Networking genetic regulation and neural computation: Directed network topology and its effect on the dynamics

Andreas Grönlund*
Department of Physics, Umeå University, 901 87 Umeå, Sweden
(Received 11 June 2004; published 20 December 2004)

Two different types of directed networks are investigated, transcriptional regulation networks and neural networks. The directed network structure is studied and is also shown to reflect the different processes taking place on the networks. The distribution of influence, identified as the the number of downstream vertices, are used as a tool for investigating random vertex removal. In the transcriptional regulation networks we observe that only a small number of vertices have a large influence. The small influences of most vertices limit the effect of a random removal to, in most cases, only a small fraction of vertices in the network. The neural network has a rather different topology with respect to the influence, which are large for most vertices. To further investigate the effect of vertex removal we simulate the biological processes taking place on the networks. Opposed to the presumed large effect of random vertex removal in the neural network, the high density of edges in conjunction with the dynamics used makes the change in the state of the system to be highly localized around the removed vertex.

DOI: 10.1103/PhysRevE.70.061908 PACS number(s): 87.10.+e, 89.75.−k, 89.70.+c, 05.65.+b

I. INTRODUCTION

In recent years complex networks have drawn a great deal of attention from the physics community. Various measures have been introduced in order to capture the function and form of specific networks. An observed feature that many networks show are a scale free or at least wide distribution of the vertex degree, which is given as a popular and well cited explanation in Ref. [1]. Other studies includes measures of clustering, assortative mixing [2], and betweenness centrality [3–5]. For a review of the recent work on networks, see Refs. [6–8]. Many of the networks appearing in the real world are directed and naturally the structure of two networks can be fundamentally different when the direction of the edges are considered, even if the overall structure might be alike when the direction of the edges is not considered. Two examples of real world networks that are naturally directed are neural networks and transcriptional regulation networks. The former is the network of neurons where neurons are connected in a directed fashion where the axons of each neuron connect to one of another neuron's dendrites, in this way building up a directed network in which signals are sent (axons) and received (dendrites) by the individual neurons. In the transcriptional regulation networks, the vertices represent proteins and the edges represent one protein's transcriptional regulation (positive and/or negative) of another protein. The cause of regulation is the attachment of a regulator protein to an operator position located on the DNA upstream of the gene coding for the regulated protein or, if more than one protein, operon. The attachment responds either in an up regulation or down regulation of the transcription rate by RNAp of the specific gene and thus the production of the protein. The reasons for regulation are many, of which one example is energy savings in a poor nutritional environment since synthesis of RNA and protein both are energy expensive processes. Another example is the regulation of enzymes. The best studied case of enzyme induction involves the enzymes of lactose degradation in *Escherichia coli*. Only in the presence of lactose the enzymes that are necessary to utilize lactose as a carbon and energy source are synthesized. But it is not just a matter of the presence of lactose. If both glucose and lactose is present *E. coli* chooses glucose. This is transcriptionally regulated via both positive and negative control.

The networks used in this paper are the neural network of the nematode *Caenorhabditis elegance* (NNCE) [9], the transcriptional regulaton network of the bacteria *E. coli* (TREC) [10], and the transcriptional regulaton network of yeast, *Saccharomyces cerevisiae* (TRSC) [11].

II. STRUCTURAL PROPERTIES

A directed edge in the literature formally termed arc also will be the term used in this paper. Because of the direction of the arcs one is able to follow directed paths in the network, representing the flow of information, the chain of command, or some other flow in the network. Depending on the system "living on" the network, the structure might look very different when the direction of the arcs in the different networks are taken into account. To get a first picture of what is going on in the networks we look at the distribution of the number of vertices with just outgoing arcs, only incoming arcs, and with both outgoing and incoming arcs. In a network in which information is flowing like the neural network a significant fraction of vertices should have both incoming and outgoing arcs in order to transport information between different parts of the network. Figure 1 shows the distribution of the three different types of vertices in the networks. The neural network consists of mostly *interneurons*, that is, neurons with both incoming and outgoing arcs, and have a low fraction of *sensory neurons* (only outgoing arcs) and

*Electronic address: gronlund@tp.umu.se

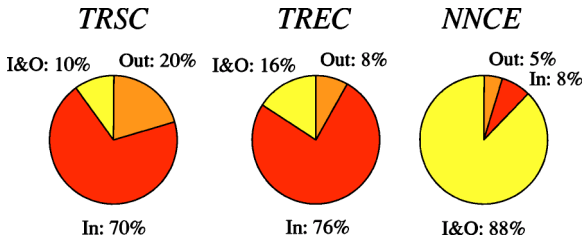

FIG. 1. (Color online) The distribution of vertices with only incoming arcs, only outgoing arcs, and with both incoming and outgoing arcs.

motor neurons (only incoming arcs). For more information on interneurons, sensory neurons and motor neurons, see Ref. [12]. The regulation networks have a different structure where the number of vertices of both incoming and outgoing arcs are suppressed and the network is dominated by vertices of only incoming links.

Since the different neurons play a different role and have different functions, a natural question to ask would be if this is reflected in the degree of the different vertices. In the neural network the sensory neurons receive their signals not from other neurons but from *receptors*. The motor neurons transmit signals not to other neurons but to one or more *effectors* igniting chemical reactions like the ones responsible for the contraction of muscles. The sensory neurons collect information from the outside world which is passed on via the interneurons to various parts of the network. This defines the state of the system which is visible via a response in the motor neurons. The "end station" of an input is not necessarily a specific motor neuron. The inputs are collectively setting the whole network in different states, and thus produces different responses to different inputs.

In Fig. 2 the degree distribution of the different networks are plotted and one can observe that the transcriptional regulation networks (TRSC and TREC) are somewhat similar in the sense that the degree of the vertices with only incoming arcs is lower than the rest of the vertices. This indicates that the proteins with no control and with a position in analogy of a laborer tend to be controlled by a few proteins and often

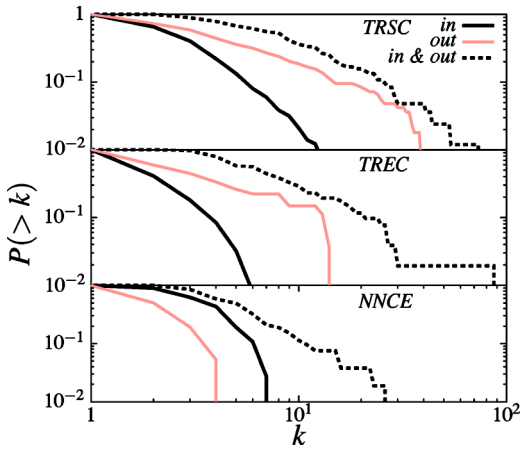

FIG. 2. (Color online) The cumulative degree distribution of vertices with only incoming arcs, only outgoing arcs, and with both incoming and outgoing arcs.

just one protein which often has a high degree [13] and with mostly just outgoing arcs, a global controller. In the neural network the situation is precisely the opposite; the sensors have in general very few outgoing arcs, in fact often just one. The sensory neurons are in most cases connected to an interneuron of a relative high degree of incoming links from which it collects information from a number of sensory neurons. In only a few cases the sensory neurons are connected directly to a motor neuron.

III. ROBUSTNESS VIA STRUCTURE

Many networks are believed to have a modular structure with functional modules in which communication is more present than between the modules. In addition to just having separate functions this also minimizes the influence of a random change of the network. The modularity has been studied and detected in undirected networks [14–16] and network models [17]. In nature there are many things found or believed to be modular [18,19] where the separate modules are responsible for different functions and together serve as a unit in a larger system. The modularity of the transcriptional regulation network of *S. cerevisiae* (TRSC) has been studied [20], and also the robustness in Ref. [21]. Some modules are more important than others, and by removing a unit more or less of the function of the total system is removed. Since the transcriptional regulation networks serves as regulating systems of the production of various proteins with different tasks they need to be constructed to retain most of the functions even if subjected to random removal or random changes of proteins. Random changes are naturally present via mutations in the DNA. Besides the fact that the DNA contains "garbage" which reduces the probability of removing important functions, one could ask if the transcriptional regulation networks have evolved to a structure which is robust to mutations and if it is possible to reveal and quantify the robustness with some measure of the structure.

In the literature there are a number of different measures of prestige and influence, see, e.g., Ref. [22]. Let D_i be the number of downstream vertices of a vertex v_i, and define the influence I_i of the vertex v_i to be the fraction of vertices in the network which is downstream of vertex v_i,

$$I_i = \frac{D_i}{N-1}. \tag{1}$$

The distribution of the influence $P(I)$ of the vertices in the network provides information on how the influence and control are distributed in the network. Moreover, the distribution $P(I)$ also provides information on how large a fraction of the network is maximally (and typically) affected by random change. For the function of the network to be stable to changes, the structure has to be designed in a way where most vertices only influence a low fraction of the vertices in the network. But even though the stability is important, the function of the network might anyway need some vertices of great influence or control, global controller proteins. The global controllers are needed for the response to nutritional elements C, O, N, P, heat shock, growth rate, and more.

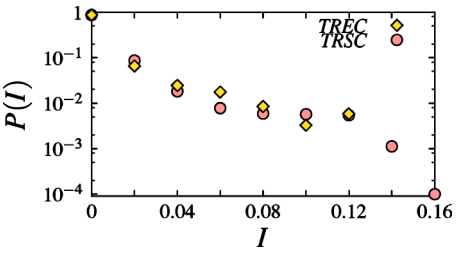

FIG. 3. (Color online) The distribution of influence $P(I)$ of TREC and TRSC. The data are binned.

In Fig. 3 the distribution $P(I)$ of TRSC and TREC is plotted. From the plot one can see that in TREC and TRSC most of the proteins control only a very little part of the network. The implication of this is that a random mutation or deletion of a vertex (protein) affects only a very small part of the network. In the case of NNCE, the situation is the reverse as can be seen in Table I.

A random removal or damage of a neuron can possibly affect the whole network. How and to which extent probably depends on the exact dynamics of the network and the situation. The network is still likely to be connected after a random vertex removal because of the high density of arcs, however, because of the high influence of the vertices a random removal of a neuron will possibly change the state of a large fraction of the other neurons, and thus the response to different inputs/stimuli. This is analyzed in the next section.

IV. ROBUSTNESS VIA DYNAMICS

To analyze whether the influence I of a vertex is of importance when considering vertex removal, two simple models are used, where one captures the nature of the interactions of the trancriptional networks and the other the neural networks. In Ref. [23], the transcriptional network of *S. cerevisiae* (TRSC) is analyzed in terms of Boolean network models with the aim of determining feasible rule structures. In their paper they find that many of the generated networks are shown to have a substantial part which is frozen in the sense that the final state is the same regardless of the initial states. As described before, the vertices in the trancriptional networks consist of proteins and the arcs represent one protein's regulation of another, in which the regulation can be either a positive regulation, an activator protein, or a negative regulation, a repressor. Also a study of the robustness of transcriptional regulation networks with the use of neural networks are done in Ref. [24].

TABLE I. The influence I, the fraction of neurons of influence $P(I)$, and the effect ΔS_f of removal of a neuron of influence I.

I	$P(I)$	ΔS_f
0.0	0.075	0.0
0.007	0.014	0.0030(0)
0.939	0.854	0.0052(6)
0.942	0.043	0.0035(8)
0.946	0.014	0.0051(7)

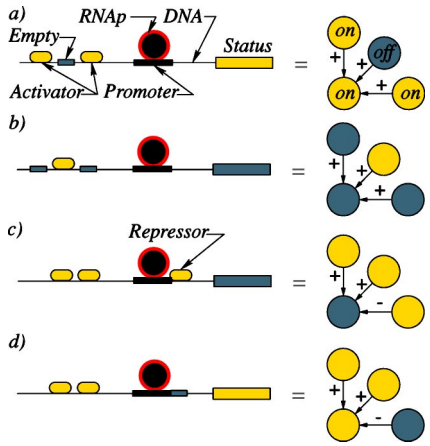

FIG. 4. (Color online) The update rules for the protein regulation networks. The positively regulating proteins, activator proteins, are treated with a majority rule illustrated in (a) and (b) and is overriden by a negative regulation, repressor, which is illustrated in (c) and (d).

In the model that we use to simulate the transcriptional regulation, the state of a gene coding for a specific protein v_i has two values, expressed or not expressed; *on* or *off*. If the state of the gene coding for a specific protein is *off* there is no production, or at most a very small production, of the protein and is therefore not considered to be present in the system. If the state is *on* there is a production enough for the protein considered to be present in the system. The state of a gene coding for a protein v_i is determined and regulated by the proteins v_j with arcs pointing towards v_i. The proteins with no incoming arcs are determined from the initiation and can be considered as different environmental settings. The rules for how the update is done can be summarized as follows.

(1) All vertices are randomly initiated with the value *on* or *off*.

(2) The vertices v_i are then updated sequentially with the following rule until a final state is achieved:

(a) The state of all vertices v_j pointing at v_i are determined.
(b) If the state of a protein v_j with negative regulation (repressor) is *on*, the state of protein v_i is *off*.
(c) If no negative regulation is present, the regulation follows a majority rule and the state of the protein v_i is *on/off* if the *majority* of the state of the positive regulating proteins v_j are *on/off*.

The update is illustrated in Fig. 4. The motivation for the model follows from the nature of the interactions. Negative regulation by a repressor blocks the production of a protein by binding to an operator downstream of the promoter of the gene that codes to the specific protein. When the repressor sits downstream of the promoter it stops the transcription of the downstream gene from RNAp. Positive regulation by an activator enhances the probability of RNAp to attach to the promoter of the gene and thereby the transcription of the gene that codes for the specific protein.

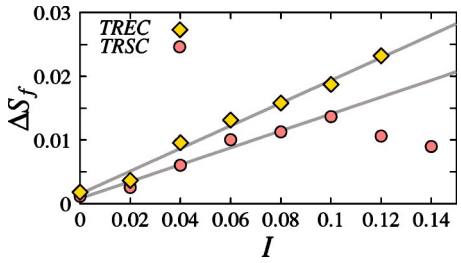

FIG. 5. (Color online) The difference of the final state ΔS_f versus vertex removal of a vertex with influence I. The lines are straight line approximations to guide the eye and the plotted data are binned. The errors are smaller than the symbol sizes.

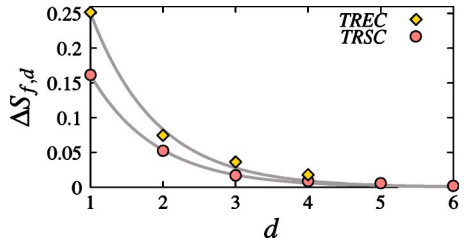

FIG. 6. (Color online) The fraction of vertices with a different state $\Delta S_{f,d}$ plotted against distance of the removed vertex. The data are fitted to exponential functions. The errors are smaller than the symbol sizes.

The state of the total system is represented by a vector of dimension N, $(S_1, S_2, \ldots, S_N)$. To investigate the effect of a random vertex deletion and to which extent the influence of the vertex plays a role for the state of the system, the final state obtained from an initial configuration is compared with the corresponding final state from the same initial configuration but with a vertex deleted from the network. The relationship of the influence of the removed vertex and the effect of the influence of the final state is demonstrated in Fig. 5. ΔS_f is the fraction of vertices (proteins) in the network having a different final state after the removal of a vertex v. Only the initial configurations that converge to a final state is considered. The fraction of initial states that converge to a final state is approximately 1 for TREC and 0.7 for TRSC. The initial configurations that do not converge to a final state ends up in an oscillatory state, and they are not considered or investigated here. As one can see, the overall behavior is that the difference in the final state ΔS_f has a somewhat linear behavior of the influence I of the vertex v being removed. The plot of TRSC does not follow the straight line approximation for larger values of I, which might indicate that the measure of influence used here is not perfectly suited for the applied dynamics.

Since the networks have a fixed structure, the values for the difference in the final states ΔS_f for the different influences I do not all follow the approximated straight line as can be observed, but the calculated error for the individual values of ΔS_f are, nevertheless, small. Since the dynamics incorporates a majority rule, the effect of a vertex removal decays with the distance from the removed vertex and therefore only a fraction of the downstream vertices get a different state after the removal. How large the fraction is depends therefore on the structure of the network and the typical distance to the downstream vertices from the removed one.

Figure 6 shows the fraction of all vertices with distance d from the removed vertex which have a different final state $\Delta S_{f,d}$ compared with the final state before the removal. Except from the exponential decay, one can also observe that the longest directed path is only of four steps in TREC and of six in TRSC. As a comparison, the diameter in TREC is 13 and in TRSC 14. Since the fraction of vertices with a different state drops exponentially with distance, a refinement of the measure of influence used here would be the measure of *proximity prestige* (see Ref. [22]), which is a vertex's number of downstream vertices normalized with their average distance to the vertex.

NNCE is applied to a similar dynamics as the transcriptional regulation networks, with the only difference that there are no negative regulations which overrule the positive ones. The neurons are thus treated as *McCullough-Pitts* neurons [25,26] with binary states *on/off*, and with equal and positive synaptic coupling strength (excitatory) and with a threshold of $n_i/0.5$, where n_i are the number of inputs. All neurons are considered to be excitatory, that is, in a state *off* when the input is below the threshold and *on* if the input is above the threshold. The update of the state of each neuron is therefore simply a majority rule, that is, the state S_i of a neuron v_i is *on/off* if the majority of the states of the incoming signals are *on/off*. If there is no majority, that is, the number of *on* inputs are equal to the number of *off* inputs, the state of the neuron is defined to be *on*. Since the influence of the neurons in the network is concentrated to a value around $I=0.94$ a linear dependence of the difference in the state ΔS_f to the influence I is thus not achievable. However, we can still get information of the effect of a vertex removal just by looking at the average difference in the final state from the deletion of a neuron. As before, the results are averaged by a number of different initial configurations and different vertex removals.

The results of the simulations done for NNCE are summarized in Table I. As one can see, the change in the state of the system is very small even if the influence of the removed vertex is large. The fact that the network is very dense and that the dynamics follow a majority rule implies that the change in the state of the system from the deletion of an individual neuron is small. The change in the state from the removal of a single neuron is simply averaged out in most cases, but there are of course changes in the state of some neurons located nearby the removed neuron. A further study would be to see how the state of the system responds when deleting a neighborhood of neurons to resemble a more realistic physical damage. One of the conclusions one can make is that even if the influence of most vertices is large, the dynamics put on the network results in a situation where the network is not very affected by a random removal of a single neuron. Figure 7 shows the decay of the fraction of changed states with the distance, and like the transcriptional regulation networks the decay fits well to an exponential decay.

V. SUMMARY

The two directed types of networks analyzed here are shown to have a different structure in various measures

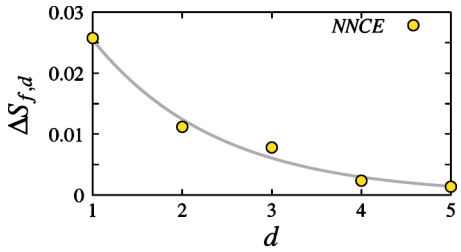

FIG. 7. (Color online) The fraction of vertices with a different state $\Delta S_{f,d}$ plotted against distance of the removed vertex. The data are fitted to an exponential function.

which incorporates the direction of the edges. The neural network of *C. elegance* consists of mostly vertices with both incoming and outgoing arcs, interneurons, possibly due to the fact that it is an information network in which information is processed and spread between different parts of the network. There is also a fraction of vertices with only outgoing arcs, sensory neurons, that feed the network with external information. Finally, there is a small fraction of vertices with only incoming arcs, motor neurons, responsible for igniting chemical reactions like contraction of muscles.

We find that the protein regulation networks of *E. coli* and *S. cerevisae* both show a small fraction of proteins with both incoming and outgoing arcs, and the dominating part of the network consists of proteins that only have incoming arcs, which we term laborers in the analogy of a human laborer which only has a small influence in the system he or she works in. The laborers are possibly used as building blocks or as components in different biochemical processes. There are also proteins that have a large number of outgoing arcs and thus globally control the production of many proteins, where most of them are laborers.

The influence of the vertices, defined for a vertex as the fraction of vertices in the network situated downstream of the vertex, are limited for most vertices in the transcriptional regulation networks. The neural network is showing a substantial part of the network to influence almost all vertices in the network. We simulate the biological processes on the networks and we investigate relationship of the influence of a removed vertex with the change in the state of the biological system. In the protein regulation networks the effect of a random removal are limited in most cases due to the fact that the influence of most proteins are restricted to a small number of proteins. However, the removal of a protein of great influence changes the state of the system more since the change in the state is shown to increase, with some exceptions, linearly with the influence. In the case of the neural network where the influence of most vertices are fairly large, the great number of arcs suppresses the effect of the removal of an individual neuron since the state of each neuron obeys a majority rule of the states of the incoming signals. Since the changes in the states of the vertices due to a vertex removal is shown to decay exponentially with the distance from the removed vertex, a proper refinement of the measure of influence could be to include the typical distance to the downstream vertices in the measure of influence, just like the measure of *proximity prestige*. A remark with the previous observations in mind would therefore be that the influence or prestige in a network (directed or not) probably to a large extent depends on the dynamics applied to the network, and therefore every investigation of prestige or influence should be in conjunction with the dynamics applied to the network.

ACKNOWLEDGMENTS

Many thanks to Petter Holme and Martin Rosvall for helpful comments and beneficial discussions contributing to this paper.

[1] A.-L. Barabási and R. Albert, Science **286**, 509 (1999).
[2] M. E. J. Newman, Phys. Rev. Lett. **89**, 208701 (2002).
[3] L. Freeman, Sociometry **40**, 35 (1977).
[4] M. Barthélemy, Eur. Phys. J. B **38**, 163 (2004).
[5] M. E. J. Newman, e-print cond-mat/0309045.
[6] M. E. J. Newman, SIAM Rev. **45**, 167 (2003).
[7] S. N. Dorogovtsev *et al.*, Adv. Phys. **51**, 1079 (2002).
[8] S. Bornholdt and H. G. Schuster, *Handbook of Graphs and Networks—From the Genome to the Internet* (Wiley-VCH, Berlin, 2002).
[9] J. G. White, E. Southgate, J. N. Thompson, and S. Brenner, Philos. Trans. R. Soc. London, Ser. A **314**, 1 (1986).
[10] S. Shen-Orr *et al.*, Nat. Genet. **31**, 64 (2002).
[11] T. I. Lee, *et al.*, Science **298**, 799 (2002).
[12] *Neuroscience*, 3rd ed., edited by D. Purves *et al.* (Sinauer Associates Inc., 2004).
[13] S. Maslov and K. Sneppen, Science **296**, 910 (2002).
[14] E. Ravasz *et al.*, Science **297**, 1553 (2002).
[15] M. Girvan and M. E. J. Newman, Proc. Natl. Acad. Sci. U.S.A. **99**, 7821 (2002).
[16] F. Radicchi *et al.*, Proc. Natl. Acad. Sci. U.S.A. **101**, 2658 (2004).
[17] A. Grönlund and P. Holme, Phys. Rev. E **70**, 036108 (2004).
[18] L. H. Hartwell *et al.*, Nature (London) **402**, C47 (1999).
[19] M. E. Csete and J. C. Doyle, Science **295**, 1664 (2002).
[20] J. Ihmels *et al.*, Nat. Genet. **31**, 370 (2002).
[21] S. Bornholdt, Biol. Chem. **382**, 1289 (2001).
[22] N. Lin, *Foundations of Social Research* (McGraw-Hill, New York, 1976).
[23] K. Stuart, P. Carsten, S. Björn, and T. Carl, Proc. Natl. Acad. Sci. U.S.A. **100**, 14 796 (2003).
[24] K. S, and W. JN, in *Systems, Man and Cybernetics, 2003 IEEE International Conference* (Washington, DC, 2003), Vol. 4, pp. 3969–3975.
[25] W. S. McCulloch and W. Pitts, Bull. Math. Biophys. **5**, 115 (1943).
[26] J. Hertz, A. Krogh, and R. G. Palmer, *Introduction to the Theory of Neural Computation* (Addison-Wesley, Redwood City, CA, 1991).