Silencing Indirect Correlations for Link Prediction

Supplementary Material

Baruch Barzel[1,2] & Albert-László Barabási[1,2,3]

1. Center for Complex Network Research and Departments of Physics, Computer Science and Biology, Northeastern University, Boston, Massachusetts 02115, USA.
2. Center for Cancer Systems Biology, Dana-Farber Cancer Institute, Harvard Medical School, Boston, Massachusetts 02115, USA.
3. Department of Medicine, Brigham and Women's Hospital, Harvard Medical School, Boston, Massachusetts 02115, USA.

Contents

I. Derivation of the silencing method

1. Derivative based derivation

Consider a system of N components (nodes), each node i being characterized by activity x_i. The meaning of x_i is system dependent – in a genetic network it could be the expression level of a gene, in a metabolic network the concentration of a reactant and in a social context it could represent the probability of being infected by a contagious disease. The topology of the system is described by the adjacency matrix A_{ij} accounting for all direct interactions in the system. The dynamics of the system is captured by

$$\frac{dx_i}{dt} = F_i(x_1, \cdots, x_N), \tag{S1}$$

which under steady state condition, namely $dx_i/dt = 0$, yields

$$x_i = f_i(x_1, \cdots x_{i-1}, x_{i+1}, \cdots, x_N). \tag{S2}$$

Here $f_i(x_1, \cdots x_{i-1}, x_{i+1}, \cdots, x_N)$ is an implicit function in which the steady state activity of x_i is described in terms of all other x_j. To derive the *local response matrix* for this system we induce a small perturbation ∂x_j on the steady state activity of j, and examine the response of i, providing

$$S_{ij} = \frac{\partial x_i}{\partial x_j} = \frac{\partial f_i}{\partial x_j}. \tag{S3}$$

The partial derivative "∂" means that *no other node's activity was allowed to change in i and j's neighborhood*, hence $S_{ij} \neq 0$ only if $A_{ij} = 1$. Indeed, for j to impact i through a longer network path, one or several transitive nodes must also be perturbed. Note that by definition $S_{ii} = 0$ since x_i is absent in f_i. Hence self-feedback loops are not accounted for by S_{ij}.

In a real experimental setting the partial correlation matrix is typically inaccessible. In reality, had the activity x_j been perturbed, one would observe the response of all other node activities, so that the resulting impact on i is a convolution of all direct and indirect paths connecting i and j, leading to the *global response matrix*

$$G_{ij} = \frac{dx_i}{dx_j}. \tag{S4}$$

The formal relationship between G_{ij} and S_{ij} is given by the identity

$$\frac{dx_i}{dx_j} = \frac{\partial x_i}{\partial x_j} + \sum_{\substack{k=1 \\ k\neq j}}^{N} \frac{\partial x_i}{\partial x_k} \cdot \frac{dx_k}{dx_j}, \tag{S5}$$

valid for any $i \neq j$. Clearly for $i = j$ we have $dx_i/dx_i = 1$, which together with (S5), written in terms of G_{ij} and S_{ij}, leads to [1]

$$\begin{aligned} G_{ii} &= 1 \\ G_{ij} &= \sum_{k=1}^{N} S_{ik} G_{kj} \quad i \neq j' \end{aligned} \tag{S6}$$

a closed set of $N \times N$ linear algebraic equations, from which S_{ij} can be extracted.

The problem with (S6) is that it consists of N^2 coupled equations, making the calculation of S_{ij} computationally infeasible. Indeed, for a network of, say, $N{\sim}10^4$ nodes, the number of equations in (S6) is $L{\sim}10^8$, beyond the feasibility limit of most computational resources. We thus derive an approximate solution to (S6). Consider the matrix multiplication $S \cdot G$. Using the second equation in (S6) we find that its off-diagonal terms satisfy

$$[S \cdot G]_{ij} = G_{ij} \quad i \neq j. \tag{S7}$$

For the diagonal terms we write

$$[S \cdot G]_{ii} = \sum_{k=1}^{N} S_{ik} G_{ki}, \tag{S8}$$

which we approximate as

$$[S \cdot G]_{ii} \approx \sum_{k=1}^{N} G_{ik} G_{ki} - 1 = [(G - I)G]_{ii}, \tag{S9}$$

where the subtraction of the identity matrix, I, accounts for the fact that while the diagonal terms of G_{ij} are all one, these terms are zero in S_{ij}. The meaning of the approximation in (S9) is that for a pair of directly interacting nodes, i and k, $G_{ik} \approx S_{ik}$, stating that for nearest neighbors the correlation is dominated by the direct interaction, and indirect contributions have only a marginal effect. Note that we only rely on this approximation in evaluating the impact of the N diagonal terms (S8), but not in the construction of the $N(N-1)$ equations for the off-diagonal terms (S7). Thus we expect this approximation to have only a marginal impact on the accuracy of the resulting S_{ij} (for a more rigorous analysis of this approximation see Sec. S.I.3). From Eqs. (S7) – (S9) we have

$$S \cdot G = G - I + \mathcal{D}(S \cdot G) \approx G - I + \mathcal{D}\big((G - I)G\big), \tag{S10}$$

where $\mathcal{D}(M)$ is a matrix whose diagonal terms are taken from M and whose off-diagonal terms are set to zero. Multiplying both sides of (S10) by G^{-1} we arrive at our key result

$$S \approx \Big(G - I + \mathcal{D}\big((G - I)G\big)\Big)\, G^{-1}, \tag{S11}$$

allowing us to approximate S_{ij} using standard scalable matrix operations, taking only G_{ij} as input. While (S11) provides an approximate solution for all off-diagonal terms of S_{ij}, it does not guarantee that the diagonal terms satisfy $S_{ii} = 0$. As explained above S_{ij} is not designed to infer self-feedback loops, so in order to avoid the prediction of such spurious loops, after applying (S11), we manually set all diagonal terms of S_{ij} to zero.

2. Derivation using network paths

Below we re-derive Eq. (S6), the premise from which (S11) is reached, using a graph theoretical approach, providing insight on the network origins of the proposed silencing methodology. Consider a perturbation in the activity of a *source* node j and its impact, G_{ij}, on a *target* node i. We first evaluate the impact of j on the target's nearest neighbors k, and then find the impact of the neighbors on the target itself. The former is given by G_{ik}, providing the perturbation of the target's neighbors via all network paths between j and k. Yet once the perturbed state of the target's neighbors k is known, their impact on i should be calculated using S_{ik}. Indeed all paths leading from j to i must pass through one of i's nearest neighbors, so once the state of the neighbors is set, only their direct impact on i matters. All indirect effects have already been accounted for when the perturbations of the neighbors were considered (Fig. S1). This leads to an equation of the form

$$\begin{pmatrix}\text{Full}\\ \text{impact of}\\ \text{source on}\\ \text{target}\end{pmatrix} = \sum_{\substack{\text{All of target's}\\ \text{nearest}\\ \text{neighbors}}} \begin{pmatrix}\text{Full impact}\\ \text{of source}\\ \text{on target's}\\ \text{neighbors}\end{pmatrix} \begin{pmatrix}\text{Partial}\\ \text{impact of}\\ \text{neighbors}\\ \text{on target}\end{pmatrix}, \tag{S12}$$

where full impact describes all direct and indirect effects and partial impact describes only direct effects. Clearly, (S12) translates to

$$G_{ij} = \sum_{k=1}^{N} S_{ik} G_{kj}, \tag{S13}$$

the second equation in (S6). Hence (S13) and (S6) are exact equations providing G_{ij} through summing over all network paths leading from j to i.

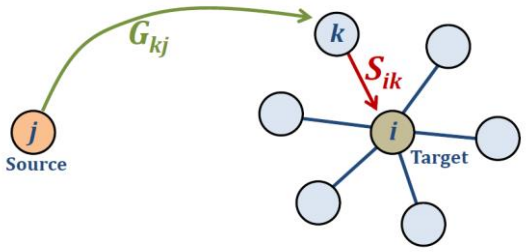

Figure S1. Construction of G_{ij} by summing over network paths. The impact of the source on the target is derived by first evaluating the impact on the target's neighbors (G_{kj}) and then the neighbor's impact on the target (S_{ik}).

3. The validity of the approximation

To evaluate the approximation

$$\mathcal{D}(S \cdot G) \approx \mathcal{D}\big((G - I)G\big), \tag{S14}$$

used in the derivation of (S11), we explicitly write the terms of the two matrices

$$[\mathcal{D}(S \cdot G)]_{ij} = \begin{cases} \sum_{k=1}^{N} S_{ik} G_{ki} & i = j \\ 0 & \text{otherwise} \end{cases} \tag{S15}$$

and

$$\big[\mathcal{D}\big((G - I)G\big)\big]_{ij} = \begin{cases} \sum_{k=1}^{N} G_{ik} G_{ki} - 1 & i = j \\ 0 & \text{otherwise} \end{cases}. \tag{S16}$$

Since $S_{ik} \neq 0$ only if i and k are directly linked, the number of elements in the sum of (S15) is exactly k_i, the degree of node i. Moreover, as this sum includes only nearest neighbors of i, we write

$$\langle [\mathcal{D}(S \cdot G)]_{ii} \rangle = \sum_{k \text{ n.n. of } i} \langle S_{ik} G_{ki} \rangle_i \approx \langle k \rangle \langle S \rangle_{\text{n.n.}} G(1), \tag{S17}$$

where $\langle \cdot \rangle_i$ denotes an average over i, $\langle k \rangle$ is the average degree and $G(d)$ is the average response between nodes at distance d. The average nearest neighbor local response, $\langle S \rangle_{\text{n.n.}}$, describes the direct impact between nearest neighbor nodes. Assuming that the impact between all pairs of nodes is dominated by the shortest paths between them, we neglect the contribution of indirect paths to the impact between nearest neighbors. This allows us to write $G(1) \approx \langle S \rangle_{\text{n.n.}}$, namely there is no significant difference between S_{ij} and G_{ij} among nearest neighbors (as is clearly observed in Fig. 1a in the paper). We thus write

$$\langle [\mathcal{D}(S \cdot G)]_{ii} \rangle \approx \langle k \rangle G^2(1). \tag{S18}$$

Similarly we evaluate the sum in (S16), decomposing it into groups of i, k pairs at distance d

$$\langle \left[\mathcal{D}\big((G-I)G\big)\right]_{ii} \rangle = \sum_{k=1}^{N} \langle G_{ik} G_{ki} - 1 \rangle_i \approx \sum_{d=0}^{d_{\max}} K(d)\, G^2(d) - 1, \tag{S19}$$

where $K(d)$ is the average number of neighbors at distance d from a node. The $d = 0$ term in the summation is $K(0)G^2(0) = 1 \times 1 = 1$, and the $d = 1$ term is $K(1)G^2(1) = \langle [\mathcal{D}(S \cdot G)]_{ii} \rangle$ (S18). Thus (S18) is actually the leading term of (S19), and hence the validity of the approximation (S9) depends on the convergence properties of the sum in (S19), namely

$$\langle \left[\mathcal{D}\big((G-I)G\big)\right]_{ii} \rangle \sim \sum_{d=1}^{d_{\max}} K(d) G^2(d). \tag{S20}$$

While $K(d)$ tends to increase with d, as nodes have a growing number of neighbors at increasing distances, $G(d)$ tends to decrease with d, as perturbations decay with distance. Thus the terms of the sum (S20) depend on the growth rate of $K(d)$ coupled with the decay rate of $G(d)$. For a typical small world network we have [2]

$$K(d) \sim e^{\alpha d}, \tag{S21}$$

in which case it can be shown that [3]

$$G(d) \sim e^{-\xi \alpha d}, \tag{S22}$$

where $\xi \geq 1$. This indicates that the decay of $G(d)$ is equal to (or more rapid than) the inflation of $K(d)$. Equation (20) becomes

$$\langle \left[\mathcal{D}\big((G-I)G\big)\right]_{ii} \rangle \sim \sum_{d=1}^{d_{\max}} e^{(1-2\xi)\alpha d} = e^{-(2\xi-1)\alpha} + \mathcal{O}\left(\left(e^{-(2\xi-1)\alpha}\right)^2\right), \tag{S23}$$

in which the leading term in the summation, as shown above, is equal to $\langle [\mathcal{D}(S \cdot G)]_{ii} \rangle$. The error, given by the remaining terms in the summation, is of order

$$\frac{e^{-2(2\xi-1)\alpha}}{e^{-(2\xi-1)\alpha}} \ll 1, \tag{S24}$$

confirming the approximation (S9).

The approximation (S9) will fail if the decay of $G(d)$ is slower than the expansion of $K(d)$, namely when $\xi < 1$. While this is not strictly prohibited, we argue that for systems of interest in the context of network inference this is extremely unlikely. The case in which $\xi < 1$ describes a system where the impact of a local perturbation gains strength as it propagates through the network: each target node is perturbed by $G(d)$, and the number of exposed nodes at d is $K(d)$. Hence using (S21) and (S22) we obtain the overall impact at distance d

$$\Gamma(d) = K(d)G(d) \sim e^{\alpha d(1-\xi)} \tag{S25}$$

which inflates exponentially with d for $\xi < 1$. This describes an unstable system, in which local perturbations do not decay. Rather they have a growing impact as they reach more distant nodes, driving the entire system into a new state. While interesting on its own right, this kind of systems is obviously not a good candidate for network inference using local perturbations. Naturally, perturbation experiments are meaningful when the system is stable, and the perturbations decay as they propagate along network paths. Indeed, the premise of inferring links from correlations is that correlations are strong only in the vicinity of the perturbed node, a condition which is only met when the approximation (S9) is satisfied.

4. From theory to empirical realization

Equation (S11) is designed to detect direct links from perturbation experiments. In the developed framework a perturbation of node i is realized by an external intervention, which in effect changes the ith equation from (S1) to

$$x_i = x_i^{ss} + \partial x_i, \tag{S26}$$

where x_i^{ss} is the steady state activity of i and ∂x_i is the induced perturbation on x_i. All the other equations in (S1) remain unchanged. Hence the experimenter forces a perturbation on node i, in effect, shifting node i's dynamics from (S1) to (S26). This perturbation will, in turn, propagate to all other nodes in the network through the rest of the equations.

The above is a rather accurate depiction of the common practice of gene overexpression. In practice, however, the empirically obtained G_{ij} could be a result of a broader set of experimental realizations. For instance, experimental perturbations might not be small, violating the infinitesimal limit under which our method is derived. In fact, an entire gene could be knocked out altogether, a rather large perturbation, which could be realized by substituting (S26) with the equation $x_i = 0$. Moreover, the correlations between nodes are not always measured directly from the response terms dx_i/dx_j (S4). Most commonly, one uses statistical correlations, *e.g.* Pearson or Spearman correlation coefficients, as a proxy from which to construct G_{ij}. Our experimental results, presented in the paper (Fig. 2) are,

in fact, no exception (see Sec. S.IV), as they rely on statistical similarity measures and not on direct measurement of (S4). Still, the silencing method, rigorously derived to treat the global response matrix (S4) is demonstrated to successfully apply to these empirically accessible proxies. The different methods by which one practically measures G_{ij} would be significant if we aimed to predict the specific response of one gene to a perturbation of another, namely a specific G_{ij} term. However our goal is different: inferring the structure of the network does not depend on the specific method by which the G_{ij} terms were acquired, but rather on the global relationships between them. The idea is that the global structure of G_{ij} reflects in some way the patterns of propagation of the perturbations along the network paths. The silencing method uses linear algebra, network science and differential tools to help interpret this global structure of G_{ij} and turn it into local link predictions via (S11). Hence what matters is the patterns of flow that are encrypted within the relationships between the G_{ij} terms. These flow patterns are inherent to the underlying network structure, and should not depend too strongly on the experimental realization. For instance, a cascade $i \to j \to k$ will be characterized by a decreasing correlation propagating along the arrows – a large correlation between i and j and a weaker one with k. The specific magnitude of these correlations might depend on the size or form of i's perturbation or on the statistical measure by which they were evaluated, but the decaying pattern, which is what is needed to infer the cascade structure, is an inherent property of the network flow. In this sense, our analysis relying on perturbations of the form of (S26) should be seen as a mathematical convenience. It allows us to rigorously derive the silencing method relying on the fundamental tools of network dynamics, but its implications are relevant for many practical empirical realizations. We exemplify this in Fig. S2, where we show numerically that silencing is observed, regardless of the magnitude or type of the induced perturbations.

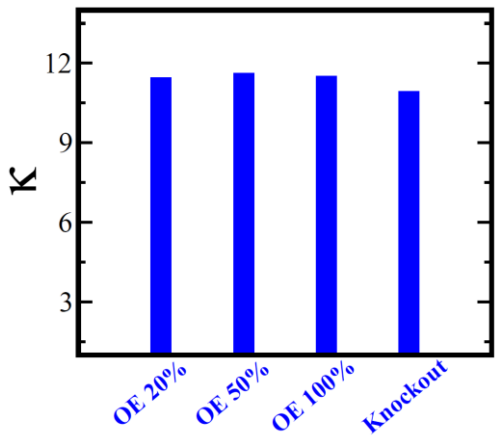

Figure S2. Silencing under different types of perturbations. To test the performance of the method with large perturbations we numerically constructed G_{ij} with overexpression (OE) perturbations increasing in size from 20 – 100% (here extracting Pearson correlations from the resulting expression patterns). The silencing, κ, is observed to be unaffected by the perturbation size. Even constructing G_{ij} from gene-knockout experiments does not harm the silencing.

II. Directed vs. Undirected Networks

Silencing is most crucial in systems where indirect correlations play an important role. Indeed, when strong indirect correlations persist it becomes impossible to distinguish direct from indirect links without the silencing effect of Eq. (S11). Below we show that *directed networks give rise to stronger transitive effects than undirected networks*, and hence their inference strongly depends on the silencing method. While here we demonstrate this using a specific model, the results we obtain are, in fact, quite general [1,3]. Consider a system whose dynamics (S1) is given by

$$\frac{dx_i}{dt} = 1 - x_i - \sum_{j=1}^{N} A_{ij} x_i x_j, \tag{S27}$$

describing a constant influx contributing to x_i; a degradation term; and negative regulation by i's nearest neighbors. We examine the structure of G_{ij} as a result of applying (S27) on a directed versus an undirected Bethe lattice (Fig. S3a). In a Bethe lattice, each node is assigned a *level*, denoting its hierarchical position in the tree-like structure, so that a node from the nth level is linked to one node at level $n-1$ and to $k-1$ nodes at level $n+1$. Denoting the average steady state activity of nodes at level n by x_n, we write (S27) under steady state condition as

$$1 - x_n - x_{n-1}x_n - (k-1)x_n x_{n+1} = 0 \tag{S28}$$

for the undirected network and

$$1 - x_n - x_{n-1}x_n = 0 \tag{S29}$$

for the directed network. In (S29) the level $n+1$ nodes do not impact nodes from level n. Following the procedure of (S3) we induce a small constant perturbation ∂x_{n-1} and test the cascading effect on x_n, x_{n+1} and so on. For the undirected network we have

$$1 - (x_n + \partial x_n) - (x_{n-1} + \partial x_{n-1})(x_n + \partial x_n) - (k-1)(x_n + \partial x_n)x_{n+1} = 0, \tag{S30}$$

while for the directed network we have

$$1 - (x_n + \partial x_n) - (x_{n-1} + \partial x_{n-1})(x_n + \partial x_n) = 0. \tag{S31}$$

Note that x_{n+1} was not perturbed in (S30), only x_{n-1} and x_n, following the "∂" notation of (S3), which excludes indirect effects. Incorporating the steady state assumption ((S28) and (S29)) and taking only linear terms in ∂x_n we find

$$\begin{aligned} S_{\text{Undir}} &= \frac{\partial x_n}{\partial x_{n-1}} = \frac{\langle x \rangle}{k\langle x \rangle - 1} \\ S_{\text{Dir}} &= \frac{\partial x_n}{\partial x_{n-1}} = \frac{\langle x \rangle}{\langle x \rangle - 1} \end{aligned} \tag{S32}$$

for the undirected and directed networks respectively. Here $\langle x \rangle$ denotes the average steady state activity of all nodes in the system. Equation (32) indicates that the direct impact between nodes is greater in the directed network than in the undirected network. The impact at larger distances d, roughly given by $G(d) \approx S^d_{\text{Undir}}$ $\left(G(d) \approx S^d_{\text{Dir}}\right)$ is thus also expected to be greater in the undirected network (an exact solution to this problem appears in [1]). Such large impact at a distance will make the directed network's G_{ij} less predictive, and hence the silencing method more effective.

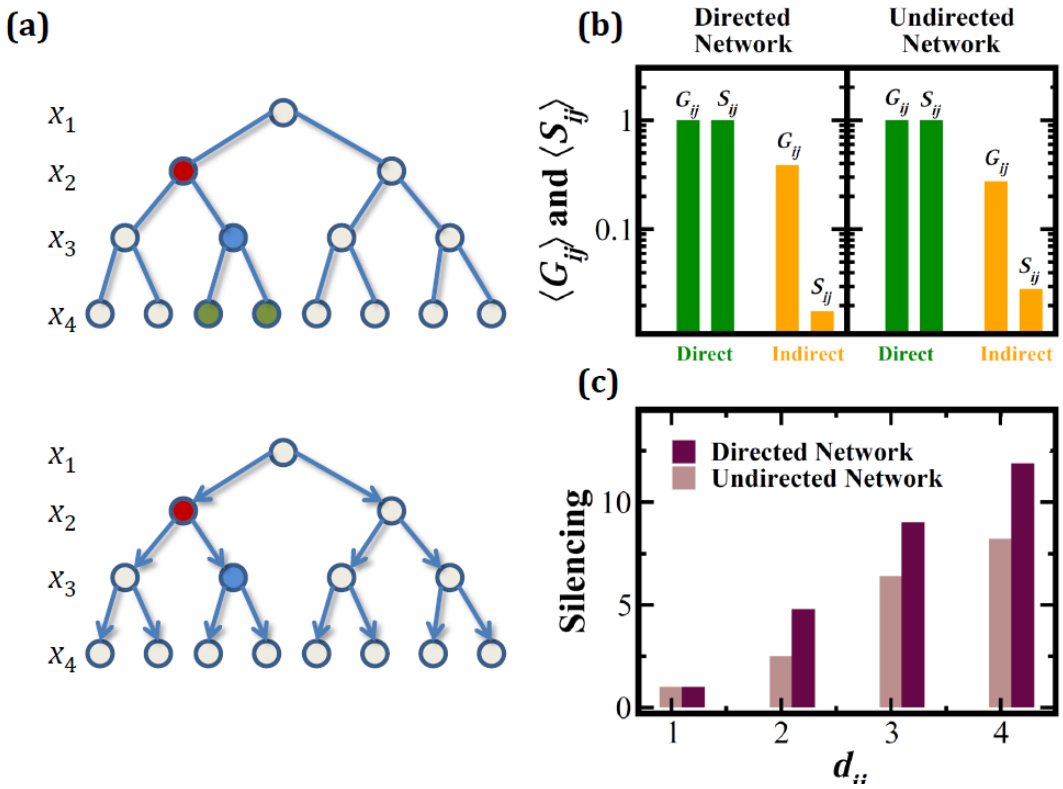

Figure S3. **Silencing in directed vs. undirected networks**. (a) An undirected versus a directed Bethe lattice. The red node has been perturbed, impacting the blue node, one level below. In the undirected network the green nodes at level 4 also regulate x_3, while in the directed network the impact is governed by the red x_2 node alone. (b) The silencing of indirect correlations is stronger in the directed network than in the undirected network, as shown by the increased difference between direct (green bars) and indirect (orange bars) terms in S_{ij} vs. G_{ij}. (c) As the distance between nodes is increased, the silencing becomes more significant. The effect is more emphasized in the directed network.

To test the silencing effect on directed/undirected networks we numerically simulated the dynamics of (S27) and obtained the global and local response matrices, G_{ij} and S_{ij}, via (S11). In Fig. S3b we compare the silencing effect in both networks, finding that indeed stronger silencing is observed for the directed network. We also find that the silencing of correlations at distance d increases more rapidly for the directed network than the undirected network (Fig. S3c).

III. Numerical Support

To test the predictive power of the method (S11) we used it to infer the network of a numerically simulated system, providing a controlled environment, where we could test its precision. We used a scale-free network consisting of $N = 5{,}000$ nodes and $L = 20{,}000$ links. The dynamical equations (S1) where taken to follow Michaelis-Menten dynamics [4 - 5]

$$\frac{dx_i}{dt} = -x_i + \sum_{j=1}^{N} A_{ij} \frac{x_j}{1 + x_j}, \tag{S33}$$

which we solved using a fourth-order Runge-Kutta stepper. After reaching steady state we numerically induced a constant perturbation on the activity of all nodes as shown in (S26). The system then reached its new, perturbed, steady state, providing the global response matrix, G_{ij} (S4). To derive S_{ij} we used (S11), achieving 100% accuracy in predicting the system's topology: all non-interacting node pairs were found to be associated with terms satisfying $S_{ij} \to 0$.

IV. Empirical evidence

1. The data

We used the DREAM5 datasets, which are described in detail in [6]. The input data includes a compendium of 805 microarray experiments, in which the expression levels of 4,511 genes were measured under different experimental conditions, giving rise to an $805 \times 4{,}511$ expression matrix M_{ij}. Of the 4,511 genes a total of 5% were added by the DREAM5 organizers as *decoy genes*, so that only 4,297 entries in the expression matrix represent real experimental data. In this dataset the genes are anonymized so that links could only be inferred from the given expression matrix. In addition to the expression matrix the input data also includes a list of 141 genes, designated as potential transcription factors.

We validated our predictions using the gold standard benchmark for *E. coli* provided by the DREAM5 challenge. This benchmark includes 2,066 transcriptional interactions constructed from RegulonDB (Release 6.8), all of which are supported by at least one strong type of evidence (as annotated by RegulonDB). We used this gold standard to construct the $4{,}511 \times 4{,}511$ directed adjacency matrix A_{ij}.

2. Analysis and evaluation

Using the expression matrix, M_{ij}, we constructed G_{ij} between all genes resulting in a $4{,}511 \times 4{,}511$ global response matrix (see below). We then used Eq. (S11) to obtain S_{ij}. The terms of G_{ij} and S_{ij} were used to evaluate the likelihood of a transcriptional link between all 141 transcription factors and 4,511 potential target genes, amounting to $141 \times 4{,}511 = 636{,}051$ potential links. Ranking these links based on their likelihood we obtained two lists L_G and L_S, representing the link predictions of G_{ij} and S_{ij} respectively. For example, the first n entries in the L_G list represent the n most likely links, as predicted by G_{ij}.

ROC curves: we define the *true positive rate* as [6]

$$TPR(n) = \frac{TP(n)}{P}, \tag{S34}$$

where $TP(n)$ denotes the number of true positives in the top n predictions and P is the number of positives in the gold standard. Similarly the false positive rate is defined as

$$FPR(n) = \frac{FP(n)}{Q}, \tag{S35}$$

where $FP(n)$ is the number of false positives among the top n predictions and Q is the number of negatives in the gold standard.

Discrimination ratio and silencing: an important role of an inference method is to discriminate between direct and indirect interactions. For a given prediction list, say L_G, we averaged over all entries G_{ij} associated with direct links (2,066 terms in the DREAM5 dataset) to obtain $\langle G_{ij} \rangle_{\text{Dir}}$, and over all the remaining entries to obtain $\langle G_{ij} \rangle_{\text{Indir}}$. The power of G_{ij} to discriminate between direct and indirect links is then given by the *discrimination ratio*

$$\Delta_G = \frac{\langle G_{ij} \rangle_{\text{Dir}}}{\langle G_{ij} \rangle_{\text{Indir}}}, \tag{S36}$$

where the larger is Δ_G the more discriminative is G_{ij}. By silencing indirect correlations Eq. (S11) leads to $\langle S_{ij}\rangle_{\text{Indir}} \ll \langle G_{ij}\rangle_{\text{Indir}}$, and hence a larger Δ_S. The *silencing* is given by

$$\kappa = \frac{\Delta_S}{\Delta_G}, \tag{S37}$$

capturing the extent to which S_{ij} is more discriminative than G_{ij}.

Motif analysis: in addition to the overall silencing described above, we also focused on the silencing of two specific motifs, cascades and co-regulatory sequences, which are highly prevalent in biology, and pose a challenge to most inference techniques (see Fig. 2e - f in the paper). Each of these motifs is associated with three response terms: G_{XI} (or G_{IX} in the case of co-regulation) and G_{IY}, which are associated with existing links, and G_{XY}, which is a result of an indirect interaction. To evaluate the discrimination ratio for these motifs we calculated

$$\Delta_G^{\text{Cascade}} = \frac{\langle \frac{1}{2}(G_{XI} + G_{IY})\rangle}{\langle G_{XY}\rangle} \tag{S38}$$

and

$$\Delta_G^{\text{Coregulation}} = \frac{\langle \frac{1}{2}(G_{IX} + G_{IY})\rangle}{\langle G_{XY}\rangle}, \tag{S39}$$

where the average is over all relevant motifs in the network. The same was calculated for S_{ij}, and the ratio was used to obtain the motif silencing as in(S37).

3. Method for calculating G_{ij}

We used three standard methodologies to calculate the global response matrix G_{ij}, as described below:

(i) **Pearson correlations:** We denote by M_i the ith column of the expression matrix M_{ni}. This vector lists the expression levels of gene i under all $n = 1, \ldots, 805$ experimental conditions. For every transcription factor j we calculated Pearson's correlation coefficient with all target genes i as

$$G_{ij} = \frac{\langle M_i M_j\rangle - \langle M_i\rangle\langle M_j\rangle}{\sqrt{\langle M_i^2\rangle - \langle M_i\rangle^2}\sqrt{\langle M_j^2\rangle - \langle M_j\rangle^2}}. \tag{S40}$$

(ii) **Spearman rank correlations:** To calculate Spearman's correlation between the transcription factors and all target genes we first transform the expression vectors M_i and M_j into rank vectors m_i and m_j. The nth entry in the rank vector denotes the rank of the corresponding entry in the expression vector. The Spearman correlation is then calculated by applying (S40) to the rank vectors m_i and m_j.

(iii) **Mutual information:** We constructed G_{ij} using the mutual information between M_i and M_j, defined as

$$G_{ij} = \sum_{x_i,x_j} P(x_i,x_j) \log \frac{P(x_i,x_j)}{P(x_i)P(x_j)}, \tag{S41}$$

where $P(x_i,x_j)$ is the probability that randomly selected terms from M_i and M_j are equal to x_i and x_j, and $P(x_i)$ ($P(x_j)$) are the relevant marginal probability distributions. Mutual information has been shown to be a superior link predictor compared to Pearson/Spearman correlations, as is also supported by our results (Fig. 2c) [7-9].

4. Additional processing of the correlation data

As Eq. (S11) indicates, the silencing method requires that the input matrix, G_{ij}, is invertible. This may not be guaranteed when using statistical measures such as Pearson or Spearman correlations. For instance, if the number of experimental conditions (columns in the expression matrix) is small compared to the number nodes (rows in the expression matrix), the resulting correlation matrix is likely to be singular. This is a rather common problem in dealing with statistical correlations applied to limited datasets and several solutions are summarized in [10]. Below we focus on two solutions of particular relevance:

(i) In the analysis of the DREAM5 datasets, described above, we did not have to construct the complete $4{,}511 \times 4{,}511$ correlation matrix, as the 141 transcription factors were already known. Hence, following the DREAM5 protocol, we only constructed the $141 \times 4{,}511$ matrix of the correlations between the transcription factors and their potential targets. To obtain a square matrix we then set the diagonal terms for all the rest of the nodes to unity. The resulting matrix, in which the correlations for 141 nodes were obtained from 805 experimental conditions is expected (and indeed found) to be invertible.

(ii) A more statistically significant, and likely invertible, correlation matrix could also be obtained by additional processing of the raw Pearson/Spearman

correlation matrix. First we compute the correlation between the activity vectors M_i and M_j for each pair of nodes, i and j. We then calculate the statistical significance of the obtained correlation by randomizing M_i and M_j, and assigning a P-value for the measured correlation. This results in a *significance* matrix, likely to be invertible, even if the raw correlation matrix was singular.

V. Robustness of the method

We now turn to analyze the robustness of the method against uncertainty in the data. We first examine the method's performance in the presence of noise, and then refer to case of *hidden nodes*, where a fraction of the nodes are experimentally inaccessible.

1. Robustness against noise

Clearly, all experimental data is subject to noise, so that the global response matrix, G_{ij}, used as input for silencing, is characterized by a certain level of uncertainty. Hence we ran extensive numerical simulations in order to test the impact of noise on the silencing effect. As a first step we constructed a linear cascade of 15 nodes in which the response between all neighboring nodes was set to g. The global response matrix in this case is

$$G_{ij} = g^{d_{ij}}, \tag{S42}$$

where d_{ij} is the distance between i and j. Applying (S11) to (S42) we obtain S_{ij}, which provides a perfect reconstruction on the original cascade network. To explore the impact of uncertainty we introduced Gaussian noise into G_{ij}. Assuming that each term in the noiseless G_{ij}, represents an expectation value $\mu = G_{ij}$, we allow the *measured* G_{ij} to express some variability around μ. We achieve this by drawing the i,j term of the global response matrix from a normal distribution with an average of μ and a variance of $\sigma^2 = \theta\mu$, leading to a noisy G_{ij} with a signal to noise ratio given by θ. We then measured the silencing, κ, for increasing values of θ, finding that the silencing decays as (Fig. S4a)

$$\kappa \sim \theta^{-1}. \tag{S43}$$

This power law dependence represents a rather slow decay of the method's performance (*i.e.* not exponential). Silencing is completely lost when the noise reaches a critical level, θ_C, for which $\kappa = 1$ (dashed line in Fig. S4a). We find that for most values of g we have $\theta_C \approx 1$, so that silencing is lost only when the signal becomes completely driven by noise. Taken together, the slow decay of the silencing (S43) and its persistence up to $\theta_C \approx 1$, both indicate the system's high robustness against noise.

As g becomes larger, the decay of the terms of G_{ij} with distance becomes slower (S42). Hence direct and indirect responses becomes less distinctive and the discrimination ratio of G_{ij} decreases. As long as $g < 1$, however, the ranking of the G_{ij} terms is ensured, that is

indirect response terms are guaranteed to be smaller than direct ones. Silencing will further lower the indirect terms, and (S11) will remain accurate. In the limit where $g \lesssim 1$, the noise can lead to a situation where indirect responses exceed direct ones. For instance for $g = 0.9$ we have for next nearest neighbors $G_{ij} = g^2 = 0.81$. Under these circumstances, even a small amount of noise could cause the indirect terms to seem greater than the direct ones, confounding S_{ij}'s ability to properly identify and silence the indirect responses. Thus the greater is the direct impact between nodes, g, the higher is the method's sensitivity to noise and the lower is θ_C. In Fig. S4b we show θ_C vs. g, finding that indeed for most values of g the critical noise level is at a signal to noise ratio of about unity, so that the method is highly robust against noise. As predicted, when g approaches unity we observe a sharp drop in the method's robustness.

Following a similar procedure we further tested the performance of (S11) with noise on the numerical G_{ij} described in Sec. S.III. As expected silencing is unharmed for small values of θ, representing low noise levels (Fig. S4c). For high noise levels ($\theta \gtrsim 0.2$) we find that the complex network displays similar behavior to what we found for the linear cascade, with the silencing decreasing as a power of the noise level ($\sim\theta^{-0.9}$), becoming completely void at $\theta_C \lesssim 1$.

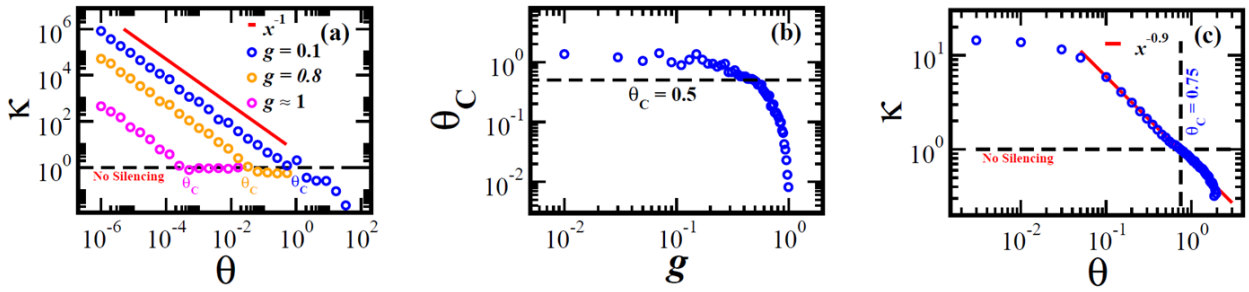

Figure S4. Silencing in a noisy environment. In order to test the method's performance with noisy input we introduced Gaussian noise into the numerically obtained G_{ij}, and measures the silencing, κ, vs. the signal to noise ratio θ. (a) For a linear cascade network we find that the silencing decays as θ^{-1}, representing a gradual decay in performance due to noise. Silencing is lost at $\theta = \theta_C$, for which $\kappa = 1$ (dashed line). When perturbations decay rapidly ($g < 1$) we have $\theta_C \approx 1$ (blue). As g is increased θ_C becomes smaller and the system is more sensitive to noise (orange and magenta). (b) The critical noise level, θ_C, vs. g. For most values of g silencing persists even when the noise levels are high, failing only at $\theta_C \approx 1$. When g approaches unity, a sharp drop is observed, indicating that S_{ij} becomes sensitive to noise. The dashed line marks $\theta_C = 0.5$. (c) We tested the resilience of the method to noise for a scale free network (Sec. S.III). For low noise levels ($\theta \lesssim 0.2$) silencing is unharmed. As the noise level is increased the silencing decreases as $\kappa \sim \theta^{-0.9}$, similarly to the behavior of the cascade network in (a). Silencing is lost at $\theta_C \approx 0.75$.

2. Robustness against hidden nodes

Another source of uncertainty is the case of hidden nodes, in which the input matrix G_{ij} includes only a subset of all the nodes acting in the system. The remaining *hidden nodes* impact the system's observed correlations, but are inaccessible experimentally. To evaluate the impact of such hidden nodes on the performance of the method we analyze the dependence of the silencing, κ, on the fraction η of hidden nodes. As explained in Sec. S.I.2 the silencer transformation is designed to detect the network paths and silence all but the direct ones. It assumes that the observed G_{ij} is a convolution of all paths in the network and uses self-consistency (Eq. (S11)) to reveal the unique solution for the underlying network. Silencing will thus fail if such a self-consistent solution cannot be found. This will occur when the fraction of hidden nodes exceeds a threshold η_c, for which most paths between nodes become *hidden*. Equation (S11) will not silence those *hidden paths* and will leave their correlations untouched. Below we show that for dense networks, even under many hidden nodes, most paths remain unhidden, and hence $\eta_c \to 0$, namely even with a large fraction of hidden nodes S_{ij} maintains its predictive power.

Consider a network of N nodes, of which a fraction η are hidden. Our experimental access to the system is through the $n \times n$ global response matrix G_{ij}, where $n = (1-\eta)N$. Our goal is to infer the underlying sub-network of the n unhidden nodes using (S11). First we consider the scenario in which the sub-network constitutes a giant connected component, namely that there exists a path connecting all or most pairs of nodes in the observed n node system. In this case, as all paths are present, the self-consistent solution of (S11) can be satisfied and the transformation will successfully silence the indirect paths. However, if a pair of nodes, i and j, becomes isolated in the sub-network, that is the path (or paths) linking them passes through the hidden nodes, the pertinent G_{ij} term will not be silenced. Indeed, in the absence of the (i,j) path, the only self-consistent solution is that the non-zero G_{ij} term originates in a direct link (Fig. S5a).

To illustrate this we consider the simple example of a linear cascade $j \to k \to i$, in which k is a hidden node, and all we are offered is experimental access to the indirect response of i to j, G_{ij}. Clearly, under these circumstances, Eq. (S11) will not be able to classify the i,j link as indirect. Indeed, in this case, absent any other node or information about the system, it is *mathematically impossible* to classify this link as direct or indirect, as there is no information in the observed response matrix from which the existence of node k could be inferred. This is exemplified in Fig. S5b where we present the silencing for all pairs of nodes alongside the silencing of isolated pairs, which cannot be connected via a finite path. Indeed, for pairs of isolated nodes, we observe no silencing at all ($\kappa \approx 1$). Hence the performance of (S11) with hidden nodes depends on the number of hidden paths, namely the fraction of node pairs that are connected by a finite path in the original N node network, but become isolated in the unhidden n node sub-network.

To test this conclusion we used the numerical experiment described in Sec. S.III, taking the original 5000×5000 node G_{ij}, and eliminating a randomly selected fraction η of the nodes.

We then applied (S11) to the resulting $n \times n$ global response matrix to predict S_{ij}, and calculated the silencing ratio. For each of the sub-networks we also calculated

$$\rho = \frac{P_{\mathrm{Fin}}}{P_{\infty}} \tag{S44}$$

where P_{Fin} is the number of node pairs that are linked via a finite path and P_{∞} is the number of isolated node pairs. Fig. S5c indicates that ρ, capturing the ratio of finite vs. infinite paths among the unhidden nodes, is indeed a predictor of the method's performance. For $\rho < 1$ silencing is lost and S_{ij} is expected to have a comparable predictive power to G_{ij}. As ρ is increased, the fraction of isolated node pairs decreases, silencing becomes stronger and the method's performance is improved.

Typically, eliminating nodes from a random network results in the emergence of a giant connected component: a fraction f of the nodes remain linked by finite paths, while the remaining $1 - f$ become isolated [2]. Hence the number of finite paths in the resulting sub-network is proportional to the number of pairs of nodes in the giant connected component, namely

$$P_{\mathrm{Fin}} \propto \frac{f^2}{2}. \tag{S45}$$

A pair of isolated nodes could either be composed of one node from the connected component and one that is not, or when both nodes are out of the giant component. This provides

$$P_{\infty} \propto f(1-f) + \frac{(1-f)^2}{2} = \frac{1}{2}(1-f^2), \tag{S46}$$

which together with (S44) and (S45) yields

$$\rho = \frac{f^2}{1-f^2}. \tag{S47}$$

As we have shown above, for silencing to persists one must have $\rho > \rho_{\mathrm{c}} \approx 1$, which in (S47) provides $f_{\mathrm{c}} = \sqrt{1/2} \approx 0.7$ (see Fig. S5d). Hence the critical fraction of hidden nodes, η_{c}, for which silencing fails and S_{ij} loses its predictive power is when the largest connected component enters below a threshold of $\sim 70\%$. Below we show that for a typical, biologically relevant network, this allows a rather large fraction of hidden nodes.

For a random graph with an average degree $\langle k \rangle$ it has been shown that the size of the largest connected component satisfies [2]

$$f = 1 - e^{-\langle k \rangle f}, \tag{S48}$$

from which we obtain the critical average degree

$$\langle k_{\mathrm{c}}\rangle = \frac{1}{f_c}\ln\left(\frac{1}{1-f_c}\right) = \sqrt{2}\ln\left(\sqrt{2}+2\right) \approx 1.7. \tag{S49}$$

Hiding nodes from the network eliminates the edges associated with them, and thus decreases the average degree. When a randomly selected fraction η of nodes has been hidden, each of the remaining nodes is expected to lose a similar fraction of its associated edges, so that the average degree of the remaining sub-network becomes

$$\langle k(\eta)\rangle = \langle k\rangle(1-\eta). \tag{S50}$$

This is clearly demonstrated in Fig. (S5e). Substituting the l.h.s. of (S50) with the critical degree (S49) we find that the critical fraction of hidden nodes is

$$\eta_{\mathrm{c}} = 1 - \frac{\langle k_{\mathrm{c}}\rangle}{\langle k\rangle}. \tag{S51}$$

Equation (S51) describes the critical fraction of hidden nodes, above which the silencing ratio is one, and S_{ij} becomes no more predictive than G_{ij}. For dense networks, with a large $\langle k\rangle$, we have $\eta_{\mathrm{c}} \to 1$, indicating that the silencing method is extremely robust against hidden nodes. For instance for $\langle k\rangle = 10$, a realistic value in the biological context, we have $\eta_{\mathrm{c}} \approx 0.8$, namely indirect paths will still be silenced even if the fraction of hidden nodes approaches 80%. In our numerical example $\langle k\rangle = 4$, for which (S51) predicts $\eta_{\mathrm{c}} \approx 0.57$. Indeed, as Fig. S5f indicates silencing is terminated for all $\eta > \eta_{\mathrm{c}}$, and for $\eta < \eta_{\mathrm{c}}$ the silencing ratio starts to increase, so that the method begins to gain predictive power.

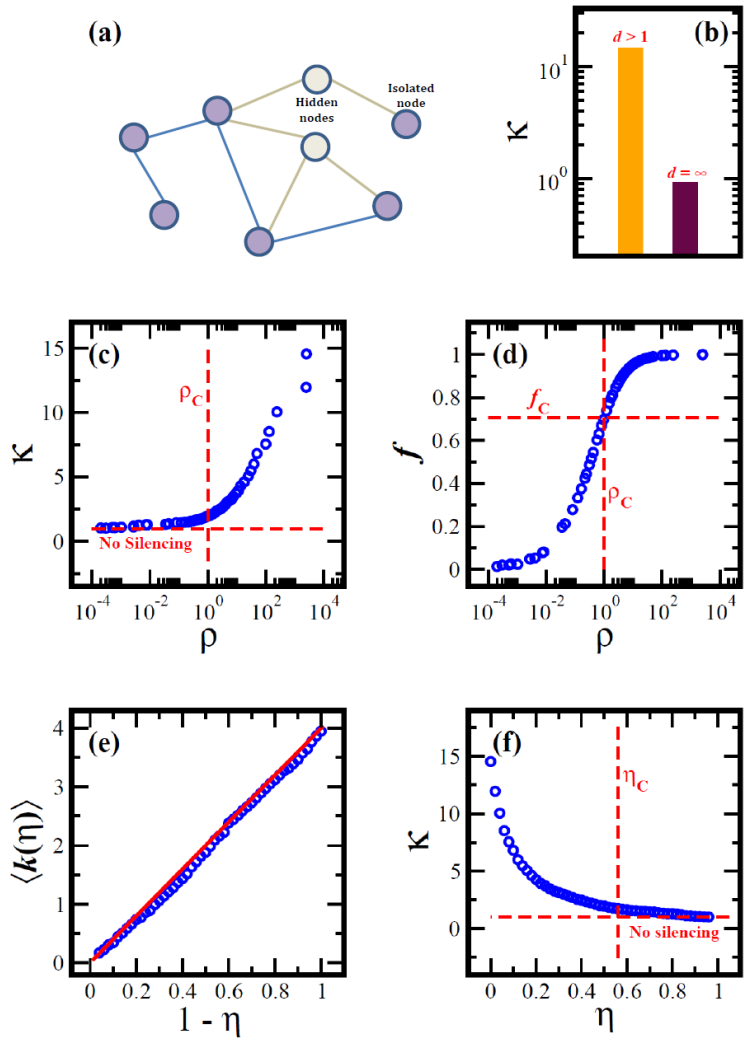

Figure S5. Performance of the method with hidden nodes. (a) A network with $N = 8$ nodes of which a fraction $\eta = 1/4$ are hidden. The resulting sub-network has $n = 6$ nodes (light blue), five of which constitute a connected component and one which is isolated. Equation (S11), applied to the sub-network will successfully silence the indirect correlations associated with the unhidden paths of the connected component. However the correlations between the isolated node and the rest of the network, which cannot be associated with an existing indirect path, will not be silenced. (b) The silencing of correlations associated with finite paths ($d > 1$) is large (orange bar), while for correlations between isolated nodes ($d = \infty$) it is $\kappa = 1$, namely no silencing (purple bar). Hence for the node pairs that are connected via hidden paths S_{ij} is not more predictive than G_{ij}. Thus as long as the isolated node pairs (connected via hidden paths) are a minority S_{ij} maintains its advantage, but if the majority of nodes become isolated S_{ij} becomes comparable to G_{ij} and the silencing approaches $\kappa = 1$. (c) To exemplify this we present κ vs. ρ, which describes the ratio of finite (existing) to infinite (hidden) paths (S44). Silencing is observed as long as most node pairs are connected by finite paths ($\rho > 1$), but when the hidden paths dominate ($\rho < 1$) silencing no longer plays a significant role. (d) The value of ρ is determined by the size of the largest connected component, f, as given by (S47). The critical value of $\rho_C = 1$ corresponde to $f_C = \sqrt{1/2}$ (dashed red lines). Hence the critical fraction of hidden nodes, η_C, is reached when the largest connected component has size f_C. (e) The average degree of the unhidden nodes vs. the fraction of unhidden nodes, $1 - \eta$ (circles) is in perfect agreement with Eq. (S50) (solid line). (f) Silencing vs. the fraction of hidden nodes. Silencing is maintained as long as $\eta < \eta_C$. Above the critical fraction η_C, given by (S51), silencing becomes insignificant and S_{ij} shows no advantage over G_{ij}. For this network we have $\langle k \rangle = 4$, providing $\eta_C \approx 0.57$.

References

1. B. Barzel and O. Biham. Quantifying the connectivity of a network: The network correlation function method. *Phys. Rev. E* **80**, 046104 (2009).
2. M.E.J. Newman. Networks - an introduction. Oxford University Press, New York, (2010).
3. B. Barzel and A.L. Barabási. Universality in network dynamics. To be published.
4. U. Alon. An introduction to systems biology: design principles of biological circuits. Chapman & Hall, London, U.K., (2006).
5. G. Karlebach and R. Shamir. Modeling and analysis of gene regulatory networks. *Nature Reviews* **9**, 770–780 (2008).
6. D. Marbach *et al*. Wisdom of the crowds for robust gene network inference. *Nature Methods* **9**, 796-804 (2012).
7. A.J. Butte and I.S. Kohane. Mutual Information relevance networks: functional genomic clustering using pairwise entropy measurements. *Pacific Symposium on Biocomputing* **5**, 415-426 (2000)
8. A.A. Margolin *et al*. ARACNE: An algorithm for the reconstruction of gene regulatory networks in a mammalian cellular context. *BMC Bioinformatics* **7**,S7, (2006).
9. J.J. Faith *et al*. Large-scale mapping and validation of *Escherichia coli* transcriptional regulation from a compendium of expression profiles. *PLoS Biol* **5**(1), e8 (2007).
10. G.H. Tucci and K. Wang. New Methods for handling singular sample covariance matrices. arXiv:1111.0235v1 (2011).