

# Hierarchy Measure for Complex Networks

Enys Mones<sup>1</sup>, Lilla Vicsek<sup>2</sup>, Tamás Vicsek<sup>1,3\*</sup>

**1** Department of Biological Physics, Eötvös Loránd University, Budapest, Hungary, **2** Institute of Sociology and Social Policy, Corvinus University of Budapest, Budapest, Hungary, **3** Biological Physics Research Group of Hungarian Academy of Sciences, Budapest, Hungary

## Abstract

Nature, technology and society are full of complexity arising from the intricate web of the interactions among the units of the related systems (e.g., proteins, computers, people). Consequently, one of the most successful recent approaches to capturing the fundamental features of the structure and dynamics of complex systems has been the investigation of the networks associated with the above units (nodes) together with their relations (edges). Most complex systems have an inherently hierarchical organization and, correspondingly, the networks behind them also exhibit hierarchical features. Indeed, several papers have been devoted to describing this essential aspect of networks, however, without resulting in a widely accepted, converging concept concerning the quantitative characterization of the level of their hierarchy. Here we develop an approach and propose a quantity (measure) which is simple enough to be widely applicable, reveals a number of universal features of the organization of real-world networks and, as we demonstrate, is capable of capturing the essential features of the structure and the degree of hierarchy in a complex network. The measure we introduce is based on a generalization of the *m*-reach centrality, which we first extend to directed/partially directed graphs. Then, we define the global reaching centrality (GRC), which is the difference between the maximum and the average value of the generalized reach centralities over the network. We investigate the behavior of the GRC considering both a synthetic model with an adjustable level of hierarchy and real networks. Results for real networks show that our hierarchy measure is related to the controllability of the given system. We also propose a visualization procedure for large complex networks that can be used to obtain an overall qualitative picture about the nature of their hierarchical structure.

**Citation:** Mones E, Vicsek L, Vicsek T (2012) Hierarchy Measure for Complex Networks. PLoS ONE 7(3): e33799. doi:10.1371/journal.pone.0033799

**Editor:** Stefano Boccaletti, Technical University of Madrid, Italy

**Received:** January 2, 2012; **Accepted:** February 17, 2012; **Published:** March 28, 2012

**Copyright:** © 2012 Mones et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Funding:** This work was supported by the EU FP7 COLLMOT Grant No: 227878. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

**Competing Interests:** The authors have declared that no competing interests exist.

\* E-mail: vicsek@hal.elte.hu

## Introduction

The last decade has witnessed an explosive growth of interest in the analysis of complex natural, technological and social systems that permeate many aspects of everyday life. These systems are typically made of many units. Complexity arises from either the structure of the interactions between very similar units or, alternatively, the units and the interactions themselves can have specific characteristics. In both cases, the abstract representation of a complex system can be achieved by a collection of nodes (units) and edges (representing interactions between the units) forming a network (or graph).

Research on networks has considerably profited from using both the standard and novel techniques developed in the field of statistical mechanics [1–3]. Although a remarkable body of knowledge has accumulated about the statistical properties of networks [4], a number of questions are still open. The issue of hierarchy has attracted the attention of a great number of social and natural scientists [5]. It has been argued that hierarchy is present in a wide range of complex systems: such as physical, chemical, biological, and social systems [6]. Recent empirical findings demonstrate that hierarchy is present in many of the related networks: in the dominant-subordinate hierarchy among animals [7], in the hierarchy of the leader-follower network of pigeon flocks [8], in rhesus macaque kingdoms [9], in the structure of the transcriptional regulatory network of *Escherichia coli* [10], or in a wide range of social and technological networks [5]. All of

these examples suggest that hierarchy is an important feature of natural, artificial and social networks.

It is important to distinguish between the three major types of hierarchies: the *order*, the *nested* and the *flow* hierarchies. In case of an order hierarchy, hierarchy is regarded to be basically only an “ordered set”, and it is understood to be “equivalent to an ordering induced by the values of a variable defined on some set of elements” [11] (i.e., generally there is no network behind this concept). In case of a nested hierarchy higher level elements consist of and contain lower level elements, or, as [12] has formulated “larger and more complex systems consist of and are dependent upon simpler systems and essential system-component entities”. When a network is structured in a flow hierarchy (mostly directed graphs), the nodes can be layered in different levels so that the nodes that are influenced by a given node (are connected to it through a directed edge) are at lower levels.

Our observation is that the notions of “hierarchy” and the “level of hierarchy” are very closely related. In fact, without a proper measure of hierarchy the notion of hierarchy cannot be complete. Indeed, there are various definitions of hierarchy, or, in other words, there is no unique, widely accepted definition of the notion of hierarchy itself. Correspondingly, we propose that a good measure of hierarchy can serve as a starting point for finding the best definition of hierarchy.

In this paper, we are interested in flow hierarchy for the following reasons. First, order hierarchy is a single-valued function

over the population and there is no underlying network of interactions attached to the hierarchy. Secondly, uncovering a nested hierarchy is analogous to community detection, for which there are known methods [13,14]. Finally, both order and nested hierarchies can be converted to flow hierarchies. In an order hierarchy, a directed edge can be assigned to each pair of adjacent members in the hierarchy and this produces a chain of directed edges. In a nested hierarchy, a virtual node is assigned to every subgraph, and if a subgraph contains another, then the two corresponding virtual nodes are connected with a directed link, which produces a flow hierarchy on the network of virtual nodes.

Among the many exciting questions related to hierarchy [5] is concerned with its origin. Several studies have approached this problem from a historical viewpoint [15,16] but without any quantitative description. The best known quantitative model for the evolution of hierarchies is the Bonabeau model [17]. According to this model, a hierarchy can emerge as the result of the outcomes of competitions between pairs of participating units, and a hierarchy itself is defined by a rank (*order*) assigned to each participating unit [17]. Another interesting result comes from game theory: simulations of prisoners dilemma type dynamics on adaptive networks showed that cooperation combined with imitation can lead to a hierarchical structure [18]. Note, however, that in this model every node can imitate at most one other, and therefore, the emerging hierarchy is by definition a directed tree.

Usually, a hierarchy is the consequence of the different roles, significances and histories of the nodes [17,19]. In other words, if the influence of the nodes on others (and thence, on the whole system) differs, then a hierarchy can emerge. Nodes with the strongest influence can denote the leaders of a group (as in the structure of a company or hidden groups [20,21]; or amongst homing pigeons [8]), central proteins in transcription regulatory networks [10,22] or opinion leaders [23,24]. These nodes can have a major impact on the system, and thus, finding them and quantifying the extent of hierarchy at the same time is an important step in the understanding of functionality and controlling of networks.

In most cases networks contain all sorts of edges (both directed and undirected, various edge weights [strength]) making the detection of hierarchy a difficult challenge. When one looks at real-life networks the picture is often much more complicated than for the simple treelike hierarchy: there can be (i) relations between entities on the same level, (ii) “shortcuts” when a step in the hierarchy is bypassed, (iii) ties which, instead of going downward on the hierarchy, go upward, (iv) even cycles of connected nodes [25] and (v) clusters [26], etc. It can even happen that some or all of the levels of hierarchy cannot be clearly defined (are not well-separated).

The hierarchy measures proposed so far have various undesirable properties that make their application to all classes of complex networks problematic: they (i) use free parameters that are unknown for many networks [20,27], (ii) quantify only the deviation of the network from the tree and penalize loops or multiple edges [28], and (iii) are applicable only to fully directed or fully undirected graphs [20,27–29]. Here we are aiming at introducing a measure which can be equally used for all sorts of networks and thus, used for uncovering universal features of the hierarchical organization of the relations within a complex system.

Visualizing the structure of networks has been a widely used approach to obtain a qualitative picture about some of their features (e.g., clusters/modules). At present, the hierarchical visualization of networks is mostly based on the Sugiyama method [30], which offers an informative and clear hierarchical layout for small networks. However, (i) for networks with more than 2–300

nodes the generated layout becomes difficult to understand; (ii) the meaning of the levels is not defined at all; (iii) independently of the presence or absence of a hierarchy in the given network, the method generates a hierarchical layout that is often misleading; (iv) all steps of the Sugiyama method are NP-complete or NP-hard [31,32], which makes the usage of several different heuristics necessary and thus, results become less well-defined.

Clearly, there is a need for (a) a measure of hierarchy that is free of the above-mentioned undesired properties and (b) a method for the hierarchical visualization of networks that is unbiased, unambiguous and easily applicable even to large graphs. Thus, the two main goals of our paper are to provide a universally applicable measure and a visualization technique of the hierarchical structure of complex large networks.

## Results

### Definition of the global reaching centrality

**Unweighted directed networks.** We are looking for a measure that is expected to satisfy the following natural and reasonable conditions:

1. Absence of free parameters and *a priori* metrics in the definition.
2. The definition should be for unweighted directed graphs (digraphs) and it should be easily extendable to both weighted and undirected graphs.
3. The hierarchy measure should be helpful for generating a layout of the graph.

To arrive at an appropriate definition, we quantify the concept of flow hierarchy, where nodes contribute to the dynamics of the system differently. We first define the *local reaching centrality* of node  $i$  in an unweighted directed graph,  $G$ , as the generalization of the *m-reach centrality* [33] to  $m = N$  (where  $N$  is the number of nodes in  $G$ ). The local reaching centrality,  $C_R(i)$ , of node  $i$  is the proportion of all nodes in the graph that can be reached from node  $i$  via outgoing edges. In other words,  $C_R(i)$  is the number of nodes with a finite positive directed distance from node  $i$  divided by  $N - 1$ , i.e., the maximum possible number of nodes reachable from a given node. We aim to define hierarchy as a heterogeneous distribution of the local reaching centrality. Thus, in graph  $G$  we denote by  $C_R^{max}$  the highest local reaching centrality and define the *global reaching centrality* (GRC) as:

$$GRC = \frac{\sum_{i \in V} [C_R^{max} - C_R(i)]}{N - 1} \quad (1)$$

Here,  $V$  denotes the set of nodes in  $G$ . For normalization, the sum is divided by  $N - 1$ , as this is the maximal value of the enumerator. In the  $GRC = 1$  case the graph has only one node with nonzero local reaching centrality (i.e., it is a star graph). Throughout this paper, for the model networks and real networks we use this directed, unweighted type of  $C_R$ .

It is worth mentioning that in the special case of a tree graph, a recursive equation can be derived for  $C_R(i)$ . This equation has some formal similarities with the one for the *complexity measure* introduced by Huberman et al. [34,35], but with some important differences related to the motivation, details of the recursive equations involved, etc.

**Weighted and undirected networks.** Generalizations to weighted or undirected graphs are straightforward based on the definition of the local reaching centrality. For the generalization of the GRC to weighted directed graphs, we introduce a simple

variant of the local reaching centrality:

$$C'_R(i) = \frac{1}{N-1} \sum_{j:0 < d^{out}(i,j) < \infty} \left( \frac{\sum_{k=1}^{d^{out}(i,j)} \omega_i^{(k)}(j)}{d^{out}(i,j)} \right) \quad (2)$$

Here  $d^{out}(i,j)$  is the length of the directed path that goes from  $i$  to  $j$  via out-going edges and  $\omega_i^{(k)}(j)$  is the weight of the  $k$ -th edge along this path (link weight is assumed to be proportional to connection strength). If nodes  $i$  and  $j$  are connected by more than one directed shortest path, then the one with the maximum weight (i.e., maximum strength) should be used. This extension of the local reaching centrality measures the average weight of a given directed path starting from node  $i$  in a weighted directed graph. If we set  $\omega_i^{(k)}(j)=1$  for every  $i, j$  and  $k$ , then the original local reaching centrality (defined for unweighted directed graphs) is recovered.

To generalize the local reaching centrality to undirected unweighted graphs, we remove the  $\sum_{k=1}^{d^{out}(i,j)} \omega_i^{(k)}(j)$  term from the previous definition and obtain

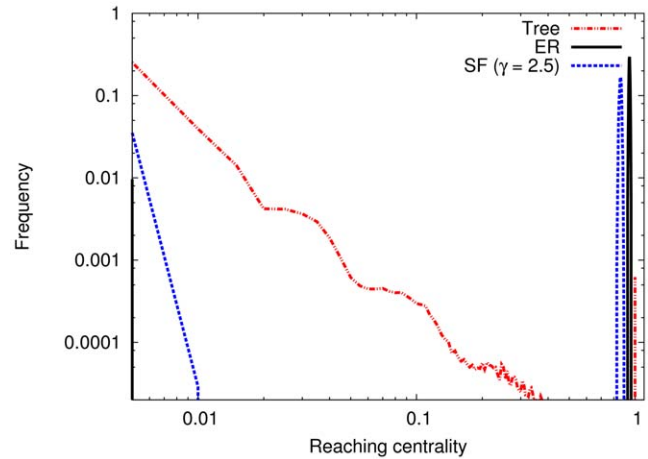
$$C''_R(i) = \frac{1}{N-1} \sum_{j:0 < d(i,j) < \infty} \frac{1}{d(i,j)} \quad (3)$$

This quantity is very similar to the local *closeness centrality* defined by Sabidussi in [36]. In fact, this is equivalent to the generalization of the closeness centrality for disconnected graphs given by Opsahl [37].

### Classical random networks

In order to demonstrate the basic features of the GRC, we briefly discuss its behavior for a few well-known network types. For Erdős-Rényi (ER) graphs [38,39], scale-free (SF) [40–42] graphs and directed trees (more precisely arborescences with random branching number [43,44]), the distribution of  $C_R$  is markedly different (the curves in Figure 1 are averages for 1000 random graphs of each type). In every case, the exponent for the SF networks was set to  $\gamma=2.5$ . For the directed tree, the distribution follows a power-law that is distorted due to the random branching numbers. Directed trees have a maximally heterogeneous distribution of  $C_R$ , thus, based on our arguments above, they are maximally hierarchical. Note that the hierarchical tree (directed tree) has very few nodes with local reaching centrality close to 1.

This is in contrast with the ER and SF graphs in which most of the nodes have a large local reaching centrality. Since almost every node has the same centrality, the contribution of the nodes in Eq. 1 for the ER and SF graphs is negligible. Note that not only the GRC, but also the standard deviation of  $C_R$  increases with the heterogeneity of the graph. The values of GRC are shown in Table 1 together with the standard deviation of the distribution. However, the GRC itself is more suitable for quantifying the heterogeneity of the graph for two reasons. On the one hand, the accuracy of the standard deviation of  $C_R$  is worse than that of the GRC (it has larger deviation on the ensemble of graphs). On the other hand, the standard deviation of  $C_R$  is much smaller for the directed tree than for the ER, which is in contrast to our definition making the tree maximally hierarchical. In summary, we find that, based on their reaching centralities, ER graphs are not hierarchical at all, as expected, and SF graphs are slightly hierarchical.



**Figure 1. An adjustable hierarchical network with the different edge types.** The blue edges belong to the original arborescence graph that is used as the backbone of the adjustable hierarchical (AH) network. There are three type of possible edges added to the graph: down edges (green), horizontal edges (orange) and up edges (red). They have different effects on the hierarchical structure of the directed tree. Down edges conserve the hierarchy, horizontal edges has a slight influence and up edges make strong changes in the structure. doi:10.1371/journal.pone.0033799.g001

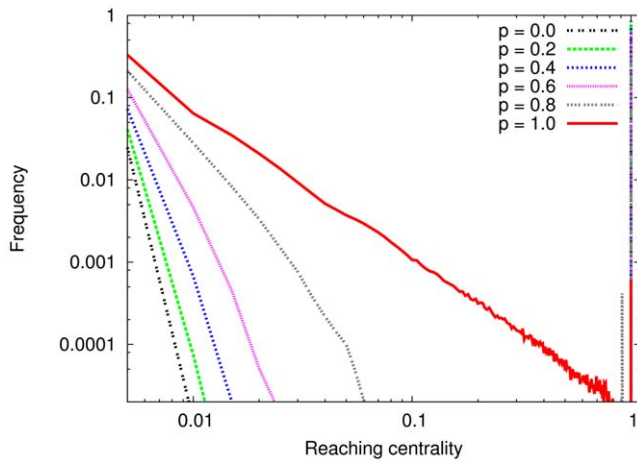
### Adjustable hierarchical network

We study the behavior of the GRC in a model with adjustable hierarchy as well (see Methods for a detailed description of the model). The parameter  $p$  tunes between the completely random and the totally hierarchical states. In the  $p=0$  limit, the topology of the AH graph is close to that of an ER graph, but, as one can see, the distribution of the local reaching centrality values of the AH is similar to that of the SF network (Figure 2): a little wider at small centralities than in the ER case. By increasing  $p$ , the distribution further widens around the origin and at  $p=1$ , it resembles the one for the directed tree, but it is even closer to a power-law. The global reaching centrality as function of the parameter  $p$  is shown in Figure 3. The GRC monotonously increases with  $p$  and sweeps through the (0,1) interval in the synthetic model, indicating that it is suitable for measuring the level of hierarchy. As seen in the figures, the global reaching centrality at a given value of  $p$  is less for larger average degrees. This observation is confirmed with the results on ER and SF networks (Figure 4). For large densities the GRC vanishes for both the ER and the SF networks.

**Table 1. Heterogeneity of the distribution of the local reaching centrality for different network types.**

Graph	GRC	$s(C_R)$
ER	$0.058 \pm 0.005$	$0.222 \pm 0.010$
SF	$0.127 \pm 0.008$	$0.300 \pm 0.009$
Tree	$0.997 \pm 0.001$	$0.031 \pm 0.004$

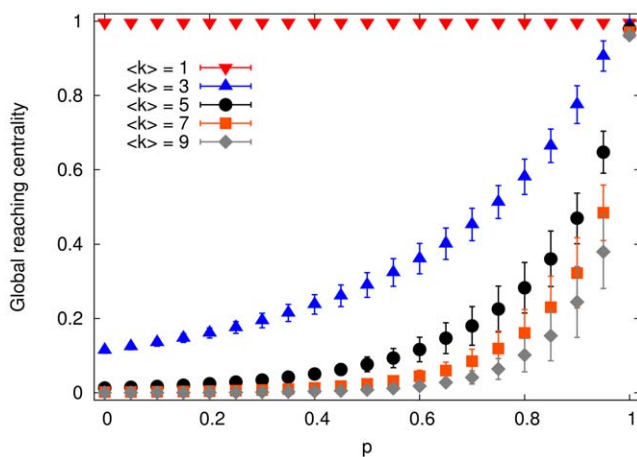
The two measures of heterogeneity presented here are the global reaching centrality ( $GRC$ ) and  $\sigma(C_R)$  (standard deviation of  $C_R$ ). Means and variances are shown for an ensemble of 1000 networks. doi:10.1371/journal.pone.0033799.t001



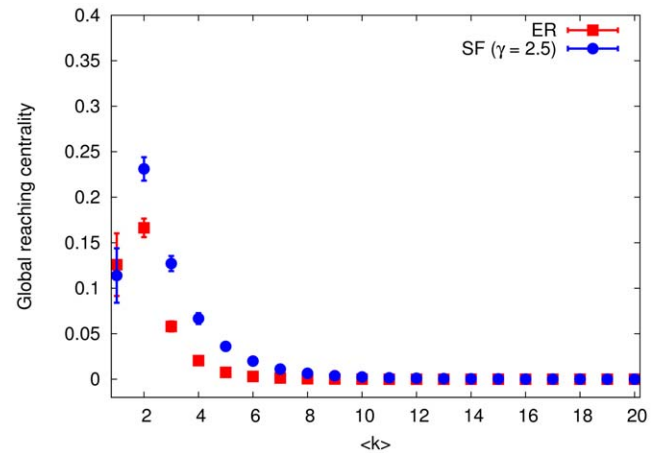
**Figure 2. Distribution of the local reaching centrality for the adjustable hierarchical network.** Distribution of the local reaching centrality in the adjustable hierarchical (AH) network model at different  $p$  parameter values. Each distribution is averaged over 1000 AH networks with  $N=2000$  and  $\langle k \rangle=3$ . The standard deviations of the distributions are comparable to the averages only for relative frequencies less than 0.002. Note that from the  $p=0$  (highly random) to the  $p=1$  (fully hierarchical) state the distribution changes continuously and monotonously with  $p$ .  
doi:10.1371/journal.pone.0033799.g002

### Real networks

We now turn our attention to the hierarchical properties of real networks. The global reaching centralities for different types of networks are shown in Table 2. For each network we show the average degree ( $\langle k \rangle$ ) and the GRC of the real network. It is important to point out that the direction of the edges in real networks had to be well-defined before calculating the reaching centrality. In every case, the networks were directed so that the source of an edge had a larger effect on the target than conversely. This choice of directedness originates in the observation that the higher a node is in the hierarchy, the more impact it has on the network. According to Table 2, the GRC can have values from a



**Figure 3. The global reaching centrality at different  $p$  values in the adjustable hierarchical model.** All curves show averages over an ensemble of 1000 networks with  $N=2000$  and different average degrees. Standard deviations grow with  $p$ , but they are clearly below the average values of the GRC. Note that for larger density, it is less likely to obtain the same level of hierarchy.  
doi:10.1371/journal.pone.0033799.g003



**Figure 4. The global reaching centrality versus average degree in the Erdős-Rényi and scale-free networks.** Dots show averages for 1000 graphs with  $N=2000$  nodes. In the Erdős-Rényi and scale-free networks, standard deviations of the GRC are comparable with its averages only for  $\langle k \rangle > 7$  and  $\langle k \rangle > 12$ , respectively.  
doi:10.1371/journal.pone.0033799.g004

broad range, depending on the average degree and the structure of the networks. For graphs with higher average degree, the GRC is usually smaller. This indicates that for a dense network it is harder to achieve a large reaching centrality, as seen with the ER, SF and AH graphs. The value of the GRC shows how hierarchical the structure of the network is. Food webs have the largest GRC and networks of intra-organizational trust have the smallest. This is in good agreement with the extremely low number of loops in food webs and the high number of loops in email-based organizational networks.

While the actual value of the GRC provides information about the hierarchical properties of the network, we can also compare the results to the randomized versions of the original networks to see how consistent the value we obtained is with the expectations. In order to do this, for each network we generated 100 random networks with the same degree (the details of randomization is explained in the Methods section): the mean values of the global reaching centralities for these randomized networks are shown in Table 2 ( $GRC^{rand}$ ). The color of the networks' names indicates the relation of each original network to its randomized version: the names of statistically significantly (with a confidence interval of 98%) hierarchical networks are in red while the names of non-hierarchical ones (same confidence) are in blue. Apart from the actual GRC values, the comparison to randomized networks by  $GRC/GRC^{rand}$  shows slight differences between the analyzed network types. For the food webs  $GRC/GRC^{rand}$  is remarkably high. Although the electronic circuits have low GRC values, they are significantly more hierarchical than their randomized versions. In contrast, although the Internet networks have larger reaching centralities than most other listed networks, these values do not differ significantly from the values of the corresponding randomized networks. Also note that the regulatory networks are significantly less hierarchical, mostly because biochemical systems contain many feedbacks keeping the processes stabilized.

The emergence of hierarchy in many human-made organizations and networks raises the question whether conscious control over these systems plays a role in the origin of hierarchy? In order to investigate this question, we compared the global reaching centralities with the controllability of networks as defined by Liu et al. [45]. They show that the minimal number of *driver nodes* ( $N_D$ ) is

**Table 2.** Hierarchical properties of real networks.

Type	Meaning of A→B	Network	N	$\langle k \rangle$	GRC	GRC <sup>rand</sup>
Food web	A eats B	Ythan [48]	135	4.452	0.814	0.507
		Seagrass [49]	49	4.612	0.723	0.253
		LittleRock [50]	183	13.628	0.811	0.045
		GrassLand [48]	88	1.557	0.961	0.695
Electric	B depends on the value at A	s1488 [51]	667	2.085	0.482	0.298
		s1494 [51]	661	2.116	0.482	0.289
		s5378 [51]	2993	1.467	0.231	0.062
		s9234 [51]	5844	1.4	0.424	0.050
		s35932 [51]	17828	1.683	0.459	0.015
Metabolic	B is an end product of A	<i>C. elegans</i> [52]	1173	2.442	0.048	0.052
		<i>E. coli</i> [52]	2275	2.533	0.043	0.058
		<i>S. cerevisiae</i> [52]	1511	2.537	0.037	0.042
Neuronal	A synapse goes from A to B	<i>C. elegans</i> [53,54]	297	7.943	0.133	0.023
		Macaque brain [55]	45	10.289	0.000	0.000
Internet	A communicates with B	p2p-1 [56,57]	10876	3.677	0.598	0.597
		p2p-2 [56,57]	8846	3.599	0.600	0.599
		p2p-3 [56,57]	8717	3.616	0.607	0.605
Organization	B trusts in A	Enron [58,59]	156	10.699	0.038	0.044
		Consulting [60]	46	19.109	0.043	0.032
		Manufacturing [60]	34	18.935	0.013	0.013
	B knows A	Freemans-1 [61]	34	18.971	0.028	0.041
		Freemans-2 [61]	77	24.412	0.000	0.000
Trust	B trusts in A	WikiVote [62]	7115	14.573	0.494	0.534
		College [63,64]	32	3	0.275	0.273
		Prison [64,64]	67	2.716	0.172	0.111
Language	B follows A	English [65]	7724	5.992	0.128	0.238
		French [65]	9424	2.578	0.657	0.875
		Spanish [65]	12642	3.57	0.951	0.939
		Japanese [65]	3177	2.613	0.054	0.206
Regulatory	A regulates B	TRN-Yeast-1 [66]	4441	2.899	0.934	0.968
		TRN-Yeast-2 [67]	688	1.568	0.116	0.670
		TRN-EC [67]	419	1.239	0.261	0.679

We show the order ( $N$ ), average degree ( $\langle k \rangle$ ), and global reaching centrality for the original ( $GRC$ ) and for the randomized networks ( $GRC^{rand}$ ). References to data sources are included. Suits next to the GRC values show comparison to the randomized networks: whether the original networks are more hierarchical than their randomization (club suit) or they are more egalitarian (diamond suit) with a 98% confidence level. The meaning of edges is also indicated. doi:10.1371/journal.pone.0033799.t002

related to the maximum matching of the network and they also provide an algorithm for determining  $N_D$ . In a network with  $N$  nodes the relative number of driver nodes is  $n_D^{Liu} = N_D/N$ . Driver nodes are the nodes that have to be controlled in order to take full control over the network. Full control means that one can drive the system from any initial state to any other desired final state. Since the networks listed in Table 2 have different original functions (food web, electric, etc.), and in many cases their controllability and hierarchical properties are not yet well understood, we compared these two quantities separately within each group of networks. The Pearson correlations of the GRC and  $n_D^{Liu}$  are shown in Table 3. In most of the listed real networks, the correlation is above 0.5, which is a relatively small value but still indicates a weak relation between the two quantities. Next, we compared the hierarchy measure, GRC, to the ratio of driver nodes in our synthetic model. Interestingly, for high link densities

( $\langle k \rangle \geq 5$ ) the ratio of driver nodes is very close to the value of the GRC and they differ significantly only for highly hierarchical graphs (i.e., for  $p > 0.85$ ). In an easily (hardly) controllable network, i.e., where  $n_D$  is low (high), few (many) nodes need to be controlled for a total control over the network. According to the results shown in Table 3 for real graphs and the results with the synthetic model (for a wide range of  $p$ ) the GRC and  $n_D^{Liu}$  are moderately positively correlated. In other words, *hierarchical networks are harder to control*. This result contradicts our initial intuitive concept that hierarchy emerges because it is the optimal structure with respect to controllability. This contradiction can be traced back to an assumption in the node-based definition of controllability given in [45] where each node is assumed to send the same signal to all of its neighbors. If, however, the network's dynamics is defined on the edges [46], then the definition of controllability differs from the definition by Liu et al. Therefore, as



**Table 3.** The Pearson correlation of the GRC and  $n_D$  defined by Liu et al.

Type of the networks	$r(\text{GRC}, n_D^{\text{Liu}})$
Regulatory	0.843
Trust	0.974
Food web	0.69
Metabolic	-0.225
Electric	0.503
Internet	0.632
Organizational	0.337
Language	0.933

With only one exception, all correlations are positive and many of them are above 0.6, i.e., the GRC and  $n_D^{\text{Liu}}$  are positively correlated.  
doi:10.1371/journal.pone.0033799.t003

an alternative, we compared hierarchy to controllability defined under the *switchboard dynamics* [46] (correlations are shown in Table 4). In the case of switchboard dynamics edges are controlled and nodes are simple devices converting the signals arriving on their in-edges to signals leaving on their out-edges. The driver nodes in this dynamics are those that one has to control for controlling the state of every edge. Based on the correlations between the GRC and the number of driver nodes, we conclude that *under the switchboard dynamics hierarchical networks are better controllable*.

To show how the generalized reaching centralities can be applied to undirected networks, we tested our method on the networks of terrorists investigated by Memon et al. Our results are similar to those of [21]: the top of the hierarchy related to the Bojinka case contains Isamudin and K. S. Mehmood (known as Mohammed). In the London Bombings network [21] found that the mastermind of the 7/7 bombings was H. R. Awwat; he was identified by our analysis (based on  $C''_R$ ) as a leader and M. S. Khan and I. M. Said as additional important participants. These results suggest that the above extensions of the local reaching centrality are effective quantities for the description of undirected graphs.

**Table 4.** Pearson correlation of the GRC and  $n_D$  in the switchboard dynamics.

Type of the networks	$r(\text{GRC}, n_D^{\text{SBD}})$
Regulatory	-0.922
Trust	-0.983
Food web	-0.406
Metabolic	-0.916
Electric	-0.969
Internet	0.57
Organizational	-0.674
Language	-0.812

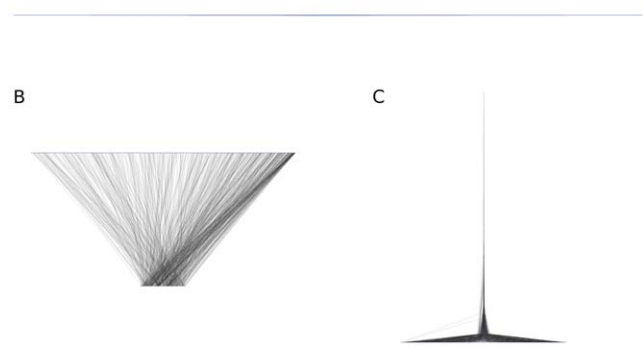
The correlations are all negative (except for the Internet networks) and most of them are very close to -1. Thus, under the switchboard dynamics the GRC (strength of hierarchy) and  $n_D^{\text{SBD}}$  are strongly negatively correlated.  
doi:10.1371/journal.pone.0033799.t004

## Visualization of large networks

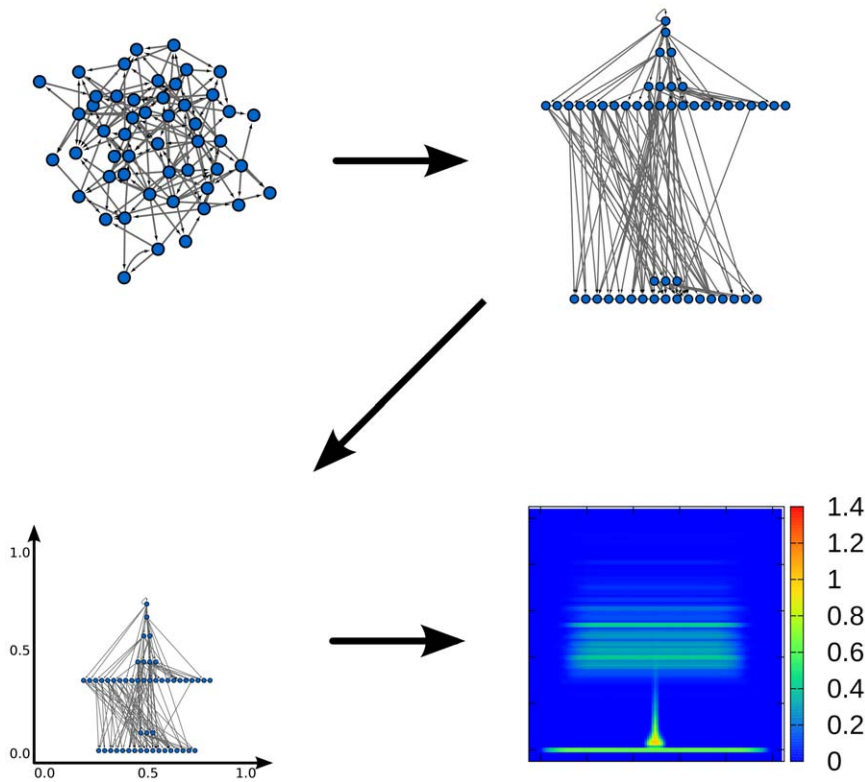
We use the method introduced in the Methods section for the hierarchical visualization of unweighted digraph by setting  $x_i = C_R(i)$ . Since the local reaching centrality takes discrete values on the graph, we use  $z = \varepsilon$ , that is, nodes that have local reaching centralities very close to each other are in the same level. Figure 5 shows the layout of various graphs. ER graphs have only two layers close to each other and most of their nodes are in the top layer indicating an almost equal impact of every node and the absence of hierarchy. As opposed to this, an arborescence has many layers, the distances between the layers vary and the layers contain different numbers of nodes. At the topmost layer there is only one node and it is far from the other nodes. This structure is due to the fact that the roles of nodes in the graph vary on a wide range, in other words, the distribution of the local reaching centrality is strongly heterogeneous. The hierarchical structure of an SF graph is between those of an ER graph and an arborescence: although it has only a few layers, these layers are clearly separated.

Note that different realizations (single graphs) of the same graph model (e.g., the SF model) usually have different hierarchical layouts. In order to eliminate this bias and to compare the graph models themselves (instead of single graphs from each model), we apply the hierarchical layouts of single graphs to define the drawing (image) of graph ensembles. To do this, first we rescale the hierarchical layout of each single graph to unit height and width and center it in the unit square (Figure 6). Next, we overlay the hierarchical layouts of graphs from the same model. For each graph model the result of this process is a density distribution of the nodes (in the unit square) averaged over the different realizations of the given model. Figure 7 shows graph ensemble drawings: the ER model is visualized as a thin horizontal line at the bottom of the box, while the SF model has more levels and it is similar to the AH(0.3) network. The ensemble of arborescences is visualized in a small concentrated region at the bottom of the unit square indicating the presence of many close levels. The transition from egalitarianism to hierarchy can be clearly seen on the visualization of the AH graphs. At small  $p$  (proportion of edges pointing to a lower level) there is mostly one level, then with increasing  $p$  more and more other levels emerge, and finally, the network splits into two groups of levels that are moving away from

A



**Figure 5. Visualization of three network types based on the local reaching centrality.** Visualization of (A) an Erdős-Rényi (ER) network, (B) a scale-free (SF) network and (C) a directed tree with random branching number between 1 and 5. All three graphs have  $N = 1000$  nodes and the ER and SF graphs have  $\langle k \rangle = 3$ . In each network  $z$  was set to  $2/N$ .  
doi:10.1371/journal.pone.0033799.g005



**Figure 6. Diagram illustrating the process of visualizing an ensemble of networks.** First, we compute the layout based on the selected  $x_i$  local quantity for each graph in the ensemble (top right). Next, we separate the levels logarithmically and scale each layout into the unit square (bottom left). Last, we overlay all rescaled layouts and plot the obtained density of nodes in the unit square (bottom right, see color scale also). In the heat maps, the color scale shows  $\log(\log(\rho(x,y)+1)+1)$ , where  $\rho(x,y)$  is the average density of the ensemble. doi:10.1371/journal.pone.0033799.g006

each other. To illustrate the usefulness of our visualization method, we show results for four real graphs as well (Figure 8). The GrassLand network is highly hierarchical, while the Enron network is very egalitarian (only very few nodes are much lower than the majority). This is in good agreement with the global reaching centrality values. The electrical circuit and the biological regulatory network are between the two extreme cases. The first contains two major levels (further subdivided into smaller levels). In contrast, the regulatory network has only one wide bottom level and a few nodes in the top and they are close to each other.

## Methods

### Synthetic model

In order to show the behavior of GRC, we introduce a synthetic network model with tunable extent of hierarchy. The construction of the network is the following:

1. In a directed tree assign a level ( $\ell$ ) to every node. The level of the root node is equal to the number of levels. If and only if a node has level  $\ell$ , then the level of its children will be  $\ell-1$ . These levels denote the natural layers in the hierarchy of the directed tree (the nodes at the bottom have  $\ell=1$ ).
2. We put a given number of additional random directed edges in the graph according to the following rule.  $1-p$  proportion of the edges is totally random, i.e. we choose two nodes randomly ( $A$  and  $B$ ) and if they are not already connected in the given ( $A \rightarrow B$ ) direction, we connect them. By  $p$  proportion of the edges, we put the  $A \rightarrow B$  edge only if  $\ell_A > \ell_B$ . In this way,  $p$

proportion of the random edges will not change the hierarchical structure of the directed tree.

An example of a generated network with the different edge types is shown in Figure 9. Hereafter, we will refer to this synthetic model as the *adjustable hierarchical network (AH)*.

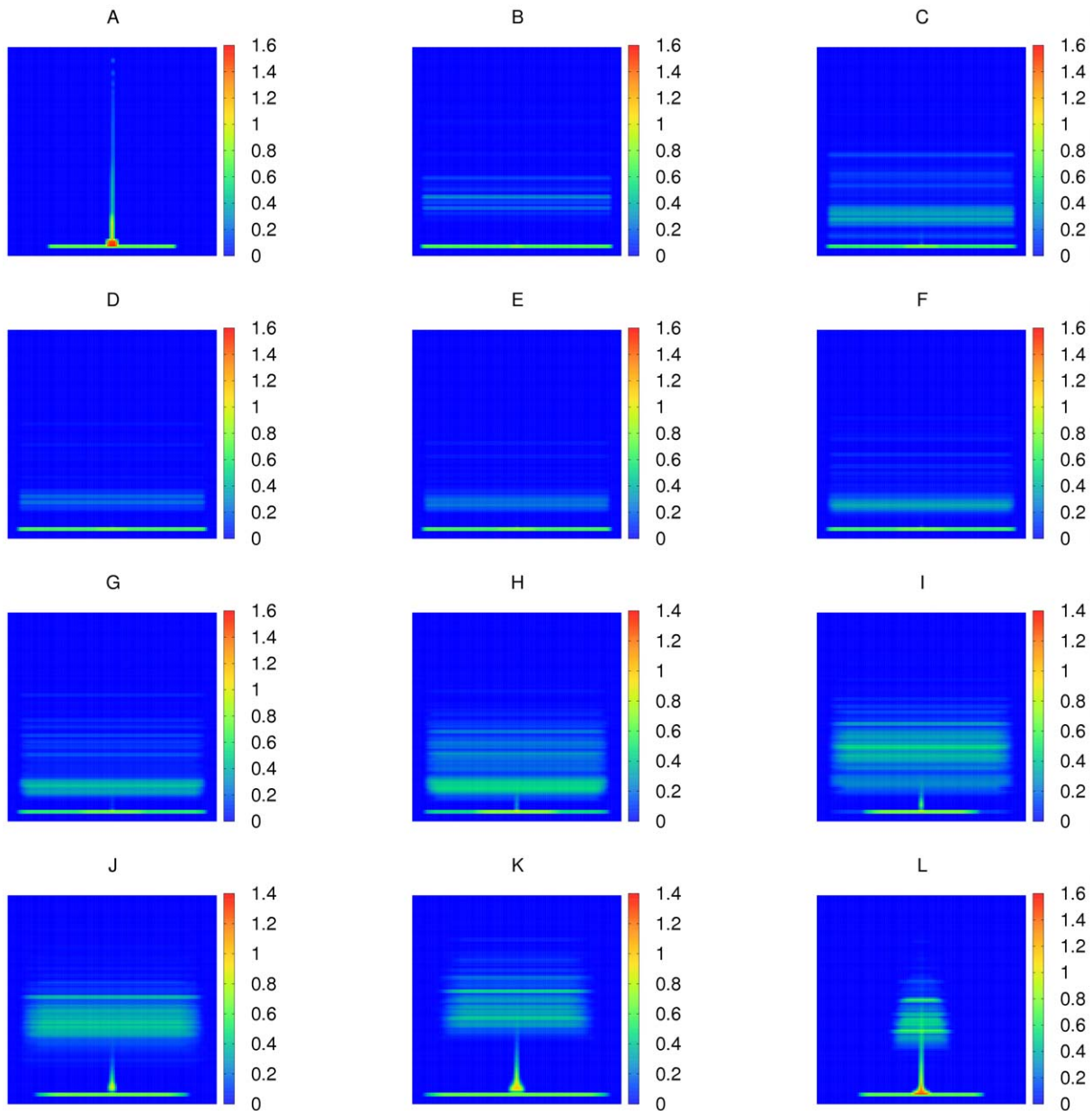
### Randomization of real networks

During the analysis of the results with real networks, we also calculated the GRC after randomizing them: first, we generated a random network with the same in and out degree distribution according to the configuration model. The generated network is further randomized in the following way: we choose two random edges ( $A \rightarrow B$  and  $C \rightarrow D$ ) and change the endpoints of them (so that we get  $A \rightarrow D$  and  $C \rightarrow B$ ). In every case, the number of rewired edge pairs was ten times the number of edges.

### Visualization

We also propose a visualization method using an arbitrary local quantity on the graph. The algorithm is as follows:

1. Grade the nodes according to the local quantity  $x_i$ .
2. Add nodes to the first (lowermost) level of the layout in the increasing order of their  $x_i$  values as long as  $\sigma_L < z \cdot \sigma_G$ . Here  $\sigma_L$  is the standard deviation of  $x_i$  within the current (first) level,  $\sigma_G$  is the standard deviations of  $x_i$  within the whole graph, and  $z$  is an adjustable coefficient.
3. When  $\sigma_L \geq z \cdot \sigma_G$  is reached, start a new level.
4. Repeat 2nd and 3rd steps until every node is put in levels.



**Figure 7. Visualization of network ensembles.** Visualizations of the (A) Erdős-Rényi, (B) scale-free, (C) directed tree and (D)–(L) AH network ensembles (subfigures (D)–(L) are for different values of the model parameter:  $p=0.1, \dots, 0.9$ ). In each case the color scale shows  $\log(\log(\rho(x,y)+1)+1)$  where  $\rho(x,y)$  is the density averaged over 1000 graphs.  $N=2000$  and  $\langle k \rangle=3$  were set. In every network,  $z$  was set to  $3/N$ . The corresponding GRC values are: 0.997 (A), 0.058 (B), 0.127 (C), 0.135 (D), 0.161 (E), 0.194 (F), 0.238 (G), 0.290 (H), 0.361 (I), 0.452 (J), 0.581 (K) and 0.775 (L).

doi:10.1371/journal.pone.0033799.g007

- For horizontal arrangement, align the center of every level to the same vertical line. In other words, in each level, the average of the horizontal positions of the nodes is the same:

$$X_{\ell_1} = X_{\ell_2} = 0 \quad \text{for all } \ell_1 \text{ and } \ell_2$$

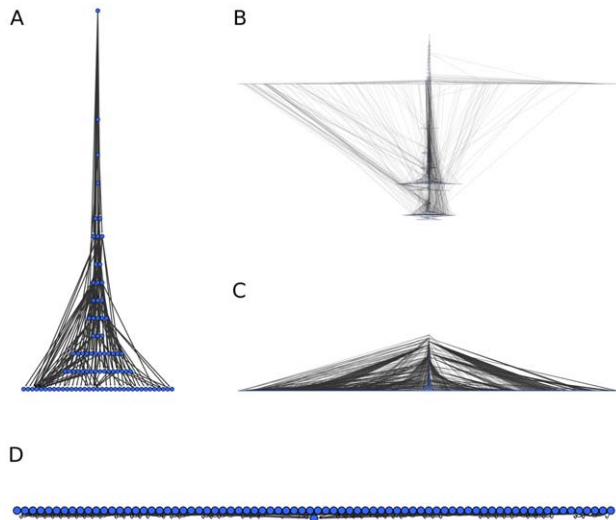
Here,  $X_\ell$  is the horizontal center of mass of level  $\ell$ .

- The levels are arranged vertically so that the distances between adjacent levels are proportional to the logarithm of the differences in the averages inside the corresponding levels, i.e.

$$(Y_{\ell+1} - Y_\ell) \propto \ln [\langle x \rangle_{\ell+1} - \langle x \rangle_\ell]$$

where  $Y_\ell$  is the vertical position of the  $\ell$ -th level and  $\langle x \rangle_\ell$  is the average of  $x_i$  inside this level. First, set the vertical distances of levels proportionally to the differences between their average values of  $x_i$  such that the smallest distance will be set to a given length (this length is the same as the horizontal distance between two adjacent nodes). Finally, set the distances to be proportional to the logarithm of the original differences so that the height of the graph is kept unchanged.





**Figure 8. Visualization of real networks.** The hierarchy-based visualization of (A) the GrassLand food web, (B) the electrical circuit benchmark s9234, (C) the transcriptional regulatory network of yeast and (D) the core of the Enron network. In every network  $z$  was set to  $2/N$ . doi:10.1371/journal.pone.0033799.g008

In the above steps we use the standard deviation in order to get clearly different layouts for different distributions of  $x_i$ . In a network with a localized distribution of  $x_i$  the method produces few levels that are very close to each other. But if the distribution of  $x_i$  is non-localized, the network will have many levels and a large vertical extension. If the distribution of  $x_i$  is continuous, then we can use  $z$  to adjust the extent to which every level contributes to the total variance. In other words, for large graphs,  $z$  tunes the vertical extension of the layout. If the distribution of  $x_i$  is discrete, then we can assign a level to each of its different values, which is mathematically equivalent to  $z=0$ . In practice, we set  $z$  to a sufficiently small value,  $\epsilon$ .

### Implementation

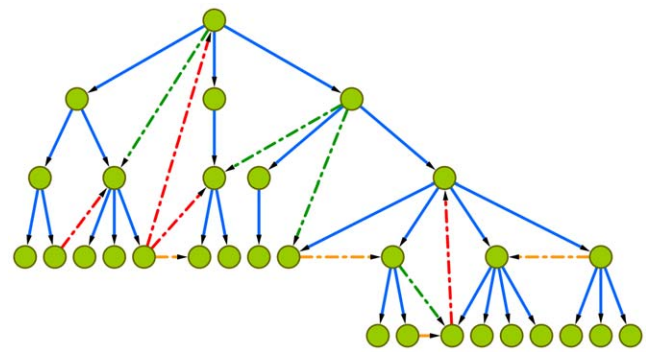
For the graph generations, randomizations and shortest path calculations presented in this paper, we used the already implemented functions in the *igraph* software package [47]. An open-source implementation of the local and global reaching centrality calculations is provided at <http://hal.elte.hu/~enys/grc.htm>.

### Discussion

Hierarchy is an essential feature of many natural and human-made networks and therefore, it is of high importance to have a

### References

- Castellano C, Fortunato S, Loreto V (2009) Statistical physics of social dynamics. *Phys Rev Lett* 81: 591–646.
- Vicsek T, Zafiris A (2010) Collective motion. arxiv:1010.5017.
- Pastor-Satorras R, Vespignani A (2004) Evolution and Structure of the Internet. Cambridge: Cambridge University Press.
- Albert R, Barabási AL (2002) Statistical Mechanics of Complex Networks. *Phys Rev Lett* 74: 47–97.
- Pumain D, ed (2006) Hierarchy in Natural and Social Sciences. Dordrecht, The Netherlands: Springer. pp 1–12.
- Huseyn L, Whetten DA (1984) The Concept of Horizontal Hierarchy and the Organization of Interorganizational Networks: a Comparative Analysis. *Social Networks* 6: 31–58.



**Figure 9. Distributions of the local reaching centrality for different network types.** For each network type  $N=2000$  and for the Erdős-Rényi (ER) and scale-free (SF) networks  $\langle k \rangle=3$ . All curves show averages of the distributions over an ensemble of 1000 graphs. Standard deviations are comparable with the averages only near the peaks in the ER and SF models. Although the standard deviations at the peaks are large, they do not change the positions of the peaks, and thus, do not affect the distributions. doi:10.1371/journal.pone.0033799.g009

measure quantifying it. Here we proposed a measure based on the assumption that the rank of the nodes should be related to their impact on the whole network, which is proportional to the number of all nodes reachable from them (local reaching centrality). The quantity we introduced, i.e., the global reaching centrality (GRC), measures the heterogeneity of the local reaching centrality distribution on the whole graph. In contrast to formerly proposed measures, the GRC does not penalize loops and undirected edges, but takes them into account by making bidirectionally connected pairs of nodes ( $A \rightarrow B$ ,  $B \rightarrow A$ ) equivalent in the hierarchy. There are neither free parameters in the method, nor optimization, and the ranks of the nodes are a natural result of the GRC. Since the controllability (according to the switchboard dynamics) and the extent of hierarchy are positively correlated, our calculations indicated that hierarchical structures are more easily controllable.

### Acknowledgments

We thank Illés Farkas and Gábor Vásárhelyi for their helpful comments on the early version of the manuscript. We also thank Tamás Nepusz for his technical and theoretical advices and suggestions on the simulations.

### Author Contributions

Conceived and designed the experiments: EM LV TV. Performed the experiments: EM. Analyzed the data: EM TV. Wrote the paper: EM LV TV.

12. Wimberley ET (2009) Nested ecology. The place of humans in the ecological hierarchy. Baltimore, MD: John Hopkins University Press.
13. Girvan M, Newman ME (2002) Community structure in social and biological networks. *PNAS* 99: 7821–7826.
14. Palla G, Derényi I, Farkas I, Vicsek T (2005) Uncovering the overlapping community structure of complex networks in nature and society. *Nature* 435: 814–818.
15. Smaje C (2000) *Natural Hierarchies. The Historical Sociology of Race and Caste*. Hoboken, NJ: Blackwell Publishers.
16. Dubreuil B (2010) *Human Evolution and the Origins of Hierarchies*. Cambridge: Cambridge University Press.
17. Theraulaz G, Bonabeau E, Deneubourg JL (1995) Self-organization of Hierarchies in Animal Societies: The Case of the Primitively Eusocial wasp *Polistes dominulus* Christ. *Journal of Theoretical Biology* 174: 313–323.
18. Eguluz VM, Zimmermann MG, Cela-Conde CJ, Miguel MS (2005) Cooperation and the Emergence of Role Differentiation in the Dynamics of Social Networks. *American Journal of Sociology* 110: 977–1008.
19. Bonabeau E, Theraulaz G, Deneubourg JL (1999) Dominance Orders in Animal Societies: The Self-organization Hypothesis Revisited. *Bulletin of Mathematical Biology* 61: 727–757.
20. Rowe R, Creamer G, Hershkop S, Stolfo SJ (2007) Automated social hierarchy detection through email network analysis. In: *WebKDD/SNA-KDD 07: Proceedings of the 9th WebKDD and 1st SNA-KDD 2007 workshop on Web mining and social network analysis ACM*. pp 109–117.
21. Memon N, Larsen HL, Hicks DL, Harkiolakis N (2008) Detecting Hidden Hierarchy in Terrorist Networks: Some Case Studies. In: *Proceedings of the IEEE ISI 2008 PAISI, PACCF, and SOCO international workshops on Intelligence and Security Informatics*. New York: Springer-Verlag. pp 477–489.
22. Bhardwaj N, Kim PM, Gerstein MB (2010) Rewiring of transcriptional regulatory networks: hierarchy, rather than connectivity, better reflects the importance of regulators. *Science Signaling* 3.
23. Song X, Chi Y, Hino K, Tseng BL (2007) Identifying Opinion Leaders in the Blogosphere. In: *Proceedings of the sixteenth ACM conference on Conference on information and knowledge management ACM, CIKM '07*. pp 971–974.
24. Mak V (2008) The Emergence of Opinion Leaders in Social Networks. Available: <http://ssrn.com/abstract=1157285>. Accessed 2012 Feb 24.
25. Hummon NP, Fararo TJ (1995) Actors and networks as objects. *Social Networks* 17: 1–26.
26. Johnsen EC (1985) Network macrostructure models for the Davis-Leinhardt set of empirical sociomatrices. *Social Networks* 7: 203–224.
27. Carmel L, Haren D, Koren Y (2002) Drawing Directed Graphs Using One-Dimensional Optimization. Heidelberg: Springer. pp 193–206.
28. Krackhardt D (1994) Graph theoretical dimensions of informal organizations. Mahwah, NJ: Lawrence Erlbaum Associates Inc.
29. Trusina A, Maslov S, Minnhagen P, Sneppen K (2004) Hierarchy measures in complex networks. *Phys Rev Lett* 92: 178702.
30. Sugiyama K, Tagawa S, Toda M (1981) Methods for visual understanding of hierarchical system structures. In: *IEEE Transactions in Systems, Man and Cybernetics*, volume 11. pp 109–125.
31. Garey MR, Johnson DS (1979) *Computers and Intractability: A Guide to the Theory of NP-Completeness*. New York: W. H. Freeman and Company.
32. Healy P, Nikolov NS (2004) *Hierarchical drawing algorithms*. St Helier, NJ: CRC Press.
33. Borgatti SP (2003) *The Key Player Problem*. Washington, D.C.: National Academy of Sciences Press. pp 241–252.
34. Huberman BA, Hogg T (1986) Complexity and adaptation. *Physica D* 22: 376–384.
35. Ceccatto HA, Huberman BA (1988) The complexity of hierarchical systems. *Physica Scripta* 37: 145.
36. Sabidussi G (1966) The centrality index of a graph. *Psychometrika* 31: 581–603.
37. Opsahl T, Agnessens F, Skvoretz J (2010) Node centrality in weighted networks: Generalizing degree and shortest paths. *Social Networks* 32: 245–251.
38. Erdős P, Rényi A (1960) On the evolution of random graphs. *Publ Math Inst Hung Acad Sci* 5: 17–60.
39. Bollobás B (2001) *Random Graphs*. Cambridge: Cambridge University Press.
40. Barabási AL, Albert R (1999) Emergence of scaling in random networks. *Science* 286: 509–512.
41. Goh KI, Kahng B, Kim D (2001) Universal behavior of load distribution in scale-free networks. *Phys Rev Lett* 87.
42. Chung F, Lu L (2002) Connected component in random graphs with given expected degree sequences. *Annual Combinatorics* 6: 125–145.
43. Tutte WT (2001) *Graph Theory*. Cambridge: Cambridge University Press.
44. Grinstead CM, Snell JL (1997) *Introduction to Probability*, Second Revised Edition. Washington, D.C.: American Mathematical Society.
45. Liu YY, Slotine JJ, Barabási AL (2011) Controllability of complex networks. *Nature* 473: 167–173.
46. Nepusz T, Vicsek T (2011) Controlling edge dynamics in complex networks. arxiv:1112.5945.
47. Csárdi G, Nepusz T (2006) The igraph software package for complex network research. *InterJournal Complex Systems* 1695. Available: <http://igraph.sf.net>. Accessed 2012 Mar 8.
48. Dunne JA, Williams RJ, Martinez ND (2002) Food-web structure and network theory: The role of connectance and size. *Proceedings of the National Academy of Sciences of the United States of America* 99: 12917–22.
49. Christian RR, Luczkovich JJ (1999) Organizing and understanding a winter's seagrass foodweb network through effective trophic levels. *Ecological Modelling* 117: 99–124.
50. Martinez N (1991) Artifacts or attributes? Effects of resolution on the Little Rock Lake food web. *Ecological Monographs* 61: 367–392.
51. Source: <http://courses.engr.illinois.edu/ece543/iscas89.html>. Accessed 2012 Feb 24. Networks available at <http://hal.elte.hu/~enys/data.htm>. Accessed 2012 Feb 24.
52. Jeong H, Tombor B, Albert R, Oltvai ZN, Barabási AL (2000) The large-scale organization of metabolic networks. *Nature* 407: 651–4.
53. Achacoso TB, Yamamoto WS (1992) *AY's Neuroanatomy of C. elegans for Computation*. First edition. Boca Raton, FL: CRC Press.
54. Watts DJ, Strogatz SH (1998) Collective dynamics of 'small-world' networks. *Nature* 393: 440–2.
55. Négyessy L, Nepusz T, Kocsis L, Bazsó F (2006) Prediction of the main cortical areas and connections involved in the tactile function of the visual cortex by network analysis. *European Journal of Neuroscience* 23: 1919–1930.
56. Leskovec J, Faloutsos C (2005) Graphs over time: densification laws, shrinking diameters and possible explanations. In: *Proceedings of the eleventh ACM SIGKDD international conference on Knowledge discovery in data mining ACM*. pp 177–187.
57. Ripceanu M, Foster I, Iamnitchi A (2002) Mapping the Gnutella network: Properties of large-scale peer-to-peer systems and implications for system design. *IEEE Internet Computing Journal* 6: 50–57.
58. Leskovec J, Lang K, Dasgupta A, Mahoney M (2009) Community structure in large networks: Natural cluster sizes and the absence of large well-defined clusters. *Internet Mathematics* 6: 29–123.
59. Klimt B, Yang Y (2004) Introducing the Enron corpus.
60. Cross R, Parker A (2004) *The Hidden Power of Social Networks*. Boston, MA: Harvard Business School Press.
61. Freeman S, Freeman L (1979) *Social science research reports* 46. Technical report, University of California, Irvine, CA.
62. Leskovec J, Huttenlocher D, Kleinberg J (2010) Signed networks in social media. In: *Proceedings of the 28th international conference on Human factors in computing systems ACM*. pp 1361–1370.
63. Van Duijn MAJ, Huisman M, Stokman FN, Wasseur FW, Zeggelink EPH (2003) Evolution of sociology freshmen into a friendship network. *Journal of Mathematical Sociology* 27: 153–191.
64. Milo R, Itzkovitz S, Kashtan N, Levitt R, Shen-Orr S, et al. (2004) Superfamilies of evolved and designed networks. *Science* 303: 1538–42.
65. Cancho RF, Solé RV (2001) The small world of human language. *Proceedings of the Royal Society of London Series B: Biological Sciences* 268: 2261–2265.
66. Balaji S, Babu MM, Iyer LM, Luscombe NM, Aravind L (2006) Comprehensive analysis of combinatorial regulation using the transcriptional regulatory network of yeast. *Journal of Molecular Biology* 360: 213–27.
67. Milo R, Shen-Orr S, Itzkovitz S, Kashtan N, Chklovskii D, et al. (2002) Network motifs: simple building blocks of complex networks. *Science* 298: 824–7.