

# Identifiability and experimental design in perturbation studies

Torsten Gross<sup>1,2,3</sup> and Nils Blüthgen<sup>1,2,3</sup>

<sup>1</sup>Charité - Universitätsmedizin Berlin, Institut für Pathologie, Berlin, Germany,

<sup>2</sup>IRI Life Sciences, Humboldt University, Berlin, Germany

<sup>3</sup>Berlin Institute of Health, Berlin, Germany

## Abstract

**Motivation:** A common strategy to infer and quantify interactions between components of a biological system is to deduce them from the network’s response to targeted perturbations. Such perturbation experiments are often challenging and costly. Therefore, optimising the experimental design is essential to achieve a meaningful characterisation of biological networks. However, it remains difficult to predict which combination of perturbations allows to infer specific interaction strengths in a given network topology. Yet, such a description of identifiability is necessary to select perturbations that maximize the number of inferable parameters.

**Results:** We show analytically that the identifiability of network parameters can be determined by an intuitive maximum flow problem. Furthermore, we used the theory of matroids to describe identifiability relationships between sets of parameters in order to build identifiable effective network models. Collectively, these results allowed to devise strategies for an optimal design of the perturbation experiments. We benchmarked these strategies on a database of human pathways. Remarkably, full network identifiability was achieved with on average less than a third of the perturbations that are needed in a random experimental design. Moreover, we determined perturbation combinations that additionally decreased experimental effort compared to single-target perturbations. In summary, we provide a framework that allows to infer a maximal number of interaction strengths with a minimal number of perturbation experiments.

**Availability:** IdentiFlow is available at [github.com/GrossTor/IdentiFlow](https://github.com/GrossTor/IdentiFlow).

**Contact:** [nils.bluehgen@charite.de](mailto:nils.bluehgen@charite.de)

## Introduction

Rapid technological progress in experimental techniques allows to quantify a multitude of cellular components in ever increasing level of detail. Yet, to gain a mechanistic understanding of the cell requires to map out causal relations between molecular entities. As causality cannot be inferred from observational data alone (Pearl, 2009), a common approach is to observe the system’s response to a set of localised perturbations (Sachs *et al.*, 2005) and reconstruct a directed interaction network from such data. A recurring idea within the large body of according network inference methods (Marbach *et al.*, 2010) is to conceive the system as ordinary differential equations and describe edges in the directed network by the entries of an inferred Jacobian matrix (Gardner *et al.*, 2003; Bonneau *et al.*, 2006; Tegner *et al.*, 2003; Kholodenko, 2007; Bruggeman *et al.*, 2002; Timme, 2007). Such methods have been successfully applied to describe various types of regulatory networks in different organisms (Ciofani *et al.*, 2012; Arrieta-Ortiz *et al.*, 2015; Lorenz *et al.*, 2009; Klinger *et al.*, 2013; Brandt *et al.*, 2019). They are continuously improved, e.g. to reduce the effect of noise, incorporate heterogeneous data sets, or allow for the analysis of single cell data (Greenfield *et al.*, 2013; Santra *et al.*, 2018;

Klinger and Blüthgen, 2018; Santra *et al.*, 2013; Kang *et al.*, 2015; Dorel *et al.*, 2018) and have thus become a standard research tool. Nevertheless, identifiability (Hengl *et al.*, 2007; Godfrey and DiStefano, 1985) of the inferred network parameters within a specific perturbation setup has not yet been rigorously analysed, even though a limited number of practically feasible perturbations renders many systems underdetermined (De Smet and Marchal, 2010; Meinhäuser *et al.*, 2016; Bonneau *et al.*, 2006). Some inference methods do apply different heuristics, such as network sparsity, to justify parameter regularisation (Gardner *et al.*, 2003; Bonneau *et al.*, 2006; Tegner *et al.*, 2003), or numerically analyse identifiability through an exploration of the parameter space using a profile likelihood approach (Raue *et al.*, 2009). Yet, neither approach provides a structural understanding on how parameter identifiability relates to network topology and the targets of the perturbations. However, such structural understanding is required to systematically define identifiable effective network models and to optimize the sequence of applied perturbations. The latter is of particular interest because perturbation experiments are often costly and laborious, which demands to determine the minimal set of perturbations that reveals a maximal number of network parameters. To address these challenges, this work derives analytical results that explain the identifiability of network parameters in terms of simple network properties which allow to optimize the experimental design.

## Methods

We consider a network of  $n$  interacting nodes whose abundances,  $\mathbf{x}$ , evolve in time according to a set of (unknown) differential equations

$$\dot{\mathbf{x}} = \mathbf{f}(\mathbf{x}, \mathbf{p}). \quad (1)$$

The network can be experimentally manipulated by  $p$  different types of perturbations, each represented by one of the  $p$  entries of parameter vector  $\mathbf{p}$ . We only consider binary perturbations that can either be switched on or off. Without loss of generality, we define  $\mathbf{f}(\mathbf{x}, \mathbf{p})$  such that the  $k$ -th type of perturbation changes parameter  $p_k$  from its unperturbed state  $p_k = 0$  to a perturbed state  $p_k = 1$ .

The main assumption is that after a perturbation the observed system relaxes into stable steady state,  $\varphi(\mathbf{p})$ , of Equation 1. Stability arises when the real parts of the eigenvalue of the  $n \times n$  Jacobian matrix,  $J_{ij}(\mathbf{x}, \mathbf{p}) = \partial f_i(\mathbf{x}, \mathbf{p}) / \partial x_j$ , evaluated at these fixed points,  $\mathbf{x} = \varphi(\mathbf{p})$ , are all negative within the experimentally accessible perturbation space (no bifurcation points). This implies that  $J(\varphi(\mathbf{p}), \mathbf{p})$  is invertible, for which case the implicit function theorem states that  $\varphi(\mathbf{p})$  is unique and continuously differentiable, and

$$\frac{\partial \varphi_k}{\partial p_l} = - [J^{-1} S]_{kl}, \quad (2)$$

where  $n \times p$  Sensitivity matrix entry,  $S_{ij} = \partial f_i(\mathbf{x}, \mathbf{p}) / \partial p_j$ , quantifies the effect of the  $j$ -th perturbation type on node  $i$ . Dropping functions’ arguments is shorthand for the evaluation at the unperturbed state,  $\mathbf{x} = \varphi(\mathbf{0})$  and  $\mathbf{p} = \mathbf{0}$ .

## A linear response approximation

A perturbation experiment consists of  $q$  perturbations, each of which involves a single or a combination of perturbation types, represented by binary vector  $\mathbf{p}$ , which forms the columns of the  $p \times q$  design matrix  $P$ . The steady states after each perturbation,  $\varphi(\mathbf{p})$ , are measured and their differences to the unperturbed steady state form the columns of the  $n \times q$  global response matrix  $R$ . Assuming that perturbations are sufficiently mild, the steady state function becomes nearly linear within the relevant parameter domain,

$$\varphi_k(\mathbf{p}) - \varphi_k(\mathbf{0}) \approx \sum_{l=1}^p \frac{\partial \varphi_k}{\partial p_l} p_l. \quad (3)$$

Replacing the partial derivative with the help of Equation 2 and writing the equation for all  $q$  perturbations yields

$$R \approx -J^{-1} S P. \quad (4)$$

This equation relates the known experimental design matrix,  $P$ , and the measured global responses,  $R$ , to quantities that we wish to infer: the nodes' interaction strengths,  $J$ , and their sensitivity to perturbations,  $S$ .

A dynamic system defined by rates  $\dot{\mathbf{f}}(\mathbf{x}, \mathbf{p}) = W \mathbf{f}(\mathbf{x}, \mathbf{p})$ , with any full rank  $n \times n$  matrix  $W$ , has the same steady states but different Jacobian and sensitivity matrices, namely  $WJ$  and  $WS$ , as the original system, defined by Equation 1. It is thus impossible to uniquely infer  $J$  or  $S$  from observations of the global response alone, and prior knowledge in matrices  $J$  and  $S$  is required to further constrain the problem. In the following, we assume that prior knowledge exists about the network topology, i.e. about zero entries in  $J$ , as they correspond to non-existent edges. Likewise, we assume that the targets of the different types of perturbations are known, which implies known zero entries in  $S$  for non-targeted nodes. In line with prior studies (Kholodenko, 2007), we also fix the diagonal of the Jacobian matrix

$$J_{ii} = -1.$$

Thus, for the  $i$ -th row of  $J$  we can define index lists  $\bar{\mu}_i$  and  $\hat{\mu}_i$  to identify its known and unknown entries. The first indicates missing edges or the self loop and the second edges going into node  $i$ . These lists have  $|\bar{\mu}_i|$  and  $|\hat{\mu}_i|$  entries, respectively, with

$$|\bar{\mu}_i| + |\hat{\mu}_i| = n. \quad (5)$$

Analogously, for the  $i$ -th row of  $S$  we define index lists  $\bar{\nu}_i$  and  $\hat{\nu}_i$ , with

$$|\bar{\nu}_i| + |\hat{\nu}_i| = p, \quad (6)$$

to report its unknown and known entries. These describe the perturbations that do not target or respectively target node  $i$ .

We show in Supplementary Material S1 that Equation 4 can be repartitioned to obtain a system of linear equations for each row in  $J$  and  $S$ , exclusively in the

$$\mathbf{u}_i = |\hat{\mu}_i| + |\hat{\nu}_i|$$

unknown parameters, which we collect in vector  $\mathbf{x}_i$ . Thus, there is a  $\mathbf{u}_i \times \mathbf{d}_i$  matrix  $V_i$ , such that

$$\mathbf{x}_i = V_i \mathbf{w} + \tilde{\mathbf{x}}_i, \quad \forall \mathbf{w} \in \mathbb{R}^{\mathbf{d}_i}, \quad (7)$$

where  $\tilde{\mathbf{x}}_i$  is some specific solution to the equation system. We further show in Supplementary Material S1 that  $V_i$  is a basis of the kernel of

$$\Psi_i = \begin{bmatrix} \hat{S}_i \hat{J}_i^{-1} & I_{|\hat{\nu}_i|} \\ \bar{S}_i \hat{J}_i^{-1} & 0_{|\bar{\nu}_i|, |\hat{\nu}_i|} \end{bmatrix}, \quad (8)$$

where  $I_{|\hat{\nu}_i|}$  and  $0_{|\bar{\nu}_i|, |\hat{\nu}_i|}$  are the identity and zero matrix of annotated dimensionality. The  $n \times |\hat{\mu}_i|$  matrix  $\hat{J}_i^{-1}$  consists of the

columns of  $(J^{-1})^T$  that are selected by indices in  $\hat{\mu}_i$ . Finally,  $|\bar{\nu}_i| \times n$  matrix  $\bar{S}_i$  and  $\hat{\nu}_i \times n$  matrix  $\hat{S}_i$  shall be formed by taking rows of  $S^T$  according to indices in  $\bar{\nu}_i$  and  $\hat{\nu}_i$ . These matrix partitionings are demonstrated for a toy example in Supplementary Figure S1. Furthermore, in Supplementary Material S1 we derive the following expression for the solution space dimensionality

$$\mathbf{d}_i = |\hat{\mu}_i| - \text{rank}(\bar{S}_i \hat{J}_i^{-1}).$$

## Identifiability conditions

The system is underdetermined when  $\mathbf{d}_i > 0$ . But independent of  $\mathbf{d}_i$ , a parameter is identifiable if the solution space is orthogonal to its according axis direction. This idea can be expressed as algebraic identifiability conditions. Accordingly, we show in Supplementary Material S1 that the unknown interaction strength  $J_{i\hat{\mu}_i j}$  is identifiable if and only if

$$1 + \text{rank}(\bar{S}_i \hat{J}_{i\setminus j}^{-1}) = \text{rank}(\bar{S}_i \hat{J}_i^{-1}), \quad (9)$$

where  $\hat{J}_{i\setminus j}^{-1}$  is matrix  $\hat{J}_i^{-1}$  with the  $j$ -th column removed. Furthermore, the unknown sensitivity  $S_{i\hat{\nu}_i j}$  is identifiable if and only if

$$\text{rank} \left( \begin{bmatrix} \bar{S}_i \\ \hat{S}_i^j \end{bmatrix} \hat{J}_i^{-1} \right) = \text{rank}(\bar{S}_i \hat{J}_i^{-1}), \quad (10)$$

where  $\hat{S}_i^j$  denotes the  $j$ -th row of matrix  $\hat{S}_i$ . However, the ranks depend on the unknown network parameters themselves and can thus not be directly computed. Yet, we can show how a reasonable assumption makes this possible and allows to express the identifiability conditions as an intuitive maximum flow problem.

First, we rewrite the identity  $J^{-1} J = I_n$  as

$$[J^{-1}]_{kl} = \sum_{m \neq l} [J^{-1}]_{km} [J]_{ml} - \delta_{kl},$$

with  $\delta_{kl}$  being the Kronecker delta (recall that  $J_{ll} = -1$ ). We can view this equation as a recurrence relation and repeatedly replace the  $[J^{-1}]_{km}$  terms in the sum. The sum contains non-vanishing terms for each edge that leaves node  $l$ . Therefore, each replacement leads to the next downstream node, so that eventually one arrives at

$$[J^{-1}]_{kl} = l \rightarrow k [J^{-1}]_{kk}, \quad \text{with} \\ l \rightarrow k = \sum_{\omega \in \Omega_{l \rightarrow k}} \prod_{m=1}^{|\omega|-1} [J]_{\omega_{m+1} \omega_m},$$

where the set  $\Omega_{l \rightarrow k}$  contains elements,  $\omega$ , for every path from node  $l$  to node  $k$ , each of which lists the nodes along that path. Strictly speaking, these elements are walks rather than paths because some nodes will appear multiple times if loops exist between  $l$  and  $k$ . In fact, with loops,  $\Omega_{l \rightarrow k}$  contains an infinite number of walks of unbounded lengths. But as the real part of all eigenvalues of  $J$  are assumed negative, the associated products of interaction strengths converges to zero with increasing walk length.

To simplify our notation, we want to expand the network by considering perturbations  $\bar{\nu}_i$  as additional nodes, each with edges that are directed towards that perturbation's targets. Furthermore, letting the interaction strength associated with these new edges be given by the appropriate entries in  $S$  we can rewrite the matrix product

$$[\bar{S}_i \hat{J}_i^{-1}]_{kl} = \bar{\nu}_{ik} \rightarrow \hat{\mu}_{il} [J^{-1}]_{\hat{\mu}_{il} \hat{\mu}_{il}}$$

where  $\hat{\mu}_{il}$  and  $\bar{\nu}_{il}$  denote the  $l$ -th entry in  $\hat{\mu}_i$  and  $\bar{\nu}_i$ , respectively. As every finite-dimensional matrix has a rank decomposition, we can further write

$$\bar{S}_i \hat{J}_i^{-1} = \Upsilon_i Y_i, \quad (11)$$

where  $|\bar{\nu}_i| \times \text{rank}(\bar{S}_i \hat{J}_i^{-1})$  matrix  $\Upsilon_i$  and  $\text{rank}(\bar{S}_i \hat{J}_i^{-1}) \times |\hat{\mu}_i|$  matrix  $Y_i$  have full rank. Finding such a decomposition therefore reveals the rank of  $\bar{S}_i \hat{J}_i^{-1}$ . To this end, we propose

$$[\Upsilon_i]_{kn} = \bar{\nu}_{ik} \mapsto y_{in}, \text{ and } [Y_i]_{nl} = y_{in} \mapsto \hat{\mu}_{il} [J^{-1}]_{\hat{\mu}_{il} \hat{\mu}_{il}},$$

where  $y_{in}$  denotes the  $n$ -th component of a certain list of nodes  $\mathbf{y}_i$ . In order for Equation 11 to hold, it must be possible to split each path from any perturbation  $\bar{\nu}_{il}$  to any node  $\hat{\mu}_{il}$  into a section that leads from the perturbation to a node in  $\mathbf{y}_i$  and a subsequent section that leads from this node to  $\hat{\mu}_{il}$ . For an extended graph that includes an additional source node, with outgoing edges to each perturbation in  $\bar{\nu}_i$ , and an additional sink node, with incoming edges from all nodes in  $\hat{\mu}_i$  (see Figure 1B),  $\mathbf{y}_i$  thus constitutes a vertex cut whose removal disconnects the graph and separates the source and the sink node into distinct connected components. Next, we want to show that if  $\mathbf{y}_i$  is a minimum vertex cut, the rank of  $\bar{S}_i \hat{J}_i^{-1}$  equals the size of  $\mathbf{y}_i$ . Because Equation 11 is a rank decomposition this is equivalent to showing that the according matrices  $\Upsilon_i$  and  $Y_i$  have full rank. To do so, we apply Menger's theorem (Menger, 1927), which states that the minimal size of  $\mathbf{y}_i$  equals the maximum number of vertex-disjoint paths from the source to the sink node. This also implies that each of these vertex-disjoint paths goes through a different node of the vertex cut  $\mathbf{y}_i$ . Recall that entries in  $\Upsilon_i$  constitute sums over paths from perturbation to vertex cut nodes, so that we could write

$$\Upsilon_i = \bar{\Upsilon}_i + \hat{\Upsilon}_i,$$

where  $\bar{\Upsilon}_i$  only contains the vertex-disjoint paths and  $\hat{\Upsilon}_i$  the sums over the remaining paths. As each of these vertex disjoint paths ends in a different vertex cut node, any column in  $\bar{\Upsilon}_i$  can contain no more than a single non-zero entry. Furthermore, as a consequence of Menger's theorem there are exactly  $|\mathbf{y}_i|$  non-zero columns. Because these paths are indeed vertex disjoint also no row in  $\bar{\Upsilon}_i$  has more than a single non-zero entry. Thus, the non-zero columns are independent, showing that  $\bar{\Upsilon}_i$  has full rank. The crucial assumption we want to make now is that the values of the interaction strengths lie outside a specific algebraic variety, which would render  $\bar{\Upsilon}_i + \hat{\Upsilon}_i$  rank deficient. This would for example be the case if for a given vertex disjoint path there also is an alternative path whose associated product of interaction strengths has the same magnitude as that of the vertex disjoint path but opposite sign, making their sum vanish. This effect corresponds to a perfectly self-compensating perturbation. Most biological networks however cannot fine-tune their interactions to such a degree that they could achieve perfect self-compensation, which justifies this non-cancellation assumption.

The same line of reasoning will demonstrate a full rank for matrix  $Y_i$  as well, which implies that indeed

$$\text{rank}(\bar{S}_i \hat{J}_i^{-1}) = |\mathbf{y}_i|, \quad (12)$$

where  $\mathbf{y}_i$  is a minimum vertex cut between source and sink node. This equation has the crucial benefit that  $|\mathbf{y}_i|$  does not depend on any unknown parameters and can be computed as the maximum flow from source to sink node with all nodes having unit capacity (Ahuja *et al.*, 1993), as detailed in Figure 1B. This maximum flow problem can be solved in only  $\mathcal{O}(n^{1/2} \sum_i^n (|\hat{\mu}_i| + |\hat{\nu}_i|))$ , as shown in Theorem 6.3 in (Even and Even, 2012). More importantly though, it allows to express the algebraic identifiability conditions 9 and 10 in terms

of network properties, providing an intuitive relationship between network topology, perturbation targets and identifiability. Specifically,  $J_{i\hat{\mu}_{ij}}$  is identifiable if and only if the removal of the edge from node  $\hat{\mu}_{ij}$  to the sink node reduces the maximum flow of the network, see Figure 1C, and  $S_{i\hat{\nu}_{ij}}$  is identifiable if the maximum flow does not increase when an additional edge connects the source node with perturbation node  $\hat{\nu}_{ij}$ , see Figure 1D.

## Identifiability relationships

Often, network inference is an underdetermined problem (De Smet and Marchal, 2010; Gross *et al.*, 2019). Thus, to achieve identifiable effective network models, certain parameters have to be set to constant values, such that the remaining parameters become uniquely determinable. This requires an understanding of the identifiability relationships between parameters, i.e. we need to know which parameter becomes identifiable when other parameters are fixed. Supplementary Equation 16 formally relates these relationships to the ranks of certain linear subspaces of the range of  $V_i^T$  as defined in Equation 7. It shows that for each network node there is a set of parameters amongst which identifiability relationships can exist. Such a set contains those interaction strengths that quantify the edges, which target the associated node, and the associated node's sensitivities to perturbations. Furthermore, we show in Supplementary Material S2 that the identifiability relationships of such parameter groups can be described as a matroid (Whitney, 1935). Matroids can be defined in terms of their circuits. Here, a circuit is a set of parameters with the property that any of its parameters becomes identifiable after fixing all of the others. Therefore, circuits describe all minimal parameter subsets that could be fixed to obtain an identifiable network.

We enumerated the set of circuits with an incremental polynomial-time algorithm (Boros *et al.*, 2003). This algorithm requires an independence oracle that indicates linear dependence of subsets of columns of  $V_i^T$ . Supplementary Material S2 shows that we can construct such an oracle by considering linear dependence within the dual matroid, which amounts to determining

$$\text{rank} \left( \begin{bmatrix} \bar{\mathcal{P}}_2^T \bar{S}_i \\ \bar{S}_i \end{bmatrix} \hat{J}_i^{-1} \mathcal{P}_1 \right).$$

Matrices  $\mathcal{P}_2$  and  $\mathcal{P}_1$  are truncated identity matrices defined in Supplementary Equations 16 and 17. Yet, the crucial point of this expression is that it has the same form as the left hand side of Equation 12. We can therefore conveniently determine it by solving a simple maximum-flow problem.

Supplementary Material S2 shows how to transform the circuits into cyclic flats. These provide a more convenient representation of the identifiability relationships, which we clarify at an example in the next section. Finally, certain scenarios constrain the choice of fixable parameters, for example when quantifying multiple isogenic cell lines (Bosdriesz *et al.*, 2018). Supplementary Material S2 describes a greedy algorithm that takes such preferences into consideration.

## Experimental design strategies

We assume that we are given a set of  $p$  perturbations, each of which targets a different subset of nodes. From these we want to select one or multiple perturbation sequences, according to a certain strategy,  $s$ . By means of our understanding of identifiability, we can determine  $\xi_i$ , the number of identifiable edges after having performed the first  $i$  perturbations in such a sequence. Our goal is to find a strategy for which this number of identifiable edges increases fastest. Thus, as a measure of optimality of  $s$ , we can define an identifiability area under the

curve

$$\frac{1}{p} \sum_{i=1}^p \frac{\langle \xi_i \rangle_s}{\xi}, \quad (13)$$

where  $\langle \xi_i \rangle_s$  denotes the average of  $\xi_i$  over the ensemble of  $s$ -associated sequences, and  $\xi$  is the number of edges in the network. For any network and set of perturbations this score ranges between zero and one.

An obvious approach to design an optimal computational strategy is to simply consider the entire set of possible sequences and to select the ones that maximize the identifiability area under the curve. This is what we refer to as the *exhaustive* strategy. Clearly, it quickly becomes computationally intractable when the set of perturbations becomes large.

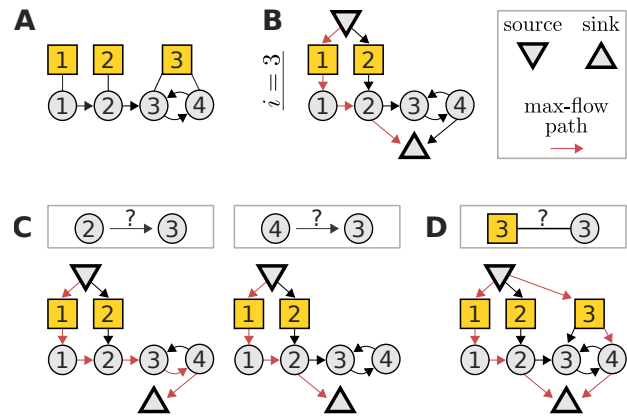
This is why we also propose strategies that build up the perturbation sequence in a stepwise manner and are therefore computationally efficient. At each step, the next perturbation is selected from a small set of perturbations that are deemed optimal according to a chosen strategy. Thus, identifying perturbation sequences in such a greedy optimisation approach amounts to performing a graph search over perturbation sets. We thus implemented a depth-first search to enumerate all strategy-conforming sequences. This sequence set might still be excessively large, when the number of possible perturbations is big. Therefore, we also implemented the option to randomly sample a fixed number of conforming sequences. A dynamic programming approach that avoids solving the same maximum-flow problem multiple times optimizes the search performance. Amongst the set of conforming strategies we eventually select the ones with the best performance according to Equation 13 (also see Figure 3A).

We suggest three different strategies to determine the set of the best next perturbations. The *single-target* approach selects those perturbations that minimize the overall dimensionality of the solution space,  $\sum_i d_i$ , amongst the perturbations that maximize the number of identifiable edges. An alternative idea that does not make use of solution space considerations, is to assume that perturbations are the most informative when they cause a response at a maximal number of nodes. Accordingly, the *naïve* strategy chooses perturbations that reach the largest number of nodes per perturbation target. Finally, we also want to allow for a combination of perturbations in a single experiment. Thus, we propose a *multi-target* strategy, which is similar to the single-target approach, except that it not only considers a single but any combination of perturbations. However, the entire power set of perturbation combinations might be too large to consider in practice. Therefore, we implemented an option for the multi-target strategy to build up combinations in a step-wise manner, where additional perturbations are only added to a combination if the enlarged combination increases performance. Additional details are provided in the documentation of the *IdentiFlow* package.

## Results

### Identifiability and identifiability relationships

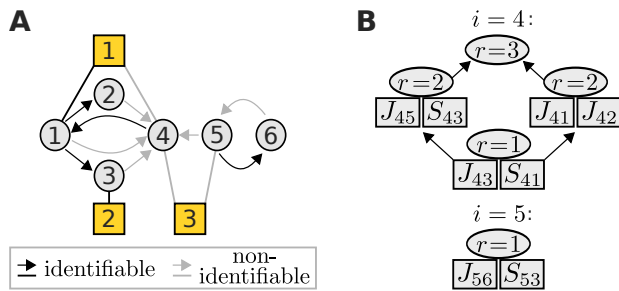
Perturbation experiments are frequently used to infer and quantify interactions in biological networks. But whether a given network edge can indeed be uniquely quantified depends on the specific targets of the perturbations and the topology of the network. Yet in order to build interpretable network models and guide experimental design, we need to elucidate this identifiability status of the network parameters. Here, we view a biological system as a weighted directed network, and assume that perturbations are sufficiently mild to cause a linear steady state response. This allows to relate the interaction strengths between nodes (i.e. the entries in the Jacobian matrix  $J$ ) and



**Figure 1:** A maximum flow problem determines the identifiability of interaction strengths and perturbation sensitivities when reconstructing a network from perturbation data. (A) Example network with three perturbations (yellow squares) to illustrate the algorithm. (B) The corresponding flow network to determine the identifiability of the edges into node 3 and the sensitivity of node 3 to perturbations. A path carrying the maximal flow of one is denoted in red (note that it is not unique). (C) The interaction strength between a given node and node 3 is identifiable if and only if the maximum flow is reduced after removing that node's edge to the sink node. In this example, there are alternative max-flow paths that re-establish a unit-flow after removal of the according edges. Thus, the respective interaction strengths are non-identifiable. (D) Similarly, the sensitivity of node 3 to perturbation 3 is identifiable, if and only if the depicted extension of the flow network does not increase the maximum flow. In this example, the maximum flow is increased by one, again revealing non-identifiability. Note that such flow representations provide an intuitive understanding on how alterations in the network or perturbation setting affect identifiability. For example, it is obvious that if the toy model would not contain an edge from node 3 to 4, the edge from 2 to 3 would become identifiable.

the sensitivity to perturbations (i.e. the entries in the Sensitivity matrix  $S$ ) to the measured responses (Equation 4), an approach that is widely known as Modular Response Analysis (Kholodenko, 2007). We derived analytical identifiability conditions (Equations 9 and 10) that describe whether this relation allows to uniquely determine the network parameters for the given network topology and the experimental setting. However, these conditions can not be directly evaluated, as they depend on the (unknown) network parameters themselves. But instead, they can be reformulated as intuitive maximum flow problems, if one disregards singular conditions of self-cancelling perturbations.

The derivation and details are given in the Methods section but briefly, to determine the identifiability of either the interaction strength from node  $j$  to node  $i$ , or the sensitivity of node  $i$  to perturbation  $p$ , the following flow network is considered: The original network is extended by (i) adding a node for each perturbation that does not target node  $i$  and connecting it to the respective perturbation's target(s), (ii) adding a "source" node that connects to all those perturbation nodes, and (iii) having all nodes that target node  $i$  connect to an additional "sink" node, see Figure 1B. Furthermore, all nodes (except source and sink) and all edges have a flow capacity of one. To reveal identifiability, we need to determine the network's maximum flow from source to sink. This is a classic problem in computer science, which we solve using the Edmonds-Karp algorithm (Dinic, 1970; Edmonds and Karp, 1972) as implemented in the Networkx package (Hagberg et al., 2008). Then, the interaction strength from node  $j$  to node  $i$  is identifiable if and only if the removal of the edge from node  $j$  to the sink node reduces the maximum flow, see Figure 1C. Similarly, node  $i$ 's sensitivity to perturbation  $p$  is identifiable if and only



**Figure 2:** (A) An example network with three perturbations (yellow squares), where nodes 4 and 5 are associated with non-identifiable parameters (grey). (B) Their identifiability relationships are represented by the lattices of cyclic flats of rank  $r$ . Each cyclic flat consists of the annotated elements in addition to elements from its preceding cyclic flats. All parameters of a cyclic flat with rank  $r$  become identifiable if at least  $r$  independent flat parameters are fixed.

if the maximum flow does not increase after linking the source to an additional node that is in turn connected to all targets of perturbation  $p$ , see Figure 1D.

Often, experimental settings do not allow determining all unknown parameters (De Smet and Marchal, 2010; Gross *et al.*, 2019). Nevertheless, they constrain the solution space such that after fixing one or multiple parameters, others become identifiable. We found that such identifiability relationships can be described by matroids, which are combinatorial structures that generalize the notion of linear dependence (see Methods). This is demonstrated for an example perturbation experiment on the network displayed in Figure 2A.

Each node is associated with a set of parameters amongst which identifiability relationships can exist. Such a set contains those interaction strengths, which quantify the edges that target the associated node, and that node’s sensitivities to perturbations. Here, nodes 4 and 5 are associated with sets of non-identifiable parameters. For example for node 5, these are  $J_{56}$  and  $S_{53}$ . We represent the matroid for such a parameter set as a hierarchy (lattice) of cyclic flats, as show in Figure 2B. A cyclic flat is a set of parameters with an associated rank  $r$ . It has the property that all of its parameters become identifiable, if amongst them at least  $r$  independent parameters are fixed. Parameters are independent if none of them becomes identifiable after fixing the others. For node 5, parameters  $J_{56}$  and  $S_{53}$  only form a single cyclic flat with  $r = 1$ , and thus fixing either one parameter makes the other identifiable. The identifiability relationships among the six parameters associated with node 4 are more complex. For example,  $J_{43}$  and  $S_{41}$  form a cyclic flat with  $r = 1$  and thus fixing one, fixes the other. Yet together with  $J_{45}$  and  $S_{43}$ , they form a cyclic flat with  $r = 2$ , thus fixing e.g.  $S_{41}$  and  $S_{43}$  will allow unique determination of  $J_{43}$  and  $J_{45}$ . In contrast, fixing  $J_{43}$  and  $S_{41}$  does not render any other parameter identifiable because they are not independent. This illustrates how the matroid description allows to generate effective models, i.e. models where a minimum number of parameters has to be set to fixed values to allow for a unique estimation of all other parameters. Importantly, the lattice of cyclic flats can be derived without specifying unknown parameters by solving a sequence of maximum flow problems (see Methods).

Collectively, our results provide a concise framework to algorithmically determine identifiability of network parameters and to construct identifiable effective networks when the experimental setting does not suffice to uniquely determine the original network structure.

## Experimental design

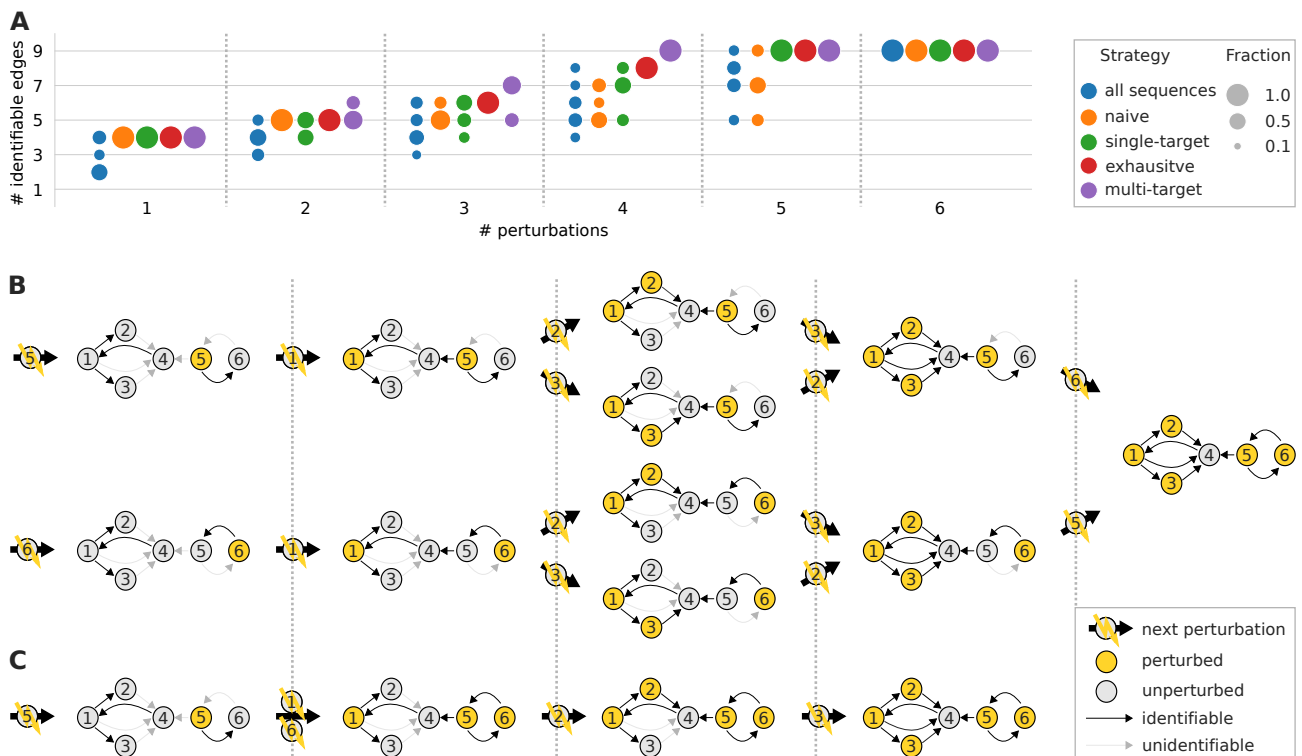
Next, we applied our identifiability analysis to optimize experimental design, i.e. to minimize the number of perturbation experiments that is required to uniquely determine a network’s interaction strengths. For this, we designed the following strategies to determine an optimal sequence from a set of available perturbations: The *exhaustive* strategy considers all possible sequences and selects the best performing amongst them. As this approach entails a prohibitive computational effort for larger networks, we also designed approaches that select perturbation sequences in a step-wise manner: First, the *single-target* strategy chooses the next perturbation such that it increases the number of identifiable edges most. Second, the *multi-target* strategy is similar to the single-target strategy except that it not only considers a single but any combination of perturbations. We then compared these strategies to a *naive* strategy that does not use our identifiability analysis. Rather, it chooses perturbations first that cause a response at a larger number of nodes (see Methods for details).

We first scrutinised the proposed experimental design strategies on the example network shown in Figure 2. We defined six different types of perturbations, each of which targets a (different) single node, or any combination of such for the multi-target strategy. Figure 3A shows how the number of identifiable edges increases with the number of performed perturbations for each strategy. A single strategy might propose multiple sequences, as described in Methods. Accordingly, Figure 3A shows the performance distribution over all such conforming sequences. In practice, we would only select the best performing sequence amongst them. Nevertheless, the depicted distributions are informative because for larger networks we can no longer enumerate all but only a (random) subset of conforming sequences.

When comparing the methods, we found that on average all strategies outperform randomly chosen sequences (“all sequences” distribution). Moreover, the “naive” strategy that did not use our framework mostly required all six perturbations to fully identify all parameters, whereas the single-target and exhaustive strategies only needed five, and the multi-target strategy only four perturbations. Figure 3B and Figure 3C display all perturbation sequences conforming to the single-target strategy, and one sequence conforming to the multi-target strategy respectively, and illustrate which network edge becomes identifiable at which step in the sequence.

To systematically analyse if and how our approach improves experimental design, we benchmarked the different strategies on all 267 nontrivial human KEGG (Kanehisa *et al.*, 2019) pathways, ranging from 5 to 120 nodes (see Supplementary Material S3 for details). Again, we assumed that perturbations can target (all) single nodes. For each network, we sampled 10 conforming sequences per strategy and compared against the performance of 10 randomly chosen sequences. As a performance measure, we considered the number of identified parameters as a function of the number of perturbations and computed a normalised area under the curve, as defined in Equation 13. Figure 4A shows the result of this benchmark, and confirms the trend already observed for the example in Figure 3A: Compared to choosing perturbations randomly, the naive strategy improved identifiability. Performance was further increased when we applied our single-target strategy, yet the multi-target strategy clearly performed best. An exhaustive enumeration of all sequences is not feasible for all KEGG networks. However, we found for a subset of small networks that there is no performance difference between the exhaustive and the single-target strategy, as shown in Supplementary Figure S4A.

Furthermore, we determined the average number of perturbations that is required for full network identifiability and computed the fraction between a given strategy and the random



**Figure 3:** (A) The same network topology as in Figure 2 was subjected to a set of perturbations that target each node individually. Shown are distributions of numbers of identifiable edges for different experimental design strategies and an increasing number of perturbations. (B) All optimal (exhaustive) single-target perturbation sequences and (C) one multi-target sequence.

sequences, see Figure 4B. We found that the average number of required perturbations can be reduced to less than one third or even less than a quarter, when using a single-target or multi-target strategy, respectively.

We next investigated which network properties led to a performance increase using our strategies. Intuitively, perturbations might be more informative if their response propagates to large parts of the network. We therefore hypothesised that a careful experimental design is particularly beneficial when networks contain many isolated nodes with little connection to the rest of the network because, in contrast to a random choice, a good strategy could then avoid perturbing such non-informative targets. On the contrary, the sequence of perturbations is irrelevant in the extreme case of a fully connected network. To investigate this hypothesis we defined a network’s isolation score as

$$1 - \sum_{ij} \frac{\pi_{ij}}{n(n-1)}, \text{ with } \pi_{ij} = \begin{cases} 1, & \exists \text{ path } i \rightarrow j \\ 0, & \nexists \text{ path } i \rightarrow j \end{cases} \quad (14)$$

Figure 4C shows that indeed the isolation score negatively correlates with the previously defined fraction of perturbations required for full network identifiability. Furthermore, we also observed a positive correlation between isolation score and the difference in the identifiability AUC between non-random and random strategies, as shown in Supplementary Figure S4B. This suggests that indeed our experimental design strategies increase their performance with increasing network isolation.

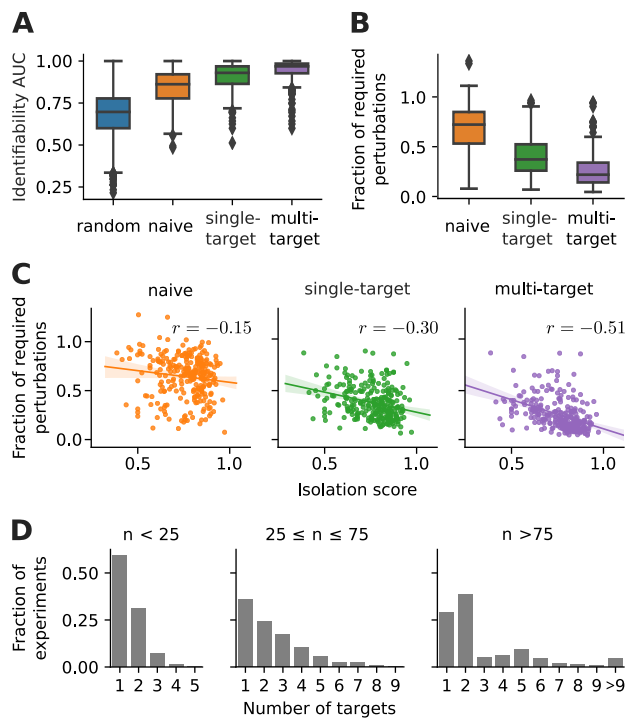
When response signals converge at a node, the individual contribution from each incoming edge can not be distinguished. Thus, the advantage of a multi-target perturbation to potentially track signal propagation through larger parts of the network is counter-balanced if it leads to more convergent signal propagation. This is prevented when the (combined) perturbations target isolated parts of the network. Therefore the

strongest correlation in Figure 4C is found for the multi-target strategy because with higher isolation score we can expect to find more such isolated subnetworks. And indeed, Figure 4D shows that the multi-target strategy typically suggest combinations of multiple single target perturbations, especially in larger networks.

In summary, we have developed an algorithmic approach to determine structural identifiability for a given network. This approach allows to derive experimental design strategies that drastically reduce experimental effort in perturbation studies. In particular, the multi-target strategy proved most efficient. Potentially, this finding has practical relevance because in many experimental contexts it easy to combine perturbations, e.g. by pooling ligands or inhibitors.

## Discussion

We have shown analytically that parameter identifiability in linear perturbation networks can be described as a simple maximum flow problem (summarised in Figure 1). This intuitive result not only explains how to achieve fully identifiable effective network models (Figure 2), but also enables us to optimize the design of perturbation experiments (Figure 3). As a test case, we examined all human KEGG pathways and found that our method typically allows to cut down the number of perturbations required for full identifiability to one fourth compared to choosing perturbation targets randomly (Figure 4). We provide a python implementation of our results [github.com/GrossTor/Identiflow](https://github.com/GrossTor/Identiflow), which allows to determine identifiability, perform matroid computations that display identifiability relationships between parameters, and optimize experimental design. The package relies on standard maximum flow algorithms from the Networkx package (Hagberg et al., 2008).



**Figure 4:** Performance of different experimental design strategies on 267 human KEGG pathways. (A) Identifiability AUC, defined as area under the number of identified nodes vs. number of perturbation curve, see Eqn. 13, (B) For each network and strategy, the average number of perturbations required for full identifiability is shown relative to the average number required for a random strategy. (C) The fraction of required perturbations correlated against the isolation score of a network (Eqn. 14),  $r$ : Spearman’s rank correlation. (D) The fraction of multi-target perturbations with a specific number of targets to all multi-target perturbations (experiments) in KEGG networks of the annotated size range.

Technically, it would be possible to cope with non-identifiabilities numerically, as was done previously (Gardner *et al.*, 2003; Bonneau *et al.*, 2006; Tegner *et al.*, 2003; Dorel *et al.*, 2018) or even through the analysis of example networks. For the latter, we could set unknown parameters to random values and numerically compute the according ranks in the identifiability conditions Equation 9 and Equation 10. The idea is that a random example system is representative of all systems with the same topology and perturbation set-up. This approach would require to define certain thresholds to detect rank deficiency and the validity of the non-cancellation assumption. Even though it is therefore not guaranteed to work in general, we would still expect it to correctly determine a parameter’s identifiability in most cases. However, the crucial benefit of the maximum flow perspective is that identifiability can be intuitively understood in relation to the network topology and the targets of the perturbations. This means that instead of requiring numerical procedures on a case by case basis, we can directly see how the maximum flow depends on the network topology and perturbation set-up. This provides a comprehensive overview on which edges become identifiable under which perturbations. For one, this permits a straightforward optimisation of the experimental design, as shown before. But even in a situation where the set of perturbations is *a priori* fixed because of experimental constraints, our approach concisely reveals which network topologies are in principle amenable to a meaningful analysis. Thereby, it maps out the range of biological questions that are actually answerable with a given experimental set-up.

It is central to our analysis to assume mild perturbations that induce a linear steady state response. This also implies that the rates in Equation 1 depend linearly on the magnitude of the other nodes and the perturbations, as shown in Supplementary Material S1. But clearly, biological systems generally break linearity assumptions in varying degrees, which bears asking how useful our description is. In principle, we could expand the steady state function Equation 3 to higher orders and attempt to also infer nonlinear rate terms, which are products of different node and perturbation magnitudes. However such products no longer have any meaningful network interpretation, as they cannot be reasonably assigned to any edge. Therefore we argue that the linearity assumption is essential to derive a useful effective network description, if we choose to interpret the biological systems in terms of ordinary differential equations. Thus the identifiability analysis presented here stays valid even when the underlying system is highly nonlinear. On the downside, the biological meaning of interaction strengths becomes increasingly obscure the more the system violates the linearity assumption (Prabakaran *et al.*, 2014). Also, our assertions no longer hold if nonlinearities violate the non-cancellation assumption. Especially combinations of perturbations, as suggested by the multi-target experimental design strategy, might push the system into saturation and thus break our analysis. We therefore need to carefully consider such biological constraints as well.

Finally, we want to stress that our analysis solely describes structural identifiability. In contrast, practical identifiability concerns situations where parameters cannot be adequately determined because of a limited amount or a poor quality of experimental data (Raue *et al.*, 2011). This means that even when the identifiability condition for a specific parameter holds, it does not necessarily mean that its value can be reliably estimated. The maximum flow approach is currently agnostic to information about noise that could potentially render a structurally identifiable parameter practically non-identifiable. Similarly, it cannot handle situations where a measurement of a node’s steady state response is not only noisy but entirely missing. As this is a common challenge in novel single cell perturbation studies (Jaitin *et al.*, 2016; Datlinger *et al.*, 2017), we consider an according generalisation of our analysis as an important line of future research.

## Funding

This work has been supported by the Deutsche Forschungsgemeinschaft, RTG2424, CompCancer.

## References

- Ahuja, R. K., Magnanti, T. L., and Orlin, J. B. (1993). *Network Flows: Theory, Algorithms, and Applications*. Prentice Hall, Englewood Cliffs, N.J.
- Arrieta-Ortiz, M. L., Hafemeister, C., Bate, A. R., Chu, T., Greenfield, A., Shuster, B., Barry, S. N., Gallitto, M., Liu, B., Kacmarczyk, T., Santoriello, F., Chen, J., Rodrigues, C. D. A., Sato, T., Rudner, D. Z., Driks, A., Bonneau, R., and Eichenberger, P. (2015). An experimentally supported model of the *Bacillus subtilis* global transcriptional regulatory network. *Mol. Syst. Biol.*, 11(11), 839.
- Bonneau, R., Reiss, D. J., Shannon, P., Facciotti, M., Hood, L., Baliga, N. S., and Thorsson, V. (2006). The Inferelator: An algorithm for learning parsimonious regulatory networks from systems-biology data sets de novo. *Genome Biol.*, 7(5), R36.
- Boros, E., Elbassioni, K., Gurvich, V., and Khachiyan, L. (2003). Algorithms for Enumerating Circuits in Matroids. In T. Ibaraki, N. Katoh, and H. Ono, editors, *Algorithms and Computation*, Lecture Notes in Computer Science, pages 485–494, Berlin, Heidelberg, Springer.

- Bosdriesz, E., Prahallad, A., Klinger, B., Sieber, A., Bosma, A., Bernardis, R., Blüthgen, N., and Wessels, L. F. A. (2018). Comparative Network Reconstruction using mixed integer programming. *Bioinformatics*, **34**(17), i997–i1004.
- Brandt, R., Sell, T., Lüthen, M., Uhlitz, F., Klinger, B., Riemer, P., Giesecke-Thiel, C., Schulze, S., El-Shimy, I. A., Kunkel, D., Fauler, B., Mielke, T., Mages, N., Herrmann, B. G., Sers, C., Blüthgen, N., and Morkel, M. (2019). Cell type-dependent differential activation of ERK by oncogenic KRAS in colon cancer and intestinal epithelium. *Nat. Commun.*, **10**(1), 2919.
- Bruggeman, F. J., Westerhoff, H. V., Hoek, J. B., and Kholodenko, B. N. (2002). Modular response analysis of cellular regulatory networks. *J. Theor. Biol.*, **218**(4), 507–520.
- Ciofani, M., Madar, A., Galan, C., Sellars, M., Mace, K., Pauli, F., Agarwal, A., Huang, W., Parkhurst, C. N., Muratet, M., Newberry, K. M., Meadows, S., Greenfield, A., Yang, Y., Jain, P., Kirigin, F. K., Birchmeier, C., Wagner, E. F., Murphy, K. M., Myers, R. M., Bonneau, R., and Littman, D. R. (2012). A validated regulatory network for Th17 cell specification. *Cell*, **151**(2), 289–303.
- Datlinger, P., Rendeiro, A. F., Schmidl, C., Krausgruber, T., Traxler, P., Klughammer, J., Schuster, L. C., Kuchler, A., Alpar, D., and Bock, C. (2017). Pooled CRISPR screening with single-cell transcriptome readout. *Nat. Methods*, **14**(3), 297–301.
- De Smet, R. and Marchal, K. (2010). Advantages and limitations of current network inference methods. *Nat. Rev. Microbiol.*, **8**(10), 717–729.
- Dinic, E. A. (1970). Algorithm for solution of a problem of maximum flow in networks with power estimation. *Sov. Math Dokl.*, **11**, 1277–1280.
- Dorel, M., Klinger, B., Gross, T., Sieber, A., Prahallad, A., Bosdriesz, E., Wessels, L. F. A., and Blüthgen, N. (2018). Modelling signalling networks from perturbation data. *Bioinformatics*, **34**(23), 4079–4086.
- Edmonds, J. and Karp, R. M. (1972). Theoretical Improvements in Algorithmic Efficiency for Network Flow Problems. *J. ACM*, **19**(2), 248–264.
- Even, S. and Even, G. (2012). *Graph Algorithms*. Cambridge University Press, Cambridge, NY, 2nd ed edition.
- Gardner, T. S., di Bernardo, D., Lorenz, D., and Collins, J. J. (2003). Inferring genetic networks and identifying compound mode of action via expression profiling. *Science*, **301**(5629), 102–105.
- Godfrey, K. and DiStefano, J. (1985). Identifiability of Model Parameter. *IFAC Proceedings Volumes*, **18**(5), 89–114.
- Greenfield, A., Hafemeister, C., and Bonneau, R. (2013). Robust data-driven incorporation of prior knowledge into the inference of dynamic regulatory networks. *Bioinformatics*, **29**(8), 1060–1067.
- Gross, T., Wongchenko, M. J., Yan, Y., and Blüthgen, N. (2019). Robust network inference using response logic. *Bioinformatics*, **35**(14), i634–i642.
- Hagberg, A., Swart, P., and S Chult, D. (2008). Exploring network structure, dynamics, and function using networkx. Technical Report LA-UR-08-05495; LA-UR-08-5495, Los Alamos National Lab. (LANL), Los Alamos, NM (United States).
- Hengl, S., Kreutz, C., Timmer, J., and Maiwald, T. (2007). Data-based identifiability analysis of non-linear dynamical models. *Bioinformatics*, **23**(19), 2612–2618.
- Jaitin, D. A., Weiner, A., Yofe, I., Lara-Astiaso, D., Keren-Shaul, H., David, E., Salame, T. M., Tanay, A., van Oudenaarden, A., and Amit, I. (2016). Dissecting Immune Circuits by Linking CRISPR-Pooled Screens with Single-Cell RNA-Seq. *Cell*, **167**(7), 1883–1896.e15.
- Kanehisa, M., Sato, Y., Furumichi, M., Morishima, K., and Tanabe, M. (2019). New approach for understanding genome variations in KEGG. *Nucleic Acids Res.*, **47**(D1), D590–D595.
- Kang, T., Moore, R., Li, Y., Sontag, E., and Bleris, L. (2015). Discriminating direct and indirect connectivities in biological networks. *Proc. Natl. Acad. Sci. U.S.A.*, **112**(41), 12893–12898.
- Kholodenko, B. N. (2007). Untangling the signalling wires. *Nat. Cell Biol.*, **9**(3), 247–249.
- Klinger, B. and Blüthgen, N. (2018). Reverse engineering gene regulatory networks by modular response analysis - a benchmark. *Essays Biochem.*, **62**(4), 535–547.
- Klinger, B., Sieber, A., Fritsche-Guenther, R., Witzel, F., Berry, L., Schumacher, D., Yan, Y., Durek, P., Merchant, M., Schäfer, R., Sers, C., and Blüthgen, N. (2013). Network quantification of EGFR signaling unveils potential for targeted combination therapy. *Mol. Syst. Biol.*, **9**, 673.
- Lorenz, D. R., Cantor, C. R., and Collins, J. J. (2009). A network biology approach to aging in yeast. *Proc. Natl. Acad. Sci. U.S.A.*, **106**(4), 1145–1150.
- Marbach, D., Prill, R. J., Schaffter, T., Mattiussi, C., Floreano, D., and Stolovitzky, G. (2010). Revealing strengths and weaknesses of methods for gene network inference. *Proc. Natl. Acad. Sci. U.S.A.*, **107**(14), 6286–6291.
- Meinshausen, N., Hauser, A., Mooij, J. M., Peters, J., Versteeg, P., and Bühlmann, P. (2016). Methods for causal inference from gene perturbation experiments and validation. *Proc. Natl. Acad. Sci. U.S.A.*, **113**(27), 7361–7368.
- Menger, K. (1927). Zur allgemeinen Kurventheorie. *Fundam. Math.*, **10**(1), 96–115.
- Pearl, J. (2009). *Causality: Models, Reasoning and Inference*. Cambridge University Press, New York, NY, USA, 2nd edition.
- Prabakaran, S., Gunawardena, J., and Sontag, E. (2014). Paradoxical results in perturbation-based signaling network reconstruction. *Biophys. J.*, **106**(12), 2720–2728.
- Raue, A., Kreutz, C., Maiwald, T., Bachmann, J., Schilling, M., Klingmüller, U., and Timmer, J. (2009). Structural and practical identifiability analysis of partially observed dynamical models by exploiting the profile likelihood. *Bioinformatics*, **25**(15), 1923–1929.
- Raue, A., Kreutz, C., Maiwald, T., Klingmüller, U., and Timmer, J. (2011). Addressing parameter identifiability by model-based experimentation. *IEE Syst Biol*, **5**(2), 120–130.
- Sachs, K., Perez, O., Pe’er, D., Lauffenburger, D. A., and Nolan, G. P. (2005). Causal protein-signaling networks derived from multiparameter single-cell data. *Science*, **308**(5721), 523–529.
- Santra, T., Kolch, W., and Kholodenko, B. N. (2013). Integrating Bayesian variable selection with Modular Response Analysis to infer biochemical network topology. *BMC Syst Biol*, **7**, 57.
- Santra, T., Rukhlenko, O., Zhernovkov, V., and Kholodenko, B. N. (2018). Reconstructing static and dynamic models of signaling pathways using Modular Response Analysis. *Current Opinion in Systems Biology*, **9**, 11–21.
- Tegner, J., Yeung, M. K. S., Hasty, J., and Collins, J. J. (2003). Reverse engineering gene networks: Integrating genetic perturbations with dynamical modeling. *Proc. Natl. Acad. Sci. U.S.A.*, **100**(10), 5944–5949.
- Timme, M. (2007). Revealing network connectivity from response dynamics. *Phys. Rev. Lett.*, **98**(22), 224101.
- Whitney, H. (1935). On the Abstract Properties of Linear Dependence. *American Journal of Mathematics*, **57**(3), 509.