

ARTICLE

DOI: 10.1038/s41467-017-00587-4

OPEN

The interdependent network of gene regulation and metabolism is robust where it needs to be

David F. Klosik¹, Anne Grimbs ², Stefan Bornholdt¹ & Marc-Thorsten Hütt²

Despite being highly interdependent, the major biochemical networks of the living cell—the networks of interacting genes and of metabolic reactions, respectively—have been approached mostly as separate systems so far. Recently, a framework for interdependent networks has emerged in the context of statistical physics. In a first quantitative application of this framework to systems biology, here we study the interdependent network of gene regulation and metabolism for the model organism *Escherichia coli* in terms of a biologically motivated percolation model. Particularly, we approach the system's conflicting tasks of reacting rapidly to (internal and external) perturbations, while being robust to minor environmental fluctuations. Considering its response to perturbations that are localized with respect to functional criteria, we find the interdependent system to be sensitive to gene regulatory and protein-level perturbations, yet robust against metabolic changes. We expect this approach to be applicable to a range of other interdependent networks.

¹Institute for Theoretical Physics, University of Bremen, Hochschulring 18, 28359 Bremen, Germany. ²Department of Life Sciences and Chemistry, Jacobs University, Campus Ring 1, 28759 Bremen, Germany. Correspondence and requests for materials should be addressed to S.B. (email: bornholdt@itp.uni-bremen.de) or to M.-T.H. (email: m.huett@jacobs-university.de)

A main conceptual approach of current research in the life sciences is to advance from a detailed analysis of individual molecular components and processes towards a description of biological systems and to understand the emergence of biological function from the interdependencies on the molecular level. Supported by the diverse high-throughput ‘omics’ technologies, the relatively recent discipline of systems biology has been the major driving force behind this new perspective, which becomes manifest, for example, in the effort to compile extensive databases of biological information to be used in genome-scale models^{1–3}. Despite its holistic ‘game plan’, however, systems biology frequently operates on the level of subsystems: Even when considering cell-wide transcriptional regulatory networks, as, e.g., in network motif analysis⁴, this is only one of the cell’s networks. Likewise, the popular approach to studying metabolic networks in systems biology, constraint-based modelling, accounts for steady-state predictions of metabolic fluxes of genome-scale metabolic networks⁵, which again, is only one of the other networks of the cell.

In the analysis of such large networks, systems biology draws its tools considerably from the science of complex networks, which, by combining the mathematical subdiscipline of graph theory with methods from statistical physics, greatly contributed to the understanding of, e.g., the percolation properties of networks⁶, potential processes of network formation⁷ or the spreading of disease on networks⁸. In the early 2000s, gene regulation and metabolism have been among the first applications of ‘network biology’⁹. The most prominent findings on the gene regulatory side concerned the statistical observation and functional interpretation of small over-represented subgraphs (network motifs)^{10, 11} and the hierarchical organization of gene regulatory networks¹². On the metabolic side, the broad degree distribution of metabolic networks stands out¹³, with the caveat, however, that ‘currency metabolites’ (as ATP and H₂O) can severely affect network properties¹⁴, as well as the hierarchical modular organization of metabolic networks^{15, 16}.

Recently, the field of complex networks moved its focus from single networks to the interplay of networks that interact with and/or depend on each other (multilayer networks, networks of networks). Strikingly, it turned out that explicit interdependency between network constituents can fundamentally alter the percolation properties of the resulting interdependent networks, which can show a discontinuous percolation transition in contrast to the continuous behavior in single-network percolation^{17–23}. Also, contrary to the isolated-network case, networks with broader degree distribution become remarkably fragile as interdependent networks¹⁷.

However, this set of recent developments in network science about fragility due to interdependence still lacks application to systems biology. In Reis et al.²⁴ the question of robustness in multilayer biological networks has been raised for the first time (see also Bianconi²⁵). Specifically, it has been shown how specific correlations of the intra- and interlayer connections reduce cascading failures and thus provide a robust multilayer network architecture. Relevant progress in the application of concepts of multilayer networks has been made, for instance, in transportation infrastructures²⁶ and brain networks²⁴. In such applications, the discovery of new mechanisms went along with advances in the theoretical foundation in exploring dynamical processes in multilayer networks, for example, diffusion processes²⁷, spreading processes²⁸ and message passing^{29–31}.

Arguably, the most prominent representative of interdependent networks in a biological cell is the combined system of gene regulation and metabolism, which are interconnected by various forms of protein interactions, e.g., enzyme catalysis of biochemical reactions couples the regulatory to metabolic

network, while the activation or deactivation of transcription factors by certain metabolic compounds provides a coupling in the opposite direction.

Although it is well-known that gene regulatory and metabolic processes are highly dependent on one another, only few studies addressed their interplay on a larger, systematic scale^{32–34}. The first two studies (Covert et al.³² and Shlomi et al.³³) aimed at finding consistent metabolic-regulatory steady states by translating the influence of metabolic processes on gene activity into metabolic flux predicates and incorporating high-throughput gene expression data. This can be considered as an extension of the constraints-based modelling framework beyond the metabolic subsystem. In the paper of Samal and Jain³⁴, on the other hand, the transcriptional regulatory network of *Escherichia coli* (*E. coli*) metabolism has been studied as a Boolean network model with flux predicates represented by additional interactions.

The formalism of interdependent networks now allows us to go beyond these important pioneering works on integrative models, by analysing the robustness of the combined system in terms of the maximal effect of a small perturbation. In particular, the findings can be interpreted in the context of cascading failures and percolation theory.

We here formulate a first application of the new methodological perspective to the combined networks of gene regulation and metabolism in *E. coli*. Using various biological databases, particularly EcoCyc^{35, 36} as the main core, we have compiled a graph representation of gene regulatory and metabolic processes of *E. coli*, including a high level of detail in the structural description, distinguishing between a comparatively large number of vertex and edge types according to their biological functionality.

A structural analysis of this compilation reveals that, in addition to a small set of direct edges, the gene regulatory and the metabolic domains are predominantly coupled via a third network domain consisting of proteins and their interactions. This rich structural description, together with a purpose-built, biologically motivated percolation model allows us to assess the functional level with methods derived from percolation theory. More precisely, we investigate cascading failures in the three-domain system, emanating from small perturbations, localized in one of the domains. We observe below that (i) randomized versions of the graph are much less robust than the original graph and (ii) that the integrative system is much more susceptible to small perturbations in the gene regulatory domain than in the metabolic one.

Results

The core object of our investigation is an *E. coli* network representation of its combined gene regulation and metabolism, which can be thought of as functionally divided into three domains. Aside from the obvious domains emerging from gene regulatory and metabolic processes, these processes are connected by an intermediate domain that models both, the enzymatic influence of genes on metabolic processes, as well as signalling-effects of the metabolism on the activation or inhibition of the expression of certain genes. Accordingly, the underlying interaction graph $G = (V, E) = \{G_R, G_I, G_M\}$ with its set of vertices V and edges E consists of three interconnected subgraphs, the gene regulatory domain G_R , the interface domain G_I and the metabolic domain G_M . A sketch of the network model and some of its properties are presented in Fig. 1. Further information and a detailed characterization are given in the ‘‘Methods’’ section (as well as in A.G., D.F.K., S.B., and M.-T.H., manuscript in preparation).

Of central importance for modelling spreading dynamics in our system are the functionally different roles of the edges in the

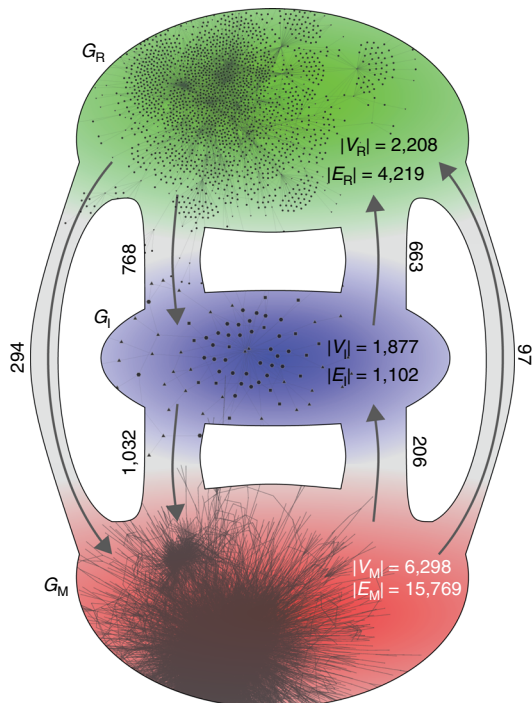


Fig. 1 Sketch of the domain organization of the integrative *E. coli* network. The gene regulatory domain, G_R , (top) is predominantly connected to the metabolic domain, G_M , (bottom) via a protein-interface layer, G_I . For each domain, the number of vertices, $|V|$, and the number of edges, $|E|$, are given as well as the number of inter-domain edges (attached to the arrows). For illustrative purposes, snapshots of the largest weakly-connected components of G_R , G_I and G_M are included

system, which we capture in three logical categories of an edge (LCE):

C, ‘conjunct’: The target vertex of an edge with this logical AND property depends on the source vertex, i.e., it will fail once the source vertex fails. For example, for a reaction to take place, all of its educts have to be available.

D, ‘disjunct’: Edges with this logical OR property are considered redundant in the sense that a vertex only fails if the source vertices of all of its incoming D-edges fail. For instance, a compound will only become unavailable once all of its producing reactions have been canceled.

R, ‘regulation’: Edges of this category cover 14 different kinds of regulatory events (described in detail in Supplementary Note 2). As shown below, in terms of the propagation dynamics we treat these edges similar to the ‘conjunct’ ones.

As outlined in the “Methods” section the conjunct and disjunct edge properties in our study have been derived from the underlying biological mechanisms. The disjunct edges may be examples of the redundant interdependencies, which have been discussed in Radicchi and Bianconi³¹ as a possible foundation of the robustness of multilayer (specifically, multiplex) networks.

In order to assess the effect of perturbations in this three-domain organized system, we devise a biologically motivated percolation model that determines the sets of potentially affected vertices (see “Methods” section). An illustration of the dynamics is shown in Fig. 2.

As a side remark, the spreading of a perturbation according to the rules defined above could also be considered as an epidemic process with one set of connections with a very large, and a second set of connections with a very low probability of infection³⁷.

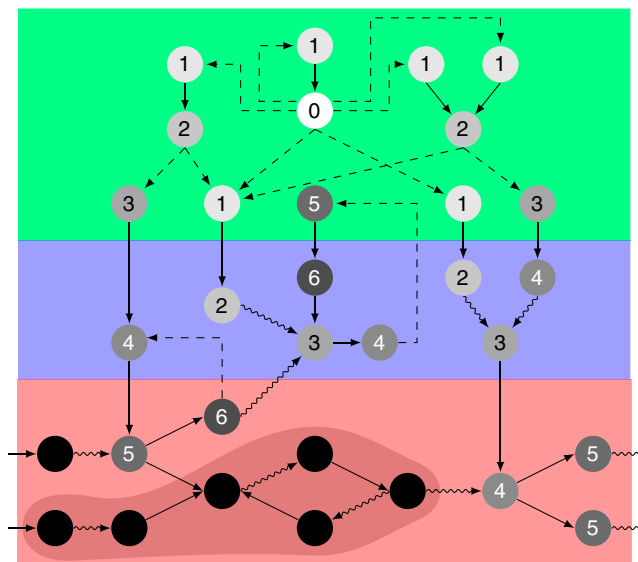


Fig. 2 Sketch of the propagation dynamics of a single-vertex perturbation. The perturbation in the small sample three-domain organized network with gene regulatory domain (top, green), (protein-)interface (middle, blue) and metabolic domain (bottom, red) is initiated in vertex 0 and spreads according to the percolation model given in Eqs. (1) and (2). The numbers and the grayscale indicate the time step in which the respective vertex has been turned off; black vertices are not affected. The arrow line styles denote the three logical categories of edges: C: curly, D: solid, and R: dashed. In the subsequent steps of the analysis, we consider the largest weakly connected component of the frozen network state (vertices in the dark area). A more detailed illustration is given in Supplementary Fig. 7

Comparison with previous approaches. In systems that can be described without explicit dependencies between its constituents, but with a notion of functionality that coincides with connectivity, percolation theory is a method of first choice to investigate the system’s response to perturbations of a given size that can be modelled as failing vertices or edges^{6, 38}. The fractional size of the giant connected component as a function of the occupation probability p of a constituent typically vanishes at some critical value p_c , the percolation threshold. In the following, we will mostly use the complementary quantity $q = 1 - p$ so that $q_c = 1 - p_c$ can be interpreted as the critical size of the initial attack or perturbation of the system. The strong fluctuations of the system’s responses in the vicinity of this point can serve as a proxy for the percolation threshold, which is especially useful in finite systems in which the transition appears smoothed out. In our analysis, the susceptibility $\tilde{\chi} = \langle S^2 \rangle - \langle S \rangle^2$, where S is the size of the largest cluster, is used³⁹.

Upon the introduction of explicit dependencies between the system’s constituents, the percolation properties can change dramatically. The order parameter no longer vanishes continuously but typically jumps at p_c in a discontinuous transition^{18, 19} as cascades of failures fragment the system. A broader degree distribution now enhances a graph’s vulnerability to random failures, in opposition to the behavior in isolated graphs^{17, 40}. Details of the corresponding theoretical framework have been worked out by Parshani et al.¹⁸, Son et al.²⁰, Baxter et al.⁴¹ and more recently the notions of ‘networks of networks’^{42–45} and network recovery^{46, 47} have been included. There have also been attempts to capture this class of models in a general framework of multilayer networks⁴⁸. Important contributions to the theoretical understanding of percolation phenomena in multilayered structures come from studying variants of message passing algorithms on multiplex networks. Cellai et al.²⁹, for example, show how the

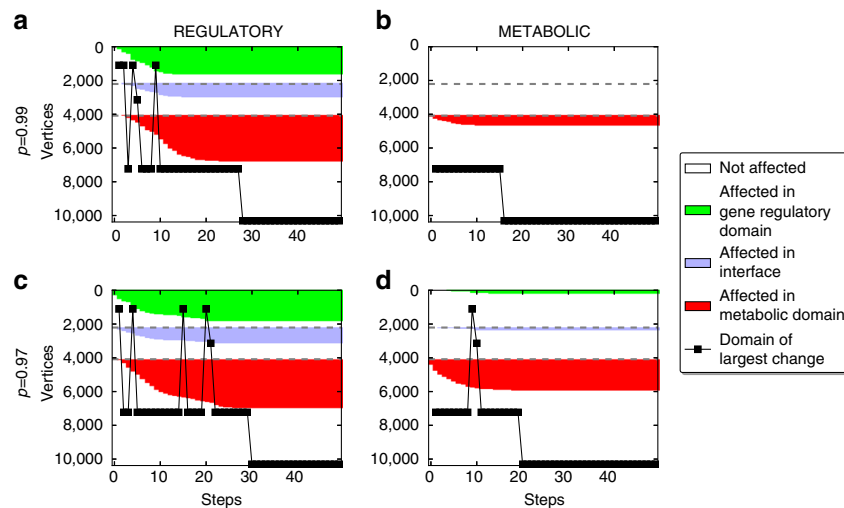


Fig. 3 Sample trajectories of the integrative *E. coli* network. Initial perturbations differ in size, $q = 1 - p = 0.01$ (**a, b**) and $q = 0.03$ (**c, d**), as well as in localization: perturbations are initiated in the gene regulatory domain (**a, c**) and the metabolic domain (**b, d**). In each panel, the vertex states are plotted against the update steps according to Eqs. (1) and (2). The vertices are grouped according to the domain affiliation, starting from gene regulatory domain (top, green) via protein-interface domain (middle, blue) towards metabolic domain (bottom, red). Within each domain they are arranged with respect to the update step, in which they were affected in the respective run (vertex ordering within the domains is thus different from run to run). Affected vertices are shown in the color of their domain, unaffected vertices are shown last for each run and in white. The horizontal dashed lines denote the domain boundaries and the black square indicates the domain with the largest number of newly affected vertices in each step. Connections between black squares are only shown to guide the eye

presence of link overlap between layers can be considered in terms of two similar, but conceptually different, processes and analyse the behavior of the corresponding order parameters.

In addition to random vertex failure other procedures for initial vertex removal have been explored, e.g., vertex removal with respect to their degree (targeted attacks)⁴⁹ or localized attacks, for which currently two notions have been described. Localized attacks of the first sort are defined on spatially embedded networks and are ‘local’ with respect to a distance in this embedding, i.e., in a ‘geographical’ sense^{50, 51}. The second approach considers locality in terms of connectivity: around a randomly chosen root vertex, neighbors are removed layer by layer^{52, 53}. In contrast, as described below in our approach, attacks are localized with respect to the three network domains, while within the domains vertices are chosen randomly.

At this point, we would like to shortly comment on the applicability of the mathematical concepts of interdependent networks to real-world data. Clear conceptual categories such as the distinction between dependency edges and connectivity edges have been instrumental in gaining theoretical insight into the properties of interdependent systems. Similarly, in the theoretical work on multilayer networks (see, e.g., De Domenico et al.⁵⁴, Boccaletti et al.⁵⁵), it is assumed that the assignment of vertices and edges to layers is defined a priori. It is worth mentioning, however, that in order to adapt these frameworks to the situation at hand several of these categories require adaptation. We envision that for many real-life applications a certain diversity of vertex and edge attributes will be required and the multilayer structure will rather emerge from the arrangement of such different edge types among the different vertex types in the system. Effectively, some classes of edges may then represent simple connectivity, while others can rather be seen as dependency edges. In biology, such dependencies are typically mediated by specific molecules (e.g., a small metabolite affecting a transcription factor, or a gene encoding an enzyme catalysing a biochemical reaction). Such implementations of dependency edges are no longer just associations and it is hard to formally distinguish them from the functional edges.

With the explicitly alternating ‘percolation’ and ‘dependency’ steps in typical computational models of the kind originally introduced in Buldyrev et al.¹⁷ in mind we would like to point out that in our directed model edges of all types are updated in every time step. If the dependencies of a vertex are fulfilled (here: only intact C- and R-neighbors), it will only fail once it loses all connectivity (here: via D-edges), irrespective of the current component structure, which is only considered after the dynamics is frozen.

Network response to localized perturbations. The main feature of our reconstructed network, the three-domain structure based on the biological role of its constituents, allows us to study the influence of localizing the initial perturbation (see “Methods” section). Thus, although we will not focus on (topological) details of the graph here (which will be presented in (A.G., D.F.K., S.B., and M.-T.H., manuscript in preparation)), already from the vertex and edge counts in Fig. 1, we see that the domains are of different structure. While the regulatory and the metabolic subgraphs, G_R and G_M have average (internal) degrees of about 1.9 and 2.5, the interface subgraph, G_I , is very sparse with $\langle k \rangle \approx 0.6$ and we can expect it to be fragmented. Hence, in the following, we decide to only perturb in G_R and G_M .

In a first step, we sample some full cascade trajectories in order to check our expectation of different responses of the system to small perturbations applied in either G_R or G_M ; two rather large values of q are chosen and the raw number of unaffected vertices is logged during the cascade. Indeed, already this first approach implies a different robustness of the gene regulatory and the metabolic domains in terms of the transmission of perturbation cascades to the other domains. Cascades initiated in the metabolic domain of the network tend to be rather restricted to this domain, while the system seems much more susceptible to small perturbations applied in the gene regulatory domain. This effect can be seen both in the overall sizes of the aggregated cascades as well as in the domain which shows the largest change with respect to the previous time step (largest set of newly affected vertices), which we indicate by *black markers* in Fig. 3. More sample

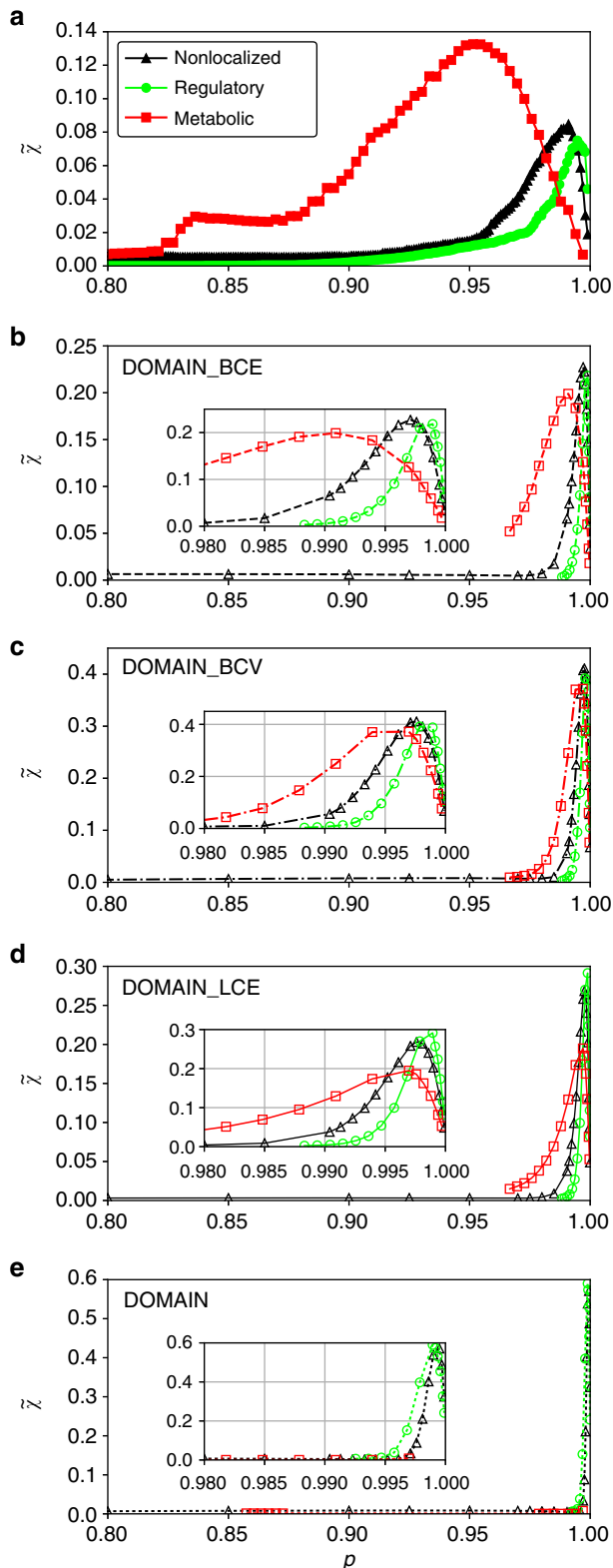


Fig. 4 Susceptibility of the integrative *E. coli* network and its randomized versions for distinct localized perturbations. In each panel, the average over 500 runs for an initial perturbation size, $q = 1 - p$, of the original network **a** and the four randomization schemes in decreasing stringency order **b–e** is shown. For each scheme 500 graph realizations have been considered. While non-localized perturbations are presented in black, perturbations initiated in gene regulatory domain and metabolic domain are given in green and in red, respectively

trajectories are shown in Supplementary Figs. 1–5 and, although they illustrate occasionally large fluctuations between the behaviors of single trajectories, they are consistent with this first observation. They also show that considerably larger metabolic perturbations are needed for large cascades and back-and-forth propagation between domains to emerge.

After this first glance at the system, we aim for a more systematic approach and apply our analysis as described above: we compute cascade steady-states σ_∞ but now we choose the largest (weakly) connected component $B(X, q)$ as the order parameter and compute the susceptibility according to Eq. (3), the peak-position of which, when considered as a function of $q = 1 - p$, we use as a proxy for the perturbation size at which the interdependent system breaks down.

The results for different initially perturbed domains illustrate that, indeed, a considerably lower p_c (i.e., larger critical perturbation size q_c) is estimated in the case of metabolic perturbations compared to regulatory or non-localized ones (Fig. 4a). For each point, we average 500 runs for the corresponding set of parameters.

In order to assess whether the above-described behavior is due to specific properties of the network, we use sets of randomized graphs. For each of the four randomization schemes, described in the “Methods” section, we prepared 500 graph instances and repeated the analysis for each of them as done before for the single original graph. The corresponding results for the susceptibility (Fig. 4b–e) yield two major observations: firstly, metabolic perturbations still lead to, albeit only slightly, higher $q_c = 1 - p_c$ estimates (with exception of DOMAIN randomization). Thus, the system’s tendency to be more robust towards metabolic perturbations is largely preserved. Secondly, we see that overall the original network seems to be much more robust than the randomized networks; very small perturbations are sufficient to break the latter ones. The robustness of the original graph, thus, cannot be solely due to the edge and vertex properties kept in the randomization schemes.

Finally, let us focus on the practical aspect of these findings. Beyond the careful statistical analysis described above, a quantity of practical relevance is the average size of the unaffected part of the system under a perturbation of given size q . For this purpose, we examine the fractions of unaffected vertices, $A(X, q)$, after cascades emanating from perturbations of different sizes and initiated in different domains, regardless of the resulting component structure and for both, the original graph and the shuffled ones (Fig. 5).

The number of unaffected vertices for the real network is much larger than for all four randomization schemes, suggesting a strong overall robustness of the biological system. Distinguishing, however, between the metabolic and the gene regulatory components reveals that the metabolic part is substantially more robust than the regulatory part (for not too large initial perturbations, $p > 0.94$).

Discussion

We investigated the spreading of perturbations through the three domains of a graph representation of the integrative system of *E. coli*’s gene regulation and metabolism. Our results quantify the resulting cascading failures as a function of size and localization of the initial perturbation.

Our findings show that the interdependent network of gene regulation and metabolism unites sensitivity and robustness by showing different magnitudes of damage dependent on the site of perturbation.

While the interdependent network of these two domains is in general much more robust than its randomized variants

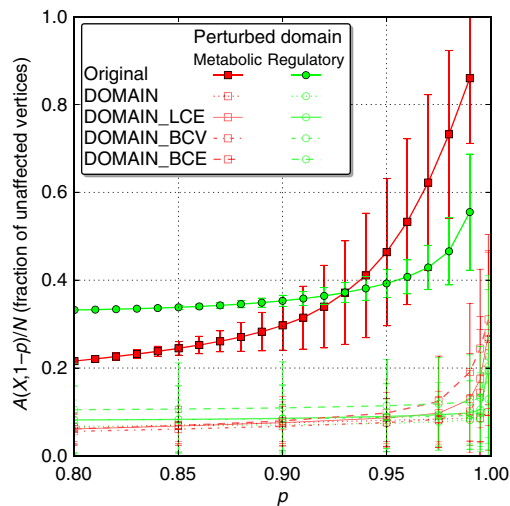


Fig. 5 Fractions of unaffected vertices of the integrative *E. coli* network and its randomized versions for distinct localized perturbations. Per initial perturbation size, $q = 1 - p$, 5000 runs were averaged for the original network (filled markers) and the four randomization schemes (open markers), where in turn, 500 realizations are considered for each scheme. Initial perturbations in the gene regulatory domain are presented as green dots, while perturbations initiated in the metabolic domain are given as red squares. The error bars indicate s.d. (results for randomized realizations have been combined)

(retaining domain structure, degree sequence, and major biological aspects of the original system), a pronounced difference between the gene regulatory and metabolic domain is found: Small perturbations originating in the gene regulatory domain typically trigger far-reaching system-wide cascades, while small perturbations in the metabolic domain tend to remain more local and trigger much smaller cascades of perturbations.

In order to arrive at a more mechanistic understanding of this statistical observation, we estimated the percolation threshold of the system, p_c , and found that it is much lower (i.e., larger perturbations, $q_c = 1 - p_c$, are required) for perturbations initiated in the metabolic domain than for those applied to the gene regulatory domain.

This is in accordance with the intuition that the metabolic system is more directly coupled to the environment (via the uptake and secretion of metabolic compounds) than the gene regulatory domain. The distinct percolation thresholds therefore allow for implementing a functionally relevant balance between robustness and sensitivity: The biological system can achieve a robustness towards environmental changes, while—via the more sensitive gene regulatory domain—it still reacts flexibly to other systemic perturbations.

Discovering this design principle of the biological system required establishing a novel method of analysing the robustness of interdependent networks, the network response to localized perturbations: An interdependent network can have markedly different percolation thresholds, when probed with perturbations localized in one network component compared to another.

Lastly, we would like to emphasize that the application of the theoretical concepts of interdependent networks to real-life systems involves several non-trivial decisions:

In the vast majority of theoretical investigations, interdependent networks are defined via a distinction between dependency edges and connectivity-representing edges^{17, 19, 56}. Often, in specific applications, this distinction of connectivity edges and dependency edges is not immediately clear (see

examples from neuroscience⁵⁷ and systems biology⁵⁸). In Morone et al.⁵⁹, an elegant way of mapping such real-life systems to the formal description of connectivity edges and dependency edges has been worked out for the case of brain networks by emphasizing the different roles of intramodule and intermodule edges.

Here the consideration of the biological data with respect to different levels of description (an overview of the corresponding vertex and edge categories is given in Supplementary Note 2) results in a graph representation with a functional three-domain organization: Vertices involved in gene regulation, metabolic vertices, and vertices associated with the (protein) interface between these two main domains are interconnected with (functionally) different classes of edges. These edge classes are a key ingredient of our biologically motivated percolation model. As a consequence, our investigation required an adjustment of the original percolation model described in Buldyrev et al.¹⁷ to the vertex and edge classes employed here. We expect that application of the notion of interdependent networks to real-life systems is likely to require adjustments of the percolation model, even though the main features of the original theory remain fully applicable.

As mentioned above, when dealing with (possibly incomplete) biological data, one major task is to find the right balance between radical simplifications of systemic descriptions and an appropriate level of detail, still allowing for a meaningful evaluation of the biological information, that is, to abstract from the minor but keep the essential details⁶⁰. We have outlined that, in this study, we chose to incorporate high level of detail in the structural description, distinguishing between a comparatively large number of vertex and edge types. This rich structural description, which includes information about the functional relationships between vertices allows us to assess the dynamical/functional level with the comparatively simple methods derived from percolation theory.

Even though the framework of multiplex networks does not strictly apply to the example of a multilayer network discussed here, as there is no one-to-one mapping from vertices in one layer to vertices in the other layers, several theoretical findings on multiplex networks nevertheless can help us to put the results of our investigation in a broader context. It has been recognized in a previous study³¹ that understanding the robustness of multilayer networks may require adapting the standard percolation model, where a vertex is switched off, if at least one of the interdependent vertices has been switched off as well. This percolation model is at the core of many of the investigations of catastrophic failures of interdependent networks, e.g., Buldyrev et al.¹⁷. In this scenario, a larger number of layers will typically lead to more vulnerable networks. The interesting question, how an increasing number of layers might actually enhance the robustness of the system, has been addressed in Radicchi and Bianconi³¹. In our case, the cascading failure, and hence the percolation model, is determined by the Boolean update rules in Eqs. (1) and (2), which have been derived from the underlying biological processes as described in the corresponding “Results” and “Methods” sections (and in more detail in the Supplementary Information).

This biologically motivated percolation model used here is comparable to the model of percolation in multiplex networks introduced in Radicchi and Bianconi³¹. Even though intracellular molecular network systems, as the one studied here, tend to not adhere to the formal requirement of a one-to-one mapping of vertices between different layers, it is likely that the main finding from³¹—that an addition of layers enhances systemic robustness by providing redundant interdependencies—may still hold. It would be interesting to further explore in what way the multilayer structure of intracellular molecular networks indeed enhances their robustness in this way.

Furthermore, an important question is, whether the analysis of the fragmentation of such a network under random removal of vertices can provide a reliable assessment of functional properties, since the response of such a molecular network clearly follows far more intricate dynamical rules than the percolation of perturbations can suggest.

A future step could include the construction of a Boolean network model for the full transcriptional regulatory network and the connection of this model to flux predictions obtained via flux balance analysis, a first attempt of which is given in Samal and Jain³⁴ (where the model of Covert et al.³² with still fewer inter-dependence edges has been used).

Our perturbation spreading approach might help bridging the gap between theoretical concepts from statistical physics and biological data integration: Integrating diverse biological information into networks, estimating ‘response patterns’ to systemic perturbations and understanding the multiple systemic manifestations of perturbed, pathological states is perceived as the main challenge in systems medicine (see, e.g., Bauer et al.⁶¹). Concepts from statistical physics of complex networks may be of enormous importance for this line of research^{62, 63}.

While the simulation of the full dynamics is still problematic as our knowledge of the networks is still incomplete, our present strategy extracts first dynamical properties of the interdependent networks. At a later time point, we can expect qualitative advances from full dynamical simulations, however, dependent on the quality of the data sets.

On the theoretical side, future studies might shift the focus onto recasting the system into an appropriate spreading model, e.g., in the form of an unordered binary avalanche^{64, 65}, or as an instance of the Linear Threshold model³⁷ with a set of edges with a very high and a second set with a very low transmission probability (C/R- and D-edges, respectively).

Also, it would be interesting to investigate in more detail, whether the pattern of dependency edges, which in Reis et al.²⁴ has been associated with robustness of interdependent networks, are more prominent in the metabolism-to-regulation direction of our system than in the regulation-to-metabolism direction.

Radicchi²³ presents an approach for the investigation of the percolation properties of finite size interdependent networks with a specific adjacency matrix with the goal of loosening some of the assumptions underlying the usual models (e.g., infinite system limit, graphs as instances of network model). While this formalism allows for the investigation of many real-world systems, there are still restrictions as to the possible level of detail. In our special case, for instance, a considerable amount of information would be lost if the system was restricted to vertices with connections in both the C/R- and D-layers.

The existence of different percolation thresholds for localized perturbations in interdependent networks may reveal itself as a universal principle for balancing sensitivity and robustness in complex systems. The application of these concepts to a wide range of real-life systems is required to make progress in this direction.

Methods

Network. The integrative network of *E. coli* has been assembled using the EcoCyc database^{35, 36} (version 18.5), which offers both the data about metabolic processes and (gene) regulatory events incorporated from the RegulonDB⁶⁶ (release 8.7). The extensive metadata allows for the assignment of the vertices to one of the three functional domains. Details of this process and a detailed characterization of the resulting model are described in (A.G., D.F.K., S.B., and M.-T.H., manuscript in preparation). The corresponding graph representation consists of 10,383 vertices and 24,150 directed edges.

Since we are interested in the propagation of a signal between the domains, in the following we will refer to the domains of the source and target vertices of the edge $e_k = (v_s^{(k)}, v_t^{(k)})$ as source domain, SD, and target domain, TD, respectively.

The metadata can be used to assign properties to the vertices and edges of the graph beyond the domain structure, some of which are used in the following analysis, namely in the compilation of the percolation model and the randomization schemes.

We distinguish between biological categories of edges (capturing the diverse biological roles of the edges) and the logical categories (determining the rules of the percolation process). According to their biological role in the system, both vertices and edges are assigned to a biological category; we abbreviate the biological category of a vertex as BCV and the biological category of an edge as BCE (for details see Supplementary Note 2). Each of the eight BCEs can then be mapped uniquely to one of only three LCE,

$$e_k^{\text{LCE}} \in \{C, D, R\}.$$

Percolation model. In the dynamical rules describing the propagation of an initial perturbation in the network, we distinguish between the different roles a given edge has in the update of a target vertex in terms of the LCE.

In our percolation model, every vertex is assigned a Boolean state variable $\sigma \in \{0, 1\}$; since we intend to mimic the propagation of a perturbation (rather than simulate a trajectory of actual biological states), we identify the state 1 with ‘not yet affected by the perturbation’, while the state 0 corresponds to ‘affected by the perturbation’. We stress that the trajectory $\sigma(t)$ does not correspond to the time evolution of the abundance of gene products and metabolic compounds, but the rules have been chosen such that the final set of affected vertices provides an estimate of all vertices potentially being affected by the initial perturbation. A vertex not in this set is topologically very unlikely of being affected by the perturbation at hand (given the biological processes contained in our model).

A stepwise update can now be defined for vertex i with in-neighbors Γ_i^- in order to study the spreading of perturbations through the system by initially switching off a fraction q of vertices:

$$\sigma_i(t + 1) = f_i(\sigma_j(t) | \sigma_j \in \Gamma_i^-) \tag{1}$$

$$f_i = \begin{cases} 1 & \text{if } \left[\sum_j c_{ij}(1 - \sigma_j) = 0 \right] \wedge \\ & \left[\sum_j d_{ij}\sigma_j > 0 \vee \sum_j d_{ij} = 0 \right] \wedge \\ & \left[\sum_j |r_{ij}|(1 - \sigma_j) = 0 \right] \\ 0 & \text{otherwise} \end{cases} \tag{2}$$

where c_{ij} is 1 if v_j is connected to v_i via a C-edge and 0 otherwise; d_{ij} and r_{ij} are defined analogously.

Thus, a vertex will be considered unaffected by the perturbation if none of its in-neighbors connected via either a C- or an R-edge have failed (regardless of the type of regulatory interaction, activation or inhibition), and at least one of its in-neighbors connected via a D-edge is still intact. With an additional rule, it is ensured that an initially switched off vertex stays off. The choice of the update rules ensures that the unperturbed system state is conserved under the dynamics: $f(\mathbf{1}) = \mathbf{1}$.

Network response to localized perturbation analysis. Due to the functional three-domain partition of our *E. coli* gene regulatory and metabolic network reconstruction, we have the possibility to classify perturbations not only according to their size, but also with respect to their localization in one of the domains comprising the full interdependent system, thereby enabling us to address the balance of sensitivity and robustness of the interdependent network of gene regulation and metabolism.

Here we introduce the concept of network response to localized perturbations analysis. This analysis will reveal that perturbations in gene regulation affect the system in a dramatically different way than perturbations in metabolism. Thus, we study the response to localized perturbations, which we denote by $\text{Per}(X, q)$, where X is the domain, in which the perturbation is localized ($X \in \{R, I, M, T\}$, with T representing the total network G , i.e., the case of non-localized perturbations). The perturbation size $q = 1 - p$ is measured in fractions of the total network size $N = |G|$. Hence, $\text{Per}(M, 0.1)$ is a perturbation in the metabolic domain with (on average) $0.1|G|$ vertices initially affected. Note that sizes q of such localized perturbations are limited by the domain sizes, e.g., $q|G| < |G_R|$ for a perturbation in the gene regulatory domain.

After the initial perturbation of a fraction q of the vertices in the domain X the stepwise dynamics described above will lead to the deactivation of further vertices and run into a frozen state σ_∞ , in which only a fraction $A(X, q)$ of the vertices are unaffected by the dynamics (i.e., are still in state 1). In addition to being directly affected by failing neighbors, in the process of network fragmentation vertices may also become parts of small components disconnected from the network’s core, and could in this sense be considered non-functional; we therefore also monitor the

relative sizes of the largest (weakly) connected component in the frozen state, $B(X, q)$, for different initial perturbation sizes.

In the limit of infinite system size, we could expect a direct investigation of $B(X, q)$ as a function of q to yield the critical perturbation size $q_c = 1 - p_c$ at which B vanishes. In our finite system, however, we have to estimate q_c ; following Radicchi and Castellano³⁹, and Radicchi⁶⁷, we measure the fluctuation of $B(X, q)$ which serves as our order parameter and look for the peak position of the susceptibility

$$\bar{\chi}(X, q) = \langle B_\infty^2 \rangle - \langle B_\infty \rangle^2 \quad (3)$$

as a function of parameter q in order to estimate the transition point from the finite system data. Supplementary Fig. 6 schematically illustrates the analysis (also see Supplementary Note 1).

Randomization schemes. In order to interpret the actual responses of a given network to perturbations, one usually contrasts them to those of suitably randomized versions of the network at hand. Thereby, the often dominant effect of the vertex degree distribution of a network can be accounted for and the effects of higher-order topological features that shape the response of the network to perturbations can be studied systematically.

The same is true for the localized perturbation response analysis introduced here. In fact, due to the substantially larger number of edges from gene regulation to metabolism (both, directly and via the interface component of the interdependent network) than from metabolism to gene regulation we can already expect the response to such localized perturbations to vary.

Here we employ a sequence of ever more stringent randomization schemes to generate sets of randomized networks serving as null models for the localized perturbation response analysis. In all of the four schemes, the edge-switching procedure introduced by Maslov and Sneppen⁶⁸ is employed which conserves the in- and out-degrees of all vertices.

Our most flexible randomization scheme (DOMAIN) only considers the domains of the source and target vertices of an edge (SD and TD): only pairs of edges are flipped which share both, the source and the target domain (e.g., both link a vertex in the metabolic domain to a vertex in the interface). The remaining three randomization schemes all add an additional constraint. The DOMAIN_LCE randomization further requires the edges to be of the same logical categories of an edge (i.e., C, D, or R), while the DOMAIN_BCV scheme only switches edges whose target vertices also share the same biological category of a vertex, BCV. The strictest randomization, DOMAIN_BCE, finally, only considers edges with, additionally, the same biological category of an edge, BCE. A tabular overview of the four schemes is given in Supplementary Table 1.

Data availability. The data that support the findings of this study are available from the authors upon reasonable request.

Received: 8 February 2017 Accepted: 11 July 2017

Published online: 14 September 2017

References

- Kitano, H. Systems biology: a brief overview. *Science* **295**, 1662–1664 (2002).
- Ideker, T., Galitski, T. & Hood, L. A new approach to decoding life: Systems biology. *Annu. Rev. Genomics Hum. Genet.* **2**, 343–372 (2001).
- Aderem, A. Systems biology: Its practice and challenges. *Cell* **121**, 511–513 (2005).
- Milo, R. et al. Network motifs: Simple building blocks of complex networks. *Science* **298**, 824–827 (2002).
- Llaneras, F. & Pic, J. Stoichiometric modelling of cell metabolism. *J. Biosci. Bioeng.* **105**, 1–11 (2008).
- Cohen, R., ben Avraham, D. & Havlin, S. Percolation critical exponents in scale-free networks. *Phys. Rev. E* **66**, 036113 (2002).
- Dorogovtsev, S. N. & Mendes, J. F. F. Evolution of networks. *Adv. Phys.* **51**, 1079–1187 (2002).
- Newman, M. E. J. Spread of epidemic disease on networks. *Phys. Rev. E* **66**, 016128 (2002).
- Barabási, A.-L. & Oltvai, Z. N. Network biology: Understanding the cell's functional organization. *Nat. Rev. Genet.* **5**, 101–113 (2004).
- Shen-Orr, S., Milo, R., Mangan, S. & Alon, U. Network motifs in the transcriptional regulation network of *Escherichia coli*. *Nat. Genet.* **31**, 64–68 (2002).
- Alon, U. Network motifs: Theory and experimental approaches. *Nat. Rev. Genet.* **8**, 450–461 (2007).
- Yu, H. & Gerstein, M. Genomic analysis of the hierarchical structure of regulatory networks. *Proc. Natl Acad. Sci. USA* **103**, 14724–14731 (2006).
- Jeong, H., Tombor, B., Albert, R., Oltvai, Z. N. & Barabási, A.-L. The large-scale organization of metabolic networks. *Nature* **407**, 651–654 (2000).
- Ma, H. & Zeng, A.-P. Reconstruction of metabolic networks from genome data and analysis of their global structure for various organisms. *Bioinformatics* **19**, 270–277 (2003).
- Ravasz, E., Somera, A. L., Mongru, D. A., Oltvai, Z. N. & Barabási, A. L. Hierarchical organization of modularity in metabolic networks. *Science* **297**, 1551–1555 (2002).
- Guimera, R. & Amaral, L. Functional cartography of complex metabolic networks. *Nature* **433**, 895–900 (2005).
- Buldyrev, S. V., Parshani, R., Paul, G., Stanley, H. E. & Havlin, S. Catastrophic cascade of failures in interdependent networks. *Nature* **464**, 1025–1028 (2010).
- Parshani, R., Buldyrev, S. V. & Havlin, S. Interdependent networks: Reducing the coupling strength leads to a change from a first to second order percolation transition. *Phys. Rev. Lett.* **105**, 048701 (2010).
- Parshani, R., Buldyrev, S. V. & Havlin, S. Critical effect of dependency groups on the function of networks. *Proc. Natl Acad. Sci. USA* **108**, 1007–1010 (2011).
- Son, S.-W., Bizhani, G., Christensen, C., Grassberger, P. & Paczuski, M. Percolation theory on interdependent networks based on epidemic spreading. *EPL* **97**, 16006 (2012).
- Radicchi, F. & Arenas, A. Abrupt transition in the structural quality of interconnected networks. *Nat. Phys.* **9**, 717–720 (2013).
- Zhou, D. et al. Simultaneous first- and second-order percolation transitions in interdependent networks. *Phys. Rev. E* **90**, 012803 (2014).
- Radicchi, F. Percolation in real interdependent networks. *Nat. Phys.* **11**, 597–602 (2015).
- Reis, S. D. et al. Avoiding catastrophic failure in correlated networks of networks. *Nat. Phys.* **10**, 762–767 (2014).
- Bianconi, G. Multilayer networks: Dangerous liaisons? *Nat. Phys.* **10**, 712–714 (2014).
- De Domenico, M., Solé-Ribalta, A., Gómez, S. & Arenas, A. Navigability of interconnected networks under random failures. *Proc. Natl Acad. Sci. USA* **111**, 8351–8356 (2014).
- Gomez, S. et al. Diffusion dynamics on multiplex networks. *Phys. Rev. Lett.* **110**, 028701 (2013).
- De Domenico, M., Granello, C., Porter, M. A. & Arenas, A. The physics of spreading processes in multilayer networks. *Nat. Phys.* **12**, 901–906 (2016).
- Cellai, D., Dorogovtsev, S. N. & Bianconi, G. Message passing theory for percolation models on multiplex networks with link overlap. *Phys. Rev. E* **94**, 032301 (2016).
- Bianconi, G. & Radicchi, F. Percolation in real multiplex networks. *Phys. Rev. E* **94**, 060301 (2016).
- Radicchi, F. & Bianconi, G. Redundant interdependencies boost the robustness of multiplex networks. *Phys. Rev. X* **7**, 011013 (2017).
- Covert, M. W., Knight, E. M., Reed, J. L., Herrgard, M. J. & Palsson, B. O. Integrating high-throughput and computational data elucidates bacterial networks. *Nature* **429**, 92–96 (2004).
- Shlomi, T., Eisenberg, Y., Sharan, R. & Ruppin, E. A genome-scale computational study of the interplay between transcriptional regulation and metabolism. *Mol. Syst. Biol.* **3**, 101 (2007).
- Samal, A. & Jain, S. The regulatory network of *E. coli* metabolism as a Boolean dynamical system exhibits both homeostasis and flexibility of response. *BMC. Syst. Biol.* **2**, 21 (2008).
- Keseler, I. M. et al. EcoCyc: A comprehensive database resource for *Escherichia coli*. *Nucleic Acids Res.* **33**, D334–D337 (2005).
- Keseler, I. M. et al. EcoCyc: Fusing model organism databases with systems biology. *Nucleic Acids Res.* **41**, D605–D612 (2013).
- Watts, D. J. A simple model of global cascades on random networks. *Proc. Natl Acad. Sci. USA* **99**, 5766–5771 (2002).
- Callaway, D. S., Newman, M. E. J., Strogatz, S. H. & Watts, D. J. Network robustness and fragility: Percolation on random graphs. *Phys. Rev. Lett.* **85**, 5468–5471 (2000).
- Radicchi, F. & Castellano, C. Beyond the locally treelike approximation for percolation on real networks. *Phys. Rev. E* **93**, 030302 (2016).
- Albert, R., Jeong, H. & Barabási, A.-L. Error and attack tolerance of complex networks. *Nature* **406**, 378–382 (2000).
- Baxter, G. J., Dorogovtsev, S. N., Goltsev, A. V. & Mendes, J. F. F. Avalanche collapse of interdependent networks. *Phys. Rev. Lett.* **109**, 248701 (2012).
- Gao, J., Buldyrev, S. V., Stanley, H. E. & Havlin, S. Networks formed from interdependent networks. *Nat. Phys.* **8**, 40–48 (2011).
- Gao, J., Li, D. & Havlin, S. From a single network to a network of networks. *Natl Sci. Rev.* **1**, 346–356 (2014).
- Bianconi, G. & Dorogovtsev, S. N. Multiple percolation transitions in a configuration model of a network of networks. *Phys. Rev. E* **89**, 062814 (2014).
- Kenett, D. Y., Perc, M. & Boccaletti, S. Networks of networks - An introduction. *Chaos Solitons Fractals* **80**, 1–6 (2015).
- Majdandzic, A. et al. Spontaneous recovery in dynamical networks. *Nat. Phys.* **10**, 34–38 (2014).
- Majdandzic, A. et al. Multiple tipping points and optimal repairing in interacting networks. *Nat. Commun.* **7**, 10850 (2016).

48. Kivela, M. et al. Multilayer networks. *J. Comp. Netw* **2**, 203–271 (2014).
49. Huang, X., Gao, J., Buldyrev, S. V., Havlin, S. & Stanley, H. E. Robustness of interdependent networks under targeted attack. *Phys. Rev. E* **83**, 065101 (2011).
50. Berezin, Y., Bashan, A., Danziger, M. M., Li, D. & Havlin, S. Localized attacks on spatially embedded networks with dependencies. *Sci. Rep.* **5**, 8934 (2015).
51. Zhao, J., Li, D., Sanhedrai, H., Cohen, R. & Havlin, S. Spatio-temporal propagation of cascading overload failures in spatially embedded networks. *Nat. Commun.* **7**, 10094 (2016).
52. Yuan, X., Shao, S., Stanley, H. E. & Havlin, S. How breadth of degree distribution influences network robustness: Comparing localized and random attacks. *Phys. Rev. E* **92**, 032122 (2015).
53. Shao, S., Huang, X., Stanley, H. E. & Havlin, S. Percolation of localized attack on complex networks. *New. J. Phys.* **17**, 023049 (2015).
54. De Domenico, M. et al. Mathematical formulation of multilayer networks. *Phys. Rev. X* **3**, 041022 (2013).
55. Boccalletti, S. et al. The structure and dynamics of multilayer networks. *Phys. Rep.* **544**, 1–122 (2014).
56. Danziger, M. M., Shekhtman, L. M., Bashan, A., Berezin, Y. & Havlin, S. in *Interconnected Networks*, 79–99 (Springer, 2016).
57. Deco, G., Jirsa, V. K. & McIntosh, A. R. Emerging concepts for the dynamical organization of resting-state activity in the brain. *Nat. Rev. Neurosci.* **12**, 43–56 (2011).
58. Menche, J. et al. Uncovering disease-disease relationships through the incomplete interactome. *Science* **347**, 1257601 (2015).
59. Morone, F., Roth, K., Min, B., Stanley, H. E. & Makse, H. A. Model of brain activation predicts the neural collective influence map of the brain. *Proc. Natl Acad. Sci. USA* **114**, 3849–3854 (2017).
60. De Domenico, M., Nicosia, V., Arenas, A. & Latora, V. Structural reducibility of multilayer networks. *Nat. Commun.* **6**, 6864 (2015).
61. Bauer, C. R. et al. Interdisciplinary approach towards a systems medicine toolbox using the example of inflammatory diseases. *Brief. Bioinformatics* **18**, 479 (2016).
62. Barabási, A.-L., Gulbahce, N. & Loscalzo, J. Network medicine: a network-based approach to human disease. *Nat. Rev. Genet.* **12**, 56–68 (2011).
63. Hütt, M.-T. Understanding genetic variation—the value of systems biology. *Br. J. Clin. Pharmacol.* **77**, 597–605 (2014).
64. Samuelsson, B. & Socolar, J. E. S. Exhaustive percolation on random networks. *Phys. Rev. E* **74**, 036113 (2006).
65. Gleeson, J. P. Mean size of avalanches on directed random networks with arbitrary degree distributions. *Phys. Rev. E* **77**, 057101 (2008).
66. Salgado, H. et al. RegulonDB v8.0: omics data sets, evolutionary conservation, regulatory phrases, cross-validated gold standards and more. *Nucleic Acids Res.* **41**, D203–D213 (2013).

67. Radicchi, F. Predicting percolation thresholds in networks. *Phys. Rev. E* **91**, 010801 (2015).
68. Maslov, S. & Sneppen, K. Specificity and stability in topology of protein networks. *Science* **296**, 910–913 (2002).

Acknowledgements

S.B. and M.-T.H. acknowledge the support of Deutsche Forschungsgemeinschaft (DFG), grants BO 1242/6 and HU 937/9.

Author contributions

M.-T.H. and S.B. designed and supervised the study; D.K. and A.G. performed the reconstruction, simulations and analyses; and D.K., A.G., S.B. and M.-T.H. wrote the manuscript.

Additional information

Supplementary Information accompanies this paper at doi:10.1038/s41467-017-00587-4.

Competing interests: The authors declare no competing financial interests.

Reprints and permission information is available online at <http://npg.nature.com/reprintsandpermissions/>

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



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

© The Author(s) 2017