Morpho-statistical description of networks through graph modelling and Bayesian inference

Quentin Laporte-Chabasse, Radu S. Stoica, Marianne Clausel, François Charoy, Gérald Oster

Morpho-statistical description of networks through graph modelling and Bayesian inference

Quentin Laporte-Chabasse *, Radu S. Stoica †, Marianne Clausel †, François Charoy *, Gérald Oster *
*Université de Lorraine, CNRS, Inria, LORIA, F-54000 Nancy, France
{quentin.laporte-chabasse, francois.charoy, gerald.oster}@loria.fr
†Université de Lorraine, CNRS, IECL, F-54000 Nancy, France
{radu-stefan.stoica, marianne.clausel}@univ-lorraine.fr

Abstract—Collaboration graphs are relevant sources of information to understand behavioural tendencies of groups of individuals. The study of these graphs enables figuring out factors that may affect the efficiency and the sustainability of cooperative work. For example, such a collaboration involves researchers who develop relationships with their external counterparts to address scientific challenges. As relations and projects change over time, the evolution of social structures must be tackled. We propose a statistical approach considering different structural collaboration patterns and captures the dynamic of the relational structures over the years. Our approach combines spatial processes modelling and Exponential Random Graph Models used to analyse social processes. Since the normalising constant involved in classical Markov Chain Monte Carlo (MCMC) approaches is intractable, the inference remains challenging. To overcome this issue, we propose a Bayesian tool that relies on the recent ABC Shadow algorithm. The method is illustrated on real data sets from an open archive of scholarly documents. Through a simple formalism, our approach highlights the interactions between the different types of social relations at stake in the collaboration network.

1 Introduction

Networks are widely studied mathematical objects [1], [2]. They describe molecular interactions, relationships between individuals in a social application, collaboration links among organisations, etc.

For example, when different organisations collaborate to produce new scientific results, a part of these results are presented through scientific papers. The publication process induces a network describing interactions among the organisations involved in this process. The network is made of the co-authorship relation of researchers belonging to the different organisations. This representation of collaboration among scientists has already been deeply investigated [3]–[5]. Most of the previous work in the state of the art had a global approach considering communities of researchers. They highlighted some properties related to small-world model and preferential attachment mechanisms. The co-authoring graph is used here for a slightly different purpose. Our work focuses on more structural aspects of the co-authoring graph. The approach we present, aims to characterise the occurrence of relational links among researchers.

The data set serving as a paradigm here is the set of scientific publications produced by LORIA[1], over the three years : 2017, 2018 and 2019. The laboratory is organised in 28 scientific teams. The data was gathered from the open publication archive HAL (https://data.archives-ouvertes.fr). We collected all the publications submitted in the period 2017-2019 with at least one author member of LORIA. The co-authorship networks of each year are represented by a graph structure. The network captured for the year 2018 is shown in Figure 1 as the main picture. Thumbnails of the three co-authorship networks (one for each year) are represented on the right-hand side. The nodes of the graph are researchers. An edge of the graph represents the link between two researchers who collaborated in 2018. Nodes are coloured according to researchers affiliation. LORIA members are coloured in yellow. All the other institutions have their own dedicated colour. The graph representing collaborations in 2018 is composed of 616 collaborators (nodes). The number of nodes for the graphs in 2017 and 2019 are 731 and 1090 respectively.

Several connected components are visible among the three networks. This tends to reflect the team oriented activity developed by the Lab. Looking at a single connected component or at a single research team raises several questions:

- What determines the occurrence of a collaboration link? The link between two researchers is not a random connection phenomenon in a social network. The resulting graph components may look more "clustered" or more "repulsive" than in a purely random network.
- How cooperation relation between individuals can be characterised? Inside a research team, people cooperate with members of the same team or from other institutes. Some of the researchers are able to maintain both types of cooperation. We call them "hubs".

1. The equivalents in French for "Lorraine Research Laboratory in Computer Science and its Applications" https://www.loria.fr

- How to characterise the cooperative patterns of a research team over time? The structure and the type of interactions, the presence of hubs evolve as the projects and the stakeholders change over time.

The aim of this paper is to propose a "morpho-statistical" methodology approach for network description that will answer these questions. We propose a model which captures the morphological aspects of the observed graph. Each parameter of the model gives a concrete meaning related to the strength of relational links among collaborators from different organisations. The underlying motivation is to study how researchers from different labs interact in their own lab and with the outside. The evolution of the parameters fitting observations for different years illustrates the dynamic of the structure encompass by the model. The stochastic model we propose is inspired by Exponential Random Graph Models (ERGMs) and Markov random graph modelling. Difficulties related to the estimation of such model, leads us to adopt a Bayesian inference strategy based on Monte-Carlo simulations.

Bayesian inference allows us to sample the posterior distribution of parmeters which represents the distribution of models that can explain the observation. This is a complementary statistical analysis tool to the classical maximum likelihood methods. The proposed approach allows statistics computations and tests for each parameter, while including the prior knowledge available for them.

The structure of the paper is as follows. Section 2 presents the modelling of the network as a line graph, obtained by transforming the nodes of the initial graph into edges, and the previous edges into nodes. Our application considers the network as a graph with edges given by the researchers and the nodes given by the co-authorship link. This underscores the collaboration over the people.

Networks seen as labelled graphs are complex systems. The different labels illustrate the diversity of relational ties, but they induce an extremely high number of configurations. Stochastic modelling allows us to deal with this situation. The approach we propose considers an appropriate version of Exponential Random Graph Models (ERGMs) to represent collaborations initiated by a community of researchers.

The model presented in Section 2.3.1 is inspired by Potts or Ising like models. The model distribution exhibits a normalising constant that is intractable. Therefore, we use Monte Carlo methods to perform statistical inference. We provide at the end of Section 2 a presentation of the simulation algorithms: the Metropolis-Hastings (MH) dynamics and the Gibbs sampler. Next, in Section 3, we describe the ABC Shadow algorithm [6] used to build posterior-based inference. Section 4 demonstrates the relevance of ABC Shadow on simulated data.

The remainder of the paper (Section 5) is dedicated to the practical application based on real data analysis. The case study handles the structures of scientific collaborations of research teams from the LORIA laboratory. We consider three years of publication (2017, 2018 and 2019) as they can illustrate the evolution of collaborations over the years. The ABC Shadow algorithm is applied to this dataset providing the whole a posteriori distribution of the model. Thereafter, the output results are used to perform parameter estimation, statistical tests and classification procedures, in order to analyse and characterise the collaboration patterns within this institution. We propose here an approach based on a fixed time step to study the dynamic of the collaboration structures over the years. The posterior distributions of each year are put together so as to assess the evolution of structural features controlled by the model.

Finally, in Section 6, conclusions and perspectives are depicted. Source code, notebooks and instructions used for this paper are provided in a GIT repository https://github.com/quentinl-c/ABCShadow_article_assets.

2 Modelling social networks

Graphs have been used to model social networks in sociology [7], [8]. We propose to understand intra and inter relation between organisations based on participant collaboration network.

2.1 Related work

The study of interactions between groups of individuals requires adapted modelling. For instance,it must encode group affiliation information. We can represent these intra and inter-group links with several hierarchically organised networks also called multi-level networks [9], [10]. Each level represents an observation of the social structure at a different scale. Let us take the case of a company that is structured in departments [11]. The lowest level represents the interactions between the employees of the company. The top level shows the links between the different departments. Finally, an intermediate level represents the affiliation of individuals in the different departments. This kind of representation is suitable for the study of nested social structures, but it is very complex as a variety of interactions within and between these different layers are at stake.

Another approach is to take into account higher order interactions [12]. Instead of representing social structures only by pairwise interactions, we can consider interactions between more complex structures (groups of people e.g.). These kinds of representations involves multi-dimensional mathematical objects such as hypergraphs [13] or simplicial complexes [14]. These higher order interactions bring us closer to the complexity of the observed world. However, they do not allow us to intuit the individual and his or her direct contribution to the topology of collaborative links.

From the collaboration graph illustrated by Figure 1, we will associate a more relevant one, considering the relation as the object of primary interest taking into account that collaboration links can be internal or external. This is what we explain in the next subsection.

2.2 Network representation through line graphs

Usually, social structure studies are conducted on graphs whose vertices are individuals and links represent social ties, as in Figure 1. Here we use a representation relying on the dual graph of the network, the so-called *line graph*. This graph is obtained from the initial graph by transforming edges into nodes, and nodes into edges, as in [15]. This principle is illustrated in Figures 2a and 2b. The first example in

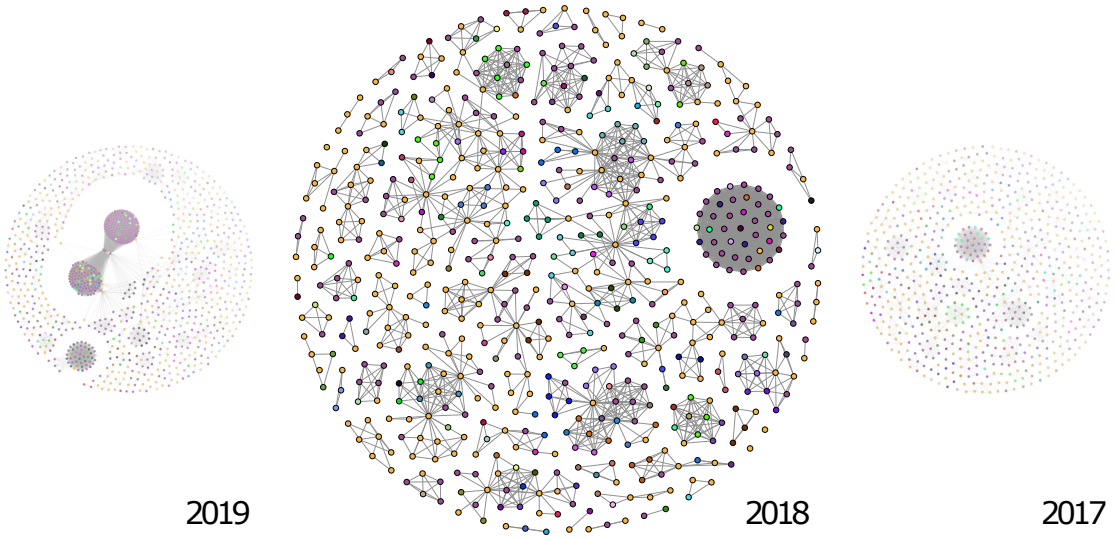
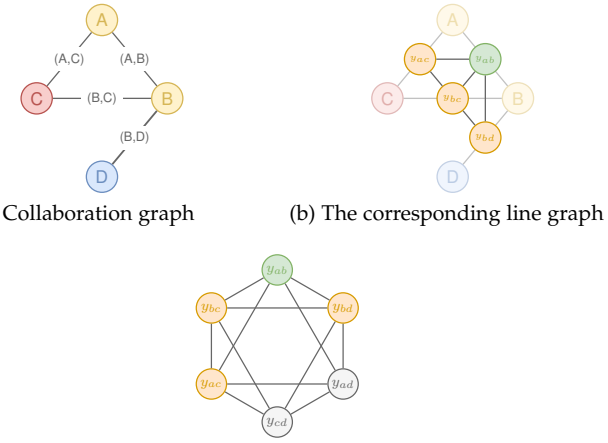


Fig. 1. Collaborations among researchers within the LORIA laboratory during the period 2017-2019 – The main network represents the co-authorship graph in 2018. Each node of the graph represents a researcher, the edges are collaboration links and nodes' colour represent the affiliation to a laboratory. For example, all LORIA members are coloured in yellow, while the members of the other labs are differently coloured. We observe highly connected nodes and patterns that evolve over the years. How to quantify this evolution?

Figure 2a shows the dual transformation of a graph towards its dual. The second example shows that the dual graph is not necessarily a complete graph. The line graph provides a representation of the network that emphasises relationships over people and allows us to reason on these relationships and the structure they propose.

Throughout this paper, we assess the extent to which inter and intra-organisational links occur. In the example presented in Figure 2a, A and B represent researchers working in the organisation of interest -in our case LORIA- while C and D are researchers working at other institutes. The augmented line graph in Figure 2c describes the structure of the type of interactions as follows. The green node is an intra-organisational relation, the orange ones are inter-organisational relations, while the grey ones represent a nil relation. This last type of relation represents two researchers potentially connected that do not work together at all.



(a) Collaboration graph

(b) The corresponding line graph

(c) Augmented line graph representation *taking nil edges into account*

Fig. 2. An example of collaboration graph and its line graph representation . In the collaboration graph (Figure 2a), the nodes represent individuals and coloured according to the affiliation : organisation 1 ●, organisation 2 ● and organisation 3 ●. In the line graph (Figures 2b and 2c), the nodes represent the relations (*i.e.* links in the collaboration graph) and are coloured according to which type the relation is : intra-organisational ● or inter-organisational ●

2.3 Markov Random Fields on graphs.

The example of Figure 2c illustrates our two main questions. The first one is the morpho-statistical description of different interactions in a social network. The second one is the description of the labelling distribution of the nodes in a graph. To each vertex of the line graph is associated a label, depending on the kind of link of the corresponding edge of the social graph *nil, intra_organisational, inter_organisational*.

The uncertain and dynamic nature of the individuals' behaviour recommends stochastic modelling of social interactions [16]. Within this context we propose a random graph model whose parameters provide a meaningful description of the social network of interest. Markov Random Fields (MRFs) provide a framework to deal with this type of problems [17]–[19]. They are also known in literature related to social networks modelling under the name of ERGMs [20], [21].

The graph illustrated by the the Figure2c can be more formally defined as a MRF. Let $\mathcal{G}$ be the considered line graph, with $\mathcal{V} = \{1, \ldots, n\}$ the vertices index set, $\mathcal{E} = \{e_{ij} | i \sim j, \forall i, j \in \mathcal{V}\}$ the set of its edges and $\mathcal{L} = \{(\ell_1, \cdots, \ell_m\}$ the set of possible labels. The structure of $\mathcal{L}$ was chosen discrete for the sake of the simplicity and for the purpose on the application on hand. Its description using more general measurable spaces is perfectly possible. Following [17], a random field Y is associated with $\mathcal{G}$, via the labels in a phase space that we denote $\mathcal{L}$ that have been attached to each vertex. A realisation of the random field Y is denoted by y. The set $\mathcal{L}^{\mathcal{V}}$ of all possible label configurations is denoted Ω and called the state space. Note that the nodes of the graph in Figure 2c are labelled

to underline their correspondence with the edges of the original graph (Figure 2a). In the following, since we only consider the line graph and for the sake of clarity, nodes will be labelled y_i where $i \in \mathcal{V}$.

In Section 2.3.1, general notions on MRFs applied to social network analysis are given. The related simulation and inference procedure are given in Section 2.4.1. For a thorough and rigorous presentation of MRFs we recommend and the references within [19].

2.3.1 Markov Random Fields models and social network analysis

The MRFs were applied for social networks analysis by [15], [20], [21]. This class of models enables us to take into account dependencies between vertices assuming *local* interactions associated with the graph nodes.

In order to specify a MRF we need a neighbourhood relation. Here, two vertices i and j are neighbours, $i \sim j$, if there is a direct edge linking them. Following [17], the probability function of a MRF Y is described by a Gibbs distribution of the form:

$$p(Y = y|\theta) = \frac{\exp(U(y|\theta))}{\kappa(\theta)} = \frac{\exp(\langle\theta, t(y)\rangle)}{\kappa(\theta)}, \qquad (1)$$

where:

- $\theta \in \Theta \subseteq \mathbb{R}^d$ is the vector of d model parameters associated with the vector of sufficient statistics $t(y)$.
- $U(\cdot|\cdot)$ is the energy function
- $\kappa(\theta)$ the normalising constant.

The difficulty with this class of model is that $\kappa(\theta)$, the normalising constant is intractable. This requires special procedures for simulation and inference. Still, their advantage is that through local specifications they allow the modelling of complex systems.

2.3.2 A Potts-like model for characterising interactions on social networks

For the problem in hand, the aim is to characterise interactions between researchers. Let us consider the following MRF model:

$$\begin{aligned} p(Y = y|\theta) = \frac{1}{\kappa(\theta)} \exp\Big[&\theta_{11} \sum_{i\sim j} \mathbb{1}\{y_i = 1, y_j = 1\} \\ &+ \theta_{12} \sum_{i\sim j} \mathbb{1}\{y_i = 1, y_j = 2\} \\ &+ \theta_{22} \sum_{i\sim j} \mathbb{1}\{y_i = 2, y_j = 2\}\Big]. \end{aligned} \qquad (2)$$

where y is the realisation of the graph representation given by the labels $\{0, 1, 2\}$ (which means that $m = 3$) associated with each node. They correspond respectively to *nil*, *intra-organisational* and *inter-organisational* links. The sufficient statistics vector is given by

$$\begin{aligned} t(y) =& [\, t_{11}(y), t_{12}(y), t_{22}(y) \,] \\ =& \Big[\sum_{i\sim j} \mathbb{1}\{y_i = 1, y_j = 1\}, \\ & \sum_{i\sim j} \mathbb{1}\{y_i = 1, y_j = 2\}, \\ & \sum_{i\sim j} \mathbb{1}\{y_i = 2, y_j = 2\}\Big]. \end{aligned} \qquad (3)$$

The condition in $\mathbb{1}\{y_1 = 1, y_j = 1\}$ is verified whenever a researcher cooperates with two members of his team. It means that the statistic t_{11} indicates how the researchers interact within their own team. The condition $\mathbb{1}\{y_i = 1, y_j = 2\}$ is checked whenever a researcher cooperates with a member of his own team and a member of a different team. The statistic t_{12} indicates how the researchers exhibit a hub behaviour, since they interact with both kinds of teams, their own and different ones. Finally, $\mathbb{1}\{y_i = 2, y_j = 2\}$ is checked whenever a researcher cooperates with two members not belonging to his own team. Then, the statistic t_{22} indicates how the researchers interact with other teams. To sum up, the vector $\theta = [\theta_{11}, \theta_{12}, \theta_{22}]$ controls the "weight" of the previous statistics. If $\theta_{ij} > 0$ then the model tends to favour configurations with a high value for the statistic t_{ij}.

This model colours the line graph associated with a network in a similar manner as the Potts model does it. If important patches of $(1, 1)$ appear this means that there is an important tendency that the researchers on the network cooperate within their teams. Similar interpretation can be given, for the patches $(1, 2)$ and $(2, 2)$. The weight, the importance of these patches, hence of the general behaviour of the members of the network is given by the model parameters.

2.4 Simulation and inference procedures.

In this section, we review the state-of-art of inference of random graphs, briefly beginning with simulation procedures, since it is a key part of the inference process presented in Section 2.4.1.

The presence of $\kappa(\theta)$ in (2) imposes special strategies for the sampling of the model, Markov chains Monte Carlo methods. The best known sampling algorithms are the MH and the Gibbs sampler. In this paper, the data structure and the model construction made us opt for the Gibbs sampler [22], [23].

2.4.1 Inference procedures

Parameter estimation of MRFs (2) is not trivial due to the presence of an intractable normalising constant:

$$\kappa(\theta) = \sum_{y\in\Omega} \exp(\langle\theta, t(y)\rangle).$$

where $\cdot$ represents the scalar product between the parameters and sufficient vector, respectively.

The frequentist approach to dealing with the presence of an intractable likelihood normalising constant is to use Monte Carlo Maximum Likelihood estimation [24]–[26]. Let y_{obs} be an observed graph and let us consider θ_0 a given parameter value. The log-likelihood function can be written as:

$$l_{\theta_0}(\theta) = \langle(\theta - \theta_0), t(y_{obs})\rangle - log\left[\frac{\kappa(\theta)}{\kappa(\theta_0)}\right]. \qquad (4)$$

[25], [27] and [28] show that the ratio of the normalising constants is

$$\frac{\kappa(\theta)}{\kappa(\theta_0)} = \mathbb{E}_{\theta_0} \exp(\langle(\theta - \theta_0), t(Y)\rangle). \qquad (5)$$

In practice, estimation of (4) is achieved by gathering a sample of random graphs $y'_1 \dots y'_n \sim f(\cdot|\theta_0)$ via forward simulation [29], yielding the approximation :

$$\frac{\kappa(\theta)}{\kappa(\theta_0)} \approx \frac{1}{n}\sum_{i=0}^{n-1} \exp(\langle(\theta-\theta_0), t(y_i)\rangle) \tag{6}$$

where the $\{y_i\}_{0\leq i<n}$ are realisations of $\{Y_i\}_{0\leq i<n}$ i.i.d. sampled from $p(y|\theta_0)$.

The simulated chain exhibits convergence properties (irreducibility, recurrence, ergodicity) [30]–[33] In fact (6) is plugged into (4) and the Monte Carlo likelihood is obtained

$$\begin{aligned} l_{n,\theta_0}(\theta) =& \langle(\theta-\theta_0), t(y_{obs})\rangle \\ & - log\left[\frac{1}{n}\sum_{i=0}^{n-1}\exp(\langle(\theta-\theta_0), t(y_i)\rangle)\right]. \end{aligned} \tag{7}$$

For the exponential family models, the log-likelihood is concave [25], [27]. This motivates to compute the gradient and the Hessian of (7). The approximated gradient and Hessian can be easily computed via importance sampling. These quantities are consistent estimators of their exact counterparts, respectively, that are computed from the original log-likelihood. Finally, using these quantities a Monte Carlo Newton Raphson (MCNR) local optimisation method can be implemented.

This method exhibits convergence results and two asymptotics explaining the estimation error can be computed. The first error is the Monte Carlo Standard Error may be interpreted as the difference between the true model parameters and the Maximum Likelihood Estimate, that are both unknown. The second error is the Monte Carlo Maximum Likelihood Error that approximates the difference between the Maximum Likelihood Estimate (which is unknown) and the Monte Carlo Maximum Likelihood Estimate, the result given by the MCNR method.

The drawback of the MCNR method is that it requires θ_0 to be close to the final estimate. This is due to the fact that the computation of the importance sampling weights needed in the evaluation of the gradient and the Hessian are not stable from a numerical point of view. Several strategies are available. Among them, the most robust is to resample the model $p(y|\theta)$ whenever the difference between the current value of the parameters and θ_0 exceeds a given threshold. Due to the concavity of the log-likelihood function, this strategy leads towards a convergent method but with a high computational cost. This question is still an open problem [24], [25], [34].

Markov models can suffer from degeneracy problems. This is a complex issue, especially for multidimensional models. [26] gave a strategy to plot the support of the field of sufficient statistics. The calculation of the asymptotic presented in Sections 4 and 5 are indicators of the convergence of the inference algorithm.

3 Posterior-based inference

According to Bayes's theorem, with $p(\theta)$ the prior knowledge on parameter distribution, the posterior distribution $p(\theta|y)$ is:

$$p(\theta|y) \propto p(y|\theta)\cdot p(\theta) = \frac{\exp\langle t(y),\theta\rangle}{\kappa(\theta)} p(\theta). \tag{8}$$

The difference between maximum likelihood that we described in Section 2.4.1 and posterior inference is the following. In the first case, under the assumption of a parametric model and with no prior knowledge regarding these parameters, the most probable model is proposed as an explanation of the observed data. The posterior-based inference also assumes a parametric model and it uses prior knowledge with respect to these parameters. But any model belonging to the family may explain the data. The quality of this explanation is given by the posterior distribution that weights each model within the considered parametric family. Posterior-based inference is much more informative. It can also be seen as a generalisation of the maximum likelihood approach. Whenever $p(\theta)$ is the uniform distribution of the parameter space Θ, both the Bayesian and the frequentist approach are strictly equivalent.

Despite the interest in performing posterior-based inference, this is not done often, since sampling the posterior or the likelihood is far from being a trivial task. A straightforward application of Monte Carlo sampling strategies such as MH or Gibbs dynamics requires the computation of the normalising constants ratio (5).

The authors in [35] give a very elegant solution to this problem. They propose a MH dynamics based on auxiliary variables. The use of the auxiliary variables requires appropriate proposal distributions. The proposal distributions can be tailored to cancel the computation of the normalising constants within the acceptance ratio of the MH algorithm. The authors indicate themselves that their rigorous mathematical solution cannot prevent the resulting chain from poor mixing.

The work presented in [36] propose the Exchange Algorithm, which implements a strategy similar to the auxiliary variable solution of [35]. Essentially, this algorithm uses a swap mechanism among the variables of the distribution of interest. The resulting algorithm is also a Metropolis-Hastings dynamics which samples from the augmented distribution: $p(\theta', x, \theta|y)$. The exchange algorithm convergence requires exact simulation for the auxiliary variable and it depends of the choice of the proposal distribution. An implementation of this algorithm dedicated to the ERGMs is proposed by [37].

Approximate Bayesian Computation (ABC) algorithms are methods used to approximately sample from the posterior distributions of the models that are intractable. They are easy to implement, but the control of the algorithm is done on a rather heuristic base. Among, the pointed drawbacks regarding these methods, one indicates the choices of the distance between the observed statistics and the outputs of the algorithm statistics, the selection of these statistics and the setup of the control error threshold [38]–[41]. Still in many situations these choices may be considered rather natural, especially whenever sufficient statistics are available.

The ABC Shadow method proposed by [6] is directly inspired by the previous ideas, auxiliary variable based simulation and approximate sampling, while trying to solve some of their mentioned drawbacks. The ABC Shadow is an approximate sampling method for posterior distribution, exhibiting better numerical properties than the auxiliary variable method and offering a more robust control than the ABC classical framework. Recent work [42] builds a

simulated convergent annealing process based on a ABC Shadow dynamics.

The ABC Shadow algorithm is presented in Algorithm 1. For all the technical details and mathematical proofs the reader has to refer to [6]. The method is general in the sense that it can be applied to sample posterior distributions, assuming only their continuous differentiability with respect to the model parameters. The algorithm needs for initialisation the observed graph y_{obs}, the initial value θ_0 of θ, Δ an error control parameter and m the number of steps the algorithm runs. The Δ parameter supports the proposal distribution whose form is given line 4 of Algorithm 1. All theoretical details about the construction of the proposal can be found in [6]. First the algorithm samples an auxiliary graph x according to the chosen model. Then for each step in the loop it proposes a new parameter value θ' that is accepted with the probability α (see line 7 of Algorithm 1). If this new state is not accepted, the algorithm remains in its previous state. The distribution of the output of the algorithm follows approximately $p(\theta|y_{obs})$ with an error limits controlled by m and Δ. The value of Δ has to be tuned in a fine way, since there is an acceptable compromise to reach between quality of approximation and good mixing properties of the chain. If the number of steps m is too large, the algorithm goes away from the posterior of interest whereas if m is too small the mixing property is negatively impacted. Hence, a reasonable value for these two parameters is needed. In [6] is proved that for any fixed value m there exists a positive value Δ so that the outputs of the ABC Shadow algorithm are distributed as close as desired from the posterior distribution of interest. If more than one sample from the posterior is needed, this can be obtained by iterating the ABC Shadow algorithm (Algorithm 1).

Algorithm 1 ABC Shadow algorithm

1: **function** ABC_SHADOW(θ_0, y_{obs}, m, Δ)
▷ Where θ_0 - initial parameters, y_{obs} - observation
2: $x \sim p(x|\theta_0)$ ▷ Choose an auxiliary variable
3: **for** i = 1 to m **do**
4: $\theta' \sim \mathcal{U}_\Delta(\theta_{i-1} \to \theta')$
5: $\alpha \leftarrow min\{1, \exp[(t(y_{obs})-t(x))(\theta'-\theta_{i-1})]\frac{p(\theta')}{p(\theta_{i-1})}\}$
6: $accepted \leftarrow \mathcal{U}(0,1)$
7: **if** $\alpha > accepted$ **then**
8: $\theta_i \leftarrow \theta'$
9: **else**
10: $\theta_i \leftarrow \theta_{i-1}$
11: **end if**
12: **end for**
13: **return** θ_m
14: **end function**

ABC Shadow is an alternative strategy between classical ABC methods and the Metropolis-Hastings sampling algorithms based on auxiliary variables. Its practical implementation is similar to the one proposed by [37], while providing the needed theoretical and practical control of the output and preventing from poor mixing. Note that ABC Shadow explicitly allows the control of the approximation by instrumenting the proposal distribution with the parameter Δ. The control over the approximation provided by ABC Shadow is shown on known models in the following section.

4 ABC SHADOW IN PRACTICE : ILLUSTRATION ON SYNTHETIC DATA

The use of the ABC Shadow algorithm requires the set-up of its parameters. In order to chose m and Δ the ABC Shadow algorithm was run on known models, with controllable expected results. Whenever it was possible, the outputs of the ABC Shadow algorithm were compared with a classical Monte Carlo sampler of the posterior, the MH algorithm and the Exchange algorithm [36].

4.1 Binomial distribution

Let y be generated by a Binomial distribution of parameters n and p. This may correspond to the independent random labelling, following a Bernoulli distribution with the parameter p, of a bi-coloured graph of size of n. We know the parameter n and we want to estimate p. Within this context the likelihood reads :

$$\begin{aligned} p(y|\theta) &= \binom{n}{y} p^y (1-p)^{(N-y)} \\ &= \exp\left[y\theta - n\log(1+e^\theta) + log\binom{n}{y} \right] \end{aligned} \tag{9}$$

with $\theta = log(p/(1-p))$. For our experiment $n = 20$, $p = 0.4$ ($\theta = -0.405$) and $m = 100$. The observed Binomial variable obtained with these values was $y = 8$. The MH algorithm is set up to sample from the distribution (9). Within the context of posterior sampling of a binomial distribution, the main difference between the implementation of the ABC Shadow and the Exchange algorithm is given by the choice of the proposal distribution. For the ABC Shadow, the chosen proposal is uniform over the interval $[-100, 100]$ of width $\Delta = 0.005$ centred on the current value. While for the Exchange Algorithm the adopted proposal is a normal distribution with mean corresponding to the current value and $\sigma = \frac{\Delta}{2}$. The procedures were executed to sample 1.002×10^6 posteriors. The first 2×10^3 samples were cut off and a sample were kept every 100 iterations. This resulted in a chain $(\theta^{(t)})_{t=1,\cdots,T}$ of 10^4 samples.

For the ABC Shadow, the proposal distribution is the same as the one of the MH algorithm. The auxiliary variable is simulated from 100 samples following (9). The procedure described in Algorithm 1 is implemented and applied to our simulated data with $m = 100$ and repeated $iters = 1.002 \times 10^6$ times. Like the MH, the output is a chain $(\theta^{(t)})_{t=1,\cdots,T}$ that we subsample, keeping only every 100. It improves the mixing properties of the chain. In addition, we skipped the first 2×10^3 samples of the chain $(\theta^{(t)})_{t=1,\cdots,T}$. To illustrate the robustness of these two algorithms, the initial value of the chain of samples $\theta^{(0)}$ is chosen far from the true value of θ. We set $\theta^{(0)} = 1$.

Figure 3 represents the three distributions respectively obtained with the standard implementation of the MH algorithm, the ABC Shadow and the Exchange Algorithm. According to the box plot and the quantile-quantile plot schemas, the three distributions are very close to each

other. It is worth noticing that the three algorithms (especially ABC) converge toward the true parameter value $\theta = -0.405$, although the initial value of the chain ($\theta^{(0)} = 1$) is quite far from the truth. Statistics, Maximum A Posteriori (MAP) and errors of the three distributions are summarised in Table 1. The three methods indicate equivalent performances for the provided inference. These results encourage us, in the following, to give preference for the ABC Shadow algorithm, since exact simulation is not required, for its control.

4.2 Posterior sampling on the Potts Model

We now consider the Potts model involved in the description of our application context. Due to the normalising constant, the Potts model (described in Section 2.3.2) is not directly tractable with the traditional MH algorithm as previously performed in Section 4. To circumvent this problem and following the strategy in [6], we tested the Potts model by comparing the maximum of the approximated a posteriori distribution with the true parameter of the model previously simulated.

In the first experiment, all interaction parameters were fixed to $0 : \theta_{11} = \theta_{12} = \theta_{22} = 0$, so that interaction effects are annihilated. Since we have three type of patterns, this leads to a Bernoulli graph model with an occurrence probability for each pattern equal to $\frac{1}{3}$. The observation represents an artificial collaboration involving 12 members of the same organisation and 8 collaborators from the outside i.e. with a $size = (12, 8)$. It was generated from $N = 10^3$ samples yielded by a Gibbs sampler. By averaging sufficient statistics we obtain $\bar{t}(y) = [164.747, 263.495, 83.7645]$ (see Equation 3) from the ABC Algorithm. In the ABC algorithm, the prior distribution $p(\theta)$ was a uniform distribution on the interval $[-4, 4] \times [-4, 4] \times [-4, 4]$. The parameters n and Δ were respectively set to $n = 200$ and $\Delta = [0.01, 0.01, 0.01]$ according to [6]. As in Section 4.1, the ABC Shadow was executed to yield $iters = 1.002 \times 10^6$ samples. We subsampled keeping every 100 value and rejected the 2×10^3 first burn in samples. At each iteration the auxiliary variable x was updated using 200 steps of a Gibbs sampler. Error metrics were computed: the asymptotic standard deviation $\hat{\sigma}_\theta$ and the Monte Carlo standard deviation $\hat{\sigma}_\theta^{MC}$. The numerical values of both metrics are depicted in Table 2.

Figure 4a represents the histograms of the posterior distributions provided by the ABC Shadow of each parameter as well as two-dimensional posterior distributions for each couple of parameters. Blue lines mark the MAP for each parameter's distribution computed by taking the maximum of the kernel density estimation: $\hat{\theta} = [-0.0262, 0.036, -0.0436]$. The green lines are the true parameter values : $\theta = [0, 0, 0]$.

We now consider a model with repulsion effects. To that end, we set $\theta_{11} = -0.5$, $\theta_{12} = 0.2$, $\theta_{22} = 0.3$ and we simulate $N = 10^3$ samples with a $size = (12, 8)$ using a Gibbs sampler as we did before. The vector Δ has been appropriately selected following the strategy given by [6]. It was set to $\Delta = [0.005, 0.005, 0.005]$. The generated observation yielded the following averaged sufficient statistics: $\bar{t}(y) = [78.8842, 360.732, 295.548]$. Figure 4b represents the resulting posterior distribution. Blue lines representing the MAP are aligned on $\hat{\theta} = [-0.5374, 0.2086, 0.2916]$ which is close to the true parameter $\theta = [-0.5, 0.2, 0.3]$ represented with green lines. The dashed lines are respectively the first quartile, the median and the third quartile. The mean and the median of the posterior estimates are respectively: $[-0.702, 0.293, 0.258]$ and $[-0.658, 0.274, 0.265]$. The error metrics, respectively the asymptotic standard deviation and the Monte Carlo standard deviation are presented in Table 2. The traces of the sampled distributions are given in Appendix.

We showed on simulated examples how to control and set up the ABC Shadow on models similar to the one used for inference from real data.

5 Application

The research works illustrated by scientific productions (such as conference papers and articles) rely on collaborative and social links. We want to identify the structural patterns formed by those links and capture their dynamic over time. The collaboration networks of three different years (2017, 2018, 2019) are obtained using the HAL publication database. In those graphs, a node represents a researcher. Two researchers are connected if they have at least a common publication during the considered year. We collected metadata of publications deposited by the members of LORIA in the period 2017-2019. The dataset is available at [43].

The aim of the study is to fit the model defined in Section 2.3.2 to the graph associated with each team of LORIA. Comparing the structural aspects of those graphs through posterior analysis enables the identification of patterns characteristic of these scientific collaborations.

For each team, the graph is constructed as follows. Figure 5 exhibits the different steps of the processing. First, two kinds of nodes were distinguished, the members of LORIA and the other researchers who had no affiliation with LORIA, the external stakeholders (Figure 5a). We took the point of view of each team and studied the way they collaborate with internal and external stakeholders. This means that only edges linking at least one member of LORIA are considered (Figure 5b). In addition, we only took into account interactions between edges linked by a member of LORIA (Figure 5c). Following the framework of Section 2.2, the line graph representation encodes the different types of research collaboration. An inter-organisational link connects one member of LORIA with an external collaborator whereas intra-organisational links connect two collaborators who are affiliated to LORIA. Under the hypothesis of the model, the sufficient statistics were computed. The results are presented in Table 3 (in Appendix).

The number of authors from LORIA as well as external stakeholders are different according to each team and for each year. Statistics of both quantities are given in Table 4. Compared to the means (μ), the standard deviations are important : approximately $0.5 \times \mu$ for LORIA's members and strictly greater than $0.5 \times \mu$ for external collaborators. This indicates that the sizes of the different collaborations are distributed on an important range.

We identified 11 teams with a sufficiently large number of publications in the three considered years. The ABC

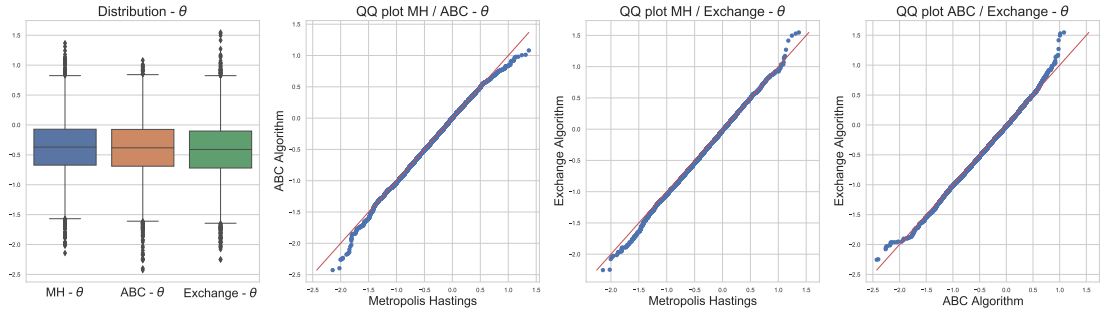


Fig. 3. Posterior sampling of a Bernoulli distribution using Metropolis Hasting, ABC Shadow and the Exchange algorithm

TABLE 1
Statistics on the posterior of Binomial distribution

	Q_{10}	Q_{25}	Q_{50}	$mean$	Q_{75}	Q_{95}	MAP	$\hat{\sigma}_\theta$	$\hat{\sigma}_\theta^{MC}$
ABC (θ)	-0.992	-0.69	-0.383	-0.392	-0.075	0.345	-0.4257	0.453	4.2×10^{-4}
MH (θ)	-0.961	-0.672	-0.371	-0.377	-0.071	0.353	-0.3775	0.452	4.3×10^{-4}
Exchange (θ)	-1.009	-0.721	-0.41	-0.415	-0.102	-0.344	-0.445	0.455	4.3×10^{-4}

TABLE 2
Statistics on the posteriors of the Potts model under two configurations : $\theta_{11} = \theta_{12} = \theta_{22} = 0$ and $\theta_{11} = -0.5$, $\theta_{12} = 0.2$, $\theta_{22} = 0.3$.

		$\hat{\theta}_{11}$	$\hat{\theta}_{12}$	$\hat{\theta}_{22}$	$\hat{\sigma}_{\theta_{11}}$	$\hat{\sigma}_{\theta_{12}}$	$\hat{\sigma}_{\theta_{22}}$	$\hat{\sigma}_{\theta_{11}}^{MC}$	$\hat{\sigma}_{\theta_{12}}^{MC}$	$\hat{\sigma}_{\theta_{22}}^{MC}$
$\theta = [0, 0, 0]$	ABC (θ)	-0.0262	0.036	-0.0436	0.08	0.093	0.144	3.80×10^{-7}	5.80×10^{-7}	1.98×10^{-6}
$\theta = [-0.5, 0.2, 0.3]$	ABC (θ)	-0.5374	0.2086	0.2916	0.364	0.169	0.116	7.106×10^{-5}	1.433×10^{-5}	5.370×10^{-6}

Shadow algorithm was launched with the same initial conditions for every team and every year. The ABC Shadow algorithm was setup to generate $iters = 10^7$ samples, the number of iterations of the shadow chain and the volume bound were set respectively to $m = 200$ and $\Delta = [0.005, 0.005, 0.005]$. The auxiliary variable x was sampled with 50 iterations of the Gibbs sampler. The first 9×10^6 burn in samples were discarded. In addition, a subsampling procedure kept every 10^3 value of each remaining chain yielded by the ABC Shadow. Consequently, for each team the size of the corresponding chain was 10^3 samples.

Figures 6, 7 and 8 show the box plots of posteriors sampled by the ABC Shadow respectively for the parameters θ_{11}, θ_{12} and θ_{22} for each team and each year. The box plots are grouped by teams and differently coloured according to the year so as to observe the evolution of parameter values over the years for a given team. In complement to these three Figures, Table 5 in Appendix present the mean, the median and the estimated MAP of the posterior distribution of θ for each team and each year.

A positive parameter indicates a positive inclination to observe the corresponding pattern compared to a pure random process. Conversely, a negative parameter means that the corresponding pattern has a weaker probability to occur than a pure random process. The parameter values are good indicators on the way researchers interact with their counterparts.

The value ranges of parameters are near to zero, even lower than zero for the majority of teams. Relatively to all possible connections, this reflects a weak global tendency for a researcher to co-author with all other researchers whether he belongs to the same lab or not. At the scale of teams this means that the collaboration graph is sparse. Putting this observation in the context of publication activities, this corroborates the intuition that every researcher does not co-author with everyone else. Co-authoring a paper implies that all stakeholders are involved in the same scientific work. These are demanding tasks. It restricts the number of publications and the underlying potential collaborations a researcher is able to undertake.

At first sight, there is no clear evidence that teams share the same evolution pattern regarding the posterior distributions of the parameters. For some teams and some parameters, the box plots tend to decrease over the years, this suggests a diminishing occurrence probability of the corresponding controlled pattern. We can also observe the opposite phenomenon which is illustrated by ascending box plots. However, the evolution of one parameter over time must be studied in combination with the evolution of the other parameters. Let us take the CARAMBA team as a first example. The range of values of the θ_{11} parameter drops drastically in 2019. Such evolution is not observed for parameters θ_{12} and θ_{22}. Moreover, the sufficient statistics (Table 3) decrease over time, until 0 (the $1 \leftrightarrow 1$ pattern is no longer observed in 2019). This allows us to conclude that there is a real decrease in the tendency to collaborate within the CARAMBA team. But the decreasing box plot over time does not necessarily mean that the corresponding tendency is falling down. The team CAPSID is a good example. As the team CARAMBA, the parameter θ_{11} drop drastically in the last year. But conversely, the parameter θ_{12} increases in positive values. More than a decreasing tendency to observe the

pattern $1 \leftrightarrow 1$, this suggests a change in the collaboration dynamic. The balance between intra-organisational links and inter-organisational links shifts, illustrating a much more outward-looking collaboration behaviour.

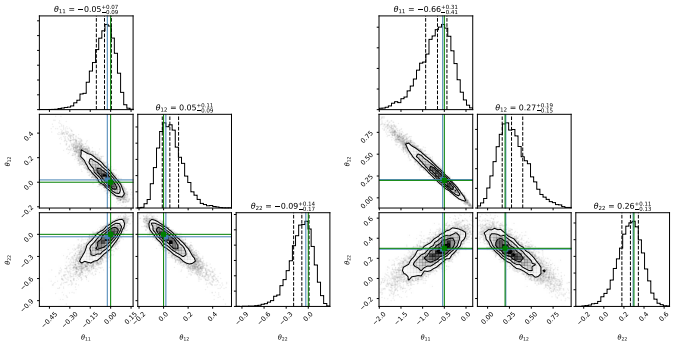
Regarding Table 5 in Appendix, both the median and the mean are close to the estimated MAP. Similarly to [6], [44], we computed the asymptotic standard deviation and the Monte Carlo standard deviation. To that end, for each estimated model we performed a simulation providing 10,000 samples (samples were uncorrelated by keeping every 10^3 sample out of the 10^7 simulated). Given the Monte Carlo standard deviation, we can determinate the 95% confidence interval. Error measures and 95% confidence intervals are reported in Table 6 (in Appendix).

The results in Tables 5 and 6 show the good performances of the parameter estimation procedure. Nevertheless these results should be interpreted with care. Model degeneracy is an important problem that should be taken into account whenever using MRF [26]. In Appendix B these results are further investigated through simulations using the estimated parameters. In particular, pathological situation with teams exhibiting no collaboration inside the team or outside the team are considered.

The closeness of value ranges to 0, especially for the posteriors of the parameter θ_{11}, raises the question of their significance. Are the three studied patterns more likely to occur in the collaboration than pure randomness? To answer this question we applied for each parameter a t-test to determine if the expectations of the posteriors equal 0. The null hypothesis and the alternative hypothesis are written as follows for each parameter:

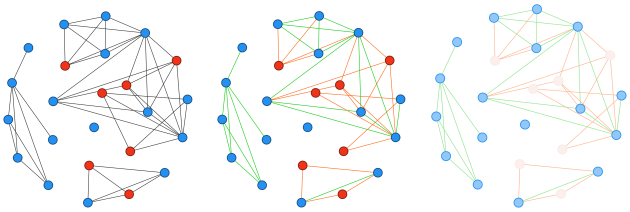
$$\mathcal{H}_0 : \mathbb{E}[\theta] = 0,$$
$$\mathcal{H}_1 : \mathbb{E}[\theta] \neq 0.$$

The results are shown in Table 7 (in Appendix). For most of the teams, the parameters are significantly non-zero since the associated p-values of the t-test are very small. There are only four observed networks for which the p-value is greater than the usual 5% level of significance. In this case, the rejection of the null hypothesis is not relevant. In conclusion, for almost all networks (except the two latter mentioned), the likelihood of link creation is not merely due to chance.



(a) (b)

Fig. 4. Corner plots of marginal distributions of posterior sampling for the Potts model using an ABC algorithm. *(Blue lines mark the MAP of each parameter, green lines correspond to the true parameter values)*



(a) *step 1:* Distinguish two kinds of nodes, members of LORIA (in blue) and external stakeholders (in red).

(b) *step 2:* Keep edges involving at least one member of LORIA.

(c) *step 3:* Consider only interactions between edges linked by a member of LORIA.

Fig. 5. An example of a pre-processing performed on a team's collaborative graph. The graph in Figure 5a illustrated co-authoring links involving the members of the team COAST in 2018. Blue nodes ● represents the members of LORIA, whereas red nodes ● are external collaborators. The inter-organisational links are coloured in orange ●, while the intra-organisational links are in green ●.

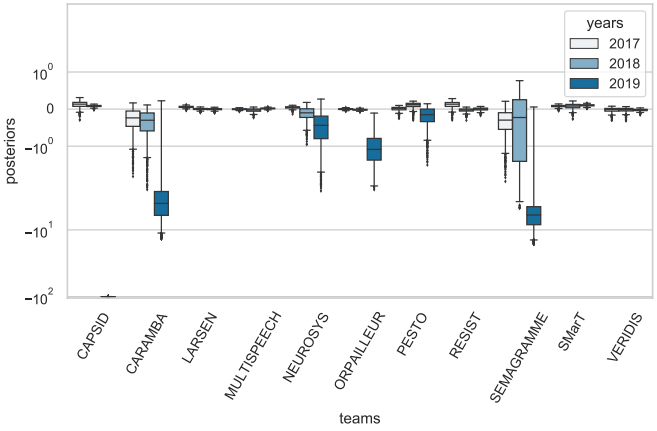


Fig. 6. Box plots of the posterior distributions for the parameter θ_{11} estimated from the collaboration graphs

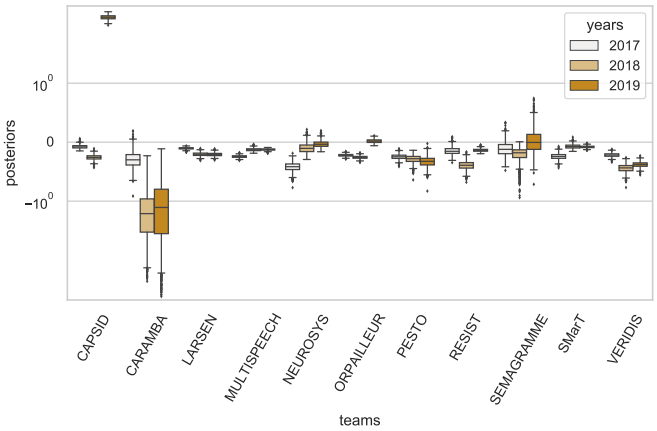


Fig. 7. Box plots of the posterior distributions for the parameter θ_{12} estimated from the collaboration graphs

Figure 9 presents the three 2d projections of the Potts model parameters. Each team is associated with a colour and each year with a different marker. Grey dashed lines set

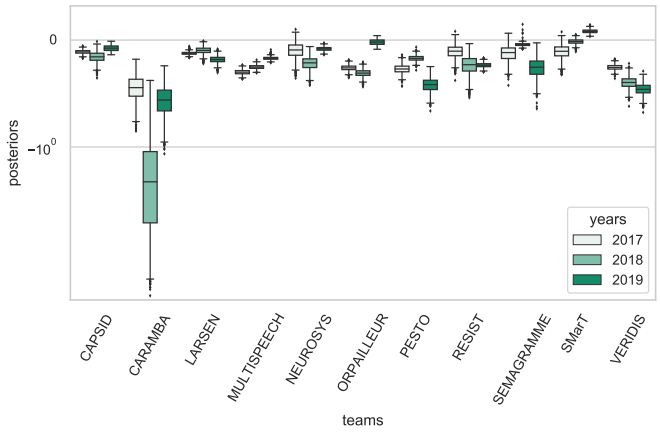


Fig. 8. Box plots of the posterior distributions for the parameter θ_{22} estimated from the collaboration graphs

limits between positive and negative trends. For instance, considering (θ_{11}, θ_{22}), the vertical line delimits the positive and negative tendencies that a pattern linking two intra-organisational ties occurs, while the horizontal line is about the occurrence of inter-organisational links. Depending on projections, we have an overview of trends followed by teams.

The major part of the estimated MAPs is concentrated in the same region. For the parameter θ_{11} the MAPs are distributed closely around 0. For θ_{12} and θ_{22}, they are mostly negative. This observation refines our analysis. The latter two, show that hub patterns and collaboration links with the outside are less likely to occur in collaboration graphs. This strengthens the prior idea that collaborations with external teams are complex to set up and maintain. The weak presence of hubs in the collaboration means that only few researchers are connected at the same time with members of their team and researchers from other labs. If a hub leaves, the ties between the corresponding organisations break [45]. This is a serious concern that should be carefully addressed in the design of collaborative applications to ensure the availability of the collaboration against the churn.

Some outliers presenting different structural features are identifiable. For instance, the team SMaRT in 2019, located at the top right-hand corner of the first frame (Figure 9), shows a slight positive tendency for the pattern $2 \leftrightarrow 2$ to occur. The team CAPSID in 2019 is also an outlier with respect to other parameters values: θ_{11} and θ_{12}, as we can see at the top left-hand corner of the third frame (Figure 9). The MAP of the parameter θ_{11} is extremely low while the MAP of θ_{12} is high. This parameters configuration suggests that hub ($1 \leftrightarrow 2$) patterns are likely to occur while the emergence probability of ($1 \leftrightarrow 1$) patterns is weak. In other words, the collaborations are outward looking for that team. Looking at the network statistics, the number of external researchers is very high compared to the number of referenced authors from LORIA (Table 3 in Appendix). This fosters the emergence of inter-organisational links to the detriment of intra-organisational links. Figure 10 illustrates the co-authorship graph of the team CAPSID. The two noticeable clusters are related to two articles involving a large number of authors. The few authors of theses articles affiliated to LORIA and connected to a few other LORIA's members generated themselves a very large number of hub patterns. This is a special configuration met only in the co-authorship graph of the team ORPAILLEUR in 2019 which closely worked with CAPSID as they shared numerous publications.

Figure 9 shows some overlapping points or points very close to each other. This suggests that some teams share with each other similar structural characteristics. By relying not only on the MAPs but on the whole posterior distributions, we aim to verify these observations.

An unsupervised hierarchical classification was performed, from the Kolmogorov-Smirnov distance computed between all posterior distributions of the three parameters: θ_{11}, θ_{12} and θ_{22}. The results in Figure 11 are shown in the form of dendrograms. The branches' height of the dendrogram gives indications about the proximity of the sub-clusters : shallower is a branch, closer are the sub-clusters and vice-versa. The few identified clusters correspond to the coloured branches.

The lab is organised in 5 departments gathering teams working on the same research thematic. The team's names are coloured according to the affiliation to one of these departments. The clustered team's names are not similarly coloured and then, don't necessarily work on the same topic. Consequently, structural patterns are not a feature specific to the research thematic. We also noticed that the closeness between two teams can be related to the fact that one originates from the other. It is not unusual that a researcher keep signing with an old affiliation a long time after the creation of a new team. Also, when a team splits in new teams, members of teams keep collaborating. This means that both teams keep intrinsic collaboration links affecting their collaboration networks. This requires to pay particular attention to the real-life context in particular to the teams' life cycle : birth, split, death.

6 Conclusion

In this paper, we proposed a method to make inference on structural aspects of collaboration networks. The work we present is embedded in the context of inter-organisational collaborations, a topic yet sparsely addressed by the state of the art. For instance, researchers from different organisation often collaborate to conduct research and write publications. We extracted the collaboration network among researchers by considering the co-authorship of publications from the French open-archive HAL as collaboration links between authors.

First we presented the representation of the collaboration graph. We relied on the line graph instead of taking the collaboration network directly as the observation of our study. Considering this alternate representation as fixed random field, we considered link creation in the collaboration as a labelling issue respecting Markov's properties. We were able to better encompass structural interactions not between individuals but among relations themselves. We used a generalisation of the Ising model, the Potts model, to describe the interactions between relations. As for all exponential models, the inference remains difficult due to

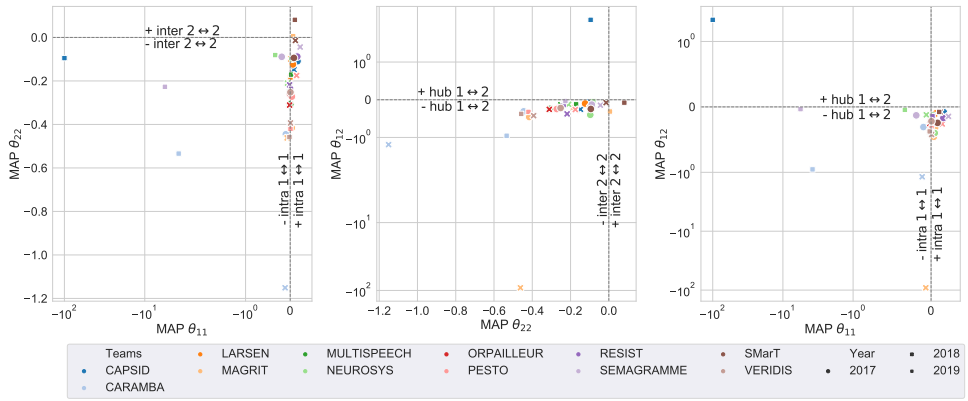


Fig. 9. Scatter plots of estimated MAPs representing the positioning of teams with respect to the different collaborative tendencies controlled by $\theta_{01}, \theta_{02}, \theta_{12}$

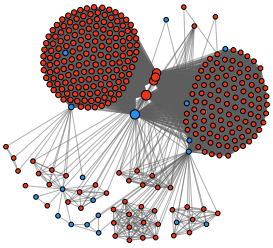


Fig. 10. Co-authorship graph of the team CAPSID. Blue nodes ● represents the members of LORIA, whereas red nodes ● are external collaborators. Size of nodes is proportional to their degree. The two red clusters represent two publications involving a very large number of authors.

the intractable normalising constant. To that end, we used a Bayesian tool, the ABC Shadow algorithm which was firstly tuned on tractable model and simulated data. We applied it on collaboration networks of different research teams for three different years. The main aim was to characterise and classify collaborations among researchers in their publication activities and their evolution over time. First of all, we observed that links formation between collaborators are not mere coincidence but the result of tendencies for almost all teams. From the posterior distributions provided by the ABC Shadow, we showed that a few actors play a key role since they connect collaborators of their organisation toward the outside. The combined evolution of the model's parameter allows us to assess the dynamics of the collaborations over time. Given the posteriors, we also demonstrated how to classify the way the different teams collaborate and conclude that structural features at stake are not related to the scientific topic addressed.

The sizes of the teams are relatively disparate and may affect the prevalence of the observed patterns. One of the first perspectives to pursue is to come up with a normalising procedure to put all the observed graph on the same level.

Hub in the collaboration are points of failure who can endanger the inter-organisational collaboration if they leave. This is a concern that must be addressed in the design of collaborative applications, to better support inter-organisational scenarios. Study of the dynamic of the network [46] enables the assessment of the churn and the detection of breaks over time.

The approach we propose to observe the dynamics of collaborative networks is based on a fixed time step and relies on networks involving different actors over time. This raises several issues. The observations are difficult to compare with each other, since they do not have the same size and do not necessarily involve the same individuals. Moreover, the time step used is important and does not allow the gradual evolution of interactions to be captured. The development of a longitudinal approach is part to our future work. It requires both adapted observations and a modelling integrating the time dependency [47], [48].

The selection of the model is a very sensitive aspect of our approach which might influence the relevance of the estimates in regard to the observation. This concern should be further investigated in a future work [49]. For instance, other classes of models such as the Markov Connected Component Fields [50] might be good candidates for structural graph pattern analysis.

Finally, the approach we proposed here can be applied in different contexts. Extending this study to other collaborative contexts is required to acquire a comprehensive understanding of features inherent to inter-organisational collaborations.

References

[1] A.-L. Barabási and M. Pósfai, *Network Science*. Cambridge University Press, 2016.

[2] B. Bollobás, *Modern graph theory*. Springer Science & Business Media, 2013, vol. 184.

[3] D. De Stefano, V. Fuccella, M. P. Vitale, and S. Zaccarin, "The use of different data sources in the analysis of co-authorship networks and scientific performance," *Social Networks*, vol. 35, no. 3, pp. 370–381, 2013.

[4] A. Ferligoj, L. Kronegger, F. Mali, T. A. B. Snijders, and P. Doreian, "Scientific collaboration dynamics in a national scientific system," *Scientometrics*, vol. 104, no. 3, pp. 985–1012, 2015.

[5] P. Ji and J. Jin, "Coauthorship and citation networks for statisticians," *Annals of Applied Statistics*, vol. 10, no. 4, pp. 1779–1812, 2016.

[6] R. S. Stoica, A. Philippe, P. Gregori, and J. Mateu, "Abc shadow algorithm: a tool for statistical analysis of spatial patterns," *Statistics and computing*, vol. 27, no. 5, pp. 1225–1238, 2017.

[7] J. Scott, "Social network analysis," *Sociology*, vol. 22, no. 1, pp. 109–127, 1988. [Online]. Available: https://doi.org/10.1177/0038038588022001007

[8] G. Robins, P. Pattison, Y. Kalish, and D. Lusher, "An introduction to exponential random graph (p*) models for social networks," *Social networks*, vol. 29, no. 2, pp. 173–191, 2007.

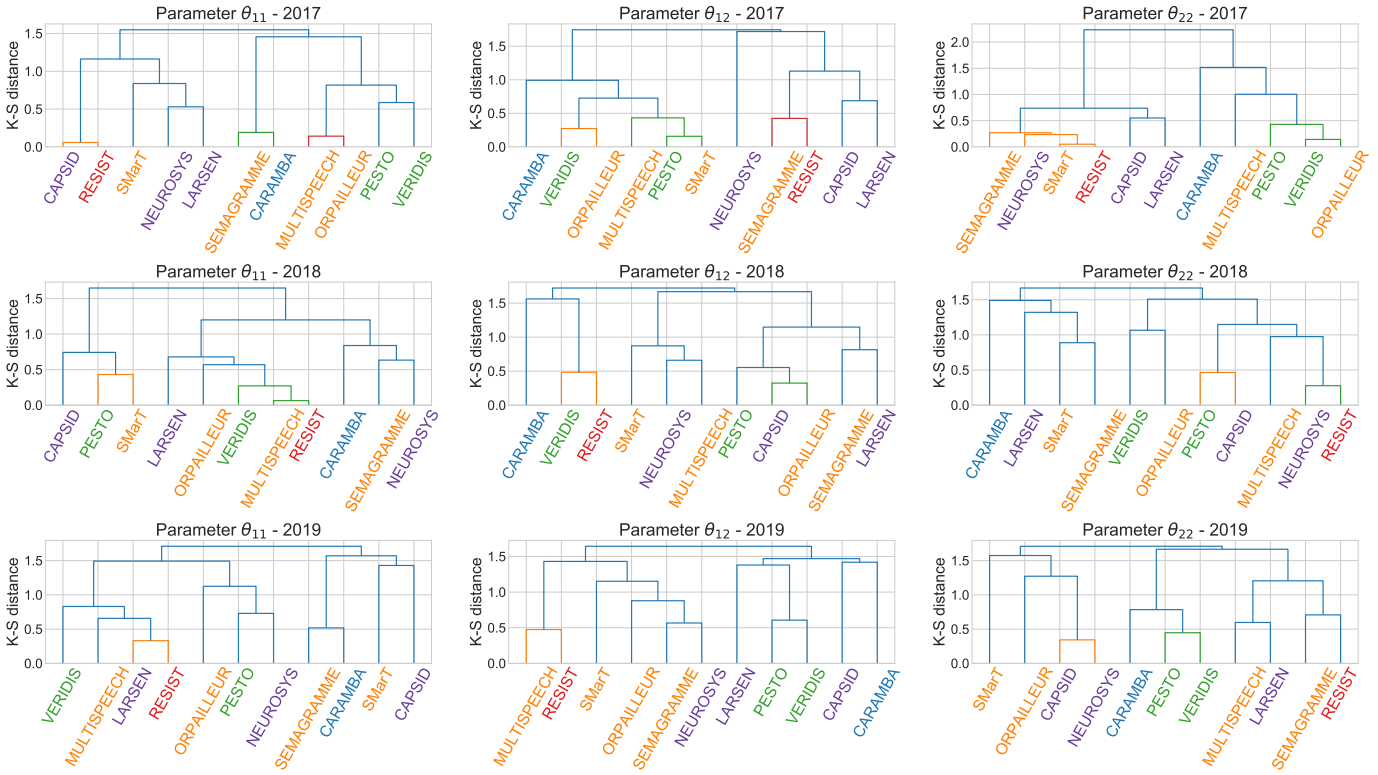


Fig. 11. Hierarchical classification throughout the Kolmogorov-Smirnov distance of posteriors. The label are coloured according to the research thematic addressed by each team : Algorithms, Computation, Image & Geometry, Formal methods, Networks, Systems and Services, Natural Language Processing & Knowledge Discovery, Complex Systems, Artificial Intelligence and Robotics.

[9] P. Wang, G. Robins, P. Pattison, and E. Lazega, "Exponential random graph models for multilevel networks," *Social Networks*, vol. 35, no. 1, pp. 96–115, 2016. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0378873313000051

[10] ——, "Social selection models for multilevel networks," *Social Networks*, vol. 44, pp. 346–362, 2016. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0378873314000781

[11] P. Zappa and A. Lomi, "Knowledge sharing in organizations: A multilevel network analysis," in *Multilevel Network Analysis for the Social Sciences*. Springer, 2016, pp. 333–353.

[12] F. Battiston, E. Amico, A. Barrat, G. Bianconi, G. Ferraz de Arruda, B. Franceschiello, I. Iacopini, S. Kéfi, V. Latora, Y. Moreno *et al.*, "The physics of higher-order interactions in complex systems," *Nature Physics*, vol. 17, no. 10, pp. 1093–1098, 2021.

[13] G. F. de Arruda, G. Petri, and Y. Moreno, "Social contagion models on hypergraphs," *Physical Review Research*, vol. 2, no. 2, p. 023032, 2020.

[14] O. T. Courtney and G. Bianconi, "Generalized network structures: The configuration model and the canonical ensemble of simplicial complexes," *Physical Review E*, vol. 93, no. 6, p. 062311, 2016.

[15] O. Frank and D. Strauss, "Markov graphs," *Journal of the american Statistical association*, vol. 81, no. 395, pp. 832–842, 1986.

[16] A. Rezvanian and M. R. Meybodi, "Stochastic graph as a model for social networks," *Computers in Human Behavior*, vol. 64, pp. 621 – 640, 2016. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0747563216305222

[17] J. Besag, "Spatial Interaction and the Statistical Analysis of Lattice Systems," *Journal of the Royal Statistical Society. Series B (Methodological)*, vol. 36, no. 2, pp. 192–236, 1974. [Online]. Available: https://www.jstor.org/stable/2984812

[18] J. E. Besag, "Nearest-Neighbour Systems and the Auto-Logistic Model for Binary Data," *Journal of the Royal Statistical Society. Series B (Methodological)*, vol. 34, no. 1, pp. 75–83, 1972. [Online]. Available: https://www.jstor.org/stable/2985051

[19] G. Winkler, *Image Analysis, Random Fields and Markov Chain Monte Carlo Methods: A Mathematical Introduction*. Springer Science & Business Media, 2013.

[20] S. Wasserman and P. Pattison, "Logit models and logistic regressions for social networks: I. an introduction to markov graphs andp," *Psychometrika*, vol. 61, no. 3, pp. 401–425, 1996.

[21] T. A. Snijders, P. E. Pattison, G. L. Robins, and M. S. Handcock, "New specifications for exponential random graph models," *Sociological methodology*, vol. 36, no. 1, pp. 99–153, 2006.

[22] S. Geman and D. Geman, "Stochastic Relaxation, Gibbs Distributions, and the Bayesian Restoration of Images," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. PAMI-6, no. 6, pp. 721–741, 1984.

[23] ——, "Stochastic relaxation, gibbs distributions, and the bayesian restoration of images," in *Readings in computer vision*. Elsevier, 1987, pp. 564–584.

[24] C. J. Geyer and E. A. Thompson, "Constrained Monte Carlo maximum likelihood for dependent data," *Journal of the Royal Statistical Society. Series B (Methodological)*, pp. 657–699, 1992.

[25] C. J. Geyer, *Likelihood inference for spatial point processes. In 0. E. Barndorff-Nielsen, WS Kendall, and MNM van Lieshout, editors, Stochas-tic Geometry: Likelihood and Computation*. Chapman and Hall, 1999.

[26] M. S. Handcock, "Assessing Degeneracy in Statistical Models of Social Networks," *Journal of the American Statistical Association*, no. 76, pp. 33–50, 2003.

[27] A. Monfort, *Cours de statistique mathématique*. Economica, 1997.

[28] J. Møller and R. P. Waagepetersen, *Statistical Inference and Simulation for Spatial Point Processes*. CRC Press, 2004.

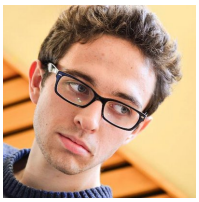
[29] D. R. Hunter, M. S. Handcock, C. T. Butts, S. M. Goodreau, and M. Morris, "Ergm: A Package to Fit, Simulate and Diagnose Exponential-Family Models for Networks," *Journal of Statistical Software*, vol. 24, no. 3, p. nihpa54860, 2008. [Online]. Available: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2743438/

[30] G. O. Roberts and A. F. Smith, "Simple conditions for the convergence of the gibbs sampler and metropolis-hastings algorithms," *Stochastic processes and their applications*, vol. 49, no. 2, pp. 207–216, 1994.

[31] G. O. Roberts and R. L. Tweedie, "Geometric convergence and central limit theorems for multidimensional Hastings and Metropolis algorithms," *Biometrika*, vol. 83, no. 1, pp. 95–110, 1996.

[32] G. O. Roberts and J. S. Rosenthal, "Optimal scaling for various

Metropolis-Hastings algorithms," *Statistical Science*, vol. 16, no. 4, pp. 351–367, 2001.
[33] ——, "General state space Markov chains and MCMC algorithms," *Probability Surveys*, vol. 1, no. 0, pp. 20–71, 2004.
[34] C. J. Geyer, "On the convergence of monte carlo maximum likelihood calculations," *Journal of Royal Statistical Society, Series B*, vol. 54, no. 1, pp. 261–274, 1994.
[35] J. Møller, A. N. Pettitt, R. Reeves, and K. K. Berthelsen, "An efficient markov chain monte carlo method for distributions with intractable normalising constants," *Biometrika*, vol. 93, no. 2, pp. 451–458, 2006.
[36] I. Murray, Z. Ghahramani, and D. J. C. MacKay, "MCMC for doubly-intractable distributions," in *Proceedings of the Twenty-Second Conference on Uncertainty in Artificial Intelligence*, ser. UAI'06. Arlington, Virginia, USA: AUAI Press, 2006, pp. 359–366.
[37] A. Caimo and N. Friel, "Bayesian inference for exponential random graph models," *Social Networks*, vol. 33, no. 1, pp. 41–55, 2011.
[38] Y. F. Atchadé, N. Lartillot, and C. Robert, "Bayesian computation for statistical models with intractable normalizing constants," *Brazilian Journal of Probability and Statistics*, vol. 27, no. 4, pp. 416–436, 2013. [Online]. Available: https://projecteuclid.org/euclid.bjps/1378729981
[39] M. A. Beaumont, J.-M. Cornuet, J.-M. Marin, and C. P. Robert, "Adaptive approximate Bayesian computation," *Biometrika*, vol. 96, no. 4, pp. 983–990, 2009. [Online]. Available: https://academic.oup.com/biomet/article/96/4/983/220502
[40] A. Grelaud, C. P. Robert, J.-M. Marin, F. Rodolphe, and J.-F. Taly, "ABC likelihood-free methods for model choice in Gibbs random fields," *Bayesian Analysis*, vol. 4, no. 2, pp. 317–335, 2009. [Online]. Available: https://projecteuclid.org/euclid.ba/1340370280
[41] J.-M. Marin, P. Pudlo, C. P. Robert, and R. J. Ryder, "Approximate Bayesian computational methods," *Statistics and Computing*, vol. 22, no. 6, pp. 1167–1180, 2012. [Online]. Available: https://doi.org/10.1007/s11222-011-9288-2
[42] R. Stoica, M. Deaconu, A. Philippe, and L. Hurtado-Gil, "Shadow Simulated Annealing: A new algorithm for approximate Bayesian inference of Gibbs point processes," *Spatial Statistics*, vol. 43, p. 100505, 2021. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S2211675321000154
[43] Q. Laporte-Chabasse, R. S. Stoica, F. Charoy, M. Clausel, and G. Oster, "Co-authoring graphs of research teams in a laboratory in computer science," 2020. [Online]. Available: https://doi.org/10.5281/zenodo.4309704
[44] M. Van Lieshout and R. S. Stoica, "The candy model: properties and inference," *Statistica Neerlandica*, vol. 57, no. 2, pp. 177–206, 2003.
[45] A. Datta, S. Buchegger, L.-H. Vu, T. Strufe, and K. Rzadca, "Decentralized Online Social Networks," in *Handbook of Social Network Technologies and Applications*, B. Furht, Ed. Boston, MA: Springer US, 2010, pp. 349–378.
[46] X. Guyon and C. Hardouin, "Markov chain markov field dynamics: Models and statistics," *Statistics: A Journal of Theoretical and Applied Statistics*, vol. 36, no. 4, pp. 339–363, 2002.
[47] C. Gaetan and X. Guyon, *Spatial Statistics and Modeling*, ser. Springer Series in Statistics. New York: Springer-Verlag, 2010.
[48] T. A. Snijders and J. Koskinen, "Longitudinal Models," in *Exponential Random Graph Models for Social Networks: Theory, Methods, and Applications*, D. Lusher, J. Koskinen, and G. Robins, Eds. New York, N.Y., Royaume-Uni de Grande-Bretagne et d'Irlande du Nord: Cambridge University Press, 2013.
[49] A. Caimo and N. Friel, "Bayesian model selection for exponential random graph models," *Social Networks*, vol. 35, no. 1, pp. 11–24, 2013.
[50] J. Møller and R. P. Waagepetersen, "Markov connected component fields," *Advances in Applied Probability*, vol. 30, no. 1, pp. 1–35, 1998.



Quentin Laporte-Chabasse is teacher at CESI Engineering School in Orléans, France. He received his PhD Degree in computer science from the University of Lorraine. His research interests are related to the study of collaborative and collective work through the lens of social network analysis, social interactions modelling and inference.

Radu S. Stoica is professor in mathematics at University of Lorraine. His research interests are related to probabilistic modeling and statistical inference for pattern analysis within spatio-temporal data. His theoretical work proposes Markov models, Monte Carlo algorithms and inference procedures adapted able to characterize and detect structured patterns hidden in the data. The application domains aimed are : astronomy, geosciences, image analysis and network sciences.

Marianne Clausel is Professor at University of Lorraine, in the Probability and Statistics department. Her research interests are data analysis and statistical learning. She received her PhD degree in Applied Mathematics from the University of Paris-Est Créteil, France and Habilitation Thesis in Statistics from Grenoble-Alpes University.

François Charoy is professor in Computer Science at University of Lorraine. His research interests are related to service computing and the engineering of collaborative systems. He is interested in how trust is built between people during collaboration and how collaboration links are established through collaborative systems.

Gérald Oster is an Associate Professor at LORIA, Inria Nancy - Grand Est, University of Lorraine. His research work explores distributed collaborative systems with a focus on optimistic content replication mechanisms. He is currently investigating the applicability of conflict-free replicated data-types (CRDTs), a novel class of optimistic replication algorithms, in diverse domains.

APPENDIX A
TRACES OF DISTRIBUTIONS SAMPLED FROM SYNTHETIC DATA WITH FIXED AND KNOWN PARAMETERS

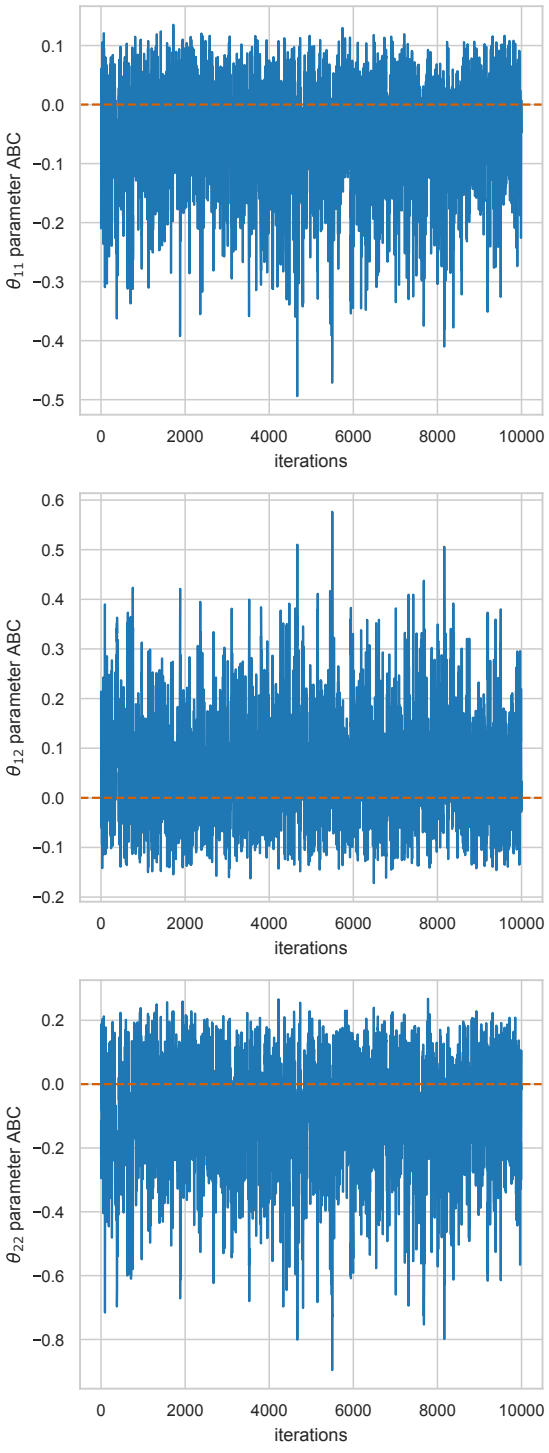


Fig. 12. Traces of parameters sampled with ABC Shadow from synthetic data. The true parameters are $\theta = [0, 0, 0]$ and are represented by dashed red lines.

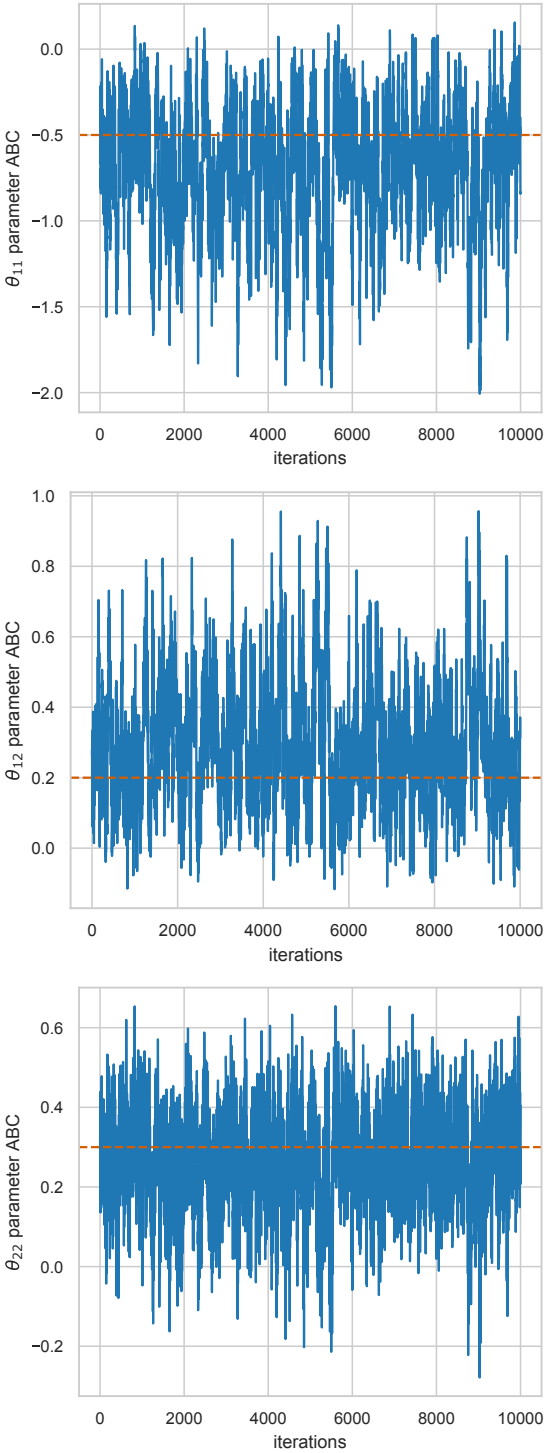


Fig. 13. Traces of parameters sampled with ABC Shadow from synthetic data. The true parameters are $\theta = [-0.5, 0.2, 0.3]$ and are represented by dashed red lines.

APPENDIX B
DIAGNOSTIC OF THE MODEL DEGENERACY

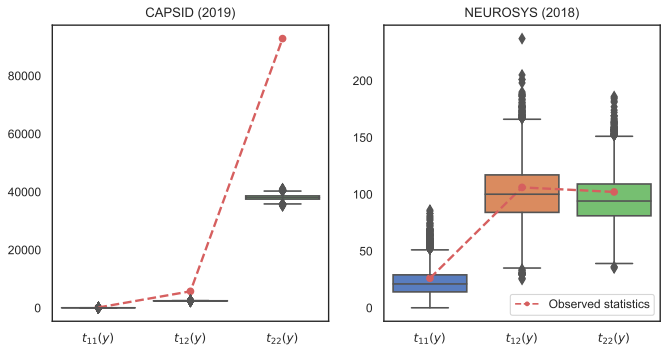


Fig. 14. The box plots represent the distributions of sufficient statistics, respectively t_{11}, t_{12} and t_{22}, generated from the MAP. The red points linked by dashes show the observed sufficient statics.

Figure 14 presents a comparison of the observed statistics with those obtained through simulations using the estimated MAP. In the first plot (on the left-hand side), the observed statistics do not match the sufficient statistics generated from the estimated MAP (Table 5). This situation is symptomatic of a model degeneracy [26]. The CAPSID team in 2019 shows very few researcher acting as "intra-organisational bridge" ($1 \leftrightarrow 1$ pattern) compared to the other two patterns ($1 \leftrightarrow 2$ and $2 \leftrightarrow 2$). The second plot exhibit analysis results for which statistics generated from the MAP correspond on average to the observed statistics. At our best knowledge, formulating an analytical solution for the degeneracy problem, is still an open question. As immediate perspective strategies for dealing with pathological situations, we mention model conditioning and tailored simulation dynamics. Regarding the inference, we mention recent developments presented in [42] allowing a variable Δ whenever sampling the posterior and stronger theoretical control of the obtained result.

APPENDIX C
NUMERICAL RESULTS AND EMPIRICAL STATISTICS

TABLE 3
Sufficient statistics of each observed collaboration graph

Team	Year	t_{11}	t_{12}	t_{22}	# LORIA members	# External collaborators
CAPSID	2017	54	291	463	10	45
	2018	104	111	79	13	16
	2019	123	5697	92796	19	311
CARAMBA	2017	6	28	38	10	15
	2018	3	4	5	9	9
	2019	0	2	26	8	15
LARSEN	2017	145	606	1446	21	78
	2018	173	259	171	21	16
	2019	176	295	199	23	22
MULTISPEECH	2017	238	345	543	33	65
	2018	141	561	648	26	70
	2019	332	1009	2268	36	120
NEUROSYS	2017	73	45	42	13	9
	2018	26	106	102	11	18
	2019	8	139	474	10	33
ORPAILLEUR	2017	280	402	314	29	39
	2018	324	374	202	31	33
	2019	308	5946	91322	34	332
PESTO	2017	49	116	146	16	26
	2018	18	36	282	12	33
	2019	6	30	88	12	28
RESIST	2017	40	95	73	9	13
	2018	54	60	40	15	11
	2019	146	491	500	23	56
SEMAGRAMME	2017	8	48	64	9	11
	2018	1	15	801	7	45
	2019	0	21	49	8	12
SMarT	2017	98	102	57	12	12
	2018	78	301	325	11	19
	2019	213	588	582	14	16
VERIDIS	2017	47	177	332	21	39
	2018	29	42	112	17	26
	2019	95	124	104	23	30

TABLE 4
Statistics on the number of internal and external stakeholders accounted for each team

	Year	Mean	Median	Standard deviation
Number of LORIA's members	2017	16.64	13.0	7.97
Number of external collaborator	2017	32.00	26.0	22.42
Number of LORIA's members	2018	15.73	13.0	7.11
Number of external collaborator	2018	26.91	19.0	17.09
Number of LORIA's members	2019	19.09	19.0	9.34
Number of external collaborator	2019	88.64	30.0	113.60

TABLE 5
Summary of estimates obtained from the collaboration networks of teams for the parameters θ_{11}, θ_{12} and θ_{22}.

Team	Year	$\overline{\theta_{11}}$	$Q_{50}\ \theta_{11}$	MAP θ_{11}	$\overline{\theta_{12}}$	$Q_{50}\ \theta_{12}$	MAP θ_{12}	$\overline{\theta_{22}}$	$Q_{50}\ \theta_{22}$	MAP θ_{22}
CAPSID	2017	0.119	0.133	0.163	-0.077	-0.079	-0.081	-0.111	-0.110	-0.109
	2018	0.082	0.084	0.083	-0.258	-0.257	-0.254	-0.158	-0.155	-0.148
	2019	-99.915	-99.946	-99.974	2.323	2.327	2.331	-0.077	-0.078	-0.095
CARAMBA	2017	-0.309	-0.237	-0.098	-0.294	-0.300	-0.308	-0.452	-0.444	-0.445
	2018	-0.413	-0.305	-0.114	-1.273	-1.217	-1.190	-1.410	-1.324	-1.151
	2019	-4.738	-4.014	-3.785	-1.277	-1.107	-0.955	-0.572	-0.562	-0.534
LARSEN	2017	0.059	0.061	0.060	-0.105	-0.103	-0.100	-0.123	-0.124	-0.124
	2018	0.001	0.004	0.011	-0.205	-0.205	-0.206	-0.097	-0.096	-0.098
	2019	-0.004	-0.001	0.006	-0.208	-0.209	-0.209	-0.183	-0.182	-0.184
MULTISPEECH	2017	-0.003	-0.002	0.001	-0.243	-0.242	-0.242	-0.300	-0.301	-0.307
	2018	-0.030	-0.024	-0.002	-0.127	-0.129	-0.132	-0.252	-0.252	-0.251
	2019	0.017	0.018	0.020	-0.126	-0.125	-0.121	-0.169	-0.171	-0.171
NEUROSYS	2017	0.043	0.046	0.047	-0.418	-0.414	-0.400	-0.096	-0.092	-0.097
	2018	-0.126	-0.100	-0.061	-0.098	-0.107	-0.121	-0.218	-0.214	-0.210
	2019	-0.542	-0.443	-0.336	-0.032	-0.037	-0.045	-0.082	-0.082	-0.081
ORPAILLEUR	2017	-0.000	0.000	-0.003	-0.222	-0.221	-0.222	-0.260	-0.260	-0.260
	2018	-0.018	-0.017	-0.014	-0.257	-0.256	-0.255	-0.309	-0.309	-0.311
	2019	-1.097	-1.087	-1.065	0.017	0.018	0.023	-0.020	-0.021	-0.030
PESTO	2017	0.009	0.020	0.037	-0.248	-0.247	-0.246	-0.271	-0.270	-0.272
	2018	0.102	0.121	0.143	-0.287	-0.278	-0.260	-0.171	-0.171	-0.175
	2019	-0.214	-0.156	0.006	-0.331	-0.326	-0.322	-0.420	-0.417	-0.420
RESIST	2017	0.116	0.132	0.158	-0.148	-0.154	-0.169	-0.110	-0.104	-0.088
	2018	-0.029	-0.023	-0.014	-0.392	-0.391	-0.375	-0.236	-0.228	-0.218
	2019	-0.000	0.006	0.018	-0.137	-0.139	-0.143	-0.234	-0.234	-0.236
SEMAGRAMME	2017	-0.367	-0.300	-0.190	-0.115	-0.121	-0.129	-0.125	-0.118	-0.089
	2018	-0.629	-0.226	0.223	-0.215	-0.179	-0.145	-0.042	-0.043	-0.044
	2019	-6.792	-5.984	-5.613	0.009	-0.005	-0.034	-0.261	-0.253	-0.227
SMarT	2017	0.079	0.083	0.083	-0.242	-0.243	-0.244	-0.109	-0.105	-0.094
	2018	0.074	0.087	0.116	-0.072	-0.076	-0.077	-0.015	-0.014	-0.014
	2019	0.100	0.102	0.104	-0.081	-0.080	-0.077	0.082	0.082	0.081
VERIDIS	2017	-0.024	-0.014	0.005	-0.218	-0.219	-0.219	-0.255	-0.256	-0.252
	2018	-0.022	-0.011	0.006	-0.441	-0.434	-0.422	-0.398	-0.397	-0.393
	2019	-0.026	-0.023	-0.018	-0.379	-0.377	-0.375	-0.461	-0.459	-0.458

TABLE 6
Error of the estimations: Asymptotic standard deviation and Monte Carlo standard deviation. Ranges of confidence intervals 95% for estimated MAPs

Team	Year	$\hat{\sigma}_{\theta_{11}}$	$\hat{\sigma}_{\theta_{12}}$	$\hat{\sigma}_{\theta_{22}}$	$\hat{\sigma}^{MC}_{\theta_{11}}$	$\hat{\sigma}^{MC}_{\theta_{12}}$	$\hat{\sigma}^{MC}_{\theta_{22}}$	CI 95% θ_{11}	CI 95% θ_{12}	CI 95% θ_{22}
CAPSID	2017	9.67e-02	3.11e-02	1.75e-02	6.40e-04	1.84e-04	5.99e-05	0.163 ± 1.3e-03	-0.081 ± 3.7e-04	-0.109 ± 1.2e-04
	2018	6.74e-02	5.17e-02	4.31e-02	2.61e-04	1.73e-04	1.49e-04	0.083 ± 5.2e-04	-0.254 ± 3.5e-04	-0.148 ± 3.e-04
	2019	6.74e-02	5.17e-02	4.31e-02	2.61e-04	1.73e-04	1.49e-04	-99.974 ± 5.2e-04	2.331 ± 3.5e-04	-0.095 ± 3.e-04
CARAMBA	2017	3.06e-01	1.38e-01	1.17e-01	6.75e-03	2.06e-03	1.43e-03	-0.098 ± 1.3e-02	-0.308 ± 4.1e-03	-0.445 ± 2.9e-03
	2018	3.19e-01	4.07e-01	3.8e-01	8.94e-03	1.49e-02	1.32e-02	-0.114 ± 1.8e-02	-1.19 ± 3.e-02	-1.151 ± 2.6e-02
	2019	3.19e-01	4.07e-01	3.8e-01	8.94e-03	1.49e-02	1.32e-02	-3.785 ± 1.8e-02	-0.955 ± 3.e-02	-0.534 ± 2.6e-02
LARSEN	2017	1.04e-01	2.07e-02	8.72e-03	8.83e-04	1.56e-04	2.80e-05	0.06 ± 1.8e-03	-0.1 ± 3.1e-04	-0.124 ± 5.6e-05
	2018	2.78e-02	3.31e-02	3.58e-02	5.77e-05	8.30e-05	1.01e-04	0.011 ± 1.2e-04	-0.206 ± 1.7e-04	-0.098 ± 2.0e-04
	2019	2.78e-02	3.31e-02	3.58e-02	5.77e-05	8.30e-05	1.01e-04	0.006 ± 1.2e-04	-0.209 ± 1.7e-04	-0.184 ± 2.0e-04
MULTISPEECH	2017	8.68e-02	3.03e-02	1.9e-02	5.44e-04	1.45e-04	4.94e-05	0.001 ± 1.1e-03	-0.242 ± 2.9e-04	-0.307 ± 9.9e-05
	2018	5.97e-02	2.89e-02	2.03e-02	1.96e-04	7.83e-05	4.41e-05	-0.002 ± 3.9e-04	-0.132 ± 1.6e-04	-0.251 ± 8.8e-05
	2019	5.97e-02	2.89e-02	2.03e-02	1.96e-04	7.83e-05	4.41e-05	0.02 ± 3.9e-04	-0.121 ± 1.6e-04	-0.171 ± 8.8e-05
NEUROSYS	2017	2.68e-02	6.86e-02	7.37e-02	1.05e-04	4.42e-04	4.48e-04	0.047 ± 2.1e-04	-0.4 ± 8.8e-04	-0.097 ± 9.e-04
	2018	1.90e-01	8.58e-02	6.16e-02	2.36e-03	8.35e-04	4.74e-04	-0.061 ± 4.7e-03	-0.121 ± 1.7e-03	-0.21 ± 9.5e-04
	2019	1.90e-01	8.58e-02	6.16e-02	2.36e-03	8.35e-04	4.74e-04	-0.336 ± 4.7e-03	-0.045 ± 1.7e-03	-0.081 ± 9.5e-04
ORPAILLEUR	2017	4.39e-02	2.96e-02	2.5e-02	9.45e-05	5.2e-05	4.78e-05	-0.003 ± 1.9e-04	-0.222 ± 1.0e-04	-0.26 ± 9.6e-05
	2018	2.30e-02	2.94e-02	3.67e-02	4.94e-05	6.88e-05	1.11e-04	-0.014 ± 9.9e-05	-0.255 ± 1.4e-04	-0.311 ± 2.2e-04
	2019	2.30e-02	2.94e-02	3.67e-02	4.94e-05	6.88e-05	1.11e-04	-1.065 ± 9.9e-05	0.023 ± 1.4e-04	-0.03 ± 2.2e-04
PESTO	2017	9.52e-02	5.37e-02	4.16e-02	5.37e-04	2.21e-04	1.54e-04	0.037 ± 1.1e-03	-0.246 ± 4.4e-04	-0.272 ± 3.1e-04
	2018	8.05e-01	7.36e-02	2.41e-02	6.17e-02	3.90e-03	1.99e-04	0.143 ± 1.2e-01	-0.26 ± 7.8e-03	-0.175 ± 4.e-04
	2019	8.05e-01	7.36e-02	2.41e-02	6.17e-02	3.90e-03	1.99e-04	0.006 ± 1.2e-01	-0.322 ± 7.8e-03	-0.42 ± 4.e-04
RESIST	2017	6.67e-02	5.71e-02	6.45e-02	2.28e-04	1.91e-04	3.15e-04	0.158 ± 4.6e-04	-0.169 ± 3.8e-04	-0.088 ± 6.3e-04
	2018	6.73e-02	7.77e-02	8.21e-02	3.33e-04	4.90e-04	5.58e-04	-0.014 ± 6.7e-04	-0.375 ± 9.8e-04	-0.218 ± 1.1e-03
	2019	6.73e-02	7.77e-02	8.21e-02	3.33e-04	4.90e-04	5.58e-04	0.018 ± 6.7e-04	-0.143 ± 9.8e-04	-0.236 ± 1.1e-03
SEMAGRAMME	2017	3.72e-01	1.15e-01	6.79e-02	1.09e-02	2.59e-03	9.77e-04	-0.19 ± 2.2e-02	-0.129 ± 5.2e-03	-0.089 ± 2.e-03
	2018	5.9e+00	7.36e-02	9.37e-03	3.47e+00	2.16e-02	1.64e-04	0.223 ± 6.9e+00	-0.145 ± 4.3e-02	-0.044 ± 3.3e-04
	2019	5.9e+00	7.36e-02	9.37e-03	3.47e+00	2.16e-02	1.64e-04	-5.613 ± 6.9e+00	-0.034 ± 4.3e-02	-0.227 ± 3.3e-04
SMarT	2017	4.76e-02	5.27e-02	5.51e-02	1.52e-04	2.04e-04	2.42e-04	0.083 ± 3.0e-04	-0.244 ± 4.1e-04	-0.094 ± 4.8e-04
	2018	6.39e-02	3.65e-02	2.84e-02	1.79e-04	8.52e-05	6.58e-05	0.116 ± 3.6e-04	-0.077 ± 1.7e-04	-0.014 ± 1.3e-04
	2019	6.39e-02	3.65e-02	2.84e-02	1.79e-04	8.52e-05	6.58e-05	0.104 ± 3.6e-04	-0.077 ± 1.7e-04	0.081 ± 1.3e-04
VERIDIS	2017	1.23e-01	4.19e-02	2.53e-02	1.11e-03	2.88e-04	9.59e-05	0.005 ± 2.2e-03	-0.219 ± 5.8e-04	-0.252 ± 1.9e-04
	2018	2.49e-01	8.48e-02	5.15e-02	5.04e-03	1.07e-03	3.35e-04	0.006 ± 1.0e-02	-0.422 ± 2.1e-03	-0.393 ± 6.7e-04
	2019	2.49e-01	8.48e-02	5.15e-02	5.04e-03	1.07e-03	3.35e-04	-0.018 ± 1.0e-02	-0.375 ± 2.1e-03	-0.458 ± 6.7e-04

TABLE 7
Results of the t-test applied on each parameter to check if the parameter are significant against pure chance. Here the score of the test as well as the corresponding p-value are presented. Except for four observed networks marked by *, the parameters are significant.

Team	Year	TS($\theta_{11}, 0$)	p-val1	$TS(\theta_{12}, 0)$	p-val2	$TS(\theta_{22}, 0)$	p-val3
CAPSID	2017	43.125	$\leq 10^{-6}$	-80.089	$\leq 10^{-6}$	-193.181	$\leq 10^{-6}$
	2018	111.135	$\leq 10^{-6}$	-199.204	$\leq 10^{-6}$	-101.268	$\leq 10^{-6}$
	2019	-34830.327	$\leq 10^{-6}$	699.856	$\leq 10^{-6}$	-95.875	$\leq 10^{-6}$
CARAMBA	2017	-28.036	$\leq 10^{-6}$	-69.123	$\leq 10^{-6}$	-126.066	$\leq 10^{-6}$
	2018	-31.093	$\leq 10^{-6}$	-89.480	$\leq 10^{-6}$	-90.430	$\leq 10^{-6}$
	2019	-46.805	$\leq 10^{-6}$	-57.079	$\leq 10^{-6}$	-123.379	$\leq 10^{-6}$
LARSEN	2017	87.699	$\leq 10^{-6}$	-199.394	$\leq 10^{-6}$	-273.585	$\leq 10^{-6}$
	2018*	1.662	9.683e-02	-224.536	$\leq 10^{-6}$	-95.604	$\leq 10^{-6}$
	2019	-5.170	$\leq 10^{-6}$	-243.612	$\leq 10^{-6}$	-180.908	$\leq 10^{-6}$
MULTISPEECH	2017	-6.971	$\leq 10^{-6}$	-402.422	$\leq 10^{-6}$	-437.728	$\leq 10^{-6}$
	2018	-21.629	$\leq 10^{-6}$	-170.839	$\leq 10^{-6}$	-418.720	$\leq 10^{-6}$
	2019	36.814	$\leq 10^{-6}$	-258.178	$\leq 10^{-6}$	-356.691	$\leq 10^{-6}$
NEUROSYS	2017	51.231	$\leq 10^{-6}$	-177.527	$\leq 10^{-6}$	-41.829	$\leq 10^{-6}$
	2018	-22.524	$\leq 10^{-6}$	-38.494	$\leq 10^{-6}$	-111.927	$\leq 10^{-6}$
	2019	-35.630	$\leq 10^{-6}$	-17.616	$\leq 10^{-6}$	-147.183	$\leq 10^{-6}$
ORPAILLEUR	2017*	-1.495	1.353e-01	-362.590	$\leq 10^{-6}$	-322.979	$\leq 10^{-6}$
	2018	-36.256	$\leq 10^{-6}$	-347.376	$\leq 10^{-6}$	-288.879	$\leq 10^{-6}$
	2019	-81.610	$\leq 10^{-6}$	18.218	$\leq 10^{-6}$	-24.374	$\leq 10^{-6}$
PESTO	2017	5.266	$\leq 10^{-6}$	-188.814	$\leq 10^{-6}$	-211.214	$\leq 10^{-6}$
	2018	42.418	$\leq 10^{-6}$	-141.148	$\leq 10^{-6}$	-207.922	$\leq 10^{-6}$
	2019	-24.260	$\leq 10^{-6}$	-118.935	$\leq 10^{-6}$	-209.319	$\leq 10^{-6}$
RESIST	2017	42.744	$\leq 10^{-6}$	-75.633	$\leq 10^{-6}$	-56.265	$\leq 10^{-6}$
	2018	-21.262	$\leq 10^{-6}$	-169.616	$\leq 10^{-6}$	-86.393	$\leq 10^{-6}$
	2019*	-0.192	8.475e-01	-185.076	$\leq 10^{-6}$	-349.118	$\leq 10^{-6}$
SEMAGRAMME	2017	-31.127	$\leq 10^{-6}$	-31.295	$\leq 10^{-6}$	-53.773	$\leq 10^{-6}$
	2018	-17.907	$\leq 10^{-6}$	-52.145	$\leq 10^{-6}$	-89.476	$\leq 10^{-6}$
	2019*	-60.880	$\leq 10^{-6}$	1.474	1.409e-01	-88.590	$\leq 10^{-6}$
SMarT	2017	92.722	$\leq 10^{-6}$	-160.254	$\leq 10^{-6}$	-54.549	$\leq 10^{-6}$
	2018	32.730	$\leq 10^{-6}$	-65.715	$\leq 10^{-6}$	-20.331	$\leq 10^{-6}$
	2019	132.546	$\leq 10^{-6}$	-141.465	$\leq 10^{-6}$	154.850	$\leq 10^{-6}$
VERIDIS	2017	-13.685	$\leq 10^{-6}$	-229.062	$\leq 10^{-6}$	-316.399	$\leq 10^{-6}$
	2018	-12.669	$\leq 10^{-6}$	-206.286	$\leq 10^{-6}$	-237.182	$\leq 10^{-6}$
	2019	-31.123	$\leq 10^{-6}$	-275.823	$\leq 10^{-6}$	-270.959	$\leq 10^{-6}$