



Structural controllability of unidirectional bipartite networks

Jose C. Nacher¹ & Tatsuya Akutsu²

¹Department of Information Science, Faculty of Science, Toho University, Miyama 2-2-1, Funabashi, Chiba 274-8510, Japan, ²Bioinformatics Center, Institute for Chemical Research, Kyoto University, Uji, 611-0011, Japan.

SUBJECT AREAS:

COMPUTER SCIENCE

COMPLEX NETWORKS

COMPUTATIONAL BIOLOGY AND
BIOINFORMATICS

SYSTEMS BIOLOGY

Received
12 March 2013

Accepted
28 March 2013

Published
10 April 2013

Correspondence and requests for materials should be addressed to J.C.N. (nacher@is.sci.toho-u.ac.jp) or T.A. (takutsu@kuicr.kyoto-u.ac.jp)

The interactions between fundamental life molecules, people and social organisations build complex architectures that often result in undesired behaviours. Despite all of the advances made in our understanding of network structures over the past decade, similar progress has not been achieved in the controllability of real-world networks. In particular, an analytical framework to address the controllability of bipartite networks is still absent. Here, we present a dominating set (DS)-based approach to bipartite network controllability that identifies the topologies that are relatively easy to control with the minimum number of driver nodes. Our theoretical calculations, assisted by computer simulations and an evaluation of real-world networks offer a promising framework to control unidirectional bipartite networks. Our analysis should open a new approach to reverting the undesired behaviours in unidirectional bipartite networks at will.

Networks represent the backbones of technological, societal and natural systems. In recent years, statistical physics has provided the tools to understand the universal principles that govern the structure of networks^{1,2}. However, the network knowledge of real systems as well as the proposed mean-field theories and approaches are still far from being complete. Detailed knowledge of the structure has not yet led to control networks with fewer input signals. For example, we still do not have a theoretical framework that allows us to reprogram the entire gene regulatory network in a living cell to transition from a disease state to a normal state using a small set of drugs. The efforts in this direction are recent and have permeated different fields, with a combination of tools from control theory and statistical physics^{3–11}. A dynamical system can be controlled if the system can be driven from any initial state to any final state by an external set of signals within a finite amount of time. Although the processes that occur in real-world networks are mostly non-linear, canonical linear, time-invariant nodal dynamics⁸ has been proposed for studying the controllability of networks (see the Supplementary Information (SI) for details) in which a vector of input signals $u(t)$ is coupled to a set of nodes (drivers) that control the interactions between the entire system's nodes defined by the transpose of the weighted adjacency matrix A . The state of a system $x(t)$ of N nodes at time t , can indicate the positive/negative opinions or high/low expression levels that change with time. This simplification is applied to the modelling of non-linear systems because the structural controllability¹² of a given system is equivalent to the controllability of a continuum of linearised systems; therefore, the analytical results could provide sufficient controllability conditions for most nonlinear systems^{8,13}.

To address structural controllability, models based on nodal⁸ and edge dynamics¹⁰ were recently proposed. In nodal dynamics, the minimum number of input signals necessary to control the whole network is determined by finding the maximum matching in a bipartite graph obtained from the original network. The number of unmatched nodes is the number of driver nodes. In this approach, the input signals (or driver nodes) tend to avoid the high-degree nodes⁸. As a result, the random networks in which hubs are absent are easier to control. An alternative view tackles the problem by evaluating a dynamical process that is defined on the edges of a network rather than in the nodes¹⁰. In the edge dynamics approach, each node i acts as a simple switchboard-like device, mathematically represented as a mixing matrix M_i with rows (columns) equal to the out-degree (in-degree), that receives information through its inbound edge and transmits the outcome or decisions to its neighbouring nodes by means of the outbound edges. In sharp contrast to the nodal dynamics⁸, this approach concludes that the scale-free degree distributions, where hubs are present, are easier to control. However, despite the fact that bipartite networks represent a network type that is often used to represent the interactions of distinct units in real-world systems, both the nodal and edge dynamics frameworks only address simple (unipartite) graphs.

Here we address the controllability of unidirectional bipartite networks, but instead of using nodal dynamics we attempt the problem from a different angle by, considering a modified version of the *minimum dominating set*



(MDS)¹¹. A set $S \subseteq V$ of nodes in a graph $G = (V, E)$ is a dominating set if every node $v \in V$ is either an element of S or is adjacent to an element of S ¹⁴. This *dominant set* (DS) of nodes plays the role of the set of driver nodes in the sense of Ref. 8,10. In our companion work, the MDS was suggested as a method to investigate the controllability of complex networks under the assumption that each node can control its outgoing edges separately¹¹. The presented conceptual approach, based on edges rather than nodes, was similar to the edge dynamics that was independently proposed in Ref. 10. Our findings showed that as the network degree distribution becomes increasingly heterogeneous, the entire system also becomes easier to control.

Here, we exploit the powerful framework of the MDS, which in bipartite graphs is known as the *Set Covering Problem*, to tackle the controllability of unidirectional bipartite networks. This combinatorial optimisation problem has found many applications in disparate areas that are not related to dynamics or control, such as transportation systems and airline crew scheduling, vehicle routing (e.g., Vehicle Routing Problem with Time Windows (VRPTW) is one of the important problems in distribution and transportation), facility location (e.g., how to optimise the location of a terrestrial cellular network of base stations (cell sites) to cover all the mobile phones), and even probe selection in hybridisation experiments in DNA sequencing^{15–19}. More recently, *hitting set* formulation, which is equivalent to set cover, has been used to uncover 14 anticancer drug combinations using data from 60 tumour derived cell lines²⁰.

Several novel and promising methods aimed to control networks, such as node⁸ and edge dynamics¹⁰, were recently proposed. Our theoretical analysis and simulations suggest that the edge dynamics based on the MDS approach provides an alternative viewpoint to investigate the controllability of complex networks, a goal that is still far from real-world applications. Our study agrees with that developed in the field of edge dynamics¹⁰ in that the shift in perspective from nodes to edges may offer new ways to tackle complex system problems; hence, this approach is worth exploring further. Note that the shift in perspective from nodes to edges has been explored in other areas of network science, with community detection being one of the most recent examples in which the shift is showing promising results²¹.

Our model assumes that more powerful control is possible (because each driver node can control its outgoing links independently), which has the possible drawback of requiring higher costs and may not be possible in some kinds of networks (e.g., metabolic networks, PPI networks).

The developed analytical tools, combined with the evaluation of real-world networks from socio-technical and biological systems, offer a promising framework to control unidirectional bipartite networks with the minimum number of driver nodes. Our analysis unveils the role of the maximum degree H in addressing the network controllability, and how this dependence significantly changes when the power-law degree exponent (γ_1) of the set of nodes that exert the control is above or below the value 2. The theoretical analysis shows that the maximum degree has a significant influence on the size of the DS. Additionally, the analysis also derives the order of nodes (*upper bound*) necessary to control the network. Among all of the topologies, unidirectional bipartite networks with scale-free degree distribution with $\gamma_1 = 1.5$ lead to a smaller upper bound of the number of nodes to be controlled. The dynamics model corresponding to the MDS approach for unidirectional bipartite networks is shown in SI Section IV-A.

To illustrate the MDS approach, consider as an example the human drug-target protein network. Figure 1 shows a small sub-network with five target proteins (disease-gene products) related to cardiovascular disorders. Although up to eleven drugs can interact with these targets, only three drugs, the so-called *cover set* or dominating set for the bipartite network, can potentially control and regulate all of the protein targets simultaneously.

Results

The dominating set and structural controllability in bipartite networks. We consider a bipartite graph $G(V_T, V_\perp; E)$ in which V_T is a set of top nodes, V_\perp is a set of bottom nodes, and E is a set of edges ($E \subseteq V_T \times V_\perp$). Note that all of the edge directions are from V_T to V_\perp in this definition. As discussed later in more detail when analysing real-world networks, this assumption is reasonable in certain cases, such networks as drug-target protein networks because the activities of the nodes in then V_T (drugs) are usually not affected by those in the V_\perp (target proteins). However, this assumption is not reasonable for bipartite networks such as metabolic networks in which nodes in V_T and nodes in V_\perp appear alternately in each path. (See the SI Section IV for details and Fig. S19).

In this work, we use a modified version of the dominating set, in which a set must be selected from V_T ; the set also is sufficient to dominate all of the nodes in V_\perp (i.e., for all nodes $w \in V_\perp$, there exists a node $v \in V_T$ such that $(v, w) \in E$). This version corresponds to a *set cover* problem by associating a set $S_v = \{w | (v, w) \in E\}$ for each $v \in V_T$. We use MDS to denote the minimum dominating set (i.e., the dominating set with the minimum number of nodes) in the above sense. Additionally, our theoretical analysis for bipartite networks also includes the cut-off as observed in real-world networks to address network controllability, a feature that was absent in Ref. 8,10,11.

As proven in Ref. 11, a unipartite network is structurally controllable if a dominating set is selected as a set of control nodes under the assumption that each control node can control its outgoing edges separately. Then, we can consider the structural controllability under the assumption given in Ref. 8, i.e., that each driver node can control only its own value. In such a case, the number of driver nodes is determined by the number of nodes in V^R not appearing in a maximum matching of the adjunct bipartite graph $G'(V^L, V^R; E')$ ⁸. However, in this case, all of the nodes in V^R corresponding to V_T remain unmatched because there is no edge connection to any of these nodes (see Figure 2 and SI Section IV-B for details on the construction of this graph). Figure 2 shows that in contrast with the predictions from nodal dynamics⁸, the MDS approach requires fewer nodes to control the network. Because $|V_T|$ is usually a very large number, in this work, we focus on the structural controllability in terms of the MDS. (See the SI Section IV for a Proposition.) However, it is to be noted that the MDS model assumes that more powerful control is possible (because each driver node can control its outgoing links independently), which has the disadvantage that the cost for control is higher.

Additionally, it is worth mentioning that structural controllability only guarantees that there exists some set of weights rendering the system controllable. The studied real and artificial networks may not correspond to one of these weights. This is particularly true for homogenous networks with many common edge weights. However, our results hold even if the same weights are assigned to all edges. In particular, the control aspect of the problem follows directly from the fact that an edge applying a unique signal to a single integrator node makes the node controllable. Consequently, structural controllability is not even needed. This is one of the merits of our model because it suggests that our model can be applied to a certain kind of nonlinear and/or discrete models. However, as mentioned in SI Section IV-A, our model still has a link with a linear control model.

Theoretical analysis of the MDS size in bipartite networks. We assume that the degree distributions of V_T and V_\perp follow $P_T(k) \propto k^{-\gamma_1}$ and $P_\perp(k) \propto k^{-\gamma_2}$, respectively. We let $n_1 = |V_T|$ and $n_2 = |V_\perp|$. We also assume that all of the nodes in a dominating set DS must be selected from V_T , and that it is necessary to dominate all of the nodes in V_\perp (it is not necessary to dominate the nodes in V_T), which means that DS is a *set cover* for V_\perp . We divide our analysis into two parts

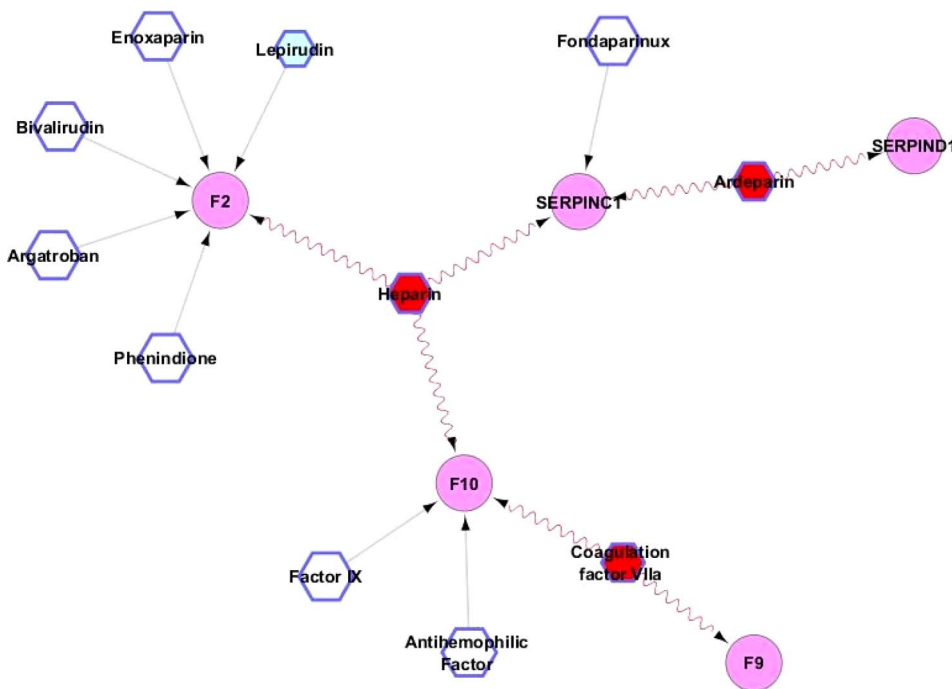


Figure 1 | A small component of the drug-target protein network consisting of 11 drugs (hexagons) interacting with five disease-gene products corresponding to cardiovascular disorder class. Although 11 drugs interact with these proteins, only three drugs (red hexagons) are required to control the system simultaneously. Drugs belonging to the DS are indicated in red. These three drugs are called the cover set (or dominating set DS) of the network. Interactions between the drugs from the DS and the disease-gene products are represented as wavy red arrows.

based on the value of the exponent γ_1 . First, for the case of $\gamma_1 > 2$, we assume $P_T(k)$ follows $\alpha_1 k^{-\gamma_1}$ with a cut-off at $k = n_1$. Let S be the set of nodes with a degree greater than or equal to K . Note that S is chosen so that the total degree (i.e., the number of edges incident to S) is maximised among the sets with the same cardinality. We first estimate the size of $\Gamma(S)$, which will allow us to find an upper bound for the minimum degree K (see the SI for the exact derivation). Then, the size of S is estimated as

$$|S| > \left[\frac{\gamma_1 - 1}{\gamma_1 - 2} \right]^{\frac{\gamma_1 - 1}{\gamma_1 - 2}} \cdot \left(\frac{n_2}{n_1} \right)^{\frac{\gamma_1 - 1}{\gamma_1 - 2}} \cdot n_1. \tag{1}$$

This expression gives us a lower bound of the size of S . From this inequality and the fact that V_T is a trivial dominating set, we can see that the size of the minimum dominating set is $\Theta(n_1)$ (for a fixed γ_1 if n_2 is the same order as n_1).

Next, we consider the case of $1 < \gamma_1 < 2$. Here we focus on the degree distribution for V_T , and thus, we let $n = n_1$, and $m = n_2$. We

assume that the maximum degree (i.e., cut-off) is H . After some calculations (see the SI for the analytic derivation), the upper bound of the size of the dominating set is estimated as

$$O\left(\frac{n^{2-\gamma_1} \cdot m^{\gamma_1-1}}{H^{(2-\gamma_1)(\gamma_1-1)}}\right). \tag{2}$$

If $H = n$, $m = cn$ and c is a constant, the upper bound takes the minimum order ($O(n^{0.75})$) when $\gamma_1 = 1.5$. This result gives insights into which bipartite scale-free network is easier to control with the minimum number of driver nodes. Moreover, this result also identifies the role of the maximum degree for network controllability, a feature that has not been explored in the previous approaches^{8,10,11}.

Computer simulation analysis of MDS in artificial bipartite networks.

To verify the analytical calculations presented above, we examine the size of the MDS using artificially generated bipartite networks. For given values of $n = n_1$, γ_1 , γ_2 and H , we generate random bipartite networks with different sizes, ranging up to 100,000 nodes. The algorithm developed to build bipartite networks with scale-free distributions for both the top and bottom nodes, with specific values of the degree exponents γ_1 and γ_2 and a specific cut-off value, is described in the Methods section. As mentioned above when defining the structural controllability of bipartite networks, the MDS computation of a bipartite network is equivalent to the computation of a minimum set cover. Although the minimum set cover problem is known for NP-hard and, thus, greedy-type approximation algorithms have been proposed²²⁻²⁴, we could obtain optimal solutions for all of the networks examined in this work using integer linear programming (see Methods section). We have verified that the optimal solution is efficiently obtained in scale-free bipartite networks of up to approximately 110,000 nodes.

First, we consider the case in which $\gamma_1 < 2$, and we evaluate the relationship between H and the cover size under the conditions that $\gamma_1 = \gamma_2$ and $|V_T| = |V_\perp| \approx 20,000$. The simulation results shown in Figure. 3 suggest that the maximum degree of the nodes in the

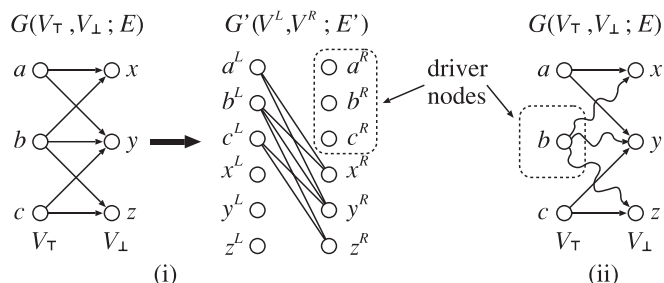


Figure 2 | Comparison of the model in Ref. 8 (i) with the MDS model (ii) for bipartite networks. In this example, $\{b\}$ is the dominating set (i.e., set cover) of $G(V_T, V_\perp; E)$, whereas $\{a^R, b^R, c^R\}$ cannot appear in the maximum matching of $G'(V^L, V^R; E')$ and thus $\{a, b, c\}$ must be the set of driver nodes in the sense of Ref. 8. We see that the MDS approach requires fewer driver nodes (only b) than Ref. 8 to structurally control the network.



network has a significant impact on the cover size. By increasing the maximum degree, the network can be easily controlled. It is worth mentioning that the effect of the cut-off on networks had not been highlighted previously. For different values of γ_1 , the results are in agreement with the behaviour expected by the theoretical analysis.

In particular, the exponent takes the minimum value at $\gamma_1 = 1.5$, which is in good agreement with our theoretical findings (see SI Section IV-C for details). Note that some deviations for large values of H can be observed for higher values of the degree exponent, such as $\gamma_1 = 1.8$. Because we are assuming $m = n$ here, $B = H^{2-\gamma_1}$ holds (see SI), where we assume in the theoretical analysis (see SI for the details) that a dominating set consists of the nodes of a degree between B and H . If γ_1 is close to 2, the value of B becomes a small number (close to 1). Because B is assumed to be the node degree (i.e., an integer), such a small B might lead to an inaccurate estimate. Additionally, because we approximate $B^{2-\gamma_1} - 1$ by $B^{2-\gamma_1}$ (see SI for details), $\gamma_1 \approx 2$ also would lead to an inaccurate estimate. A comparison of the analytically predicted H exponent and that calculated with computer simulations is shown in Figure S16. Because Figure 3 only considers the case in which $\gamma_1 < 2$, we also have performed computations of the case in which $\gamma_1 > 2$, as shown in Figure S17. In this case, dependence does not exist with respect to the maximum degree H . This different behaviour contrasts with the scaling-law observed in Figure 3 for $\gamma_1 < 2$. Our analytical results, as shown in SI, indicate that the dependence with H is $H^{2-\gamma_1}$ for $\gamma_1 > 2$, which vanishes for large H , in agreement with results of the computer simulations.

In Figure S18, the relationship between the cover size and n_2 under the condition that $H = 100$ also shows the log-log scaling for a variety of degree exponents from $\gamma_1 = 1.1$ to $\gamma_1 = 1.9$ that is in fair agreement with the theoretical predictions, except for the case in which $\gamma_1 = 1.1$, where the observed exponent 0.45 is significantly larger than the theoretically estimated exponent (0.1). However, we observe that, for a very large m , the exponent is smaller than 0.45 and is closer to 0.10. If γ_1 is close to 1 and m is small, $B = \frac{n}{m} \cdot H^{2-\gamma_1}$ might be larger than H . Because we assumed in our theoretical analysis that B is no larger than H , such a large value of B might lead to a non-accurate estimate.

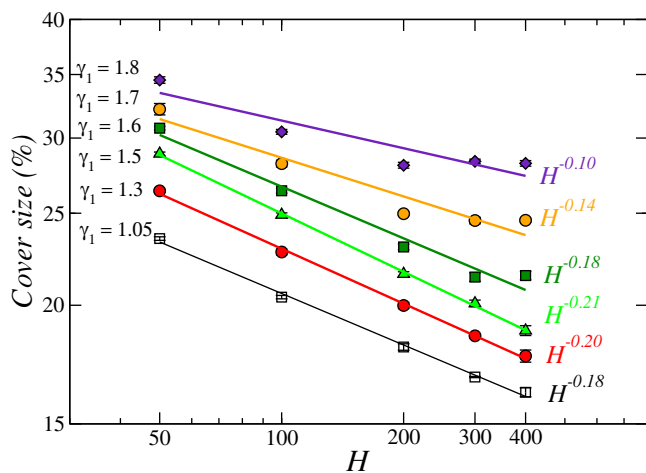


Figure 3 | The dependence between the cover size and H under the condition that $\gamma_1 = \gamma_2$ and $|V_T| = |V_\perp| \approx 20,000$ nodes. The fitted functions for each degree exponent γ are from top to bottom as follows: $H^{(-0.097 \pm 0.022)}$ ($r = 0.9266$), $H^{(-0.137 \pm 0.021)}$ ($r = 0.9652$), $H^{(-0.181 \pm 0.017)}$ ($r = 0.9870$), $H^{(-0.206 \pm 0.003)}$ ($r = 0.9996$), $H^{(-0.195 \pm 0.004)}$ ($r = 0.9991$) and $H^{(-0.180 \pm 0.006)}$ ($r = 0.9984$). The results are averaged over ten realizations. Statistical errors of the exponents are shown together with the correlation coefficients r in parentheses. In most cases error bars are smaller than the symbols in the figure.

Next, we evaluate our theoretical results by considering the case in which $\gamma_1 > 2$ in larger networks. We constructed two sets of bipartite networks with $\gamma_1 = 3$ and $\gamma_2 = 1.5, 2.5$ and 3. In the first set, $|V_T| \approx 50,000$ nodes and in the second set, $|V_T| \approx 100,000$ nodes. The results are shown in Table 1. When Eq. 1 is computed, it can be seen that in all cases, the lower bound gives a smaller value than the size of the cover computed using the optimal algorithm. Note that when only the giant connected component is considered (see Table 1) the cover set is smaller because isolated components are absent in the network.

We have compared the cover set computed in scale-free networks with that from random networks that obey the Poisson distribution. We generated scale-free networks with a variety of degree exponents as shown in Table S1. Because the optimal solution is hard to find in random networks with very large number of nodes, we performed computer simulations using 100 nodes for n_1 and n_2 respectively. The random networks were constructed with the same average degree as that of the corresponding scale-free networks. The results show that in the vicinity of $\gamma_1 = 1.5$, the cover set is significantly smaller than that from the random networks. A detailed mathematical analysis of the random bipartite networks following the Poisson distribution for both the top and the bottom set of nodes is presented in SI Section IV-D. We also computed the analytically derived equations that give an upper bound (Eq. 27 in SI) and lower bound (Eqs. 30 and 33 in SI) for the cover set in random networks. Hence, we verified that the upper bounds predicted for the cover set in random networks are larger than those of the observed cover size, and that the lower bounds predicted are lower in all cases, except the first case which is 26 (observed) vs. 26.5 (predicted), in simulated networks (See Table S1). Although the difference of the MDS size between bipartite scalefree and random networks is not large (due to small n_1 and n_2), it would be much larger for large n_1 and n_2 .

Controllability analysis of the real-world bipartite networks. We have analysed ten real-world bipartite networks from social and biological systems as shown in Table 2. Here, we briefly describe the results for the Facebook-like forum²⁵, the firms-world city network²⁶, the cond-mat scientific collaboration²⁷ and the human drug-target protein network²⁸. For the data sources, the details of the statistics, and the analysis of all of the networks, see the SI Sections II and III. As stated above, a network is structurally controllable if a dominating set is selected as a set of the control nodes under the assumption that each driver node can control its outgoing edges separately. In the Facebook network, V_T is the users set, and each user can decide the topic on which a new message is posted. The cover size for this network represents 10% of all of the users (see Table 2). This small set of nodes could influence the opinions circulating in the forums and with the possibility to induce rapid changes.

Next, we focus on the firms-world cities network. These data reflect the services of 100 global firms distributed across 315 cities worldwide. The computation of the cover set in this bipartite network, in which V_T represents the firms, shows that a very small set of the firms (8%) offers services in all cities. This result suggests that these firms play a prominent role in controlling the socio-economic developments in the world. Again, each firm is able to establish its offered services separately and satisfies the structural controllability assumption for the DS.

The cond-mat scientific collaboration consists of a network of scientists and research papers. Here, we have again satisfied the structural controllability condition because each scientist can choose to investigate each research subject independently. Therefore, the scientists are the V_T set in this network. The results show that 25% of the scientists may induce research opinions and new scientific routes by leading and participating in all of the research performed within the field. However, this cover set size is the largest among the



Table 1 | Computational results of cover set size for computer generated bipartite networks with $H = 200$. The results for the giant connected component (GCC) of each complete network (CN) are also shown. The results were averaged over ten realizations. The standard error of the mean (s.e.m.) is shown in parentheses. The analytical predictions are displayed for $\gamma_1 > 2$ (see Eq. 1). Note that the observed cover size is always larger than the lower bound prediction

Type	γ_1	$ V_T $	Cover size (s.e.m.)	Lower bound	γ_2	$ V_\perp $
CN	3.0	49,997	4,265(12)	190	1.5	6,171
GCC		46,939	2,128(7)	85		4,006
CN	3.0	49,997	27,774(22)	7,305	2.5	38,222
GCC		20,102	4,554(11)	1,716		11,747
CN	3.0	49,997	36,437(18)	12,499	3.0	49,997
GCC		7,319	2,245(6)	1,820		7,300
CN	3.0	100,013	8,454(33)	383	1.5	12,385
GCC		93,886	4,230(14)	175		8,119
CN	3.0	100,013	54,735(82)	14,338	2.5	75,737
GCC		41,856	9,380(33)	3,488		24,167
CN	3.0	100,013	72,674(93)	25,003	3.0	100,013
GCC		15,597	4,651(18)	3,777		15,351
CN	1.5	49,940	10,674(76)	—	1.5	49,940
GCC		49,153	9,949(28)	—		49,147
CN	1.5	9,903	7,138(20)	—	2.5	59,841
GCC		7,570	4,825(18)	—		56,940
CN	1.5	9,903	7,134(8)	—	3.0	75,069
GCC		6,221	4,644(10)	—		69,626

three analysed networks. Note that, in this case, the V_\perp set of nodes does not follow a sufficiently clear power-law and instead exhibits exponential decay (see Figs. S1 for the degree distributions).

The network features of the relationships between all of the drugs and drug targets are important to organise the current knowledge of the relationships between drug targets and disease-gene products, as well as even human therapies^{28–31}. This kind of network representation could aid drug discovery because newly developed drugs could target the disease-gene product that molecularly links the roots of the distinct disorders³². Here, we raise questions on the controllability of the human drug-target protein (DT) network, the system's minimum number of *driver drugs* and their topological role in the network.

In this bipartite network of drug-protein interactions, a drug and a protein are connected to each other if the protein is a known target of the drug. Here, the V_T set of nodes is represented by the drugs that can alter the activity of the targeted protein. The numbers of approved drugs with known human protein targets and drug targets are 888 and 394, respectively. A small fraction of validated disease genes encodes drug-target proteins. The drug targets were assigned a human disorder class if the protein was a disease-gene product. Each gene was assigned to a disorder class as shown in SI, Table S3 in Ref. 32. This information is available in the OMIM database, which reports on the topics of human disorders and disease-related genes. The target proteins encoded by disease genes are colored based on the disorder class to which they belong.

A complete map of the giant component of the bipartite network with a mapping of the identified dominating set of drugs is shown in Figure 4. Figure S5 shows the isolated components of the network. To satisfy the controllability assumption for DS, we have assumed that each drug is designed to interact with specific targets and that these interactions are independently to some extent. The computation of the cover size shows that only 21% of the approved drugs could control the entire known druggable proteome.

Based on the linkage of the genes to disparate disease pathophenotypes³² and the reduced size of the *driver drugs* identified, we suggest that a relatively small number of drugs could address the common genetic origin of these diseases. Moreover, if we consider only the giant connected component, the fraction of the drugs required to control the network is significantly reduced to 8%. Although the average degree of the drugs is 1.81, nonetheless,

the drug cover set shows $\langle k \rangle = 2.2$, indicating that the DS consists, on average, of high-degree drugs (see also Fig. S6). Furthermore, this value increases to 3.59 when only the giant component is considered.

A classification of disorders based on the fraction of the disease-gene products targeted by the cover set is shown in Figure S7. A large fraction of disease-gene products belonging to specific disorder classes are covered by the 12 most highly connected drugs in the DS. These disorder classes include, among others, dermatological, neurological and psychiatric disorders. In contrast, the proteins that belong to a different group of complex disorders such as cancer, immunology and renal disorders tend to be targeted by low-degree drugs in the DS.

To shed light on the topological role of the cover set in the drug-target network, we projected the giant component of the network onto the drug space and computed several network metrics (see Figs. S8–S14). The results reveal that the cover set tends to select the nodes with a high betweenness centrality. This finding suggests that the topological roles played by the cover set and the so-called influential spreaders are somewhat similar³³. The computation of a k -shell decomposition³⁴ (Fig. S15) shows that an important fraction of the most highly connected drugs in the DS occupy the core (or higher shells) of the network. In contrast, fewer drugs are allocated in the periphery of the network.

Even when the cover set is calculated optimally, as in our case, more than one optimal covering can result with the same number of driver nodes. This issue is related to the nature of the optimisation problem in networks rather than to the selected algorithm. The maximum matching used in⁸ does not identify a unique set of driver nodes either. However, it was applied to real networks to understand that the maximum matching identifies the driver nodes whose mean degree is smaller than the mean degree of all of the nodes in real and model networks, showing that in real systems the hubs are avoided by the driver nodes in their model. Here, we also have characterised the role of the driver nodes identified by the optimal dominating set algorithm and we have evaluated their topological properties in a similar manner. Even though a different optimal covering is selected, the topological properties of the nodes in each set do not change significantly because all of the nodes in each optimal set must satisfy the same connectivity conditions.



Table 2 | Computational results of cover set size for real-world networks. Degree exponents and network data are also shown. Degree distributions and their fits are shown in Figs. S1–S4. When the degree distribution does not follow a power-law, γ and x_{min} values are absent. For data sources and references, see SI Section II

Category	Name	γ_1	x_{min}	$ V_T $	Cover size	γ_2	x_{min}	$ V_{\perp} $	Edges
Social	Firms-World Cities	3.27 ± 0.36	77.8 ± 18.6	100 firms	8 (8%)	2.40 ± 0.31	16.6 ± 6.8	315 cities	7,518
	Facebook Forum UCA	3.25 ± 0.23	14.8 ± 3.5	899 users	93 (10%)	2.98 ± 0.29	19.1 ± 3.9	522 topics	7,089
	Davis' Southern Women Club	—	—	18 women	2 (11%)	—	—	14 events	89
	Cond-Mat Sci. Coll.	3.50 ± 0.01	21.3 ± 1.3	16,726 scientists	4,332 (25%)	—	—	22,016 papers	58,595
	Graph Book Bibliography	2.31 ± 0.10	1.1 ± 0.2	314 authors	174 (55%)	—	—	360 papers	613
Biological	The Dutch Elite	—	—	3,810 persons	549 (14%)	3.33 ± 0.16	7.1 ± 1.0	551 administrations	5,221
	Drugs-Targets	2.96 ± 0.15	2 ± 0.4	888 approved drugs	194 (21%)	1.84 ± 0.08	1.1 ± 0.5	394 target proteins	1,593
	Transcriptional network (Yeast)	2.45 ± 0.39	8.0 ± 3.3	131 TFs	75 (57%)	—	—	592 genes	1,079
	ncRNA-protein network (human)	2.05 ± 0.46	3.5 ± 2.0	34 ncRNAs	25 (71%)	—	—	154 proteins	249
	ncRNA-proteins (6 organisms)	1.97 ± 0.31	2.8 ± 1.7	98 ncRNAs	67 (68%)	—	—	425 proteins	647

Discussion

We have presented a methodology to address the previously unexplored structural controllability in bipartite networks. The developed theoretical tools, assisted by computer simulations and the analysis of real-world networks from social and biological systems, allow for a deeper understanding of bipartite networks and shows how to structurally control the complex systems represented by this ubiquitous type of networks.

Our results demonstrate that to control the network, the MDS tends to select the high degree nodes in the bipartite network. The theoretical results together with the analysis of several configurations in model networks shows that a V_T with $\gamma_1 = 1.5$ and a maximum degree of the network H taking a high value, minimise the order of the required number of driver nodes.

These results are very relevant in two respects. First, the finding that a minimum dominating set significantly depends on the maximum degree H unveils a new tool for the control of networks. Second, for $\gamma_1 < 2$, the value of $\gamma_1 = 1.5$ was computed using an upper bound; therefore, the exact value could differ slightly and might be in the vicinity of 1.5, which shows the non-trivial nature of this result. It also is interesting to see how the behaviour of the MDS significantly changes as γ_1 increases above 2 (see Fig. 3 and Fig. S17).

Our theoretical results suggest that the use of edge dynamics, based on the MDS approach, is able to control large scale-free bipartite networks with exponents in the vicinity of $\gamma_1 = 1.5$ using a relatively small set of driver nodes. Whereas previously the genes that are linked to two distant diseases were thought to be important to understand the deep roots of complex disorders³², we found that the minimum set of the approved drugs acting on the target proteins share unique properties, which could be used to develop future drugs. For example, the drugs that belong to the DS occupy core locations in the network so that these drugs bridge multiple disease-gene products, with many of the shortest paths crossing through these drugs. This shows that these drugs have specific chemical features for treating distant disorders.

Our analysis of drug-target interactions illustrates how small the number of drugs can be to address the entire known human drug-gable proteome. Multi-target and optimal combinations of drugs in cancer and HIV have been suggested as new approaches to deal with such complex disease modularity and aid in drug design^{20,35,36}. The drawbacks could exist in the possible antagonistic drug combinations in which the strength of two drugs in the same treatment is weaker than that of either drug alone as well as in the unwanted side effects³⁷.

Although we have assumed that no interactions exist between the drugs used in the drug-target network, it might be possible to adapt the MDS algorithm to include constraints such that certain drugs cannot appear in the MDS simultaneously, which is left as future work. Although edge dynamics¹⁰, as well as our MDS-based edge dynamics, which requires the control of individual edges, are still far from real-world applications, the set cover algorithmic framework has already found many applications^{15–19}, some of which are related to clinical analysis. A recent work has already explored optimal drug combination using a minimal hitting set algorithm, which is equivalent to cover set, that successfully targets the whole population of 60 tumour derived cell lines, uncovering 14 anticancer drug combinations²⁰.

With respect to cellular networks, although both transcriptional and metabolic networks can be represented by bipartite networks, only the former allows the direct application of our methodology because the bidirectional nature of the latter prevents the direct usage of our approach to metabolic pathways (see SI Section V). Although the problem of controllability in unidirectional and bidirectional bipartite networks present several technical differences and challenges, it might be possible to extend our MDS approach to tackle this type of network and contribute to future work in this key issue.

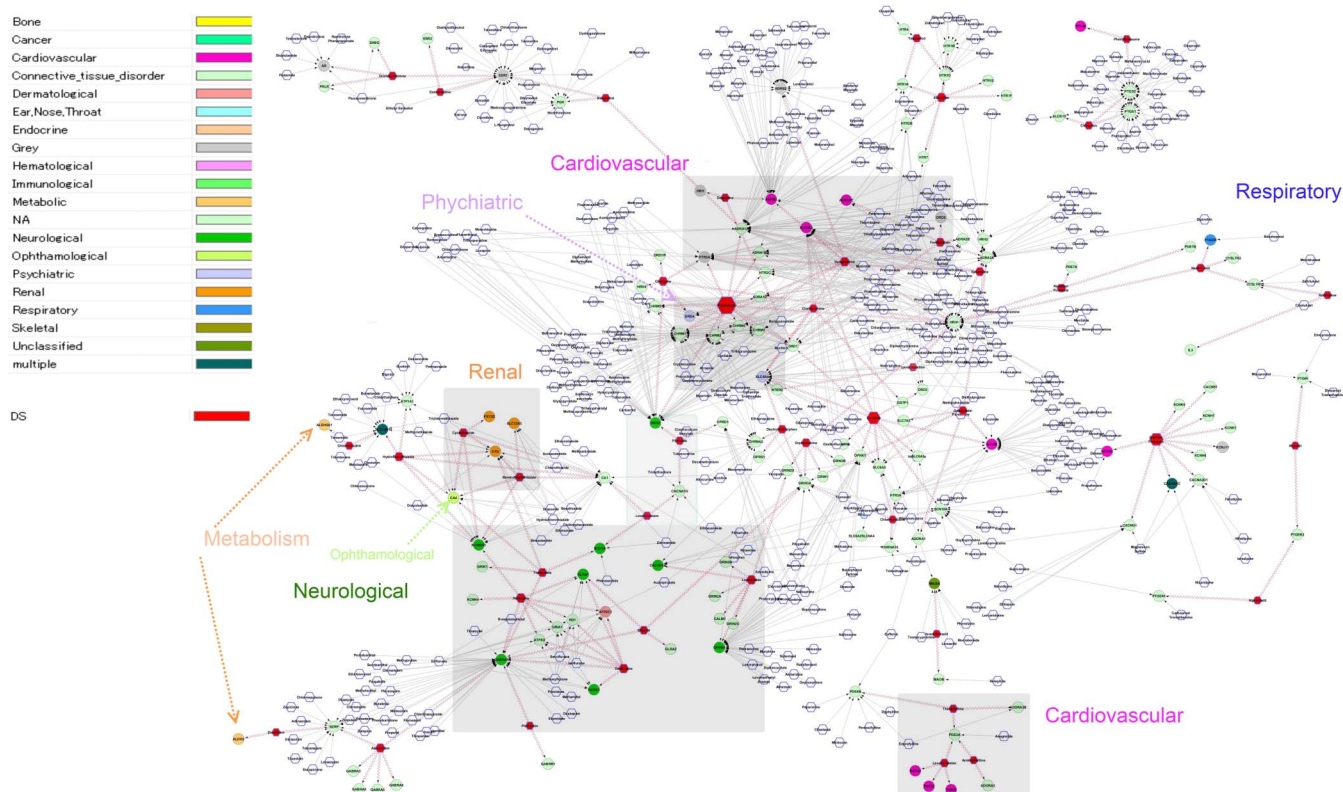


Figure 4 | The full giant component of the drug-target protein network. Drugs (hexagons) belonging to the DS are indicated in red. The interactions between DS drugs and target proteins (circles) are denoted by wavy red arrows. Each node is colored based on the disorder class to which the disease-gene product belongs. The class of a disease gene is assigned if all the related diseases to the gene are of the same class. Otherwise, it is assigned *Grey* class. Disorders having distinct multiple clinical features are assigned to the *multiple* class. When there was insufficient information available for a clear assignment, it was annotated as *unclassified* class. When a gene is not directly implicated in any disorder it was annotated as *NA* (not assigned) class. The rest of legends are self-explanatory (see also Ref. 32).

Methods

Algorithm for generating artificial bipartite network. For given $n = n_1, \gamma = \gamma_1, \gamma_2$ and H , we generate a random bipartite network in the following way.

- (1) For each node $v \in V_T$, generate half edges $e_i = (v, u_i)$ (u_i is a virtual node) according to the degree distribution $\alpha_1 k^{-\gamma_1}$ and the degree cutoff H where α_1 is selected so that the number of nodes in V_T is almost n_1 .
- (2) For each node $w \in V_\perp$, generate half edges $e'_j = (u'_j, w)$ (u'_j is a virtual node) according to the degree distribution $\alpha_2 k^{-\gamma_2}$ and the degree cutoff H where α_2 is selected so that the number of e'_j s is equal to the number of e_j s.
- (3) Randomly connect e_i s and e'_j s in a one-to-one manner.

It is to be noted that n_2 (the number of nodes of V_\perp) is determined automatically in step 2 to satisfy the condition on edge numbers. Although multiple edges between the same node pair may appear, the number of such edges is small and thus these edges should have almost no influence on the results of computer simulation.

Analytical solutions for the size of DS in scale-free bipartite networks. We consider a bipartite graph $G(V_T, V_\perp; E)$, where V_T is a set of top nodes, V_\perp is a set of bottom nodes, and E is a set of edges ($E \subseteq V_T \times V_\perp$). The directions of the edges are considered from V_T to V_\perp . Therefore, the set of driver nodes will be a subset of V_T . We then consider the case of having scale-free degree distributions for both the top and bottom nodes. That is, the degree distribution of V_T and V_\perp follow $P_T(k) \propto k^{-\gamma_1}$ and $P_\perp(k) \propto k^{-\gamma_2}$, respectively. A similar analysis could be conducted for the asymmetric case, where one of the distribution shows an exponential decay. Furthermore, a cut-off is assumed for the degree distributions as observed in real-world networks. Then, the analytic derivations for the expected fraction of the minimum driver nodes that control bipartite networks are presented in the SI Section IV.

Computation of the MDS in bipartite networks. We have introduced the structural controllability of bipartite networks and showed that the computation of a MDS of a bipartite network is equivalent to the computation of a minimum set cover. Although it is an NP-hard problem, we have verified that the optimal solution is obtained in networks with power-law distributions of up to approximately 110,000 nodes within a few seconds. The computation was formalized as the following Integer Linear Programming (ILP) problem

$$\begin{aligned} & \text{minimize} \sum_{v \in V_T} x_v \\ & \text{subject to} \sum_{\{v, u\} \in E} x_v \geq 1 \quad \text{for all } u \in V_\perp \\ & x_v \in \{0, 1\} \quad \text{for all } v \in V_T \end{aligned} \quad (3)$$

where the optimal solution was calculated using 'glpsol' solver (<http://www.gnu.org/software/glpk>).

By using artificially generated bipartite networks and the optimal cover algorithm, we examined the following properties: (1) the relationship between H and the cover size under the conditions that $\gamma_1 = \gamma_2$ and $|V_T| = |V_\perp| \approx 20,000$ and (2) the relationship between n_2 and the cover size under the condition that $H = 100$, where n_2 is controlled by varying γ_2 .

1. Newman, M. E. J. *Networks: An Introduction* (Oxford University Press, New York, 2010).
2. Barabási, A.-L. & Oltvai, Z. N. Network biology: understanding the cell's functional organization. *Nature Review Genetics* **5**, 101–113 (2004).
3. 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).
4. Lombardi, A. & Hornquist, M. Controllability analysis of networks. *Phys. Rev. E* **75**, 56110 (2007).
5. Liu, B., Chu, T., Wang, L. & Xie, G. Controllability of a leader-follower dynamic network with switching topology. *IEEE Trans. Automat. Contr.* **53**, 1009–1013 (2008).
6. Rahmani, A., Ji, M., Mesbahi, M. & Egerstedt, M. Controllability of multi-agent systems from a graph-theoretic perspective. *SIAM J. Contr. Optim.* **48**, 162–189 (2009).
7. Kim, D.-H. & Motter, A. E. Slave nodes and the controllability of metabolic networks. *New Journal of Physics* **11**, 113047 (2009).
8. Liu, Y.-Y., Slotine, J.-J. & Barabási, A.-L. Controllability of complex networks. *Nature* **473** 167–173 (2011).



9. Müller, F.-J. & Schuppert, A. Few inputs can reprogram biological networks. *Nature* **478**, E4 (2011).
10. Nepusz, T. & Vicsek, T. Controlling edge dynamics in complex networks. *Nature Physics* **8**, 568–573 (2012).
11. 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).
12. Lin, C. Structural controllability. *IEEE Trans. Automat. Contr.* **19**, 201–208 (1974).
13. Slotine, J.-J. & Li, W. *Applied Nonlinear Control* (Prentice-Hall, 1991).
14. Haynes, T. W., Hedetniemi, S. T. & Slater, P. J. *Fundamentals of Domination in graphs* (Chapman and Hall/CRC Pure Applied Mathematics, New York, 1998).
15. Bramel, J. & Simchi-Levi, D. On the effectiveness of set covering formulations for the vehicle routing problem with time windows. *Operations Research* **45** (2), 295–301 (1997).
16. Amaldi, E., Capone, A., Malucelli, F. & Mannino, C. *Optimization problems and models for planning cellular networks* (Handbook of Optimization in Telecommunications, 2006).
17. Balas, E. & Padberg, W. Set partitioning: A survey. *SIAM Review* **18**, 710–716 (1976).
18. Borneman, J., Chrobak, M., Vedova, G. D., Figueroa, A. & Jiang, T. Probe selection algorithms with applications in the analysis of microbial communities. *Bioinformatics* **17**, S39–48 (2001).
19. Umetani, S. & Yagiura, M. Relaxation heuristics for the set covering problem. *Journal of the Operations Research Society of Japan* **50**, 350–375 (2007).
20. Vázquez, A. Optimal drug combinations and minimal hitting sets. *BMC Systems Biology* **3**, 81 (2009).
21. Ahn, Y.-Y., Bagrow, J. P. & Lehmann, S. Link communities reveal multi-scale complexity in networks. *Nature* **66**, 761–764 (2010).
22. Chvátal, V. A greedy heuristic for the set-covering problem. *Mathematics of Operations Research* **4**, 233–235 (1979).
23. Lovász, L. On the ratio of optimal integral and fractional covers. *Discrete Mathematics* **13**, 383–390 (1975).
24. Johnson, D. S. Approximation algorithms for combinatorial problems. *Journal of Computer and System Sciences* **9**, 256–278 (1974).
25. Opsahl, T. Triadic closure in two-mode networks: Redefining the global and local clustering coefficients. *Social Networks* **34**, doi:10.1016/j.socnet.2011.07.001. (2012).
26. Taylor, P. J. *World city network: a global urban analysis* (London: Routledge, 2000).
27. Newman, M. E. J. The structure of scientific collaboration networks. *Proc. Natl. Acad. Sci. USA* **98**, 404–409 (2001).
28. Yildirim, M. A. *et al.* Drug-target network. *Nature Biotech.* **25**, 1119–1126 (2007).
29. Barabási, A.-L., Gulbahce, N. & Loscalzo, J. Network medicine: a network-based approach to human disease. *Nature Reviews Genetics* **12**, 56–68 (2011).
30. Spiro, Z., Kovacs, I. A. & Csermely, P. Drug therapy networks and the prediction of novel drug targets. *Journal of Biology* **7**, 20 (2008).
31. Nacher, J. C. & Schwartz, J.-M. A global view of drug-therapy interactions. *BMC Pharmacology* **8**, 5 (2008).
32. Goh, K.-I. *et al.* The human disease network. *Proc. Natl. Acad. Sci. USA* **104**, 865–8690 (2007).
33. Kitsak, M. *et al.* Identification of influential spreaders in complex networks. *Nature Physics* **6**, 888–893 (2010).
34. Carmi, S., Havlin, S., Kirkpatrick, S., Shavitt, Y. & Shir, E. A model of Internet topology using k-shell decomposition. *Proc. Natl. Acad. Sci. USA* **104**, 11150–11154 (2007).
35. Yang, K. *et al.* Finding multiple target optimal intervention in disease related molecular networks. *Mol. Syst. Biol.* **4**, 228 (2008).
36. Csermely, P., Agoston, V. & Pongor, S. The efficiency of multi-target drugs: the network approach might help drug design. *Trends Pharmacol. Sci.* **26**, 178–182 (2005).
37. Motter, A. E. Improved network performance *via* antagonism: From synthetic rescues to multidrug combinations. *Bioessays* **32**, 236–245 (2010).

Acknowledgments

T.A. was partially supported by MEXT, Japan (Grant-in-Aid 22650045) and J.C.N. was partially supported by MEXT, Japan (Grant-in-Aid 23700352).

Author contributions

J.C.N. and T.A. designed the research and participated in the analytical and numerical calculations. J.C.N. analysed the empirical data and wrote the manuscript, where T.A. wrote a part of the SI. Both authors read and reviewed the manuscript.

Additional information

Supplementary information accompanies this paper at <http://www.nature.com/scientificreports>

Competing financial interests: The authors declare no competing financial interests.

License: This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivs 3.0 Unported License. To view a copy of this license, visit <http://creativecommons.org/licenses/by-nc-nd/3.0/>

How to cite this article: Nacher, J.C. & Akutsu, T. Structural controllability of unidirectional bipartite networks. *Sci. Rep.* **3**, 1647; DOI:10.1038/srep01647 (2013).