing line 4 with $\mathcal{S} \leftarrow \mathcal{N}_{in}(c)\cup\mathcal{N}_{out}(c)$. Figure 8



(a) In blue, the quasi-regular subgraph detected with algorithm 4.




(b) Reconstruction of the average of the quasi-regular subgraph obtained

Fig. 9. Illustration of the q-RIS approach. Our algorithm is able to detect a subgraph (a) which is enough regular to make the reconstruction of its average quite good (b).

implements this solution. See [26] for a discussion on the lowering of the solution due to the connectedness constraint.

*Simulation* 2. We present in this paragraph a simulation of Algorithm 4. We designed a graph endowed with a particularly regular subgraph to emphasize the ability of the algorithm to detect it. In this simulation the parameters are $\epsilon_0 = 0.1$ and $\beta = 300$ which means that at each step we conserve the 300 best candidates. The subgraph obtained is presented in fig. 9(a). We compute then the reconstruction of the average in this subgraph. Figure 9(b) shows the actual average and the estimation made by measuring the nodes outside the subgraph.

We have seen how to detect a regular or quasi-regular subgraph in order to estimate their average. We propose to generalize this approach, to detect several subgraphs and estimate different averages. The next section presents an extension of the current results to the multiple subgraphs case.

### 3.4 Extension to Multiple quasi-Regular Induced Subgraphs (mq-RIS)

In the previous problems we tried to find one regular or quasi-regular subgraph in order to estimate its average. We wonder, now, to which extent it is possible to detect several regular or quasi-regular subgraphs and estimate their respective average. Considering the limitations posed by the regularity case evoked in section 2.6, we focus only on the quasi-regular problem. However a similar generalisation can be led for the regularity case. In the RIS and q-RIS problems we wanted to have the minimum number of nodes to measure leading to the minimisation problems 2 and 3. Here again we have the same objective to minimize the number of measured nodes. Thus, we want to find disjoint quasi-regular induced subgraphs $G_{I_1}, \dots, G_{I_m}$ maximizing the cardinality of the union of all the subgraph. We denote $\mathcal{I} = [I_1, \dots, I_m]$ the set of the subsets. Moreover these subgraphs can not share any successors, i.e. nodes outside the subgraph and pointed by a node of the subgraph. This is because the successors of a subgraph are measured to estimate the average value of the subgraph. If a node is pointed by two nodes belonging to two different subgraphs, the condition for the reconstruction does not hold (see [9]).

**Problem 4** (Multi quasi-regular induced subgraph detection)**.** *Let $G$ be a graph and $\epsilon_0 > 0$ a threshold for quasi-regularity. We look for a set of quasi-regular subsets $\mathcal{I}$ minimizing the number of nodes to measure, which is:*

$$\begin{aligned} &\max_{\mathcal{I}=[I_1,\dots,I_m]} \left| \bigcup_j I_j \right|, \\ &s.t.\ \forall i,\ \epsilon(G_{I_i}) < \epsilon_0 \\ &\quad \forall i, j,\ (I_i \cup \mathcal{N}_{out}(I_i)) \cap (I_j \cup \mathcal{N}_{out}(I_j)) = \emptyset \end{aligned} \tag{30}$$

The second constraint translates the non-overlapping of the subgraphs and their successors. The quasi-RIS detection algorithm 4 presented in the previous section can be extended almost straightforwardly to the multiple subgraphs case as follows: a first quasi-regular subgraph is detected, the subgraph and its neighborhood (which are the nodes to measure) are removed from the graph and the process is repeated with the new graph. To limit the number of nodes to measure it is interesting to limit the number of neighbors of the subgraph selected at each iteration. To this aim, we propose to find at each iteration the subgraph maximizing $|I|/|\mathcal{N}_{out}(I)|$ i.e. the ratio between its size and the size of its neighborhood instead of the subgraph minimizing $\epsilon(G_I)$ in Algorithm 4.

The algorithm for multi quasi-RIS detection is described in Algorithm 5 where quasi-RIS$^\star$ refers to the algorithm 4 where line (9) has been replaced by

$$\text{quasi-RIS} = \underset{c\in\Omega,\, \epsilon(G_c)<\epsilon_0}{\arg\max} \frac{|c|}{|\mathcal{N}_{out}(c)|} \tag{31}$$

**Algorithm 5** Multi quasi-RIS detection

**Require:** $G$: graph with $n$ nodes, $\epsilon_0$ maximum acceptable error, $\beta$ beam width, $i_{max}$ maximum number of subgraph detected
1: $\mathcal{I} = []$
2: $I = \text{quasi-RIS}^\star(G, \epsilon_0, \beta)$
3: **while** $I$ is not empty **do**
4: $\quad \mathcal{I} = \mathcal{I} \cup I$
5: $\quad G$ = remove $I \cup \mathcal{N}_{out}(I)$ from $G$
6: $\quad I = \text{quasi-RIS}^\star(G, \epsilon_0, \beta)$
7: **end while**
**Ensure:** $\mathcal{I}$ set of subsets inducing multiple quasi-RIS

*Simulation* 3. We propose here a simulation of Algorithm 5 for the mq-RIS detection. To this end, we consider an initial graph (fig. 10) designed with five zones more regular than the average. This aims to test the capacity of the algorithm to detect regular subgraphs which are almost invisible to the naked-eye. The result of the simulation is displayed in fig. 11. In this case, the quasi-RIS detection algorithm is applied five times before no more satisfying subgraphs can be found. At each step, we can see that the subgraph and its out-neighborhood found at the previous step is removed and the algorithm is applied with the graph obtained.

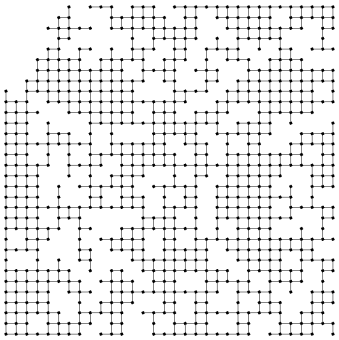


Fig. 10. The initial graph is a grid with some irregularities making certain zones more regular. The algorithms aims to find these zones.

## 4 Application to network epidemiology

In this last section we apply the mq-RIS algorithm to a real-world case: we consider the spreading of a disease over a contact network of the main cities in France and we aime to estimate the evolution of the proportion of infected people in different areas.

### 4.1 The epidemiological model

Several models have been developed to capture the mechanisms of disease spreading, the most common ones are the so-called compartmental models. In these models, the population is divided in compartments representing the state (infected, susceptible, recovered, ...) of the individuals [27]. Moreover these type of models may have an underlying network structure. In this case, nodes represent individuals or group of individuals and edges represent interactions between individuals or between the groups. A state is assigned to each node and a dynamical equation describes the evolution in function of the state of the node, the state of its neighbors and the parameters of the disease. Among this network compartmental models, we consider here one of the most commonly used: the SIS model. Within this model, each node can be susceptible (S) or infected (I) and can pass from one state to another with a certain probability: as shown in fig. 12, an infected individual may recover with probability $\delta$ (the recovery rate) and a susceptible individual may be infected with probability $\beta$ (the infection rate) scaled by the state of its neighborhood. The dynamics is detailed hereafter:

$$\begin{aligned} X_i : 0 \to 1 \quad &\text{with rate } \beta \sum_{\mathcal{N}_{in}(i)} X_j \\ X_i : 1 \to 0 \quad &\text{with rate } \delta \end{aligned} \tag{32}$$

where $X_i = 0$ means that $i$ is susceptible and $X_i = 1$ means that node $i$ is infected and where $\mathcal{N}_{in}(i)$ is the set of predecessors of $i$. We consider here a mean-field approximation of the SIS model introduced in [28]. In this case, the nodes of the network does not represent individuals but groups of people. The nodes are not in a determined state $S$ or $I$ but have a proportion $p$ of people infected. The dynamics of $p$ is then:

$$\dot{p}(t) = (AB - \Delta)p(t) - P(t)ABp(t), \quad p(0) = p_0 \tag{33}$$

where $p = [p_1, \dots, p_n]$ are the proportions of infected people in each group (and $P = \mathrm{diag}(p)$), $A$ is the adjacency matrix of the underlying network, $B = \mathrm{diag}(\beta_1, \dots, \beta_n)$, $\Delta = \mathrm{diag}(\delta_1, \dots, \delta_n)$ are the parameters of the epidemic and $p_0$ is the initial proportion of infected people in each group. In order to use the approach developed before, we consider a linearized version [27] of the dynamics in (33):

$$\dot{p}(t) = \underbrace{(AB - \Delta)}_{\mathcal{A}} p(t), \quad p(0) = p_0 \tag{34}$$

To complete the system we add some inputs:

$$\dot{p}(t) = \mathcal{A}p(t) + \mathcal{B}u(t), \quad p(0) = p_0 \tag{35}$$

The matrix $\mathcal{B} \in \{0;1\}^{n \times b}$ identifies some nodes which are in contact with a source of infection[3], and the function $u(t) \in \mathbb{R}^{b \times 1}$ gives the dynamics of each source of infection. We consider that $\forall i,\ \beta_i = \beta$ and $\delta_i = \delta$ (the rate of infection and recovery are the same in the whole network) resulting in an adjacency matrix with the following form:

$$\mathcal{A} = \begin{pmatrix} -\delta & \beta & 0 & \beta & 0 & 0 & 0 \\ 0 & -\delta & 0 & 0 & \beta & \beta & 0 \\ \beta & 0 & -\delta & 0 & 0 & 0 & 0 \\ \beta & 0 & \beta & -\delta & 0 & \beta & 0 \\ 0 & 0 & 0 & 0 & -\delta & 0 & \beta \\ 0 & 0 & 0 & \beta & \beta & -\delta & 0 \\ 0 & 0 & 0 & 0 & 0 & \beta & -\delta \end{pmatrix} \tag{36}$$

This adjacency matrix correspond to a system which falls in the scope of positively uniform system with a large negative self-loop as discussed in Remark 1.

### 4.2 The estimation problem

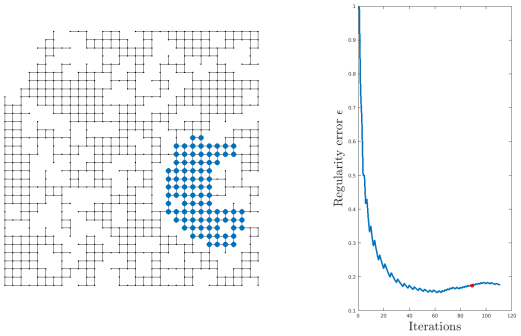
It is often relevant to estimate the evolution of a disease: as it is very costly to determine the state of each individual, one needs methods to reconstruct the spreading of the epidemic. Based on such methods, it is then possible to propose an heatmap as in fig. 13 allowing for example to take appropriate sanitary measures, or to study the efficiency of a treatment.

We propose here to use our approach to estimate the evolution of the proportion of infected people in different areas, by measuring the state of few groups. Thus, we consider here the propagation of a disease over a network representing a contact network of the main French cities (detailed after). By means of the mq-RIS approach presented above, we aim to estimate the evolution of the epidemic with few measured nodes.

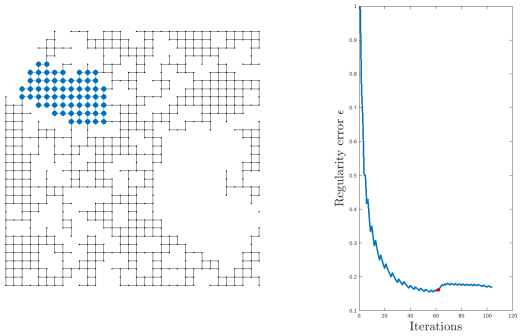
### 4.3 Construction of the graph

We consider a graph of interaction between groups of individuals at a country scale. The graph is structured at two different scales: a level within the cities and a level between cities. It is known that at the level of a city, individuals are strongly interconnected and tend to form clusters [29], [30]. Here we use the Watts-Strogatz model [31] which is known to well capture the features of social networks. At the level of the country, there are fewer connections between groups of different cities and the number of connections between two cities is proportional to their number of inhabitants and inversely correlated to their distance. We generate such a network by considering twenty two of the most populated cities in France (fig. 14). The details of the construction of the graph are given in table 1.

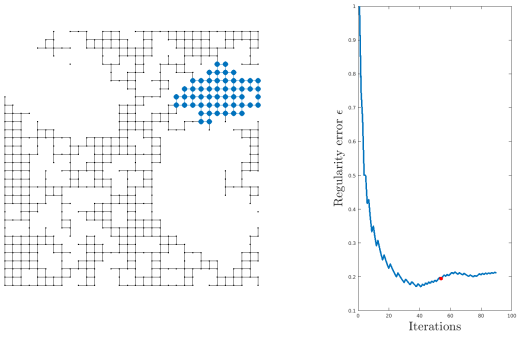
3. Sources of infection may be a polluted area, insect bites, a population of infected animals, ...



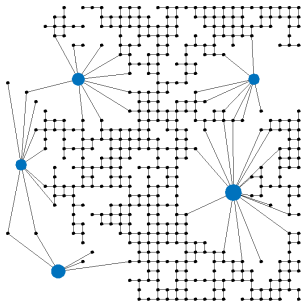

(a) A first subgraph is found (in blue). On the right the evolution of the regularity error through the iterations. The red dot corresponds to the iteration where the subgraph maximizing (31) is found.




(b) The first subgraph found and its successors are removed and the algorithm is applied again to detect a second subgraph (in blue).




(c) The second subgraph found and its successors are removed and the algorithm is applied again to detect a third subgraph (in blue).



(d) The graph obtained when merging the subgraphs obtained at each steps. By measuring the remaining nodes, the average of the subgraphs detected can be estimated.

Fig. 11. **(a-c)** left: in blue the subgraph within the current graph; right: the evolution of the regularity error of the best candidate is displayed through the iteration. Moreover a red dot shows the iteration of the selected subgraph. Two other subgraphs are found but it is not shown here. **(d)** represents the graph obtained by merging the different subgraphs. This reduced graph offers an estimation of the initial graph, and thus can be seen as an aggregation of it.

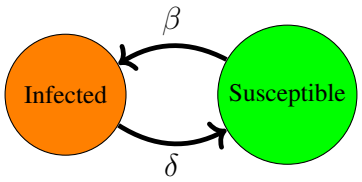



Fig. 12. Sketch of the SIS model

| **Population** | |
|---|---|
| Number of groups (nodes) | 1404 |
| Pop. per groups | 5000 |
| Number of cities | 22 |
| **Graph** | |
| Model within cities | Watts-Strogatz |
| Mean degree $K$ | 10 |
| Prob. rewire $\beta$ | 0.1 |
| Prob. connection inter-cities | $\frac{e^{d^2/10}}{850}$ |
| Number of input | 323 |
| **SIS model** | |
| Infection rate $\beta$ | 0.05 |
| Recovery rate $\delta$ | 0.98 |

TABLE 1
Parameters for the graph of interaction and the SIS model

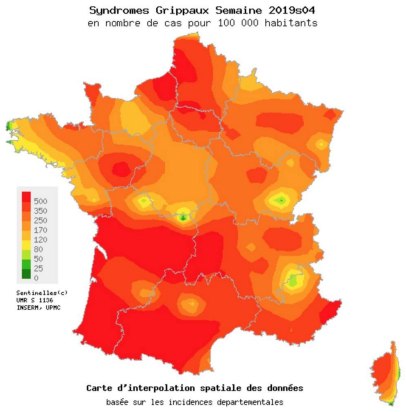



Fig. 13. Interpolating map of the number of individual infected with flu for 100000 inhabitants in January 2019 in France. Map available on www.sentiweb.fr. ©Inserm

### 4.4 Simulation

Figure 15 presents the partition found with algorithm 5. The graph has been divided in 11 subgraphs containing 1112 nodes in total. Thus, only 292 nodes remains to measure which represents only 20.80% of the nodes. Some of the parts fit cities while others include a whole region. The figure gives also the regularity error $\epsilon$,

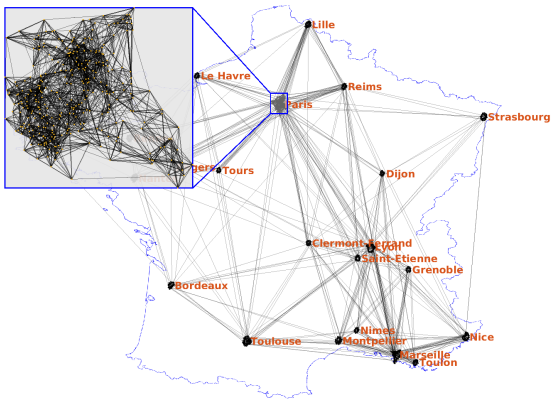



Fig. 14. Graph of interaction of the main cities of France. It is composed of 1404 nodes, each representing a population of 5000 individuals. The subgraphs within the cities is based on the Watts-Strogatz model while the graph between cities is a random graph where the probability of connection between two nodes decreases exponentially with the distance.

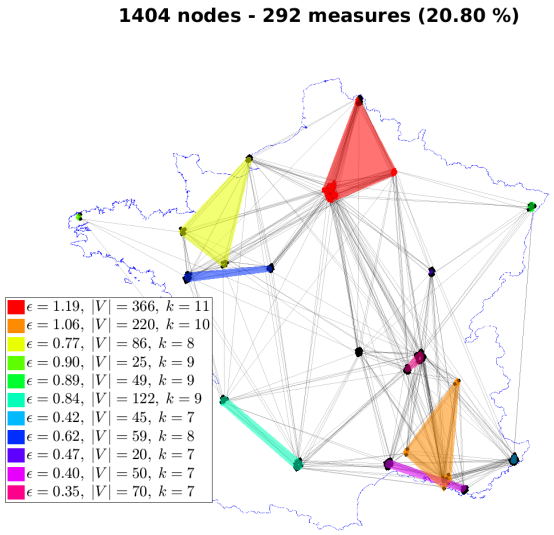



Fig. 15. Partition obtained via the mq-RIS algorithm 5. The legend gives the error of regularity, the number of nodes and the mean degree for each subgraph detected.

the number of nodes $|\mathcal{V}|$ and the mean degree $k$ for each subgraph.

Now that we have this partition where each subgraph is quasi-regular it is possible to estimate the value of the average inside the subgraphs. We chose $\beta = 0.05$ and $\delta = 0.98$ and arbitrarly add 323 sources of infection randomly distributed in the territory. To be close to the reality we use as initial conditions the situation presented in fig. 13 and available on *www.sentiweb.fr*. Figure 16 shows the evolution of the proportion of infected individuals inside each subgraph and the estimation made with the open-loop observer $\hat{x}_2^{av}$ described in (7). The solid lines are the actual averages while the dotted lines are the estimated average. Figure 17 shows the evolution of the absolute error for each subgraph. We observe that, as expected, the estimations errors decrease quickly and remain relatively small. However, as we did not find exact regular subgraphs, the system is not detectable and the error does not converge to zero. Instead there is a small residual error which is acceptable.

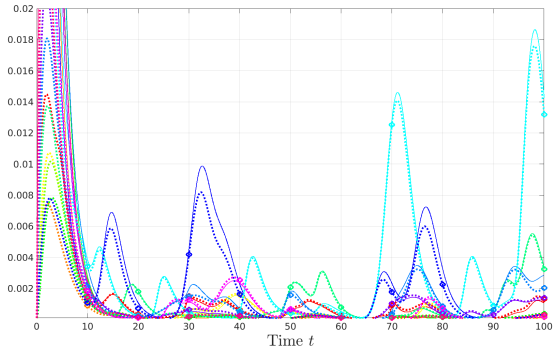



Fig. 16. Proportion of infected individuals within each subgraph. The solid lines are the ground-truth values and the dotted lines are the average estimated from the measurements.

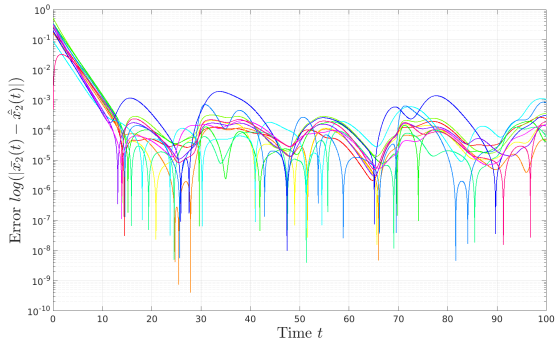



Fig. 17. Semi-log representation of absolute errors between the ground-truth value and the estimated value for each subgraph. The errors decrease but they do not converge to zero.

## 5 Conclusion

Based on the novel notion of average detectability we proposed here three algorithms providing measured node placement in order to estimate the average of a subset of the system. Considering a particular type of system, the first algorithm finds regular induced subgraph to reach full average detectability, i.e. the estimation of the average is asymptotically unbiased. Due to the limitation of this first problem, we proposed a relaxation: we focused on the detection of *quasi-regular* induced subgraph which results in an estimation of the average with a bias which depends on the quasi-regularity. The second algorithm achieves this task. The third algorithm allows to detect several quasi-regular induced subgraphs to estimate the averages of different subsets of the system. Finally, we apply the third algorithm to estimate the evolution of a disease spreading in a contact network on French territory.

## Acknowledgements

This project has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation program (grant agreement n° 694209)

## References

[1] F. F. Wu and A. Monticelli, "Network observability: theory," *IEEE Transactions on Power Apparatus and Systems*, no. 5, pp. 1042–1048, 1985.

[2] B. Gou and A. Abur, "An improved measurement placement algorithm for network observability," *IEEE Transactions on Power Systems*, vol. 16, no. 4, pp. 819–824, 2001.

[3] S. H. H. J. X. Xin, "Algorithm research for selecting measurement nodes in measuring network traffic based on adjacency matrix," *Journal of Southeast University (Natural Science Edition)*, p. S1, 2008.

[4] F. L. Iudice, F. Sorrentino, and F. Garofalo, "On node controllability and observability in complex dynamical networks," *IEEE Control Systems Letters*, 2019.

[5] Y.-Y. Liu, J.-J. Slotine, and A.-L. Barabási, "Controllability of complex networks," *nature*, vol. 473, no. 7346, p. 167, 2011.

[6] G. Notarstefano and G. Parlangeli, "Controllability and observability of grid graphs via reduction and symmetries," *IEEE Transactions on Automatic Control*, vol. 58, no. 7, pp. 1719–1731, 2013.

[7] A. Rahmani, M. Ji, M. Mesbahi, and M. Egerstedt, "Controllability of multi-agent systems from a graph-theoretic perspective," *SIAM Journal on Control and Optimization*, vol. 48, no. 1, pp. 162–186, 2009.

[8] M. Ji and M. Egerstedt, "Observability and estimation in distributed sensor networks," in *2007 46th IEEE Conference on Decision and Control*. IEEE, 2007, pp. 4221–4226.

[9] M. U. B. Niazi, C. Canudas de Wit, and A. Kibangou, "Average observability of large-scale network systems," in *ECC 2019 - European Control Conference*, Naples, Italy, Jun. 2019, pp. 1–6. [Online]. Available: https://hal.archives-ouvertes.fr/hal-02073668

[10] I. M. Bomze, M. Budinich, P. M. Pardalos, and M. Pelillo, "The maximum clique problem," in *Handbook of combinatorial optimization*. Springer, 1999, pp. 1–74.

[11] Y. Chi, R. R. Muntz, S. Nijssen, and J. N. Kok, "Frequent subtree mining–an overview," *Fundamenta Informaticae*, vol. 66, no. 1-2, pp. 161–198, 2005.

[12] M. M. SysŁ *et al.*, "The subgraph isomorphism problem for outerplanar graphs," *Theoretical Computer Science*, vol. 17, no. 1, pp. 91–97, 1982.

[13] S. Imane, T. Baptiste, and R. Frédéric, "Discrete linear functional observer for the thermal estimation in power modules," in *2018 IEEE 18th International Power Electronics and Motion Control Conference (PEMC)*. IEEE, 2018, pp. 812–817.

[14] L. Lovász, "On the Shannon capacity of a graph," *IEEE Transactions on Information theory*, vol. 25, no. 1, pp. 1–7, 1979.

[15] D. M. Cardoso, M. Kamiński, and V. Lozin, "Maximum k-regular induced subgraphs," *Journal of Combinatorial Optimization*, vol. 14, no. 4, pp. 455–463, 2007.

[16] Y. Asahiro, H. Eto, T. Ito, and E. Miyano, "Complexity of finding maximum regular induced subgraphs with prescribed degree," *Theoretical Computer Science*, vol. 550, pp. 21–35, 2014.

[17] V. V. Lozin and R. Mosca, "Maximum regular induced subgraphs in 2p3-free graphs," *Theoretical Computer Science*, vol. 460, pp. 26–33, 2012.

[18] D. Kobler and U. Rotics, "Finding maximum induced matchings in subclasses of claw-free andp 5-free graphs, and in graphs with matching and induced matchingof equal maximum size," *Algorithmica*, vol. 37, no. 4, pp. 327–346, Dec 2003. [Online]. Available: https://doi.org/10.1007/s00453-003-1035-4

[19] F. V. Fomin, F. Grandoni, and D. Kratsch, "Measure and conquer: a simple o(20.288n) independent set algorithm," in *SODA*, 2006.

[20] J. M. Robson, "Algorithms for maximum independent sets," *Journal of Algorithms*, vol. 7, no. 3, pp. 425–440, 1986.

[21] S. Gupta, V. Raman, and S. Saurabh, "Fast exponential algorithms for maximum r-regular induced subgraph problems," in *International Conference on Foundations of Software Technology and Theoretical Computer Science*. Springer, 2006, pp. 139–151.

[22] A. Agra, G. Dahl, T. A. Haufmann, and S. J. Pinheiro, "The k-regular induced subgraph problem," *Discrete Applied Mathematics*, vol. 222, pp. 14–30, 2017.

[23] C. Luz, "Improving an upper bound on the size of k-regular induced subgraphs," *J. Comb. Optim.*, vol. 22, pp. 882–894, 11 2011.

[24] C. J. Luz, "An upper bound on the independence number of a graph computable in polynomial-time," *Operations Research Letters*, vol. 18, no. 3, pp. 139 – 145, 1995. [Online]. Available: http://www.sciencedirect.com/science/article/pii/0167637795000429

[25] Y. Asahiro, H. Eto, and E. Miyano, "Inapproximability of maximum r-regular induced connected subgraph problems," *IEICE Transactions on Information and Systems*, vol. E96.D, no. 3, pp. 443–449, 2013.

[26] N. Martin, P. Frasca, T. Ishizaki, J.-I. Imura, and C. Canudas-De-Wit, "The price of connectedness in graph partitioning problems," in *2019 18th European Control Conference (ECC)*. IEEE, 2019, pp. 2313–2318.

[27] C. Nowzari, V. M. Preciado, and G. J. Pappas, "Analysis and control of epidemics: A survey of spreading processes on complex networks," *IEEE Control Systems Magazine*, vol. 36, no. 1, pp. 26–46, 2016.

[28] P. Van Mieghem, "The n-intertwined SIS epidemic network model," *Computing*, vol. 93, no. 2-4, pp. 147–169, 2011.

[29] M. Boguná, R. Pastor-Satorras, A. Díaz-Guilera, and A. Arenas, "Models of social networks based on social distance attachment," *Physical review E*, vol. 70, no. 5, p. 056122, 2004.

[30] R. Toivonen, J.-P. Onnela, J. Saramäki, J. Hyvönen, and K. Kaski, "A model for social networks," *Physica A: Statistical Mechanics and its Applications*, vol. 371, no. 2, pp. 851–860, 2006.

[31] D. J. Watts and S. H. Strogatz, "Collective dynamics of small-world networks," *Nature*, vol. 393, no. 6684, p. 440, 1998.


**Nicolas Martin** received the M.Sc. degree in applied mathematics from Institut National des Sciences Appliquées (INSA) of Toulouse, France in 2016. In 2015, he made an internship at the European Southern Observatory (ESO) in Chile, Santiago. In 2016, he made an internship at the Centre National d'Etude Spatiale (CNES) in Toulouse, France. Since 2016 he is a Ph.D. student in GIPSA-lab at Grenoble, France, under the supervision of Carlos Canudas-de-Wit and Paolo Frasca.

**Paolo Frasca** (M'13, SM'18) received the Ph.D. degree in Mathematics for Engineering Sciences from Politecnico di Torino, Italy, in 2009. From 2013 to 2016, he has been an Assistant Professor at the University of Twente, the Netherlands. Since October 2016, he is a CNRS researcher at GIPSAlab, Grenoble, France. His research interests are in the theory of network systems and cyber-physical systems, with applications to robotic, sensor, infrastructural, and social networks.

**Carlos Canudas-de-Wit** was born in Villahermosa, Mexico, in 1958. He received the B.S. degree in electronics and communications from the Monterrey Institute of Technology and Higher Education, Monterrey, Mexico, in 1980, and the M.S. and Ph.D. degrees in automatic control from the Department of Automatic Control, Grenoble Institute of Technology, Grenoble, France, in 1984 and 1987, respectively. He is currently a Directeur de recherche (Senior Researcher) with CNRS, Grenoble, where he is the Leader of the NeCS Team, a joint team of GIPSA-Lab (CNRS) and INRIA, on networked controlled systems. Dr. Canudas-de-Wit is an IEEE-CSS Fellow and an IFAC Fellow. He is an Associate Editor of the IEEE TRANSACTIONS ON AUTOMATIC CONTROL, the Automatica, the IEEE TRANSACTIONS ON CONTROL SYSTEMS TECHNOLOGY, the Asian Journal of Control, and the IEEE TRANSACTIONS ON CONTROL OF NETWORK SYSTEMS. He served as the President of the European Control Association from 2013 to 2015, and the IEEE Board of Governors of the Control System Society from 2011 to 2014. He holds the ERC Advanced Grant Scale-FreeBack from 2016 (1999–2002).