Controllability and observability analysis for vertex domination centrality in directed networks

Bingbo Wang1,2, Lin Gao1, Yong Gao3, Yue Deng1 & Yu Wang1

1School of Computer Science and Technology, Xidian University, Xi’an, 2School of Computer Science and Technology, Xi’an University of Technology, Xi’an, 3Department of Computer Science, Irving K. Barber School of Arts and Sciences, University of British Columbia Okanagan, Kelowna.

Topological centrality is a significant measure for characterising the relative importance of a node in a complex network. For directed networks that model dynamic processes, however, it is of more practical importance to quantify a vertex’s ability to dominate (control or observe) the state of other vertices. In this paper, based on the determination of controllable and observable subspaces under the global minimum-cost condition, we introduce a novel direction-specific index, domination centrality, to assess the intervention capabilities of vertices in a directed network. Statistical studies demonstrate that the domination centrality is, to a great extent, encoded by the underlying network’s degree distribution and that most network positions through which one can intervene in a system are vertices with high domination centrality rather than network hubs. To analyse the interaction and functional dependence between vertices when they are used to dominate a network, we define the domination similarity and detect significant functional modules in glossary and metabolic networks through clustering analysis. The experimental results provide strong evidence that our indices are effective and practical in accurately depicting the structure of directed networks.
but this process embodies only one aspect of a vertex’s power in dominating a system. If the state variables are looped back, the feedback signal can then control a system within itself. State feedback is therefore the ability to examine the role that a vertex plays in both its upstream system through feedback, and this process reflects the mechanics of directed network topology and function by providing insights into the extent to which a vertex can influence the system.

In this paper, we focus on the domination centrality (DC) index to assess the capabilities of vertices in directed networks. Intuitively, domination centrality includes two aspects: control capability and observation capability. Under the minimum-control-cost condition, for a single vertex, the control capability captures the dimension of the controllable subspace and quantifies the extent of influence that can be exerted on the downstream subnetwork through this vertex. Similarly, the observation capability captures the dimension of the observable subspace and quantifies the intervention that can be exerted on the upstream subnetwork through this vertex.

The purpose of emphasizing the global minimum cost is to determine the responsibility and capability of each individual vertex cooperating with others in dominating the entire system. Mathematically, domination centrality is the harmonic mean of these two capabilities and represents the capability of a vertex synthetically. This approach is in good agreement with our original notion regarding the “power” of a vertex in dominating the entire network. Inspired by this general consideration, we perform statistical studies of the index DC for several types of real-world directed networks, including citation, metabolic, glossary and synthetic networks, and analyse the underlying topological factors by which the distribution of DC is primarily determined. To uncover DC and functions of vertices, a clustering analysis is presented based on the intuitive assumption that vertices that control and observe the same subspace tend to serve identical functions in a network. Our domination centrality index bridges the concepts of directed network topology and function by providing useful insights into the extent of the former on the latter from the viewpoint of cybernetics.

**Results**

**Domination centrality.** Consider the linear time-invariant dynamic system \( \dot{X}(t) = AXX(t) + B \cdot u(t) \), \( Y(t) = CXX(t) \) with the state vector \( X \in \mathbb{R}^n \), the adjacency matrix \( A \in \mathbb{R}^{n \times n} \), the input matrix \( B \in \mathbb{R}^n \), and the output matrix \( C \in \mathbb{R}^m \) and the output vector \( Y \in \mathbb{R}^m \). The underlying directed network of this system is denoted by \( G(A) \), with vertex set \( V \) and link set \( L \). The rank of the \( n \times nm \) controllability matrix \( Q_C = [B, AB, A^2B, \ldots, A^{n-1}B] \), which is denoted by \( \text{rank}(Q_C) \), provides the dimension of the controllable subspace of the structural system \((A, B, C)^{\text{\text{\dagger\dagger}}12,13} \). \((A, B, C)\) is completely controllable if \( \text{rank}(Q_C) = n \). Analogously, the rank of the \( n \times n \) observability matrix \( Q_{O} = [C]^{\tau}, \ [CA]^{\tau}, [CA^2]^{\tau}, \ldots, \ [CA^{n-1}]^{\tau} \), which is denoted by \( \text{rank}(Q_{O}) \), provides the dimension of the observable subspace of this system. \((A, B, C)\) is completely observable if \( \text{rank}(Q_{O}) = n \). Furthermore, the duality theorem\(^{12}\) indicates that system \((A, B, C)\) is completely controllable if and only if system \((A^\tau, C^\tau, B^\tau)\) is completely observable, and vice versa.

Liu’s Minimum Input Theorem\(^{16}\) states that the minimum number of driver vertices \((N_D)\) required to fully control a network \(G(A)\) is one if there is a perfect matching in \(G(A)\). Otherwise, it is equal to the number of unmatched vertices with respect to any maximum matching, \(N_D = \max(n - |M|, 1)\). A maximum matching is a link set \( M \subseteq L \) with maximum cardinality (size), and no two links in \( M \) may share a common starting vertex or a common ending vertex. A vertex is matched if it is an ending vertex of a link in \( M \). \(|M|\) denotes the size of the maximum matching.

For a given maximum matching link set \( M \) of a directed network \(G(A)\), the minimum-control-cost configuration \(CF(V, M, \cup AL)\) carries the structural information of completely control\(^{17,18}\). \(CF\) is a spanning subnetwork of \(G(A)\), with vertex set \( V \) and link set \( M \cup AL \subseteq L \). \(M\) is a stem-cycle disjoint cover of \(G(A)\) and indicates the directed routes along which the input control signals are transmitted. \(AL\) is the set of additional links that begin in vertices of stems (except the top vertices) and end in vertices of cycles. The \( n \times n \) adjacent matrix \(A(M)\) is used to indicate the wiring diagram of the spanning subnetwork \(CF\) that corresponds to the maximum matching link set \( M \) of \(G(A)\). As an example, in figure 2(a), the red links are elements of a maximum matching. When vertices are connected by red links, the network thus constructed is composed of vertex-disjoint stems (two in shades of green) and cycles (four in shades of red); \(l_{3,7} \) and \(l_{4,9} \) are the additional links that connect the stems and cycles.

The controllability of a complex network concentrates on the interaction structure in which the pattern of influence may be known, but not the specific extent of influence. In response to unknown or uncertain edge weights, the controllability is used to uncover the generic properties of systems, independent of parameter values. The cactus is the most economical topology-structure pattern to propagate control influence, since the cactus is a minimal structure with maximum matching. A maximum matching shows the important links by which we can construct the cactus structures efficiently in a complex system. Therefore, the maximum matching not only reveals the minimum driver set but also consists of a backbone of the key control routes, which are a stem-cycle cover of the original network. The minimum-control-cost configuration \(CF\) is just constructed for showing the backbone of the propagation of control influence.

To quantify the control capability of a single vertex \(i\) under the minimum-control-cost condition, \(B\) reduces to the vector \(b^i\) with a single non-zero entry, and \(A\) reduces to the matrix \(A(M)\). Then, the control capability of a single vertex \(i\) can be defined as

\[
\text{rank}(Q_C(M)) = \text{rank} \left[ b^i, A(M)b^i, (A(M)^2)b^i, \ldots, (A(M)^{n-1})b^i \right].
\]  

Lin’s theorem\(^{11}\) has demonstrated that a linear control system \((A, B)\) is structurally controllable if and only if the associated digraph \(G(A)\) can be spanned by cacti. A cactus is a subnetwork in the form of a distinct stem or a stem connected to several buds. A stem is simply an elementary path that originates from an input vertex. The initial (or terminal) vertex of a stem is known as the root (or top) of the stem. A bud is an elementary cycle with an additional link that ends, but does
not begin, in a vertex of the cycle, and the top vertex of the stem is not 
the initial vertex of any additional link. The network can be spanned 
by cacti using links of $A(M)$. Thus, $A(M)$ demonstrates the manner in 
which vertices control the entire network under the minimum-con-
trol-cost condition. When the vertex $i$ is taken as an input vertex, the 
subspace that is accessible from vertex $i$ in the spanning subnetwork 
$CF$ is cactus-structured and structurally controllable. A vertex 
j is called accessible if there is at least one directed path that passes from 
the input vertex $i$ to vertex $j$. For example, in figure 2(b), the accessible 
subspace of vertex 1 is highlighted in bold purple and spanned by 
the links of this $CF$ in the form of a cactus.

We can therefore use the size of the accessible subspace of vertex $i$ 
as an accurate measure of $\text{rank}(Q(M))$. Thus, equation (1) can be 
represented by 

$$\text{rank}(Q(M)) = |\text{CS}(M)|,$$  \hspace{1cm} (2)
of $V, \mathcal{M}^T \cup \mathcal{A}L'$, where $\mathcal{A}L'$ is the set of additional links in $G(A')$, and $A'(M')$ can be used to indicate the wiring diagram of $OF$ that corresponds to $M'$ in $G(A')$. As an example, in figure 2(a), $I_{12,18}$ is the only additional link in $\mathcal{A}L'$. To quantify the observation capability of a single vertex $i$ under the minimum-observation-cost condition, the output matrix $B^r$ reduces to the vector $(B^o)^T$ with a single non-zero entry, and $A^r$ reduces to the matrix $A'(M')$. Then, the observation capability can be represented by the size of the observable subspace $OS(M')$ of vertex $i$ in $OF(V, \mathcal{M}^T \cup \mathcal{A}L')$ and can be accurately measured as follows:

$$\text{rank}(Q_i(M')^T) = \text{rank}\left[ (b_i^0)^T, (b_i^0)^T A^T(M'), \right.$$

$$\left. (b_i^0)^T (A^T(M'))^2, \ldots, (b_i^0)^T (A^T(M'))^{n-1} \right]^T,$$

$$\text{rank}(Q_i(M')^T) = |OS(M')|,$$

where $OS(M') = \{ij|j\text{ is accessible from vertex } i \}
\text{in } OF(V, \mathcal{M}^T \cup \mathcal{A}L') \}$.

Considering the role that a vertex plays in both controlling the downstream subspace and observing the upstream subspace, the domination centrality ($DC$) index for the assessment of the capabilities of vertices in directed networks can be synthetically defined as the harmonic mean of a vertex's control and observation capability. The domination centrality of vertex $i$ is represented by

$$DC = \frac{2}{\text{rank}(Q_i(M))^T + \text{rank}(Q_i(M')^T)}.$$

The $DC$ index is used to detect the most powerful vertex through which we can not only control but also observe a network. Therefore, as the harmonic mean of the control capability and observation capability of the vertex, $DC$ will be significant only when the control capability and observation capability attain high values simultaneously. In figure 2(b), for the given maximum matching $M_1$, the controllable subspace of vertex 1, $CS(M_1)$, is highlighted by a purple dotted line and has $\text{rank}(Q_i^1(M_1)) = 10$, and the observable subspace, $OS(M_1^T)$, is highlighted by a green dotted line and has $\text{rank}(Q_i^1(M_1)) = 7$; thus, the domination centrality of vertex 1 is $DC = 2/(1/10 + 1/7) \approx 8.2$, and vertex 1 is powerful in dominating the network. By contrast, vertex 14 has the highest value of control capability but a very small observation capability, meaning that $DC_{14} = 2/(1/13 + 1/1) \approx 1.9$. Thus, vertex 1 has a stronger overall ability to dominate the network than does vertex 14. In the worst case, when a vertex $i$ can only control and observe itself, $DC = 1$.

Furthermore, we note that there are multiple different maximum matchings (N! matchings for a complete connected network). Each one illustrates a unique manner in which vertices may control and observe the entire network under a minimum-cost condition. Therefore, in combination with other vertices, a vertex may play several different roles in dominating a network. Thus, we may ask this question: in all possible minimum-cost configurations, if two vertices can perform similar control and observation functions, does that fact indicate that they can also play similar functions in intervening in the system? To answer this question, the domination similarity (DS) is defined as

$$DS(i,j) = \text{agm}(JC(i,j), JO(i,j)),$$

where $\text{agm}(x, y)$ is the arithmetic-geometric mean of two positive real numbers $x$ and $y$. We calculate the Jaccard similarity coefficient of the complete controllable subspaces of $i$ and $j$ to determine their control-function similarity. Meanwhile, the Jaccard similarity coefficients of the complete observable subspaces of $i$ and $j$ are calculated to determine their observation-function similarity. The complete controllable subspace $CS = \bigcup_{k=1}^K CS(M_k)$ and the complete observable subspace $OS = \bigcup_{k=1}^K OS(M_k^T)$, where $K$ is the number of different maximum matchings. $JC(i, j) = \text{Jaccard}(CS, CS), JO(i, j) = \text{Jaccard}(OS, OS)$. $\text{agm}(x, y)$ is a number between the geometric and arithmetic means of $x$ and $y$; thus, $DS(i, j)$ will be significant only when $JC(i, j)$ and $JO(i, j)$ attain high values simultaneously. Furthermore, in the case that there is a large difference between the two quantities, $\text{agm}(x, y)$ yields a more reasonable result than the arithmetic or harmonic mean.

Figure 2 vividly illustrates this concept. $M_1$ and $M_2$ are two different maximum matchings of this toy network, with links highlighted in red in figure 2(a) and figure 2(c), respectively. We concentrate on the domination capabilities of vertices 1 and 2. In figure 2(b), for vertex 1, the controllable subspace $CS(M_1)$ in $CF(A(M_1))$ is indicated in purple, and the observable subspace $OS(M_1^T)$ in $OF(A(M_1^T))$ is indicated in green. Similarly, the situation for vertex 2 is illustrated in figure 2(d), with orange and blue colours corresponding to $M_2$. A great deal of information regarding the functions of these two vertices can be determined based on the overlapping of their controllable and observable subspaces, as shown in figure 2(e).

**Distribution of domination centrality.** If a structural system can be shown to be controllable for almost all weight combinations and the dimension of the controllable subspace is stable, in the sense that for almost any set of system parameters, the dimension is equal to some maximal constant (the generic rank of the controllability matrix), all these properties also hold for observability. Thus, to some extent, domination centrality and domination similarity can be calculated without assessing the link weights. This property is one of the greatest advantages of controllability-based topological measures: they are robust to uncertainty in link weights, which frequently arises in networks constructed from real data, such as biological networks.

In this section, we perform statistical studies of the domination centrality on several types of real-world directed networks, including citation, glossary, metabolic and synthetic scale-free networks, as summarised in table 1. We have manually reconstructed the global human enzyme-centric network based on data available in the August 2009 release of the Kyoto Encyclopedia of Genes and Genomes (KEGG). The citation and glossary networks are drawn from Pajek datasets and can be downloaded at http://vlado.fmf.uni-lj.si/pub/networks/data/. The synthetic scale-free networks were constructed using the method of Fan et al. In table 1, we provide the statistical values of the numbers of vertices ($n$), links ($m$) and minimum driver vertices ($N_{\Delta}$) for the original networks.

We first consider the distribution of the domination centrality. For a given network, any existing algorithm can be used to compute a maximum matching $M$. For this $M$, the domination centrality reveals the responsibility and capability of each individual vertex in controlling and observing the system with the global minimum cost. Figure 3 presents the distribution of the domination centrality for the synthetic scale-free networks listed in table 1. In double-logarithmic coordinates, the relation between the $DC$ value and the probability $P(DC)$ is nearly linear, suggesting the coexistence of a few powerful
Table 1 | Summarized statistics for the original representative networks

| Type                          | Name           | n     | m     | $N_0$ |
|-------------------------------|----------------|-------|-------|-------|
| Glossary                      | GlossTG        | 67    | 122   | 32    |
| Citation                      | SmaGri         | 1024  | 4918  | 511   |
| Metabolic(enzyme-centric)     | Homo sapiens   | 689   | 2382  | 149   |
| Synthetic Scale Free ($k_0$)  | SF $\gamma = 2.1$ | 5000  | 14972 | 2059  |
|                               | SF $\gamma = 2.4$ | 5000  | 14989 | 1583  |
|                               | SF $\gamma = 3$  | 5000  | 14996 | 1007  |
|                               | SF $\gamma = 4$  | 5000  | 14997 | 532   |

In addition, we calculate the mean, the average of absolute deviation and the relative entropy for the distribution of the domination centrality in each real network and their random counterparts in Table 3. Compared to the real networks, the rand-Degree counterparts yield similar mean values, similar averages of absolute deviation and small relative entropies. The same indices of the rand-ER counterparts differ significantly in comparison. From all these observations, we conclude that domination centrality is, to a great extent, encoded by the degree distribution of the underlying network.

Another interesting phenomenon observed in this study is that the hubs (vertices of high degree) do not tend to play more important roles in dominating a system. We divide the vertices into three groups of equal size according to their degree $k$ (low, medium and high) and calculate the average values of DC among the low-degree, medium-degree and high-degree vertices. As Table 2 demonstrates, for real networks and two random network models (Erdős–Rényi and scale-free), the average DC value of the set of low-degree vertices is not significantly lower than that of the set of hubs in each case. Figure 4(a) graphically represents the values for the Homo sapiens networks. In figure 4(b, c, d), as expected, in all cases, a low-degree vertex can also have a significant domination centrality. For a vertex with a degree equal to 1, either the control capability or the observation capability must also be equal to 1; thus, as the harmonic mean of these two capabilities, the domination centrality must be less than 2. Intuitively, a vertex with a degree of 1 must have either no downstream space it can control or no upstream space it can observe. This is the reason why the hubs are observed to attain slightly larger DC values than the low-degree vertices. To conclude, this experimental study demonstrates that there is no obvious correlation between the degree and the DC. This result is very useful in the following sense: the most effective method by which we can

Figure 3 | The distribution of the domination centrality in double-logarithmic coordinates. The results for scale-free synthetic directed networks with $N = 5000$, $(k_0) = (k_{\text{rand}}) = (k)/2 = 3$, $p(k_{\text{in}}) \sim k_{\text{in}}^{-\gamma}$ and $p(k_{\text{out}}) \sim k_{\text{out}}^{-\gamma}$ are shown.
intervene in a system’s dynamics is to identify vertices with great domination capability, which are not restricted to hubs alone.

**Clustering analysis.** In fact, a consensus among the topological criteria for measuring the functional similarity of vertices is often lacking in directed networks. Domination similarity is a direction-specific index and concentrates on quantifying the unique relation between the upstream and downstream subspaces of vertices in directed networks. Independent of the weights of the links used in the calculation, the domination similarity is a parameter-free index for analysing data with noise. With the global minimum-cost limitation, the domination similarity represents the ability of vertices to work synergistically with others and provides guidance for dominating a system using multiple vertices operating cooperatively at the minimum cost. In this section, we apply the DS index to detect and analyse functional modules in a glossary network and the enzyme-centric network of *Homo sapiens*.

We utilise the DS values as the input of the AP algorithm to identify the functional modules in the glossary network. In this case, we test the performance on a directed word network that has also been recently introduced by Newman and Boccaletti. The network represents the connections among a set of technical terms, such as “Tree” and “Digraph”, contained in a glossary of network jargon. Vertices represent terms, and a directed link from one vertex to another exists in the network iff the second term is used to describe the meaning of the first term. Because circular definitions are unhelpful and are normally avoided, most links in the network are not reciprocal. The statistics for this network are provided in table 1.

![Figure 4](https://www.nature.com/scientificreports)

**Figure 4** A schematic diagram illustrating the domination centrality in *Homo sapiens* networks. (a): The average values of domination centrality among low-, medium- and high-degree vertices. The scatter plots of the domination centrality versus the vertex in-degree, out-degree and degree are presented in panel (b), panel (c) and panel (d), respectively. The green, blue and purple plots represent the real network, rand-Degree network and rand-ER network, respectively.

---

| Table 2 | Summarized statistics for the domination centrality values in representative networks |
|---------|-------------------------------------------------|
| Network                      | $N_D^{\text{DS}}$ (change rate) | $N_D^{\text{ER}}$ (change rate) | DC | $D_{\text{DS}}^{\text{DS}}$ (change rate) | $D_{\text{ER}}^{\text{DS}}$ (change rate) |
|-----------------------------|---------------------------------|---------------------------------|-----|---------------------------------|---------------------------------|
| **GlossTG**                 | 32 (0.00%)                      | 10 (32.84%)                     | 1.31/1.44/2.47 | 1.25/1.74/1.85 | 2.94/3.31/5.14 |
| **SmaGri**                  | 458 (5.18%)                     | 8 (49.12%)                      | 1.19/1.68/1.95 | 1.35/1.91/2.59 | 41.04/42.72/44.92 |
| **SciNet**                  | 1075 (2.97%)                    | 81 (39.39%)                     | 1.84/1.34/1.76 | 2.11/2.42/2.73 | 20.08/23.52/23.50 |
| **Kohonen**                 | 2039 (2.01%)                    | 178 (51.35%)                    | 1.19/1.41/1.68 | 1.21/1.50/1.88 | 11.20/14.58/14.86 |
| **Homo sapiens**            | 174 (3.63%)                     | 24 (18.14%)                     | 2.59/3.77/5.36 | 2.32/3.94/4.84 | 16.58/20.21/19.65 |
| **SF γ = 2.1**              | 2103 (0.88%)                    | 360 (24.46%)                    | 1.29/1.67/2.31 | 1.26/1.61/2.25 | 8.66/11.3/12.3 |
| **SF γ = 2.4**              | 1633 (1.00%)                    | 1 (0.00%)                       | 1.6/2.34/3.06 | 1.59/2.27/2.93 |                         |
| **SF γ = 3**                | 982 (0.50%)                     | 517 (5.18%)                     | 2.44/3.71/4.18 | 2.51/4.07/4.74 |                         |
| **SF γ = 4**                | 4.97/7.06/7.56                  | 517 (0.30%)                     | 4.95/6.80/7.33 |                         |                         |

*The domination centrality average values of Low/Medium/High degree vertices.*
interventional effect of enzymes in the metabolic system and detect the relevant modules. We cluster the network into modules using the AP algorithm augmented by our DS index to identify the pathways that correspond to metabolic functions. In total, 63 modules are detected by our method. We measure the biological quality of the clustering result by means of Gene Ontology (GO) enrichment and use the tool GO TermFinder to compute the functional enrichment p-values of components with respect to their biological process annotations. In the results, 28 modules are annotated by GO terms with p-values ≤ 0.01 (the most significant p-value = 3.45E-16), which means that these modules represent significant biological functions in the metabolic system. In figure 6, certain representative modules are depicted with their corresponding GO terms and p-values. We note that not only dense subnetworks but also functional modules, with distinctive circle and path structures, are detected. These experiments provide compelling evidence that the DS is a meaningful and practical indicator in accurately depicting the structures of directed networks.

Discussion
The domination capability of a vertex reflects the vertex’s ability to interfere in dynamical control processes in many directed complex systems. The key task is to explain the manner in which a vertex

| Network          | Meana | Average Deviationa | Relative Entropyb |
|------------------|--------|--------------------|-------------------|
| GlossTG          | 1.77/1.62/3.76 | 0.69/0.46/1.61 | 0.0386/0.5244 |
| SmaGri           | 1.61/1.96/42.87 | 0.46/0.87/18.80 | 0.1047/3.2409 |
| SciMet           | 1.73/1.93/22.37 | 0.51/0.71/13.56 | 0.043/2.7694 |
| Kohonen          | 1.43/1.54/13.58 | 0.31/0.43/8.21 | 0.02/2.8511 |
| Homo sapiens     | 3.96/3.736/18.826 | 2.42/2.33/11.01 | 0.0732/1.3426 |
| SF γ = 2.4       | 2.32/2.26/10.94 | 1.02/0.98/7.18 | 0.0049/1.5441 |

*a* In Original/randDegree/rand-ER networks.
*b* RandDegree/rand-ER relative to the original networks.

Figure 5 | Module-detection results in the directed glossary network. The modules are labelled with colours.
intervenes in its downstream and upstream spaces in the implementation of dynamical functions. In this paper, based on the determination of controllable and observable subspaces under the minimum-cost condition, we introduced the DC index to assess the capabilities of vertices in directed networks. The results of our statistical studies demonstrate that the domination centrality is, to a great extent, encoded by the degree distribution of the underlying network, yet there is no discernible correlation between the degree and DC of a vertex. This result provides guidelines for the selection of the most effective means through which we can intervene in a system’s dynamics. Furthermore, to analyse the cooperative relations among vertices in the domination of an entire network, we defined the domination similarity, and we were able to detect significant functional modules in glossary and metabolic networks through clustering. As direction-specific and parameter-free indexes, DC and DS are effective and practical in accurately depicting the structures of directed networks. In our future studies, we intend to investigate the most effective approach to intervening in the dynamical functions of complex systems through selected vertices.

Methods

Enumerating all possible different maximum matching link sets $M_s$ is infeasible when calculating $DS$, as in the worst case scenario, there may be an exponential number of them. However, we note that there are many “redundant” links in real networks that may never appear in any maximum matching. Based on their role in the $M_s$, links can be classified into three categories: “critical” links must appear in all $M_s$, “redundant” links may never appear in any one of them and “ordinary” links play roles in some, but not all, $M_s$. In combination with the sparseness of real networks, we can approximate the control subspaces and observation subspaces of vertices via an optimisation routine. As shown in figure 7, we observe the beneficial phenomenon that in a real network, there always exists some consistent set of control subspaces and observation subspaces of vertices induced by different maximum matchings. This observation supports the feasibility of using a small number of maximum matchings to approximate the complete control and observation subspaces of vertices. A random optimisation can be performed rather quickly using a Markov sampling process.

Algorithm for $DS$.
Input network $G(A)$

\[
\theta = 1, \tau = 1, t = 0
\]

do

Markov random sampling to produce a maximum matching $M$

\[
CS(M) = \{ jj \text{ is accessible from vertex } i \text{ in } CF(V,(M \cup AL)) \}
\]

\[
OS(M^T) = \{ jj \text{ is accessible from vertex } i \text{ in } OF(V,(M^T \cup AL')) \}
\]

\[
CS' = CS \cup CS(M), OS' = OS \cup OS(M^T)
\]

if $|\theta - t|/\tau \leq t = t + 1$ else $t = 0$

$\tau = 0$

While $t \leq \psi$

Calculate Domination Similarity

Figure 6 | Module-detection results in the *Homo sapiens* network. Certain representative modules are marked with colours and the corresponding GO terms and p-values are given.
In every loop, we randomly produce a maximum matching \( M \) and update the complete control subspace \( CS \) and complete observation subspace \( OS \) of each vertex by merging the additional accessible vertices introduced in this \( M \). \( CS \) and \( OS \) are added but never deleted throughout the entire procedure. \( \theta \) is the sum of all \( CS \) and \( OS \). If the rate of increase of \( \theta \) is less than \( \varepsilon \) for \( \psi \) continuous loops, the random optimisation procedure terminates, and we then calculate the Jaccard similarity coefficients of the \( CS \) and \( OS \) of two arbitrary vertices. The growth rates of \( \theta \) during the random optimisation procedures for the GlossTG, Homo sapiens and SmaGri networks are presented in figure 7. Clear improvement in \( \theta \) is achieved for 90, 7845 and 4527 maximum matchings in 247, 27995 and 15172 random samples for GlossTG, Homo sapiens and SmaGri, respectively. We note that the growth rate of \( \theta \) rapidly decreases to nearly 0 as the sampling number increases. This observation supports the appropriateness of using only a certain number of \( M \)s to approximate the domination similarities of vertices in real networks. We set \( \varepsilon = 0.000001 \) and \( \psi = 50 \) in the clustering-analysis case studies.

A Markov process, as described by Jia et al.\(^36,37\), performs unbiased random sampling among all maximum matchings and can be used to estimate the role of each vertex in controlling the network. This algorithm randomly chooses a vertex in a given \( M \), enumerates all alternative maximum matchings that include all other elements except this vertex by removing all its links, then randomly chooses one of these alternative maximum matchings as the current \( M \) and repeats the process.

However, removing a vertex in a random sampling may not be effective for calculating \( DS \). Usually, we identify a maximum matching in a bipartite graph and attempt to increase the matching size via an augmenting path that begins at a matched vertex, ends at an unmatched vertex and alternates between unmatched and matched links on the path. For example, in figure 8, a bipartite graph that is separated into the \( out \) and \( in \) sets; the red link set \( \{l_{1,2},l_{2,3},l_{3,4}\} \) forms a maximum matching, \( M_1 \); and the blue path is an augmenting path when the matched link \( l_{1,2} \) is removed. (c): A new maximum matching \( M_2 \) constructed by alternating the blue augmenting path.

The growth rates of the sum of the complete control and observation subspaces as functions of the number of random samples of maximum matchings in the GlossTG, Homo sapiens and SmaGri networks are shown.

Figure 7 | The influence of the number of random samples among all maximum matchings on the domination-similarity result. The growth rates of the sum of the complete control and observation subspaces as functions of the number of random samples of maximum matchings in the GlossTG, Homo sapiens and SmaGri networks are shown.

In fact, the maximum matchings reveal the functions and the roles that the vertices and links play for controlling the whole network with minimum cost. Different combinations of ordinary links constitute different maximum matchings and produce different choices of minimum-control-cost configurations. Ordinary links are alternatives for constructing the backbone of the propagation of control influence. Therefore, we consider each combination of ordinary links (COL) to be one state. The set of ordinary links of a network is \( \{l_1,l_2,...,l_v\} \), where \( v \) is the number of ordinary links. The Markov chain can be characterised by a transition matrix \( P \) with the elements \( P_{ij} = T_{ij} \times (1 - Q_j) \times Q_i \), where \( Q_i \) is the probability of ordinary link \( l_i \) being included in an \( M \). The transition from state \( i \) to state \( j \) requires the choice of a matched link from an \( M \), with a probability of \( T_{ij} = 1/|M| \); the choice of a COL set that excludes \( l_i \), with a probability of \( 1 - Q_i \); and the choice of a COL set that includes \( l_i \), with a probability of \( Q_i \). Clearly, \( P_{ij} \neq P_{ji} \); our algorithm is not guaranteed to choose each set.

Figure 8 | A schematic diagram illustrating the process of random sampling. (a): The original network. (b): A bipartite graph separated into the \( out \) and \( in \) sets; the red link set \( \{l_{1,2},l_{2,3},l_{3,4}\} \) forms a maximum matching, \( M_1 \). Proceeding from this maximum matching, we randomly choose vertex 2 and leave the current matched vertices and links unchanged. Instead of removing all links of vertex 2, we delete only the matched link \( l_{1,2} \). Then, we can identify an augmenting path that begins at the relevant matched vertex 1 and ends at the presently unmatched vertex 2; in the figure, this path is indicated by a blue line. Finally, by alternating between unmatched and matched links on this blue path, we obtain a new maximum matching \( M_2 \) in which the matching of vertex 2 has been replaced, as shown in figure 8(c). By contrast, because it removes all links of vertex 2, the method of Jia et al.\(^37\) cannot produce any new maximum matching from the given \( M_2 \). Nevertheless, we use the Markov process defined by these authors to perform unbiased random sampling among all maximum matchings to estimate \( DS \). The only difference is that we also enumerate all alternative maximum matchings that include all other elements except the matching link of the chosen vertex.

In fact, the maximum matchings reveal the functions and the roles that the vertices and links play for controlling the whole network with minimum cost. Different combinations of ordinary links constitute different maximum matchings and produce different choices of minimum-control-cost configurations. Ordinary links are alternatives for constructing the backbone of the propagation of control influence. Therefore, we consider each combination of ordinary links (COL) to be one state. The set of ordinary links of a network is \( \{l_1,l_2,...,l_v\} \), where \( v \) is the number of ordinary links. The Markov chain can be characterised by a transition matrix \( P \) with the elements \( P_{ij} = T_{ij} \times (1 - Q_i) \times Q_j \), where \( Q_i \) is the probability of ordinary link \( l_i \) being included in an \( M \). The transition from state \( i \) to state \( j \) requires the choice of a matched link from an \( M \), with a probability of \( T_{ij} = 1/|M| \); the choice of a COL set that excludes \( l_i \), with a probability of \( 1 - Q_i \); and the choice of a COL set that includes \( l_i \), with a probability of \( Q_i \). Clearly, \( P_{ij} \neq P_{ji} \); our algorithm is not guaranteed to choose each set.

Figure 9 | The distribution of the counts in each of the 164 COLs in the GlossTG network. (a): A matched link \( l_i \) is randomly selected from an \( M \) with a probability of \( 1/|M| \). (b) The selection probability is adjusted based on the number of alternative COLs that are enumerated by our sampling procedure.
of matched links with equal probability. For example, in figure 9(a), for the real network "GlossTG" with 364 COLs, we perform 239,000 iterations of our sampling algorithm and count the number of times that each COL is picked. We find that a few COLs are sampled many times, but it is very difficult to ensure that all COLs are sampled at least once. Thus, we adjust the transition matrix $P$ and construct a new transition matrix $P'$ with the elements $P'_{ij} = P_{ij} \times (1 - Q_i) \times Q_j$. If we set $T_i^1 = T_i^0 \times (1 - Q_i) \times Q_j$, then $P_{ij}' = P_{ij}$, meaning that the transition matrix $P$ is symmetric and the steady-state distribution possesses equal probabilities for all states. However, $Q_i$ cannot be determined effectively; in practice, $u_i / \sum u_i$ is used to approximate $Q_i$, where $u_i$ is the average number of all alternative COLs that can be enumerated by removing $i$, the first $|M|^2$ iterations of our sampling procedure. Intuitively, if any alternative COLs can be enumerated by removing $l$, then the probability of choosing $l$ from an $M$ should be increased. With this modification, the sampling procedure becomes more efficient, and the 164 COLs in "GlossTG" can be obtained within 20,000 iterations. As shown in figure 9(b), we perform our modified sampling algorithm 193,500 iterations and count the number of times that each COL is picked. The result demonstrates that this procedure provides a more even-handed random sampling among all maximum matchings.

1. Newman, M. Networks: an introduction. (Oxford University Press, Oxford, 2009).
2. Freeman, L. C. A set of measures of centrality based on betweenness. Sociometry 40, 35–41 (1977).
3. Barthelemy, M. Betweenness centrality in large complex networks. Eur. Phys. J. B 38, 163–168 (2004).
4. Sabidussi, G. The centrality index of a graph. Psychometrika 31, 581–603 (1966).
5. Bonacich, P. & Lloyd, P. Eigenvector-like measures of centrality for asymmetric relations. Social Networks 23, 191–201 (2001).
6. Estrada, E. & Rodriguez-Velazquez, J. A. Subgraph centrality in complex networks. Phys. Rev. E 71, 056103 (2005).
7. Brin, S. & Page, L. The anatomy of a large-scale hypertextual Web search engine. Comput. networks and ISDN syst. 30, 107–117 (1998).
8. Delvenne, J.-C. & Libert, A.-S. Centrality measures and thermodynamic formalism for complex networks. Phys. Rev. E 83, 046117 (2011).
9. Fortunato, S. & Flammini, A. Random walks on directed networks: the case of PageRank. Int. J. Bifurcation Chaos 17, 2343–2353 (2007).
10. Liu, Y.-Y., Slotine, J.-J. & Barabasi, A.-L. Controllability of complex networks. Nature 473, 167–173 (2011).
11. Lin, C.-T. Structural controllability. IEEE T. Automat. Contr. 19, 201–208 (1974).
12. Luenberger, D. Introduction to dynamic systems: theory, models, and applications. (John Wiley & Sons, New York, 1979).
13. Hermann, R. & Krener, A. Nonlinear controllability and observability. IEEE T. Automat. Contr. 22, 728–740 (1977).
14. Liu, Y.-Y., Slotine, J.-J. & Barabasi, A.-L. Observability of complex systems. Proc. Natl Acad. Sc. USA 110, 2460–2465 (2013).
15. Liu, Y.-Y., Slotine, J.-J. & Barabasi, A.-L. Control centrality and hierarchical structure in complex networks. PloS one 7, e44459 (2012).
16. Wang, B., Gao, L. & Gao, Y. Control range: a controllability-based index for node significance in directed networks. J. Stat. Mech-Theor. E. 2012, P04011 (2012).
17. Kalman, R. E. Mathematical description of linear dynamical systems. SIAM, Ser. A: Contr. 1, 152–192 (1963).
18. Poljak, S. On the generic dimension of controllable subspaces. IEEE T. Automat. Contr. 35, 367–369 (1990).
19. Hosoe, S. Determination of generic dimensions of controllable subspaces and its application. IEEE T. Automat. Contr. 25, 1192–1196 (1980).
20. Kanehisa, M. et al. From genomics to chemical genomics: new developments in KEGG. Nucleic Acids Res. 34, D354–D357 (2006).
21. Chung, F. & Lu, L. Connected components in random graphs with given expected degree sequences. Ann. Comb. 6, 125–145 (2002).
22. Kuhn, H. W. The Hungarian method for the assignment problem. Res. Log. 2, 83–97 (1955).
23. Hopcroft, J. E. & Karp, R. M. An $n^{5/2}$ algorithm for maximum matchings in bipartite graphs. SIAM J. on computing 2, 225–231 (1973).
24. Bollobas, B. Random graphs. (Cambridge university press, Cambridge, 2001).
25. Barabasi, A.-L. & Albert, R. Emergence of scaling in random networks. Science 286, 509–512 (1999).
26. Ruppert, D. Statistics and data analysis for financial engineering. (Springer, New York, 2011).
27. Kullback, S. & Leibler, R. A. On information and sufficiency. Ann. Math. Stat. 22, 79–86 (1951).
28. Lancichinetti, A. & Fortunato, S. Community detection algorithms: a comparative analysis. Phys. Rev. E 80, 056117 (2009).
29. Borwein, J. M. & Borwein, P. B. The arithmetic-geometric mean and fast computation of elementary functions. SIAM rev. 26, 351–366 (1984).
30. Frey, B. J. & Dueck, D. Clustering by passing messages between data points. Science 315, 972–976 (2007).
31. Newman, M. E. The structure and function of complex networks. SIAM rev. 45, 167–256 (2003).
32. Boccaletti, S., Latora, V., Moreno, Y., Chavez, M. & Hwang, D.-U. Complex networks: Structure and dynamics. Phys. Rep. 424, 175–308 (2006).
33. Duarte, N. C. et al. Global reconstruction of the human metabolic network based on genomic and bibliometric data. Proc. Natl Acad. Sci. USA 104, 1777–1782 (2007).
34. Ashburner, M. et al. Gene Ontology: tool for the unification of biology. Nat. Genet. 25, 25–29 (2000).
35. Boyle, E. L. et al. GO::TermFinder—open source software for accessing Gene Ontology information and finding significantly enriched Gene Ontology terms associated with a list of genes. Bioinformatics 20, 3710–3715 (2004).
36. Jia, T. et al. Emergence of bimodality in controlling complex networks. Nat. Commun. 4, 2002 (2013).
37. Jia, T. & Barabasi, A.-L. Control Capacity and A Random Sampling Method in Exploring Controllability of Complex Networks. Sci. Rep. 3, 2354 (2013).
38. Wang, B., Gao, L., Gao, Y. et al. Maintain the structural controllability under malicious attacks on directed networks. Europhys. Letts. 101, 58003 (2013).
39. Yuan, Z., Zhao, C., Di, Z., Wang, W. X. & Lai, Y. C. Exact controllability of complex networks. Nat. Commun. 4, 2447 (2013).
40. Cornelius, S. P., Kath, W. L. & Motter, A. E. Realistic control of network dynamics. Nat. Commun. 4, 1942 (2013).
41. Sun, J. & Motter, A. E. Controllability transition and nonlocality in network control. Phys. Rev. Lett. 110, 208701 (2013).
42. Ruths, J. & Ruths, D. Control Profiles of Complex Networks. Science 343, 1373–1376 (2014).

Acknowledgments

This work was supported by the National Natural Science Foundation of China (Grant No. 91130006, 61303122, 61101057 & 61303118) and the Fundamental Research Funds for the Central Universities (Grant No. BDZ201404).

Author contributions

B.-W. designed and performed experiments, analyzed data and wrote the paper; L.G. and Y.G. designed experiments and wrote the paper; Y.D. and Y.W. designed experiments and wrote the paper.

Additional information

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

How to cite this article: Wang, B.B., Gao, L., Gao, Y., Deng, Y. & Wang, Y. Controllability and observability analysis for vertex domination centrality in directed networks. Sci. Rep. 4, 5399; DOI:10.1038/srep05399 (2014).

This work is licensed under a Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International License. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in the credit line; if the material is not included under the Creative Commons license, users will need to obtain permission from the license holder in order to reproduce the material. To view a copy of this license, visit http://creativecommons.org/licenses/by-nc-sa/4.0/