NETWORKS AND HETEROGENEOUS MEDIA
©American Institute of Mathematical Sciences
Volume **7**, Number **3**, September **2012**
doi:10.3934/nhm.2012.7.441
pp. **441–461**

ON CONGRUITY OF NODES AND ASSORTATIVE INFORMATION CONTENT IN COMPLEX NETWORKS

Mahendra Piraveenan

The Centre for Complex Systems Research, Project Management Graduate Programme
School of Civil Engineering, University of Sydney, NSW 2006, Australia

Mikhail Prokopenko

CSIRO Information and Communications Technologies Centre
Locked Bag 17, North Ryde, NSW 1670, Australia

Albert Y. Zomaya

The Centre for Distributed and High Performance Computing
School of Information Technologies, University of Sydney, NSW 2006, Australia

Abstract. Many distributed systems lend themselves to be modelled as networks, where nodes can have a range of attributes and properties based on which they may be classified. In this paper, we attempt the task of quantifying varying levels of similarity among nodes in a complex network over a period of time. We analyze how this similarity varies as nodes implement their functional logic and node states vary accordingly. We then use information theory to analyze how much Shannon information is conveyed by such a similarity measure, and how such information varies with time. We also propose node congruity as a measure to quantify the contribution of each node to the network's scalar assortativity. Finally, focussing on networks with binary states, we present algorithms (logic functions) which can be implemented in nodes to maximize or minimize scalar assortativity in a given network, and analyze the corresponding tendencies in information content.

1. **Introduction.** Networks are ubiquitous in today's world. Communication networks such as world wide web, telephone networks and mobile phone networks are changing the way we live and we interact with other people. Social networks built on top of these, such as Facebook and Twitter, are redefining ways of keeping in touch. Vast airline and rail networks have given us access to the remotest parts of the world and reduced travel times by orders of magnitude. Our survival depends on the functioning of a number of biological and ecological networks. The energy needed for our domestic and industrial use is supplied by electric power networks. Indeed, the interest and awareness about networks are not only a trend in scientific research but also a social and cultural phenomenon of this age [6].

Nodes of complex networks may have a number of properties [6, 1, 16]. Some of these properties may be boolean in nature, taking one of two states. For example, in a social network, each node may have a gender (male or female), or in a neural network, each neuron may be spiking or not spiking at a given time [10, 22, 11, 8]. Other properties may take integer values, such as the number of friends to a person

2000 *Mathematics Subject Classification.* Primary: 91D30, 05C82, 90B18; Secondary: 92C42.
Key words and phrases. Complex networks, graph theory, assortativity, information content, congruity.

in a social network, and yet other properties may be continuous real numbers, such as the age of people in a social network [5] or reading of sensors in a sensor network. In each of these cases, the node-states have a distribution. For example, let us say that in a binary network, the node-state distribution is such that most nodes have state '1' rather than '0'. One may then ask, does that mean that if we pick a random neighbour of a node, that neighbour is more likely to have state '1'? That is not always the case.

To illustrate this, consider a star network with the central node (the hub) having state '0', and all other nodes having state '1'. While most nodes have state '1', neighbours of most nodes have state '0' (since the hub is the sole neighbour for all other nodes). In other words, even though most nodes have state '1', if we pick a random link, finding a '1' at an end of this link is not more likely than finding a '0'. In fact these likelihoods are equal, and are influenced by the topology.

The knowledge about the likelihood of finding a given state at the end of a link is quite important to understanding a complex network and its dynamics [23]. For instance, in a sensor network, relatively high readings of temperature in a chain of direct neighbours may point to a potential fault line. In a social network, we may be interested to know whether people who are directly connected are in similar age groups. In a neural network, it may be important to understand if all the neurons which spike at the given time are directly connected [11]. A number of other examples could be provided from other domains of complex networks. In short, measuring the tendency in a network where directly connected nodes have similar properties is critically important in understanding the network's dynamics. In this paper we analyze this tendency, by generalizing and extending the concept of scalar assortativity, as described below.

Assortative mixing in complex networks has been one of the much explored areas in network science and graph theory in recent years [14, 15, 18, 17, 4]. Assortative mixing quantifies the tendency of nodes in a network to make links with similar nodes. Accordingly, assortativity is defined by a correlation coefficient, whereby perfect correlation means the network is perfectly assortative where all nodes are 'identical', while perfect anti-correlation means the network is disassortative where no instance of similar nodes being connected can be found. While similarity between nodes can be interpreted in many ways, assortativity has been primarily defined by similarity of degrees of nodes [14, 15, 18, 17]. Thus, the assortativity coefficient is related to network topology and is constant for that network while the topology remains unaltered. The concept of assortativity was extended by Newman [15] to measure similarity of scalar attributes of nodes (other than degree) - this was called scalar assortativity by [15].

While network topology is an important aspect of a complex network, studying the network dynamics gives us a much deeper understanding of its functionality. The state of the node, whether it is a boolean, discrete or continuous quantity, is an attribute of node, and similarity of nodes can be interpreted in terms of this attribute. Moreover, unlike node degree, the node state will change with time, therefore when similarity is defined in terms of node-states, the assortativity coefficient of a network varies with time as well. Therefore, it is possible to measure scalar assortativity over time and analyze its tendencies as a way of understanding the dynamics of the network.

In this paper, we analyze scalar assortativity coefficient as a function of time, based on node-states. We focus on binary states, while pointing out that the work

could be extended to multi-state models. Network scalar assortativity measures the tendency of nodes to have direct links with similar nodes in terms of their states, and can vary with time. We show that network scalar assortativity carries information about the network's dynamics that cannot be described by either the topology alone or by the state distribution alone, and we quantify this information using information theoretic measures. Furthermore, following [17], we also define a local contribution of an individual node to the global scalar assortativity, which we call node *congruity*. We analyze a number of model and real-world networks and their dynamics using these concepts.

2. **Definitions and terms.** In this section, we review the concepts related to degree-based assortativity. Let us consider a network with N nodes (vertices) and M links (edges). Let us say that the probability of a randomly chosen node having degree k is p_k, where $1 \leq k \leq N_p$. The distribution of such probabilities is called the *degree distribution* p_k of the network. Let us now consider a randomly chosen link in an undirected network. We may denote the probability of the node at a random end of this link having remaining-degree k as q_k. We call the distribution of such probabilities as the *excess degree distribution* q_k of the network. This distribution is biased in favour of nodes of high degree, since more links end at a high-degree node than at a low-degree one [14]. It is related to the original degree distribution as follows:

$$q_k = \frac{(k+1)p_{k+1}}{\sum_1^{N_p} kp_k} \quad , \quad 1 \leq k \leq N_p \tag{1}$$

Now let us consider an undirected link having a node with excess-degree j on one end and a node with excess-degree k on the other end. Following [4] and Newman [14], we can define the quantity $e_{j,k}$ to be the joint probability distribution of the degrees of the two nodes at either end of a randomly chosen link.

2.1. **Network assortativity.** The measure proposed in [15, 14] defines assortativity as a correlation function in terms of degrees at the network level [4, 14]. This correlation function yields zero for non-assortative mixing and positive or negative values for assortative or disassortative mixing respectively. In the case of undirected networks, If no preferential mixing occurs, then $e_{j,k} = q_j q_k$. Therefore the correlation can be defined as

$$r = \frac{1}{\sigma_q^2}\left[\sum_{jk} jk\left(e_{j,k} - q_j q_k\right)\right] \tag{2}$$

where $e_{j,k}$ is the joint probability distribution of the remaining degrees of the two nodes at either end of a randomly chosen link. σ_q is the standard deviation of the remaining degree distribution of the network, q_k. Similarly, the term $\sum_j jq_j$ can be understood as μ_q, the expected value or mean of the remaining degree distribution. Therefore network assortativity r can be defined also as:

$$r = \frac{1}{\sigma_q^2}\left[(\sum_{jk} jke_{j,k}) - \mu_q^2\right] \tag{3}$$

where μ_q and σ_q are both constants for the network.

Here r lies between -1 and 1, whereby $r = 1$ means perfect assortativity, $r = -1$ means perfect disassortativity, and $r = 0$ means no assortativity (random linking).

This definition was extended in [15] to any scalar attribute of a network. Accordingly, scalar assortativity in [15] is defined as

$$r = \frac{1}{1 - \sum_{jk} a_j b_k} \left[\sum_{jk} (e_{j,k} - a_j b_k) \right] \tag{4}$$

where a_j and b_k are the fraction of each type of end (source or target) of a link that is attached to node of type j and node of type k . In undirected networks, where there is no 'source' or 'target' node, $a_j = b_k$. As before, $e_{j,k}$ is the fraction of links which have type j of node at source and type k of nodes at target.

Let us say that a given node at a given time t is in state y^t. For simplicity, let us assume that y^t takes only integer values (though the concept of scalar assortativity is applicable to continuous node states with appropriate binning). Most of the examples we present in this paper, in fact, assume binary node states. Following the excess degree distribution, let us define distribution q_y^t as the probability distribution of finding a link with node state y^t at an end of a link at time t. Similarly, let us define distribution $e_{y,z}^t$ as the probability distribution of finding a link with node state y^t at one end of the link and node state z^t at the other end of the link. Let us also say the expectation of q_y^t at a given instant t is denoted as μ_q^t and the standard deviation of the same distribution at time t as σ_q^t, Then network scalar assortativity $\mathscr{L}^t$ is defined as [15]:

$$\mathscr{L}^t = \frac{1}{(\sigma_q^t)^2} \left[\sum_{yz} yz \left(e_{y,z}^t - q_y^t q_z^t \right) \right] \tag{5}$$

Equivalently, we can also write

$$\mathscr{L}^t = \frac{1}{(\sigma_q^t)^2} \left[\left(\sum_{yz} yz e_{y,z}^t \right) - (\mu_q^t)^2 \right] \tag{6}$$

where μ_q^t is the expected value of node-state at the end of a link at time t.

Let us note that if the scalar assortativity $\mathscr{L}^t = 1$, it means all links in the network have same node-states at either side of the link (In a non-fragmented network, this also means that all nodes must be on the same state). If $\mathscr{L}^t = -1$, it means that all links have nodes with dissimilar states on either side of them.

If $\mathscr{L}^t = 0$, it means that a link is equally likely to have similar or dissimilar node-states on either side of the link.

We should note that $\mathscr{L}^t = 0$ *does not* imply a random distribution of nodes states. Indeed, scalar assortativity is a measure of the influence of topology in the 'expected' node state at the end of a link. Therefore, if the expected value of the node-state distribution is equal to the expected value of q_z^t, then scalar assortativity should be zero. The following examples with some model networks will illustrate this point further.

3. **Synthetic network models.**

Before analyzing real-world networks, we consider the scalar assortativity of some simple canonical networks with some simple node-state distributions. For simplicity, let us consider binary node states, where node state can be either 1 or 0. Note that regardless of the number of possible

states and their discrete/continuous nature, network scalar assortativity will take continuous values. This is the case even when the node states are binary.

Perfect positive scalar assortativity $\mathscr{L}^t = 1.0$ is possible if and only if all nodes are in the same state, regardless of the topology (unless the network is fragmented). Therefore let us concentrate on the cases which show perfect negative scalar assortativity $\mathscr{L}^t = -1.0$.

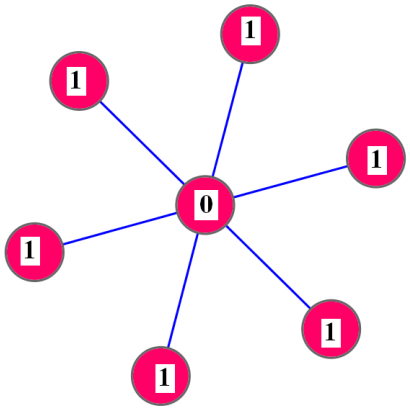

FIGURE 1. Star network with scalar assortativity $\mathscr{L}^t = -1$

First, let us consider a Star network with dissimilar node as a hub. Star topology is an important motif in many real world networks including communication networks, Local Area Networks (LAN), and regulatory networks [23]. We alluded to this network at the start of this paper in discussing the motivation for this work. The star network shown in Figure 1 with binary nodes states has scalar assortativity of $\mathscr{L}^t = -1$. It should be noted that even though most nodes have similar states, the scalar assortativity shows extreme negative correlation. This is the simplest case with perfect negative scalar assortativity. $\mathscr{L}^t = -1$ is not possible for all network topologies however. In scale free networks, it may not be possible to achieve $\mathscr{L}^t = -1$ for any combination of node states, simply due to the topology. However, if the scale free networks is a tree, then a set of node states can be found such that $\mathscr{L}^t = -1$.

In general, the exact values of $\mathscr{L}^t$ depend not only on the topology but also on the number of possible states.

Now let us consider a ring network with nodes having alternating states. The ring network, as shown in Figure 2, also shows $\mathscr{L}^t = -1$.

As an example for the simple scale-free network (in this case, also a network with a tree topology) showing perfect negative scalar assortativity, we present the network in Figure 3. As the figure shows, this network with the given node-states has perfect negative scalar assortativity.

A random or scale-free network, with randomly distributed binary node states, would display $\mathscr{L}^t \approx 0$.

We have utilized a number of topologies above to demonstrate the occurrences of extreme scalar assortativity values ($\mathscr{L}^t = 1.0$, $\mathscr{L}^t = -1.0$) and $\mathscr{L}^t = 0$. However, it is important to note that network scalar assortativity is not determined by

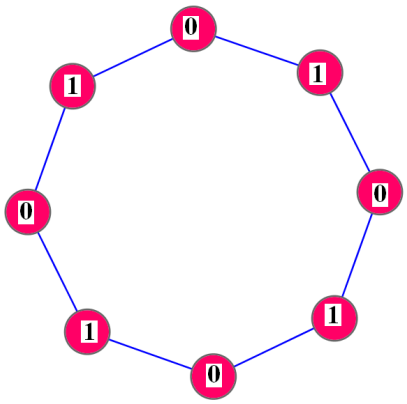

FIGURE 2. Ring network with scalar assortativity $\mathscr{L}^t = -1$

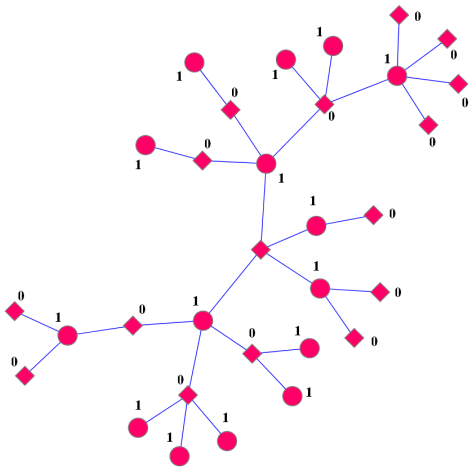

FIGURE 3. A scale-free network with scalar assortativity $\mathscr{L}^t = -1$

topology alone. Indeed, even for very simple topologies, the whole range of scalar assortativity values are possible. To demonstrate this, let us consider the simple 'benzene-ring' like topology in Figure 4. Assume that, in seven time steps, the nodes take the node-states shown in Table 1. As Table 1 also shows, as the '1' states propagate the scalar assortativity goes from 1.0 to −1.0, while the topology remains the same. We will show that large fluctuations in scalar assortativity are possible in other topologies also, including scale-free networks. Thus, network scalar assortativity provides more information about the node-states and dynamics of networks than the network's degree-based assortativity (a correlation measure of network topology) or statistical measures such as the standard deviation of node-states (correlation measures on node-state distribution) can by themselves provide.

4. **Scalar assortativity in Random Boolean Networks.** To understand scalar assortativity as a function of time, we simulated network dynamics on a number of

Node	T=1	T=2	T=3	T=4	T=5	T=6	T=7
1	0	1	1	1	1	1	1
2	0	0	0	0	0	0	0
3	0	0	1	1	1	1	1
4	0	0	0	0	0	0	0
5	0	0	0	1	1	1	1
6	0	0	0	0	0	0	0
7	0	0	0	0	1	1	1
8	0	0	0	0	0	0	0
9	0	0	0	0	0	1	1
10	0	0	0	0	0	0	0
11	0	0	0	0	0	0	1
12	0	0	0	0	0	0	0
scalar assortativity	1.0	-0.14	-0.33	-0.60	-0.71	-0.84	-1.0

TABLE 1. The states for seven time steps and corresponding scalar assortativity for the network shown in Figure 4

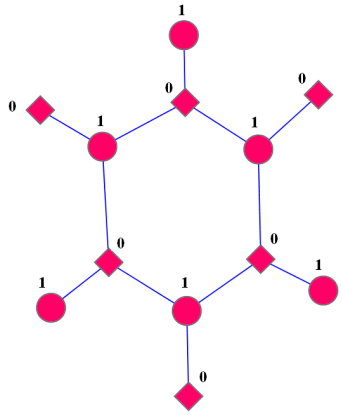

FIGURE 4. A benzene-ring like topology shows scalar assortativity ranging from $\mathscr{L}^t = 1$ to $\mathscr{L}^t = -1$ depending on node-states (The states in the figure corerspond to $T = 7$ in Table 1).

boolean networks and measured their scalar assortativity against time. We used the topologies of a number of real world networks (eg: *E. coli* transcription network), but in simulating the dynamics, we assumed that their node state would be either 'zero' or 'one' (i.e we considered them as boolean networks). Particularly, we utilised the topologies of Gene Regulatory Networks and transcription networks, since it has been shown that boolean networks are good models for these types of real world networks [3, 2]. The interpretation of the boolean states is *expressed* or *not-expressed* states of the genes. We implemented a number of logic functions in the nodes to simulate the dynamics, as listed below.

1. logic f_1: The nodes are simply assigned a boolean state ('0' or '1') with probabilities $1 - p$ and p. The previous state of the node considered or other

nodes do not influence the current state. If $p = 0.5$, the node will be randomly assigned '1' or '0' with equal likelihood.

2. logic f_2: The nodes follow the 'average state' of all their neighbours with probability p. Specifically:
   - If node state $y_v = 0$ and average neighbour state $\overline{z} > 0.5$, then with probability $p = A\overline{z}$, the node changes state to $y_v = 1$. A is a parameter of the logic.
   - If node state $y_v = 1$ and average neighbour state $\overline{z} < 0.5$, then with probability $p = A\overline{z}$, the node changes state to $y_v = 0$
3. logic f_3: The nodes 'oppose' the states of their neighbours with probability p. Since a node will have a number of neighbours, a node will choose one node from its neighbours and change its state to 'oppose' that neighbours state. The probability of a neighbour being chosen is proportional to the neighbour's degree. That is, nodes with more connections are more likely to be 'opposed' by their neighbours. Formally, for the concerned node v with degree d_v, choose a node w among the neighbours with probability p_w such that

$$p_w = \frac{k_w}{\sum_1^{d_v} k_w} \tag{7}$$

and change the node state of node v such that $y_v \neq y_w$.

We ran a number of simulation experiments, implementing the above logical functions in the nodes of networks. In a given simulation experiment, all nodes had identical logical behaviour. However, the logic function that is run on nodes could change with time (e.g., nodes implementing f_1 for T_1 time-steps and then f_2 for T_2 time-steps, periodically.). The node states were synchronously updated, with the updating order random and shuffled for each time step. The simulation results for the boolean network with the *E. coli* transcription network topology are given below.

Logic f_1: When node states are randomly assigned (with $P(1) = p$), the scalar assortativity remains close to zero for any number of time steps. We tried changing the value p periodically, so that the proportion of '1' states changes with time. The result of such a simulation experiment is shown in Figure 5, where parameter p is periodically changed from $p = 0.2$ to $p = 0.8$. We see that despite the change in the proportion of '1' states, the scalar assortativity remains close to zero. A similar example is shown in Figure 6, where parameter p is periodically and *linearly* (rather than like a step-function) changed from $p = 0.2$ to $p = 1.0$. Again, we see that the scalar assortativity remains close to zero throughout the simulation time.

These results are easy to explain. Scalar assortativity does not depend only on the distribution of states, but it depends also on the placement of states topologically. If the topological assignment is random, then despite the variations in the state distribution, the scalar assortativity will be close to zero. It can be shown that this result is valid for networks with any number of states, not just binary state networks.

Logic f_2: As seen above, logic f_2 is implemented in such a way that nodes tend to (stochastically) follow the states of their neighbours. Intuitively, this should mean that the scalar assortativity must increase, since links are increasingly likely to have nodes with similar states at each end over time. The result of a simulation for 200 time-steps where nodes implement this logic is shown in Figure 7. The

nodes are initialised randomly. Indeed, we could see from this figure that the scalar assortativity starts from close to zero and increases exponentially and stabilises at $\mathscr{L} = 0.63$. We observe that comparatively the proportion of '1's do not change much. The result of another run of the same simulation is shown in Figure 8 where the proportion of '1' states actually decreases, while scalar assortativity still increases exponentially.

From these results, it is clear that scalar assortativity can vary by orders of magnitude while the distribution of states remain nearly unchanged. In this case, the implemented dynamics, which encourages neighbouring nodes to have similar states, is responsible for the eventual high (positive) scalar assortativity.

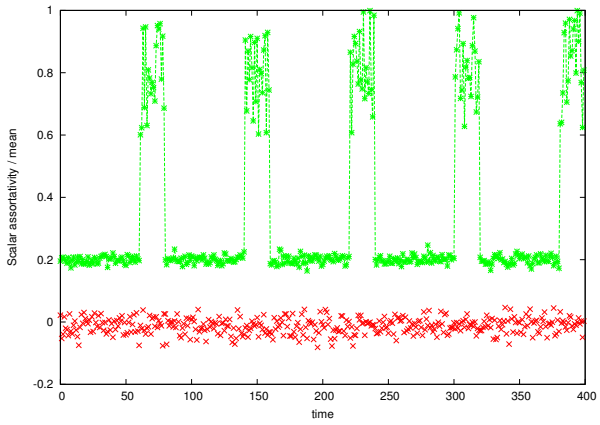

FIGURE 5. Variation of scalar assortativity vs time for a boolean network with *E.coli* transcription topology. The nodes implement a simple logic to randomly assign node states. The state distribution is varied, periodically having a high proportion of '1' states. While the state distribution peaks periodically, scalar assortativity does not change much and stays close to zero. Stars: mean of state distribution. Crosses: network scalar assortativity.

Logic f_3: Logic f_3 is implemented in such a way that nodes stochastically 'oppose' the state of their neighbours, with the neighbours with the highest degree having more likelihood to be 'opposed'. Intuitively, this should mean that the scalar assortativity must decrease from zero, since links are increasingly likely to have nodes with opposite states at each end. The results of two separate simulation runs for 200 time-steps where nodes implement this logic are shown in Figure 9 and 10. Again, the nodes are initialised randomly. We could see from these figures that scalar assortativity indeed decreases from zero and stabilises on considerably negative values (around $\mathscr{L} = -0.5$). In Figure 9, the proportion of '1's slightly increases with time, whereby in Figure 10, the proportion of '1's slightly decreases with time; however, in both cases, the scalar assortativity decreases by an order of magnitude. These results further confirm that scalar assortativity can vary by orders of magnitude while the distribution of states remain nearly unchanged. In this case, the implemented dynamics, which discourages neighbouring nodes to have similar states, is responsible for the eventual negative scalar assortativity.

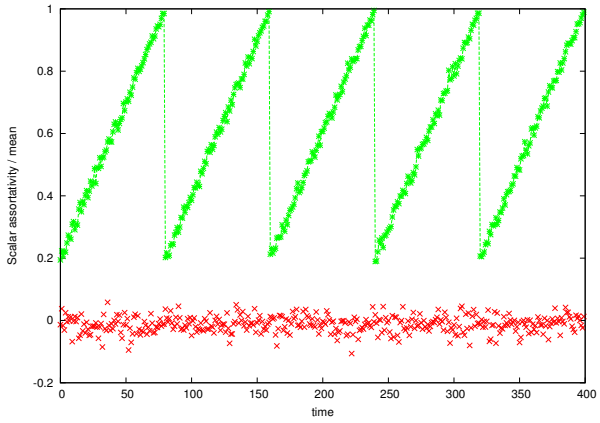

FIGURE 6. Variation of scalar assortativity vs time for a boolean network with *E.coli* transcription topology. The nodes implement a simple logic to randomly assign node states. The state distribution is varied, periodically having a high proportion of '1' states. While the state distribution peaks periodically, scalar assortativity does not change much and stays close to zero. Stars: mean of state distribution. Crosses: network scalar assortativity.

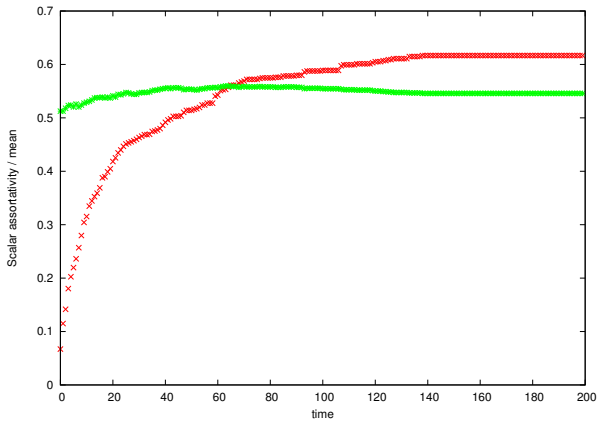

FIGURE 7. Variation of scalar assortativity vs time for a boolean network with *E.coli* transcription topology. The nodes implement a simple logic to probabilistically follow the average state of their neighbours. The initial node states are random. Note that while the state distribution does not change much, scalar assortativity increases exponentially from near zero (scalar non-assortativity) to positive scalar assortativity. Stars: mean of state distribution. Crosses: network scalar assortativity.

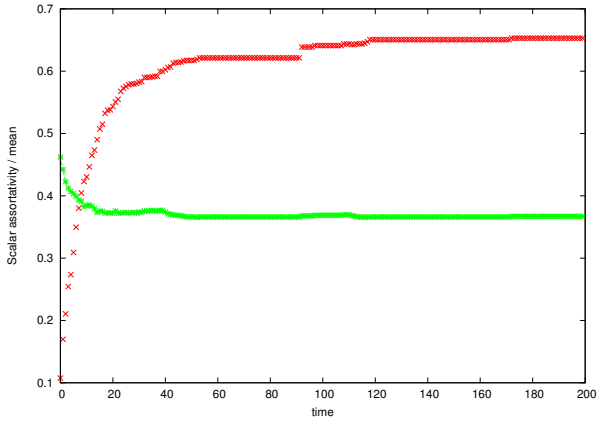

FIGURE 8. Variation of scalar assortativity vs time for a boolean network with *E.coli* transcription topology. The nodes implement a simple logic to probabilistically follow the average state of their neighbours. The initial node states are random. While the state distribution does not change much, scalar assortativity increases exponentially from near zero (scalar non-assortativity) to positive scalar assortativity. Stars: mean of state distribution. Crosses: network scalar assortativity.

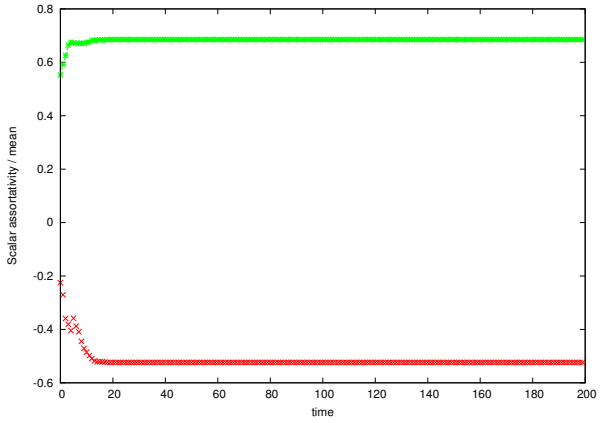

FIGURE 9. Variation of scalar assortativity vs time for a boolean network with *E.coli* transcription topology. The nodes implement a simple logic to probabilistically oppose the state of their neighbours. The initial node states are random.While the state distribution does not change much, scalar assortativity decreases exponentially from near zero (scalar non-assortativity) to negative scalar assortativity. Stars: mean of state distribution. Crosses: network scalar assortativity.

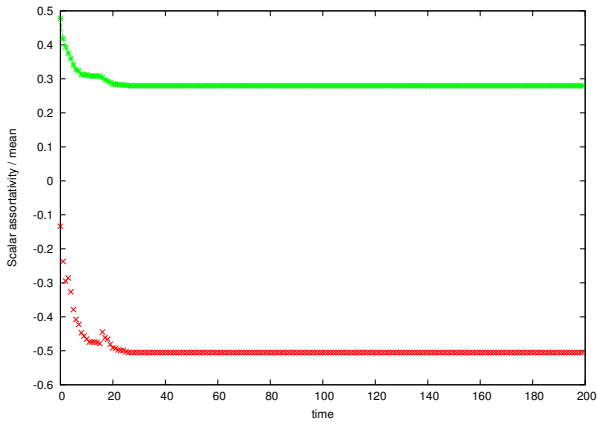

FIGURE 10. Variation of scalar assortativity vs time for a boolean network with *E.coli* transcription topology. The nodes implement a simple logic to probabilistically oppose the state of their neighbours. The initial node states are random. While the state distribution does not change much, scalar assortativity decreases exponentially from near zero (scalar non-assortativity) to negative scalar assortativity. Stars: mean of state distribution. Crosses: network scalar assortativity.

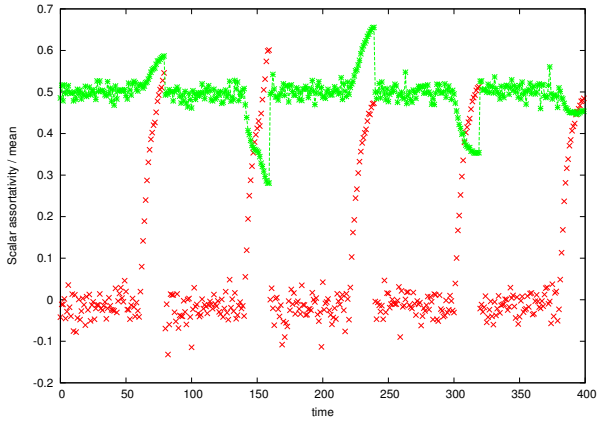

FIGURE 11. Variation of scalar assortativity vs time for a boolean network with *E.coli* transcription topology. The nodes alternatively implement two logics, one assigning node states randomly and the other logic where nodes probabilistically follow their neighbour's states. When the second logic is implemented, scalar assortativity increases rapidly, and when the first logic is implemented, scalar assortativity drops back to near zero. The proportion of nodes with state '1' comparatively does not change much. Stars: mean of state distribution. Crosses: network scalar assortativity.

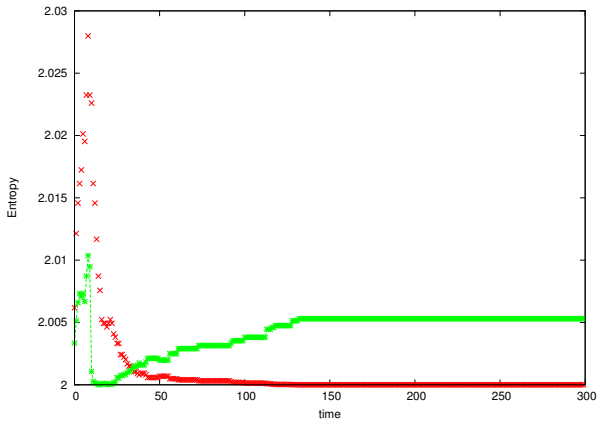

FIGURE 12. Variation of entropy vs time for a boolean network having the topology of *E. coli* transcription network, simulated with logic f_2. Note that as scalar assortativity increases with time (even though not shown in this figure, logic f_2 ensures that it will increase with time as we have seen before), the entropy $H(u^t)$ decreases but the entropy $H(q^t)$ increases. Crosses: $H(u^t)$. Stars: $H(q^t)$.

Combination of logical functions: To further verify the results above, we combined the logic functions mentioned above along the time axis. For example, logic f_1 was implemented on all nodes for $t = 60$ time steps followed by logic f_2 for $t = 20$ time steps on all nodes. This process is repeated to create a periodic combination of logic f_1 and logic f_2. The results of such an experiment are shown in Figure 11. We may see that when random logic f_1 is implemented, (with $p = 0.5$), the scalar assortativity remains close to zero. When logic f_2 is implemented though, scalar assortativity raises by an order of magnitude. The proportion of '1' states either increases or decreases depending on the node state distributions when the logic is flipped (from logic f_1 to logic f_2), but in all cases the change in the proportion of '1' states is small compared to the change in scalar assortativity. When the logic is flipped again (from logic f_2 to logic f_1), scalar assortativity drops back immediately close to zero. We combined logic f_3 with logic f_1 and logic f_2 and obtained similar results. These results confirm that scalar assortativity is highly influenced by topological placement of node states (node values), and as such provides information about the network dynamics that cannot be obtained by just analysing the node state distributions of the network. In the next section, we attempt to quantify the information provided by scalar assortativity.

5. **Scalar assortativity and information content.** In the previous sections we have seen that scalar assortativity can convey more information about the states of the network than just conveyed by the network's state-distribution. How can we quantify this? In other words, what is the relationship between scalar assortativity and the information contained in the network in terms of its node states? To answer this, we should define the information content of a network in terms of node states.

Shannon information $I(q)$ is a more generic measure of dependence than the correlation functions that measure linear relations [12, 9, 13]. In [23, 18, 21], the

entropy and information content were defined with respect to the degree distribution and joint degree distribution — purely in topological terms, irrespective of node states. Now we will attempt to define these in terms of node states in a network.

At first glance, one may wish to define the entropy of a network, using the probability distribution u^t that is defined via the probabilities u^t_z of encountering a node at the state z^t *anywhere* in the network, at time t:

$$H(u^t) = -\sum_z u^t_z \log u^t_z \tag{8}$$

In this paper however, we are interested in node states as well as the topology, and therefore, shall define entropy and information content in terms of node state distributions that depend on the link distribution. Such network entropy can be defined as

$$H(q^t) = -\sum_z q^t_z \log q^t_z \tag{9}$$

where q^t_z is, at time t, the probability (proportion) of links with a node (at one end) in the state z. Since q^t_z is dependent on link distribution, the entropy defined by Equation (9) also depends on the network topology, and is not just the entropy of node states, defined by Equation (8).

The defined entropy measures are contrasted in Figure 12. Note that the boolean network having the topology of *E. coli* transcription network is simulated here, with logic f_2 implemented in nodes. As seen before, the logic f_2 will ensure that scalar assortativity of the network will increase with time until it stabilises at a maximum value. We may note that the entropy $H(u^t)$ decreases but the entropy $H(q^t)$ increases with time. This is due to the fact that the former is not dependent on topology, and simply reflects the proportion of zeros and ones, while the latter depends on topology and reflects the scalar assortativity of the network.

Similarly, mutual information in terms of node states can be defined as:

$$I\left(q^t\right) = \sum_y \sum_z e^t_{y,z} \log \frac{e^t_{y,z}}{q^t_y q^t_z} \tag{10}$$

where $e^t_{y,z}$ is the proportion of links connecting, at time t, the nodes with states y, z respectively; q^t_y is the proportion of links, at time t, with a node (at one end) in the state y; and similarly, q^t_z is the proportion of links, at time t, with a node (at one end) in the state z.

Now we can analyse how this mutual information changes with scalar assortativity. To do so, we looked at the two logical functions (other than f_1) of random boolean networks, plotting network mutual information as well as scalar assortativity. The results are given in the Figures 13, 14 respectively. From the figures we may see that the information content matches the absolute values of scalar assortativity. That is, the more assortative or disassortative the network is, the more information it contains about expected states at the end of links. The beginning of each simulation where scalar assortativity is close to zero contains the least amount of information. We also note that there is no evidence for *just* positive scalar assortativity containing more information or vice-versa. Therefore, as suggested in [18] for assortativity and information content regarding degrees, we postulate that information content in a network regarding note states has a positive correlation with the *absolute value* of scalar assortativity. A detailed study of this correlation,

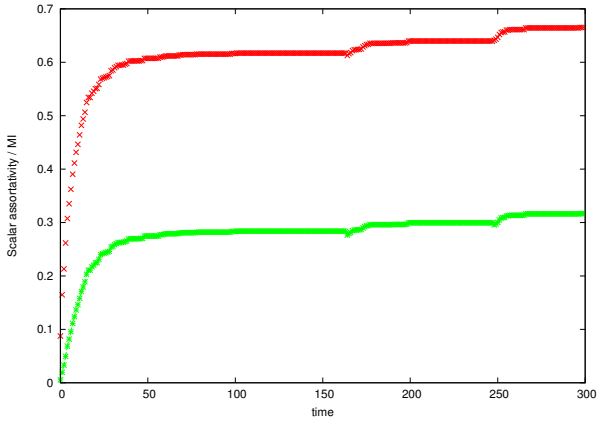

FIGURE 13. Variation of scalar assortativity and mutual information vs time for a boolean network having the topology of *E. coli* transcription network. Note that information content increases with the increase in positive scalar assortativity. Simulated with logic f_2. Stars: mutual information. Crosses: network scalar assortativity.

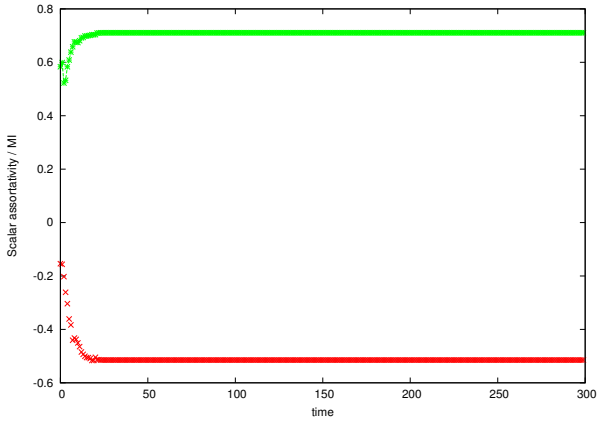

FIGURE 14. Variation of scalar assortativity and mutual information vs time for a boolean network having the topology of *E.coli* transcription network. Note that information content increases with the increase in negative scalar assortativity. Simulated with logic f_3. Stars: mutual information. Crosses: network scalar assortativity.

as done in [18] for information content and degree-based assortativity, is a subject of future research.

6. **Node congruity.** The concept of local assortativity [17] was introduced to quantify the contribution of an individual node to network assortativity. Since the scalar assortativity $\mathscr{L}^t$ measures similarity of nodes globally, the *local scalar*

assortativity, denoted λ^t, can be defined for each node as the node's contribution to the scalar assortativity $\mathscr{L}^t$, at time t. We choose for simplicity to call this local property λ^t the *node congruity*. We believe that *congruity* is a suitable term as it quantifies the extent to which a node is similar (congruent) to its neighbours. In this section we derive the expression for node congruity λ^t.

Following [17] and [20], we propose to derive node congruity as the contribution of a given node to the network scalar assortativity, which means we need to determine how much each node v contributes to the term

$$\frac{1}{\sigma_q^2}\left[\left(\sum_{yz} yze_{y,z}^t\right) - (\mu_q^t)^2\right]$$

Let us first look at the term $\sum_{yz} yze_{y,z}^t$ (which is calculated over node states) and the contribution of the node v in the state y_v to this term.

Suppose we visit all the nodes in a network, and from each node in turn we visit all the links of that node. In a network with N nodes and M links, the total visits we will thus make will be $2M$, since each link will be visited twice, once from each end. Suppose we build up the probability distribution $e_{y,z}^t$ as we make these visits. Each link will add a probability of $(1/2M)$ to the pair of (y, z) where y and z are the node states of nodes at each end of the link. Thus, each visit to a link will contribute $yz/2M$ to the sum $\sum_{yz} yze_{y,z}^t$. Therefore, if we examine the node v with state y_v and degree d_v which is connected to nodes with states $z_1, z_2, \ldots z_{d_v}$, it will contribute $(y_v z_1/2M) + (y_v z_2/2M) + \ldots + (y_v z_{d_v}/2M) = \frac{y_v}{2M}\sum_{i=1}^{d_v} z_i$ to the sum $\sum_{yz} yze_{y,z}^t$. Let us denote the average of node states of a node's neighbours as $\overline{z} = \frac{1}{d_v}\sum_{i=1}^{d_v} z_i$. Then we can represent the individual node's contribution, α_v, to the sum $\sum_{yz} yze_{y,z}^t$ as

$$\alpha_v = \frac{y_v}{2M}\sum_{i=1}^{d_v} z_i = \frac{y_v}{2M} d_v \overline{z} \tag{11}$$

Now let us consider a node's contribution to the term $(\mu_q^t)^2$. To do so, let us first examine the definition of μ_q^t:

$$\mu_q^t = \frac{1}{2M}\sum_{m=1}^{2M} y_m = \frac{1}{2M}\sum_{w=1}^{N} d_w y_w \tag{12}$$

where m is an end of a link; y_m is the state of the node at the end m; while d_w is a node-degree and y_w a node state. The equivalence of the two representations used in this definition, is yielded by the replacement of every set of links connected to some node by that node scaled by its degree. It follows that

$$(\mu_q^t)^2 = \left(\frac{1}{2M}\sum_{w=1}^{N} d_w y_w\right)^2 \tag{13}$$

$$(\mu_q^t)^2 = \frac{1}{4M^2}(d_1 y_1 + d_2 y_2 + ... + d_N y_N)^2 \tag{14}$$

Now, let us consider the node v (without loss of generality, let it be the node 1 with node state y_1), and its contribution to the expression above. The terms with

index 1 are:

$$\frac{1}{4M^2}((d_1y_1)^2 + 2d_1y_1(d_2y_2 + d_3y_3 + ... + d_Ny_N)) \tag{15}$$

Among these, terms such as $2d_1y_1d_jy_j$ have to be 'divided' between node 1 and node j respectively. These are multiplication terms, and we assume that an equal division is appropriate. Therefore, we can consider that contribution of node 1 is:

$$\frac{1}{4M^2}((d_1y_1)^2 + d_1y_1(d_2y_2 + d_3y_3 + ... + d_Ny_N)) \tag{16}$$

$$= \frac{1}{4M^2}(d_1y_1(d_1y_1 + d_2y_2 + d_3y_3 + ... + d_Ny_N)) \tag{17}$$

$$= \frac{1}{4M^2}\left(d_1y_1\sum_{v=1}^{N} d_vy_v\right) \tag{18}$$

Therefore, the contribution of a given node v to the term $(\mu_q^t)^2$ can be given by:

$$\beta_v = \frac{1}{4M^2}\left(d_vy_v\sum_{w=1}^{N} d_wy_w\right) \tag{19}$$

$$\beta_v = \frac{1}{2M}\left(d_vy_v\mu_q^t\right) \tag{20}$$

The standard deviation is already a scaling term, and we need not worry about a single node's contribution to it. Combining Equations (11) and (20) we formally define congruity of a node.

Congruity of a node λ_v^t is given by

$$\lambda_v^t = \frac{\alpha_v - \beta_v}{(\sigma_q^t)^2} = y_vd_v\frac{(\overline{z} - \mu_q^t)}{2M(\sigma_q^t)^2} \tag{21}$$

Congruity can be interpreted as a scaled difference between (i) the average *state* of the node's neighbours, and (ii) the average *state* across the whole network (i.e., the expected global or network-level state). If the node's local neighbours are in the states that are comparatively 'higher' than the globally expected value, then the node's congruity is positive. On the other hand, if the neighbours are in the states that are comparatively 'lower' than the globally expected value, then the congruity of the node is negative. Thus, congruity also quantifies the extent of how much the states of the node's immediate neighbours differ to the network as a whole. From the definition and derivation of congruity, it also follows that the sum of congruities λ_v^t over all nodes is equal to network scalar assortativity $\mathscr{L}^t$, at any time t. That is,

$$\mathscr{L}^t = \sum_{v=1}^{N}\lambda_v^t \tag{22}$$

7. **Distributions of node congruity.** Since congruity is a property of a node, it is possible to construct node congruity distributions for a given network, just like local assortativity distributions mentioned in [17]. We may plot node congruity values against degrees, *or* we may calculate the average node congruity value for all nodes with a given degree k. If we denote by $N(k)$ the number of nodes with degree k, the following equations hold true.

$$\mathscr{L}^t = \sum\nolimits_k N(k)\overline{\lambda^t}\,(k) \tag{23}$$

where $\overline{\lambda^t}(k)$ is the average congruity λ^t, at time t, of all nodes with degree k.

$$\mathscr{L}^t = N\sum\nolimits_k p_k\overline{\lambda^t}\,(k) \tag{24}$$

where p_k is the degree distribution of the network, being independent of time.

The Figures 15, 16, 17, 18, 19 show some examples of congruity distributions of networks. Figure 15 shows the node congruity distribution of *M. musculus* Gene Regulatory Network, simulated according to f_2 described above, until the scalar assortativity stabilises at its maximum (which was, in this case $\mathscr{L}$ =0.94). Thus, this network at the considered point in time has near perfect scalar assortativity. We note that the congruity distribution shows a strong correlation between node degree and node congruity. That is, it is the hubs which have the highest congruity. However, we may note that the relationship between node congruity and node degree is not linear. That is, there are some nodes which seems to have higher or lower congruity than predicted by a linear correlation with degrees. This shows that the overall placement of a node in the network (not merely the degree of the nodes), as well as the overall distribution of node states across the network, plays a part in node congruity.

Figure 16 shows the congruity distribution of *E. coli* transcription network, simulated according to f_3 described above, until the scalar assortativity stabilises at its minimum (which was, in this case $\mathscr{L}^t = -0.52$). Let us note that, as mentioned above, minimal scalar assortativity is harder to achieve in a network topology, since it requires neighbouring nodes to have different values, and the topology may make this harder to achieve. We note that the congruity distribution shows again strong correlation between node degree and node congruity, with the hubs having the highest negative congruity. Again, we may note that the relationship between node congruity and node degree is not linear. Figure 17 shows the node congruity distribution of *E. coli* transcription network, simulated according to random logic: logic f_1. Here the scalar assortativity remains close to scalar non-assortativity ($\mathscr{L}^t = 0.10$) and we may see that there is no recognisable correlation between node congruity and node degree. Other simulated networks confirmed the patterns in the results described above.

When a network has maximal scalar assortativity, ($\mathscr{L}^t = 0.94$) does it mean that all nodes in the network will have positive node congruity, or merely the majority of nodes will? This question cannot be answered by the plots above, since the average node congruity is plotted against degree. In the following two Figures 18, 19, we show node congruity of all individual nodes, where network scalar assortativity is either maximal or minimal. It can be noted that when scalar assortativity is maximal, all nodes have positive node congruity. However, when scalar assortativity is minimal, quite a few nodes still have positive congruity. This is again a property of congruity, since as we explained above, it is not always possible for a node to be

different from all of its neighbours. However, it could be easily similar to all of its neighbours, if almost all nodes have similar states anyway. It is important to note that this result is true only for binary states — specifically, if the number of states are comparable to, or higher than, the number of nodes, then it is much easier for the nodes to be dissimilar. Thus node congruity profiles gives us interesting insights about the interplay between node states, average neighbour-degree, and network size in a network with complex dynamics.

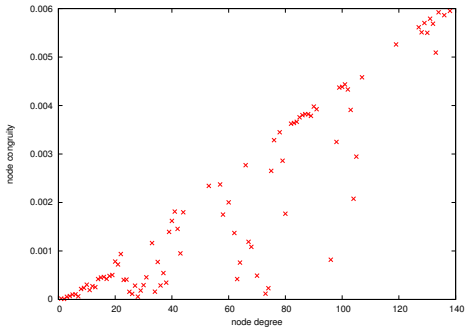

FIGURE 15. Node congruity profile. The *M. musculus* Gene regulatory network is simulated with logic f_2. $\mathscr{L} = 0.94$.

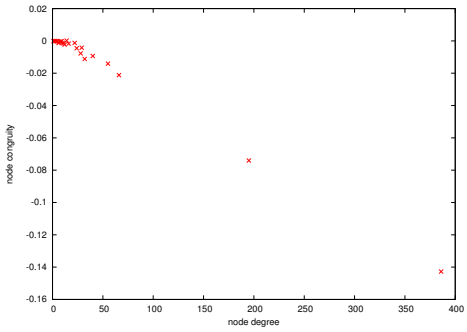

FIGURE 16. Node congruity profile. The *E. coli* transcription network is simulated with logic f_3. $\mathscr{L} = -0.52$.

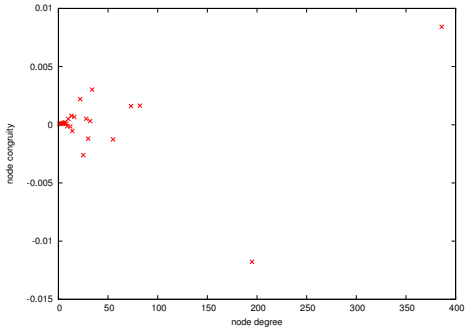

FIGURE 17. Node congruity profile: The *E. coli* transcription network is simulated with logic f_1. $\mathscr{L} = 0.10$.

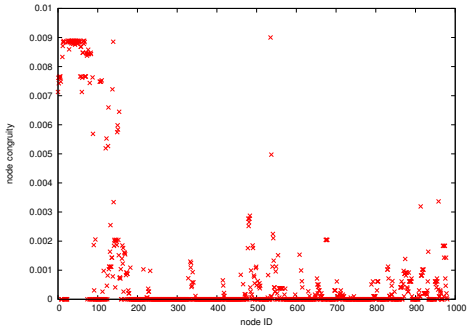

FIGURE 18. Node congruity profile with individual nodes. The *M. musculus* Gene regulatory network is simulated with logic f_2. $\mathscr{L} = 0.94$.

8. **Conclusions.** The dynamics of a network is influenced by its topology, and the topology of a network evolves as a result of its functional requirements and dynamics

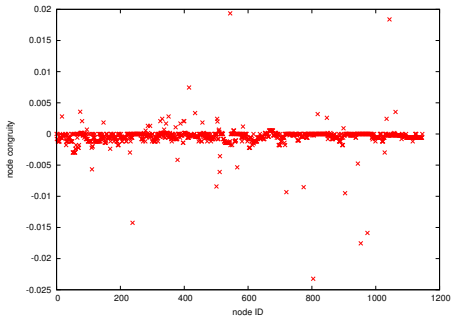

FIGURE 19. Node congruity profile with individual nodes. The *E. coli* transcription network is simulated with logic f_3. $\mathscr{L} = -0.52$.

[3, 7, 24, 25, 19]. Therefore, the patterns in a network dynamics and its topology are closely interdependent. In this paper, we considered scalar assortativity as a function of time. It measures the tendency of nodes in a network to make links with similar nodes, where similarity is interpreted in terms of node-states, rather than node-degree. High positive scalar assortativity of a network means that nodes which have direct links to each other tend to have similar node-states at the given time, whereas high negative scalar assortativity means that nodes which have direct links to each other tend to have dissimilar states.

Scalar assortativity can vary with time and show tendencies which give information about the dynamics of the networks. Using simulated boolean networks, we showed that networks which initially have nodes with random states, can in time achieve either high positive or high negative scalar assortativity, depending on the functionality of the nodes (the logic the nodes implement). We also showed that such networks can have high positive or high negative scalar assortativity, even if nodes are no more likelier to be in one state than the other (the state distribution is more or less uniform). We pointed out a number of scenarios where scalar assortativity could be used to measure dynamics of real-world networks, and quantified the relationship between network scalar assortativity and information content of networks.

Finally, we introduced node congruity as a node's contribution to scalar assortativity, and showed that a positive node congruity corresponds to comparatively high readings of node states in a node's immediate vicinity, and vice-versa. As such, we showed that congruity distributions provide an additional tool to understand a network's dynamics.

The specific contributions of this paper can be briefly listed as (i) extending the definition of assortativity as a function of time (ii) Definition of Shannon information content of a network based on both node states and topology (iii) The demonstration of logic functions which can be used to maximize or minimize scalar assortativity, while topology remains unchanged (iv) the definition of node congruity, which quantifies a single node's contribution to the scalar assortativity of a network.

We hope that the tools we introduced here will be extensively used in quantifying network properties and dynamic tendencies of complex networks.

REFERENCES

[1] R. Albert and A. L. Barabási, *Statistical mechanics of complex networks*, Reviews of Modern Physics, **74** (2002), 47–97.

[2] M. Aldana, *Boolean dynamics of networks with scale-free topology*, Physica D, **185** (2003), 45–66.

[3] U. Alon, "Introduction to Systems Biology: Design Principles of Biological Circuits," 1^{st} edition, Chapman and Hall, London, 2007.

[4] D. S. Callaway, J. E. Hopcroft, J. M. Kleinberg, M. E. J. Newman and S. H. Strogatz, *Are randomly grown graphs really random*, Physical Review E, **64** (2001).

[5] K. K. S. Chung, L. Hossain and J. Davis, *Exploring sociocentric and egocentric approaches for social network analysis*, in "KMAP 2005: Knowledge Management in Asia Pacific, Wellington, New Zealand," 2005.

[6] S. N. Dorogovtsev and J. F. F. Mendes, "Evolution of Networks: From Biological Nets to the Internet and WWW," 1^{st} edition, Oxford University Press, Oxford, 2003.

[7] R. Guimera, M. Sales-Pardo and L. A. Amaral, *Classes of complex networks defined by role-to-role connectivity profiles*, Nature Physics, **3** (2007), 63–69.

[8] B. H. Junker and F. Schreiber, "Analysis of Biological Networks (Wiley Series in Bioinformatics)," 1^{st} edition, Wiley-Interscience, 2008.

[9] A. Kaiser and T. Schreiber, *Information transfer in continuous processes*, Physica D, **166** (2002), 43–62.

[10] F. Kepes, "Biological Networks," 1^{st} edition, World Scientific, Singapore, 2007.

[11] S. Knock, A. McIntosh, O. Sporns, R. Ktter, P. Hagmann and V. Jirsa, *The effects of physiologically plausible connectivity structure on local and global dynamics in large scale brain models*, Journal of Neuroscience Methods, **183** (2009), 86–94.

[12] A. Kraskov, H. Stögbauer and P. Grassberger, *Estimating mutual information*, Physical review E, **69** (2004), 066138.

[13] D. J. MacKay, "Information Theory, Inference, and Learning Algorithms," 1^{st} edition, Cambridge University Press, Cambridge, 2003.

[14] M. E. J. Newman, *Assortative mixing in networks*, Physical Review Letters, **89** (2002), 208701.

[15] M. E. J. Newman, *Mixing patterns in networks*, Physical Review E, **67** (2003), 026126.

[16] B. O. Palsson, "Systems Biology: Properties of Reconstructed Networks," 1^{st} edition, Cambridge University Press, Cambridge, 2006.

[17] M. Piraveenan, M. Prokopenko and A. Y. Zomaya, *Local assortativeness in scale-free networks*, Europhysics Letters, **84** (2008), 28002.

[18] M. Piraveenan, M. Prokopenko and A. Y. Zomaya, *Assortativeness and information in scale-free networks*, European Physical Journal B, **67** (2009), 291–300.

[19] M. Piraveenan, M. Prokopenko and A. Y. Zomaya, *Assortativity and growth of Internet*, European Physical Journal B, **70** (2009), 275–285.

[20] M. Piraveenan, M. Prokopenko and A. Y. Zomaya, *Local assortativeness in scale-free networks—addendum*, Europhysics Letters, **89** (2010), 49901.

[21] M. Piraveenan, M. Prokopenko and A. Y. Zomaya, *Assortative mixing in directed biological networks*, IEEE/ACM Transactions on Computational Biology and Bioinformatics, **9** (2012), 66–78.

[22] M. Rubinov, O. Sporns, C. van Leeuwen and M. Breakspear, *Symbiotic relationship between brain structure and dynamics*, BMC Neuroscience, **10** (2009), 55.

[23] R. V. Sole and S. Valverde, *Information theory of complex networks: on evolution and architectural constraints*, in "Lecture Notes in Physics" (eds. E. Ben-Naim, H. Frauenfelder, and Z. Toroczkai), Springer, (2004), 650.

[24] S. Zhou and R. J. Mondragón, *Towards modelling the internet topology - the interactive growth model*, Physical Review E, **67** (2003), 026126.

[25] S. Zhou and R. J. Mondragón, *The rich-club phenomenon in the internet topology*, Physical Review E, **8** (2004), 180–182.

Received December 2011; revised June 2012.

E-mail address: mahendrarajah.piraveenan@sydney.edu.au
E-mail address: mikhail.prokopenko@csiro.au
E-mail address: albert.zomaya@sydney.edu.au