

NRC Publications Archive Archives des publications du CNRC

Determining interaction directionality in complex biochemical networks from stationary measurements

Leibovich, N.

This publication could be one of several versions: author's original, accepted manuscript or the publisher's version. / La version de cette publication peut être l'une des suivantes : la version prépublication de l'auteur, la version acceptée du manuscrit ou la version de l'éditeur.

For the publisher's version, please access the DOI link below. / Pour consulter la version de l'éditeur, utilisez le lien DOI ci-dessous.

Publisher's version / Version de l'éditeur:

<https://doi.org/10.1038/s41598-025-86332-0>

Scientific Reports, 15, pp. 1-11, 2025-01-23

NRC Publications Archive Record / Notice des Archives des publications du CNRC :

<https://nrc-publications.canada.ca/eng/view/object/?id=edc25d27-e6eb-4e7c-89dc-1a1845aa2232>

<https://publications-cnrc.canada.ca/fra/voir/objet/?id=edc25d27-e6eb-4e7c-89dc-1a1845aa2232>

Access and use of this website and the material on it are subject to the Terms and Conditions set forth at

<https://nrc-publications.canada.ca/eng/copyright>

READ THESE TERMS AND CONDITIONS CAREFULLY BEFORE USING THIS WEBSITE.

L'accès à ce site Web et l'utilisation de son contenu sont assujettis aux conditions présentées dans le site

<https://publications-cnrc.canada.ca/fra/droits>

LISEZ CES CONDITIONS ATTENTIVEMENT AVANT D'UTILISER CE SITE WEB.

Questions? Contact the NRC Publications Archive team at

PublicationsArchive-ArchivesPublications@nrc-cnrc.gc.ca. If you wish to email the authors directly, please see the first page of the publication for their contact information.

Vous avez des questions? Nous pouvons vous aider. Pour communiquer directement avec un auteur, consultez la première page de la revue dans laquelle son article a été publié afin de trouver ses coordonnées. Si vous n'arrivez pas à les repérer, communiquez avec nous à PublicationsArchive-ArchivesPublications@nrc-cnrc.gc.ca.



OPEN Determining interaction directionality in complex biochemical networks from stationary measurements

N. Leibovich

Revealing interactions in complex systems from observed collective dynamics constitutes a fundamental inverse problem in science. Some methods may reveal undirected network topology, e.g., using node-node correlation. Yet, the direction of the interaction, thus a causal inference, remains to be determined - especially in steady-state observations. We introduce a method to infer the directionality within this network only from a “snapshot” of the abundances of the relevant molecules. We examine the validity of the approach for different properties of the system and the data recorded, such as the molecule’s level variability, the effect of sampling and measurement errors. Simulations suggest that the given approach successfully infer the reaction rates in various cases.

Many complex systems in physics and biology constitute networks of dynamically interacting units¹. In these systems, especially in biological ones, the process is often described with stochastic variables^{2–4}, and their collective dynamics and functions are governed by the network structure. Determining the underlying interaction topology is essential, for example, for understanding and controlling their function, identifying new pathways in gene regulatory networks or understanding some metabolic mechanisms^{5–7}. Particularly, understanding the direction of metabolic reactions is fundamental for predicting cellular behavior, optimizing metabolic engineering strategies, and identifying potential drug targets. Revealing these interaction networks poses a great challenge. Therefore, various studies have examined methods to find the structure of interaction networks^{8–13}.

Commonly, interaction networks are constructed from data obtained by time-series measurements, or even pseudo-time trajectories^{14–23}. To do so, researchers use some mathematical and statistical tools such as Bayesian inference and maximum likelihood together with machine-learning algorithms^{17–22}. There, the collective dynamics are used for network inference—whether the interactions are approximately linear or remain nonlinear, where the latter requires further assumptions on the system^{24–27}. However, multidimensional time-dependent synchronized recorded data is not available in some scenarios.

Generally, many biological processes are presumed to reach a steady state^{21,27–33}. Importantly, while the population of molecules within a cell or organism is dynamic - constantly undergoing synthesis and degradation - the molecular abundances display time-independent characteristics. The use of steady-state data for the inference of interaction networks has been examined, yet the dynamics are assumed to be linear in the activity of the nodes of the network³⁴, or the system is assumed to be externally driven in a controlled way^{27,29}. Nevertheless, real systems are not required to fulfill these requirements.

In many cases the collective dynamics evolve via the (Markovian-) Master equation, where stochastic variables possess integer values³⁵. Specifically, one of the well known models studied using the Master equation is the stochastic birth-death process which describes stochastic biochemical reactions³⁵. In general, interaction networks of variables which evolve with Master dynamics are not limited to biochemical reaction networks, and has also been studied in other disciplines, see³⁶ and references therein. Yet, for consistency, this manuscript is oriented to bio-chemical reaction network studies, with adequate terminology.

In biochemical reaction network studies, one may infer the *functional* connectivity which gains insight into statistical dependencies that the entire set of interactions yield between pairs of units through the collective network dynamics^{37,38}. The statistical dependencies between components may be given by the well-known Pearson correlation, the mutual information, or their analyses with silencing methods^{38–40}. These, however, are symmetric metrics and as such they are blind to the direction of the interaction providing only non-directed

National Research Council of Canada, NRC-Fields Mathematical Sciences Collaboration Centre, 222 College st., Toronto, ON M5T 3J1, Canada. email: Nava.Leibovich@nrc-cnrc.gc.ca

networks. Non-symmetric metrics, such as partial correlations and dependency analyses, require recorded data from all other interacting variables within the network^{31,41–44}. That data is not available in many systems.

Here, we introduce a novel data analysis approach to infer the interaction direction, which under some conditions may be interpreted as causality. Its key advantage is that only the variables under investigation need to be measured, while the mechanistic details of regulation or degradation of all other components within the network do not need to be modeled or even to be observed, see Fig. 1. This approach exploits the global probability flux-balance equation that must be satisfied in the steady state^{45–48}. In contrast to existing works, our approach does not require temporal information, experimental perturbations, or complete observation of all components within an interaction network.

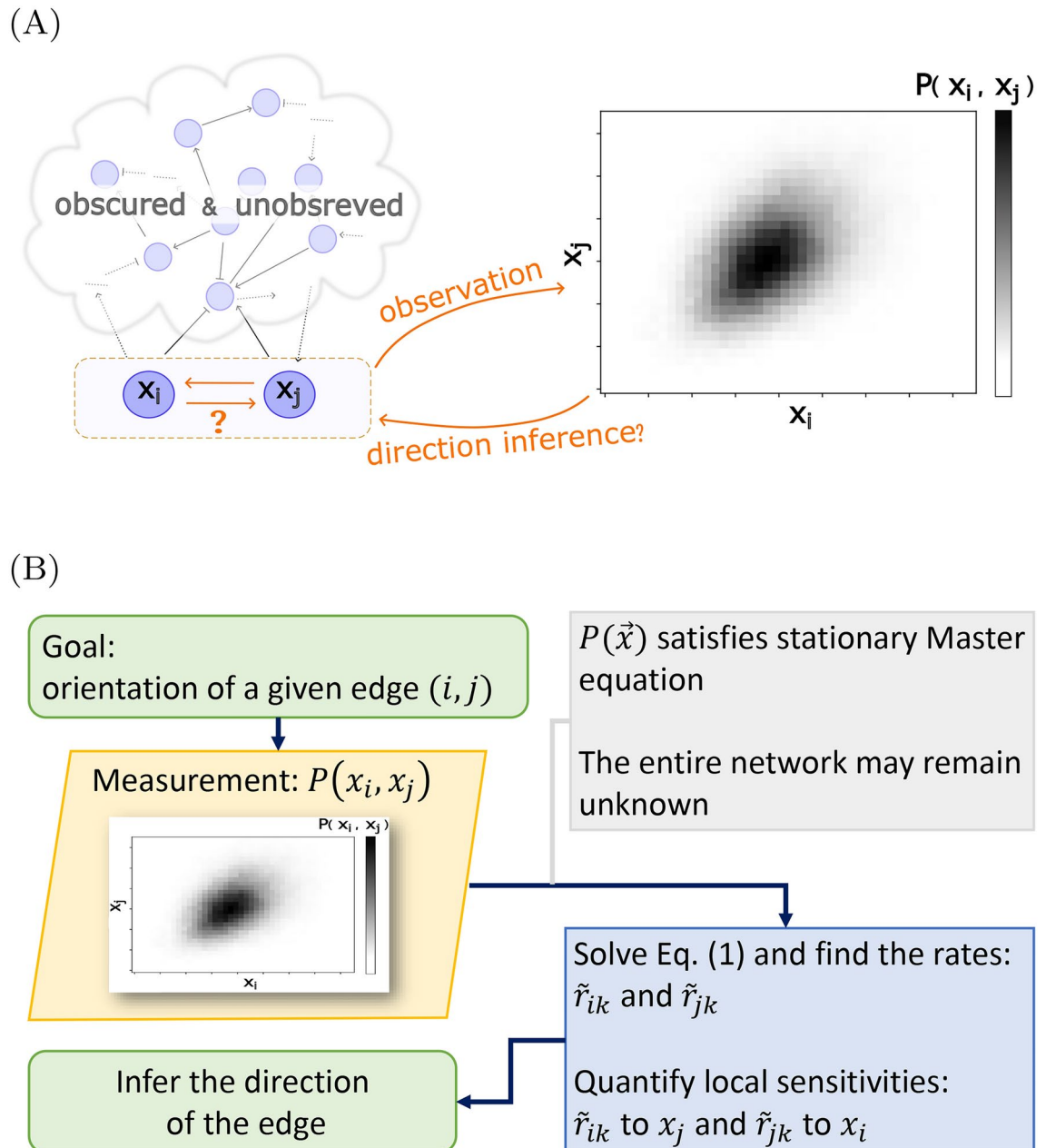


Fig. 1. Main goal and inference method flowchart. (A) Determining the direction of interaction between stochastic variables is challenged by the lack of temporal observations, further knowledge of the entire network topology, or records of other variables within that network. (B) Our method exploits the steady-state joint distribution to infer the direction of arrows using the global balance equation Eq. (1) and the local sensitivity analysis.

Results

Theoretical background and overview

Consider V variables coupled via an interaction network, i.e., a directed graph with V vertices. The variables state vector is given by $\vec{x} = \{x_1, x_2, \dots, x_V\}$, where x_i represents the number of units from type i , e.g. its expression or activity level. Each variable $x_i(t) \in \mathbb{N}_0$ and evolves with time by probabilistic events $x_i \xrightarrow{r_{ik}(\vec{x})} x_i + d_{ik}$ where the k -th reaction changes the abundance of the i -th molecule x_i by d_{ik} . The reaction rates $r_{ik}(\vec{x})$ depend on the state vector \vec{x} according to the interaction network. The stationary Master equation yields the global balance relation, which means that for every node i

$$0 = \sum_k \langle \tilde{r}_{ik}(x_i - d_{ik}, x_j) | x_i - d_{ik} \rangle P(x_i - d_{ik}) - \langle \tilde{r}_{ik}(x_i, x_j) | x_i \rangle P(x_i), \quad (1)$$

where $\tilde{r}_{ik}(x_i, x_j) \equiv \langle r_{ik}(\vec{x}) | x_i, x_j \rangle$ and the angular brackets represent conditional means. The above relation is derived from a summation of the Master equation over all other variables where consider stationarity $\partial_t P = 0$ ^{45–48}, see further derivation in “Methods” and SI.

From Eq. (1) one may obtain the reaction rate only from a given joint steady-state distribution of the two molecules $P(x_i, x_j)$, see “Methods” and⁴⁸. Our approach is based on *local* sensitivity of \tilde{r}_{ik} to x_j ; a direction of the interaction is inferred, which means an arrow is drawn in the directed graph, if a sufficient sensitivity is quantified, see details in “Methods” and SI. We remark that the suggested directional inference approach solely uses information about (x_i, x_j) , thus the directionality inference problem decomposes over pairs of nodes within the network such that the direction of each interaction can be reconstructed independently.

To evaluate the quality of the inference, we initially consider a class of random networks with Michaelis–Menten kinetics which model gene regulatory networks⁴⁹, and random networks where each unit is a Goodwin oscillator, a prototypical biological oscillator, that characterizes various biological processes such as circadian clocks^{50,51}. Additionally, we examine our inference strategy on a biological realistic system based on a subset of an *E. coli* gene regulatory network given in^{52–55}. In the following, we evaluate the performance of the suggested approach to identifying the direction of a given interaction.

Importantly, Eq. (1) is not an approximation but an exact flux balance relation at stationarity. However, a finite number of recorded points N introduces sampling errors into the distribution $P(\vec{x})$, as shown in SI. These sampling errors may affect the inference of directions. Therefore, we also examine below the impact of the number of sampling points, N , on the overall results.

Performance in general synthetic networks

Our strategy successfully infers the direction of the interaction between two nodes within an unknown network of interactions, where only the two-node joint steady-state distribution is given as the input, see Figs. 2, 3 and 4. It demonstrates that stationary probability distributions contain sufficient information to deduce the direction of a given interaction within an unknown network of interactions.

One of the most commonly used classifiers’ performance scores is the receiver operating characteristic curve (ROC)—which compares the true positive rate (TPR) to the false positive rate (FPR) at different threshold settings. A perfect classification yields TPR=1 with FPR=0, while a random one forms a diagonal line where FPR = TPR. The area under the receiver operating characteristic curve (AUC) is a single-valued score indicating overall performance, ranging from 0.5 for random to near 1 for perfect classification. We use these to evaluate the suggested direction inference strategy.

As aforementioned, our method poses a good direction classifier—it successfully infers the direction of a given edge where the relevant joint distribution is sufficiently sampled. As is shown in Fig. 2, the direction inference improves with a larger set of data points. In particular, a sufficiently large data set provides scores beyond the random classifier. It signifies that the direction of interaction is indeed encoded within the stationary snapshot of many data points. Importantly, recall that the input is simply the two-variables joint distribution, our method allows decomposition of the network. This means that within a given un-directed graph, one can learn the direction of a single edge independently, regardless of other edges, see further discussion in SI.

In the following, we discuss some key aspects that we found essential for the underlying goal. In particular, we examine some determinants that may affect probability to correctly infer directionality.

Determinants and limitations

The stationary joint probability density between x_i and all variables directly interacting with the i -th molecule must satisfy Eq. (1), hence it acts as a self-consistency test for stationarity. It is shown in⁴⁸ how this relation can be “inverted” to determine rate functions from observed probability densities. In other words, the information about the dynamical rate of a variable with its dependence on its coupled variables is indeed encoded within the joint steady-state distribution. However, while Eq. (1) must provably hold for all stationary states, empirically observed distributions carry sampling errors that can significantly impact inferred rates⁴⁸, leading to potential errors in determining the interaction direction. Our results reflect this errors’ dependence, as illustrated in Fig. 2; larger N , which is equivalent to lower sampling errors, yields increased accuracy (SI).

A further characteristic examined is the index of dispersion D , i.e. the variance to mean ratio. It quantifies the variable dependence on others, where $D \approx 1$ indicates that the variable’s dynamic is weakly dependent on other variables (SI). We show in Fig. 3A that a higher probability of success inference is found where D is far from 1. We note that similar behavior was previously reported in⁴⁸. For a given edge $x_j \rightarrow x_i$, the functional

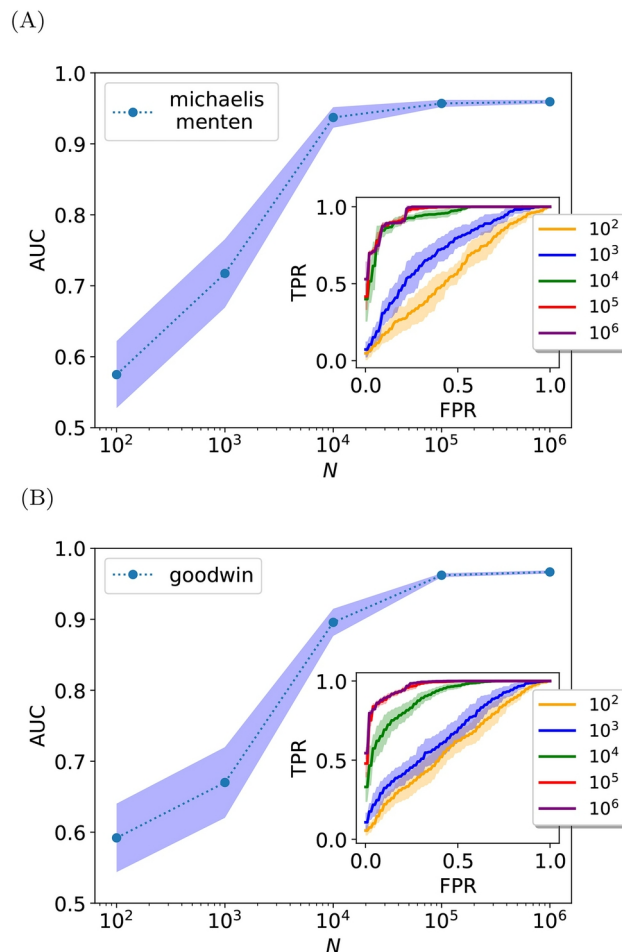


Fig. 2. Inferring interaction directions from steady state. Simulation results indicate an improvement with increasing the number of data points N . Panel (A): The AUC scores for the Michaelis–Menten model (see “Methods”). Inset: its corresponding ROC curves. Panel (B): Similar results are obtained for the coupled Goodwin oscillators. The shaded area corresponds to the standard deviation.

dependence of $r_i^+(x_j)$ in the level x_j is better determined for higher mutual information between these two variables.

Moreover, some local properties of the network may affect the performance of our direction inference method. We examine the in- and out- degrees, and the lifetimes of the molecules. Recall that we evaluate the sensitivity of $\tilde{r}_i(x_j)$ to x_j , where $\tilde{r}_i(x_j)$ is a quantity averaged over all other income edges, thus we expect that the connectivity nature possesses an influence over the performance. Additionally, the lifetime of a molecule is also expected to play an important role in the overall performance as is discussed in⁴⁸. In the results presented in Fig. 3B we have found that the in-degree of x_j presents some dependence over the performance, yet we did not find it significant, see further analysis and simulation results in SI. Moreover, we show in Fig. 3C that a lifetimes ratio far from 1 results in decreasing the probability for an accurately inferred direction of influence, which agrees with⁴⁸. Nevertheless, we comment that our network ensemble does not cover the entire possible network topologies space, an important factor that might skew the results, thus the nature of these phenomena needs further research.

***E. coli* gene regulatory network**

We aggregate the insights from the above, and apply our strategy for the direction inference on a previously proposed biologically realistic system—a model which is based on *E. coli* gene regulatory network provided in the DREAM challenge^{52–54}, see “Methods”. The system, shown in Fig. 4, poses various characteristic lifetimes and vertices’ degrees. We decompose the network and analyze each edge separately with its measured two-node joint distribution.

As shown, our inference method performs well for edges from nodes with zero in-degree with relatively long lifetimes (all edges from nodes 0 and 5), while erroneous direction emerges at edges connecting high-degree nodes with short lifetimes (nodes 1, 6 where each connected to two wrongly drawn arrows). We comment that our local two-node analysis may provide pathways that are not encoded in the true network, therefore complementary use of global analysis^{56–61} should be considered, especially when all other nodes were observed.

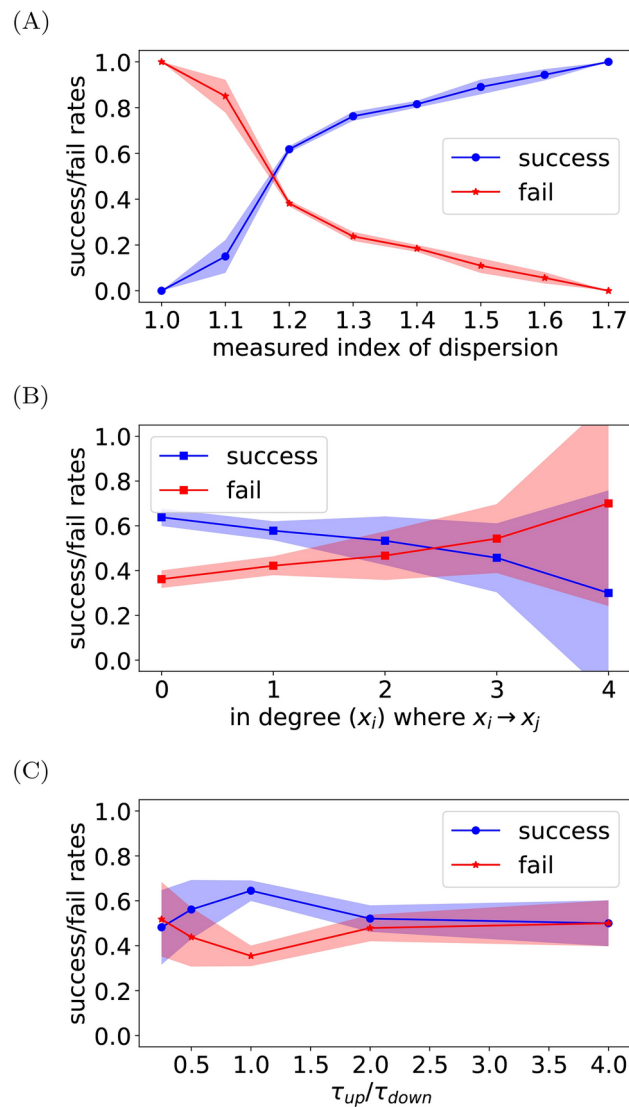


Fig. 3. Factors dictate the performance. **(A)**: the index of dispersion affects the probability of a successful inference of the direction. Results obtained the Michaelis–Menten model with $N = 10^2$. **(B)** Dependence of the degree on the directionality inference and **(C)** lifetimes dependence obtained from the Michaelis–Menten model with $N = 10^2$. For all panels the shaded area represents the standard deviation.

Observation requirements

Our model is not restricted to a specific interpretation of the fluctuating variables (nodes). Generally, the set $\{x_i\}$ could represent levels of DNA, mRNA, complexes, proteins, or metabolites. For instance, our model applies to scenarios involving the expression levels of multiple genes (e.g., obtained via microarray or RNA-seq) alongside genotypes of cis-eQTL, similar to^{62,63}. Another example is the two-dimensional data observations in³², where x_i and x_j are the measured mRNA and protein copy numbers. The authors used a yellow fluorescent protein fusion library to examine the single-cell profile of *E. coli*. Depending on the desired insights and measurable variables, one should select the most appropriate recording method.

As mentioned, we assume the process is governed by the stationary Master equation, with an interaction between the variables x_i and x_j . The inference approach requires measuring $P(x_i, x_j)$ to determine the interaction orientation between these variables. Principally, the inference approach is quite general, allowing for varied underlying processes and biological interpretations of x_i and x_j . For example, gene-regulatory networks can be modeled as discrete systems using the Master equation⁴⁹. The Master equation can also describe spatial dependencies, such as complexes at a given codon or in modeling allosteric regulation. Furthermore, models can incorporate multiple gene transcription-translation processes, feedback loops, and ribosome stalling, to name only a few. These models may be simplified through “coarse-graining” or maintained in their complexity. The inference method thus requires the measurement of the joint probability $P(x_i, x_j)$ corresponding to the interpretation of each variable, see further discussion in the SI. Note that the interaction network is assumed to be static, unlike dynamic networks^{14,62}.

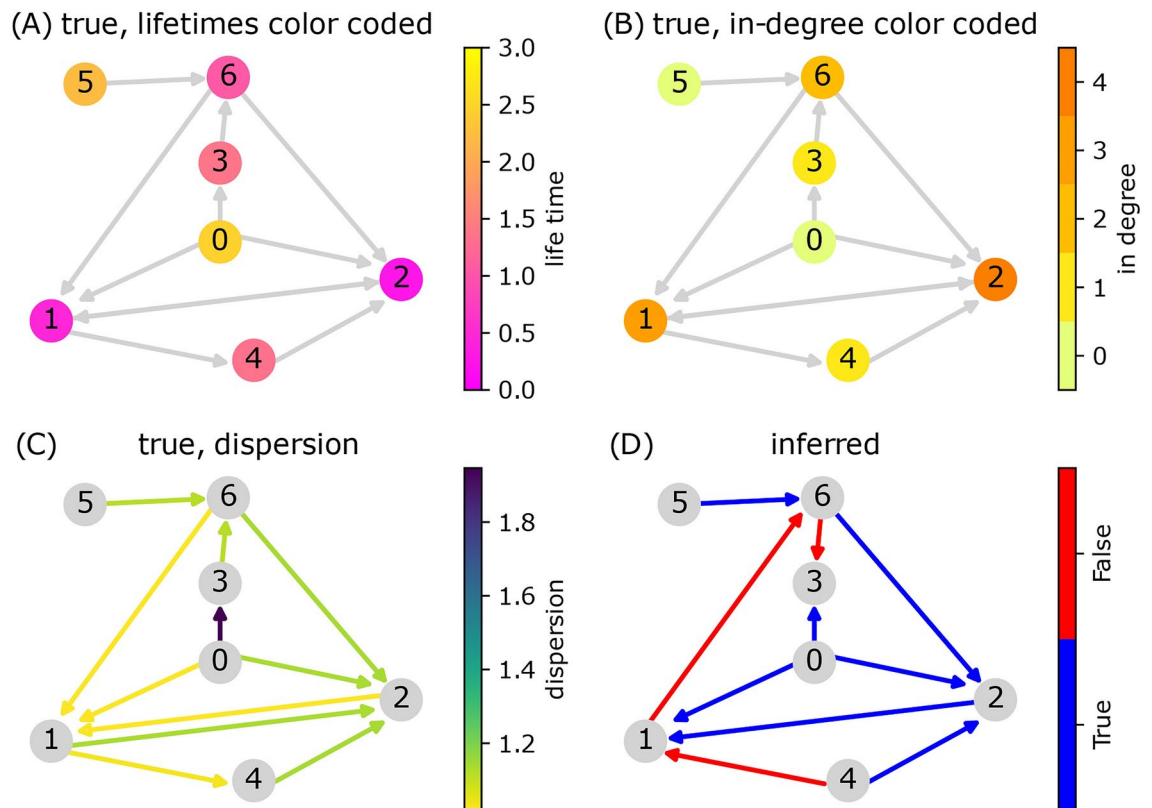


Fig. 4. Simulation results for the *E. coli* gene regulatory network model. The features of the nodes and edges are color-coded according to the sub-figure titles and the provided color bars. Panels (A–C) display the true network, with each panel highlighting a specific feature in color: lifetimes and in-degree of the nodes in panels (A) and (B), and the index of dispersion which is associated with node-node correlation is shown in panel (C). Panel (D) presents the inference results, where correctly inferred directions are shown in blue, and incorrect ones are depicted in red.

Discussion

Unveiling the interaction network among stochastic variables and delineating the direction of each interaction is crucial for understanding biological systems. Here, we present a method designed to infer the directionality within this network solely from the steady state of pertinent molecules. We assess the effectiveness of this approach across various properties of the system. Simulation results indicate that the proposed method demonstrates success in inferring the reaction direction in various cases.

As mentioned, our strategy relies on the time-invariant density, and as such possesses a benefit as explained in the following. Recorded data points can be employed without a known temporal order, taken at different sampling intervals, or aggregated from multiple experiments, as long as they were conducted under the same conditions. Hence this stationarity assumption enables the analysis of a wide range of biological systems. Importantly, Eq. (1) is not an approximation but corresponds to an exact flux balance relation at stationarity. The only dynamics excluded from our analysis is transient behavior such that stationary probability distributions are not accessible from experimental data. Even so, explicitly time-varying systems that technically never reach a stationary state, such as deterministic oscillations, may satisfy a similar form of Eq. (1) when considering their time-averaged probability distributions and rates. However, care has to be taken when comparing such time averages with population averages, see further discussion in SI.

The static snapshot provides statistical features of the stochastic variables, enabling the inference of the interaction direction. Statistical information is frequently used to gain insights from these snapshots about the dynamics and the relations between variable components^{29,38,39,64–66}. Our strategy for revealing interaction directions thus constitutes a previously unknown intermediate approach, placed between purely statistical methods used to infer effective connectivity (such as correlations) and approaches that infer physical connectivity from high-resolution, time-ordered recordings of the complete dynamics^{26,27,34,67,68}. In particular, statistical methods for inferring *directed* edges that rely on partial correlation analyses were applied for revealing directed graphs^{42,44,69,70}. These methods, however, require measurements from other components within the system and in that sense are non-local. Nevertheless, we tested the performance of the partial correlation analyses over our synthetic data and found that it is underachieved and thus inadequate for our goal (see SI).

Using the sensitivity analysis presents an equivalence with the examination of the slope of $\tilde{r}_{ik}(x_j)$ or its linear response. This linear sensitivity analysis for the interaction network identification is conceptualized

and demonstrated for stationary systems under small perturbations²⁹. There, the authors assume continuum variables that evolve via coupled differential equations with no noise or with additive noise. Conversely, our method considers discrete variables that propagate following the Master equation which effectively possess multiplicative noise. In addition, we note that our approach does not require externally driving the system in a controlled way that is not applicable in many systems²⁹.

Another possible directionality classification quantity is related to the Sum-of-Squared Errors (SSE), which might be associated with the Minimum Description Length (MDL). The SSE (or the MDL) comparison between the two directions is well mathematically established^{71–73}. Nevertheless, we show (see SI) that our directionality classification quantity based on $|J_{ij}|$ outperforms the previously proposed SSE-based one, especially in low N . In the SI we further discuss the classifying quantities, with additional simulation results.

Harnessing of the global flux balance Eq. (1) to gain insight on the interaction quantification has been suggested recently⁴⁸. There, the authors demonstrated that the steady state joint distribution indeed encapsulates information about the dynamical features such as the birth rates. They assume that the direction of the edge, together with the stationary $P(x_i, x_j)$, are given, and thus find the functional shape of $r_i^+(x_j)$ for various values of x_j . Here we exploit this approach, but instead of looking for the full functional shape we only examine a partial behavior of $r_i^+(x_j)$ - its sensitivity analysis - without assuming the direction of the edge at all, see further discussion on SI.

Furthermore, we have examined some of the performance determinants, i.e. the properties which affect the efficacy of the direction's identifier. For example, we examined the variables' degradation rate. When upstream and downstream variables fluctuate with similar lifetimes, we find a greater probability of successfully inferring their edge direction. Furthermore, we have found a connection between the index of dispersion D and the ability to infer the direction of the interaction, where larger D yields a larger success ratio. Both degradation times and index of dispersion effect on the overall performance were previously demonstrated to be with similar conclusions as is given in⁴⁸. Moreover, some of the network topological features such as the node degree were also being inspected. We have found that a lower incoming degree of the node provides better results in the inference of the direction. That is in agreement with previously published results that show that increasing the incoming degree requires more data points to gain the same quality of inference^{29,65}.

To summarize, the method performs well where the joint distribution is well sampled. That is achieved when measuring sufficiently many data points such that they satisfactorily capture the required statistical features. Nevertheless, as mentioned, here we provide only a basic algorithm that proves the principle that the direction of interaction is encoded within the stationary information.

For future research, we suggest examining more sophisticated approaches, such as deep learning methods. Previously published techniques involve artificial neural networks to study node-to-node interactions^{74–76}, while others employ stacking methods^{77–79}. Supervised learning methods, such as regression-based approaches, are commonly used to infer the global structure of gene regulatory networks^{56–61}. These methods, however, require multidimensional data that goes beyond simple two-node observations. It is important to note that for these learning methods, the training data must closely resemble the observational data under scrutiny, a condition that can be challenging to fulfill in real-world situations. Moreover, these approaches typically require very large datasets and significant training resources, including time and GPUs, limitations that might be eased in the future with advances in technology. Yet, as noted in⁵⁵, there is currently no universal method that applies to all scenarios; there is no “one-size-fits-all” solution, and the inference method should be tailored to the system under investigation.

Methods

Background theory

The well-known Master equation for V variables yields at stationarity the global balance relation, which means that for every variable i

$$0 = \sum_k \langle r_{ik}(\vec{x} - \vec{d}_{ik}) | x_i - d_{ik} \rangle P(x_i - d_{ik}) - \langle r_{ik}(\vec{x}) | x_i \rangle P(x_i). \quad (2)$$

$\vec{d}_{ik} = \{0, \dots, d_{ik}, \dots, 0\}$ where the non-zero value is located in the i -th component. Each x_i denotes the quantity of units of type i . For instance, this could represent the mRNA or protein copy number indicative of gene expression levels, a metabolite level within a cell, or the population count of a given species.

Here, we emphasize that the system contains V variables, so the state of a particular sample is given by the V -dimensional array $\{x_1, x_2, \dots, x_V\}$. Assuming the system is sampled N times, the observed dataset contains $V \times N$ points in total.

The above relation is derived from a simple summation of the stationary Master equation over all other variables, and has been derived and discussed previously^{45–48}, see also derivation in SI. The angular brackets represent conditional means, i.e. $\langle r_{ik}(\vec{x}) | x_i \rangle \equiv \sum_{x_j, j \neq i} r_{ik}(\vec{x}) P(x_1, x_2, \dots, x_{i-1}, x_{i+1}, \dots, x_N | x_i)$ ⁸⁰.

Equation (2) can be written as

$$\sum_k \langle \tilde{r}_{ik}(x_i - d_{ik}, x_j) | x_i - d_{ik} \rangle P(x_i - d_{ik}) = \langle \tilde{r}_{ik}(x_i, x_j) | x_i \rangle P(x_i) \quad (3)$$

where $\tilde{r}_{ik}(x_i, x_j) \equiv \langle r_{ik}(\vec{x}) | x_i, x_j \rangle$. Our approach is based on *local* sensitivity of \tilde{r}_{ik} to x_j , which aims to quantify the dependence of the dynamic of x_i on the level of x_j . In particular, we calculate $J_{ij} \equiv \partial r_{ik} / \partial x_j$ next to the probable abundance point. The quantities of $|J_{ij}|$ and $|J_{ji}|$ are then utilized as a classification feature. The nature of this comparison between $|J_{ij}|$ and $|J_{ji}|$, and its role in classification would depend on the specific context or system. It is partially inspired by^{71,72} where both directions were quantified using regression and then assessed such that the more probable direction is chosen.

Models

For numerical demonstrations, we generate a random directed graph \hat{G} which holds the topology of the network. Mathematically the graph is described by the adjacency matrix with elements $G_{ij} = 1$ where the state of j affects the dynamics of node i , i.e. $j \rightarrow i$, and $G_{ij} = 0$ otherwise. We examined systems with Erdős Rényi random networks. In addition, we examine more biologically realist system which is based on a subset of an *E. coli* gene regulatory network and was provided in the DREAM challenge^{52–54}.

For illustration, we use a typical reaction rate model where a molecule reaction can be either a reproduction $d_i^+ = +1$ or degradation $d_i^- = -1$. We specified that the production rate of x_i , denoted as r_i^+ , depends on other units' states via the reaction network and that each molecule within the system degrades independently with a typical degradation rate. Hence, Eq. (2) can be written as

$$\langle r_i^+(\vec{x}) | x_i \rangle P(x_i) = \frac{1}{\tau_i} (x_i + 1) P(x_i + 1) \quad (4)$$

where $r_i^+(\vec{x})$ encodes all incoming edges (all units that affect the production rate of node i), and τ_i is the known lifetime of molecule i .

Michaelis–Menten regulatory network

The birth rate of molecule i is

$$r_i^+ = \sum_{j \neq i} G_{ij} \frac{\lambda x_j}{x_j + k} + 30 \cdot \delta \left(\sum_{j \neq i} G_{ij} \right) \quad (5)$$

where for our simulation we choose $V = 10$ nodes, 10 directed edges, $\lambda = 100$ and $k = 100$. These commonly described production rates in biochemical reaction networks, where λ gives the maximal rate and k refers to the concentration which gives half of the maximal rate^{28,81}. The degradation rates follow $r_i^- = x_i$ for all i .

We note that the Michaelis–Menten model describes the dynamics that involve enzymes. This rate function holds for a quasi-steady state with low enzyme concentrations^{28,81}. In the context of gene regulation through transcription factor (TF) binding to gene regulatory sequences, the application of a Michaelis–Menten rate type is considered valid due to the typical excess TFs over their chromosome-binding sites⁸². Still, when intermediate molecules are involved, or when describing dynamics of comparable concentrations, the Michaelis–Menten-rate law needs alternatives⁸³.

Coupled Goodwin oscillators

Each Goodwin oscillator is given by three interacting variables (x_i, y_i, z_i) which are coupled through the y variables. Their proliferation rates are given in the following:

$$\begin{aligned} r_{x,i}^+ &= \frac{\lambda_x}{(z_i/k)^n + 1} \\ r_{y,i}^+ &= \lambda_y x_i + \sum_{j \neq i} G_{ij} y_j \\ r_{z,i}^+ &= \lambda_z y_i \end{aligned} \quad (6)$$

and death rates are $r_i^- = x_i$ for all i . Here, the dynamic influences network within each oscillator is known, namely the relations within the triplet (x_i, y_i, z_i) are given, hence we aim to infer only the coupling network between these triplets G_{ij} . In the simulation presented in the main text, \hat{G} has 8 edges and 8 nodes—where each node is a triplet, such as we have 24 fluctuated variables in total.

Gene regulatory interactions in *E. coli*

As mentioned, we also examine a more biologically realist model which is based on a subset of an *E. coli* gene regulatory network and was provided in the DREAM challenge^{52–55}. There, the *in-silico* network inference challenge investigated how well gene networks can be deduced from simulated data. The network is derived as

subgraphs from the recognized *E. coli* and *S. cerevisiae* gene regulation networks⁸⁴. This means that the results presented are thus biologically realistic, i.e. aiming to capture a reasonable network, but not given from a real observation, and the gene indexes are thus arbitrary.

There, the variables x_i are described as continuous and possess small values. Here, for the discrete treatment of the model, we scaled the variables by $\tilde{x}_i = x_i/100$, thus our *E. coli* gene regulatory network is thus defined and simulated with the following rates:

$$\begin{aligned} r_1^+ &= \frac{k_{0,1}}{K_{M,1} + D} \\ r_2^+ &= \frac{k_{0,2}}{K_{M,2} + \tilde{x}_1} + \frac{k_{0,3}\tilde{x}_3}{K_{M,3} + \tilde{x}_3} + \frac{k_{0,4}\tilde{x}_7}{K_{M,4} + \tilde{x}_7} \\ r_3^+ &= \frac{k_{0,5}\tilde{x}_1}{K_{M,5} + \tilde{x}_1} + \frac{k_{0,6}}{K_{M,6} + \tilde{x}_2} + \frac{k_{0,7}\tilde{x}_5}{K_{M,7} + \tilde{x}_5} + \frac{k_{0,8}\tilde{x}_7}{K_{M,8} + \tilde{x}_7} \\ r_4^+ &= \frac{k_{0,9}}{K_{M,9} + \tilde{x}_1} \\ r_5^+ &= \frac{k_{0,10}}{K_{M,10} + \tilde{x}_2} \\ r_6^+ &= \frac{k_{0,11}U}{K_{M,11} + U} \\ r_7^+ &= \frac{k_{0,12}\tilde{x}_4}{K_{M,12} + \tilde{x}_4} + \frac{k_{0,13}\tilde{x}_6}{K_{M,13} + \tilde{x}_6} \end{aligned} \tag{7}$$

and for every i

$$r_i^- = \gamma_i \tilde{x}_i. \tag{8}$$

with parameters

Gene	Parameter values
N_1	$k_{0,1} = 0.0362, K_{M,1} = 0.1259, \gamma_1 = 0.4060,$
N_2	$k_{0,2} = 1.0106, K_{M,2} = 1.7937, k_{0,3} = 0.3550,$
	$K_{M,3} = 1.2069, k_{0,4} = 0.7472, K_{M,4} = 1.2858,$
	$\gamma_2 = 2.1362$
N_3	$k_{0,5} = 2.4007, K_{M,5} = 0.8218, k_{0,6} = 0.8511,$
	$K_{M,6} = 1.7099, k_{0,7} = 2.8247, K_{M,7} = 1.6656,$
	$k_{0,8} = 0.6081, K_{M,8} = 0.0202, \gamma_3 = 3.8740,$
N_4	$k_{0,9} = 0.0903, K_{M,9} = 0.069, \gamma_4 = 0.7256$
N_5	$k_{0,10} = 0.5264, K_{M,10} = 0.9600, \gamma_5 = 0.7466$
N_6	$k_{0,11} = 0.6541, K_{M,11} = 1.0891, \gamma_6 = 0.4525$
N_7	$k_{0,12} = 0.0090, K_{M,12} = 0.5191, k_{0,13} = 1.1236,$
	$K_{M,13} = 0.4986, \gamma_7 = 0.9473$

These parameters are taken from^{52–54}. The ‘external disturbance’ is modeled by uniform distribution, means $D, U \sim \text{Uniform}[0, 1]$.

Data availability

All data are included in the main article and its supplementary information files.

Received: 22 May 2024; Accepted: 9 January 2025
Published online: 23 January 2025

References

1. Strogatz, S. H. Exploring complex networks. *Nature* **410**, 268 (2001).
2. Bower, J. M. & Bolouri, H. *Computational Modeling of Genetic and Biochemical Networks* (MIT press, 2004).
3. Wilkinson, D. J. *Stochastic Modelling for Systems Biology* (CRC Press, 2011).
4. Mazza, C. & Benaïm, M. *Stochastic Dynamics for Systems Biology* (Chapman and Hall, 2016).
5. Corneliuș, S. P., Kath, W. L. & Motter, A. E. Realistic control of network dynamics. *Nat. Commun.* **4**, 1942 (2013).
6. Kolch, W., Halasz, M., Granovskaya, M. & Kholodenko, B. N. The dynamic control of signal transduction networks in cancer cells. *Nat. Rev. Cancer* **15**, 515 (2015).
7. Karrer, T. M. et al. A practical guide to methodological considerations in the controllability of structural brain networks. *J. Neural Eng.* **17**, 026031 (2020).
8. Rosato, A. et al. From correlation to causation: Analysis of metabolomics data using systems biology approaches. *Metabolomics* **14**, 1 (2018).

9. Laubach, Z. M., Murray, E. J., Hoke, K. L., Safran, R. J. & Perng, W. A biologist's guide to model selection and causal inference. *Proc. R. Soc. B* **288**, 20202815 (2021).
10. Loskot, P., Atitey, K. & Mihaylova, L. Comprehensive review of models and methods for inferences in bio-chemical reaction networks. *Front. Genet.* **10**, 549 (2019).
11. Kundu, S. A mathematically rigorous algorithm to define, compute and assess relevance of the probable dissociation constants in characterizing a biochemical network. *Sci. Rep.* **14**, 3507 (2024).
12. Kundu, S. ReDirection: An R-package to compute the probable dissociation constant for every reaction of a user-defined biochemical network. *Front. Mol. Biosci.* **10**, 1206502 (2023).
13. Sachs, K., Perez, O., Péter, D., Lauffenburger, D. A. & Nolan, G. P. Causal protein-signaling networks derived from multiparameter single-cell data. *Science* **308**, 523 (2005).
14. Su, E. Y., Spangler, A., Bian, Q., Kasamoto, J. Y. & Cahan, P. Reconstruction of dynamic regulatory networks reveals signaling-induced topology changes associated with germ layer specification. *Stem Cell Rep.* **17**, 427 (2022).
15. Guillén-Gosálbez, G., Miró, A., Alves, R., Sorribas, A. & Jiménez, L. Identification of regulatory structure and kinetic parameters of biochemical networks via mixed-integer dynamic optimization. *BMC Syst. Biol.* **7**, 1 (2013).
16. Michailidis, G. & d'Alché Buc, F. Autoregressive models for gene regulatory network inference: Sparsity, stability and causality issues. *Math. Biosci.* **246**, 326 (2013).
17. Schmidt, H., Cho, K.-H. & Jacobsen, E. W. Identification of small scale biochemical networks based on general type system perturbations. *FEBS J.* **272**, 2141 (2005).
18. Gong, H., Klinger, J., Damazyn, K., Li, X. & Huang, S. A novel procedure for statistical inference and verification of gene regulatory subnetwork. *BMC Bioinf.* **16**, 1 (2015).
19. Kim, J., Bates, D. G., Postlethwaite, I., Heslop-Harrison, P. & Cho, K.-H. Least-squares methods for identifying biochemical regulatory networks from noisy measurements. *BMC Bioinf.* **8**, 8 (2007).
20. Wang, A. & Pang, J. Iterative structural inference of directed graphs. *Adv. Neural. Inf. Process. Syst.* **35**, 8717 (2022).
21. Larvie, J. E. et al. Stable gene regulatory network modeling from steady-state data. *Bioengineering* **3**, 12 (2016).
22. Warne, D. J., Baker, R. E. & Simpson, M. J. Simulation and inference algorithms for stochastic biochemical reaction networks: From basic concepts to state-of-the-art. *J. R. Soc. Interface* **16**, 20180943 (2019).
23. Kirst, C., Timme, M. & Battaglia, D. Dynamic information routing in complex networks. *Nat. Commun.* **7**, 11061 (2016).
24. Timme, M. & Casadiego, J. Revealing networks from dynamics: An introduction. *J. Phys. A: Math. Theor.* **47**, 343001 (2014).
25. Herbach, U., Bonnaïffoux, A., Espinasse, T. & Gandrillon, O. Inferring gene regulatory networks from single-cell data: A mechanistic approach. *BMC Syst. Biol.* **11**, 1 (2017).
26. Yu, D., Righero, M. & Kocarev, L. Estimating topology of networks. *Phys. Rev. Lett.* **97**, 188701 (2006).
27. Timme, M. Revealing network connectivity from response dynamics. *Phys. Rev. Lett.* **98**, 224101 (2007).
28. Nelson, D. L., Lehninger, A. L. & Cox, M. M. *Lehninger Principles of Biochemistry* (Macmillan, 2008).
29. Nitzan, M., Casadiego, J. & Timme, M. Revealing physical interaction networks from statistics of collective dynamics. *Sci. Adv.* **3**, e1600396 (2017).
30. Tripathi, S., Kessler, D. A. & Levine, H. Minimal frustration underlies the usefulness of incomplete regulatory network models in biology. *Proc. Natl. Acad. Sci.* **120**, e2216109120 (2023).
31. Aghdam, R., Ganjali, M., Zhang, X. & Eslahchi, C. CN: A consensus algorithm for inferring gene regulatory networks using the SORDER algorithm and conditional mutual information test. *Mol. BioSyst.* **11**, 942 (2015).
32. Taniguchi, Y. et al. Quantifying *E. coli* proteome and transcriptome with single-molecule sensitivity in single cells. *Science* **329**, 533 (2010).
33. Friedman, N., Cai, L. & Xie, X. S. Linking stochastic dynamics to population distribution: An analytical framework of gene expression. *Phys. Rev. Lett.* **97**, 168302 (2006).
34. Gardner, T. S., Di Bernardo, D., Lorenz, D. & Collins, J. J. Inferring genetic networks and identifying compound mode of action via expression profiling. *Science* **301**, 102 (2003).
35. Gardiner, C. W. et al. *Handbook of Stochastic Methods* Vol. 3 (Springer, Berlin, 1985).
36. Ibe, O. *Markov Processes for Stochastic Modeling* (Newnes, 2013).
37. Pearl, J. Causal inference in statistics: An overview. *Stat. Surv.* **3**, 96. <https://doi.org/10.1214/09-SS057> (2009).
38. Peel, L., Peixoto, T. P. & De Domenico, M. Statistical inference links data and theory in network science. *Nat. Commun.* **13**, 6794 (2022).
39. Barzel, B. & Barabási, A.-L. Network link prediction by global silencing of indirect correlations. *Nat. Biotechnol.* **31**, 720 (2013).
40. Glymour, C., Zhang, K. & Spirtes, P. Review of causal discovery methods based on graphical models. *Front. Genet.* **10**, 524 (2019).
41. Paluš, M. & Vejmelka, M. Directionality of coupling from bivariate time series: How to avoid false causalities and missed connections. *Phys. Rev. E* **75**, 056211 (2007).
42. Eichler, M., Dahlhaus, R. & Sandkühler, J. Partial correlation analysis for the identification of synaptic connections. *Biol. Cybern.* **89**, 289 (2003).
43. Zhang, X., Zhao, J., Hao, J.-K., Zhao, X.-M. & Chen, L. Conditional mutual inclusive information enables accurate quantification of associations in gene regulatory networks. *Nucleic Acids Res.* **43**, e31 (2015).
44. Jacob, Y. et al. Dependency network analysis (DEPNA) reveals context related influence of brain network nodes. *Sci. Rep.* **6**, 27444 (2016).
45. Schnakenberg, J. Network theory of microscopic and macroscopic behavior of master equation systems. *Rev. Mod. Phys.* **48**, 571 (1976).
46. Haag, G., Weidlich, W. & Alber, P. Approximation methods for stationary solutions of discrete master equations. *Zeitschrift für Phys. B Condens. Matter.* **26**, 207 (1977).
47. Kelly, F. P. *Reversibility and Stochastic Networks* (Cambridge University Press, 2011).
48. Wittenstein, T., Leibovich, N. & Hilfinger, A. Quantifying biochemical reaction rates from static population variability within incompletely observed complex networks. *PLoS Comput. Biol.* **18**, e1010183 (2022).
49. Karlebach, G. & Shamir, R. Modelling and analysis of gene regulatory networks. *Nat. Rev. Mol. Cell Biol.* **9**, 770 (2008).
50. Goodwin, B. C. Oscillatory behavior in enzymatic control processes. *Adv. Enzyme Regul.* **3**, 425 (1965).
51. Ruoff, P., Vinsjevsk, M., Monnerjahn, C. & Rensing, L. reactThe Goodwin oscillator: On the importance of degradation ions in the circadian clock. *J. Biol. Rhythms* **14**, 469 (1999).
52. Schaffter, T., Marbach, D. & Floreano, D. GeneNetWeaver: In silico benchmark generation and performance profiling of network inference methods. *Bioinformatics* **27**, 2263 (2011).
53. Liu, F., Zhang, S.-W., Guo, W.-F., Wei, Z.-G. & Chen, L. Inference of gene regulatory network based on local Bayesian networks. *PLoS Comput. Biol.* **12**, e1005024 (2016).
54. Foo, M., Kim, J. & Bates, D. G. Modelling and control of gene regulatory networks for perturbation mitigation. *IEEE/ACM Trans. Comput. Biol. Bioinf.* **16**, 583 (2018).
55. Prill, R. J. et al. Towards a rigorous assessment of systems biology models: The DREAM3 challenges. *PLoS ONE* **5**, e9202 (2010).
56. Huynh-Thu, V. A., Irrthum, A., Wehenkel, L. & Geurts, P. Inferring regulatory networks from expression data using tree-based methods. *PLoS ONE* **5**, e12776 (2010).
57. Moerman, T. et al. GRNBoost2 and Arboreto: Efficient and scalable inference of gene regulatory networks. *Bioinformatics* **35**, 2159 (2019).

58. Van de Sande, B. et al. A scalable SCENIC workflow for single-cell gene regulatory network analysis. *Nat. Protoc.* **15**, 2247 (2020).
59. Pratapa, A., Jaliyal, A. P., Law, J. N., Bharadwaj, A. & Murali, T. Benchmarking algorithms for gene regulatory network inference from single-cell transcriptomic data. *Nat. Methods* **17**, 147 (2020).
60. Yuan, Q. & Duren, Z. Inferring gene regulatory networks from single-cell multiome data using atlas-scale external data. *Nat. Biotechnol.* [SPACE] <https://doi.org/10.1038/s41587-024-02182-7> (2024).
61. Osorio, D. et al. Population-level comparisons of gene regulatory networks modeled on high-throughput single-cell transcriptomics data. *Nat. Comput. Sci.* **4**, 237 (2024).
62. Zhou, X. & Cai, X. Inference of differential gene regulatory networks based on gene expression and genetic perturbation data. *Bioinformatics* **36**, 197 (2020).
63. Chaibub Neto, E., Ferrara, C. T., Attie, A. D. & Yandell, B. S. Inferring causal phenotype networks from segregating populations. *Genetics* **179**, 1089 (2008).
64. Rizvi, A. H. et al. Single-cell topological RNA-seq analysis reveals insights into cellular differentiation and development. *Nat. Biotechnol.* **35**, 551 (2017).
65. Casadiego, J., Nitzan, M., Hallerberg, S. & Timme, M. Model-free inference of direct network interactions from nonlinear collective dynamics. *Nat. Commun.* **8**, 2192 (2017).
66. Dablander, F. & Hinne, M. Node centrality measures are a poor substitute for causal inference. *Sci. Rep.* **9**, 6846 (2019).
67. Yu, D. & Parlitz, U. Driving a network to steady states reveals its cooperative architecture. *Europhys. Lett.* **81**, 48007 (2008).
68. Di Lauro, F., Croix, J.-C., Dashti, M., Berthouze, L. & Kiss, I. Network inference from population-level observation of epidemics. *Sci. Rep.* **10**, 18779 (2020).
69. Opgen-Rhein, R. & Strimmer, K. From correlation to causation networks: A simple approximate learning algorithm and its application to high-dimensional plant gene expression data. *BMC Syst. Biol.* **1**, 1 (2007).
70. Yuan, Y., Li, C.-T. & Windram, O. Directed partial correlation: Inferring large-scale gene regulatory network through induced topology disruptions. *PLoS ONE* **6**, e16835 (2011).
71. Marx, A. & Vreeken, J. Telling cause from effect using MDL-based local and global regression, in *2017 IEEE international conference on data mining (ICDM)* (IEEE, 2017) pp. 307–316.
72. Blöbaum, P., Janzing, D., Washio, T., Shimizu, S. & Schölkopf, B. Analysis of cause-effect inference by comparing regression errors. *PeerJ Comput. Sci.* **5**, e169 (2019).
73. Marx, A. & Vreeken, J. Telling cause from effect by local and global regression. *Knowl. Inf. Syst.* **60**, 1277 (2019).
74. Goudet, O., Kalainathan, D., Caillou, P., Guyon, I., Lopez-Paz, D. & Sebag, M. Learning functional causal models with generative neural networks, Explainable and interpretable models in computer vision and machine learning, 39 (2018).
75. Lee, D., Jayaraman, A. & Kwon, J. S. Development of a hybrid model for a partially known intracellular signaling pathway through correction term estimation and neural network modeling. *PLoS Comput. Biol.* **16**, e1008472 (2020).
76. Frauen, D., Imrie, F., Curth, A., Melnychuk, V., Feuerriegel, S. & van der Schaar, M. A Neural framework for generalized causal sensitivity analysis, arXiv preprint [arXiv:2311.16026](https://arxiv.org/abs/2311.16026) (2023).
77. Fonollosa, J. A. Conditional distribution variability measures for causality detection. *Cause Effect Pairs in Machine Learning*, 339 (2019).
78. Ghasemian, A., Hosseinmardi, H., Galstyan, A., Airolidi, E. M. & Clauset, A. Stacking models for nearly optimal link prediction in complex networks. *Proc. Natl. Acad. Sci.* **117**, 23393 (2020).
79. Varando, G., Catsis, S., Diaz, E. & Camps-Valls, G. Pairwise causal discovery with support measure machines. *Appl. Soft Comput.* **150**, 111030 (2024).
80. Edition, F., Papoulis, A. & Pillai, S. U. *Probability, Random Variables, and Stochastic Processes* (McGraw-Hill Europe, 2002).
81. Santillán, M. On the use of the Hill functions in mathematical models of gene regulatory networks. *Math. Model. Nat. Phen.* **3**, 85 (2008).
82. Ronen, M., Rosenberg, R., Shraiman, B. I. & Alon, U. Assigning numbers to the arrows: Parameterizing a gene regulation network by using accurate expression kinetics. *Proc. Natl. Acad. Sci.* **99**, 10555 (2002).
83. Kim, J. K. & Tyson, J. J. Misuse of the Michaelis–Menten rate law for protein interaction networks and its remedy. *PLoS Comput. Biol.* **16**, e1008258 (2020).
84. Marbach, D., Schaffter, T., Mattiussi, C. & Floreano, D. Generating realistic in silico gene networks for performance assessment of reverse engineering methods. *J. Comput. Biol.* **16**, 229 (2009).

Additional information

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1038/s41598-025-86332-0>.

Correspondence and requests for materials should be addressed to N.L.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2025