


**Gene regulatory and signaling networks exhibit distinct topological distributions of motifs**Gustavo Rodrigues Ferreira,<sup>1</sup> Helder Imoto Nakaya,<sup>2</sup> and Luciano da Fontoura Costa<sup>1,\*</sup><sup>1</sup>*São Carlos Institute of Physics, University of São Paulo, São Carlos, Brazil*<sup>2</sup>*School of Pharmaceutical Sciences, University of São Paulo, São Paulo, Brazil* (Received 24 October 2017; revised manuscript received 18 December 2017; published 27 April 2018)

The biological processes of cellular decision making and differentiation involve a plethora of signaling pathways and gene regulatory circuits. These networks in turn exhibit a multitude of motifs playing crucial parts in regulating network activity. Here we compare the topological placement of motifs in gene regulatory and signaling networks and observe that it suggests different evolutionary strategies in motif distribution for distinct cellular subnetworks.

DOI: [10.1103/PhysRevE.97.042417](https://doi.org/10.1103/PhysRevE.97.042417)**I. INTRODUCTION**

In mathematics, cellular biological processes can be represented through concepts of graph theory [1–4]. In these models, proteins and genes are depicted as nodes, and the chemical reactions or regulatory interactions between them as edges. Such representations of molecular systems, called networks, highlight the extensive crosstalk between a cell's components and the complex ways in which it regulates itself. To function, cells recruit or silence specific subsets of nodes and edges that have to be both spatially and temporally coordinated [5].

This organization inside the cell serves the purpose of integrating and propagating hundreds of distinct signals and stimuli. This complex process, known as signal transduction, involves two different types of networks [Fig. 1(a)]: signaling networks (SNs) in the cytosol and gene regulatory networks (GRNs) in the nucleus. The former consists of a series of biochemical reactions that activate or inactivate proteins, channels, and transcription factors, generally starting with the binding of a ligand molecule to a receptor protein. In an SN, nodes are biochemical species that undergo the aforementioned reactions, and an edge from species  $X$  to  $Y$  indicates that  $X$  triggers or ends the activity of  $Y$ .

The GRN is a network composed of transcription factors (proteins) that enhance or inhibit the translation of other genes, including themselves. Signal propagation in a GRN usually initiates with the translocation of an activated transcription factor to the nucleus, where it activates the transcription of specific targets [see Fig. 1(a)]. Both kinds of networks show a common feature: Signal propagation starts from a specific origin, which we dub a receptor, input, or upstream node.

The different combinations of activated SNs and GRNs will determine the cell's response to one or more stimuli [see Fig. 1(b)]. In fact, different activation patterns for the same receptor can also induce distinct cellular responses [6,7], making for an extremely diverse signal processing system. For instance, ligands such as EGF and TGF- $\beta$  induce cell proliferation [8], whereas IFN- $\gamma$  and IL-4 induce B lymphocytes to secrete antibodies [9].

Thus, understanding the networks' topology, or structure, is crucial to grasping the pathway's qualitative responses. For instance, a cell's ability to endure deleterious mutations in transcription factors has been shown to evolve gradually with changes in its GRN's topology [10]. Its resilience to total collapse (here meaning a sudden irreversible transition to a state where all cellular activity ceases) is also known to be defined by a delicate interplay between its biochemical parameters and GRN structure [11].

Of particular interest, then, are certain ubiquitous interaction patterns (or subgraphs) termed motifs [2,12]. The switch, the feedforward motif, and the feedback loop, shown in Fig. 2, are examples of this class, and all of them have been shown to possess special dynamical properties regarding signal transduction and transmission [13–16].

Previous works based on this premise have been successful in identifying signaling motifs that participate in signal transduction and cellular decision-making [17]. Furthermore, it was shown [18] that motifs tend to organize themselves in clear regions around cell receptors, suggesting both a role in signal processing and the importance of their precise topological placement in SNs. In this paper we compare the distribution of motifs in both the GRNs and SNs to assess their roles in signal propagation during cell differentiation. We use publicly available data from online databases to construct the networks and analyze them using local concepts from network theory, emphasizing the characterization of motifs along topological neighborhoods [18–20]. The results indicate two main types of organization. The motifs in SNs tend to organize in symmetric, concentric layers around the receptor. On the other hand, in the case of GRNs, the motifs spread out in an asymmetric fashion along the hierarchical layers.

**II. METHODS****A. Network construction**

The two types of networks analyzed here, SNs and GRNs, have been made available in public databases. For the GRNs, the RegNetwork [21] database is a knowledge-based collection of regulatory interactions between transcription factors, microRNAs, and target genes. It combines and synthesizes

\*ldfcosta@gmail.com

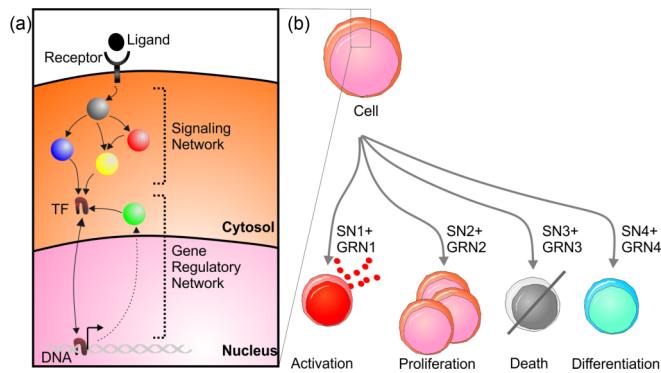


FIG. 1. Signal transduction in eukaryotic cells. The signaling cascade initiated by (a) a ligand-receptor binding drives (b) cellular decision-making depending on the activated receptor.

information from curated databases such as GenBank, BioGrid, and Ensembl.

The signaling pathways were taken from Reactome [22] and processed with the rBiopaxParser package [23]. This processing entails extracting the network structure from a Biopax annotation file, used in such public databases, and converting it to an edge list suited to our needs. Reactome is a database of curated interactions maintained by a collaboration among several research institutes and also integrates orthology-based information from Ensembl.

For each signal transduction pathway considered here, a GRN subnetwork consisting of the genes involved in the associated biological process was extracted. This serves the double purpose of reducing the computational load of our analyses and focusing our attention on the biological entities that are actually relevant to the process.

More specifically, for each pathway, an RNA sequencing (RNA-seq) study concerning an associated differentiation process was used to identify differentially expressed genes (DEGs). The preprocessed expression profiles were taken from the sequence read archive [24] (SRA), and DEGs were selected in a two-stage process, variance control and subsequent filtering of highly variant genes, as described below.

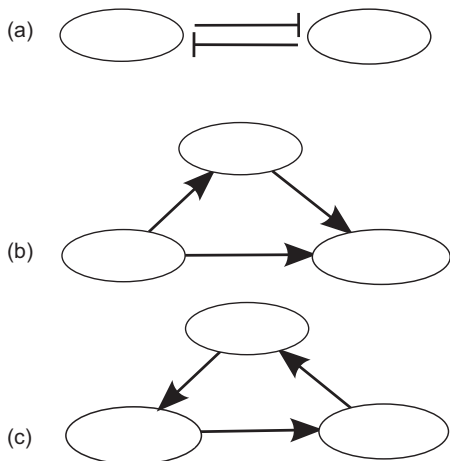


FIG. 2. Examples of motifs found in cellular systems: (a) the switch, (b) the feedforward loop, and (c) the feedback loop.

In an RNA-seq expression profiling, gene expression (under a null hypothesis of no change between samples) is modeled as a negative binomial distribution [25]. This distribution is characterized by a strong dependence between mean and variance, which affects subsequent variance-based analyses. Thus, the expression profiling is transformed according to the variance stabilizing transformation for negative binomial data, derived by Guan [26]. More specifically, Guan proves that if  $X$  is a negative-binomial-distributed random variable with parameters  $r$  and  $p$ , then the transformed variable  $Y = \sqrt{r} \sinh^{-1} \sqrt{X/r}$  has no dependence between mean and variance.

After removing the mean-variance dependence, the variance between different sample phenotypes (obtained from the SRA) was used as a criterion for DEG selection. We performed an analysis of variance [27], which assesses the statistical significance of the within-class variance to total variance ratio, and attributed a  $p$  value to each gene. Then we controlled the rate of false-positive discoveries with Benjamini and Hochberg’s false -iscovery rate (FDR) [28] method. The FDR orders the  $p$  values increasingly, so that  $p_1 \leq p_2 \leq \dots \leq p_n$ , and selects those satisfying  $p_k \leq \alpha k/m$ , where  $\alpha$  is the desired  $p$ -value threshold for an individual test. It is known (see, for instance, [29]) that this procedure has an expected false-discovery rate upper bounded by  $\alpha$ .

After extracting the DEGs, the subnetworks related to each signaling pathway were taken as the subgraphs of the RegNetwork GRN generated by the DEGs.

**B. Network analysis**

A concept central to our analyses is that of a node’s neighborhood [19]. Given a network, represented as a graph  $G$ , and a node  $v$  in  $G$ , the  $d$ th neighborhood of  $v$ , denoted by  $R_d(v)$ , is the set of all nodes accessible in at most  $d$  steps from  $v$ . In a cellular network, successive neighborhoods act as a proxy for the dynamics of signal propagation in the cell; since edges represent direct regulatory interactions, the chemical reactions associated with signal transduction occur along paths in the cellular network.

Associated with a node’s neighborhoods are several different measures. Of interest to us are the concepts of concentric symmetry and the motif cumulative distribution function (mCDF). The former is defined as the entropy associated with transition probabilities of a walk  $h$  steps long starting from  $v$  [20]. In other words, it quantifies how similar the possible walks of a certain length around a node can be: The higher its value, the more similar they are and thus the more symmetric the node’s neighborhood. Concentric symmetry is preferred here over the automorphism-related symmetry metrics [30,31] due to the latter’s computational intractability (in general) and poor normalization (automorphism groups have a loose upper bound on their order at around  $N!$  for a graph on  $N$  nodes, as discussed by Silva *et al.* [20]). The motif cumulative distribution addresses how a certain motif is arranged in the neighborhoods of a node. Given a motif structure, a node  $v$ , and a graph  $G$ , it is defined as the number of occurrences of the motif along successive neighborhoods  $R_d(v)$ , subsequently normalized to approach unity.

The mCDF is tied to the motif location index (MLI) metric used by Ma'ayan *et al.* [18]. For a particular instance of a motif in a network, the MLI represents where the instance lies between a membrane receptor and a given cellular machinery. More specifically, both assess the placement of motifs relative to an origin (the receptor) in a cellular network. However, our work does not focus on specific cellular machines or processes, looking instead at generic signaling pathways. Thus, our metric lacks the distance to cellular machinery component of the Ma'ayan *et al.* MLI and thus reflects the positioning of motifs without regard to a particular cellular task (for our purposes, a cellular machinery is a subset of a GRN or SN containing nodes related to a particular cellular function, e.g., translation or cell division). This relation will be crucial when comparing our results to those of Ma'ayan *et al.* in Sec. III.

Another issue is determining whether the observed distributions are relevant when compared to randomly generated networks with the same degree distribution [32,33]. We address this question by random sampling of networks through the edge-switching Monte Carlo algorithm described by Gkantsidis *et al.* [34]. Different motif distributions were compared using the supremum distance for function spaces: If  $f$  and  $g$  are two real-valued functions defined on  $X$ , the distance between them is  $d(f, g) := \sup_{x \in X} |f(x) - g(x)|$ .

Based on this distance, the distributions  $f_1, f_2, \dots, f_n$  obtained through the sampling procedure, and given the observed distribution  $f_{obs}$ , we define a  $z$  score  $Z_{obs} = d(f_{obs}, \bar{f})/s_f$ , where  $\bar{f} := (1/n)(\sum_{i=1}^n f_i)$  and  $s_f$  is the sample's standard deviation (in terms of the supremum distance). We also used a bootstrap  $p$  value based on the distance to the mean; basically,  $p$  is the proportion of times when  $d(f_i, \bar{f}) > d(f_{obs}, \bar{f})$ .

Another metric we incorporate from Ma'ayan *et al.* is the density of information processing (DIP). It is defined as the increase in the number of motifs divided by the increase in edges between consecutive neighborhoods. In the work of Ma'ayan *et al.*, this ratio is multiplied by the grid coefficient [18,35], a generalization of the clustering coefficient taking into account the formation of rectangles. However, since our considered motifs are of size smaller than 4, we opted to forego this normalization by grid coefficient. In keeping with the idea that motifs are a network's processing units, this measure indicates the proportion of signal processing activity as information propagates through the network's paths.

**III. RESULTS AND DISCUSSION**

We chose as representatives GRNs and SNs associated with three major signal transduction pathways in a mouse (*Mus musculus*) and human (*Homo sapiens*). For the former, we studied the T-cell-receptor (TCR) signaling network, which drives the differentiation of T lymphocytes [36], and the EGF receptor (EGFR) pathway, involved in cell growth and survival [37]. In humans, we studied the TGF- $\beta$  pathway, which regulates cell growth, differentiation, and apoptosis [38,39]. The data for each pathway were taken from the SRA as described in Sec. II A. The RNA-seq profilings used were, respectively, GSE48138 [40], GSE86467 [41], and GSE36552 [42].

The obtained networks were analyzed with respect to their size (see Table I) and degree distributions. We note, as expected, that signaling components are much smaller than

TABLE I. Number of nodes and edges for our representative networks. Numbers are displayed as nodes, edges.

Receptor	Network	
	SN	GRN
TCR	148, 848	3835, 13 390
EGFR	253, 1153	8092, 36 266
TGF- $\beta$	67, 233	2261, 5384

their GRN counterparts. In keeping with the current literature [3,21], degree distributions were seen to be power laws as determined by fitting a linear model to the logarithmically transformed degrees and degree probabilities; all linear models had a coefficient of determination above 0.7.

The three distinct pathways, each with two different networks (a cytosolic signaling pathway component and a gene regulatory component), were analyzed with respect to the mCDF, concentric symmetry, and DIP metrics. For the motif cumulative distribution, our results show that signaling networks employ their motifs in a much more distinct fashion (lengthwise) than GRNs (see Fig. 3). That is to say, different motifs may appear more strongly in different neighborhoods (like in the TCR signaling network [Fig. 3(c)] or a certain motif may not be present in the network (notice how the double feedforward loop is missing from the TGF- $\beta$  signaling

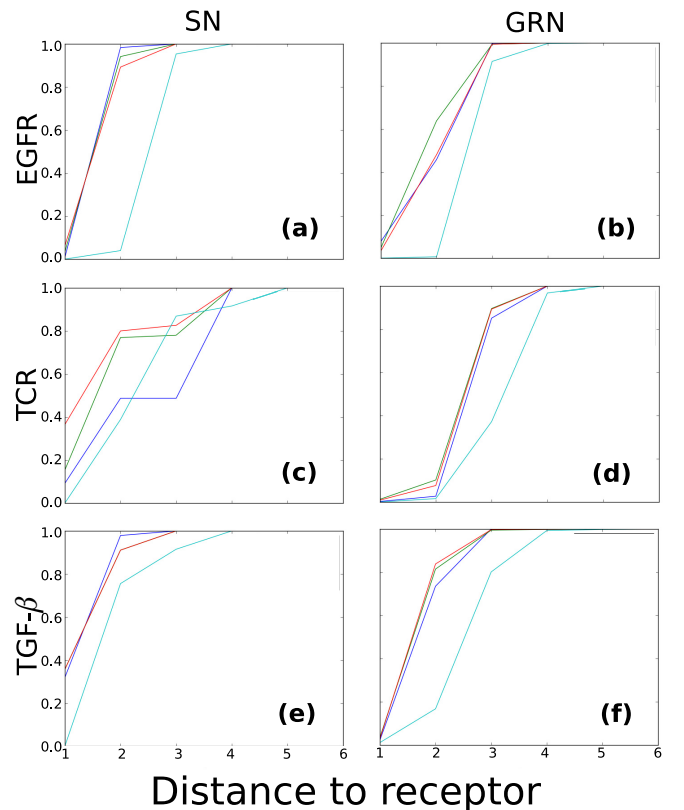


FIG. 3. The mCDFs for distinct motifs in the various networks: (a), (c), and (e) motif distributions for signaling pathways and (b), (d), and (f) mCDFs of gene regulatory networks. Each color represents a motif: blue, feedforward loop; green, double feedforward loop; red, switch; and cyan, feedback loop.

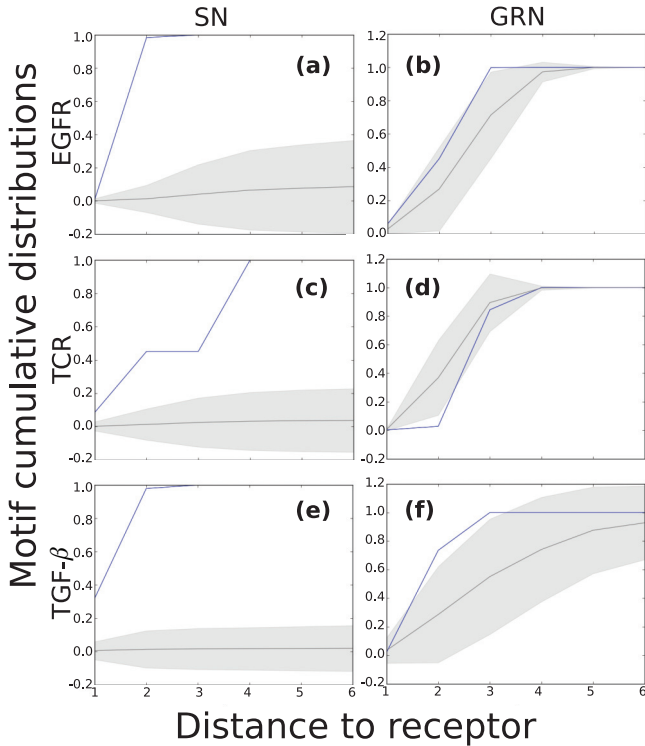


FIG. 4. Bootstrap analysis for the feedforward loop. The blue line shows the observed distribution. The gray line stands for the random ensemble mean and the shaded area represents a standard deviation above and below the mean.

network [Fig. 3(e)]. This variety of patterns in motif placement may have emerged in two opposite ways: It might indicate that the placement of motifs in signaling networks is not under any evolutionary constraint or it might suggest tailored distributions for each pathway.

We answer this question by comparing the networks’ motif cumulative distribution to random networks with the same degree distribution, generated as described in Sec. II B. Our results for the feedforward loop are displayed in Fig. 4 and Table II. It is evident that signaling networks are further removed from the distribution generated by random networks than their gene regulatory counterparts. Following the idea that deviations from the random ensemble indicate natural selection [3,13], our results suggest that motif distributions have indeed been tailored to specific pathways, insofar as signaling components are involved. As for the GRNs, there is not enough evidence to indicate evolutionary pressure; a

TABLE II. Bootstrap statistics for the motif distributions. The left number indicates the network’s  $z$  score and the right number its  $p$  value (see Sec. II B).

Receptor	Network	SN	GRN
TCR		9.27, 0.006	0.57, 0.22
EGFR		3.42, 0.002	0.034, 0.35
TGF- $\beta$		7.14, 0.008	-0.14, 0.43

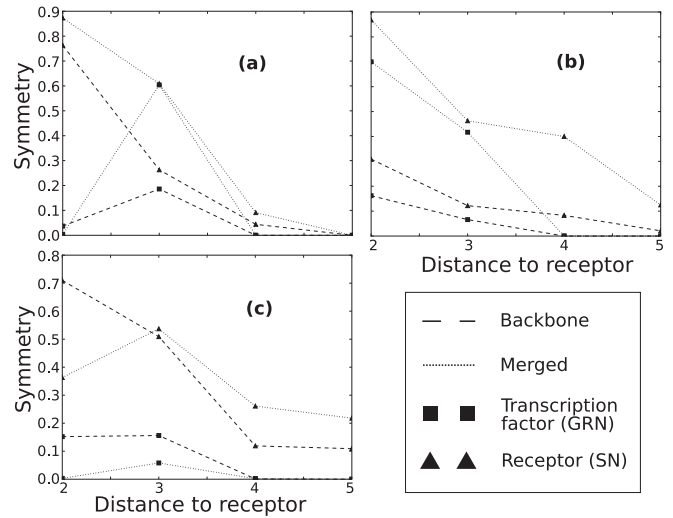


FIG. 5. Concentric symmetries for increasing neighborhoods of the receptor nodes in each studied network: (a) EGFR, (b) TGF- $\beta$ , and (c) TCR. The legend defines the labels for all symmetry plots. The transcription factors and receptors for each pathway are described in the text (see Sec. III).

possible explanation is that heavy-duty signal processing is performed in signaling pathways (see, for instance, Fig. 7 and the associated discussion in this section) and GRNs are simply responsible for carrying out the consequences of cellular signal integration.

Additionally, the feedback loop is given a special place in all networks, being placed further from the upstream node than other motifs. This tendency, already observed in signaling networks by Ma’ayan *et al.* [18], is reinforced here and extended to gene regulatory circuitry. As with other principles of cell network evolution, this has plausibly come to be due to distinct dynamical properties for the feedback loop when compared to other motifs. We cite here the roles of the feedforward and feedback loops in amplifying signals and filtering out noise, respectively, as an example [13,16].

With regard to symmetry, it is also observed that signaling networks are considerably more symmetric around their input nodes (Fig. 5). We see that the cytosolic components consistently display higher-symmetry values in neighborhoods closer to the receptor when compared to corresponding GRN components. Again, this might reflect a different usage of the motifs’ dynamical properties by the SNs when compared to

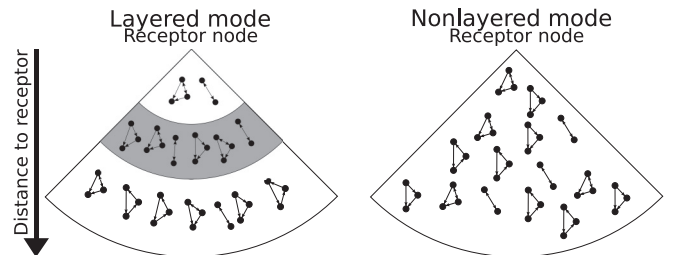


FIG. 6. The two types of motif placement found in this work. To the left is the motif belt observed in SNs and to the right, the asymmetric distribution present in GRNs.

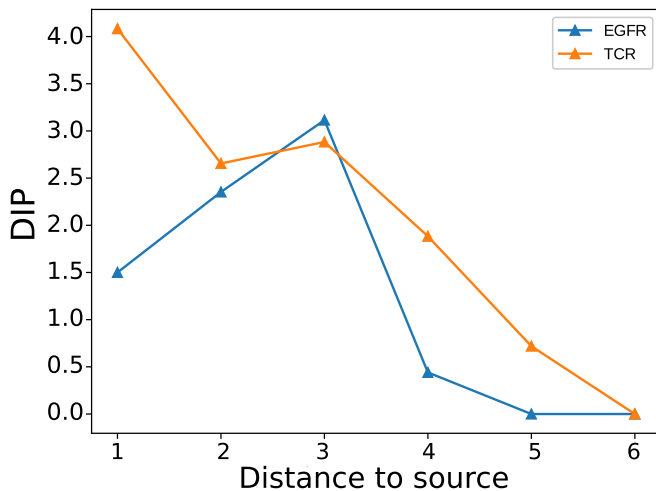


FIG. 7. Density of information processing for the TCR and EGFR signaling pathways.

GRNs. In particular, the higher symmetries around the receptor node of signaling networks suggest a uniforming constraint on the paths originating at the upstream node and consequently on the placement of motifs around it.

Next we compare our results to those of Ma'ayan *et al.* As discussed in Sec. IIB, the mCDF and MLI are both related to how motifs are arranged relative to a receptor. However, the MLI considers motif placement relative to an origin and a destination (a cellular machinery) in the network, while the mCDF considers only a starting point. In analyzing the MLI distributions for different motifs and machineries (see Fig. 5 in Ref. [18]), we may expand our conclusions by comparing with a related measure. The analyses of Ma'ayan *et al.* place most motifs roughly at a halfway point between receptor and target machinery, regardless of motif or signaling pathway. When compared to our results on the motif cumulative distribution (Fig. 3), which suggest a vast heterogeneity of motif placements depending on the network and motif in question, we are drawn to the conclusion that different machineries are found at differing distances from the input node. Thus, differences in path length in a signaling network may be related to cellular machinery activation and decision-making.

Finally, by combining both aspects of symmetry and cumulative distribution of motifs, as well as previous results in the area [18], we see the emergence of two distinct patterns of motif placement (see Fig. 6): In signaling pathways, the location of motifs is strictly constrained, leading to concentric motif belts, which may be one or more, around the receptor. For instance, the TCR signaling network shows two belts at distances 1 and 3 from its receptor, while the EGFR pathway shows only a single belt three steps away (see Fig. 7). Gene regulatory networks, on the other hand, present a more relaxed distribution when compared to random networks with the same

degree distribution and suggest a different use for the motifs' dynamical properties.

These findings reassert the notion that cellular networks evolved as modular, relatively independent solutions to distinct selective pressures [43]. Each part of a cell's machinery would then exhibit different organizational features to address their specific demands [44,45], and as a particular example we find here the difference between motif organization of signaling networks and gene regulatory networks. As sources of distinct selective pressures, we could cite the different timescales of network dynamics (proteins transition between their active and inactive states much faster than the transcription of genes; see Ref. [13]) and the need for SNs to cope with ever-changing external environments [45–47].

#### IV. CONCLUSION

The idea of motifs as regular components and processing units of biological networks is central in our understanding of cellular systems. They have been shown to be ubiquitous in situations as distinct as gene regulation, signal processing, and metabolism. Additionally, advances in characterizing the relationship between network topology and dynamics point toward special roles of particular motifs such as the feedforward and feedback loops. Despite these advances in characterizing both GRNs and SNs, studies combining both of them remain scarce. Here we compared them with respect to their characteristics pertaining signal transduction and propagation.

Our results add another layer of versatility to the functional importance of motifs by suggesting that their topological distribution differs from signaling to gene regulatory networks. More specifically, signaling pathways show one or more symmetric layers of motifs around the receptor, differing strongly from random networks with the same degree distribution. In contrast, gene regulatory networks display asymmetric motifs in a single layer around key transcription factors, on par with random networks of the same degree distribution. We remark that feedback loops are usually lagged behind other motifs, as was already noted by Ma'ayan *et al.*

Thus, our work expands on the previous notion that biological networks in different locations of the cell, or performing different functions, exhibit distinct topological features. Such diversity of topologies could conceivably have emerged as separate evolutionary answers to the different selective pressures acting on the cellular components. As such, efforts to understand exactly what the demands of each cellular subnetwork are, and how the cell addresses them, can offer great insights into both the organization of cellular circuits and the dynamics of network systems.

#### ACKNOWLEDGMENT

The authors acknowledge financial support from FAPESP Grants No. 2015/22308-2, No. 2012/19278-6, and No. 2014/19323-7.

[1] B. Alberts, A. Johnson, J. Lewis, D. Morgan, M. Raff, K. Roberts, and P. Walter, *Molecular Biology of the Cell*, 6th ed. (Garland Science, New York, 2015).

[2] U. Alon, *Nat. Rev. Genet.* **8**, 450 (2007).

[3] A. L. Barabási and Z. N. Oltvai, *Nat. Rev. Genet.* **5**, 101 (2004).

- [4] M. Buchana, G. Caldarelli, P. de los Rios, and M. Vendruscolo, *Networks in Cell Biology* (Cambridge University Press, Cambridge, 2010).
- [5] B. Kholodenko, *Nat. Rev. Mol. Cell Biol.* **7**, 165 (2006).
- [6] X. Liu and R. Bosselut, *Nat. Immunol.* **5**, 280 (2004).
- [7] T. Tomida, S. Oda, M. Takekawa, Y. Iino, and H. Saito, *Sci. Signal.* **5**, ra76 (2012).
- [8] A. Pardee, *Science* **246**, 603 (1989).
- [9] S. Nutt, P. Hodgkin, D. Tarlinton, and L. Corcoran, *Nat. Rev. Immunol.* **15**, 160 (2015).
- [10] A. Wagner, S. Ciliberti, and O. Martin, *PLoS Comput. Biol.* **3**, e15 (2007).
- [11] J. Gao, B. Barzel, and A.-L. Barabási, *Nature (London)* **530**, 307 (2016).
- [12] R. Milo, S. Shen-Orr, S. Itzkovitz, N. Kashtan, D. Chklovskii, and U. Alon, *Science* **298**, 824 (2002).
- [13] U. Alon, *An Introduction to Systems Biology: Design Principles of Biological Circuits* (CRC, Boca Raton, 2007).
- [14] A. Ghaffarizadeh, N. S. Flann, and G. J. Podgorski, *BMC Bioinformatics* **15**, S7 (2014).
- [15] D. Liu, L. Albergante, and T. Newman, *Nucleic Acids Res.* **45**, 7078 (2017).
- [16] Y.-J. Shin and L. Bleris, *PLoS ONE* **5**, e12785 (2010).
- [17] F. Diella, N. Haslam, C. Chica, A. Budd, S. Michael, N. P. Brown, G. Trave, and T. J. Gibson, *Front. Biosci.* **13**, 6580 (2008).
- [18] A. Ma'ayan *et al.*, *Science* **309**, 1078 (2005).
- [19] L. d. F. Costa and F. Silva, *J. Stat. Phys.* **125**, 841 (2006).
- [20] F. Silva *et al.*, *Inf. Sci.* **333**, 61 (2016).
- [21] Z.-P. Liu, C. Wu, H. Miao, and H. Wu, *Database* **2015**, bav095 (2015).
- [22] A. Fabregat *et al.*, *Nucleic Acids Res.* **44**, D481 (2016).
- [23] F. Kramer, M. Bayerlová, F. Klemm, A. Bleckmann, and T. Beissbarth, *Bioinformatics* **29**, 520 (2013).
- [24] NCBI, Sequence read archive, <https://www.ncbi.nlm.nih.gov/sra>, accessed: 29/05/17.
- [25] S. Ander and W. Huber, *Genome Biol.* **11**, R106 (2010).
- [26] Y. Guan, *Stat. Prob. Lett.* **79**, 1621 (2009).
- [27] X. Cui and G. A. Churchill, *Genome Biol.* **4**, 210 (2003).
- [28] Y. Benjamini and Y. Hochberg, *J. R. Stat. Soc. B* **57**, 289 (1995).
- [29] C. Giraud, *Introduction to High-Dimensional Statistics* (CRC, Boca Raton, 2015).
- [30] N. Biggs, *Algebraic Graph Theory* (Cambridge University Press, Cambridge, 1994).
- [31] Y. Xiao, M. Xiong, W. Wang, and H. Wang, *Phys. Rev. E* **77**, 066108 (2008).
- [32] P. Sah, L. Singh, A. Clauset, and S. Bansal, *BMC Bioinformatics* **15**, 220 (2014).
- [33] E. de Silva and M. Stumpf, *J. R. Soc. Interface* **2**, 419 (2005).
- [34] C. Gkantsidis, M. Mihail, and E. Zegura, in *Proceedings of the Fifth Workshop on Algorithm Engineering and Experiments*, edited by R. E. Ladner (SIAM, Philadelphia, 2003).
- [35] G. Caldarelli, R. Pastor-Satorras, and A. Vespignani, *Eur. Phys. J. B* **38**, 183 (2004).
- [36] R. Brownlie and R. Zamojska, *Nat. Rev. Immunol.* **13**, 257 (2013).
- [37] K. Oda, Y. Matsuoka, A. Funahashi, and H. Kitano, *Mol. Syst. Biol.* **1**, 2005.0010 (2005).
- [38] J. Massagué, *Nat. Rev. Mol. Cell Biol.* **1**, 169 (2000).
- [39] A. Schier and M. Shen, *Nature (London)* **403**, 385 (2000).
- [40] G. Hu, Q. Tang, S. Sharma, F. Yu, T. M. Escobar, S. A. Muljo, J. Zhu, and K. Zhao, *Nat. Immunol.* **14**, 1190 (2013).
- [41] K. Fujii, Z. Shi, O. Zhulyn, N. Denans, and M. Barna, *Nat. Commun.* **8**, 14443 (2017).
- [42] L. Yan, M. Yang, H. Guo, L. Yang *et al.*, *Nat. Struct. Mol. Biol.* **20**, 1131 (2013).
- [43] J. Jordan, E. Landau, and R. Iyengar, *Cell* **103**, 193 (2000).
- [44] N. Kashtan and U. Alon, *Proc. Natl. Acad. Sci. USA* **102**, 13773 (2005).
- [45] S. Navlakha, X. He, C. Faloutsos, and Z. Bar-Joseph, *J. R. Soc. Interface* **11**, 20140283 (2014).
- [46] M. Kaern, T. Elston, W. Blake, and J. Collins, *Nat. Rev. Genet.* **6**, 451 (2005).
- [47] J. Newman, S. Ghaemmaghami, J. Ihmels, D. Breslow, M. Noble, J. DeRisi, and J. Weissman, *Nature (London)* **441**, 840 (2006).