SCIENTIFIC REP**O**RTS

OPEN

Finding and analysing the minimum set of driver nodes required to control multilayer networks

Jose C. Nacher[1], Masayuki Ishitsuka[1], Shuichi Miyazaki[2] & Tatsuya Akutsu [3]

It is difficult to control multilayer networks in situations with real-world complexity. Here, we first define the multilayer control problem in terms of the minimum dominating set (MDS) controllability framework and mathematically demonstrate that simple formulas can be used to estimate the size of the minimum dominating set in multilayer (MDSM) complex networks. Second, we develop a new algorithm that efficiently identifies the MDSM in up to 6 layers, with several thousand nodes in each layer network. Interestingly, the findings reveal that the MDSM size for similar networks does not significantly differ from that required to control a single network. This result opens future directions for controlling, for example, multiple species by identifying a common set of enzymes or proteins for drug targeting. We apply our methods to 70 genome-wide metabolic networks across major plant lineages, unveiling some relationships between controllability in multilayer networks and metabolic functions at the genome scale.

In recent years, structural controllability and control theory approaches have been studied in depth in the context of linear and nonlinear complex systems and networks[1–4]. More recently, Liu *et al*. mapped the structural controllability problem into one that involves solving a maximum matching (MM) problem[5]. This approach has been used in several studies to investigate control features under different topologies[6,7]. One of the focuses in these studies is to analyze the number of driver nodes, where a set of nodes is called a set of driver nodes if the whole system can be driven from a given initial state to a given target state by applying control signals to only these nodes (see[5] for the details of the relationship between driver nodes and controllability). However, a large fraction of real-life processes and systems can be better represented by multilayer network structures than by single-layer networks[8,9]. Some groups have extended the MM framework and proposed others to investigate the controllability of multilayer networks. Yuan *et al*. studied the minimum number of driver nodes needed to fully control diffusion dynamics by using matrix computation, where controllers can interact with any layer[10]. Zhang *et al*. studied the controllable subspace of multilayer linear networks, where control signals are applied to only one layer[11]. Pósfai *et al*. studied the controllability of multilayer linear discrete systems with time delays, where control signals are also applied to only one layer[12]. They showed that the minimum number of driver nodes for such systems can be computed based on the maximum network flow. These approaches are useful for controlling multilayer networks in which different layers are connected by some dynamics.

However, there exist other situations in which multiple networks should be independently controlled using the same set of driver nodes. For example, consider the case of controlling biological systems. In such a case, there may exist some differences among networks according to individual differences or species differences, but we need to give the same set of controllers (e.g., the same drugs) or choose one from several sets of controllers. Clearly, for this type of approach, we do not need that these network layers are physically coupled. Menichetti *et al*. studied such type of multilayer network control[13] by extending the framework of the linear structural controllability of single-layer networks using the maximum matching approach[5].

Here, we use the minimum dominating set (MDS) approach[14] to investigate the controllability of multilayer networks instead of using the MM model. A set of nodes in an undirected network is called a dominating set if each node in the network belongs to the set or has a neighbour in the set. A dominating set with the minimum number of elements is called a minimum dominating set (MDS). It was shown in[14] that if every edge in a network

[1]Department of Information Science, Toho University, Funabashi, Chiba, 274-8510, Japan. [2]Academic Center for Computing Media Studies, Kyoto University, Kyoto, 606-8501, Japan. [3]Bioinformatics Center, Institute for Chemical Research, Kyoto University, Uji, 611-0011, Japan. Correspondence and requests for materials should be addressed to J.C.N. (email: nacher@is.sci.toho-u.ac.jp) or T.A. (email: takutsu@kuicr.kyoto-u.ac.jp)

is bi-directional and every node in a dominating set can control itself and all its outgoing edges individually, then the network is structurally controllable by selecting the nodes in a dominating set as the driver nodes. Therefore, an MDS corresponds to a minimum set of driver nodes under this framework. Since the underlying assumptions are different, there is no technical contradiction between both models. However, there are several reasons that motivated us to use the MDS model instead of the MM approach. First, while the MM approach can only guarantee its controllability for linear systems, which are not common in real problems, the MDS model can be applied to nonlinear systems[15], which exist in abundance in natural problems, because each node has at least one independent control input. Moreover, although the Feedback Vertex Set (FVS)-based control model can also handle non-linear systems, the target states are limited to steady states (including periodic ones)[4,16]. Second, although the MDS is an NP-hard problem, current integer linear programming solver and graph-reduction-based algorithms allow us to find an MDS for very large networks. We will see later that this is the case of the multilayer MDS (MDSM) problem. Third, for values of the power-law degree exponent close to $\gamma = 2$, the number of driver nodes identified using an MDS is much smaller than that using the MM approach[14]. Finally, the MDS approach has also been adopted by many different groups[17–19]; multiple biological systems have been studied using MDS, such as protein-protein interaction networks[17,20–23], drug-target networks[15], ncRNA-protein networks[24], drug-disease networks[25] and metabolic networks[26]; and relevant biological findings have been uncovered, such as the enrichment of cancer-related and virus-target genes within the MDS in protein networks[21]. Menichetti *et al.* extensively studied the distribution of the minimum number of driver nodes[13] based on linear structural controllability[5]; however, no studies on MDS-based controllability using multilayer networks have been conducted.

Here, we first mathematically provide new insights into the MDS controllability framework when multilayer networks are considered. Although the MDS is generally an NP-hard problem, we demonstrate that even in special cases of networks in which the MDS is solved in polynomial time, the multilayer MDS (MDSM) problem is still NP-hard. More importantly, by using a recursive probabilistic technique, we demonstrate that simple formulas can be used to estimate the MDSM size for *k*-regular random networks and maximally assortative scale-free networks. To our knowledge, these are the first results of giving simple formulas to estimate the number of driver nodes in multilayer networks. We also demonstrate that the size of the MDSM does not increase by much compared with that of the MDS for a single network if the difference among the layers in multilayer networks is small.

In spite of the NP-hardness of the multilayer MDS problem, we propose a new algorithm that efficiently computes an MDSM and is able to identify controllers in large-scale multilayer real-world networks. This method includes a novel preprocessing technique based on integer linear programming (ILP), which is our second main result. The algorithm is able to efficiently compute networks of up to 6 layers, with several thousand nodes in each network. Using this novel algorithm, we explore for the first time the optimal solution for the MDSM-based multilayer controllability problem for large genome-wide metabolic networks across major plant lineages. We also analytically and empirically show that the size of the MDSM tends to be close to that of the MDS when the two networks are similar. Finally—but most importantly—we validate the biological importance of the set of nodes included in the MDSMs by compiling data from 70 plant metabolic networks and computing the corresponding MDS and MDSM. The enrichments of the MDS and MDSM in each main metabolic pathway unveil for the first time a relationship between controllability in multilayer networks and metabolic functions at the genome scale.

Theoretical Results

In this work, we investigate the controllability of multilayer networks using a minimum dominating set (MDS)[14] approach and focus on a generic type of multilayer network constructed in a manner similar to that of multiplex networks (see Fig. 1). While the definition of a multiplex network states that, in each layer, the same set of nodes is connected by a different set of links[9,27], here, we allow each layer to consist of a different set of nodes. In other words, the sets of nodes in the multiple layers do not necessarily overlap completely. By considering similar type of networks, such as metabolic networks, a large fraction of nodes are overlapped. Moreover, the network layers are not physically coupled. This situation is often seen in molecular networks, for example, when comparing networks from different organisms that do not synthesize the same set of proteins or enzymes. Concepts closely related to multiplex networks have emerged rapidly and have led to a number of analytical and theoretical developments such as networks of networks, multidimensional networks, multilevel networks and interdependent networks, among others[8,27,28]. In the following, we present the main theoretical findings.

Upper bound estimation for the size of the MDSM.
Our main theoretical result is the development of methods for upper bound estimation of the size of the MDS for multilayer networks (i.e., the size of the MDSM).

Recall that for a graph $G(V, E)$, $U (U \subseteq V)$ is called a *dominating set* (DS) of G if

$$(\forall v \in V)(v \in U \vee (\exists u \in U)(\{u, v\} \in E))$$

holds. A DS with the minimum cardinality is called a *minimum dominating set* (MDS).

MDSM is defined by extending this definition of MDS. Let $\mathcal{G} = \{G_i(V_i, E_i) | i = 1, \ldots, N\}$ be a set (multiset) of undirected networks. That is, $\mathcal{G}$ corresponds to multilayer networks. Let $V = \bigcup_{i \in \{1,\ldots,N\}} V_i$. $U (U \subseteq V)$ is called a *dominating set for multilayer networks* $\mathcal{G}$ (DSM for $\mathcal{G}$) if

$$(\forall i \in \{1, \ldots, N\})(\forall v \in V_i)(v \in U \vee (\exists u \in U)(\{u, v\} \in E_i))$$

holds. A DSM with the minimum cardinality is called a *minimum dominating set for multilayer networks* (MDSM). Since an MDSM is also a DS for each G_i, if we select an MDSM as a set of driver nodes (with assuming that each driver node can control its links independently), every G_i becomes structurally controllable from the results in[14,15].
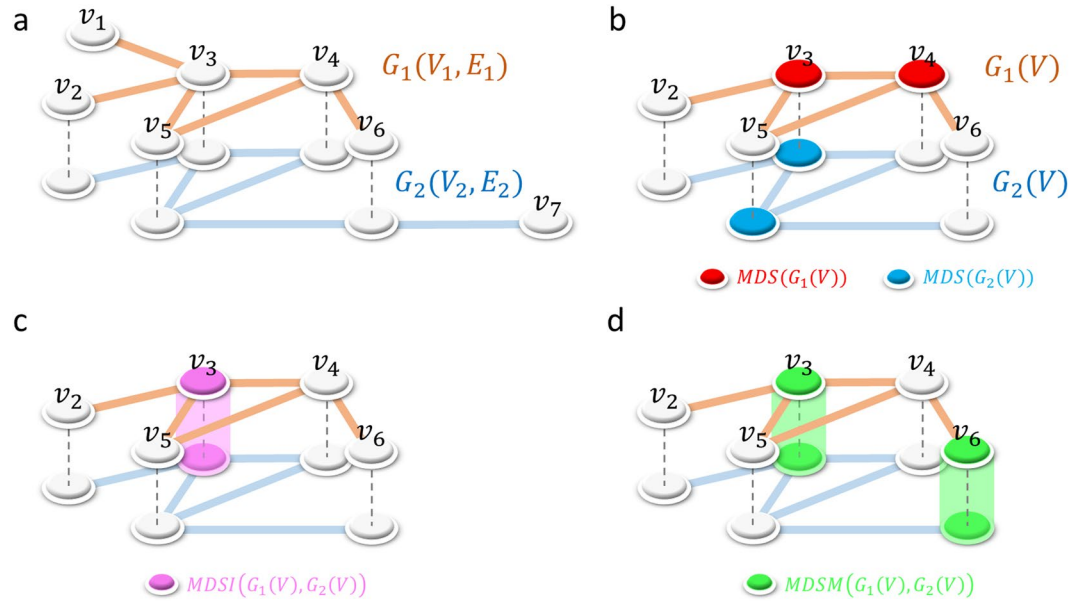
Figure 1. (**a**) A multilayer network defined by graphs $G_1(V_1, E_1)$ and $G_2(V_2, E_2)$. (**b**) The common nodes are defined as $V = V_1 \cap V_2$. From each graph G_1 and G_2, we construct the graphs $G_1(V)$ and $G_2(V)$ induced by the set of nodes V. Then, we compute $MDS(G_1(V))$ and $MDS(G_2(V))$. (**c**) The MDSI is defined as $MDS(G_1(V)) \cap MDS(G_2(V))$. (**d**) The set of nodes (two nodes) that simultaneously multicontrol both layers is defined as $MDSM(G_1(V), G_2(V))$. Note that although the MDSM is defined for any pair of graphs, $V = V_1 \cap V_2$ is considered in the analysis of metabolic networks. Note that the network layers are not physically coupled.

Next, we show a simple property. Let $V_{MDS}(G_i)$ be an MDS for G_i. Then, $S_{MDS}(G_i)$ denotes the MDS size for G_i (i.e., $S_{MDS}(G_i) = |V_{MDS}(G_i)|$). Let $V_{MDS}(\mathcal{G})$ and $S_{MDS}(\mathcal{G})$ denote an MDSM and its size for $\mathcal{G}$, respectively. Clearly, $S_{MDS}(\mathcal{G}) \geq S_{MDS}(G_i)$ holds for each G_i because an MDSM is also a DS for G_i. Conversely, $\cup_{i \in \{1,\dots,N\}} V_{MDS}(G_i)$ becomes a DSM for $\mathcal{G}$. Therefore, we have the following.

Proposition 1

$$\max_{i \in \{1,\dots,N\}}\{S_{MDS}(G_i)\} \leq S_{MDS}(\mathcal{G}) \leq \sum_{i=1}^{N} S_{MDS}(G_i).$$

Although this result is fairly obvious, it gives strict bounds in the worst case: (i) if G_is are identical, $S_{MDS}(G_i) = S_{MDS}(\mathcal{G})$ holds for all G_i; (ii) if V_is are disjoint, $S_{MDS}(\mathcal{G}) = \sum_{i=1}^{N} S_{MDS}(G_i)$ holds. This fact suggests that we should consider a special family of networks. Furthermore, as will be discussed later, the result on artificial networks suggests that the size of the MDSM for k-regular random networks is much smaller than this upper bound. Therefore, we theoretically explain this empirical finding.

We assume that graphs are given uniformly at random on the same set of nodes V with $|V| = n$ under the constraint that every node has degree k, with k being constant. Since a lot of theoretical studies have been done on these k-regular random graphs[29], it is reasonable to consider k-regular random graphs in order to investigate theoretical properties of an MDSM.

We utilize the recursive probabilistic estimation technique that was recently developed for analysis of the size of an MDS in k-regular random graphs[30], although our analysis needs additional ideas. It is shown in[30] that this technique yields very accurate estimates of the MDS size for random graphs. It is to be noted that this technique does not offer rigorous analysis methods but rather approximate analysis methods, as presented in many studies on complex networks using mean-field approximation, cavity methods, and so on.

Let $G(U)$ denote the subgraph of G that is induced by a set of vertices U, and let $N_G(U)$ denote the set of neighbours of U in G excluding U (i.e., $N_G(U) = \{v|\{u, v\} \in E, u \in U, v \notin U\}$). We consider the following virtual procedure that outputs a DSM for $\mathcal{G} = \{G_1, \dots, G_N\}$.

(i) Let DS_1 be the dominating set for G_1 obtained using the method in[30]. Let $V_1 \leftarrow V$.

(ii) For $i = 2$ to N, perform steps (iii)–(iv).

(iii) Let $V_i \leftarrow V - \bigcup_{j=1}^{i-1} DS_j - N_{G_i}\left(\bigcup_{j=1}^{i-1} DS_j\right)$. ($V_i$ is the set of vertices in G_i that are not dominated by a combined dominating set for $G_1, \dots, G_{i-1}$).

(iv) Let DS_i be the dominating set for $G_i(V_i)$, obtained using the method in[30].

(v) Output $DS_1 \cup DS_2 \cup \cdots \cup DS_N$ as a DSM.

It is obvious that this procedure outputs a correct DSM (but not necessarily an MDSM) for $\mathcal{G}$. By analysing this procedure, the size of the resulting DSM is estimated as $\alpha_N n$, where α_i is given by

$$\alpha_1 = \frac{1}{k+1},$$

$$\alpha_{i+1} = \alpha_i + \frac{1}{k+1}(1-\alpha_i)^{k+1}.$$

Although it only gives an estimate of the upper bound of the size of MDSM, comparison with the computational results (see Supplementary Information (SI)) suggests that this simple formula (for an upper bound) gives a very accurate estimate of the MDSM size for k-regular random networks. This estimation method can be modified for the analysis of maximally assortative scale-free networks in which the degree distribution follows a power law $\propto k^{-\gamma}$, where it is known that many real-world networks have both scale-free and assortative properties[30].

A network is called *maximally assortative* if an exchange of any pair of edges does not increase the assortative coefficient (see SI). It is shown in[30] that a maximally assortative network is approximately regarded as a collection of k-regular networks. By using the same virtual procedure as presented above, the size of the resulting DSM is estimated as $\beta_N n$, where β_i is given by

$$\beta_1 = \frac{\sum_{k=1}^{\infty} \frac{1}{k+1} \cdot k^{-\gamma}}{\sum_{k=1}^{\infty} k^{-\gamma}},$$

$$\beta_{i+1} = \beta_i + \left(\frac{\sum_{k=1}^{\infty} \left(\frac{1}{k+1}\right)(1-\beta_i)^{k+1} k^{-\gamma}}{\sum_{k=1}^{\infty} k^{-\gamma}} \right).$$

Comparison with the computational results (see SI) suggests that this formula gives a reasonable estimate of the MDSM size for maximally assortative scale-free networks.

Upper bound estimations and the result on artificial networks suggest that the MDSM size converges to n as the number of networks grows. We show below that this speculation is true (see SI for the proof).

Proposition 2

Suppose that each graph in multilayer networks has the minimum degree d_{min}. Then, the MDSM size is at least $n - d_{min}$ if a sufficient number of distinct graphs are given and n is sufficiently large.

Hardness of computation of MDSM. It is known that the maximum bipartite matching for networks of up to two layers can be determined in polynomial time, whereas the computation of such a matching for networks of three or more layers is NP-hard[31]. This fact suggests that the minimum set of driver nodes under linear structural controllability[5] can be obtained in polynomial time only for networks of up to two layers.

On the other hand, it is known that the computation of an MDS is NP-hard even for one network[32]. However, the situation changes if we consider special graph classes: it is known that an MDS can be computed in polynomial time if a given network is a partial k-tree (for a constant k)[32]. As a special case, it is seen that an MDS can be computed in polynomial time if networks are forests or consist of cycles and stars. We can show that the computation of an MDSM is NP-hard even for such simple networks (see SI for the proofs).

Theorem 3 The MDSM problem for two-layer networks is NP-hard even if a graph in each layer consists of cycles and at most one star.

Theorem 4 The MDSM problem for three-layer networks is NP-hard even if a graph in each layer does not contain any cycle.

These theorems suggest that the use of multilayers causes easy cases to be difficult.

Results on artificial networks. We performed numerical experiments to examine how the MDSM size changes as the number of wiring operations increases. Note that in this section, we consider similar networks, different from the Theoretical Results section. We employed the FAST-MDSM method mentioned in the Methods section using the SCIP ILP solver (http://scip.zib.de/).

We generated random graphs that follow the power-law degree distribution $k^{-\gamma}$, with each containing approximately $n = 5,000$ nodes. For each generated graph G_0, we applied K rewiring operations and obtained a graph G_K (see Methods and Fig. S1). Then, we computed $S_{MDS}(G_0, G_K)$ and $S_{MDS}(G_0)$ by using FAST-MDSM, where computation of an MDSM was completed in a few seconds per graph pair on a PC with an Intel Core (TM) i7-3517U CPU and 4 GB of RAM.

For each $\gamma = 2.1, 2.3, 2.5, 2.7$, we calculated the average ratio of $S_{MDS}(G_0, G_K)$ (obtained using FAST-MDSM) to $S_{MDS}(G_0)$ over 10 trials. The averaged results over 10 trials are shown in Fig. S2, from which it can be seen that the increase in the size of the MDSM is much smaller than K. In addition, we compared these ratios with those obtained via the theoretical estimate described in the Methods section. The results are shown in Fig. S3(a). We can see that the actual ratio does not linearly increase as N increases, whereas the theoretically estimated ratio increases almost linearly as N grows. We can also see that there is substantial discrepancy between the ratios obtained using FAST-MDS and those obtained using the theoretical estimate—especially for large K. This discrepancy is reasonable because the rewiring operations do not change the degree of each node and, thus, high degree nodes tend to remain in the MDSM.

We also compared the ratios obtained using FAST-MDSM with those obtained using the theoretical estimate described in the Methods section for the case of non-degree preservation modifications (i.e., insertion and

Plantcyc – Main component	boleracea_capitacyc	brapa_fpsccyc	oglaberrimacyc	oryzacyc	mosscyc	selaginellacyc	csubellipsoideacyc	mpusillacyc
Brassica oleracea var. capitata (boleracea_capitacyc)		Overlap : $|V|$ 1742; MDSM 266; MDSI 201	Overlap : $|V|$ 1515; MDSM 246; MDSI 141	Overlap : $|V|$ 1537; MDSM 248; MDSI 147	Overlap : $|V|$ 1359; MDSM 225; MDSI 128	Overlap : $|V|$ 1429; MDSM 237; MDSI 154	Overlap : $|V|$ 1140; MDSM 204; MDSI 110	Overlap : $|V|$ 1025; MDSM 185; MDSI 112
Brassica rapa FPsc (brapa_fpsccyc)	Network Size 1898; Computational Time 2458.5 ms		Overlap : $|V|$ 1548; MDSM 251; MDSI 165	Overlap : $|V|$ 1585; MDSM 256; MDSI 145	Overlap : $|V|$ 1403; MDSM 227; MDSI 134	Overlap : $|V|$ 1474; MDSM 240; MDSI 158	Overlap : $|V|$ 1175; MDSM 209; MDSI 126	Overlap : $|V|$ 1055; MDSM 188; MDSI 99
Oryza glaberrima (oglaberrimacyc)	Network Size 2024; Computational Time 8119.5 ms	Network Size 2047; Computational Time 4254.4 ms		Overlap : $|V|$ 1691; MDSM 269; MDSI 180	Overlap : $|V|$ 1384; MDSM 234; MDSI 118	Overlap : $|V|$ 1466; MDSM 242; MDSI 146	Overlap : $|V|$ 1180; MDSM 208; MDSI 122	Overlap : $|V|$ 1048; MDSM 187; MDSI 105
Oryza sativa japonica group (oryzacyc)	Network Size 2094; Computational Time 2282.8 ms	Network Size 2102; Computational Time 10368.2 ms	Network Size 1895; Computational Time 8294.2 ms		Overlap : $|V|$ 1427; MDSM 242; MDSI 123	Overlap : $|V|$ 1502; MDSM 252; MDSI 134	Overlap : $|V|$ 1204; MDSM 213; MDSI 113	Overlap : $|V|$ 1067; MDSM 188; MDSI 105
Physcomitrella patens (mosscyc)	Network Size 2139; Computational Time 15310.9 ms	Network Size 2151; Computational Time 4639.6 ms	Network Size 2069; Computational Time 52656.9 ms	Network Size 2118; Computational Time 109374.8 ms		Overlap : $|V|$ 1432; MDSM 237; MDSI 131	Overlap : $|V|$ 1197; MDSM 213; MDSI 108	Overlap : $|V|$ 1084; MDSM 193; MDSI 97
Selaginella moellendorffii (selaginellacyc)	Network Size 2010; Computational Time 3327.3 ms	Network Size 2021; Computational Time 14898.9 ms	Network Size 1928; Computational Time 1699.9 ms	Network Size 1984; Computational Time 11814.8 ms	Network Size 1921; Computational Time 3743.8 ms		Overlap : $|V|$ 1178; MDSM 207; MDSI 117	Overlap : $|V|$ 1065; MDSM 190; MDSI 95
Coccomyxa subellipsoidea C-169 (csubellipsoideacyc)	Network Size 2034; Computational Time 3318.9 ms	Network Size 2055; Computational Time 1323.0 ms	Network Size 1949; Computational Time 33578.6 ms	Network Size 2017; Computational Time 19373.0 ms	Network Size 1891; Computational Time 1877.4 ms	Network Size 1851; Computational Time 2247.2 ms		Overlap : $|V|$ 1043; MDSM 190; MDSI 125
Micromonas pusilla CCMP1545 (mpusillacyc)	Network Size 2015; Computational Time 804.8 ms	Network Size 2041; Computational Time 1155.3 ms	Network Size 1947; Computational Time 3130.6 ms	Network Size 2020; Computational Time 1747.1 ms	Network Size 1870; Computational Time 917.5 ms	Network Size 1830; Computational Time 1487.9 ms	Network Size 1587; Computational Time 578.3 ms	

Figure 2. A representative set of eight metabolic networks from the angiosperm plant lineage, including eudicot and monocot subgroups and the early land plant and green algae lineages. (Upper matrix) The number of common enzymes between both species is denoted as $|V|$. MDSM and MDSI numbers are indicated in the box cells. (Lower matrix) The network size and the computational time (milliseconds) required to execute the FAST-MDSM method are shown. The total number of nodes and the MDS for each single network are shown in the first column.

deletions of random edges). The results are shown in Fig. S3(b), where the average ratio over 10 trials is shown for each case. We can see a good agreement between the simulation results and the theoretical estimates for the cases of a non-large number of insertions and deletions of random edges. The reason for the much better agreement may be that the degree is not necessarily preserved in this case.

Results on real networks. To examine the usefulness of the MDSM for the analysis of biological networks, we applied the MDSM to two kinds of large-scale multilayer biological networks: plant metabolic networks and protein-protein interaction networks.

First, we applied the developed fast method of computing the MDS for multilayer networks (FAST-MDSM) to the study of genome-wide metabolic networks across major plant lineages compiled from the Plant Metabolic Network database 16.0[33,34]. To validate the biological relevance and the functional role of the nodes (enzymes) included in the MDSMs, we performed a comparative analysis of plant metabolic networks. We collected 70 species of metabolic networks corresponding to major plant lineagesâ€" from green algae and early land plants to angiosperms. The angiosperm lineage was further subdivided into major evolutionary groups: monocots (such as grasses and cereals) and eudicots. We then computed the corresponding MDSM for each pair of networks. The resulting MDSM identifies the common minimum number of enzymes that must be simultaneously controlled in both networks. To compare these with a case involving a simple overlap between networks, we consider the MDSI (minimum dominating set induced) defined in the Methods section and illustrated in Fig. 1. Note that the MDSI is a simple intersection of the MDS; therefore, its controllability role is unclear, which highlights the novelty and importance of the MDSM. The results shown in Fig. 2 (upper matrix) correspond to eight species and indicate that the size of the MDSM tends to be closer to that of the MDSI when the two networks belong to the same evolutionary group. On the other hand, Fig. 2 (lower matrix) shows the network size and the computational time in milliseconds required to execute the FAST-MDSM method using two-layer networks. See the SI files for details of all 70 species (Table S1). To illustrate the efficiency of the proposed algorithm, we also considered multilayer network analysis with more than two networks. The results for a 6-layer network analysis, in which one group is composed of up to three species, are shown in Fig. 3. We see that the computational time is the same order of magnitude of that necessary to compute a much simpler 2-layer network in many cases. The results also show that this multilayer network problem can be quickly solved using the proposed method, in spite of the fact that the MDS is an NP-hard problem (see Fig. 3 (lower matrix)). See also the SI for the complete results of the 6-layer network analysis involving 25 groups (Table S2).

Plantcyc — Main component	ntabacum_tn90cyc oilseedrapecyc boleracea_capitatacyc	klaxifloracyc chinesecabbagecyc brapa_fpsccyc	setariacyc wheatacyc oglaberrimacyc	pineapplecyc oryzacyc barleycyc	mosscyc	selaginellacyc	csubellipsoideacyc olucimarinuscyc mcommodacyc	chlamycyc mpusillacyc vcartericyc
ntabacum_tn90cyc oilseedrapecyc boleracea_capitatacyc		Overlap : |v| 1465 / MDSM 238 / MDSI 84	Overlap : |v| 1361 / MDSM 227 / MDSI 76	Overlap : |v| 1383 / MDSM 233 / MDSI 83	Overlap : |v| 1323 / MDSM 219 / MDSI 86	Overlap : |v| 1383 / MDSM 233 / MDSI 92	Overlap : |v| 905 / MDSM 173 / MDSI 48	Overlap : |v| 917 / MDSM 172 / MDSI 51
klaxifloracyc chinesecabbagecyc brapa_fpsccyc	Network Size 2246 / Computational Time 27103.0 ms		Overlap : |v| 1388 / MDSM 236 / MDSI 75	Overlap : |v| 1409 / MDSM 241 / MDSI 79	Overlap : |v| 1356 / MDSM 224 / MDSI 93	Overlap : |v| 1422 / MDSM 239 / MDSI 107	Overlap : |v| 922 / MDSM 173 / MDSI 54	Overlap : |v| 939 / MDSM 174 / MDSI 55
setariacyc wheatacyc oglaberrimacyc	Network Size 2372 / Computational Time 7937.3 ms	Network Size 2266 / Computational Time 50676.4 ms		Overlap : |v| 1455 / MDSM 244 / MDSI 68	Overlap : |v| 1308 / MDSM 225 / MDSI 82	Overlap : |v| 1388 / MDSM 236 / MDSI 92	Overlap : |v| 901 / MDSM 171 / MDSI 44	Overlap : |v| 923 / MDSM 170 / MDSI 42
pineapplecyc oryzacyc barleycyc	Network Size 2367 / Computational Time 27507.4 ms	Network Size 2268 / Computational Time 130482.7 ms	Network Size 2150 / Computational Time 9403.0 ms		Overlap : |v| 1328 / MDSM 233 / MDSI 71	Overlap : |v| 1401 / MDSM 241 / MDSI 80	Overlap : |v| 907 / MDSM 174 / MDSI 51	Overlap : |v| 915 / MDSM 170 / MDSI 59
mosscyc	Network Size 2396 / Computational Time 30356.8 ms	Network Size 2286 / Computational Time 4169.2 ms	Network Size 2252 / Computational Time 5946.2 ms	Network Size 2220 / Computational Time 7063.3 ms		Overlap : |v| 1432 / MDSM 237 / MDSI 131	Overlap : |v| 935 / MDSM 176 / MDSI 56	Overlap : |v| 959 / MDSM 179 / MDSI 56
selaginellacyc	Network Size 2260 / Computational Time 19236.1 ms	Network Size 2161 / Computational Time 287157.8 ms	Network Size 2117 / Computational Time 4793.2 ms	Network Size 2093 / Computational Time 8218.6 ms	Network Size 1921 / Computational Time 3843.0 ms		Overlap : |v| 932 / MDSM 174 / MDSI 60	Overlap : |v| 955 / MDSM 178 / MDSI 44
csubellipsoideacyc olucimarinuscyc mcommodacyc	Network Size 2485 / Computational Time 2310.4 ms	Network Size 2374 / Computational Time 2863.4 ms	Network Size 2342 / Computational Time 2125.9 ms	Network Size 2319 / Computational Time 7052.7 ms	Network Size 2067 / Computational Time 2654.5 ms	Network Size 2026 / Computational Time 1433.0 ms		Overlap : |v| 892 / MDSM 168 / MDSI 44
chlamycyc mpusillacyc vcartericyc	Network Size 2532 / Computational Time 2372.4 ms	Network Size 2446 / Computational Time 3198.1 ms	Network Size 2402 / Computational Time 2419.9 ms	Network Size 2359 / Computational Time 2569.0 ms	Network Size 2107 / Computational Time 2185.3 ms	Network Size 2097 / Computational Time 1254.2 ms	Network Size 1966 / Computational Time 3750.3 ms	

Figure 3. Same as Fig. 2 but considering a 6-layer network analysis in which each group consists of up to three species. The results for the complete set of 70 metabolic networks for the 2-layer analysis and for the set of 25 groups for the 6-layer analysis are shown in the Supplementary Information file (Tables S1 and S2, respectively).
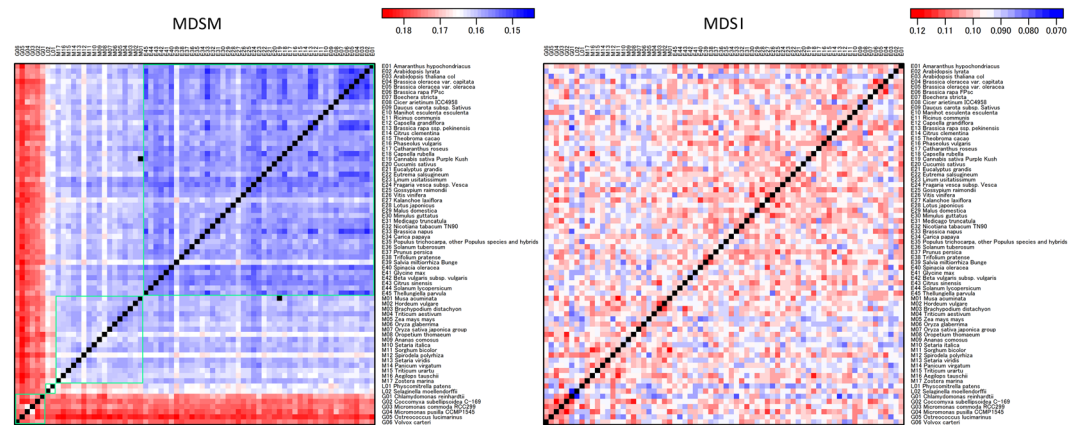

Figure 4. The heat map of the fraction of MDSM (left) and MDSI (right) for each pair of metabolic networks from green algae (G), early land plants (L), eudicots (E) and monocots (M). This result highlights the fact that the MDSMs cluster according to each major plant lineage (see green box). The MDSM size tends to be smaller within each lineage, facilitating the control of each network pair in each major group. The analysis involves 70 species.

Next, to verify the algorithm performance for even larger real-world networks, we downloaded protein-protein interaction networks from the HINT database version 4.0. The results in Fig. S4 (upper matrix) show that the size of the MDSM is still close to that of the MDSI (see *H. sapiens* vs. *D. Melanogaster* results in Fig. S4). This finding illustrates the advantages of controlling several networks simultaneously because the identified MDSM set does not significantly differ from the set required to control a single network in many cases. The results show that these large networks can be quickly solved using the proposed method, in spite of the fact that the MDS is an NP-hard problem (see Fig. S4, lower matrix). The results for 3-layer, 4-layer PPI networks are shown in the SI (Table S3).

Then, the fractions of the MDSMs and MDSIs were represented in a heat map for all 70 species using a 2-layer network analysis and a 6-layer network analysis in which one group was composed of up to three species (see Figs 4 and 5, respectively). The result highlights the fact that the MDSMs cluster according to each major plant
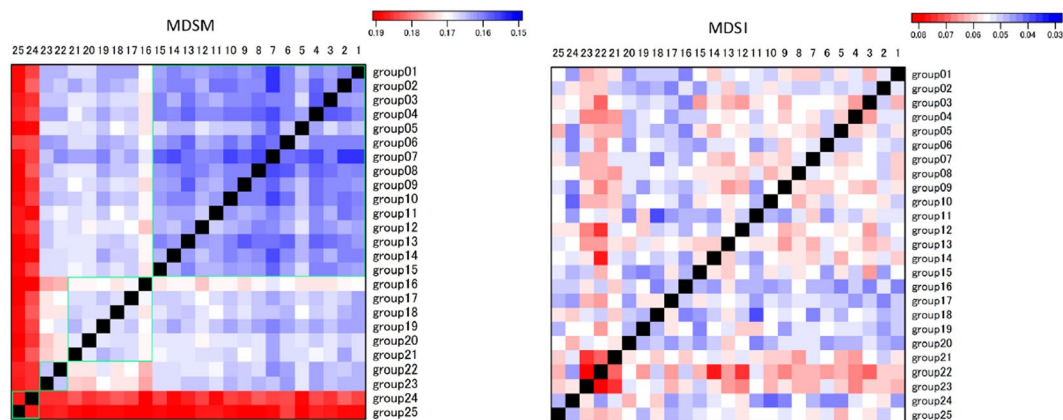
Figure 5. Same as Fig. 4 but for a 6-layer network analysis in which 25 groups consisting of up to 3 layers were analysed. This result highlights the fact that the MDSMs cluster according to each major plant lineage (see green box) whereas MDSI does not recognise the plant lineages.

lineage. In other words, species belonging to the same lineage tend to have a higher similarity of reaction node sets[33]; therefore, each MDSM tends to be smaller within each lineage, facilitating the control of each network pair.

It is worth considering the multilayer network problem as a single network. Therefore, we compute an MDS for the union of two graphs G_1 and G_2, instead of considering a two-layer network analysis. Two graphs can be combined ($G_1 \cup G_2$) as shown in Fig. S7. Then, the MDS results show that the resulting union graph does not increase notably the required MDS size to control the network. Specially, when both networks belong to the same lineage, the MDS gives similar results (see Fig. S8).

To elucidate the biological functionality and importance of the reaction node sets associated with the MDSM, we computed the average enrichment of the functional classes of the metabolic reactions classified for each plant lineage (green algae, monocots and eudicots) (see Methods and Fig. 6). We excluded early land plants from this analysis because only two samples were available. First, the results show similar control trends for monocots (e.g., grasses and cereals) and eudicots because they share a more similar metabolism than do the algae. In contrast, the differences between angiosperms and algae are more evident. A more detailed analysis was performed using the enrichment results of each functional class and major plant lineage. Fourteen functional classes were used to classify the enzymes and reactions of each species. For example, consider the specialized metabolism functional class. For the MDSM, the absolute difference in enrichment (er) between eudicots (E) and monocots (M) (both are angiosperm major groups) is much smaller than that between eudicots and green algae (A), as follows: ($|er(E) - er(M)| \ll |er(E) - er(A)|$). In contrast, these two differences have approximately the same strength for the MDSI ($|er(E) - er(M)| \approx |er(E) - er(A)|$). Therefore, by using these criteria, the results in Fig. 6 suggest that the functional classes of amino acids and specialized metabolism show a tendency to benefit from network multicontrol via MDSM rather than via MDSI. In contrast, the results for the functional classes of fatty acids and lipids and for hormone metabolism show an increased tendency to benefit from MDSI control.

Discussion

We developed a new FAST-MDS computation that allows us to efficiently identify the minimum driver nodes in large multilayer networks. The application of the developed tools to metabolism in major plant lineages showed that, as expected, for each pair of species, a subset of the functional metabolic classes could be more efficiently multicontrolled using a common subset of enzymes.

In previous studies, the abundance of chemical reactions in functional classes was investigated from an evolutionary perspective for major plant lineages, including early land plants and angiosperms. The results suggested a significant depletion of carbohydrates, amino acids, nucleotides, energy and cofactors metabolism in angiosperms. On the other hand, early plants and angiosperms showed an enrichment of carbohydrates and a specialized metabolism, respectively[33].

The results shown in Fig. 6 suggest that the specialized metabolism functional class tends to benefit from multicontrol MDSM. This metabolic pathway generates specialized metabolites that are useful for adjusting the cell state to the surroundings and environmental conditions[35]. The secondary metabolism has an inherent plasticity that facilitates its control at different levels. It has been reported that specific gene clusters are responsible for the synthesis of specialized metabolites in bacteria and fungi[36,37]. Studies have shown that the control and modification of these regulatory processes can lead to the production of desired specialized metabolites[38]. A similar approach could be applied to plants; indeed, one transcriptional regulator has already been associated with a plant metabolic gene cluster[39,40]. Our results indicate that in some cases, the control of multiple pathways and species could be even more efficient in terms of the minimum number of required driver enzymes than that required to control specific pathways and species separately.

First, the presented findings suggest future directions for controlling multiple species by identifying a common set of enzymes or genes for regulation or drug targeting. This strategy may be useful in developing drugs that
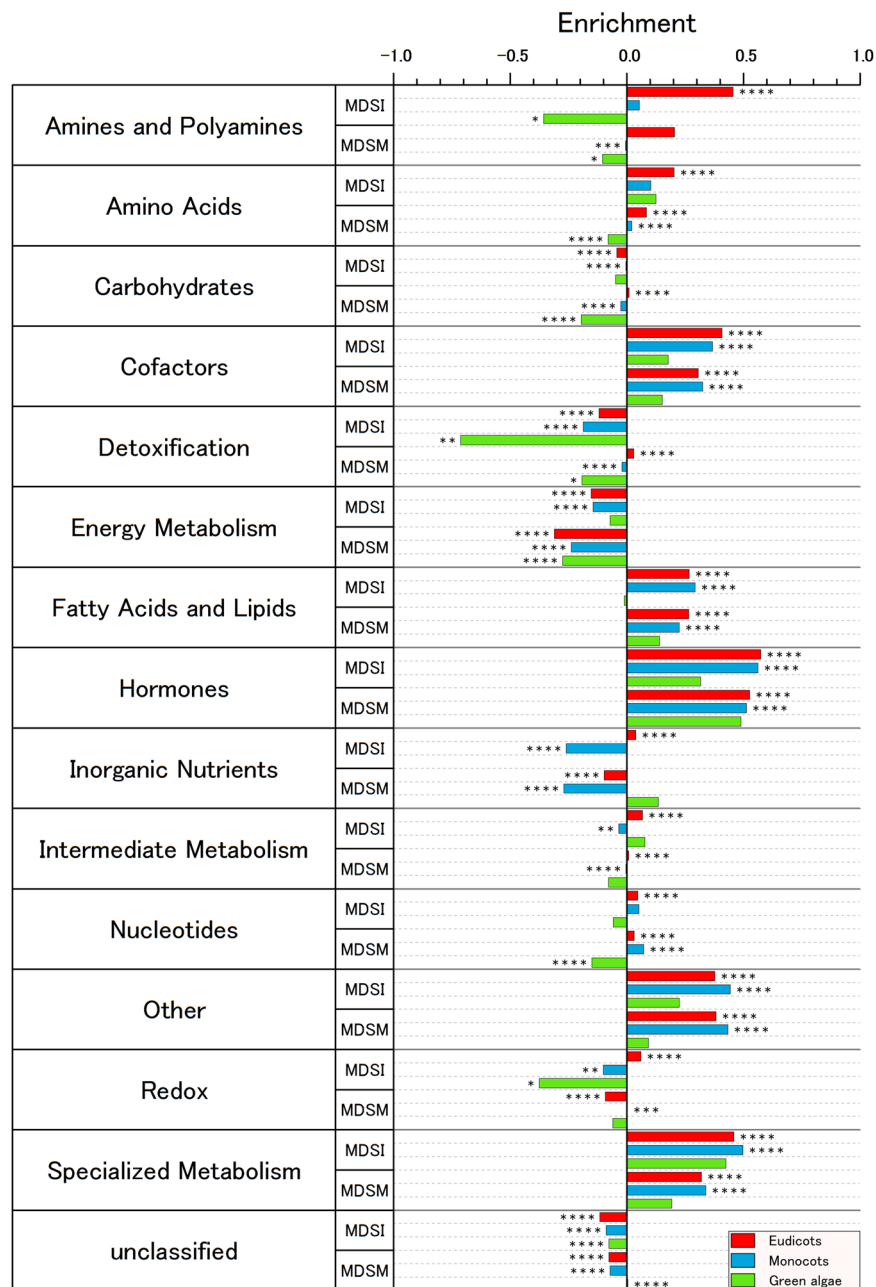
Figure 6. The average enrichment of the functional classes of metabolic reactions classified for each major lineage group. The results suggest that the functional classes of amino acids and specialized metabolism show a tendency to benefit from network multicontrol via MDSM rather than via MDSI. In contrast, the results for the functional classes of fatty acids and lipids and hormone metabolism show an enhanced tendency to benefit from MDSI control. The statistical analysis results are denoted as follows: ****p-value < 0.0001, ***p-value < 0.001, **p-value < 0.01 and *p-value < 0.05.

can kill several kinds of harmful bacteria. Second, cellular networks may differ among individuals. However, our computational findings suggest that we may be able to use the same set of drugs for many patients.

In conclusion, the tools and methodologies presented for controlling multilayer networks using a common set of nodes could lead to broad research directions, ranging from evolutionary biology to drug design and development, that deserve further exploration.

Methods

Metabolic network analysis. A set of genome-wide plant metabolic networks was compiled from the publicly available Plant Metabolic Network (PMN) database version 16.0[34]. The set included 70 species from major plant lineages: six species correspond to green algae, two species are early land plants, 17 species are monocots, and 45 species are eudicots. The analysis was performed using an enzyme/reaction-centric network assembled

from the full metabolic pathways, in which enzymes are considered as control targets. In multiple layers problem, the shared enzymes may potentially lead to identify key enzymes to control multiple organisms. Therefore, we consider the intersection of nodes among layers as shown in Fig. 1. To validate the biological importance of the nodes engaged in the MDSMs, we compared them with the MDSI, which is defined as follows (see also Fig. 1):

(1) From two species of metabolic networks defined as $G_1(V_1, E_1)$ and $G_2(V_2, E_2)$, we compute the common nodes defined as $V = V_1 \cap V_2$.
(2) From each graph G_1 and G_2, we compute the graphs $G_1(V)$ and $G_2(V)$ induced by the set of nodes V.
(3) Then, we compute $MDS(G_1(V))$, $MDS(G_2(V))$ and $MDSM(G_1(V), G_2(V))$.
(4) Finally, we compare $MDSM(G_1(V), G_2(V))$ with $MDS(G_1(V)) \cap MDS(G_2(V))$. The latter is called the MDSI.

For the above metabolic network analysis, the ILP was determined using the GLPK LP/MIP Solver v4.55. The enrichment of a multicontrol feature C (MDSI or MDSM) in a given functional class F of enzymes for a pair of networks reads as $E^C(F) = \ln{[(N_p^C(F)/N_p^C)/(N_p(F)/N_p)]}$, where N_p indicates the total number of enzymes common in both organisms and $N_p(F)$ indicates the total number of enzymes common in both organisms that also belong to the F functional class. N_p^C refers to the total number of enzymes that are in the multicontrol set C computed for both organisms (included in N_p). $N_p^C(F)$ indicates the total number of enzymes that are in the control set C computed for both organisms and that belong to the F functional class (included in $N_p(F)$).

FAST-MDSM computational procedure. Although both MDS and MDSM computation are NP-hard, optimal solutions can be obtained for large networks by using a simple integer linear programming (ILP) formulation if networks have scale-free properties. Furthermore, for two-layer networks, MDSMs can be obtained more quickly using the heuristic preprocessing method given below, which has some similarity with the generalized leaf-removal procedure proposed for single-layer networks in[19] but is substantially different to cope with multi-layer networks. In this subsection, we provide ILP-based procedures for both two-layer networks and networks with more than two layers, both of which are referred to as FAST-MDSM.

First, we determine a subset V_M of an MDSM in the following way by focusing on degree 1 nodes, where v is *observed* in G_i if v has a neighbouring node $v' \in V_M$ in G_i. Note that $d_{G_i}(v)$ indicates the degree of v in G_i.

(1) Let $V_M \leftarrow \{\}$.
(2) Repeat steps (3)–(8) until no more nodes are added to V_M.
(3) For all nodes $v_i \in V_1 \cup V_2$, perform steps (4)–(6).
(4) If there exists an unobserved $v_j \notin V_M$ in G_1 such that $d_{G_1}(v_j) = 1$, $\{v_i, v_j\} \in E_1$ and $v_j \notin V_2$ hold, then delete $\{v_i, v_j\}$ from E_1, add v_i to V_M, and let v_j be observed in G_1.
(5) If there exists an unobserved $v_j \notin V_M$ in G_2 such that $d_{G_2}(v_j) = 1$, $\{v_i, v_j\} \in E_2$ and $v_j \notin V_1$ hold, then delete $\{v_i, v_j\}$ from E_2, add v_i to V_M, and let v_j be observed in G_2.
(6) If there exists an unobserved $v_j \notin V_M$ such that $d_{G_1}(v_j) = 1$, $d_{G_2}(v_j) = 1$, $\{v_i, v_j\} \in E_1$ and $\{v_i, v_j\} \in E_2$ hold, then delete $\{v_i, v_j\}$ from E_1 and E_2, add v_i to V_M, and let v_j be observed in G_1 and G_2.
(7) For all unobserved nodes $v_i \notin V_M$ in G_1, if there exists $v_j \in V_M$ such that $\{v_i, v_j\} \in E_1$, let v_i be observed in G_1.
(8) For all unobserved nodes $v_i \notin V_M$ in G_2, if there exists $v_j \in V_M$ such that $\{v_i, v_j\} \in E_2$, let v_i be observed in G_2.

It is obvious that this procedure never puts a v_i into a V_M that is not included in an MDSM; thus, it is ensured that V_M is a subset of an MDSM. Then, we apply the following ILP, where each x_i is a binary variable (i.e., x_i takes either 0 or 1).

$$
\begin{aligned}
\text{minimize} \quad & \sum_{i=1}^{n} x_i, \\
\text{subject to} \quad & x_i = 1 \quad \text{for all } v_i \in V_M, \\
& x_i + \sum_{j:\{v_j, v_i\} \in E_1} x_j \geq 1, \quad \text{for all } v_i \in V_1 - V_M \text{ not observed in } G_1, \\
& x_i + \sum_{j:\{v_j, v_i\} \in E_2} x_j \geq 1 \quad \text{for all } v_i \in V_2 - V_M \text{ not observed in } G_2.
\end{aligned}
$$

In this ILP, $x_i = 1$ corresponds to $v_i \in V_{MDSM}$, where V_{MDSM} is the set of nodes in an MDSM computed using this ILP-based method. The first line states that the number of nodes in V_{MDSM} must be minimized. The second line states that every $v_i \in V_M$ must be included in V_{MDSM}. The third (resp., fourth) line states that every $v_i \in V_1 - V_M$ (resp., $v_i \in V_2 - V_M$) must be included in V_{MDSM} (i.e., $x_i = 1$) or must have a neighbour $v_j \in V_{MDSM}$. Since an ILP solver always outputs an optimal solution (if computation is finished), it is shown that this procedure clearly finds an MDSM.

This preprocessing method can be generalized for networks of three or more layers. However, the following simple ILP formulation works enough well for three or more layer networks as well as for two-layer networks.

$$
\begin{aligned}
\text{minimize} \quad & \sum_{i=1}^{n} x_i, \\
\text{subject to} \quad & x_i + \sum_{j:\{v_j, v_i\} \in E_k} x_j \geq 1 \quad \text{for all } v_i \in V_k, \text{ for all } k = 1, \ldots, N.
\end{aligned}
$$

There exist two variants in its implementation: (i) ignoring isolated nodes and (ii) not ignoring isolated nodes, because if an isolated node exists, it must be included in an MDSM under the original definition.

As will be shown later, the above ILP-based methods work efficiently for real biological networks. Although exact reasons are very unclear, one possible reason is that most real networks have power-law degree distributions with a low average degree and thus have many degree 1 and degree 2 nodes by which many MDSM nodes are determined based only on the local topology and/or progressively.

Degree-preserving rewiring. Here, we present a theoretical analysis by employing degree-preserving rewiring, which has been widely used in the study of complex networks, to modify a given network while preserving the degree distribution (see also Fig. S1). In this method, we randomly choose a pair of edges $\{v_i, v_j\}$ and $\{v_h, v_k\}$ in a given graph $G(V, E)$ such that $\{v_i, v_k\} \notin E$ and $\{v_h, v_j\} \notin E$. Then, we delete these edges and add two edges $\{v_i, v_k\}$ and $\{v_h, v_j\}$. We repeat this procedure K times. Clearly, the degree of each node does not change; thus, the degree distribution is preserved. Let G_0 be the original network and G_K be the network obtained by applying degree-preserving rewiring K times.

We here give a quantitative estimate of the difference between $S_{MDS}(G_0, G_0) = S_{MDS}(G_0)$ and $S_{MDS}(G_0, G_1)$. It is expected that $S_{MDS}(G_0, G_K) \leq K \cdot S_{MDS}(G_0, G_1)$ approximately holds in many cases, although it does not always hold because the size of an MDSM is sensitive to exchanged edges. Let V_{MDS} be an MDS for $G_0(V_0, E_0)$. We can obtain a dominating set for (G_0, G_1) by adding at most one node to V_{MDS}. Specifically, we may need to add one node to V_{MDS} only for the following cases (see Fig. S1):

(i) $|\{v_i, v_j, v_h, v_k\} \cap V_{MDS}| = 1$,
(ii) $|\{v_i, v_k\} \cap V_{MDS}| = 1$ and $\{v_j, v_h\} \cap V_{MDS} = \varnothing$,
(iii) $|\{v_j, v_h\} \cap V_{MDS}| = 1$ and $\{v_i, v_k\} \cap V_{MDS} = \varnothing$.

Let p be the probability that an arbitrary endpoint of a randomly chosen edge of G_0 belongs to V_{MDS}, which is estimated as

$$p \approx \frac{\sum_{v \in V_{MDS}} d_{G_0}(v)}{\sum_{v \in V_0} d_{G_0}(v)}.$$

Then, case (i) holds with probability $4p(1-p)^3$ if there was no constraint on the chosen edges, where '4' comes from the fact that any one of v_i, v_j, v_h, v_k can be a node in an MDS. Both case (ii) and case (iii) hold with probability $p^2(1-p)^2$ if there were no constraints on the chosen edges. Here, we should note that in case (i), if v_j has another neighbour in V_{MDS}, then we need not add a node to V_{MDS}. Similarly, we need not add a node in case (ii) (resp., case (iii)) if either v_j or v_h (resp., v_i and v_k) has another neighbour in V_{MDS}. Thus, we need to consider the probability q that an arbitrary node $v \notin V_{MDS}$ has more than one neighbouring node in V_{MDS}, which is given by the ratio of the number of such nodes to the total number of nodes not in V_{MDS}. Therefore, the probability P_{+1} that we need to add one node to V_{MDS} is estimated as

$$P_{+1} \approx 4(1 - q)p(1 - p)^3 + 2(1 - q)^2 p^2 (1 - p)^2. \tag{1}$$

If we apply rewiring K times, $S_{MDS}(G_0, G_K) \leq K \cdot P_{+1}$ would approximately hold for small K. In the Results on artificial networks section, we compared this estimate of $S_{MDS}(G_0, G_K)$ with its actual size using artificially generated scale-free networks. It is to be noted that although P_{+1} is calculated from the empirical distribution of k-degree nodes in an MDS, we might be able to obtain an analytical expression of P_{+1} if such a distribution could be analytically obtained.

Random insertions and deletion of edges. Next, we consider the case of the random insertions and deletion of edges. We assume that G_{i+1} is created from G_i via either edge deletion or edge insertion as follows:

- One edge is randomly selected from G_i and deleted with probability p_d;
- Otherwise, a pair of nodes (not connected by an edge) is randomly selected and connected by an edge (with probability $1 - p_d$).

Since it is difficult to estimate the change in the MDS size due to edge insertions, we focus on edge deletions. Let $\{v_i, v_j\}$ be an edge to be deleted from G_0. Suppose that $v_j \notin V_{MDS}$, $v_i \in V_{MDS}$, and v_j does not have more than one neighbouring node in V_{MDS}. Then, v_j must be added to V_{MDS}. The probability of occurrence of such a case is estimated as $(1 - q)p(1 - p)$. Since we need to consider the symmetric case and the probability p_d, the probability that one node is added to V_{MDS} is estimated as

$$P_{+1} \approx 2p_d(1 - q)p(1 - p). \tag{2}$$

Data Availability

All data related to metabolic patways analysed in this work were downloaded from the publicly available Plant Metabolic Network Database Version 16.0 (PMN) www.plantcyc.org. Because this is a public database we cannot re-upload the entire datasets in our submission. All the computational results are presented in figures and included in this manuscript and in the Supplementary Information that accompanies this paper.

References

1. Lin, C. T. Structural controllability. *IEEE T. Automat. Contr.* **19**, 201–208 (1974).
2. Akutsu, T., Hayashida, M., Ching, W.-K. & Ng, M. K. Control of Boolean networks: Hardness results and algorithms for tree structured networks. *Journal of Theoretical Biology* **244**, 670–679 (2007).
3. Kim, D.-H. & Motter, A. E. Slave nodes and the controllability of metabolic networks. *New Journal of Physics* **11**, 113047 (2009).
4. Tejeda Zañudo, J. G., Yang, G. & Albert, R. Structure-based control of complex networks with nonlinear dynamics. *Proc. Natl. Acad. Sci. USA* **114**, 7234–7239 (2017).
5. Liu, Y.-Y., Slotine, J.-J. & Barabási, A.-L. Controllability of complex networks. *Nature* **473**, 167–173 (2011).
6. Nepusz, T. & Vicsek, T. Controlling edge dynamics in complex networks. *Nature Physics* **8**, 568–573 (2012).
7. Menichetti, G., Dall' Asta, L. & Bianconi, G. Network controllability is determined by the density of low in-degree and out- degree nodes. *Phys. Rev. Lett.* **113**, 078701 (2014).
8. Berlingerio, M., Coscia, M., Giannotti, F., Monreale, A. & Pedreschi, D. Foundations of multidimensional network analysis. *International Conference on Advances in Social Networks Analysis and Mining* 485–489, https://doi.org/10.1109/ASONAM (2011).
9. De Domenico, M. *et al.* Mathematical formulation of multilayer networks. *Phys. Rev. X* **3**, 041022 (2013).
10. Yuan, Z., Zhao, C., Wang, W.-X., Di, Z. & Lai, Y.-C. Exact controllability of multiplex networks. *New J. Phys.* **16**, 103036 (2014).
11. Zhang, Y., Garas, A. & Schweitzer, F. Value of peripheral nodes in controlling multilayer scale-free networks. *Phys. Rev. E* **93**, 012309 (2016).
12. Pósfai, M., Gao, J., Cornelius, S. P., Barabási, A.-L. & D'Souza, R. M. Controllability of multiplex, multi-time-scale networks. *Phys. Rev. E* **94**, 032316 (2016).
13. Menichetti, G., Dall' Asta, L. & Bianconi, G. Control of multilayer networks. *Scientific Reports* **6**, 20706 (2016).
14. Nacher, J. C. & Akutsu, T. Dominating scale-free networks with variable scaling exponent: heterogeneous networks are not difficult to control. *New Journal of Physics* **14**, 073005 (2012).
15. Nacher, J. C. & Akutsu, T. Structural controllability of unidirectional bipartite networks. *Scientific Reports* **3**, 1647 (2013).
16. Mochizuki, A., Fiedler, B., Kurosawa, G. & Saito, D. Dynamics and control at feedback vertex sets. II: A faithful monitor to determine the diversity of molecular activities in regulatory networks. *Journal of Theoretical Biology* **335**, 130–146 (2013).
17. Nacher, J. C. & Akutsu, T. Minimum dominating set-based methods for analyzing biological networks. *Methods* **102**, 57–63 (2016).
18. Molnár, F., Sreenivasan, S., Szymanski, K. & Korniss, G. Minimum dominating sets in scale-free networks ensembles. *Scientific Reports* **3**, 1736 (2013).
19. Zhao, J.-H., Habibulla, Y. & Zhou, H.-J. Statistical mechanics of the minimum dominating set problem. *J Stat Phys* **159**, 1154–1174 (2015).
20. Milenkovic, T., Memisevic, V., Bonato, A. & Przulj, N. Dominating biological networks. *PloS One* **6**, e23016 (2011).
21. Wuchty, S. Controllability in protein interaction networks. *Proc. Natl. Acad. Sci. USA* **111**, 7156–7160 (2014).
22. Zhang, X.-F., Ou-Yang, L., Zhu, Wu, M.-Y. & Dai, D.-Q. Determining minimum set of driver nodes in protein-protein networks. *BMC Bioinformatics* **16**, 146 (2015).
23. Nacher, J. C. & Akutsu, T. Analysis of critical and redundant nodes in controlling directed and undirected complex networks using dominating sets. *Journal of Complex Networks* **2**, 394–412 (2014).
24. Kagami, H., Akutsu, T., Maegawa, S., Hosokawa, H. & Nacher, J. C. Determining associations between human diseases and non-coding RNAs with critical roles in network control. *Scientific Reports* **5**, 14577 (2015).
25. Sun, P. G. Co-controllability of drug-disease-gene network. *New Journal of Physics* **17**, 085009 (2015).
26. Basler, G., Nikoloski, Z., Larhlimi, A., Barabási, A.-L. & Liu, Y.-Y. Control of fluxes in metabolic networks. *Genome Research* **26**, 956–968 (2016).
27. Boccaletti, S. *et al.* The structure and dynamics of multilayer networks *Physics Reports* **14** (2015).
28. Gao, J., Buldyrev, S. V., Stanley, H. E. & Havlin, S. Networks formed from interdependent networks. *Nat. Phys.* **8**, 40–48 (2012).
29. Bollobás, B. *Random Graphs. Second Eddition.* (Cambridge University Press, Cambridge, UK, 2011).
30. Takemoto, K. & Akutsu, T. Analysis of the effect of degree correlation on the size of minimum dominating sets in complex networks. *PLoS One* **11**, e0157868 (2016).
31. Bredereck, R. *et al.* Assessing the computational complexity of multi-layer subgraph detection. *Lecture Notes in Computer Science* **10236**, 128–139 (2017).
32. Fomin, F. V. & Kratsch, D. *Exact Exponential Algorithms.* (Springer, Berlin, 2010).
33. Chae, L., Kim, T., Nico-Poyanco, R. & Rhee, S. Y. Genomic signatures of specialized metabolism in plants. *Science* **344**, 510–513 (2014).
34. Rhee, S. Y. Plant Metabolic Network Database Version 16.0 (PMN) www.plantcyc.org (*Date of access*: 15/05/2017).
35. Ignea, C. *et al.* Overcoming the plasticity of plant specialized metabolism for selective diterpene production in yeast. *Scientific Reports* **7**, 8855 (2017).
36. Brakhage, A. Regulation of fungal secondary metabolism. *Nat. Rev. Microbiol.* **11**, 21–32 (2013).
37. van Wezel, G. P. & McDowall, K. J. The regulation of the secondary metabolism of Streptomyces: new links and experimental advances. *Nat. Prod. Rep.* **28**, 1311–1333 (2011).
38. Bergmann, S. *et al.* Genomics-driven discovery of PKS-NRPS hybrid metabolites from Aspergillus nidulans. *Nat Chem Biol* **3**, 213–217 (2007).
39. Okada, A. *et al.* OsTGAP1 a bZIP transcription factor, coordinately regulates the inductive production of diterpenoid phytoalexins in rice. *J Biol Chem* **284**, 26510–26518 (2009).
40. Nutzmann, H.-W. & Osburn, A. Gene clustering in plant specialized metabolism. *Current Opinion in Biotechnology* **26**, 91–99 (2014).

Acknowledgements

Author Contributions

J.C.N. and M.I. performed research, analysed empirical datasets, and prepared the figures. J.C.N. and T.A. designed research, contributed new analytical and theoretical tools and wrote the paper. S.M. contributed to theoretical analysis and improvements of the presentation. All authors have reviewed and approved the manuscript content.

Additional Information

Supplementary information accompanies this paper at https://doi.org/10.1038/s41598-018-37046-z.

Competing Interests: The authors declare no competing interests.

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