Network link prediction by global silencing of indirect correlations

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

Predictions of physical and functional links between cellular components are often based on correlations between experimental measurements, such as gene expression. However, correlations are affected by both direct and indirect paths, confounding our ability to identify true pairwise interactions. Here we exploit the fundamental properties of dynamical correlations in networks to develop a method to silence indirect effects. The method receives as input the observed correlations between node pairs and uses a matrix transformation to turn the correlation matrix into a highly discriminative silenced matrix, which enhances only the terms associated with direct causal links. Against empirical data for *Escherichia coli* regulatory interactions, the method enhanced the discriminative power of the correlations by twofold, yielding >50% predictive improvement over traditional correlation measures and 6% over mutual information. Overall this silencing method will help translate the abundant correlation data into insights about a system's interactions, with applications ranging from link prediction to inferring the dynamical mechanisms governing biological networks.

The currently incomplete maps of molecular interactions between cellular components limit our understanding of the molecular mechanisms behind human disease[1–6]. Ultimately, high-throughput projects[7–10] are expected to provide the accurate maps of interactomes necessary to systematically unlock disease mechanisms. Yet, as a complete interaction map is currently not at hand, we need to develop tools that allow us to infer the structure of cellular networks from empirically obtained biological data[11,12]. Many current tools designed to infer functional and physical interactions in the cell rely on the global response matrix,

$$G_{ij} = \frac{dx_i}{dx_j} \tag{1}$$

which captures the change in node i's activity in response to changes in node j's[13]. This matrix can be measured directly from gene knockout or overexpression experiments, or inferred indirectly using related measures such as Pearson or Spearman correlations[14], mutual information[15,16] or Granger causality[17]. Traditional methods for predicting links[15,16,18,19] assume that the magnitude of G_{ij} correlates with the likelihood of a direct functional or physical link between nodes i and j. Yet G_{ij} cannot distinguish between direct and indirect relationships: a path $i \to k \to j$ can result in a measurable response observed between i and j, falsely suggesting the existence of a direct link between them (**Fig. 1a**,**b**).

Several methods to correct for such effects have been proposed. Information theory approaches evaluate the association between nodes by measuring the entropy of their mutual activities, where a low entropy indicates a statistical dependence between the node activities[16,18,20]; probabilistic models, such as the graphical Gaussian model, allow one to evaluate the correlation between i and j, while controlling for the state of node k, and thereby provide a more indicative measure of direct linkage[21–25]; other models rely on assumptions pertaining to the network topology, such as the tendency of real networks to exhibit strong degree correlations[26]. The ultimate solution, however, should enable us to fully unwind the direct from the indirect effects, providing a measure that distinctly indicates the existence of direct links. Consequently, we focus here on the local response matrix

$$S_{ij} = \frac{\partial x_i}{\partial x_j} \tag{2}$$

in which the contribution of indirect effects is eliminated. In contrast with equation (1), which allows for global changes in i and j's environment, here the "∂" indicates that S_{ij} is defined to capture only local effects, namely the response of i to changes in j when all surrounding nodes except i and j remain unchanged. Hence $S_{ij} > 0$ implies a direct link between i and j.

We derive a method for calculating the local response matrix (2) from experimentally accessible correlation measures, allowing us to mathematically discriminate direct from indirect links (**Fig. 1**). We show that the resulting S_{ij} matrix, in which the contribution of indirect paths is silenced, is more discriminative than the empirically obtained G_{ij} matrix, enhancing our ability to extract direct links from experimentally collected correlation data.

RESULTS

The silencing method

To extract S_{ij} from the experimentally accessible G_{ij}, we formally link equations (1) and (2) via

$$\begin{cases} \dfrac{dx_i}{dx_i} = 1 \\ \dfrac{dx_i}{dx_j} = \displaystyle\sum_{k=1}^{N} \frac{\partial x_i}{\partial x_k}\frac{dx_k}{dx_j} \quad i \neq j \end{cases} \tag{3}$$

[1]Center for Complex Network Research and Departments of Physics, Computer Science and Biology, Northeastern University, Boston, Massachusetts, USA. [2]Center for Cancer Systems Biology, Dana-Farber Cancer Institute, Harvard Medical School, Boston, Massachusetts, USA. [3]Department of Medicine, Brigham and Women's Hospital, Harvard Medical School, Boston, Massachusetts, USA. Correspondence should be addressed to A.-L.B. (barabasi@gmail.com).

Received 22 October 2012; accepted 23 April 2013; published online 14 July 2013; doi:10.1038/nbt.2601

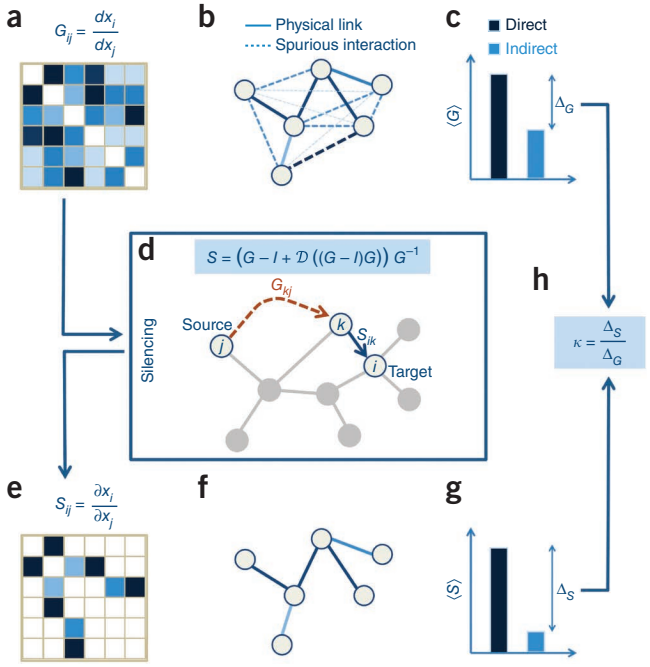

Figure 1 Silencing indirect links. (**a**) The experimentally observed global response matrix, G_{ij}, accounts for direct as well as indirect correlations, with no clear separation between them. The source of G_{ij} could be gene coexpression data, statistical correlations or genetic perturbation experiments. (**b**) In the absence of a clear separation in G_{ij} assigned to direct and indirect correlations, our ability to infer direct physical links (solid lines) is limited. Simple thresholding, that is, accepting all links for which G_{ij} exceeds a predefined threshold, is known to predict spurious links (thick dashed lines) and overlook true links (thin solid lines). (**c**) Although the average G_{ij} terms associated with direct links are higher than the average terms associated with indirect links, as captured by the discrimination ratio, Δ_G, the difference is not sufficient to fully discriminate between direct and indirect links. (**d**) Silencing is achieved through equation (5), which exploits the flow of information in the network: the flow from the source (j) to the target (i) is carried through the indirect effect G_{kj} (brown) coupled with the direct impact S_{ik} of the target's nearest neighbor k. By silencing the indirect contributions, equation (5) provides the local response matrix, S_{ij}, whose nonzero elements correspond to direct links. (**e**,**f**) In S_{ij} the terms associated with indirect links are silenced, allowing us to detect only the direct links of the underlying network. (**g**) As indirect terms become much smaller in S_{ij}, we obtain a greater discrimination ratio, Δ_S. (**h**) The degree of silencing, κ, captures the increase observed in the discrimination ratio by the transition from G_{ij} to S_{ij} (through equation (5)).

Equation (3) is exact and the sum accounts for all network paths connecting i and j (**Supplementary Note**, I.1–2). It is of limited use, however, as it requires us to solve N^2 coupled algebraic equations. In **Supplementary Note**, I.1, we show that equation (3) can be reformulated as

$$S = (G - I + \mathcal{D}(S \cdot G))G^{-1} \tag{4}$$

where I is the identity matrix and $\mathcal{D}(M)$ sets the off-diagonal terms of M to zero. To obtain an approximate solution for S, we use the fact that, typically, perturbations decay rapidly as they propagate through the network, so that the response observed between two nodes is dominated by the shortest path between them. This allows us to approximate $\mathcal{D}(S \cdot G)$ with $\mathcal{D}((G - I)G)$ (**Supplementary Note**, I.3), obtaining

$$S = (G - I + \mathcal{D}((G - I)G))G^{-1} \tag{5}$$

Equation (5), our main result, provides S_{ij} from the experimentally accessible G_{ij}. It achieves this through a 'silencing effect', in which direct response terms are preserved, whereas indirect responses are silenced. To understand this, consider a specific term in G_{ij}, documenting the response of node i to j's perturbation. As indicated by equation (3), this response is a consequence of all direct and indirect paths leading from j to i. As we document below, the transformation (5) detects the indirect paths and silences them, maintaining only the contribution of the direct paths (**Fig. 1d–f**). An alternative method to approximate $\mathcal{D}(S \cdot G)$ in equation (4) is using an iterative scheme, in which S_{ij} is evaluated first via equation (5) and then used as input in equation (4), repeating the process until sufficient accuracy is achieved (**Supplementary Note**, I.1).

Silencing in model systems

To demonstrate the predictive power of equation (5), we implemented Michaelis-Menten dynamics on a model network (**Supplementary Note**, III), as commonly used to model gene regulation[27,28]. We obtained G_{ij} by perturbing the activity of each node and then calculated S_{ij} using equation (5). **Figure 2a** shows the G_{ij} and S_{ij} terms associated with interacting and noninteracting node pairs. Although G_{ij} is higher for direct interactions, the overlap between the orange and the green symbols indicates a lack of a clear threshold q that separates direct and indirect interactions. In contrast, S_{ij} displays a clear separation between direct and indirect interactions, accurately predicting each direct link. Indeed, the receiver operating characteristic (ROC) curve derived from G_{ij} (**Fig. 2b**) has an area of AUROC = 0.91, reflecting inherent limitations in separating direct from indirect interactions based on G_{ij} only. In contrast, for S_{ij} we obtain AUROC = 0.997 (blue), where the true-positive rate reaches 100% with a false-positive rate of $<10^{-3}$. Also, as opposed to G_{ij}, for which precision increases gradually with the threshold q (**Fig. 2c**), S_{ij}'s precision jumps to 1 for $q > 10^{-4}$. Hence, in our well-controlled model system, any nonzero S_{ij} corresponds effectively to a direct link.

The performance of equation (5) is due to the silencing effect. It leaves G_{ij} unchanged if i and j are linked, whereas it systematically lowers all G_{ij} not rooted in a direct interaction. To quantify this effect we measured the discrimination ratio $\Delta_G = \langle G_{ij} \rangle_{\mathrm{Dir}} / \langle G_{ij} \rangle_{\mathrm{Indir}}$ ($\Delta_S = \langle S_{ij} \rangle_{\mathrm{Dir}} / \langle S_{ij} \rangle_{\mathrm{Indir}}$), which captures the ratio between G_{ij} (S_{ij}) terms associated with direct links and those associated with indirect links. We find that S_{ij} is much more discriminative than G_{ij} owing to its silencing of indirect responses. This effect can be quantitatively measured through the silencing metric

$$\kappa = \frac{\Delta_S}{\Delta_G} \tag{6}$$

which captures the increased power of S_{ij} to discriminate between direct and indirect links compared to G_{ij} (**Fig. 1h**). In our model system we find that $\kappa = 15$, a silencing of more than an order of magnitude (**Fig. 2d**). Furthermore, the longer the distance d_{ij} between two nodes, the larger is the silencing (**Fig. 2e**). As an illustration, consider a linear cascade in which changes in any node result in a finite response G_{ij} by all other nodes (**Fig. 2f**). Equation (5) silences all indirect responses, while leaving the response of direct links effectively unchanged, offering a discriminative measure that enables a perfect reconstruction of the original network.

Predicting molecular interactions in *E. coli*

To test the predictive power of equation (5) on real data, we used the *E. coli* data sets distributed by the DREAM5 network inference challenge[19]. The input data include a compendium of microarray

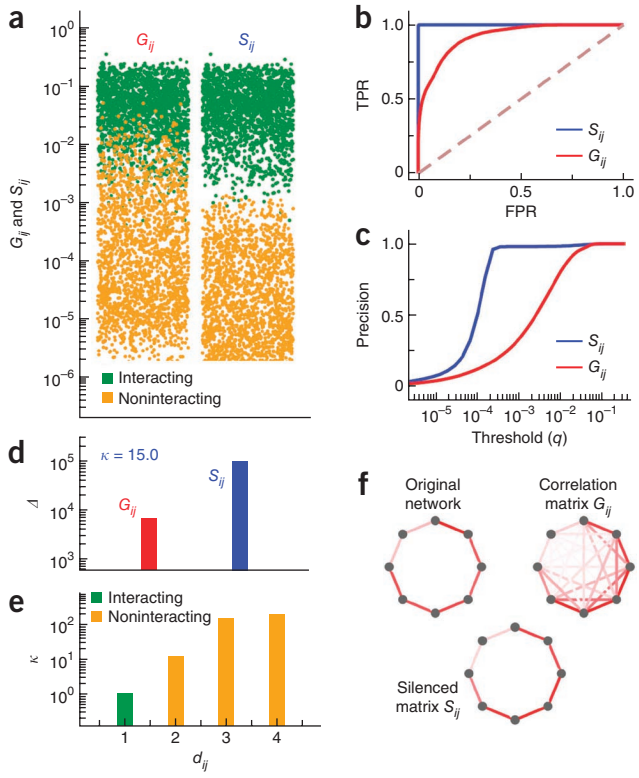

Figure 2 Network inference in model systems. We numerically simulated Michaelis-Menten dynamics on a scale-free network (refs. 40–42), extracting the correlations G_{ij} between all pairs of nodes (**Supplementary Note**, III). (**a**) G_{ij} and S_{ij} associated with interacting and noninteracting node pairs. S_{ij} silences the correlations associated with indirect interactions, resulting in a clear separation between direct and indirect interactions, a phenomenon absent from G_{ij}. (**b**) ROC curve obtained from G_{ij} (red, AUROC = 0.91) and S_{ij} (blue, AUROC = 0.997). The S_{ij} network reaches 100% accuracy with a negligible amount of false positives. TPR, true-positive rate; FPR, false-positive rate. (**c**) Precision obtained for threshold q for G_{ij} and S_{ij}. The gradual rise of the G_{ij}-based precision indicates that for a broad range of thresholds only a small fraction of the links will be identified. In contrast, the steep rise in precision for S_{ij} indicates its enhanced discriminative power between direct and indirect links; virtually any nonzero S_{ij} corresponds to a directly interacting pair. (**d**) The discrimination ratio, Δ, is much higher in S_{ij} compared to G_{ij}. This indicates that S_{ij} is a much better predictor of direct versus indirect interactions. The silencing metric (6), which captures the increase in the discrimination ratio, is $\kappa = 15.0$. (**e**) Silencing increases with the path length d_{ij} between i and j, so that the more indirect the link, the more dramatic the silencing. (**f**) The source of S_{ij}'s success is the silencing effect, here illustrated on correlations measured for a linear cascade. The reconstruction of the cascade from G_{ij} is confounded by numerous nonvanishing indirect correlations. In S_{ij} the indirect correlations are silenced, providing a perfect reconstruction.

experiments measuring the expression levels of 4,511 *E. coli* genes (141 of which are known transcription factors) under 805 different experimental conditions (**Supplementary Note**, IV.1). We constructed three separate global response matrices G_{ij} between the 141 transcription factors and their 4,511 potential target genes, based on (i) Pearson correlations, (ii) Spearman rank correlations and (iii) mutual information, which are three commonly used methods for link detection (**Supplementary Note**, IV.3). From each of the three G_{ij} matrices, we obtained S_{ij} via equation (5), and compared the performance of G_{ij} with the pertinent S_{ij}. To validate our predictions we relied on the gold standard used in the DREAM5 challenge, consisting of 2,066 established gene regulatory interactions. Measuring AUROC from G_{ij} and S_{ij}, we found an improvement of 56% for Pearson correlations (**Fig. 3a**), 67% for Spearman rank correlations (**Fig. 3b**) and a smaller improvement of 6% for mutual information (**Fig. 3c**), allowing us to improve upon the top-performing inference methods[19].

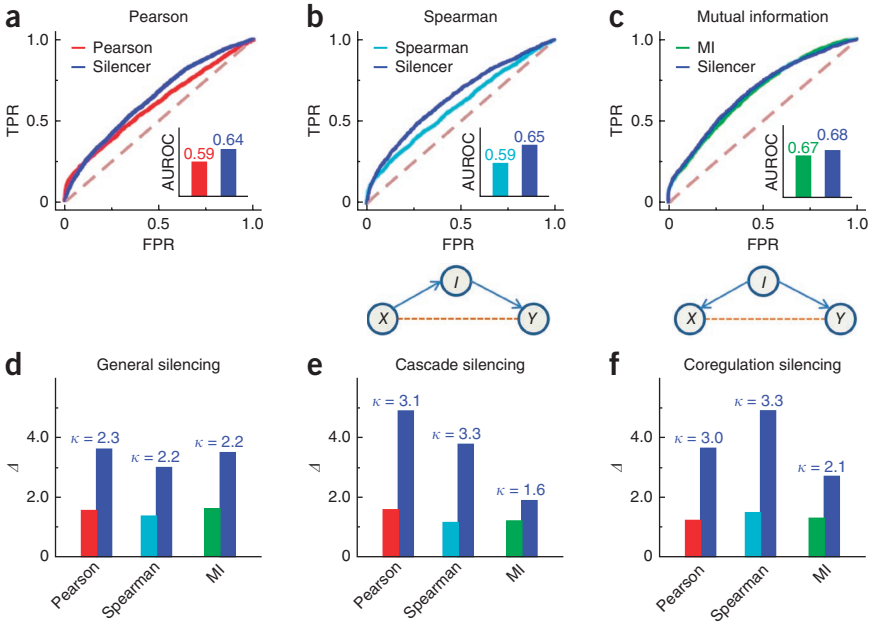

Figure 3 Inferring regulatory interactions in *E. coli*. (**a**) Starting from gene expression data, we used Pearson correlations in expression patterns to construct G_{ij} for 4,511 *E. coli* genes, obtaining S_{ij} via equation (5). We compared our predictions to a gold standard of experimentally verified genetic regulatory links[19]. The area under the ROC curve (AUROC) is increased from 0.59 to 0.64 in the transition from G_{ij} to S_{ij}, representing a 56% improvement (above the baseline of 0.5 for a random guess). TPR, true-positive rate; FPR, false-positive rate. (**b**) An improvement of 67% is observed for Spearman rank correlations. (**c**) A less dramatic improvement of 6% is shown when G_{ij} is constructed using mutual information (MI). (**d**) The discrimination ratio for all three methods compared with that obtained from the pertinent S_{ij} matrix. The transition to S_{ij} increases the discrimination between direct and indirect interactions by a factor of 2 or more, so that indirect interactions have a considerably lower expression in S_{ij}. (**e**,**f**) This observation becomes even more dramatic when focusing on two specific motifs: cascades and co-regulators. In G_{ij} the indirect correlation between X and Y, which is induced by the intermediate node, I, may lead to the false prediction of the spurious X-Y link. Thanks to silencing, the discrimination between the direct and indirect links in these motifs is increased by a factor of 3 or more for Pearson and Spearman correlations, and by a factor of ~2 for mutual information.

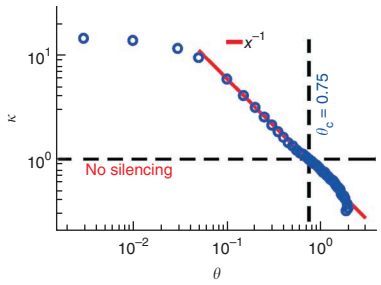

Figure 4 Silencing in a noisy environment. To test the method's performance in the presence of a noisy input we added Gaussian noise to the numerically obtained G_{ij}, and measured the silencing, κ, versus the signal-to-noise ratio θ. For low noise levels ($\theta \lesssim 0.1$), silencing is relatively unharmed. At higher noise levels, silencing decreases as $\kappa \sim \theta^{-1}$, a slow decay that supports the robustness of the method. Silencing is lost at $\theta_C \approx 0.75$, when the signal is almost fully driven by the noise.

We further tested the discrimination ratio, Δ, and the silencing, κ, for each of these methods, finding that indirect correlations are subject to an average of twofold silencing in the transition from G_{ij} to S_{ij} (**Fig. 3d**). Silencing is especially crucial in the presence of the cascade and co-regulation motifs (**Fig. 3e**,**f**), where most inference methods indicate a spurious link between X and Y, owing to the indirect correlation mediated by node I. Indeed, the transformation (5) silences these indirect correlations by a factor of three or more for Pearson and Spearman correlations and by a smaller factor (1.6 or 2.1) for mutual information, overcoming one of the most common hurdles of inference methods, which tend to over-represent triadic motifs[19].

The effects of noise and uncertainty

As all experimental data are subject to noise, the global response matrix, G_{ij}, is characterized by some degree of uncertainty. To test the performance of our methodology in the presence of noise, we repeated the numerical experiment of **Figure 2**, this time adding Gaussian noise to G_{ij}, which allows us to calculate silencing as a function of increasing the signal-to-noise ratio θ (**Fig. 4**). As expected, silencing is unaffected by small values of θ, so that κ features a plateau below $\theta \lesssim 0.1$. For large θ, silencing decays as $\kappa \sim \theta^{-1}$, demonstrating that the performance of the method decreases slowly as the signal-to-noise ratio is increased. Indeed, as opposed to a rapid exponential decay, the observed, slower, power-law dependence indicates that the method is rather tolerant to noise. Silencing is lost only when the noise reaches the critical level $\theta_C \approx 0.75$, when the signal is almost completely overridden by noise, leading to $\kappa = 1$ (**Supplementary Note**, V.1).

Hidden nodes offer another source of uncertainty. They represent the fact that in most cases we are unable to read the states of all nodes in the system[29]. To illustrate the effect of the hidden nodes on the performance of the silencing method, we consider the case of a simple cascade $i \to k \to j$, where the intermediate node k is hidden. In this scenario, equation (5) will not be able to silence the indirect $i \to j$ link, because in the observable system, the G_{ij} term cannot be attributed to any indirect path. Hence, absent any other information about the system, it is mathematically impossible to infer the indirectness of G_{ij}, as the removal of k isolated i from j[30]. This touches upon the fundamental mechanism of silencing: the silencing transformation (5) exploits the flow of information through indirect paths (**Fig. 1** and **Supplementary Note**, I.2). Consequently, if as a result of hidden nodes, the network fragments into several components such that the node pair i and j become isolated from each other, then all indirect paths between them became hidden and the pertinent G_{ij} term will not be silenced (**Fig. 5a**,**b**). Hence silencing is expected to fail only when the network breaks into many isolated components so that most node pairs become isolated. Fortunately, a fundamental property of complex networks is that with average degree $\langle k \rangle \gg 1$, one needs to remove a large fraction of the nodes to fragment the underlying giant connected component[31–34]. Therefore we can build on percolation theory, which allows us to analytically predict how the size of the largest connected component changes with the random removal of a certain fraction of nodes[35,36]. The calculation shows that silencing is maintained as long as the fraction of hidden nodes is smaller than

$$\eta_C \approx 1 - \frac{\Omega}{\langle k \rangle} \tag{7}$$

where $\Omega = \sqrt{2}\ln(\sqrt{2}+2) \approx 1.7$ (**Supplementary Note**, V.2). This equation indicates that for large $\langle k \rangle$ the method will be reliable even if a large fraction of the nodes are hidden.

To test this prediction, we revisited the numerically obtained G_{ij} analyzed in **Figure 2** and measured the degree of silencing after randomly removing an increasing fraction of nodes. In each case we also

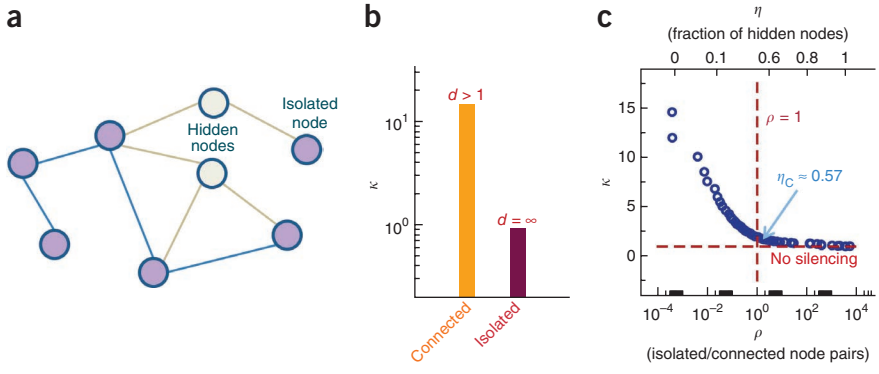

Figure 5 Performance with hidden nodes. (**a**) A network with $N = 8$ nodes, of which a fraction $\eta = 1/4$ are hidden. The observable subnetwork has 6 nodes, 5 forming a connected component (with 10 connected node pairs) and 1 isolated (6 isolated pairs). The ratio between isolated and connected node pairs here is $\rho = 6/10$. Equation (5), applied to the observable network, successfully silences the indirect G_{ij} terms among the nodes of the connected component. However, the correlations between the isolated node and the rest of the network, lacking an indirect path, are not silenced. (**b**) To test the silencing in the presence of hidden nodes, we used the numerically obtained G_{ij} (**Fig. 2**) from which we eliminated a fraction η of the nodes, obtaining an observable network with 10^4 isolated node pairs ($\rho \approx 10^{-3}$). After applying equation (5) to the remaining nodes, we found that the silencing of G_{ij} terms associated with connected node pairs is unaffected (orange bar), whereas for the isolated node pairs, silencing drops to $\kappa = 1$, namely no silencing (purple bar). Hence, for the isolated node pairs, S_{ij} is not more predictive than G_{ij}. (**c**) Increasing the fraction of hidden nodes, η (top horizontal axis), we measured κ versus ρ. As expected, silencing is observed as long as most node pairs are connected via finite paths ($\rho < 1$). However, when the number of hidden nodes is increased to the point that the isolated pairs dominate ($\rho > 1$), silencing is no longer observed ($\kappa = 1$). The critical fraction of hidden nodes, η_C, corresponds to $\rho = 1$, the point at which silencing no longer plays an important role. Here we find $\eta_C \approx 0.57$ (blue arrow), in agreement with the prediction of equation (7).

measured the ratio between isolated and connected node pairs (ρ). We found that, as predicted, the degree of silencing is driven mainly by ρ, approaching $\kappa \approx 1$ (no silencing) when $\rho \geq 1$, namely, when the isolated pairs begin to dominate the network (**Fig. 5c**). Here as $\langle k \rangle = 4$, equation (7) predicts $\eta_C \approx 0.57$, that is, the method will fail only when almost 60% of the nodes are hidden. Note that for biological networks, $\langle k \rangle$ is expected to be in the range of $\langle k \rangle \gtrsim 10$ (ref. 37), predicting $\eta_C \gtrsim 0.8$. Namely, one needs to lose access to 80% of the nodes for silencing to lose its effectiveness.

DISCUSSION

With computational complexity $\mathcal{O}(N^3)$, equation (5) is scalable and requires no assumptions about the network topology. By silencing indirect effects, it turns the raw correlation data into a predictive S_{ij} matrix, dominated by direct interactions. It is especially suited to treat perturbation data, such as genetic perturbation experiments, in which case G_{ij} describes the response of all genes (dx_i) as a consequence of the perturbation of the source gene (dx_j)[38]. In practice, however, G_{ij} could be the result of a broader set of experimental realizations where other measures are used to evaluate the association between nodes, typically statistical measures such as Pearson or Spearman correlation coefficients. Still, our empirical results (**Fig. 3**) clearly show that the transformation (5) successfully applies to these empirically accessible measures as well. Hence, silencing is largely insensitive to the specific process by which G_{ij} was constructed.

The method's broad applicability is rooted in the fact that it does not depend on the value of each specific term in G_{ij}, but rather on the global relationships between them. Indeed, the global structure of G_{ij} reflects the patterns of propagation of the perturbations along the network. Equation (5) helps uncover these paths from the raw data, disentangling the direct from the indirect effects. These patterns of information flow are inherent to the underlying network structure and should not depend on the specific experimental realization of equation (1). For instance, a cascade $i \rightarrow j \rightarrow k$ will be characterized by a decreasing correlation propagating along the arrows, a large correlation between i and j and a weaker one between i and k. Although the magnitude of these correlations might depend on the size or the form of i's perturbation as well as on the statistical measure we used to evaluate them, the decay pattern required to infer the structure of the cascade is an inherent property of the network flow and can be successfully detected by the silencing method (**Supplementary Note**, I.4).

The silencing transformation is derived from fundamental mathematical principles of dynamical correlations in networks. Hence it is expected to apply under rather general conditions. However, as equation (5) indicates, it requires that the input matrix, G_{ij}, is invertible. This imposes some limitations when constructed from statistical correlation measures. For instance, in the empirical results of **Figure 3a**, we constructed G_{ij} from Pearson correlations, using the states of 4,511 nodes measured under 805 experimental conditions. In general, if the number of experimental conditions is smaller than the number of nodes, the resulting Pearson correlation matrix may be singular. In this case, additional processing will be required before equation (5) can be applied. In this work, following the DREAM5 protocol, we only focused on the correlations between the 141 known transcription factors and the rest of the nodes, which lead to an invertible G_{ij} (**Supplementary Note**, IV). Other means to ensure G_{ij}'s invertibility are discussed in **Supplementary Note**, IV.4.

Isolating indirect effects in correlation data, a fundamental challenge of network inference, is typically approached through local probabilistic tools[12,14–18]. In contrast, the success of the silencing method is rooted in its exploitation of the global network topology[39]. It relies on the fundamental principles of network structure and dynamics to identify and silence the effects of indirect paths. The ability to extract S_{ij} from G_{ij} could also have implications for our understanding of network dynamics. Indeed, G_{ij} is a global network measure, as its magnitude is determined by the numerous indirect paths connecting i and j. Hence, for a given dynamics, the G_{ij} matrix will take a different form depending on the network topology, making it a poor predictor of the system's dynamics. By eliminating indirect effects, S_{ij} measures the effect gene i would have on gene j had they been isolated from the rest of the network. It thus helps us quantify the dynamical mechanism that governs individual pairwise interactions, avoiding the convolution of dynamical and topological effects present in experimental data. For instance, consider a set of perturbation experiments providing G_{ij}. The structure of G_{ij} reflects the microscopic mechanisms that govern the pairwise interactions, for example, genetic regulation and biochemical processes. It is difficult, however to extract this information from G_{ij} because its terms are a convolution of many interactions, reflecting the many paths leading from i to j. The transition to S_{ij}, via equation (5), allows us to treat each isolated interaction on its own, providing a direct observation of the microscopic interaction mechanism. Direct application of this fact could be the derivation of a rate equation that governs the system's dynamics from G_{ij}, as well as predicting the universality class and the scaling laws governing the system's response to perturbations. Hence equation (5) helps translate the ever-growing amount of data on global correlations into valuable local information.

METHODS

Methods and any associated references are available in the online version of the paper.

Note: Supplementary information is available in the online version of the paper.

ACKNOWLEDGMENTS

We thank B. Alipanahi and B. Frey for their valuable insights, A. Sharma, F. Simini, J. Menche, S. Rabello, G. Ghoshal, Y.-Y. Liu, T. Jia, M. Pósfai, C. Song, Y.-Y. Ahn, N. Blumm, D. Wang, Z. Qu, M. Schich, D. Ghiassian, S. Gil, P. Hövel, J. Gao, M. Kitsak, M. Martino, R. Sinatra, G. Tsekenis, L. Chi, B. Gabriel, Q. Jin and Y. Li for discussions, and S.S. Aleva, S. Morrison, J. De Nicolo and A. Pawling for their support. This work was supported by the US National Institutes of Health (NIH), Center of Excellence of Genomic Science (CEGS), Grant number NIH CEGS 1P50HG4233; and the NIH, award number 1U01HL108630-01; DARPA Grant Number 11645021; DARPA Social Media in Strategic Communications project under agreement number W911NF-12-C-0028; the Network Science Collaborative Technology Alliance sponsored by the US Army Research Laboratory under agreement number NS-CTA W911NF-09-02-0053; the Office of Naval Research under agreement number N000141010968; and the Defense Threat Reduction Agency awards WMD BRBAA07-J-2-0035 and BRBAA08-Per4-C-2-0033.

AUTHOR CONTRIBUTIONS

Both authors designed the research and wrote the paper. B.B. analyzed the empirical data, and did the analytical and numerical calculations.

COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

1. Buchanan, M., Caldarelli, G., De Los Rios, P., Rao, F. & Vendruscolo, M. (eds). *Networks in Cell Biology* (Cambridge University Press, 2010).
2. Ideker, T. & Sharan, R. Protein networks in disease. *Genome Res.* **18**, 644–652 (2008).
3. Kann, M.G. Protein interactions and disease: computational approaches to uncover the etiology of diseases. *Brief. Bioinform.* **8**, 333–346 (2007).
4. Albert, R. Scale-free networks in cell biology. *J. Cell Sci.* **118**, 4947–4957 (2005).

5. Barabási, A.-L. & Oltvai, Z.N. Network biology: understanding the cell's functional organization. *Nat. Rev. Genet.* **5**, 101–113 (2004).
6. Vidal, M., Cusick, M.E. & Barabási, A.-L. Interactome networks and human disease. *Cell* **144**, 986–998 (2011).
7. Rual, J.F. *et al.* Towards a proteome-scale map of the human protein-protein interaction network. *Nature* **437**, 1173–1178 (2005).
8. Yu, H. *et al.* High-quality binary protein interaction map of the yeast interactome network. *Science* **322**, 104–110 (2008).
9. Braun, P. *et al.* An experimentally derived confidence score for binary protein-protein interactions. *Nat. Methods* **6**, 91–97 (2009).
10. Krogan, N.J. *et al.* Global landscape of protein complexes in the yeast *Saccharomyces cerevisiae*. *Nature* **440**, 637–643 (2006).
11. Costanzo, M. *et al.* The genetic landscape of a cell. *Science* **327**, 425–431 (2010).
12. Ramani, A.K. *et al.* A map of human protein interactions derived from co-expression of human mRNAs and their orthologs. *Mol. Syst. Biol.* **4**, 180–195 (2008).
13. Barzel, B. & Biham, O. Quantifying the connectivity of a network: the network correlation function method. *Phys. Rev. E* **80**, 046104 (2009).
14. Eisen, M.B. *et al.* Cluster analysis and display of genome-wide expression patterns. *Proc. Natl. Acad. Sci. USA* **95**, 14863–14868 (1998).
15. Butte, A.J. & Kohane, I.S. Mutual information relevance networks: functional genomic clustering using pairwise entropy measurements. *Pac. Symp. Biocomput.* **5**, 415–426 (2000).
16. Margolin, A.A. *et al.* ARACNE: an algorithm for the reconstruction of gene regulatory networks in a mammalian cellular context. *BMC Bioinformatics* **7**, S7 (2006).
17. Guo, S. *et al.* Uncovering interactions in the frequency domain. *PLoS Comput. Biol.* **4**, e1000087 (2008).
18. Faith, J.J. *et al.* Large-scale mapping and validation of *Escherichia coli* transcriptional regulation from a compendium of expression profiles. *PLoS Biol.* **5**, e8 (2007).
19. Marbach, D. *et al.* Wisdom of crowds for robust gene network inference. *Nat. Methods* **9**, 796–804 (2012).
20. Lezon, T.R. *et al.* Using the principle of entropy maximization to infer genetic interaction networks from gene expression patterns. *Proc. Natl. Acad. Sci. USA* **103**, 19033–19038 (2006).
21. Ma, S. *et al.* An *Arabidopsis* gene network based on the graphical Gaussian model. *Genome Res.* **17**, 1614–1625 (2007).
22. Han, L. & Zhu, J. Using matrix of thresholding partial correlation coefficients to infer regulatory network. *Biosystems* **91**, 158–165 (2008).
23. Chen, L. & Zheng, S. Studying alternative splicing regulatory networks through partial correlation analysis. *Genome Biol.* **10**, R3 (2009).
24. Peng, J. *et al.* Partial correlation estimation by joint sparse regression models. *J. Am. Stat. Assoc.* **104**, 735–746 (2009).
25. Yuan, Y. *et al.* Directed Partial Correlation: inferring large-scale gene regulatory network through induced topology disruptions. *PLoS ONE* **6**, e16835 (2011).
26. Adamic, L.A. & Adar, E. Friends and neighbors on the web. *Soc. Networks* **25**, 211–230 (2003).
27. Alon, U. *An Introduction to Systems Biology: Design Principles of Biological Circuits* (Chapman & Hall, London, 2006).
28. Karlebach, G. & Shamir, R. Modeling and analysis of gene regulatory networks. *Nat. Rev. Mol. Cell Biol.* **9**, 770–780 (2008).
29. Caldarelli, G., Capocci, A., De Los Rios, P. & Muñoz, M.A. Scale-free networks from varying vertex intrinsic fitness. *Phys. Rev. Lett.* **89**, 258702 (2002).
30. Liu, Y.-Y., Slotine, J.-J. & Barabási, A.-L. Observability of complex systems. *Proc. Natl. Acad. Sci. USA* **110**, 2460–2465 (2013).
31. Erdős, P. & Rényi, A. On the evolution of random graphs. *Publications Math. Inst. Hungarian Acad. Sci.* **5**, 17–61 (1960).
32. Albert, R., Jeong, H. & Barabási, A.-L. Error and attack tolerance of complex networks. *Nature* **406**, 378–382 (2000).
33. Cohen, R., Erez, K., Ben-Avraham, D. & Havlin, S. Resilience of the Internet to random breakdowns. *Phys. Rev. Lett.* **85**, 4626–4628 (2000).
34. Bollobás, B. The Evolution of Random Graphs—the Giant Component. in *Random Graphs* 2nd ed. (Cambridge University Press, 2001).
35. Stauffer, D. & Aharony, A. *Introduction to Percolation Theory* (CRC Press, 1994).
36. Cohen, R. & Havlin, S. *Complex Networks: Structure, Robustness and Function* (Cambridge University Press, 2010).
37. Venkatesan, K. *et al.* An empirical framework for binary interactome mapping. *Nat. Methods* **6**, 83–90 (2009).
38. Kauffman, S. The ensemble approach to understand genetic regulatory networks. *Physica A* **340**, 733–740 (2004).
39. Marks, D.S., Hopf, T.A. & Sander, C. Protein structure prediction from sequence variation. *Nat. Biotechnol.* **30**, 1072–1080 (2012).
40. Barabási, A.-L. & Albert, R. Emergence of scaling in random networks. *Science* **286**, 509–512 (1999).
41. Albert, R. & Barabási, A.-L. Statistical mechanics of complex networks. *Rev. Mod. Phys.* **74**, 47–97 (2002).
42. Caldarelli, G. *Scale-free Networks* (Oxford University Press, 2007).