

Assortative Mixing in Directed Biological Networks

Mahendra Piraveenan, Mikhail Prokopenko, and Albert Zomaya

Abstract—We analyze assortative mixing patterns of biological networks which are typically directed. We develop a theoretical background for analyzing mixing patterns in directed networks before applying them to specific biological networks. Two new quantities are introduced, namely the in-assortativity and the out-assortativity, which are shown to be useful in quantifying assortative mixing in directed networks. We also introduce the local (node level) assortativity quantities for in- and out-assortativity. Local assortativity profiles are the distributions of these local quantities over node degrees and can be used to analyze both canonical and real-world directed biological networks. Many biological networks, which have been previously classified as disassortative, are shown to be assortative with respect to these new measures. Finally, we demonstrate the use of local assortativity profiles in analyzing the functionalities of particular nodes and groups of nodes in real-world biological networks.

Index Terms—Networks, graph theory, assortativity, systems biology, biological networks.

1 INTRODUCTION

ASSORTATIVITY is the tendency observed in networks where nodes mostly connect with similar nodes. Typically, this similarity is interpreted in terms of degrees of nodes. Many scale-free networks in real world show the tendency where highly connected nodes link with other highly connected nodes (that is, nodes mix assortatively). The reverse is also true in some networks, where highly connected nodes are more likely to make links with isolated, less connected nodes, i.e., to mix disassortatively. In both cases, the probability of creating a link depends on the degrees of both nodes. Averaging across the network, assortativity quantifies the tendency for preferential association within the network [1], [2], [3].

Recently, a new measure called local assortativity was proposed to quantify mixing patterns at the node level of a network [4], [5]. Specifically, local assortativity is defined as a node's contribution to network assortativity. It was shown that local assortativity distributions could be constructed by plotting average local assortativity against node degree, and such distributions can be used to gain insights about the topological design of networks. It was also shown that naturally existing networks can be classified based on their local assortativity distributions. Furthermore, the mechanisms and growth models that can produce each of these

local assortativity-based classes were analyzed [6]. In all of the above mentioned works, local assortativity has been defined and used for undirected networks.

Many naturally occurring networks, and biological networks in particular, are directed networks. Transcription networks [7], neural networks [8], Gene Regulatory Networks (GRNs) [9], and brain (cortical) networks [10] fall into this category. While assortativity of some biological networks, such as food webs, has been analyzed by considering them as undirected [2], we generally can get far better insights about their topologies if their directedness is taken into account. As we show in this paper, biological networks that may appear disassortative, when directedness is not considered, do in fact become assortative when they are considered as directed networks. Furthermore, as we also show, their local assortativity profiles are much more informative when directedness is taken into account. Therefore, it is necessary that a sound theoretical background is developed for analyzing assortativity and local assortativity in directed networks. In this paper, we attempt this task and use our results to analyze topological patterns in directed biological networks.

Assortativity has been defined in [3] for directed networks as a correlation function, similar to the definition for assortativity in the undirected case. However, the meaning of this definition is not so sound. In the undirected case, assortativity measures the tendency of a node to connect with other nodes that have similar degree. In directed case, the "in-degree" and "out-degree" of nodes come into play. According to the definition in [3], the assortativity for directed networks measures "the tendency of nodes to connect with nodes that have in-degrees similar to (the original) node's out-degree." It would make more sense if assortativity instead measures the tendency for nodes to connect with other nodes with similar out-degrees or similar in-degrees. With this in mind, we propose alternative definitions for assortativity in directed networks. Then we define and derive expressions for local assortativity in each case. We lay a sound background for these new definitions by analyzing some canonical networks. Only

• M. Piraveenan is with the School of Information Technologies, The University of Sydney, and the CSIRO Information and Communications Technology Centre, Locked Bag 17, North Ryde, NSW 1670, Australia. E-mail: mahendra.piraveenan@csiro.au.

• M. Prokopenko is with the CSIRO Information and Communications Technology Centre, Locked Bag 17, North Ryde, NSW 1670, Australia, and Max Planck Institute for Mathematics in the Sciences, Inselstraße 22-26, Leipzig 04103, Germany. E-mail: Mikhail.prokopenko@csiro.au.

• A. Zomaya is with the School of Information Technologies, Building J12, The University of Sydney, Sydney, NSW 2006, Australia. E-mail: a.zomaya@usyd.edu.au.

Manuscript received 19 May 2009; revised 26 Oct. 2009; accepted 1 Nov. 2009; published online 20 Aug. 2010.

For information on obtaining reprints of this article, please send e-mail to: tcbb@computer.org, and reference IEEECS Log Number TCBB-2009-05-0078. Digital Object Identifier no. 10.1109/TCBB.2010.80.

after these steps, we can and do embark on analyzing assortativity and local assortativity in real-world directed biological networks, using the derived theory, and analyze the significance of our results in terms of the functionality and topology of the networks.

2 ASSORTATIVITY AND LOCAL ASSORTATIVITY

2.1 Review of Definitions and Terminology

It is necessary to review some existing concepts before we formally define our measure for directed assortativity and directed local assortativity. The concepts of degree distribution $p(k)$ and excess degree (remaining degree) distribution $q(k)$ for undirected networks are well known [1], [2], [3]. The measures are related by

$$q(k) = \frac{(k+1)p(k+1)}{\sum_j jp(j)}, \quad (1)$$

where j, k are the excess degrees at each end of a given link. When the upper limit of the sum over links is not specified, the maximum excess degree $N_p - 1$ is implied, where N_p is the maximum degree. The term $\sum_j jq(j)$ can be understood as μ_q , the expected value or mean of the excess degree distribution.

Given $q(k)$, one can introduce the quantity $e_{j,k}$ as the joint probability distribution of the excess degrees of the two nodes at either end of a randomly chosen link. $e_{j,k}$ is symmetric for undirected networks, but not necessarily so for directed networks. Assortativity of an undirected network is defined as

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

where σ_q is the standard deviation of the excess degree distribution of the network, $q(k)$.

A similar definition has been proposed for node-based (i.e., vertex-based) assortativity coefficient in directed networks [3]. For this purpose, [3] considered the distribution of finding a target node with in-degree k at the end of a randomly selected link, $q^{in}(k)$, and the distribution of finding a source node with out-degree k at the start of a randomly selected link, $q^{out}(k)$.

Also defined is the $e_{j,k}$ distribution for directed networks, where $e_{j,k}$ is now the probability that a randomly chosen directed link leads into a node of in-degree j and out of a node of out-degree k . If no degree-related preferential mixing occurs, then $e_{j,k} = q^{in}(j)q^{out}(k)$ for all j, k , and thus, the assortativity coefficient can be defined via the divergence between $e_{j,k}$ and $q^{in}(j)q^{out}(k)$:

$$r_d = \frac{1}{\sigma_q^{in} \sigma_q^{out}} \left[\sum_{j,k} jk(e_{j,k} - q^{in}(j)q^{out}(k)) \right]. \quad (3)$$

The latter can also be written as

$$r_d = \frac{1}{\sigma_q^{in} \sigma_q^{out}} \left(\sum_{j,k} jke_{j,k} - \mu_q^{in} \mu_q^{out} \right), \quad (4)$$

where μ_q^{in} and μ_q^{out} are the means of the distributions q^{in} and q^{out} , respectively. Similarly, σ_q^{in} and σ_q^{out} are the standard deviations of the respective distributions.

2.2 Local Assortativity

Local assortativity for undirected networks can be defined as follows [4], [5]:

$$\rho = \frac{j(j+1)(\bar{k} - \mu_q)}{2M\sigma_q^2}, \quad (5)$$

where j is the node's excess degree \bar{k} is the average excess degree of its neighbors, and $\sigma_q \neq 0$. Following [4] and [5], we now define local assortativity for directed networks. We propose to define local assortativity as the contribution that each node makes to the network assortativity, where network assortativity is given by (4). That is, we need to determine how much contribution each node makes to the term (4).

First of all, we point out that the degrees j and k in this expression are understood to be the in-degree of the target node j^{in} and the out-degree of the source node k^{out} , respectively. We use k to indicate properties of "source" nodes, and j to indicate properties of "target" nodes: this is meaningful when considering a link (j, k) . When considering properties of any individual node, such as its in-degrees and out-degrees, we stay with the notation k^{out} for out-degrees and j^{in} for in-degrees, although this is not strictly necessary. That is, k^{out} and j^{out} should be understood uniquely as out-degrees, while k^{in} and j^{in} should both refer to in-degrees.

The distribution $e_{j,k}$ is the corresponding joint distribution of out-degrees of the sources and in-degrees of the targets.

Second, following [4], let us consider the contribution of each node to the term $\sum_{j,k} jke_{j,k}$. Suppose, we visit all nodes in a network, and in turn, from each node, we visit all links that depart from that node, keeping the directionality of the links in mind. In a network with N nodes and M links, the total number of visits we will thus make will be M . Suppose, we build up the probability distribution $e_{j,k}$ as we make these visits. Each link will add a probability of $\frac{1}{M}$ to the pair of (j^{in}, k^{out}) , in accordance with the definition of $e_{j,k}$. Thus, each visit to a link will contribute $j^{in}k^{out}/M$ to the sum $\sum_{j,k} jke_{j,k}$. Therefore, considering a node with out-degree k^{out} which is connected to nodes with in-degrees $j_1^{in}, j_2^{in}, \dots, j_{k^{out}}^{in}$, it will contribute

$$\alpha_1 = \frac{k^{out} j_1^{in}}{M} + \frac{k^{out} j_2^{in}}{M} + \dots + \frac{k^{out} j_{k^{out}}^{in}}{M} = \frac{k^{out}}{M} \sum_{i=1}^{k^{out}} j_i^{in}$$

to the sum $\sum_{j,k} jke_{j,k}$. Let us denote the average in-degree of a node's neighbors as $\bar{j}^{in} = \frac{1}{k^{out}} \sum_{i=1}^{k^{out}} j_i^{in}$. Then the individual node's contribution to the sum $\sum_{j,k} jke_{j,k}$ is

$$\alpha_1 = \frac{k^{out}}{M} \sum_{i=1}^{k^{out}} j_i^{in} = \frac{(k^{out})^2 \bar{j}^{in}}{M}.$$

An alternative definition is also possible, if we, in turn, consider each node and all links that come into that node. In this case, the contribution to the sum $\sum_{j,k} jke_{j,k}$ is

$$\alpha_2 = \frac{(j^{in})^2 \bar{k}^{out}}{M}, \quad (6)$$

where j^{in} is the in-degree of the target node, \bar{k}^{out} is the average neighbor out-degree, neighbors being those nodes from which this node can be reached. Therefore, let us say that

the “average” contribution α_d of a node to the term $\sum_{jk} jke_{j,k}$ is the average of the above two quantities, α_1 and α_2 ,

$$\alpha_d = \frac{(k^{out})^2 \bar{j}^{in} + (\bar{j}^{in})^2 \bar{k}^{out}}{2M}. \quad (7)$$

Finally, let us consider a node’s contribution to the term $\mu_q^{in} \mu_q^{out}$. It can be seen that the expectation of distribution $q^{in}(k)$ can be written in two equivalent forms:

$$\mu_q^{in} = \frac{1}{M} \sum_{m=1}^M k_m^{in} = \frac{1}{M} \sum_{v=1}^N (k_v^{in})^2, \quad (8)$$

where k^{in} is the in-degree of a source node of a given link m (the first form), or the in-degree of a given source node v (the second form). Similarly,

$$\mu_q^{out} = \frac{1}{M} \sum_{m=1}^M k_m^{out} = \frac{1}{M} \sum_{v=1}^N (k_v^{out})^2, \quad (9)$$

where k^{out} is the out-degree. The last two expressions lead to

$$\mu_q^{in} \mu_q^{out} = \frac{1}{M^2} \sum_{v=1}^N (k_v^{in})^2 \sum_{v=1}^N (k_v^{out})^2. \quad (10)$$

Without loss of generality, let us consider the contribution of node 1 to the above expression rewritten as

$$\frac{1}{M^2} \left((k_1^{in})^2 + \sum_{v=2}^N (k_v^{in})^2 \right) \left((k_1^{out})^2 + \sum_{v=2}^N (k_v^{out})^2 \right). \quad (11)$$

We assume that a term, such as $(k_1^{in})^2$, is contributed fully by node 1, whereas a multiplication term, such as $(k_1^{in})^2 (k_2^{in})^2$, is contributed equally by node 1 and node 2. Therefore, node 1 contributes

$$\frac{(k_1^{in})^2 (k_1^{out})^2}{M^2} + \frac{(k_1^{out})^2 \sum_{v=2}^N (k_v^{in})^2}{2M^2} + \frac{(k_1^{in})^2 \sum_{v=2}^N (k_v^{out})^2}{2M^2}. \quad (12)$$

This yields

$$\frac{1}{2M^2} \left((k_1^{out})^2 \sum_{v=1}^N (k_v^{in})^2 + (k_1^{in})^2 \sum_{v=1}^N (k_v^{out})^2 \right), \quad (13)$$

and using (8) and (9), can be further reduced to

$$\frac{1}{2M} ((k_1^{out})^2 \mu_q^{in} + (k_1^{in})^2 \mu_q^{out}). \quad (14)$$

Thus, we obtain a node’s contribution to the term $\mu_q^{in} \mu_q^{out}$ as

$$\beta_d = \frac{1}{2M} ((k^{out})^2 \mu_q^{in} + (k^{in})^2 \mu_q^{out}). \quad (15)$$

The standard deviations are already used as scaling terms, so we need not worry about their contributions. Therefore, we can now define a node’s contribution to directed assortativity of a network, represented by (4), by using α_d given by (7), and β_d given by (15):

$$\rho_d = \frac{\alpha_d - \beta_d}{\sigma_q^{in} \sigma_q^{out}}. \quad (16)$$

This results in

$$\rho_d = \frac{(k^{out})^2 (\bar{j}^{in} - \mu_q^{in}) + (k^{in})^2 (\bar{j}^{out} - \mu_q^{out})}{2M \sigma_q^{in} \sigma_q^{out}}, \quad (17)$$

where k^{out} is the out-degree of the node under consideration, k^{in} is its in-degree, \bar{j}^{in} is the average in-degree of its neighbors (from which the node can be reached), and \bar{k}^{out} is the average out-degree of its neighbors (which can be reached from the node). $\sigma_q^{in} \neq 0$, $\sigma_q^{out} \neq 0$. By including the scaling terms σ_q^{in} and σ_q^{out} , we ensure that the equation for local assortativity for a directed network satisfies the condition

$$r_d = \sum_{i=1}^N \rho_d, \quad (18)$$

where r_d is the assortativity of the directed network. It may also be illustrative to look at ρ_d as the average of two quantities:

$$\rho_1 = \frac{(k^{out})^2 (\bar{j}^{in} - \mu_q^{in})}{M \sigma_q^{in} \sigma_q^{out}} \quad (19)$$

and

$$\rho_2 = \frac{(k^{in})^2 (\bar{j}^{out} - \mu_q^{out})}{M \sigma_q^{in} \sigma_q^{out}}. \quad (20)$$

In general ρ_1 and ρ_2 are not equal for individual nodes.

2.3 Motivation for Alternative Local Assortativity Definitions

Before we proceed to analyze the local assortativity distributions in directed model and real-world networks, in terms of ρ_d , let us pause to reflect on the meaning of assortativity and local assortativity in directed networks. As we have mentioned, assortativity in undirected networks is the tendency for nodes to connect with nodes which have similar degrees. In directed networks, however, according to the definitions in [3], we are looking at the correlation between out-degrees of the source nodes and in-degrees of the target nodes (or vice versa). Therefore, assortativity here is the tendency of nodes to connect with other nodes, whose in-degrees are similar to the considered node’s out-degree. If an individual node has high local assortativity ρ_d , it means that it has high out-degree and connects to other nodes with high in-degree, or it means that this node has high in-degree and connects to other nodes with high out-degrees.

Consider a directed biological network where there are regulators and regulatees, such as gene regulatory networks. Suppose, the links in such networks are directed from regulator to regulatee. A node which has high out-degree will be a dominant regulator. However, the impact of the regulator in the network will be maximised if the nodes that this regulator regulates, in turn, regulate a lot of other nodes, i.e., they themselves have high out-degrees. Therefore, to understand the importance of such nodes in the networks, we need a quantity that favours nodes that have high out-degree and are connected to other nodes with high out-degrees. Similarly, the nodes which are most likely to have complex regulation patterns are those nodes which

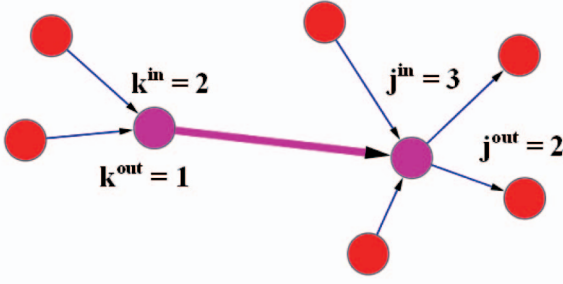


Fig. 1. In-degrees and out-degrees of nodes with respect to a link. Note the highlighted link leaves *from* a node of in-degree two, and out-degree one. It goes *into* a node of in-degree three, and out-degree two.

are regulated by many nodes, each of which, in turn, is regulated by many other nodes. To measure this tendency, we need a quantity which favours nodes with high in-degree which are (directionally) connected to other nodes with high in-degrees, as shown in Fig. 1.

Such quantities cannot be obtained by decomposing (3) for directed assortativity above, i.e., by analyzing each node's contribution to this assortativity coefficient. In other words, ρ_d fails to capture the "cascading" effect. We, therefore, need alternative assortativity coefficients, which measure the tendencies where nodes preferentially connect with other nodes with similar out-degrees to themselves, or nodes preferentially connect with other nodes with similar in-degrees. We call these tendencies as "out-assortativity" and "in-assortativity" of a network.

2.4 Out-Assortativity and In-Assortativity

Let us define out-assortativity of a network as the tendency where nodes connect with other nodes which have similar out-degrees to themselves. In-assortativity is, on the other hand, the tendency where nodes connect with other nodes with similar in-degrees. Note that these definitions fit well with the "generic" definition of assortative mixing, where similarity can be interpreted in terms of any given single quantity [2], [3]. Whereas the previous definition of node assortativity in directed networks defined similarity in terms of two different quantities (out-degree *and* in-degree) for a node pair.

To formally define out-assortativity and in-assortativity, however, we first need to define a few concepts in terms of node distributions. These definitions are similar but subtly different from the ones we have reviewed above.

We have defined $e_{j,k}$ as the probability distribution of a link going into a node with j in-degree and out of a node with k out-degree. Now, let us define $e_{j,k}^{out}$ distribution as the probability of a link going into a node of j out-degree and out of a node of k out-degree. While $q^{out}(k)$ is the probability distribution of a link going out of a node with k out-degree, as before, let us define $q^{out}(k)$ as the probability distribution of a link going *into* a node with k out-degree. Similarly, let us define $e_{j,k}^{in}$ distribution as the probability of a link going into a node of j in-degree and out of a node of k in-degree. Furthermore, while $q^{in}(j)$ is the probability distribution of a link going into a node with j in-degree, let us define $q^{in}(j)$ as the probability distribution of a link going *out* of a node with j in-degree. It is important to appreciate the subtle differences in these distributions. An example demonstrating these differences is shown in Fig. 1.

We now can define the out-assortativity and in-assortativity of a network, which we consider as more meaningful measures of assortativity in directed networks. The out-assortativity of a network is the tendency where nodes tend to connect with other nodes with similar *out-degrees*. This is formally defined as

$$r_{out} = \frac{1}{\sigma_q^{out} \sigma_{q'}^{out}} \left[\sum_{jk} j k e_{j,k}^{out} - \mu_q^{out} \mu_{q'}^{out} \right], \quad (21)$$

where σ_q^{out} is the standard deviation of q^{out} , and $\sigma_{q'}^{out}$ is the standard deviation of q'^{out} of the network. Where a network has positive r_{out} , it means that nodes with high out-degrees tend to connect to other nodes with high out-degrees. If a network has negative r_{out} , it means that nodes with high out-degrees tend to connect to nodes with low out-degrees.

Similarly, the in-assortativity of a network is the tendency whereby nodes tend to connect with other nodes with similar *in-degrees*. In-assortativity can be formally defined as

$$r_{in} = \frac{1}{\sigma_q^{in} \sigma_{q'}^{in}} \left[\sum_{jk} j k e_{j,k}^{in} - \mu_q^{in} \mu_{q'}^{in} \right], \quad (22)$$

where σ_q^{in} is the standard deviation of q^{in} , and $\sigma_{q'}^{in}$ is the standard deviation of q'^{in} of the network. If a network has positive r_{in} , it means that nodes with high in-degrees tend to connect to other nodes with high in-degrees. If a network has negative r_{in} , it means that nodes with high in-degrees tend to connect to nodes with low in-degrees.

2.5 Local Out-Assortativity and Local In-Assortativity

Now, we can define local assortativity for directed networks in terms of exclusively out-degrees and exclusively in-degrees. We define local out-assortativity of a node as a node's contribution to the network's out-assortativity. As demonstrated in Appendix, the local out-assortativity of a node is given by

$$\rho_{out} = \frac{k^{out}}{2M\sigma_q^{out}\sigma_{q'}^{out}} (k^{out}(\bar{j}^{out} - \mu_{q'}^{out}) + j^{in}(\bar{k}^{out} - \mu_q^{out})), \quad (23)$$

where k^{out} is the node's out-degree, j^{in} is the node's in-degree, \bar{j}^{out} is the average out-degree of the "target" neighbors *to* which this node has a directed link, and \bar{k}^{out} is the average out-degree of the "source" neighbors *from* which this node is reachable via a directed link. Furthermore, μ_q^{out} and $\mu_{q'}^{out}$ are the expectations of the distributions $q^{out}(k)$ and $q'^{out}(k)$, respectively; σ_q^{out} and $\sigma_{q'}^{out}$ are the standard deviations of these quantities.

Local out-assortativity can be interpreted in the following way. It is a linear combination of two terms, $(\bar{j}^{out} - \mu_{q'}^{out})$ and $(\bar{k}^{out} - \mu_q^{out})$. The first term represents the difference between the average out-degree of target nodes from this node, and the average out-degree of target nodes globally (that is, the expected out-degree of a node at the end of a directed link). Similarly, the second term represents the difference between the average out-degree of source nodes that are neighbors to this node, and the average out-degree of source nodes globally. That is, both terms compare the

TABLE 1
Summary of the Assortativity Measures

Assortativity	Global (network) level	Local (node) level	
Undirected networks	r as defined in [2], [3], [1]	ρ as defined in [4], [5]	
Directed networks	r_d as defined in [3]	ρ_d	$r_d = \sum \rho_d$
	r_{in}	ρ_{in}	$r_{in} = \sum \rho_{in}$
	r_{out}	ρ_{out}	$r_{out} = \sum \rho_{out}$

Previously introduced measures are indicated by citations. New measures introduced in this paper are without citations.

local average with the global average. The overall local out-assortativity is a scaled linear combination of these terms. Therefore, ρ_{out} is increased if local average of a node's neighbors, in terms of out-degrees, is higher than the global average. On the other hand, ρ_{out} is reduced if the global average of out-degrees is higher than the local averages around a given node. In this case, the node tends to become locally out-disassortative. This interpretation is similar to local assortativity in the undirected case, as proposed in [5].

Similarly, we may define local in-assortativity of a node as a node's contribution to the network in-assortativity, given by

$$\rho_{in} = \frac{j^{in}}{2M\sigma_q^{in}\sigma_q^{in}}(j^{in}(\bar{j}^{in} - \mu_q^{in}) + k^{out}(\bar{k}^{in} - \mu_q^{in})), \quad (24)$$

where k^{out} is the node's out-degree, j^{in} is the node's in-degree, \bar{j}^{in} is the average in-degree of the "target" neighbors to which this node has a directed link, and \bar{k}^{in} is the average in-degree of the "source" neighbors from which this node is reachable via a directed link. μ_q^{in} and μ_q^{out} are the expectations of the distributions $q^{in}(k)$ and $q^{out}(k)$, respectively; σ_q^{in} and σ_q^{out} are the standard deviations of the same quantities. Note that the interpretation of ρ_{in} , in terms of the differences between local and global averages, is similar to that given for ρ_{out} above.

The local out-assortativity and local in-assortativity indeed satisfy the sum rules:

$$r_{out} = \sum_{i=1}^N \rho_{out}, \quad (25)$$

$$r_{in} = \sum_{i=1}^N \rho_{in}. \quad (26)$$

The above mentioned and defined concepts are summarised in Table 1. Note that some of the measures have been introduced in previous studies, as indicated by citations given. Other measures (without citation details) are introduced in this paper for the first time.

3 SINGULARITY CASES OF DIRECTED LOCAL ASSORTATIVITY

When defining directed assortativity and directed local assortativity, it is important to carefully consider what happens when one of the degree distributions, $q^{in}(k)$, $q^{out}(k)$,

$q^{out}(k)$, $q^{out}(k)$, is a Kronecker δ function (i.e., only one type of degree exists), making the variance zero. A few combinations are worthy of attention here.

Let us consider the in-assortativity. Suppose $\sigma_{q^{in}}$ and $\sigma_{q^{out}}$ are both zero. Let us then look at the degree k^{in} for which these distributions are nonzero. If both these distributions are nonzero at the same point (k_{in}), then we need to obtain network assortativity equal to 1, since all nodes have the same degree. Therefore, in this case, we define local in-assortativity of a node as

$$\rho_{in} = \frac{k^{in}}{M}, \quad (27)$$

where k^{in} is the in-degree of the node concerned.

Let us note that both these distributions cannot be Kronecker δ functions and nonzero at different points (for different k^{in}). If q^{in} is a δ function, then there is only one type of in-degrees present in the network, and q^{out} also must be a δ function and be nonzero at the same point for the same value of k_{in} . For the same reason, if $\sigma_{q^{in}} = 0$, then it must follow that $\sigma_{q^{out}} = 0$ too.

Similarly, if $\sigma_{q^{out}} = 0$, then $\sigma_{q^{in}}$ would have to be zero too, and we can define ρ_{out} as

$$\rho_{out} = \frac{k^{out}}{M}. \quad (28)$$

Let us point out, however, that in real-world networks of reasonable size, these singularity conditions rarely occur.

4 DISTRIBUTIONS OF LOCAL ASSORTATIVITY

Since local assortativity is a property of a node, it is possible to construct local assortativity distributions for a given directed network, plotting local assortativity values against degrees (in-degrees or out-degrees). Following [4], we may calculate the average local assortativity value for all nodes with a given in-degree k^{in} , or a given out-degree k^{out} . We propose that ρ_{out} should be plotted against out-degree, since this quantity measures the contribution of a node to the out-degree correlation, while ρ_{in} should be plotted against in-degree, since this quantity measures the contribution of a node to the in-degree correlation. ρ_d is plotted against node degree. If we denote by $N(k^{out})$ the number of nodes with out-degree k^{out} , by $N(k^{in})$ the number of nodes with in-degree k^{in} , and by $N(k)$ the number of nodes with degree k , the following equations hold true:

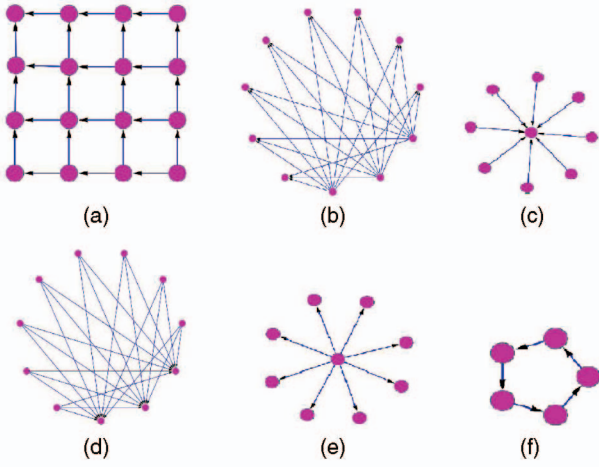


Fig. 2. Model networks: (a) grid network with links directed uniformly, (b) “inward” multistar with links directed toward the hubs, (c) “outward” multistar with links directed toward the peripheral nodes, (d) “outward” star with links directed toward peripheral nodes, (e) “inward” star with links directed toward the hub, and (f) ring with directed links with uniform orientation.

$$r_d = \sum_k N(k) \overline{\rho_d}(k), \quad (29)$$

where $\overline{\rho_d}(k)$ is the average ρ_d of all nodes with degree k ;

$$r_{out} = \sum_{k^{out}} N(k^{out}) \overline{\rho_{out}}(k^{out}), \quad (30)$$

where $\overline{\rho_{out}}(k^{out})$ is the average ρ_{out} of all nodes with out-degree k^{out} ;

$$r_{in} = \sum_{k^{in}} N(k^{in}) \overline{\rho_{in}}(k^{in}), \quad (31)$$

where $\overline{\rho_{in}}(k^{in})$ is the average ρ_{in} of all nodes with in-degree k^{in} .

In the following sections, we will mainly consider $\overline{\rho_{out}}$ versus k^{out} distributions, and $\overline{\rho_{in}}$ versus k^{in} distributions. We will not concentrate on $\overline{\rho_d}$ versus k distributions since as we pointed out earlier, the quantity ρ_d is less helpful in understanding the topological role of nodes in directed networks. However, we will give a few examples of these plots for comparison.

5 LOCAL ASSORTATIVITY IN CANONICAL NETWORKS

Before analyzing local assortativity profiles for real-world and simulated directed networks, let us look at these profiles for some important but simple topologies. Let us first consider the $\overline{\rho_{out}}$ versus k^{out} profiles. Star networks are commonly used in communication networks and are an important motif embedded in larger networks [1]. A few varieties of star topologies are possible, as shown in Fig. 2, and it is easy to understand that all of these topologies result in disassortative networks. As we show below, the local assortativity distributions in these cases are δ functions, or linear combinations of two δ functions, and all nodes are disassortative in nature. Grid layout, on the other hand, results in a perfectly assortative network, and all nodes are

assortative, even though here too, the distribution is a δ function. The ring topology results in a perfectly assortative network, with all nodes equally contributing to this assortativity, therefore, the $\overline{\rho_{out}}$ versus k^{out} is again a delta function. Similar or complimentary results are obtained for $\overline{\rho_{in}}$ versus k^{in} distributions. We formally present these observations below, which can be mathematically derived and have been verified by simulation.

Regular lattice (a)

$$\rho_{out} = \frac{1}{N} \delta_{j,m^*} \quad (32)$$

$$\rho_{in} = \frac{1}{N} \delta_{j,n^*} \quad (33)$$

where m^* is the out-degree of a node, and n^* is the in-degree.

Inward multistar (b)

$$\rho_{out} = -\frac{1}{N - n^*} \delta_{j,n^*} \quad (34)$$

$$\rho_{in} = -\frac{1}{n^*} \delta_{j,N-n^*} \quad (35)$$

Outward multistar (c)

$$\rho_{out} = -\frac{1}{n^*} \delta_{j,N-n^*} \quad (36)$$

$$\rho_{in} = -\frac{1}{N - n^*} \delta_{j,n^*} \quad (37)$$

where n^* is the number of hubs.

Outward star (d)

$$\rho_{out} = -\delta_{j,N-1} \quad (38)$$

$$\rho_{in} = -\frac{1}{N-1} \delta_{j,1} \quad (39)$$

Inward star (e)

$$\rho_{out} = -\frac{1}{N-1} \delta_{j,1} \quad (40)$$

$$\rho_{in} = -\delta_{j,N-1} \quad (41)$$

Ring (f)

$$\rho_{out} = \rho_{in} = \frac{1}{N} \delta_{j,1}. \quad (42)$$

Now, we proceed to analyze assortativity and local assortativity distributions in simulated and real-world directed scale-free networks.

6 ASSORTATIVITY OF REAL-WORLD DIRECTED NETWORKS

Table 2 shows the assortativity coefficients of a number of directed networks, including neural networks, GRNs, transcriptions networks, cortical networks, and foodwebs. r_d is the network assortativity according to (3) for directed networks, whereas r_{out} and r_{in} represent the out-assortativity and in-assortativity of networks respectively. r represents the assortativity when networks are considered undirected, and it is meaningless for directed networks and provided only for comparison. A clear tendency can immediately be observed in these values. First of all, the

TABLE 2
Assortativity in Real-World Directed Networks

Network	Size N	r		r_d		r_{out}		r_{in}	
			$I(r)$		$I(r_d)$		$I(r_{out})$		$I(r_{in})$
Neural networks									
C. elegans	297	-0.15	0.42	-0.23	0.46	0.1	1.01	-0.09	0.35
GRNs									
rat (R. norvegicus)	819	0.86	1.65	0.31	0.95	0.64	2.24	0.59	0.75
human (H. sapiens)	1452	-0.03	1.08	-0.03	0.61	0.2	1.41	-0.01	0.48
mouse (M. musculus)	981	0.66	1.92	0.2	0.75	0.53	1.98	0.49	0.62
C. elegans	581	-0.09	0.94	-0.12	0.68	0.36	1.06	0.01	0.38
A. thaliana	395	-0.04	1.07	-0.12	0.61	0.16	1.31	0.03	0.52
Transcription nets									
E. coli	1147	-0.26	0.96	0.06	0.36	0.17	1.26	0.03	0.11
C. glucamitum	539	-0.37	0.84	-0.04	0.31	0.09	0.22	-0.01	0.13
C. jeikeium	52	-1	1	undefined	0	-1	0	-1	0
C. efficiens	50	-0.64	0.86	undefined	0	-1	0	-1	0
Cortical networks									
human	994	0.17	0.19	0.13	0.19	0.17	0.19	0.17	0.19
Macaque monkey	71	0.02	0.97	-0.01	0.41	0.06	0.4	-0.01	0.39
Macaque sensory motor cortex	47	0.01	0.68	-0.02	0.4	0.03	0.45	-0.02	0.5
Cat cortex	65	0.01	0.52	-0.05	0.32	-0.03	0.37	0.09	0.4
Foodwebs									
Chesapeake Lower	170	-0.39	0.64	-0.45	0.57	0.21	0.7	-0.06	0.49
Chesapeake Upper	193	-0.33	0.38	-0.38	0.58	0.1	1.61	-0.12	0.49
Chrystal river c	106	-0.33	0.49	-0.48	0.63	0.08	1.38	-0.14	0.47
Chrystal river d	90	-0.46	0.45	-0.54	0.65	0.06	1.18	-0.18	0.34
Bay wet	2216	-0.12	0.39	-0.23	0.71	0.02	2.71	0.24	0.7
Bay dry	2248	-0.11	0.39	-0.23	0.75	0.03	2.78	0.25	0.74

The table shows assortativity coefficients calculated treating the networks as undirected and directed, and the out-degree and in-degree correlations. The table also shows corresponding mutual information quantities. The sources for the biological networks are [10], [7], [9], [8], [11].

relative level of assortativity is largely preserved over different ways of measuring it: there are no dramatic changes. Second, the networks tend to be more assortative when their in-assortativity and out-assortativity are considered than when they are combined together. For example, consider the neural network of *C. elegans*. Even though the network appears disassortative with $r_d = -23\%$, its $r_{out} = +10\%$ and $r_{in} = -9\%$, and both values shifted considerably toward the (positive) assortativity side. Similarly, if we consider *Chesapeake lower* foodweb, the network seems disassortative with $r_d = -45\%$, even though when r_{in} and r_{out} are considered separately, the values are $r_{in} = -6\%$ and $r_{out} = +21\%$, again, both being shifted considerably toward positive values of assortativity. Fig. 3 captures this tendency. Therefore, we may conclude that there is a *weaker signature* of disassortative mixing when out-degrees and in-degrees are considered separately. Indeed, most directed networks tend to be considerably disassortative when the *tendency of nodes mixing with other nodes which have in-degrees similar to their own out-degrees* is considered. However, these networks tend to be *more assortative when the tendency of nodes mixing with other nodes which have out-degrees similar to*

their own out-degrees is considered, or when the *tendency of nodes mixing with other nodes which have in-degrees similar to their own in-degrees* is considered.

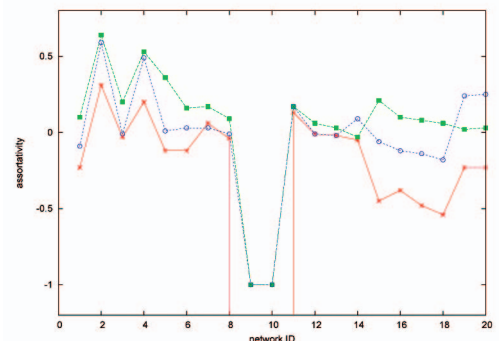


Fig. 3. Shift in assortativity coefficient when in-degrees and out-degrees are considered separately. The points correspond to networks in Table 1. Crosses: directed assortativity coefficients. Filled squares: directed out-assortativity coefficients. Circles: directed in-assortativity coefficients. Note the upward shift in values when out-assortativity and in-assortativity are compared to assortativity.

Furthermore, we note that r_{out} values tend to be even more assortative than r_{in} values. While in-assortativity of the networks considered is still slightly disassortative ($r_d < 0$), out-assortativity remains slightly or strongly on the assortative side ($r_d > 0$).

Note also that the cortical networks show comparatively less difference when r_{in} and r_{out} are considered separately, from the case when r_d is calculated as a single value. This could be because of the fact that the cortical networks have a comparatively high link density, and thus, many pairs of nodes have links connecting in both directions. Therefore, the effect of directionality is minimized, and in-degrees and out-degrees of nodes tend to be similar.

7 ASSORTATIVITY AND INFORMATION CONTENT IN DIRECTED NETWORKS

To further understand these tendencies when out-degree distributions and in-degree distributions are considered separately for assortative mixing, we have considered the tendencies in mixing-related mutual information in networks. In previous sections, we defined a number of new degrees or link distributions for directed networks. These include $e_{j,k}^{out}$, $e_{j,k}^{in}$, $q^{in}(k)$, and $q^{out}(k)$. This was in addition to the distributions already defined and used, namely $e_{j,k}$, $q^{in}(k)$, and $q^{out}(k)$ for directed networks. We can make some interesting observations by looking at the information content of various excess degree distributions of directed networks.

Solé and Valverde [1] considered (Shannon) mutual information contained in the network, as an indicator of the amount of general correlation between network nodes. Importantly, the maximum attainable information content defines the network's capacity, in analogy with the information-theoretic notion of channel capacity—the maximum mutual information for the channel over all possible distributions of a transmitted signal. In general, information content is a vital indicator of complex nonlinear behavior in self-organizing systems, and it can be associated with *predictive information*, richness of structure (i.e., *excess entropy*), and *physical complexity* [12]. It was shown that the information content within a scale-free network increases nonlinearly with the absolute value of the assortativity [13], and can be related to network's resilience under node removal or percolation [14], [4].

Formally, the network's information content (called information transfer by Solé and Valverde [1]) is defined as

$$I(q) = H(q) - H(q | q'), \quad (43)$$

where the first term is the Shannon entropy of the network, $H(q) = -\sum_{k=0}^{N_p-1} q_k \log(q_k)$, that measures the diversity of the degree distribution or the network's heterogeneity, and the second term is the conditional entropy defined via conditional probabilities of observing a node with k links leaving it, provided that the node at the other end of the chosen link has k' leaving links. Importantly, the conditional entropy $H(q | q')$ estimates correlations in the network created by connecting the nodes with dissimilar degrees—this component affects the overall diversity or the heterogeneity of the network, but does not contribute to the amount of information within it. Informally, information content within the

network is the difference between network's heterogeneity and assortative noise within it [1].

Shannon information $I(q)$ is a better, more generic measure of dependence than the correlation functions that measure linear relations. Mutual information measures the general dependence and is thus a less biased statistic [1]. Shannon information (43) can also be seen as the Kullback-Leibler divergence $K(e_{j,k} || q_j q_k)$ [12] (i.e., relative entropy) of the product of two marginal distributions q from the joint distribution e . This divergence amounts to the expected number of extra bits that must be transmitted in order to identify (on average) excess degrees of connected nodes j and k of the link (j, k) if they are assigned using only the marginal distribution q , instead of the joint distribution $e_{j,k}$.

The entropy and information content described above are defined with respect to the degree distribution and joint degree distribution, and there are alternative definitions where the entropy of the network could be characterized by higher order correlations or community structure [15]. It is well known that Shannon information contains no inherent directionality, and various alternatives have been proposed. For example, it can be argued that transfer entropy [16] is the appropriate measure for predictive information transfer in spatiotemporal systems [17]. In this work, however, we employ mutual information to represent information content (and not a directional transfer) within a network, as suggested by Solé and Valverde [1], and extend this measure to directed networks.

If one attempts to compare networks according to their information content (analogous to optimizing communication channels), assortativity becomes the main factor: the less nonassortative is the network, the more information it can contain. Hence, to compare the levels of assortative mixing in directed biological networks, it makes sense to define relevant quantities of information content and measure these in the networks under consideration.

Mutual information in terms of excess degrees is defined for undirected networks as

$$I(q) = \sum_j \sum_k e_{j,k} \log \frac{e_{j,k}}{q(j)q(k)}. \quad (44)$$

In similar vein, mutual information for directed networks could be defined as

$$I(q^{in}; q^{out}) = \sum_j \sum_k e_{j,k} \log \frac{e_{j,k}}{q^{in}(j)q^{out}(k)}. \quad (45)$$

Furthermore, we can define mutual information exclusively in terms of in-degrees or out-degrees as:

$$I^{in}(q^{in}; q^{in}) = \sum_j \sum_k e_{j,k}^{in} \log \frac{e_{j,k}^{in}}{q^{in}(j)q^{in}(k)}, \quad (46)$$

and

$$I^{out}(q^{out}; q^{out}) = \sum_j \sum_k e_{j,k}^{out} \log \frac{e_{j,k}^{out}}{q^{out}(j)q^{out}(k)}. \quad (47)$$

By defining mutual information in terms of various excess degree distributions in directed networks, we can make interesting observations about the information content in the network topology. Table 2 shows the values of the

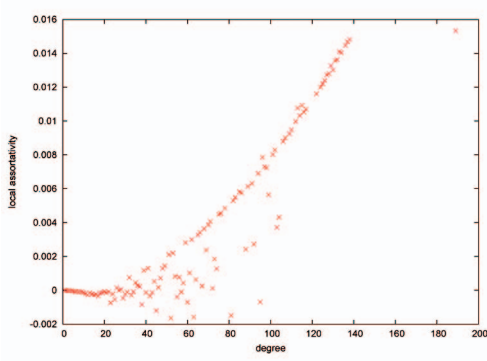


Fig. 4. Local in-assortativity distribution versus in-degree: rat GRN network.

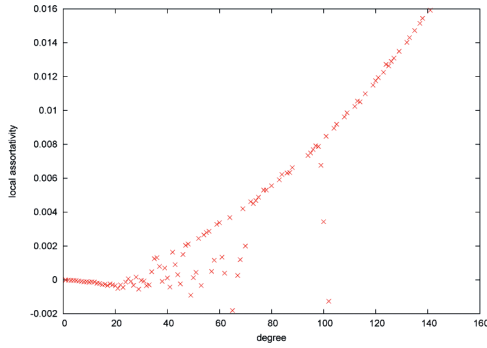


Fig. 5. Local out-assortativity distribution versus out-degree: rat GRN network.

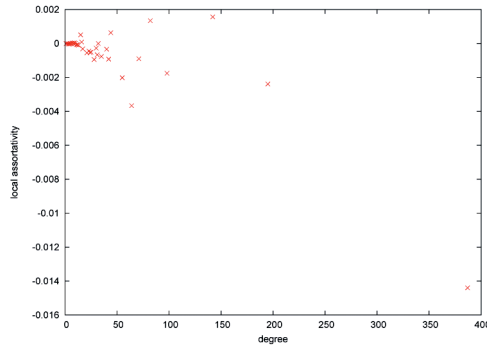


Fig. 6. Local in-assortativity distribution versus in-degree: *E. coli* transcription network.

various mutual information measures that we consider. We observe that the highest information is contained in out-degree mixing patterns. That is, it is the regulators that dominate, defining the connecting patterns of the network.

8 LOCAL ASSORTATIVITY DISTRIBUTIONS OF DIRECTED BIOLOGICAL NETWORKS

Now, let us consider local assortativity distributions of some of the networks in Table 2. We will look at both ρ_{out} versus out-degree distributions and ρ_{in} versus in-degree distributions. The local assortativity distributions of four different types of biological networks are shown in Figs. 4, 5, 6, 7, 8, 9, 10, and 11.

It can be observed that, in the rat GRN, the distributions are nonlinear and have assortative hubs. However, in some

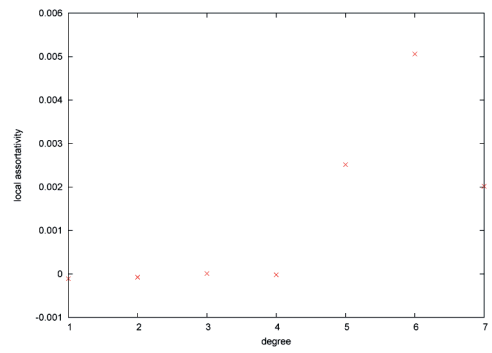


Fig. 7. Local out-assortativity distribution versus out-degree: *E. coli* transcription network.

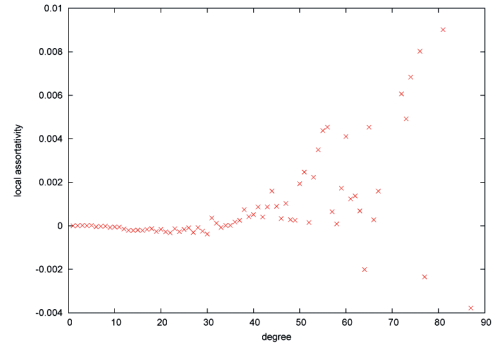


Fig. 8. Local in-assortativity distribution versus in-degree: human cortex network.

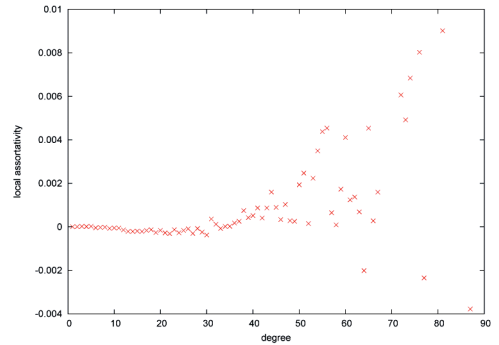


Fig. 9. Local out-assortativity distribution versus out-degree: human cortex network.

cases, such as *E. coli* transcription network (local in-assortativity) and *C. elegans* neural network (again local in-assortativity), the hubs are disassortative. Furthermore, in the human cortical network also, the largest hubs are disassortative, though the provincial hubs seem assortative. As pointed out in [5], it is possible to classify networks (directed networks in this case) based on whether hubs are assortative or disassortative. Specifically, networks can be classified as

1. assortative networks with assortative hubs,
2. disassortative networks with assortative hubs,
3. assortative networks with disassortative hubs, and
4. disassortative networks with disassortative hubs [5].

In the case of directed networks, this classification can be done for both out-degree correlations and in-degree correlations. For example, the rat GRN network would fall into the

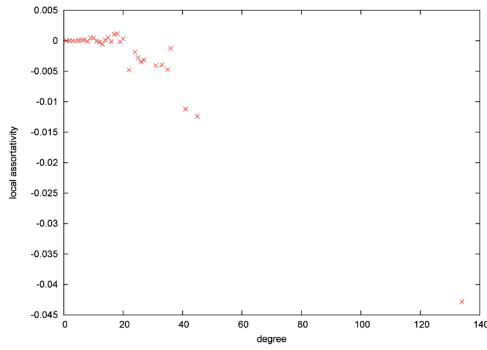


Fig. 10. Local in-assortativity distribution versus in-degree: *C. elegans* neural network.

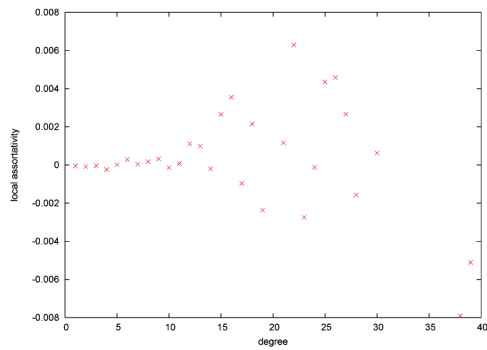


Fig. 11. Local out-assortativity distribution versus out-degree: *C. elegans* neural network.

first class, for both in-degree and out-degree correlations. Other networks, such as *E. coli* transcription, would fall into different classes depending on whether out-degree or in-degree correlation is considered. Furthermore, if the hubs with high out-degree are assortative, this means that these hubs are regulators which regulate other regulators, thereby highly influencing the expression patterns of the whole network. Similarly, if the nodes with larger in-degrees are assortative, they form “sinks” of the regulating signals. Most biological networks that we studied have assortative hubs and disassortative peripheral nodes when out-degree and in-degree are considered separately.

It is also possible to plot *individual* node degrees on X axis and local assortativity on Y axis as a scatter plot to get a different kind of local assortativity profile. This profile better highlights the individual nodes with highest ρ_{in} or ρ_{out} . For example, the ρ_{out} versus out-degree plot for *R. norvegicus* (rat) GRN network is shown in Fig. 12. The corresponding figure for *M. musculus* (mouse) is shown in Fig. 13. Note that the nodes with the highest ρ_{out} can be easily highlighted in these plots. These are the regulators of the network. Furthermore, we may observe certain “branches” in the profile, where nodes with similar degrees seem to have vastly different out-assortativity values. This highlights the fact that node degree (or out-degree) alone cannot be used to determine the regulating effect of a node upon the rest of the network.

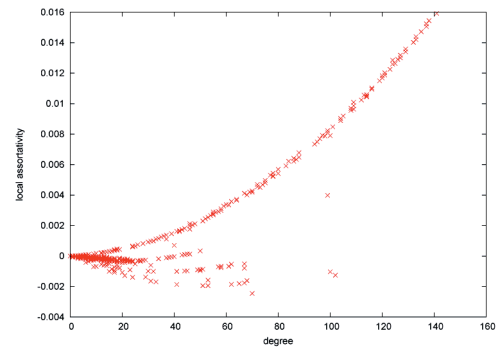


Fig. 12. *Rat* Gene Regulatory network: Scatter plot of node ρ_{out} versus out-degree. Note the several “branches” in the plot, which seem to indicate that nodes with similar degrees can have very different ρ_{out} values depending on their topological placement. The nodes at the highest branch are the ones topologically in the best position to regulate the other nodes, while the nodes in the lowest branch are in the worst position to do so. Local out-assortativity highlights this property.

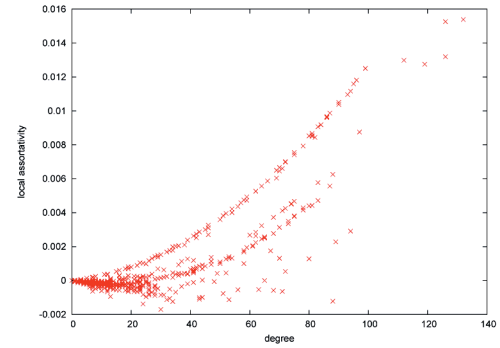


Fig. 13. *M. musculus* Gene Regulatory network: Scatter plot of node ρ_{out} versus out-degree. Again, note the several “branches” of plot indicating nodes with similar out-degree but differing local out-assortativity.

8.1 Comparing Various Local Assortativity Measures

We should note that for directed networks, the ρ_{out} and ρ_{in} profiles are most informative in understanding the network topology. Treating them as undirected networks leads to misleading impressions, while using ρ_d does not give as much information about node’s roles as ρ_{out} and ρ_{in} do. To understand these points, let us first plot the local assortativity distributions of a transcription network, treating it as undirected network. The (undirected) transcription network of *E. coli* is shown in Fig. 14.

The figure shows the local assortativity ρ versus degree distributions for the transcription network, considered as undirected network. It would seem from this profile that this network has linear local assortativity profiles, and hubs are disassortative. The mechanisms responsible for growing such networks are quite different from the mechanisms that grow networks with nonlinear profiles and assortative hubs [6]. However, as Figs. 6, 7, and 15 show, we know that this transcription network has a nonlinear assortativity profile with hubs assortative or disassortative depending on the correlations in which we are interested (out-degree or in-degree correlations). Therefore, it is clear that the (undirected) ρ distribution can be misleading when applied to directed networks, and directed local assortativity must be used.

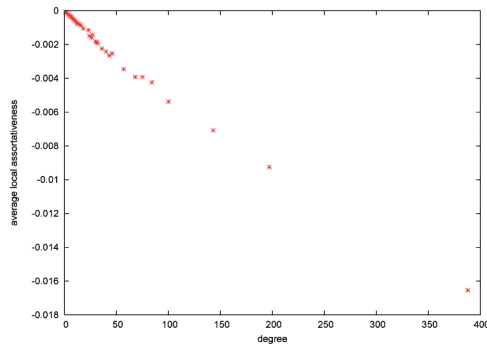


Fig. 14. Local assortativity ρ versus degree profile of *E. coli* transcription network. Here, the network was considered as undirected. Contrast with Figs. 6, 7, and 15.

Let us now look at the measure ρ_d . The distribution of this measure for *E. coli* transcription networks is given in Fig. 15. This shows nonlinear assortativity profile with assortative hubs. This measure has directedness embedded in it, but it was derived from (3). If the ρ_d of a node is high, it means that the node has high out-degree coupled with its neighbors having high in-degree, or the node has high in-degree coupled with its neighbors having high out-degree. It is a regulator surrounded by regulatees, or vice versa. Therefore, we cannot say much about the global impact of this regulator. The Fig. 15 seems to show two large hubs, one highly assortative and another slightly disassortative. In fact, the profile seems ‘split’ and moving toward these two hubs. On the other hand, when we consider ρ_{out} and ρ_{in} separately, the trend is much clearer, as is shown in Figs. 6 and 7. We have a highly disassortative in-degree hub and a highly assortative out-degree hub. Thus the quantities ρ_{out} and ρ_{in} have more utility in identifying assortative or disassortative mixing in directed networks.

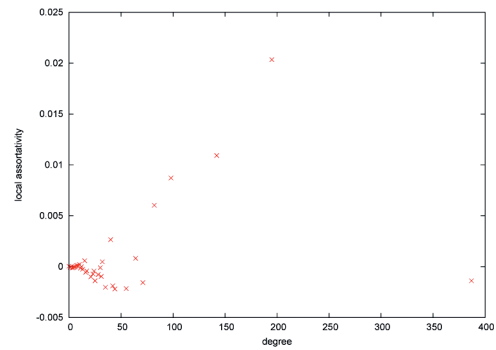


Fig. 15. Local assortativity ρ_d distribution versus degree: *E. coli* transcription network, considered directed. Contrast with the undirected case above.

In summary, we can say that local assortativity profiles of directed networks are most informative when ρ_{out} and ρ_{in} are plotted against out-degree and in-degree, respectively. Their undirected local assortativity profiles, if considered, can give misleading information about their growth mechanisms and phase of growth, while considering their ρ_d values tend to combine and confuse out-degree and in-degree trends.

8.2 Local Assortativity Profiles and Functionality of Individual Nodes

The local assortativity profiles can be used to make some interesting observations about the functionality of individual nodes in a directed biological network, and thus simulate further research about these nodes at an individual level. For example, the nodes with the highest local assortativity (both ρ_{out} and ρ_{in}) for a range of networks are given in Table 3. For comparison, the nodes with highest out-degrees and in-degrees are also listed in the table. It is

TABLE 3
A List of Biological Networks and Their Nodes with Highest ρ_{out} or ρ_{in} Values

Network name	Nodes with highest ρ_{in} [highest d_{in} in brackets]	Nodes with highest ρ_{out} [highest d_{out} in brackets]
E. coli transcription	arcA [crp]	fnr [several nodes but not fnr]
C. efficiens transcription	several nodes but not pcaR [pcaR]	all [all]
C. glutamicum transcription	arnR [glxR]	ramB [several but not ramB]
C. jeikeium transcription	all except dtxR [dtxR]	dtxR [all except dtxR]
bay wet foodweb	Respiration, water POC, raptors [raptors]	Herb. amphipods, det. amphipods, BC [pred. shrimp]
bay dry foodweb	Respiration, water POC, raptors [raptors]	Herb. amphipods, det. amphipods, BC [pred. shrimp]
chrysal c foodweb	string grey, striped anchovy, silver jelly [detritus]	BI, zooplankton, moharra [BI]
chrysal d foodweb	moharra, gold spotted killfish, silverside [detritus]	BI, bay anchovy, silverside [BI]
lower Chesapeake foodweb	pytoplankton, picoplankton, meroplankton [POC]	pytoplankton, picoplankton, meroplankton [POC]
upper Chesapeake foodweb	output, stripped bass, perch [POC]	net pytoplankton, picoplankton, ciliates [POC]
human GRN	RHOV [RHOV]	RHOA [RHOA]
house mouse GRN	LOC100046796 [LOC100046796]	Ppp3r2 [Ppp3r2]
rat GRN	LOC690845 [LOC690845]	Adcy1 [Adcy1]
A. thaliana GRN	AT5GO8100 [AT5GO8100]	AT2GO2000 [AT2GO2000]
C. elegans GRN	F23B2.13 [F23B2.13]	F23B2.13 [F23B2.13]

The nodes with highest out-degrees and in-degrees are shown in corresponding brackets for comparison. POC stands for Particulate Organic Carbon; BI stands for Benthic Invertebrates; BC stands for Benthic Crustaceans. In GRNs, the nodes are genes, while in transcription networks, the nodes are transcription factors.

apparent that the nodes with highest out-degrees or in-degrees are not necessarily the strongest regulators/regulatees in the table. This seems to be the case only among the GRNs we analyzed. Among the transcription networks and foodwebs, the node with the highest out-degree is often different from those with highest ρ_{out} . For example, in the case of *C. glaucum* transcription network, the transcription factor **arnR** is the node with the highest ρ_{out} , while the transcription factor **glxR** is the node with the lowest ρ_{out} and highest out-degree. A similar argument holds true for ρ_{in} . Therefore, local assortativity quantities can be used to gain information about node functionality that is not apparent from node degrees.

The considered examples demonstrated that local assortativity profiles of biological networks can be used to highlight interesting topological properties of individual nodes or groups of nodes.

9 CONCLUSION

In this paper, we analyzed assortative mixing in directed biological networks. We introduced new assortativity coefficients, the out-assortativity r_{out} and in-assortativity r_{in} , and showed how these can be meaningful measures in understanding network topology. Then, we extended the concept of local assortativity into directed networks, defining local out-assortativity ρ_{out} and local in-assortativity ρ_{in} . We analyzed local assortativity profiles of directed biological networks and attempted to relate out-assortativity and in-assortativity of individual nodes with their functions.

The studied directed biological networks are more assortative when in-degree mixing and out-degree mixing are considered separately, i.e., r_{in} and r_{out} are generally higher than r_d . Furthermore, out-degree mixing patterns contain the highest amount of Shannon information, suggesting that nodes with high local out-assortativity (regulators) dominate the connectivity of the network. When out-degree and in-degree mixing are considered separately, the assortativity or disassortativity of nodes especially hubs become more explicit, i.e., the nodes with relatively low ρ can have relatively high ρ_{in} or ρ_{out} and vice versa. We showed also that the local out-assortativity can be used to identify the regulators, which are most influential, since high out-degree (or in-degree) does not necessarily imply high ρ_{out} (or ρ_{in}) and vice versa. Local assortativity profiles can also be used to identify nodes and groups of nodes, which are “interestingly” placed topologically—that is, with the ability to regulate or to be regulated.

We expect that the assortative and local assortative measures, introduced in this paper, will be extensively used in analyzing directed biological networks.

APPENDIX

We set out to derive the expressions for ρ_{out} and ρ_{in} . That is, we need to analyze contributions to terms appearing in (21) and (22) for ρ_{out} and ρ_{in} , respectively.

First of all, we note that μ_q^{in} and μ_q^{out} can be equivalently defined as:

$$\mu_q^{in} = \frac{1}{M} \sum_{m=1}^M k_m^{in} = \frac{1}{M} \sum_{v=1}^N k_v^{in} k_v^{out}, \quad (48)$$

and

$$\mu_q^{out} = \frac{1}{M} \sum_{m=1}^M k_m^{out} = \frac{1}{M} \sum_{v=1}^N k_v^{out} k_v^{in}. \quad (49)$$

Note that μ_q^{in} is the “expected in-degree” when a link goes out from a node. From any node, k_v^{out} links go out, therefore, it has to be multiplied by k_v^{out} when we consider the node v . Similarly, note that μ_q^{out} is the “expected out-degree” when a link comes into a node. Any node has k_v^{in} links going into it, therefore, it has to be multiplied by k_v^{in} when we consider the node v . Therefore,

$$\mu_q^{out} = \mu_q^{in}. \quad (50)$$

However, μ_q^{out} and μ_q^{in} are not always equal.

Having defined these expected degrees, we analyze (21) for ρ_{out} . We begin by considering the contribution to its first term: $\sum_{jk} j k e_{jk}^{out}$, where j, k are out-degrees. On the one hand, it is produced via the neighbors reachable from the node:

$$\alpha_{out_1} = k^{out} \frac{k^{out} \bar{j}_{out}}{M}, \quad (51)$$

where k^{out} is the out-degree of the node considered, and \bar{j}_{out} is the average out-degree of the neighbors reachable from the node. That is, we have considered all links that depart from the node as contributing to the quantity $\sum_{jk} j k e_{jk}^{out}$.

On the other hand, we should also consider how much of all links that reach a given node contribute to it. In this case, the contribution to $\sum_{jk} j k e_{jk}^{out}$ is

$$\alpha_{out_2} = k^{out} \frac{k^{in} \bar{k}_{out}}{M}, \quad (52)$$

where k^{in} is the in-degree of node considered, and \bar{k}_{out} is the average out-degree of the neighbors from which the node can be reached.

As was done previously, we take the average of these quantities as the contribution of a given node, α_{out} , yielding

$$\alpha_{out} = k^{out} \frac{k^{out} \bar{j}_{out}}{2M} + k^{out} \frac{k^{in} \bar{k}_{out}}{2M}. \quad (53)$$

This expression captures the contribution to the term $\sum_{jk} j k e_{jk}^{out}$ in (21).

We follow by considering the contribution to the second term $\mu_q^{out} \mu_q^{out}$, obtained using (9) and (49), as follows:

$$\begin{aligned} \mu_q^{out} \mu_q^{out} &= \frac{1}{M^2} \sum_{v=1}^N k_v^{in} k_v^{out} \sum_{v=1}^N (k_v^{out})^2 \\ &= \frac{1}{M^2} \left(k_1^{in} k_1^{out} + \sum_{v=2}^N k_v^{in} k_v^{out} \right) \left((k_1^{out})^2 + \sum_{v=2}^N (k_v^{out})^2 \right). \end{aligned} \quad (54)$$

Considering a single node (without loss of generality, we choose node 1), we obtain its contribution as

$$\frac{k_1^{in} (k_1^{out})^3}{M^2} + \frac{k_1^{out} k_1^{in} \sum_{v=2}^N (k_v^{out})^2}{2M^2} + \frac{(k_1^{out})^2 \sum_{v=2}^N k_v^{out} k_1^{in}}{2M^2}. \quad (56)$$

We assume equal contribution when two nodes are involved in a term, hence division by 2. This can be further reduced to

$$\frac{1}{2M^2} \left(k_1^{out} k_1^{in} \sum_{v=1}^N (k_v^{out})^2 + (k_1^{out})^2 \sum_{v=1}^N k_v^{out} k_1^{in} \right), \quad (57)$$

yielding, for any node,

$$\beta_{out} = \frac{1}{2M} ((k_q^{out})^2 \mu_q^{out} + k_q^{in} k_q^{out} \mu_q^{out}). \quad (58)$$

Therefore, we obtain

$$\rho_{out} = \frac{\alpha_{out} - \beta_{out}}{\sigma_q^{out} \sigma_q^{out}}, \quad (59)$$

resulting in

$$\rho_{out} = \frac{k_q^{out}}{2M \sigma_q^{out} \sigma_q^{out}} (k_q^{out} (\bar{j}^{out} - \mu_q^{out}) + j_q^{in} (\bar{k}^{out} - \mu_q^{out})). \quad (60)$$

Similarly, we obtain

$$\rho_{in} = \frac{j_q^{in}}{2M \sigma_q^{in} \sigma_q^{in}} (j_q^{in} (\bar{j}^{in} - \mu_q^{in}) + k_q^{out} (\bar{k}^{in} - \mu_q^{in})). \quad (61)$$

ACKNOWLEDGMENTS

The authors thank Joseph Lizier and Dany Pradhana for motivating discussions.

REFERENCES

- [1] R.V. Solé and S. Valverde, "Information Theory of Complex Networks: On Evolution and Architectural Constraints," *Complex Networks*, Springer, 2004.
- [2] M.E. Newman, "Assortative Mixing in Networks," *Physical Rev. Letters*, vol. 89, no. 20, p. 208701, 2002.
- [3] M.E. Newman, "Mixing Patterns in Networks," *Physical Rev. E*, vol. 67, no. 2, p. 026126, 2003.
- [4] M. Piraveenan, M. Prokopenko, and A.Y. Zomaya, "Local Assortativeness in Scale-Free Networks," *Europhysics Letters*, vol. 84, no. 2, p. 28002, 2008.
- [5] M. Piraveenan, M. Prokopenko, and A.Y. Zomaya, "Classifying Complex Networks Using Unbiased Local Assortativity," *Artificial Life XII, Proc. 12th Int'l Conf. Synthesis and Simulation of Living Systems*, H. Fellermann, M. Dörr, M.M. Hanczyc, L. Ladegaard Laursen, S. Maurer, D. Merkle, P.-A. Monnard, K. Stoy, and S. Rasmussen, eds., pp. 329-336, 2010.
- [6] M. Piraveenan, M. Prokopenko, and A.Y. Zomaya, "Local Assortativity and Growth of Internet," *European Physical J. B*, vol. 70, no. 2, pp. 275-285, 2009.
- [7] "The Coryne Bacterial Transcription Networks Database," <http://www.coryneregnet.de/>, 2009.
- [8] D.J. Watts and S.H. Strogatz, "Collective Dynamics of 'Small-World' Networks," *Nature*, vol. 393, pp. 440-442, 1998.
- [9] "The Michigan Molecular Interactions Database," <http://mimi.ncibi.org/MimiWeb/main-page.js>, 2009.
- [10] "The Macaque Monkey Cortical Connectivity Data," <http://www.cocomac.org/>, 2010.
- [11] M.E. Newman, "Foodweb Data," <http://vlado.fmf.uni-lj.si/pub/networks/data/bio/foodweb/foodweb.htm/>, 2009.
- [12] M. Prokopenko, F. Boschetti, and A. Ryan, "An Information-Theoretic Primer on Complexity, Self-Organisation and Emergence," *Complexity*, vol. 15, no. 1, pp. 11-28, 2009.

- [13] M. Piraveenan, M. Prokopenko, and A.Y. Zomaya, "Assortativeness and Information in Scale-Free Networks," *European Physical J. B*, vol. 67, pp. 291-300, 2009.
- [14] M. Piraveenan, M. Prokopenko, and A. Zomaya, "Information-Cloning of Scale-Free Networks," *Proc. Ninth European Conf. Advances in Artificial Life (ECAL '07)*, 2007.
- [15] G. Bianconi, "Local Assortativeness in Scale-Free Networks," *Europhysics Letters*, vol. 81, no. 2, p. 28005, 2008.
- [16] T. Schreiber, "Measuring Information Transfer," *Physical Rev. Letters*, vol. 85, p. 461, 2000.
- [17] J.T. Lizier, M. Prokopenko, and A.Y. Zomaya, "Local Information Transfer as a Spatiotemporal Filter for Complex Systems," *Physical Rev. E*, vol. 77, no. 2, p. 026110, 2008.

Mahendra Piraveenan received the BEng degree in computer systems with a 1st class honours from the University of Adelaide, Australia, in 2004. Currently, he is a doctoral student with the School of Information Technologies, University of Sydney, Australia. After a stint in the industry including working for CSIRO as a software engineer, he commenced his PhD research in the area of biologically-inspired network design. His research interests include biological networks, bioinformatics, systems biology, graph theory, multiagent systems, and adaptive systems.

Mikhail Prokopenko received the MSc degree in applied mathematics (1988, USSR) and the PhD degree in computer science (2002, Australia). He is currently a principal research scientist at the CSIRO ICT Centre. He has a strong international reputation in the areas of complex self-organizing systems (more than 100 publications and patents). He edited a book "Advances in Applied Self-Organizing Systems" (Springer, 2008). Recently, he organised the First International Workshop on Guided Self-Organisation (GSO-08) in Sydney, followed by the Second GSO-09 Workshop held in Leipzig, Germany. He has worked on a number of international programs and organizing committees. He was a keynote speaker at the Sixth International Workshop on Agent-Based Simulation (2005); he cochaired sessions on Evolutionary and Self-Organizing Sensors, Actuators and Processing Hardware (ESOSAPH). He is an adjunct associate professor at the School of Computer Science and Engineering, University of New South Wales, Australia. He is a member of the IEEE.

Albert Y. Zomaya currently holds the chair of High Performance Computing and Networking in the School of Information Technologies at Sydney University. He is the author/coauthor of seven books and more than 300 papers, and the editor of eight books and eight conference proceedings. He serves as an associate editor for 16 leading journals. He is the recipient of the Meritorious Service Award 2000, and the Golden Core Recognition 2006, both from the IEEE Computer Society. He is a chartered engineer (CEng), a fellow of the American Association for the Advancement of Science, the IEEE, the Institution of Engineering and Technology (United Kingdom), and a distinguished engineer of the ACM.

► **For more information on this or any other computing topic, please visit our Digital Library at www.computer.org/publications/dlib.**