2-2020

Boolean Network Topologies and the Determinative Power of Nodes

Bronson W. Wacker
Portland State University

Mihaela T. Velcsov
University of Nebraska at Omaha, dvelcsov@unomaha.edu

Jim A. Rogers
University of Nebraska at Omaha, jrogers@unomaha.edu

Journal of Complex Networks: A Journal of the IMA (2020) Page 1 of 31
doi:10.1093/imaiai/drn000

Boolean Network Topologies and the Determinative Power of Nodes

BRONSON W. WACKER

Department of Mathematical Sciences, Portland State University
bwacker@pdx.edu
MIHAELA T. VELCSOV*
Department of Mathematics, University of Nebraska at Omaha, Durham Science Center 203, Omaha, NE 68130, USA
*Corresponding author: dvelcsov@unomaha.edu
AND
JIM A. ROGERS
Department of Mathematics, University of Nebraska at Omaha, Durham Science Center 203, Omaha, NE 68130, USA
jrogers@unomaha.edu

[Received on 22 September 2020]

Boolean networks have been used extensively for modeling networks whose node activity could be simplified to a binary outcome, such as *on-off*. Each node is influenced by the states of the other nodes via a logical Boolean function. The network is described by its topological properties which refer to the links between nodes, and its dynamical properties which refer to the way each node uses the information obtained from other nodes to update its state. This work explores the correlation between the information stored in the Boolean functions for each node in a property known as the determinative power and some topological properties of each node, in particular the clustering coefficient and the betweenness centrality. The determinative power of nodes is defined using concepts from information theory, in particular the mutual information. The primary motivation is to construct models of real world networks to examine if the determinative power is sensitive to any of the considered topological properties. The findings indicate that, for a homogeneous network in which all nodes obey the same threshold function under three different topologies, the determinative power can have a negative correlation with the clustering coefficient and a positive correlation with the betweenness centrality, depending on the topological properties of the network. A statistical analysis on a collection of 36 Boolean models of signal transduction networks reveals that the correlations observed in the theoretical cases are suppressed in the biological networks, thus supporting previous research results.

Keywords: Boolean network; determinative power; clustering coefficient; betweenness centrality; Pearson correlation; biological network.

2010 Math Subject Classification: 05C82, 90B10, 94A17, 90B18, 92Bxx, 92D99, 62H20

1. Introduction

A network is a collection of nodes interconnected by communication paths, or links. Boolean Networks (BN) are used for modeling networks in which the node activity, or state of the cell, can be described as a binary value: *on-off*, active-non active, 1-0, etc. This type of network has been used to examine the connections among diverse physical and engineered networks such as genetic regulatory or signal transduction networks (Kauffman [1], Shmulevich et al. [2–4], Helikar et al. [5], Kochi and Matache

[6], Conroy et al. [7], Abou-Jaoude et al. [8, 9], Mendez and Mendoza [10], Pentzien et al. [11]), or more general biological networks (Klemm and Bornholdt [12], Raeymaekers [13], Albert and Othmer [14], Saadatpour et al. [15], Correia et al. [16], Grob et al. [17], Murrugarra and Dimitrova [18]). Studying these network representations leads to predictive models of real occurrences.

The scientific community has found a broad spectrum of applicability of these models. Specific biological problems studied include cell differentiation, immune response, regulatory networks and neural networks. For cell differentiation and immune response, the basic binary element might be a chemical compound, while in neural networks it might be the state of firing of a neuron.

Although BNs are a great way to model reality in a simple way that is easy to use and understand, even such a simplification can pose challenges in assessing the dynamics of the network due to the exponential dependence of the state space on the number of nodes. One way to ease the computational burden is to reduce the network to a fairly small subset of nodes that can capture the dynamics of the whole network to a large extent. The literature offers a number of different approaches for defining and identifying the most relevant nodes in a BN.

A recent method defines relevant nodes as those with the highest determinative power (Heckel et al. [19], Matache and Matache [20]). These are considered the most powerful nodes in the network. In a paper by (Klotz et al. [21]) it is shown that canalizing Boolean functions maximize the mutual information under the same assumption as in [19], thus leading to maximal determinative power. Moreover, in [11] it is shown that a large fraction of the most powerful nodes are biologically essential genes in several BN models of biological networks and processes across a number of different organisms obtained from Cell Collective (CC, www.cellcollective.org, (Helikar et al. [22, 23])). Essential genes are those genes of an organism that are thought to be critical for its survival and are involved in crucial biological functions.

As described in [11], for a given node, the determinative power (DP) is obtained via a summation of all mutual information quantities over all nodes having the given node as a common input. The more powerful the node, the more the information gain provided by the knowledge of its state. The mutual information, as a basic concept in information theory, allows one to represent the reduction of the uncertainty or entropy of the state of a node due to the knowledge of any of its inputs. The entropy has been used in the literature to find the average mutual information of a random Boolean model of regulatory network as a way to quantify the efficiency of information propagation through the entire network (Ribeiro et al. [24]). On the other hand, the entropy of the relevant components of the network, which are comprised of nodes that eventually influence each other's state, has been used as a measure of uncertainty of the future behavior of a random state of the network (Krawitz and Shmulevich [25, 26]).

In [19] it is shown that the knowledge of the states of the most determinative nodes in the feedforward regulatory network of *E. coli* reduces the uncertainty of the overall network significantly. Similar results are observed in [20] for a model of general cell signal transduction. In [11], other models of biological processes are obtained from the Cell Collective and the most powerful nodes are identified. The authors provide a statistical analysis of the DP in relation to a number of different topological measures including in-degree, out-degree, in-closeness and out-closeness centrality, as well as betweenness centrality. They conclude that for the biological networks under consideration, the DP is mildly correlated with the out-degree, and shows no correlation with the other measures.

The goal of this paper is to explore further the relationship between the DP and two particular topological measures, namely the clustering coefficient and the betweenness centrality of nodes, under three different types of topologies and the simplified scenario of a homogeneous network where all nodes obey the same type of Boolean function. In particular we consider the following topologies which will be described in detail in the body of the paper: the Watts-Strogatz [27], the Barabási-Albert

[28, 29], and the Holme-Kim topology [30]. Similar topologies have been considered in (Hernandez et al. [31]) to test analytic results for a lower bound for the algebraic connectivity of a network. The lower bound is obtained via the betweenness centrality. In (Dzaferagic et al. [32]) it is shown that a complexity metric which quantifies the variety of structural patterns and roles of nodes in the topology of a telecommunication network is very weakly correlated to the clustering coefficient, and that the correlation decreases with an increase in network size. Our results are similar for certain network scenarios.

In general, our results show a clear negative correlation between the DP and the clustering coefficient and a positive correlation between the DP and the betweenness centrality. Certain topologies show stronger relationships, but the general result still stands. Although we consider networks in which all nodes obey the same type of dynamical rule, these results can serve as a baseline for further considerations in models of real networks.

In an effort to compare our results to those in [11], we generate a statistical analysis of 36 networks from the Cell Collective. Although different from the examples used in [11], our networks indicate a similar lack of correlation between the DP and the two topological measures under consideration.

In Section 2 we provide the basic definitions and results needed to understand the types of topologies and dynamical rules used. We provide a clear description of the Boolean networks in Section 2.1, and a description of the DP in Section 2.2. We also provide a detailed description of the topological node attributes in Section 2.3, and the layout of the topologies considered in this study in Section 2.4. The results of the research are presented in Section 3 with a comparison of the entropic features to the topological features, and a display of the levels of correlation between the entropic and topological features. In Section 4 we focus on the analysis of biological networks, while in Section 5 we summarize and discuss the results, and provide some further directions of investigation.

2. Background and Methods

This section provides a complete overview of the BN structure under consideration, including its topology and dynamical functions, together with the numerical measures whose correlations are the focus of this research.

2.1 *The Boolean Network*

There are two important aspects to consider in a BN: its topology or how the nodes are linked in the network and influence each other, and its dynamical aspects that are related to actual Boolean functions that govern the evolution of the nodes from one time step to the next. As we will see, the DP is connected to both aspects.

A BN is modelled as a set $\mathscr{N} = \{X_1, X_2, \ldots, X_N\}$ of N nodes, each node being *on* (in state 1) or *off* (in state 0). Then the vector $(X_1, X_2, \ldots X_N) \in \{0,1\}^N$ represents the state of the network. Each node X_i has an associated Boolean function $f_i : \{0,1\}^{k_i} \to \{0,1\}$ that governs the dynamics of the node as the network is iterated (evolved) from the current state to the next state. The k_i inputs of f_i represent the individual states of the nodes that influence the node X_i, also called its neighbors or links. The quantity k_i is called the connectivity, the degree, or the number of neighbors of node X_i. We use the notation $X_j \to X_i$ to indicate that node X_j is an input to node X_i (which also indicates that X_i is an output of X_j), and $\mathscr{N}_{\text{neigh}}(X_i) = \{X_j \in \mathscr{N} \setminus \{X_i\} \,|\, X_j \to X_i\}$ to denote the collection of inputs of X_i, or its neighborhood. We note that $|\mathscr{N}_{\text{neigh}}(X_i)| = k_i$ for each $X_i \in \mathscr{N}$. For the purpose of this work we will consider a non-directed network, that is a network in which if $X_j \to X_i$ then $X_i \to X_j$. In other words,

the links are bi-directional.

Furthermore, although the nodes of real networks are heterogeneous in nature since nodes are usually governed by a variety of Boolean functions, in this paper we are focusing on homogeneous networks in which all nodes obey the same Boolean function. Heterogeneous approaches will be subject for future research. As such we will simplify our approach by considering a simple threshold function for all nodes in the network. The function can be labelled as an **AVERAGE** function and is defined as

$$f(X_i) = \begin{cases} 1, & \frac{1}{k_i}\sum_j X_j > \frac{1}{2} \\ 0, & \frac{1}{k_i}\sum_j X_j \leqslant \frac{1}{2} \end{cases} \qquad X_i \in \mathscr{N} \tag{2.1}$$

where the summations are over all nodes $X_j \in \mathscr{N}_{\text{neigh}}(X_i)$. Basically, the rule means that if more than half of the inputs are *on* the node under consideration turns *on*, otherwise it turns *off*. This is an example of a Boolean linear threshold function with equal weights that is typical for neural networks. These kinds of networks were the subject of "threshold logic" in the 1960's (Hu [33]), and have been studied further, for example, by (Anthony [34]). On the other hand, it is known that biologically meaningful Boolean functions have input elements that are activators or inhibitors, which can act alone or in conjunction with other activators and/or inhibitors, as specified in [13]. A small threshold favors activation of nodes, while a large threshold favors inhibition. By keeping the threshold at $1/2$ we consider a function that does not have a built-in bias toward activation or inhibition, especially for larger connectivity.

One is usually interested in the evolution or the dynamics of the BN as it is iterated or in the long-run. Do the nodes stabilize, do they fluctuate between states, or is there no clear pattern? Observe that a BN with N nodes has 2^N possible states of the network. For small networks it may be easy to visualize the dynamics using various simulations. However, real networks such as cellular or genetic networks for example, can have hundreds or thousands of nodes, which makes dealing with the space of all possible states prohibitive. Thus, it is of interest to eliminate nodes that are not essential and focus only on smaller sub-networks of powerful or important nodes. The DP is one type of numerical measure that can be used to rank the nodes according to their importance in the network [11, 19, 20]. In [11] the authors generate a biological function analysis of nodes for some of the gene network models available in the online database Cell Collective (CC, www.cellcollective.org, [22, 23]). The analysis shows that a large fraction of the most determinative nodes are essential. Essential genes are those genes of an organism that are thought to be critical for its survival and are involved in crucial biological functions. The biological pathway analysis of the most determinative nodes performed in that paper shows that they are involved in important disease pathways.

2.2 *Determinative Power of Nodes*

The procedure for calculating the DP requires we start with some formal definitions from information theory. In Shannon's seminal work [35] we are introduced to the concept of entropy for a discrete information source. More precisely, consider two discrete random variables, X and Y. The entropy associated with a random variable is given by the following definition.

DEFINITION 2.1 Given the random variable X, the **Shannon Entropy** of X is defined as $H(X) = -E\left[\log_2 P(X)\right] = -\sum_x p_x \log_2 p_x$, where x is a value of the random variable X, $p_x = P(X = x)$, and $E\left[\log_2 P(X)\right]$ is the expected value of $\log_2 P(X)$. For a BN we have $x \in \{0,1\}$, so that formula becomes $H(X) = h(p) = -p\log_2 p - (1-p)\log_2(1-p)$ where $P(X=1) = p \Rightarrow P(X=0) = 1-p$ and $h(0) = h(1) = 0$.

DEFINITION 2.2 The **Determinative Power** of a discrete random variable X is given by $DP(X) = \sum_y MI(Y = y;X)$, where $MI(Y;X) = H(Y) - H(Y|X)$ is the mutual information of random variables X and Y and $H(Y|X)$ is the conditional entropy of Y given X.

In principle, the mutual information is a measure of the **"gain of information"**, or the DP of X over Y. The authors of [19] use the MI to introduce the DP of a node X_j of a BN as follows:

$$DP(X_j) = \sum_{i=1}^{N} MI(f_i(X);X_j). \tag{2.2}$$

which represents a summation of all "information gains" obtained from node X_j over its outputs. Note that if $X_j \notin \mathscr{N}_{\text{neigh}}(X_i)$, then $MI(f_i(X);X_j) = 0$.

In [19] the authors identify the nodes with the largest determinative power in a feedforward *E. coli* network, with the goal of finding a subnetwork whose knowledge can provide sufficient information about the entire network; in other words the entropy of the network conditional on the knowledge of that subnetwork is small enough. They show that in the *E. coli* network, one could consider a subnetwork consisting of less than half of the nodes, and that for larger subnetworks, the entropy does not improve significantly once an approximate (threshold) subnetwork size is reached. Similar results have been found in [20] for a signal transduction model in fibroblast cells, paired with a mathematical generalization of some of the results in [19] under more relaxed assumptions.

In [11], that work is extended to 36 network models available in the online database Cell Collective. For some of those 36 cellular networks it is shown that the DP is only weakly correlated to the number of outputs of a node, and it appears to be independent of other topological measures such as closeness or betweenness centrality. The work in [11] is a main trigger for this paper, since it raises the question of correlations between DP and various topological or dynamical aspects of a BN. In [11] it is noted that general low DP values is what we expect in an equilibrium situation. Prior research indicates that correlations between nodes become high only when facing a transition [36–38]. It is possible that the simple node level hierarchy coming from mutual information, in other words the DP, could benefit from a study of at least some complex graph analysis descriptors such as, among others, the betweenness centrality and the clustering coefficient of nodes. These topological measures keep track of the role played by the nodes in the system they are embedded into [39, 40].

Our focus in this paper is on the clustering coefficient, which is a local topological measure, and the betweenness centrality of nodes, which is a global topological measure. We are interested in their individual and combined impact on the magnitude of the DP values. This is done under three different types of topologies: Watts-Strogatz or small-world topology [27], Barabási-Albert or scale-free topology [28, 29], and Holme-Kim or scale-free-with-clustering topology [30]. We will describe each of them briefly.

Findings indicate, regardless of topology considered, an anti-correlation between the DP and the clustering coefficient and a positive correlation between the DP and the betweenness centrality. The strength of the correlations depends on the topology considered. There appears to be an interaction between the clustering coefficient and betweenness centrality with regards to the amount of correlation with the DP for both. This is present in the Holme-Kim topology, i.e. scale-free with tunable clustering.

2.3 *Topological Attributes*

During the 1960s and 1970s, social scientists were exploring social networks and the way information is transmitted through these networks. A property that relates information transmission and centrality

in the overall network is the betweenness centrality, first defined by Freeman in 1977 [41]. It represents the degree of which nodes stand between each other. It has been applied to a wide range of problems related to networks in biology, transport, telecommunications, social networks, scientific cooperation. For example, it has been shown that scale-free networks display a high degree of vulnerability against betweenness attacks, i.e. attacks targeting nodes with high betweenness score [42]. As observed in [31], potential applications of betweenness include the study of epidemic behaviors in the world wide web and transportation networks as in [43], which is one of the classical papers on the topic of epidemics in scale-free networks. In [31], the betweenness centrality is actually used as a mathematical tool to provide a lower bound for the algebraic connectivity of a network. The analytic results are tested on topologies similar to those used in this paper.

Consider two distinct nodes $X_j, X_k \in \mathscr{N}$. We say they are *reachable* if there is a sequential collection of links starting at X_j and ending at X_k (and viceversa given the bi-directional links) called a path between the two nodes, or *unreachable* otherwise. Note that there could be multiple paths of various lengths depending on the number of sequential links. Denote by $d(X_i, X_j)$ the *shortest path length*, i.e. the minimum number of links between node X_j and node X_k also known as a *geodesic*. Let $N_{\mathrm{sp}}(X_j, X_k)$ be the total number of distinct shortest paths between X_j and X_k, and $N_{\mathrm{sp}}^{X_i}(X_j, X_k)$ the number of shortest paths between X_j and X_k that pass through node X_i, where $X_i \neq X_j \neq X_k$. The distinct shortest paths differ in at least one node.

DEFINITION 2.3 The **betweenness centrality** of the node $X_i \in \mathscr{N}$ is the quantity

$$c_B(X_i) = \frac{2}{(N-1)(N-2)} \sum_{X_i \neq X_j \neq X_k} \frac{N_{\mathrm{sp}}^{X_i}(X_j, X_k)}{N_{\mathrm{sp}}(X_j, X_k)}. \tag{2.3}$$

Observe that due to the scaling factor $\frac{2}{(N-1)(N-2)}$ the betweenness centrality is confined to the interval $[0,1]$. We obviously consider only reachable nodes, so that $N_{\mathrm{sp}}(X_j, X_k) \neq 0$.

Our second measure reflects the degree to which nodes tend to locally cluster together, or an indication of their embeddedness in the network. More precisely, given the node $X_i \in \mathscr{N}$, we are interested in the density of links between its neighbors. One method, proposed in [27], is the probability that the neighbors of node X_i are neighbors of each other, i.e. the *clustering coefficient*.

DEFINITION 2.4 The **clustering coefficient** of the node $X_i \in \mathscr{N}$ is the quantity

$$C(X_i) = \frac{2\left|\left\{(X_j, X_k) : d(X_i, X_j) = d(X_i, X_k) = d(X_j, X_k) = 1\right\}\right|}{k_i(k_i - 1)}. \tag{2.4}$$

Observe that $C(X_i) \in [0,1]$, and it represents the proportion of links between the neighbors of X_i divided by the number of links that could possibly exist between them. In [32], the authors introduce a complexity metric (functional complexity) which quantifies the variety of structural patterns and roles of nodes in the topology of a network. They show that the complexity metric is very weakly correlated to the clustering coefficient, and that the correlation decreases with an increase in network size. We will observe similar results relating the DP and the clustering coefficient for some of our networks scenarios.

2.4 *Network Topologies and Related Parameters*

In this section we describe the three types of topology considered. The Watts-Strogatz or small-world topology accounts for high localized clustering and short path lengths observed in a variety of real networks including websites with navigation menus, electric power grids, metabolite processing networks,

networks of brain neurons, voter networks, telephone call graphs, and social influence networks. On the other hand, the Barabási-Albert or scale-free topology, accounts for the presence of hubs or highly connected nodes observed in a number of real networks, but at the same time reduces the level of clustering noticed in small-world networks. Scale-free properties have been observed in social networks, computer networks, protein-protein interaction networks, or airline networks. Finally, the Holme-Kim or scale-free-with-clustering topology combines the idea of hubs and high clustering, which was observed in a number of social networks, computer networks, metabolic networks of certain organisms, or e-mail networks with e-mail addresses as nodes and e-mail messages as links.

2.4.1 *Watts-Strogatz.* Watts and Strogatz [27] measured that many real-world networks have a small average shortest path length, but also a clustering coefficient significantly higher than expected by random chance. They developed a model that relies on rewiring a predefined ring network multiple times. The basic construction of a ring network is to place all available nodes in a circle, so that each node has a neighbor to the left and a neighbor to the right, and place links connecting the nodes into a ring structure. The number of neighbors, or the connectivity is fixed and would be $k = 2$ for this example. After the multiple random rewiring of links, the resulting network is a random network with localized clustering. For more details see for example [27, 44].

2.4.2 *Barabási-Albert.* Moving away from randomly distributed networks we come to other real world phenomena: *scale-free networks* [28, 29]. In the late 1990s, Barabási and Albert observed network properties that were not reflected in random networks. This property is the presence of hubs, or highly connect nodes, i.e. very high degree nodes. The process of attaining a network that contains hubs involves a dynamical construction of the network, rather than placing links between pairs of nodes. We start with an arbitrary network of m_0 nodes, with no specific topology other than the criteria that each node has at least one link. The network is updated, at each time step, by adding a new node with $m \leqslant m_0$ available links. The selection process for the available links is as follows: going through the total nodes in the network until the m links are assigned, the probability that a link of the new node connects to node X_i depends strictly on the weighted degree of node X_i, i.e. $\Pi(k_i) = k_i/\sum_j k_j$, which is known as the *preferential attachment* probability. Nodes with higher degrees have more likely a chance of acquiring a link.

2.4.3 *Holme-Kim.* Authors in [30] noted that the Watts-Strogatz topology and the Barabási-Albert topology had two exclusive features that were present in real world networks: (1) Watts-Strogatz has higher clustering, i.e. a node's neighbors are more likely to be neighbors themselves, and (2) Barabási-Albert has the scale-free distribution that yields hubs. In an effort to construct model networks that better represent real-world networks, the authors decided to modify the creation process of the Barabási-Albert scale-free networks to *include* more clustering. Using the ideas from Section 2.4.2, we start with an arbitrary network of m_0 nodes with each node having at least one link. Each time $t > 0$ a new node arrives with $m \leqslant m_0$ links. The m links are connected to the already present nodes by way of preferential attachment. Holme-Kim proposes the following modification to the scale-free network construction: *triad formation*. In the preferential attachment phase, after one of the m links is attached to a node X_j already present in the network, connect one of the $m-1$ remaining links to a neighbor of node X_j. If all neighbors of node X_j are already connected to the incoming node, preferentially attach to the network.

As indicated in [30], the addition of clustering into the scale-free network does not change the degree-distribution. So, we note that the scale-free topology with added clustering has the same degree

Topologies	Barabási-Albert	Watts-Strogatz	Holme-Kim
Parameters	Description and ranges of values		
N	number of nodes $\{15,16,\dots 24\}$	number of nodes $\{15,16,\dots 24\}$	number of nodes $\{15,16,\dots 23\}$
p	N/A	rewiring probability $\{0.01,0.03,0.05,0.07,0.1\}$ $\cup\{0.5,0.6,0.7,0.9,0.9\}$	triangle formation probability $\{0.1,0.2,0.3,0.4,0.5\}$
m	number of links new node brings $\{1,2,3,4,5\}$	N/A	number of links new node brings $\{2,3,4,5\}$
k_{input}	N/A	number of initial neighbors for given node 2	N/A
$F=\{f_i\}_{i=1}^{N}$	set of Boolean functions f_i as in (2.1)	set of Boolean functions f_i as in (2.1)	set of Boolean functions f_i as in (2.1)
s	set of seeds for random number generator 1000 seeds	set of seeds for random number generator 1000 seeds	set of seeds for random number generator 1000 seeds

Table 1. Model parameters for network construction. The above description, for each parameter, is related to the topology that the parameter is coupled to. The values of the parameters are indicated as well.

distribution as the Barabási-Albert topology.

2.5 *Parameters*

We describe the ranges of values of the parameters used in simulations and summarize them in Table (2.5). The choices are made based on the computational efforts associated with network construction, finding the topological attributes, and in conjunction with the goal of exploring how sensitive the entropic information is to additional links in the network.

All networks explore a similar range of values for the network size N. Due to computational complexity associated with hubs in the scale free networks, the range of node numbers under consideration is $15 \leqslant N \leqslant 24$.

Within the Watts-Strogatz topology, the rewiring probability is considered for the *small-world* domain, i.e. vanishingly small rewiring probability, and higher rewiring probabilities $p \in \{0.001,0.03,0.05,0.07,0.1\}$ $\cup\{0.5,0.6,0.7,0.8,0.9\}$. The number of initial neighbors is fixed, $k_{\text{input}} = 2$, similar to the original Watts-Strogatz model.

Within the Barabási-Albert topology, the number of links a new node brings to the network is $m \in \{1,2,3,4,5\}$.

For the Holme-Kim topology, we are blending the number of incoming links and the possibility of rewiring (with the intent of triangle formation). These parameters allow us to adjust the clustering of the scale-free network. Since we are concerned with rewiring probabilities with incoming links, an

incoming node with $m = 1$ link would have no clustering impact, so it can be discounted in the analysis. Therefore, $m \in \{2,3,4,5\}$. The probability of rewiring the remaining incoming links to form a triangle is $p \in \{0.1,0.2,0.3,0.4,0.5\}$.

Since all networks are constructed using the Python package Networkx and model networks are randomly constructed with the use of a random number generator, a fixed set of 1000 random number generator seeds is used in the construction of the model networks. In particular, networks are examined so all nodes have degree $k \geqslant 1$, i.e. we consider connected networks. The seed that yields connected networks is noted and, if necessary, used for future network construction to maintain a high quality of data.

2.6 *Pearson Correlation Coefficient*

We recall briefly that given two random variables, U and V, their Pearson correlation coefficient (PCC) is defined as

$$\rho = \frac{E[(U - E[U])(V - E[V])]}{\sqrt{E[(U - E[U])^2]}\sqrt{E[(V - E[V])^2]}} \tag{2.5}$$

where $E[U], E[V]$ are the expected values of the two random variables. It is known that $\rho \in [-1,1]$, and that: (1) U,V are anti-correlated if $\rho \in [-1,0)$; (2) U,V are not correlated if $\rho = 0$; (3) U,V are correlated if $\rho \in (0,1]$. We estimate this coefficient from the actual data obtained from simulated networks, $u_i, v_i, i = 1,2,\ldots,N$, thus generating the sample correlation coefficient r.

We determine the level of correlation between the following pairs of random variables from the data: (1) $(U,V) = (C(X_i), DP(X_i))$ and (2) $(U,V) = (c_B(X_i), DP(X_i))$. This will be done for the considered topologies in an effort to disentangle which topological features have the biggest impact on the entropic property known as the DP.

3. Results

For each topology with a given set of parameters, we compute the DP, the clustering coefficient, and the betweenness centrality for each node of the network. First we focus on graphical representations of the relationship between the DP values and the topological attributes. Then we focus on assessing the degree of correlation between these measures. We discuss the results by the types of topology under consideration. For illustration purposes, we focus on four network sizes, namely $N = 15,19,20,23$ since all the other plots are similar.

3.1 *DP versus Clustering Coefficient*

Watts-Strogatz: Figure 1 represents the graphical representation of the relationship between the DP and the clustering coefficient for the Watts-Strogatz topology. The different subplots correspond to the different network sizes. Notice the striking similarity of the graphs, which indicates that the networks size is not of significance. A quick examination for various rewiring probabilities p shows reduced ranges of values for the DP and the correlation coefficients expressed by the sparsity of the graphs. Looking across the graphs, a trend is highly present. For zero or low clustering, the DP has both a larger range of values and a larger maximum value. On the opposite end, the nodes that have the highest level of clustering, $C(X_i) = 1$, have both the smallest range of possible DP values and the smallest maximum DP values.

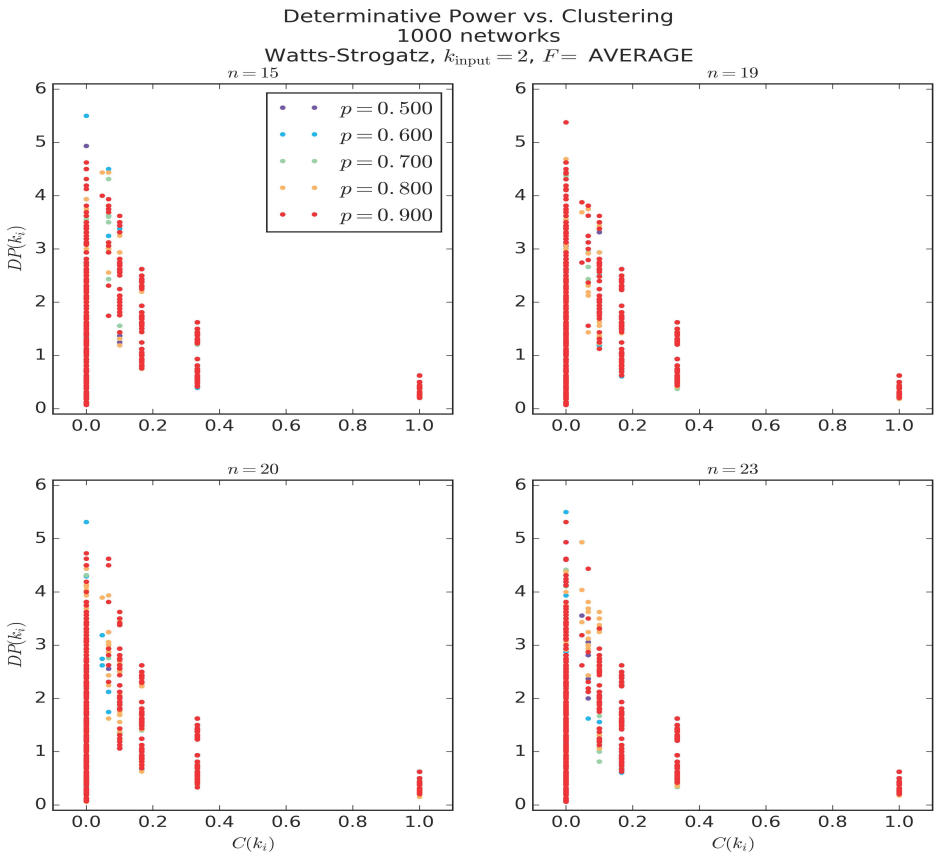

FIG. 1. DP vs. clustering coefficient for the Watts-Strogatz topology. Notice the similarities between plots for different network sizes, and the decreasing trend of DP with increased values of the clustering coefficient. The graphs are quite sparse.

Barabási-Albert: Figure 2 demonstrates the relationship for the Barabási-Albert topology. We see a clear spread of DP and clustering values in contrast to the sparsity of the graphs in Figure 1. As expected, however, for $m = 1$, meaning that the number of links a new node brings into the network is 1, there is zero clustering *for all* nodes in the network. The absence of clustering clearly demonstrates the same phenomena observed for the small-world networks: a larger range of DP values paired with the highest maximum DP values across all clustering coefficient values for all values of m. By increasing the number of incoming links, $m = 2,3,4,5$ we can observe a larger possible selection of clustering values, rather than the small set observed with the small-world topology. The nodes with the highest clustering value still have the lowest overall DP values, but, as we increase the number of incoming links, the clustering value that corresponds to the highest DP actually increases. For $m = 2$, the clustering coefficient that corresponds to the highest DP is about $C(X_i) \approx 0.15$; for $m = 3$, $C(X_i) \approx 0.22$; for $m = 4$, $C(X_i) \approx 0.32$;

and for $m = 5$, $C(X_i) \approx 0.4$. There is an overall trend for overall lower DP values as we increase m. We notice again that the network size has little impact on the observations.

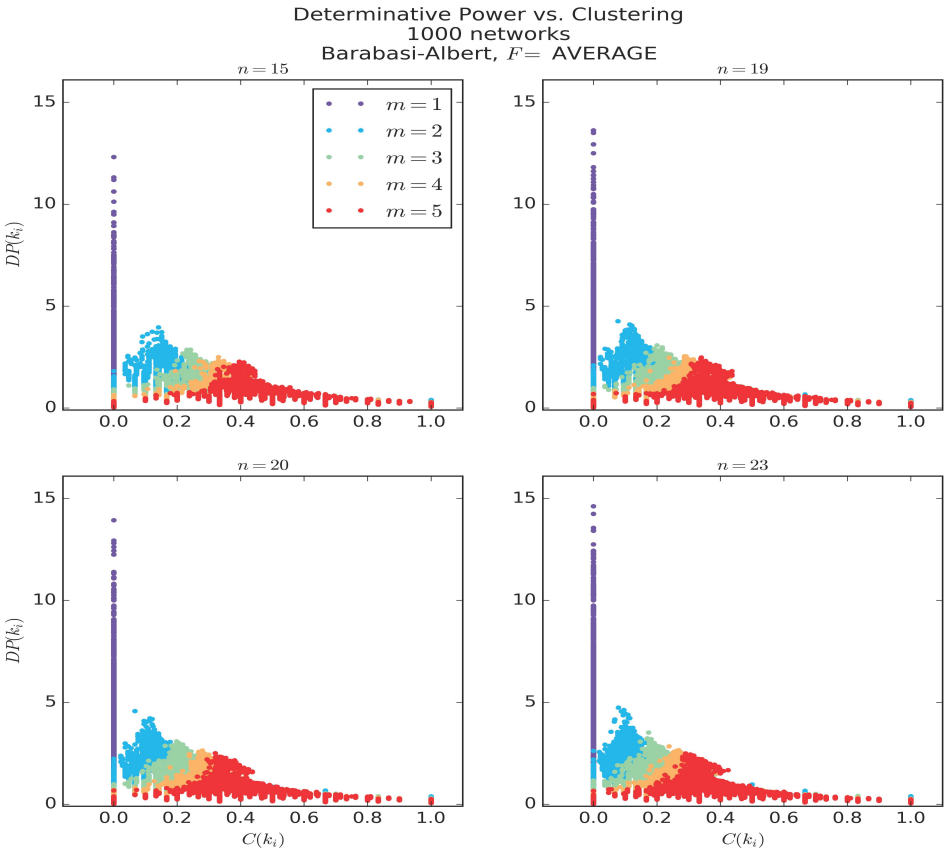

FIG. 2. DP vs. clustering coefficient for the Barabási-Albert topology. Notice the similarities between plots for different network sizes, and the overall decreasing trend of DP with increased values of the clustering coefficient or increased values of the number of links a new node brings into the network, m.

Holme-Kim: Figure 3 is representative for the Holme-Kim topology. There is still a wide range of clustering values, as was present in the scale-free topology. The graphs are similar to those for the Barabási-Albert topology, except for the case of zero clustering, where the range of DP values is significantly smaller. Looking through the various network sizes for $m = 2$ and $p = 0.1$ we can observe that the maximum DP, for fixed clustering, is increasing with network size, but only slightly. The increase becomes more observationally small as we increase the number of incoming links, $m = 3, 4, 5$. However, the trend of overall decreasing DP for increasing m holds. There is a remarkable amount of similarity in the trends as we increase the probability p of triangle formation, therefore we choose to

include only one sample figure for $p = 0.1$.

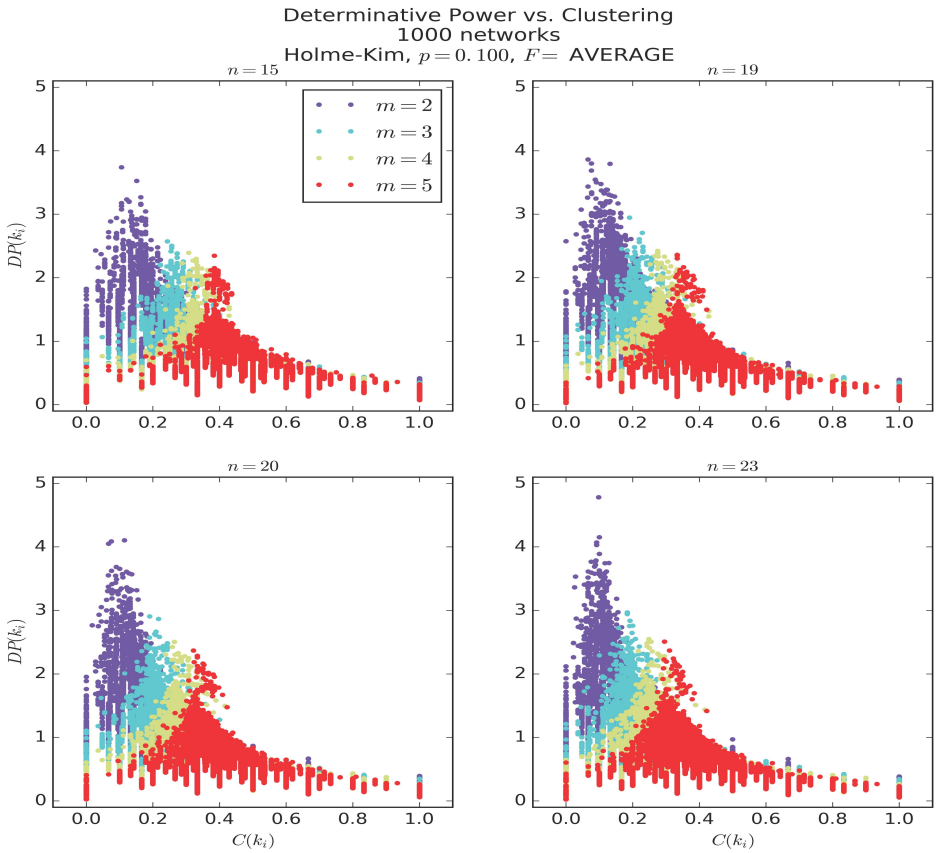

FIG. 3. DP vs. Clustering for the Holme-Kim topology with $p = 0.1$. Notice again the similarities between plots for different network sizes, and the overall decreasing trend of DP with increased values of the clustering coefficient or increased values of the number of links a new node brings into the network, m. For a fixed clustering value, the maximum DP tends to increase slightly with increased network size.

3.2 *DP versus Betweenness Centrality*

After considering the clustering coefficient as the local topological feature, we focus on the global topological feature of betweenness centrality.

Watts-Strogatz: Figure 4 shows the results for the Watts-Strogatz topology. An examination of the relationship between the DP and the betweenness centrality for this topology shows a fairly large swath of DP values for basically all betweenness centrality values. We notice again that the network size does not induce qualitative differences between plots. It appears that a large number of DP values are in the

$(0,2)$ range, regardless of the rewiring parameter p. No particular trends are observed.

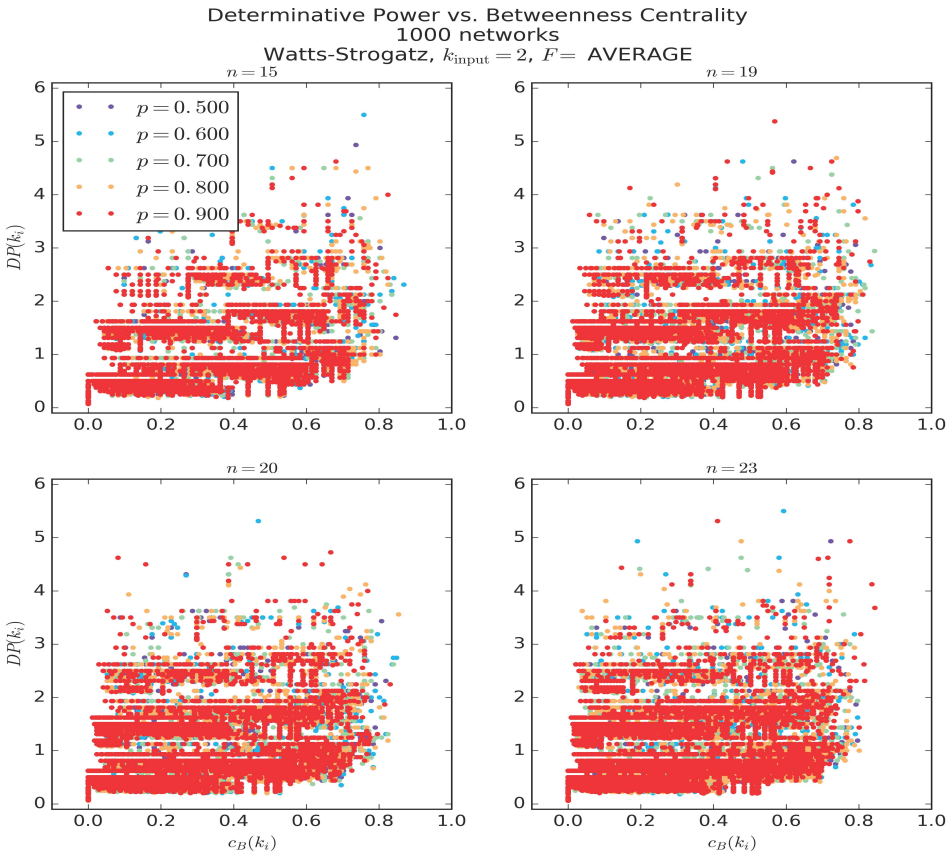

FIG. 4. DP vs. betweenness centrality for the Watts-Strogatz topology. Notice the similarities between plots for different network sizes, and that the DP values are located mostly in the interval $(0,2)$ regardless of the betweenness centrality values.

Barabási-Albert: Figure 5 shows the results for the Barabási-Albert topology. We notice a clear difference. First up is the $m = 1$ or new nodes with a single incoming link. The data set shows a clear preference for higher DP paired with higher betweenness centralities. This is independent of the network size. Once more incoming links are added, $m = 2,3,4,5$, the betweenness centrality decreases considerably, and this translates into smaller DP values. The relationship between DP and betweenness centrality appears to be linear for the larger values of incoming link number, m.

Holme-Kim: Figure 6 shows the results for the Holme-Kim topology with $p = 0.1$. The various plots, for triangle formation probability $p \in [0.1, 0.5]$ (not shown), exhibit very similar trends, in particular when we compare them to the results for the Barabási-Albert topology for $m = 2,3,4,5$. As the number

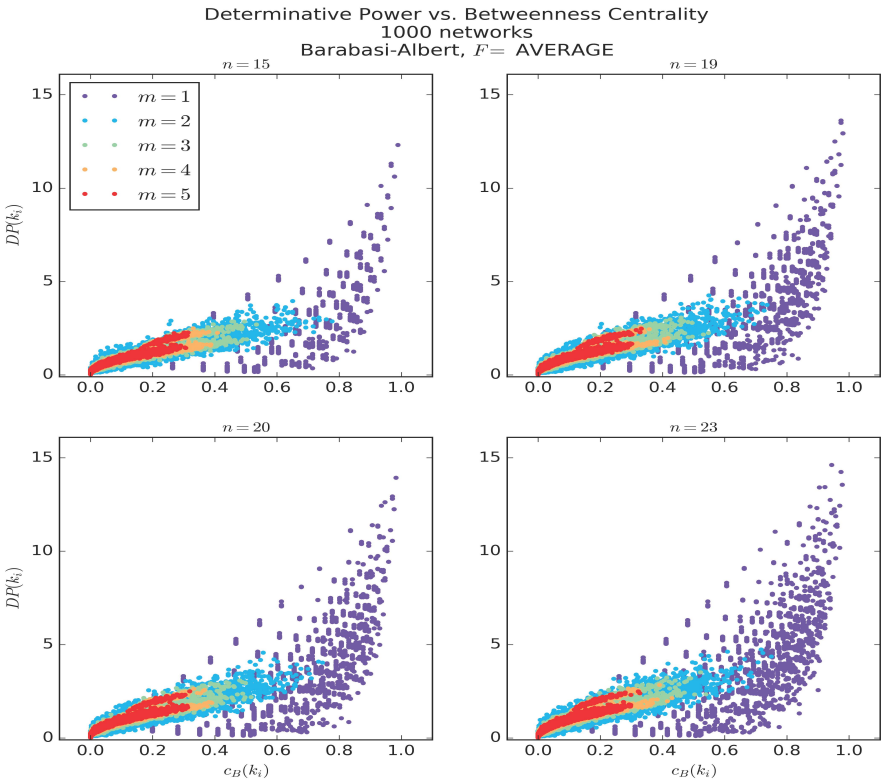

FIG. 5. DP vs. betweenness centrality for the Barabási-Albert topology. Notice the similarities between plots for different network sizes. However, the case $m = 1$ is very different than the other values of m with a wider range of values for the DP and the betweenness centrality, as well as an apparently linear relationship for larger values of m. In those cases the ranges of values decrease with increased m.

of incoming available links m increases, the overall betweenness centrality range of values decreases, as does the range for the DP. We, again, see a linear trend for the relationship between the DP and the betweenness centrality.

3.3 *DP versus Clustering Coefficient and Betweenness Centrality*

In an effort to get a better understanding of the observed trends we use a Python interpolation algorithm, matplotlib.mlab.griddata[1], to construct a color or heat map associated with the data. The goal is to see if the data *collects* in any areas, in particular, what regions of clustering *and* betweenness centrality are most likely to have the highest DP given the configuration of parameters associated with the topology. The horizontal axis represents the betweenness centrality, the vertical axis represents the clustering

[1] https://matplotlib.org/api/mlab_api.html#matplotlib.mlab.griddata

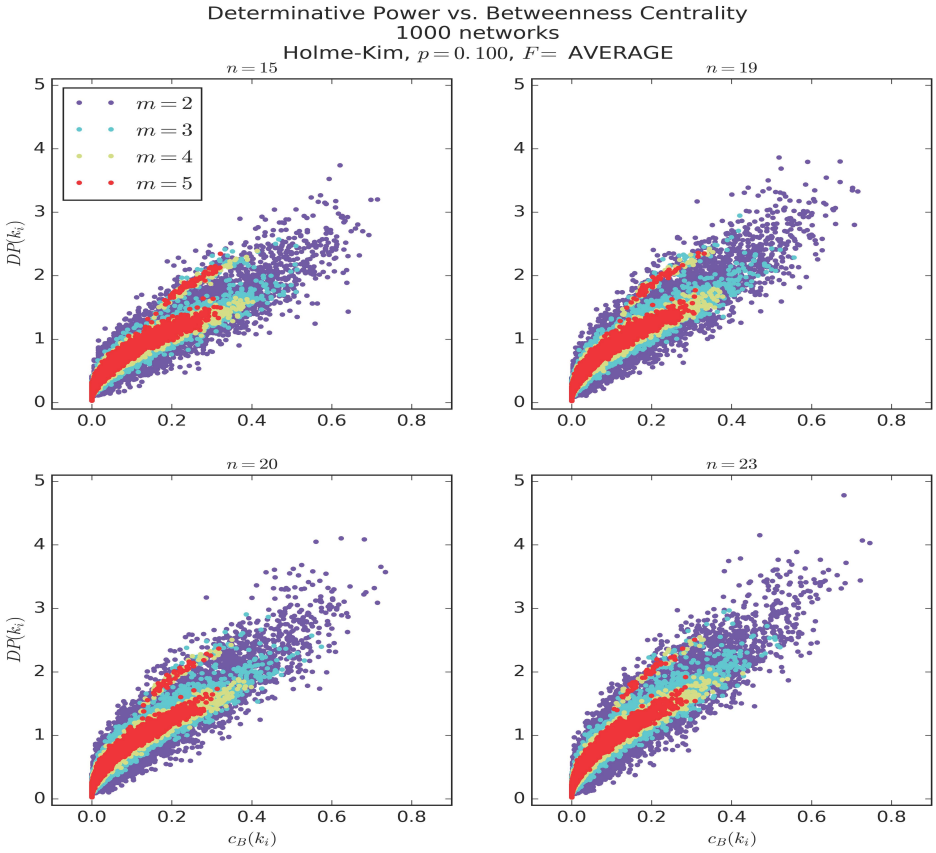

FIG. 6. DP vs. betweenness centrality for the Holme-Kim topology with $p = 0.1$. There is a clear similarity between these plots and those in Figure 5 for the Barabási-Albert topology with $m > 1$. The conclusions are similar.

coefficient, while the DP values are represented with colors ranging from dark blue (low DP) to red (high DP). The plots are obtained for multiple combinations of network size and other parameters.

Watts-Strogatz: Figure 7 shows the color maps for the Watts-Strogatz topology with the indicated parameters. Looking at the plots we see some clear regions with high DP values. For example, for $N = 15, p = 0.5$ (top left plot), we observe two peaked regions around

$$(c_B(X_i), C(X_i)) \in \{(0.4, 0.1), (0.72, 0.09)\}.$$

The clustering coefficients are similar but with different betweenness centralities. Keeping the rewiring probability p fixed and increasing the total number of network nodes we note that the regions for maximal DP have a fairly constant clustering, but, the betweenness centrality peaks at various and multiple locations. On the other hand, fixing the number of nodes, $N = 15$, and increasing the rewiring probability, from $p = 0.5$ to $p = 0.9$, the clustering region remains fairly fixed. However, the global maximum

peaks for betweenness centrality fluctuate in no discernible pattern. For $N > 15$ we see similar trends for consistent regions of higher DP associated with reduced clustering coefficient values, but varying peaks over the entire range of betweenness centrality values.

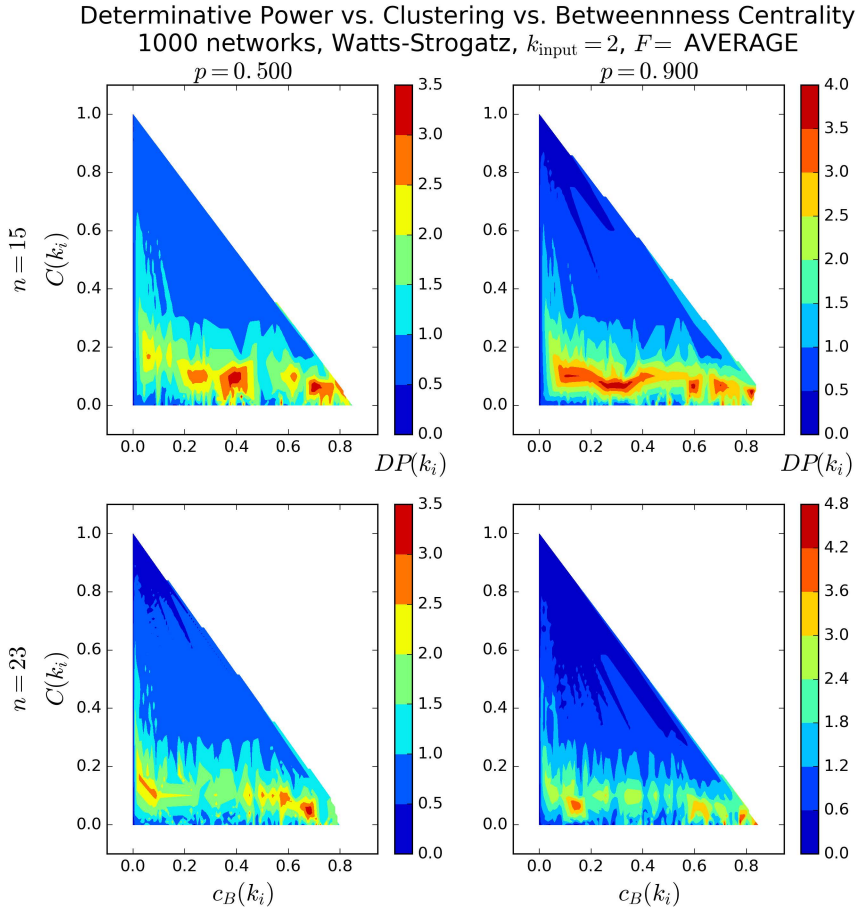

FIG. 7. DP vs. clustering vs. betweenness centrality for the Watts-Strogatz topology. Notice that high DP values are confined to the smaller clustering coefficient region at the bottom of the triangular heat maps, and they are spread over the entire range of betweenness centrality values.

Barabási-Albert: Figure 8 shows the color maps for the Barabási-Albert topology. In direct contrast to the small-worlds color maps, we find much more well-defined global maximum regions for the DP. First we note that for $m = 1$ the clustering is constant, i.e. $C(X_i) = 0$ for all nodes. Attempting to interpolate this was not possible, so the results stand for $m = 2, 3, 4, 5$. The global and local maximum DP values huddle around a peak associated with a particular region of clustering and betweenness centrality. For a fixed number of incoming links, m, the heat map, for increasing network size, is essentially the same, supporting the previous observations that the network size does not seem to have a qualitative impact and a rather insignificant quantitative impact. The highest DP values correspond to low clustering

and high betweenness centrality, for $m = 2$. As we increase the number of incoming links a new node brings, m, the slice of clustering-betweenness centrality space adjusts. As we go from $m = 2$ to $m = 5$, the clustering region with the highest DP increases in area. That means the high DP values are obtained in more cases. We also note the decreasing betweenness centrality for the higher DP. This corresponds to previous observations regarding the decrease of the betweenness centrality and DP ranges observed in Figure 5.

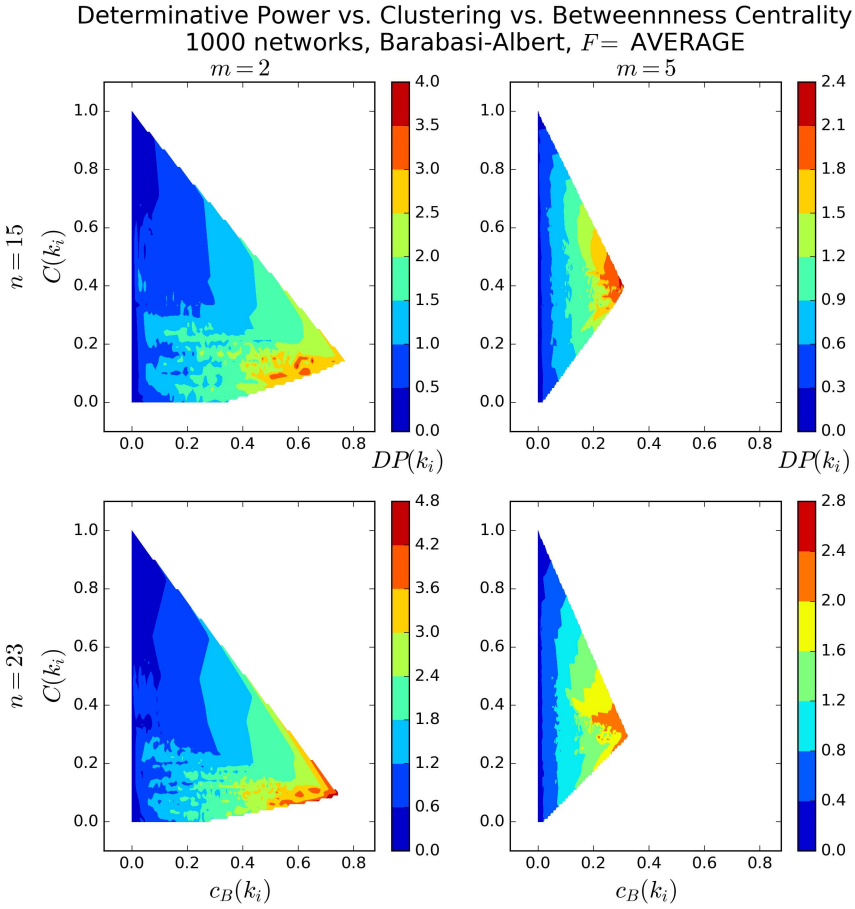

FIG. 8. DP vs. clustering vs. betweenness centrality for the Barabási-Albert topology. The network size has little to no impact, while the increase in the parameter m induces an overall reduction of values for the betweenness centrality associated with the high DP values. The peaks of the triangles are the areas of high DP.

Holme-Kim: Figure 9 shows the color maps for the Holme-Kim topology. With this topology we get essentially the same results as with the Barabási-Albert topology, regardless of increasing triangle formation probability $p \in \{0.1, 0.2, 0.3, 0.4, 0.5\}$. The overall results display a fairly specific region of high DP that has a lower clustering and high betweenness centrality, for low m. As m is increased, the region with high DP has an increasing clustering and a decreasing betweenness centrality. As the net-

work size is increased, the maximal DP region compresses slightly. This also holds true when increasing the triangle formation probability p.

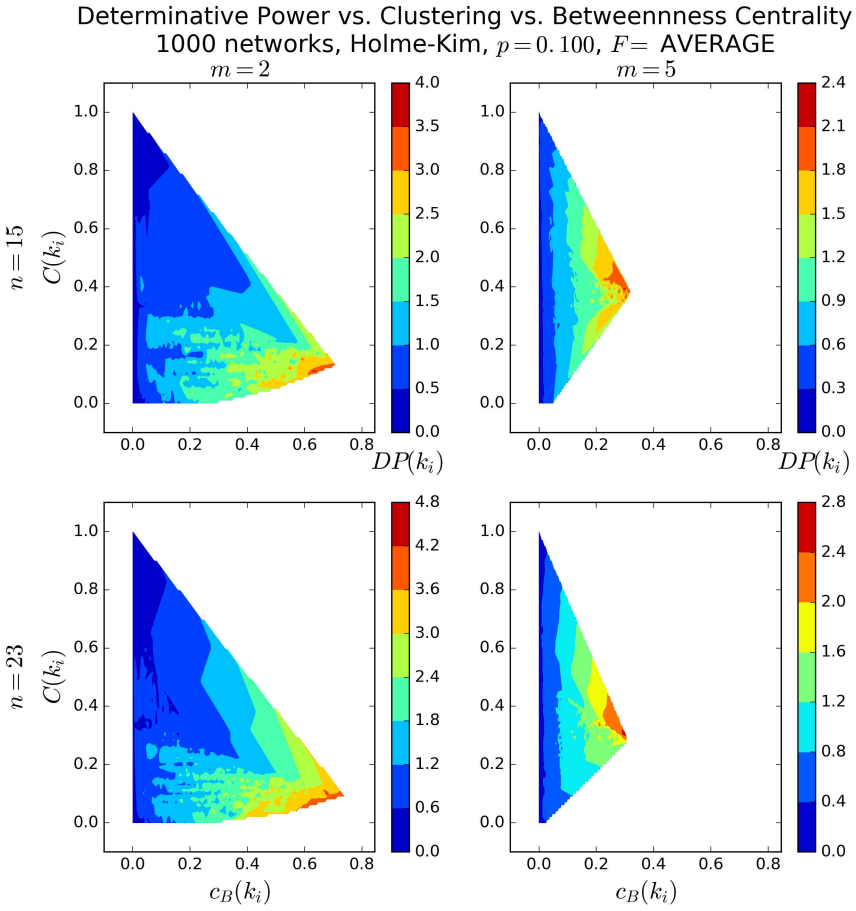

FIG. 9. DP vs. clustering vs. betweenness centrality for the Holme-Kim topology with $p = 0.1$. The results are similar to those in Figure 8.

3.4 *Correlation Analysis*

To quantify more clearly the trends observed in the previous sections, we determine the level of correlation of DP with the topological measures, that is, we find the sample correlation coefficient r for the PCC given by formula (2.5). We compute the PCC for the following pairs of random variables from the data $(C(X_i), DP(X_i))$ and $(c_B(X_i), DP(X_i))$, for $i = 1, 2, \ldots N$. This is done for all three topologies under consideration. Recall that U, V are anti-correlated if $r \in [-1, 0)$; (2) U, V are not correlated if $r = 0$; (3) U, V are correlated if $r \in (0, 1]$. The error bars correspond to a 95% confidence interval for the PCC.

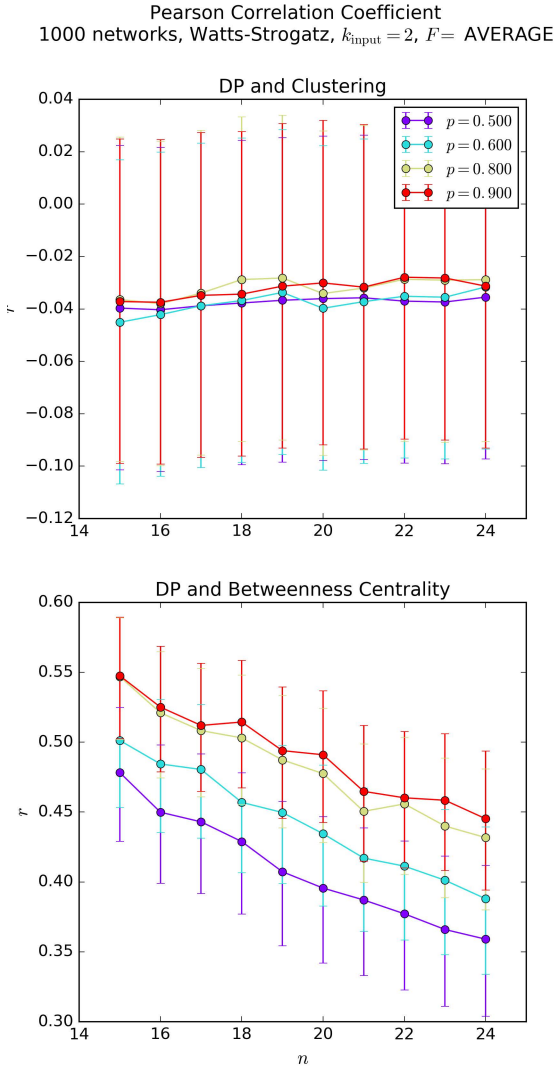

FIG. 10. PCC for the Watts-Strogatz topology. Top graph: PCC for DP and the correlation coefficient. Notice the negative values that indicate anti-correlation. However, the PCC values are very close to zero, which means the correlation is insignificant, and it becomes even weaker with increased network size. Bottom graph: PCC for DP and the betweenness centrality. The PCC values indicate a positive correlation, however it is not very strong. The correlation is stronger for higher values of the rewiring probability p and weakens with increased network size.

Watts-Strogatz: Figure 10 shows us the results of the analysis for the Watts-Strogatz topology. The top plot is the PCC for the DP and the clustering coefficient, shown as a function of the number of network nodes. We immediately see that the PCC is negative, indicating anti-correlation, which supports the observations in Figure 1. The level of anti-correlation, for all rewiring probabilities $p \in \{0.5, 0.6, 0.7, 0.8, 0.9\}$, falls in a narrow range of values: $-0.045 \leqslant r \leqslant -0.028$. A PCC very close to 0 paired with error bars that extend to positive values indicate that there is not a significant evidence for anti-correlation. Moreover, the level of anti-correlation is approaching zero as the number of network nodes increases.

The bottom plot of Figure 10 is the PCC for the DP and the betweenness centrality, also shown as a function of the number of network nodes. We see only positive values for the PCC, indicating a positive correlation. Moreover, $0.35 \leqslant r \leqslant 0.55$, showing a stronger level of correlation in comparison to the level of anti-correlation displayed by the PCC for the DP and the clustering coefficient. However, even this level of correlation is not particularly strong. There are clear differences when considering the PCCs for specific rewiring probabilities p. The higher the rewiring probability the higher the PCC, overall, regardless of the number of network nodes, i.e. for $p_1 < p_2$ we have $r_{p_1}(N) < r_{p_2}(N)$ for each N. An interesting observation is that for fixed p, the PCC is decreasing as a function of N, i.e. the overall level of correlation is weakening as we increase the number of nodes in the network.

Looking at Figure 4, we can see that for each fixed betweenness centrality, we have a corresponding large range of DP values. Moreover, for each fixed DP value, we could have low betweenness centrality or high betweenness centrality. The initial reaction to Figure 4 is that knowing anything about either the betweenness centrality or the DP will not tell us much about the other. The bottom plot of Figure 10 tells us that, as the network increases in size, the PCC for the betweenness centrality and the DP also decreases.

Barabási-Albert: Next we examine the PCC for the Barabási-Albert topology. Figure 11 shows the results: the top plot is the PCC for the DP and the clustering coefficient and the bottom plot is the PCC for the DP and the betweenness centrality. Due to the constant levels of the clustering coefficient for the $m = 1$ case, as can be seen in Figure 2, we omit the calculation. Looking at the PCC for the DP and the clustering coefficient, we immediately see the negative PCC overall, indicating anti-correlation, as observed also in Figure 2. Moreover, compared to the Watts-Strogatz PCC for the DP and the clustering coefficient, it is a stronger level of anti-correlation since $-0.45 < r < -0.35$. However, the values are still far from a very strong anti-correlation for which the PCC would be close to -1. We also note that as we increase m the overall PCC is decreasing, i.e. $m_1 < m_2$ implies that $r_{m_1} > r_{m_2}$ so the anti-correlation is strengthening as each node brings more links into the Barabási-Albert network construction. An increase in the network size does not generate significant differences, which is easy to see in Figure 2 as well.

Next we look at the bottom plot in Figure 11 which shows the PCC for the DP and the betweenness centrality. The PCC shows a very strong level of correlation since $0.935 < r < 0.955$. Examining the PCC as a function of m, we see that as m increases the overall PCC increases, i.e. $m_1 < m_2$ leads to $r_{m_1} < r_{m_2}$, which means that the larger the m the stronger the linear relationship between DP and betweenness centrality as can be seen in Figure 5. Notice again that the network size does not have a significant impact. This suggests that for larger networks we could still expect clear correlations of the DP and the betweenness centrality for this type of topology.

Holme-Kim: The Holme-Kim topology takes the Barabási-Albert topology and tweaks it by assigning a probability to extra incoming links for triangle formation in order to increase the amount of clustering present in a scale-free topology. Figure 12 shows the results of the various parameter combinations

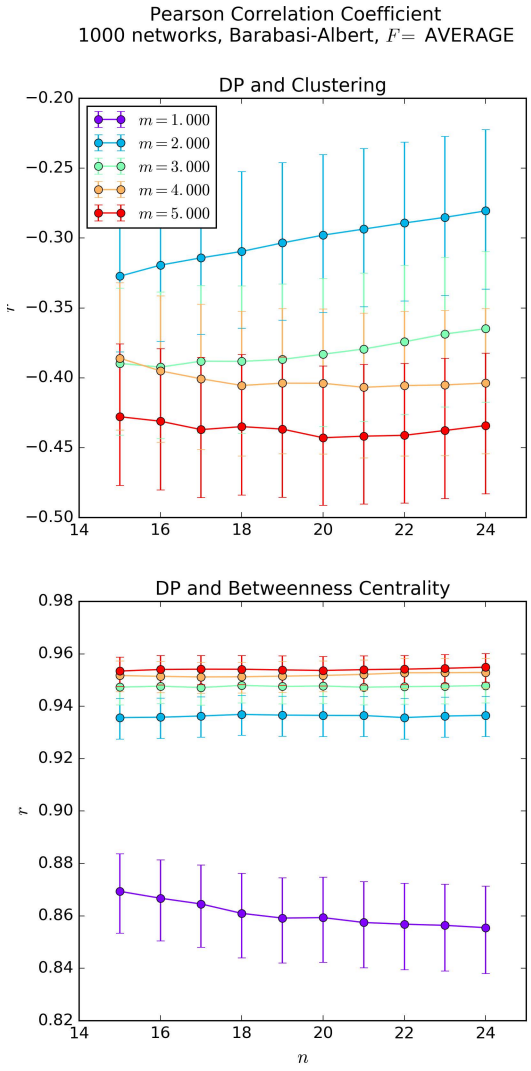

FIG. 11. PCC for the Barabási-Albert topology. Top graph: PCC for DP and the correlation coefficient. Notice the negative values that indicate anti-correlation. Like for the Watts-Strogatz topology, the PCC values are very close to zero, which means the anti-correlation is insignificant. However, there is a clear decrease of PCC values (meaning a stronger anti-correlation) with increased values of the number of new links m. Bottom graph: PCC for DP and the betweenness centrality. The PCC values indicate a strong positive correlation, which increases with an increase of m. However, an increase in network size has no significant impact on the two graphs.

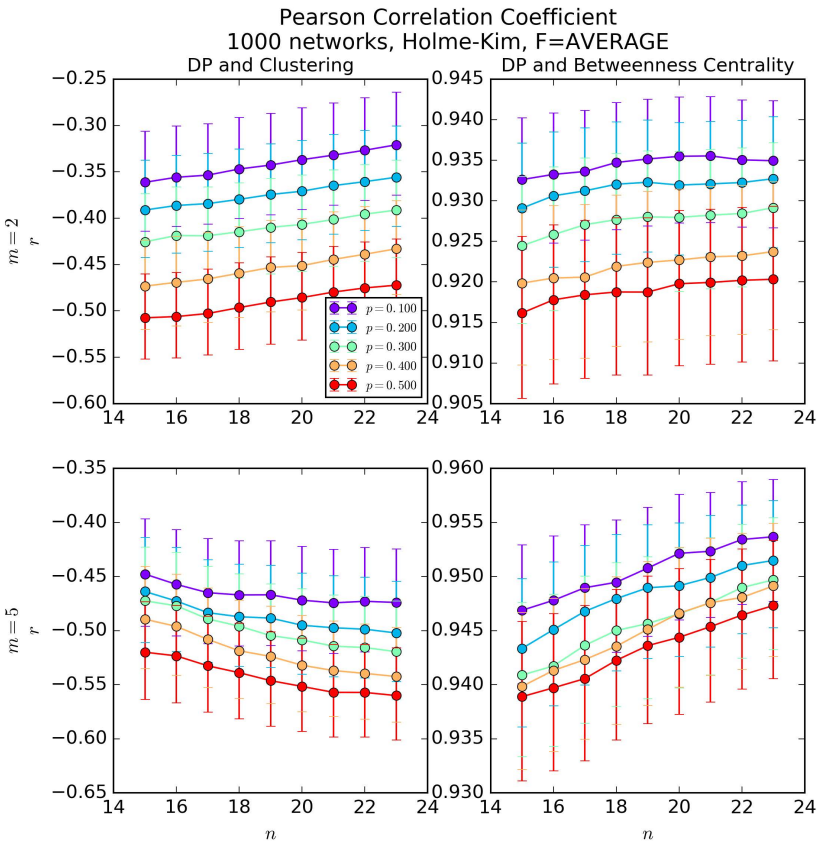

FIG. 12. PCC for the Holme-Kim topology. Left column: PCC for DP and the correlation coefficient. Notice the negative values that indicate (a weak) anti-correlation, regardless of the parameters p or m. Right column: PCC for DP and the betweenness centrality. The PCC values indicate a strong positive correlation, which increases with an increase of m, but decreases with an increased probability p.

of the incoming link number m and triangle formation probability p. The left column of plots is the PCC for the DP and the clustering coefficient and the right column is the PCC for the DP and the betweenness centrality. Starting from the top, the rows correspond to increasing the incoming link number m. Starting with the first column the weak trends of anti-correlation for the DP and clustering coefficient are maintained. For a fixed triangle formation probability p, scanning through the plots for increasing values of m we see the decreasing PCC that is displayed in Figure 11 for the Barabási-Albert topology. More incoming links lead to a stronger amount of anti-correlation between the DP and the clustering coefficient. Moreover, for a fixed m, we see a clear trend of decreasing PCC values for an increasing triangle formation probability p, i.e. anti-correlation gets stronger. For a higher number of incoming links m, this higher level of anti-correlation is more prominent in networks with a larger number of nodes. The network size has a mild increasing effect for small m and a decreasing effect for larger values of m.

Next we consider the right column which displays the PCC for the DP and the betweenness centrality

for the Holme-Kim topology. As with the Watts-Strogatz and Barabási-Albert topologies, we have positive PCC values indicating a positive correlation between the DP and the betweenness centrality. As in the scale-free case, we see that, for a fixed triangle formation probability p, the correlation increases with an increased number of incoming links m. However, for a fixed incoming number of links m, as we increase the probability of triangle formation p we see a clear decrease in the overall PCC values. For higher values of incoming links m an increased network size leads to increased PCC values. An increase in network size seems to generate an increase in the PCC values, meaning a stronger overall correlation.

4. A Biological Exploration

The initial motivation for exploring the correlation between the DP and the two topological measures in this paper is provided by the results in [11]. The authors of that paper analyze four biological networks and find no correlation between the DP and the betweenness centrality. As described in [11], those four networks are obtained from Cell Collective (CC, www.cellcollective.org, [22, 23]), an interactive platform for building and simulating logical models. The database contains over 60 peer-reviewed published models of biological networks and processes. The networks are of many sizes and represent a variety of different biological processing across a number of different organisms (e.g., yeast [45, 46], flies [47], humans [7, 10]).

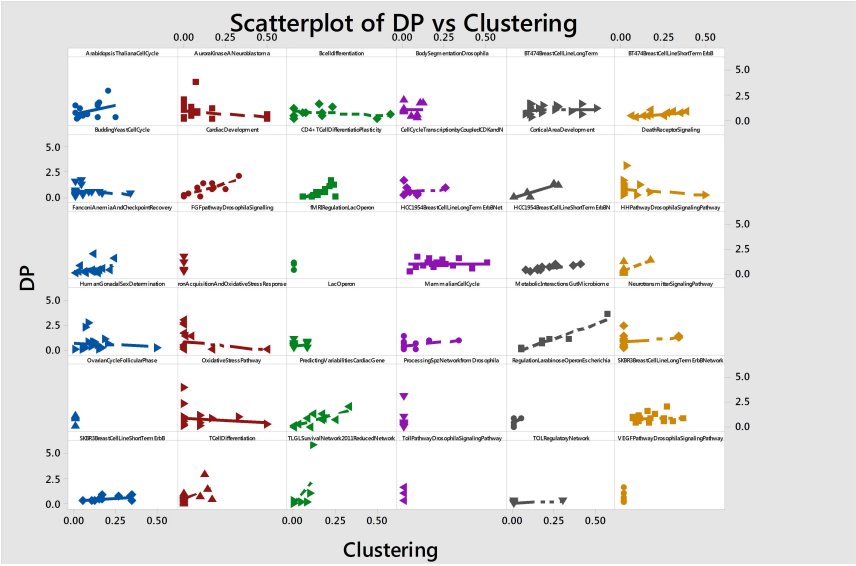

FIG. 13. Scatter plots with fitted least squares lines for DP versus the clustering coefficient by network. The 36 networks are listed in the order of the graphs. Notice that there is no observable trend or correlation.

We use the Cell Collective database to analyze all networks of sizes comparable to the results in this paper. More precisely, we find the DP, the clustering coefficient, and the betweenness centrality for 36 small networks of sizes 5 through 25. In Figure 13 we present the collection of scatter plots with fitted least squares lines for DP versus the clustering coefficient by network. Observe that there is no

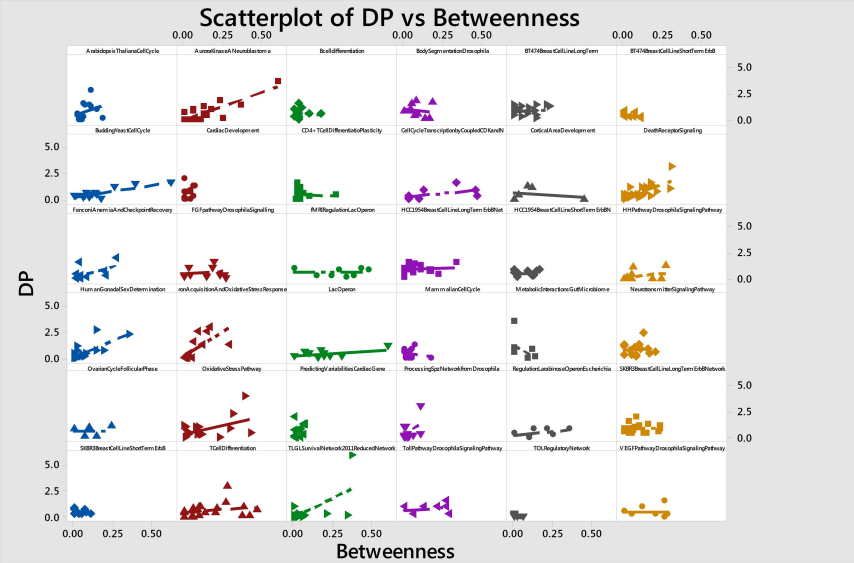

FIG. 14. Scatter plots with fitted least squares lines for DP versus the betweenness centrality by network. The 36 networks are listed in the order of the graphs. Notice that there is no observable trend or correlation.

observable trend in the data. A similar situation can be seen in Figure 14 where we present the collection of scatter plots with fitted least squares line for DP versus the betweenness centrality by network. There is no significant trend and the corresponding PCC values are fairly close to zero (not shown). We also show the corresponding scatter plots with fitted regression line for the aggregated data over all 36 networks in Figures 15 and 16. They produce a similar insignificant level of correlation. We conclude that compared to the theoretical networks analyzed in this paper, the biological networks do not exhibit a clear correlation between the DP and the two topological measures under consideration. For simplicity, in the mathematical and computational explorations of this paper we considered only simple graphs. However, self-regulation (self-loops) are important in biological networks. We will focus on the impact of self-regulation in further research.

In Figures 10-12 we note that an increase in the network size may have an impact on the PCC . We are interested to see if that is the case for the biological networks. For that purpose we group the networks in two simple categories: S, small (network size at most 16) and L, large (network at least 17). The cutoff is chosen so that the total number of nodes in all networks is divided almost equally between the two categories. There are 23 small networks with a total of 244 nodes, and 13 large networks with a total of 249 nodes.

In Figure 17 we show the scatter plots with fitted lines for both the clustering coefficient and the betweenness centrality separated by the two categories, S and L. Despite the lack of correlation, we can observe the trend given by the slope of the fitted line which matches the observations in Figure 12 for the Holme-Kim topology for higher values of incoming links m. More precisely, for the clustering coefficient a larger network size leads to slightly smaller PCC values. For the betweenness centrality a larger network size leads to larger PCC values. The same can be seen in Figure 17. Out of the three

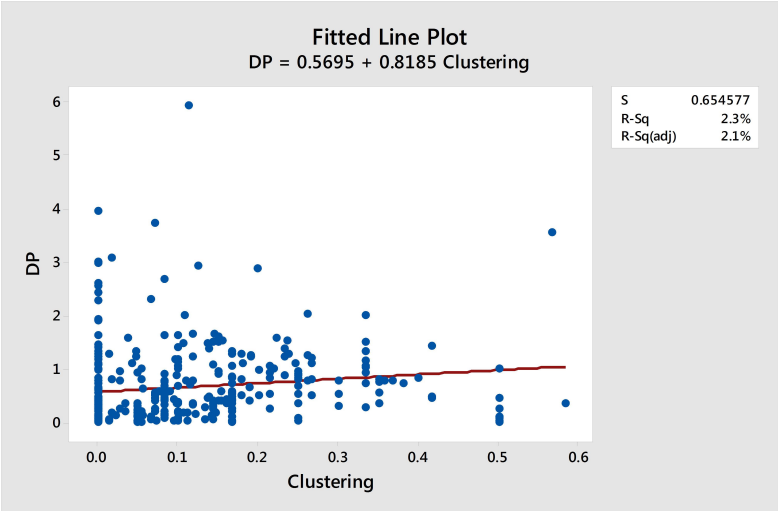

FIG. 15. Analog of Figure 13 for aggregated data over all 36 networks. The level of correlation is insignificant.

types of topologies considered in this work, the Holme-Lim topology that leads to Figure 12 exhibits similar trends when focusing on the impact of the network size. Although the correlations are irrelevant, the approximate trends related to the impact of the network size may point in the direction that the Holme-Kim topology could be closer to modeling BN representations of biological networks than the other topologies considered in this study. More research is needed to confirm these observations.

Thus, the results **of this biological exploration** indicate that the results in [11] seem to be valid for a wider range of biological networks, and that further studies are needed to clarify the correlations of the DP with other node attributes.

Note: For the completion of the biological exploration, we list all the networks used in Figures 13 and 14 using the labels from Cell Collective. We list the networks in the order they appear starting with the top left plot and ending with the bottom right plot: ArabidopsisThalianaCellCycle, AuroraKinaseANeuroblastoma, Bcelldifferentiation, BodySegmentationDrosophila, BT474BreastCellLineLongTerm, BT474BreastCellLineShortTermErbB, BuddingYeastCellCycle, CardiacDevelopment, CD4+TCellDifferentiatioPlasticity, CellCycleTranscriptionbyCoupledCDKandN, CorticalAreaDevelopment, DeathReceptorSignaling, FanconiAnemiaAndCheckpointRecovery, FGFpathwayDrosophilaSignalling, fMRIRegulationLacOperon, HCC1954BreastCellLineLongTermErbBNet, HCC1954BreastCellLineShortTermErbBN, HHPathwayDrosophilaSignalingPathway, HumanGonadalSexDetermination, IronAcquisitionAndOxidativeStressResponse, LacOperon, MammalianCellCycle, MetabolicInteractionsGutMicrobiome, NeurotransmitterSignalingPathway, OvarianCycleFollicularPhase, OxidativeStressPathway, PredictingVariabilitiesCardiacGene, ProcessingSpzNetworkfromDrosophila, RegulationLarabinoseOperonEscherichia, SKBR3BreastCellLineLongTermErbBNetwork, SKBR3BreastCellLineShortTermErbB, TCellDifferentiation, TLGLSurvivalNetwork2011ReducedNetwork, TollPathwayDrosophilaSignalingPathway, TOLRegulatoryNetwork, VEGFPathwayDrosophilaSignalingPathway.

5. Discussion

Looking at Section 3, the results indicate some connections between the DP and the topology of the network. There is a clear interaction between the local (clustering) and the global (betweenness centrality)

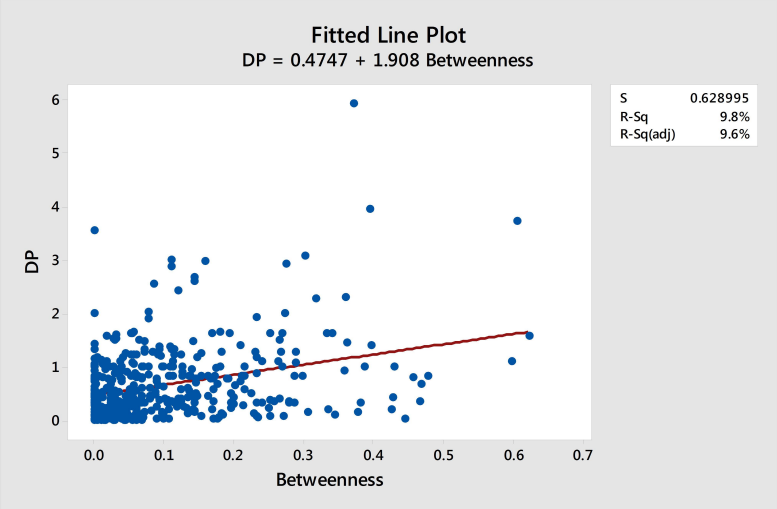

FIG. 16. Analog of Figure 14 for aggregated data over all 36 networks. The level of correlation is insignificant.

topological measures on the DP. An increase in clustering is associated with a decrease in the overall betweenness centrality of the network. The clustering coefficient is negatively correlated with the DP, however the correlation is weak at best. The betweenness centrality has a strong positive correlation with the DP.

Watts-Strogatz: The small-world model constructed by rewiring an initial network configuration to include clustering, leads to a clear region of high DP values in Figure 7 related mostly to small clustering values. At the same time, the negative PCC values in Figure 10 indicate a negative, but almost null amount of correlation between DP and the clustering coefficient, at least for the considered rewiring probabilities $p \in \{0.5, 0.6, 0.7, 0.8, 0.9\}$. This can be seen also in Figure 1 where for each value of the clustering coefficient the DP may take on a range of values.

On the other hand, Figure 7 shows large regions along the betweenness centrality axis of high DP, in particular peaks of global maxima surrounded by local maxima regions. The PCC, Figure 10, for the DP and the betweenness centrality shows somewhat stronger levels of positive correlation, however, the degree of correlation is consistently decreasing as the number of nodes in the network is increased. These trends hold for the considered values of the rewiring parameter p and provide more clarity of the swath of DP values in the plots of Figure 4.

Barabási-Albert: The scale-free model, constructed to describe networks composed of a large number of small-degree nodes coexisting with hubs or large-degree nodes, displays a wider and more clear region of high DP as a function of the clustering coefficient and betweenness centrality, Figure 8. The PCC, Figure 11, shows the anti-correlation for the DP and the clustering coefficient, with stronger anti-correlation for higher incoming number of links m. This matches the observations in Figure 2.

Figure 8 indicates that scale-free networks are associated with small betweenness centrality values and a positive correlation between the DP and the betweenness centrality. This fact becomes clear in Figure 11 where the PCC values are close to one, and supplementing the visual observations in Figure 5.

Holme-Kim: The scale-free with clustering model, attempting to combine the guiding ideas behind

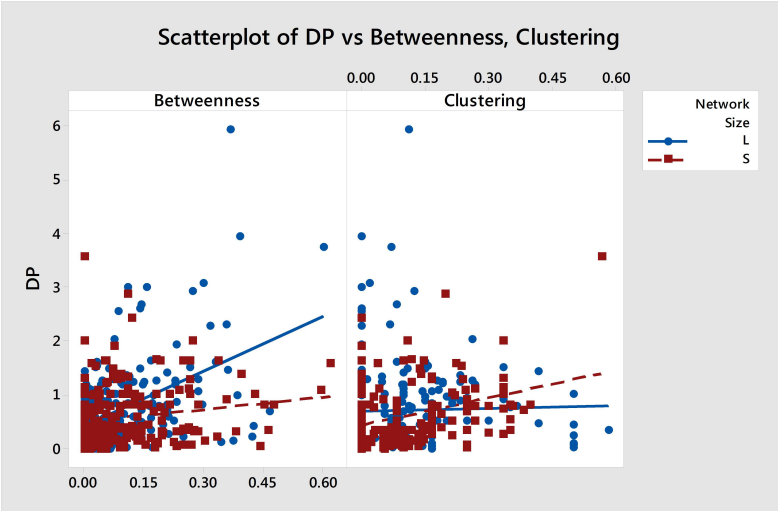

FIG. 17. Scatter plots with fitted lines for both the clustering coefficient and the betweenness centrality separated by the two categories, S small (network size at most 16) and L, large (network size 17-25). Out of the three types of topologies considered in this work, the Holme-Kim topology that leads to Figure 12 exhibits similar trends when focusing on the impact of the network size.

the Watts-Strogatz and the Barabási-Albert models, shows very similar connections between the DP, the clustering coefficient, and the betweenness centrality, Figure 9, as is displayed with the Barabási-Albert model. When considering the PCC, Figure 12, we see the general trend of mild anti-correlation of the DP and the clustering coefficient and the strong correlation of the DP and the betweenness centrality. The triangle formation probability influences the strength of correlation or anti-correlation such that for increasing triangle formation p we see a strengthening of the anti-correlation between DP and the clustering coefficient, and a weakening of the correlation between the DP and the betweenness centrality. These observations support the visual trends in Figures 3 and 6.

The results of this study indicate that for smaller size networks, the DP seems to be strongly correlated to the global topological measure of betweenness centrality and weakly or very weakly correlated to the local topological measure of clustering, depending on the underlying topology. The Barabási-Albert and the Holme-Kim topologies lead to fairly similar results, while the small-world networks generate a different type of behavior. Therefore, it seems that the existence of hubs in the network may be more important than the clustering of the nodes as far as the DP is concerned.

On the other hand, the biological exploration shows that for the networks considered in Cell Collective, those correlations are lost. There are several aspects of the theoretical models that may lead to this result: the undirected links, the choice of Boolean function, the homogeneity of the network, the actual topology models. All of these aspects are more flexible in real networks, so further work is needed to identify the right combination of node attributes, either topological or dynamical, that may impact the information gain in the network expressed via the DP.

5.1 *Future Direction*

The primary motivation behind this research is to understand the relationship between the entropic information stored in the Boolean functions and the overall topology of the network under consideration. This research focuses on undirected networks with nodes that operate by a single Boolean function. Expanding the research would require exploring more diverse network models, with a goal of better modeling biological networks and understanding the mathematics that govern real-world network phenomena.

5.1.1 *Boolean Functions.* The Boolean function that is considered for this research is the **AVERAGE** function, defined in the formula (2.1). While this is a good starting point, real-world networks don't always operate with such a strictly confining Boolean condition. A node that has k inputs will have 2^{2^k} potential Boolean functions. Having the option to select a variety of Boolean functions will greatly increase the exploratory options. Some types of functions of interest in biology may be canalizing or nested canalizing functions. Moreover, a methodological and systematic analysis of how various Boolean functions interact in a given topology could be revealing of the underlying processes that are driving real-world networks.

5.1.2 *Directed Networks.* A necessary simplification for this research is to consider only undirected networks, i.e. if node X_i is an input to node X_j, then node X_j is also an input to node X_i. This condition can be lifted to create a directed network, i.e. the links need not be bi-directional. Monitoring the in-degree and out-degree distributions and the impacts each has on the entropic information and the topological features would allow further insight into the various dependencies of real-world networks.

5.1.3 *Other Topologies.* While the three topologies considered in this paper are a good starting point for identifying the correlations between the DP and some topological measures, expanding this type of work to other topologies seems a natural next step. For example, self-regulation (self-loops are important in biological networks. Exploring the impact of self-regulation may reveal new insights. On the other hand, the concept of assortativity indicates to what extent nodes in a network associate with other similar nodes or with opposite nodes (disassortativity) [48]. Generally, the assortativity of a network is determined in conjunction to the node connectivity, so that in assortative networks nodes with large connectivity are mostly connected to nodes with large connectivity. However, it can be extended to other node characteristics such as betweenness centrality or clustering coefficient. Assortativity provides information about the structure of a network, but also about the robustness of the network to random or targeted attacks and virus spread. Constructing an assortative or disassortative network could be a more realistic approach for biological applications.

6. Acknowledgements

This work was supported by a Systems Science Collaborative Initiative Grant from the Nebraska Research Initiative [to B.W.W., J.A.R.]; and University of Nebraska at Omaha Lipton Program Fund [10289 to M.T.V.].

References

[1] Kauffman, S., (1993). The Origins of Order. *Oxford University Press*, New York, New York.

[2] Shmulevich I., Kauffman S.A., (2004). Activities and sensitivities in Boolean network models, *Physical Review Letters* 93(4), 048701.

[3] Shmulevich I., Dougherty E.R., Zhang W., (2002). From Boolean to probabilistic Boolean networks as models for genetic regulatory networks, *Proceedings of the IEEE* 90(11), 1778-1792.

[4] Shmulevich I., Lähdesmäki H., Dougherty E.R., Astola J., Zhang W., (2003). The role of certain Post classes in Boolean network models of genetic networks, *Proceedings of the National Academy of Sciences of the United States of America* 100(19), 10734–10739.

[5] Helikar T., Konvalina J., Heidel J., Rogers J.A., (2008). Emergent decision-making in biological signal transduction networks, *Proceedings of the National Academy of Sciences of the United States of America* 105(6), 1913-1918.

[6] Kochi N., Matache M.T., (2012). Mean-Field Boolean Network Model of a Signal Transduction Network, *Biosystems* 108(1-3), 14-27.

[7] Conroy, B. D. and Herek, T. A. and Shew, T. D. and Latner, M. and Larson, J. J. and Allen, L. and Davis, P. H. and Helikar, T. and Cutucache, C. E., (2014). Design, Assessment, and in vivo Evaluation of a Computational Model Illustrating the Role of CAV1 in CD4(+) T-lymphocytes, *Front. Immunol.* 5, 599.

[8] Abou-Jaoude, W. and Monteiro, P. T. and Naldi, A. and Grandclaudon, M. and Soumelis, V. and Chaouiya, C. and Thieffry, D., (2015). Model checking to assess T-helper cell plasticity, *Front. Bioeng. Biotechnol.* 2, 86.

[9] Abou-Jaoude, W. and Traynard, P. and Monteiro, P. T. and Saez-Rodriguez, J. and Helikar, T. and Thieffry, D., Chaouiya C., (2016). Logical Modeling and Dynamical Analysis of Cellular Networks, *Front. Genet.* 7, 94.

[10] Mendez, A. and Mendoza, L., (2016). A Network Model to Describe the Terminal Differentiation of B Cells, *PLoS Comput. Biol.* 12, e1004696.

[11] Pentzien T., Puniya B.L., Helikar T., Matache M.T., (2018). Identification of biologically essential nodes via determinative power in logical models of cellular processes, *Frontiers in Physiology*, 9, 1185.

[12] Klemm K., Bornholdt S., (2000). Stable and unstable attractors in Boolean networks, *Phys. Rev.* E 72, 055101.

[13] Raeymaekers L., (2002). Dynamics of Boolean networks controlled by biologically meaningful functions, *Journal of Theoretical Biology* 218, 331-341.

[14] Albert, R. and Othmer, H., (2003). The topology of the regulatory interactions predicts the expression pattern of the segment polarity genes in *Drosophila melanogaster*, *J. Theor. Bio.* 223, 1-18.

[15] Saadatpour, A. and Albert, R. and Reluga, T. C., (2013). A reduction method for Boolean network models proven to conserve attractors, *SIAM Journal on Applied Dynamical Systems* 12, 1997-2011.

[16] Correia, R.B., Gates, A.J., Wang, X. and Rocha, L.M., (2018). CANA: A Python Package for Quantifying Control and Canalization in Boolean Networks, *Frontiers in Physiology* 9, 01046.

[17] Grob, A., Kracher, B., Kraus, J.M., Kuhlwein, S.D., Pfister, A.S., Wiese, S., Luckert, K., Potz, O., Joos, T., Van Daele, D., De Raedt, L., Kuhl, M., Kestler, H., (2019). Representing dynamic biological networks with multi-scale probabilistic models, *Communications Biology* 2(12), s42003-018-0268-3.

[18] Murrugarra, D., Dimitrova, E., (2015). Molecular Network Control Through Boolean Canalization. *EURASIP Journal on Bioinformatics and Systems Biology* 1, s13637-015-0029-2.

[19] Heckel R., Schober S., Bossert M., (2013). Harmonic analysis of Boolean Networks: determinative power and perturbations. *EURASIP Journal on Bioinformatics and Systems Biology*, 1–13.

[20] Matache, M.T. and Matache, V., (2016) Logical Reduction of Biological Networks to Their Most Determinative Components, *Bull. Math. Biol.* 78(7), 1520-1545.

[21] Klotz, J.G., Kracht, D., Bossert, M., Schober, S., (2014). Canalizing Boolean functions maximize the mutual information, *IEEE Trans. Inform. Theory* 60(4), 2139 - 2147.

[22] Helikar, T. and Kowal, B. and McClenathan, S. and Bruckner, M. and Rowley, T. and Wicks, B. and Shrestha, M. and Limbu, K. and Rogers, J. A., (2012). The Cell Collective: Toward an open and collaborative approach to systems biology, *BMC Systems Biology* 6, 96.

[23] Helikar, T. and Kowal, B. and Rogers, J. A., (2013). A cell simulator platform: the cell collective, *Clinical Pharmacology And Therapeutics* 93, 393-395.

[24] Ribeiro, A. S. and Kauffman, S. A. and Lloyd-Price, J. and Samuelsson, B. and Socolar, J. E. S., (2008). Mutual information in random Boolean models of regulatory networks, *Phys. Rev.* E 77, 011901.

[25] Krawitz, P. and Shmulevich, I., (2007). Basin entropy in Boolean network ensembles, *Physical Review Letters* 98(15), 158701.

[26] Krawitz, P. and Shmulevich, I., (2007). Entropy of complex relevant components of Boolean networks, *Phys. Rev.* E 76, 036115.

[27] Watts, D. and Strogatz, S., (1998). Collective dynamics of "small-world" networks. *Nature*, Vol. 393.

[28] Barabási, A., Albert, R., and Jeong, H., (1999). Mean-field Theory for scale-free random networks. *Nature* 401, 130.

[29] Barabási, A. and Albert, R., (1999). Emergence of Scaling in Random Networks. *Science* 15 Oct 1999: Vol. 286, Issue 5439, pp. 509-512 DOI: 10.1126/science.286.5439.509.

[30] Holme, P. and Kim, B. J., (2002). Growing scale-free networks with tunable clustering. *Physical Review* E, 65, 026107.

[31] Hernandez, J. M., Li, Z., Van Mieghem, P., (2014). Weighted betweenness and algebraic connectivity, *Journal of Complex Networks* 2(3), 272-287.

[32] Dzaferagic, M., Kaminski, N., McBride, N., Macaluso, I., Marchetti, N., (2018). A functional complexity framework for the analysis of telecommunication networks, *Journal of Complex Networks* 6(6), 971-988.

[33] Hu S.-T., (1965). *Threshold logic*, University of California Press, Berkeley, 1965.

[34] Anthony M., (2003). Accuracy of Classification by Iterative Linear Thresholding, *Proceedings of the Workshop on Discrete Mathematics and Data Mining*, 3rd SIAM International Conference on Data Mining, San Francisco.

[35] Shannon, C.E., (1948). A Mathematical Theory of Communication. *Bell System Technical Journal*, 27, 379-423 & 623-656.

[36] Censi F., Giuliani A., Bartolini P., and Calcagnini G., (2011). A multiscale graph theoretical approach to gene regulation networks: a case study in atrial fibrillation, *IEEE Transactions on Biomedical Engineering* 58(10), 2943-2946.

[37] Gorban A. N., Smirnova E. V., and Tyukina T. A., (2010). Correlations, risk and crisis: From physiology to finance, *Physica A: Statistical Mechanics and its Applications*, 16(389), 3193-3217.

[38] Mojtahedi M., Skupin A., Zhou J., Castano I.G., Leong-Quong R.Y.Y., Chang H., Trachana K., Giuliani A., and Huang S., (2016). Cell fate decision as high-dimensional critical state transition, *PLoS biology* 12(14), e2000640.

[39] Csermely P., Agoston V., and Pongor S., (2005). The efficiency of multi-target drugs: the network approach might help drug design, *Trends in pharmacological sciences* 4(26), 178-182.

[40] Kovacs I. A., Palotai R., Szalay M. S., and Csermely P., (2010). Community landscapes: an integrative approach to determine overlapping network module hierarchy, identify key nodes and predict network dynamics, *PloS one* 9(5), e12528.

[41] Freeman, L. C., (1977). A set of measures based on centrality betweenness. *Sociometry*, 40: 35-41.

[42] Trajanovski, S., Martin Hernandez, J., Winterbach, W., Van Mieghem, P., (2013). Robustness Envelopes of Networks. *Journal of Complex Networks* 1(1), 44-62.

[43] Pastor-Satorras, R., Vespignani, A., (2001). Epidemic Spreading in Scale-Free Networks. *Phys Rev Lett* 86(14), 3200–3203.

[44] Menezes, M. B. C., Kim, S., and Huang, R., (2017). Constructing the Watts-Strogatz network from a small-world network with symmetric degree distribution. *PLOS One*, https://doi.org/10.1371/journal.pone.0179120.

[45] Irons, D. J., (2009). Logical analysis of the budding yeast cell cycle, *Journal of Theoretical Biology* 257(4), 543-559.

[46] Todd, R. G. and Helikar, T., (2012). Ergodic Sets as Cell Phenotype of Budding Yeast Cell Cycle, *PloS One* 7(10), e45780.

[47] Marques-Pita, M. and Rocha, L. M., (2013). Canalization and Control in Automata Networks: Body Segmentation in Drosophila melanogaster, *PLoS One* 8(3), e55946.

[48] Noldus, R., Van Mieghem, P., (2015). Assortativity in complex networks, *Journal of Complex Networks* 3(4), 507-542.