



Enumerating Dominant Pathways in Biological Networks by Information Flow Analysis

Ozan Kahramanoğulları^(✉) 

Department of Mathematics, University of Trento, Trento, Italy
ozan.kah@gmail.com

Abstract. Cells perceive and respond to their microenvironment as a part of their functioning via networks of processes resulting from molecular interactions. The complexity of such networks has been the subject of studies that address their various aspects. Some of these include static methods that focus on graph representations and their consequent properties, while others take a dynamical systems approach based on simulations. Here, we address the problem of identifying dominant pathways in biological networks that are represented as activation and repression edges. For this purpose, we propose a hybrid method that combines static graph properties with a dynamic quantification of information flow that results from stochastic simulations. We first illustrate our method on a simple example, and then apply it to the *Escherichia coli* transcription network consisting of 4639 regulatory edges.

Keywords: Dominant pathways · Biological networks · Information flow analysis

1 Introduction

The complex biological processes dedicated to sustaining life are commonly represented by various kinds of networks for formal analysis. These representations cover a broad spectrum from detailed rule-based models and chemical reaction networks to other stoichiometric representations as well as graph models with varying details. These different methods have their individual strengths in delivering new insights on a rich landscape of biological queries. However, more than often, the availability of empirical data, or the lack thereof, poses a bottleneck in the formal setting within the context of specific studies.

Despite the inherent complexity, experimental findings indicate that certain aspects and patterns are common in many biological networks. Some of these features resemble those in the networks that are observed outside the biochemical realm such as communication networks. By relying on these similarities, here we work with the consideration of networks of biochemical entities as processes that relay the incoming stimulus to response components. We aim at benefitting

from this notion especially when larger biological processes are considered and kinetic data is scarce or difficult to apply. In particular, we address the problem of identifying dominant pathways: within an organism, biological processes work together to produce an overall global flux distribution by making the flow of information accessible over different pathways [10, 12, 16, 20]. Despite the presence of many pathways, some of the available pathways are more dominant in terms of their information flow capacity.

In the light of the observations above, our method is based on the following assumptions on biological systems. Firstly, information propagates by a series of coupled biochemical reactions, whereby cellular signals are relayed predominantly through functional modules of highly connected nodes, see, e.g., [13]. In particular, in the extreme case of scale-free networks, whose degree distribution follows a power law [1], only a tiny fragment of hub nodes process a large fragment of the information. Secondly, local signal transduction tends to be stochastic, hence information propagation by individual components is subject to noise. And finally, in the presence of multiple pathways for the signal, information flow has a predilection for the pathway of least biochemical resistance.

Our analysis of biological systems combines methods from static graph theoretical considerations in the literature with those for dynamical systems based on simulations. We work with biological systems that can be represented as directed graphs with two kinds of edges, namely activation and repression. Many biological systems can be represented in this form as well as gene transcription networks that easily fit into this category. We map each edge of the network to a reaction of a chemical reaction network (CRN). The idea here is that each activation edge consumes the instance of the incoming signal at its source node to propagate the information flow by producing an instance of the target node. The signal can then be passed on to the next reaction. Each repression edge, on the other hand, consumes the instance of the incoming signal together with an instance of its target node if it is available. This way, it inhibits the further propagation of the signal from its target node.

In accordance with the considerations above, we resort to the idea that cellular signals are transmitted dominantly through pathways of highly connected components. We implement our method by computing the reaction rates as proxies of connectivity of the source and target nodes of the corresponding edge in the network. We instantiate the rates by borrowing three topological measures from the literature that are used to study the static properties of graphs. These are the *topological overlap measure* (TOM) [18], the Randić index (RI) [17], and the combined linkage index (CLI) [15]. For each network, we produce three different CRNs using these measures to instantiate the reaction rates, together with a control CRN that assigns 1.0 to all the rates.

The dynamic component of our method is realised by running stochastic simulations on the CRNs by using the Gillespie algorithm that implements mass action kinetics [5]. In previous work, we have developed a conservative extension of this algorithm that traces species fluxes during simulation [8, 9]. We compare the results with TOM, RI, CLI models and the control model by quantifying the

distance between their results. We quantify the flow of information in different models in terms of their simulation fluxes on the mean of repeated simulations. We then apply exhaustive breadth-first search on the resulting flux graphs to enumerate the pathways and rank them.

In the following, we illustrate our method on a simple network. We then apply it to the *E. coli* transcription network [4] with 4639 regulatory edges.¹

2 Information Flow in Biological Networks

2.1 Network Implementation

We work with networks consisting of directed graphs with two kinds of edges. Formally, an information flow network $\mathcal{G} = (\mathcal{V}, \mathcal{A}, \mathcal{I})$ is given with

- the set \mathcal{V} of vertices representing the biochemical molecules or events;
- the set \mathcal{A} of directed activation edges where the source activates the target;
- and the set \mathcal{I} of directed inhibition edges where the source inhibits the target.

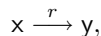
Example 1. The network depicted in Fig. 1 provides a description of a fragment of the dopamine signalling network.

The degree of a node x , denoted with $deg(x)$, is the number of edges incident to the node, with loops counted twice. For two nodes, we define $edge(x, y)$ as the number of edges, be it activation or inhibition, from x to y . We define $int(x, y)$ as the number of nodes that are connected with a single edge to both x and y .

Example 2. In the network depicted in Fig. 1, $deg(\text{GBetaGamma})=6$ and $deg(\text{PQCaCh})=2$. We have that $edge(\text{GBetaGamma}, \text{PQCaCh})=1$. Because they do not have any common neighbours, $int(\text{GBetaGamma}, \text{PQCaCh})=0$

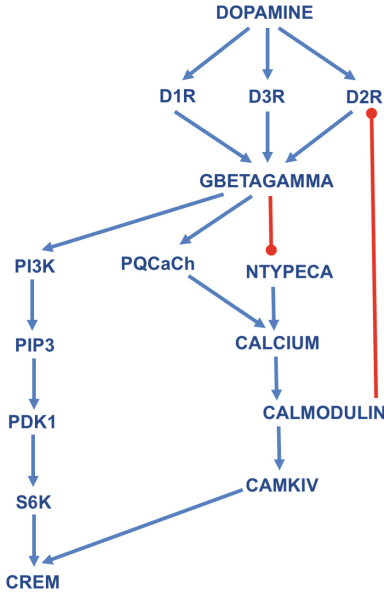
At the first step, our algorithm for computing the information flow maps the network to a chemical reaction network (CRN), whereby activation and inhibition edges are given with two different kinds of reactions.

The activation edges of the form (x, y) are mapped to reactions



which model the information flow from x to y , and r is the rate of the reaction. By relying on the notion that cellular signals are transmitted dominantly through pathways of highly connected components, we compute the reaction rates as proxies of connectivity of the source and target nodes of the corresponding edge in the network. For this, we employ three different measures from the literature.

¹ All the data and scripts are available for download at: ozan-k.com/pathways.zip.



$$\mathcal{A} = \{ (\text{Dopamine, D1R}), (\text{Dopamine, D2R}), (\text{Dopamine, D3R}), (\text{D1R, GBetaGamma}), (\text{GBetaGamma, PI3K}), (\text{GBetaGamma, PQCach}), (\text{PI3K, PIP3}), (\text{PIP3, PDK1}), (\text{PDK1, S6K}), (\text{S6K, Crem}), (\text{NTYPECA, Calcium}), (\text{Calcium, Calmodulin}), (\text{Calmodulin, Camkiv}), (\text{Camkiv, CREM}), (\text{PQCach, Calcium}), (\text{D2R, GBetaGamma}), (\text{D3R, GBetaGamma}) \}$$

$$\mathcal{I} = \{ (\text{GBetaGamma, NTYPECA}), (\text{Calmodulin, D2R}) \}$$

Fig. 1. The network given with $\mathcal{G} = (\mathcal{V}, \mathcal{A}, \mathcal{I})$, where $\mathcal{V} = \{\text{Dopamine, D1R, GBetaGama, PI3K, PIP3, PDK1, S6K, CREM, NTYPECA, Calcium, Calmodulin, Camkiv, PQCach, D2R, D3R}\}$, together with its graphical representation, whereby inhibitory edges are depicted in red and with round arrowheads. (Color figure online)

1. The *topological overlap measure* (TOM) [18], which was originally introduced to study the relationship between the network structure and the functional organisation of cellular metabolisms. We obtain the TOM rate value r as:

$$r = \frac{\text{int}(x,y)+1}{\min(\text{deg}(x), \text{deg}(y))}$$

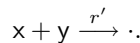
2. The Randić index (RI) [17] has been related to physical and chemical properties of organic molecules. We apply it to a single edge as follows:

$$r = \frac{1}{\sqrt{\text{deg}(x) \cdot \text{deg}(y)}}$$

3. The combined linkage index (CLI) [15] extends RI with the aim of emphasising the strongest links of each node.

$$r = \frac{\text{edge}(x,y) + \text{edge}(y,x) + 2 \cdot \text{int}(x,y)}{\sqrt{(\text{deg}(x) + 1) \cdot (\text{deg}(y) + 1)}},$$

The inhibitory edges of the form (x, y) are mapped to reactions of the form



Such a reaction models the consumption of the information at x to inhibit the further downstream flow from y to other components of the network. The reaction rate is given by $r' = r.p$, whereby r is defined as above and the constant p is the inhibition constant, which factors for these second order reactions that can have a much higher propensity in comparison to first-order activation reactions. In our analysis, we first use a default value of 1.0 for p , and also evaluate the effect of smaller values.

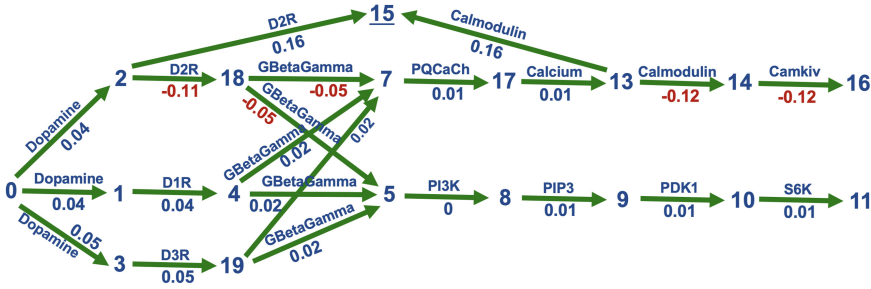
Example 3. By applying the definitions of TOM, RI and CLI to the network in Fig. 1, we obtain the CRN with the reactions listed in Fig. 2 and together with the rate values listed in Table 1.

2.2 Network Simulation

CRNs can be simulated stochastically by using Gillespie’s direct method, which is also known as the stochastic simulation algorithm (SSA) [5]. Various extensions of SSA in the literature address a variety of concerns such as increasing efficiency of simulations, simulation of rare events or others, e.g., [3, 6, 11]. In previous work [8, 9], we have presented a method that extends SSA for stochastic flux analysis of CRNs. The method, called fSSA, is a conservative extension of SSA that monitors the distribution of the network resources during simulation with respect to the causal interdependence of the reaction instances. This consideration originates from non-interleaving models of concurrent computations used in computer science [7, 14]. In such a setting, the dependencies are observed in a manner that takes into account the propensity of each reaction in terms of the resources available to that reaction. As a result of this, simulations

Table 1. The CRN obtained from the network depicted in Fig. 1, and its r values according to TOM, RI and CLI. In the simulations, we have varied the inhibitory constant p between 10^{-4} and 1.0. The resulting flux graph is depicted in Fig. 2 and normalised flux values for different p values are listed in Table 6.

| Reactions | TOM | RI | CLI | Reactions | TOM | RI | CLI |
|-----------|--------|--------|--------|-----------|--------|--------|--------|
| 1 | 0.5 | 0.41 | 0.29 | 11 | 0.5 | 0.5 | 0.33 |
| 2 | 0.33 | 0.33 | 0.25 | 12 | 0.5 | 0.41 | 0.29 |
| 3 | 0.5 | 0.41 | 0.29 | 13 | 0.33 | 0.33 | 0.25 |
| 4 | 0.5 | 0.29 | 0.22 | 14 | 0.5 | 0.41 | 0.29 |
| 5 | 0.5 | 0.29 | 0.22 | 15 | 0.33.p | 0.33.p | 0.25.p |
| 6 | 0.22.p | 0.29.p | 0.44.p | 16 | 0.5 | 0.5 | 0.33 |
| 7 | 0.5 | 0.29 | 0.22 | 17 | 0.5 | 0.41 | 0.29 |
| 8 | 0.5 | 0.5 | 0.33 | 18 | 0.33 | 0.24 | 0.19 |
| 9 | 0.5 | 0.5 | 0.33 | 19 | 0.5 | 0.29 | 0.22 |
| 10 | 0.5 | 0.5 | 0.33 | | | | |



- | | | |
|---------------------|--------------------------|-------------------------|
| 0 → 100000 Dopamine | 6 GBetaGamma + NTYPECA → | 13 Calcium → Calmodulin |
| 1 Dopamine → D1R | 7 GBetaGamma → PQCaCh | 14 Calmodulin → Camkiv |
| 2 Dopamine → D2R | 8 PI3K → PIP3 | 15 Calmodulin + D2R → |
| 3 Dopamine → D3R | 9 PIP3 → PDK1 | 16 Camkiv → CREM |
| 4 D1R → GBetaGamma | 10 PDK1 → S6K | 17 PQCaCh → Calcium |
| 5 GBetaGamma → PI3K | 11 S6K → CREM | 18 D2R → GBetaGamma |
| | 12 NTYPECA → Calcium | 19 D3R → GBetaGamma |

Fig. 2. The fluxes of the network in Fig. 1, delivered by the simulations with the CRN above and the rates listed in Table 1. The simulations are initiated with 100000 Dopamine as the system input. The node numbers are the CRN reactions. The inhibition reaction 15 is indicated with an underline. The complete flux data with TOM, RI and CLI is given in Table 6. The numbers on the edges summarise the data: each number denotes the maximum difference in normalised flux resulting from increasing the inhibition constant from $p = 10^{-4}$ to $p = 1.0$ in all cases.

resulting from our algorithm provide a quantitative view of the flow of information in the network besides the usual time series information. The flux graphs, that are output by the algorithm, reflect what fragment of system resources flow through which pathways of the network. This kind of information becomes particularly significant when a system resource is produced or consumed by multiple components. In this regard, flux graphs display which components produce and consume such resources. For example, in the network above, GBetaGamma production and consumption can follow many different pathways in the network.

We use the fSSA algorithm to run simulations on the CRNs. During these simulations, flux graphs can be obtained for the whole simulation interval as well as for arbitrary time intervals. In contrast to similar considerations with ordinary differential equations, these time intervals can be transient intervals, whereby the system has not yet reached its steady state levels, given by the ordinary differential equation simulations. Because the flow of resources can take different pathways at different intervals of the simulation, such a capability is essential for analysing the system behaviour at different stages.

For the example network in Fig. 1, we have obtained the flux graph depicted in Fig. 2. The measures described above, that is, TOM, RI and CLI, result in different reaction rates, listed in Table 1, thus they result in different values for the fluxes. However they all result in the same topology depicted in Figs. 2 and 3.

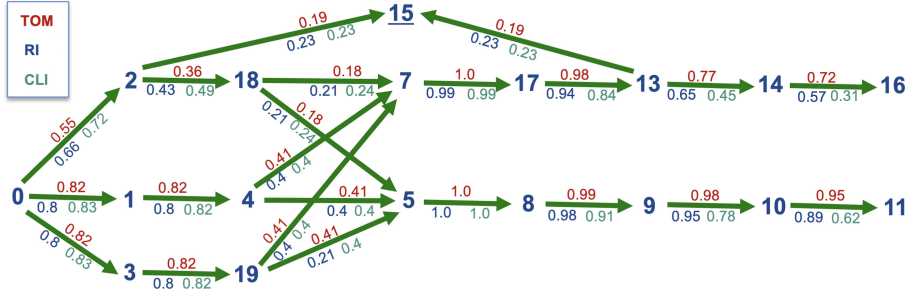


Fig. 3. The fluxes in Fig. 2 with TOM (red), RI (blue) and CLI (green) measures. (Color figure online)

Due to stochasticity, each simulation with the same CRN produces slightly different fluxes. For a systematic comparison that takes into account these variations as well as the effect of the different measures, we have first set a control network, where all the reaction rates are set to 1.0. With the inclusion of this network, we have obtained four different networks; three given by TOM, RI, and CLI, and a control network. For each one of these four networks, we have run 10 simulations. For each flux edge in a network, we computed the mean of 10 simulations, and then normalised these mean fluxes according to the maximum flux of each network.

Our simulations resulted in the normalised flux values listed in Table 6, where we have considered a spectrum of inhibition constants. Figure 2 provides a summary of the data in Table 6 with respect to the effect of varying inhibition constant from $p = 10^{-4}$ to $p = 1.0$. We observe that the inhibition constant p does not have a significant impact in general. More interestingly, the variations in p affect the versions of CRN that are instantiated with different measures similarly. Most of the fluxes are affected to an extent of 0.04% of maximum flux, and the greatest effect is on the fluxes that feed reaction 15 or compete with these fluxes, which however do not exceed 16% even with $p = 10^{-4}$, and these greater effects are pronounced at the lower end of the spectrum. Figure 3 displays the fluxes with $p = 10^{-2}$ for TOM, RI and CLI measures.

To compare the impact of the different measures on the simulations and the resulting fluxes, we have computed the distance between the results with different networks. We define this as the sum of squared distances between normalised fluxes. That is, given that \mathcal{F}_1 and \mathcal{F}_2 are flux graphs as in Fig. 3, for each flux edge from a reaction x to reaction y with w_1 in \mathcal{F}_1 and w_2 in \mathcal{F}_2 , we compute the sum of the values $(w_1 - w_2)^2$. If a flux edge does not exist, its weight is 0.

$$\sum_{\substack{w_1 \in \mathcal{F}_1 \\ w_2 \in \mathcal{F}_2}} (w_1 - w_2)^2$$

For this network, we observe in Table 2 that the inhibition constant does not play a significant role in distinguishing the effect of different measures,

Table 2. The distances between flux graphs for the Dopamine network with the measures TOM, RI, CLI and control, which assigns the rate $r = 1.0$ to all the reactions.

| p | TOM & 1.0 | RI & 1.0 | CLI & 1.0 | TOM & RI | TOM & CLI | RI & CLI |
|-----------|-----------|----------|-----------|----------|-----------|----------|
| 10^{-4} | 0.24 | 0.76 | 0.14 | 0.19 | 0.65 | 1.42 |
| 10^{-2} | 0.24 | 0.78 | 0.15 | 0.2 | 0.66 | 1.47 |
| 1.0 | 0.24 | 0.8 | 0.14 | 0.21 | 0.64 | 1.46 |

Table 3. The ranking (#) of pathways delivered by measures RI, TOM, CLI and 1.0.

| RI # | Pathway | RI flux | TOM # | CLI # | 1.0 # |
|------|------------------------------------|---------|-------|-------|-------|
| 1 | 0 → 3 → 19 → 5 → 8 → 9 → 10 → 11 | 0.83 | 1 | 1 | 1 |
| 2 | 0 → 1 → 4 → 5 → 8 → 9 → 10 → 11 | 0.83 | 2 | 2 | 2 |
| 3 | 0 → 3 → 19 → 7 → 17 → 13 → 14 → 16 | 0.74 | 3 | 6 | 5 |
| 4 | 0 → 1 → 4 → 7 → 17 → 13 → 14 → 16 | 0.74 | 4 | 7 | 4 |
| 5 | 0 → 2 → 18 → 5 → 8 → 9 → 10 → 11 | 0.73 | 5 | 5 | 3 |
| 6 | 0 → 3 → 19 → 7 → 17 → 13 → 15 | 0.69 | 6 | 3 | 8 |
| 7 | 0 → 1 → 4 → 7 → 17 → 13 → 15 | 0.69 | 7 | 4 | 7 |
| 8 | 0 → 2 → 18 → 7 → 17 → 13 → 14 → 16 | 0.64 | 8 | 9 | 6 |
| 9 | 0 → 2 → 18 → 7 → 17 → 13 → 15 | 0.58 | 9 | 8 | 9 |
| 10 | 0 → 2 → 15 → | 0.45 | 10 | 10 | 10 |

which confirms our observations above. We observe that RI measure provides the greatest distinction from the control network and CLI provides the smallest distinction. The much larger distance between CLI and RI confirms this observation. Moreover, RI and TOM appear similar. Based on these observations, Table 3 enumerates the flux pathways by ranking them according to their mean fluxes, where RI measure is used as reference. As indicated by the observations in Table 2, all measures agree on the first two rankings and RI and TOM have the same rankings, which are different from those with the control network.

3 A Case Study: *Escherichia coli* Transcription Network

We have applied our method to the *Escherichia coli* transcription network version 10.5 reported in the RegulonDB [4] with the date 13 September 2018, which is depicted in Fig. 4. In this network, the distribution of the nodes with respect to their frequency in regulations follows a power-law, whereby 1610 of the 1886 proteins participate in not more than 5 regulations. As listed in Table 7, CRP has the highest frequency as it participates in 585 regulations, followed by FNR with 322, IHF with 259 and H-NS with 195 regulations.

We applied the four measures given by TOM, RI and CLI as well as the control model with all the rates set to 1.0, and considered the inhibition constants

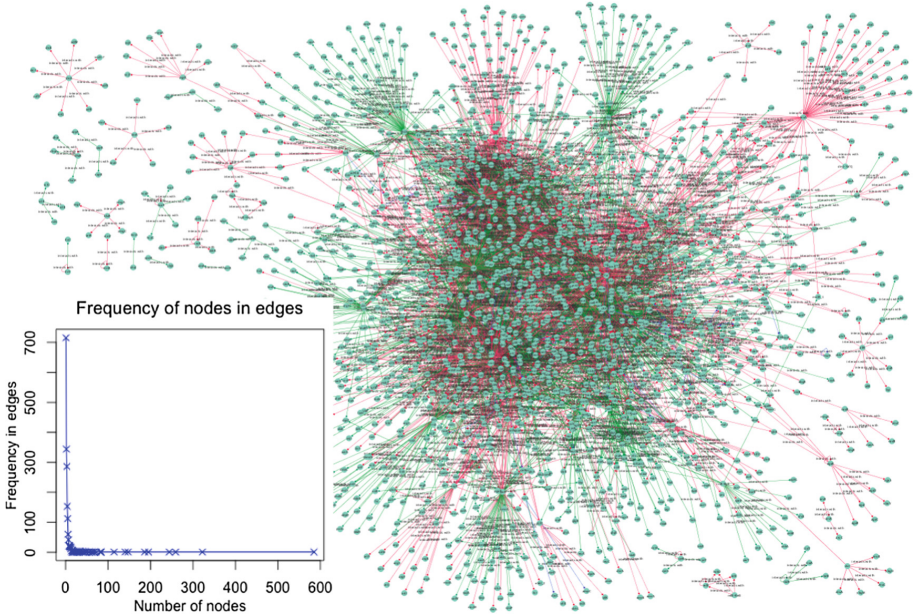


Fig. 4. Escherichia coli transcription network as reported in [4], rendered by Cytoscape [19]. The network consists of 1886 nodes, which are regulated by 4639 edges. The 2338 activation edges are denoted by green, whereas the 2301 repression edges are denoted by red. 207 of the nodes are transcription factors and 63 are at the root position. The graph in the corner displays the frequency of nodes in the edges. (Color figure online)

$p = 10^{-2}$ and $p = 1.0$, and this way obtained 8 different versions of the model. Each of these networks consist of 4639 reactions with 1886 species. We focused our investigation on the pathways initiated by CRP as this transcription factor has the highest frequency among all the 63. We performed 10 simulations with an initial value of 10000 CRP molecules for all of the 8 cases. Each of these simulations resulted in 1000 to 1500 flux edges. For each case, we took the mean of each flux edge given by the 10 simulations. We then normalised each one of the 8 flux graphs with the maximum flux in that graph.

For a comparison, we first computed the squared distance between the 8 flux graphs. To emphasise the effect of different measures, we have taken the sum of the fluxes for each species in flux graphs and computed the squared distance on these sums. The differences between TOM, RI, CLI and 1.0 model for each of the $p = 10^{-2}$ and $p = 1.0$ values are listed in Table 4. The differences between $p = 10^{-2}$ and $p = 1.0$ for each of TOM, RI and CLI and 1.0 are listed in Table 5. We observe in Tables 4 and 5 that the inhibition constant does not play a significant role in distinguishing the effect of different measures as before with the exception of 1.0 network. This observation confirms that the rates provided by the measures plays a more significant role in determining the fluxes in comparison to the inhibition constant. However, in the control model, the inhibition constant plays a greater role in determining the system behaviour.

Table 4. The distances between flux graphs of the E. coli network with the measures TOM, RI, CLI and control, which assigns the rate $r = 1.0$ to all the reactions.

| p | TOM & 1.0 | RI & 1.0 | CLI & 1.0 | TOM & RI | TOM & CLI | RI & CLI |
|-----------|-----------|----------|-----------|----------|-----------|----------|
| 1.0 | 17620 | 21417 | 91637 | 191 | 28897 | 24470 |
| 10^{-2} | 19579 | 24360 | 104359 | 267 | 33543 | 27897 |

Table 5. The distances between flux graphs of the E. coli network with the inhibition constants $p = 1.0$ and $p = 10^{-2}$.

| p | TOM | RI | CLI | 1.0 |
|-----------------|-----|-----|-----|--------|
| 1.0 & 10^{-2} | 282 | 204 | 13 | 106662 |

Table 4 indicates that CLI measure provides the greatest distinction from the control model and TOM provides the smallest distinction. The large distance between TOM and CLI and the one between RI and CLI as well as the much smaller distance between RI and TOM confirm this observation.

The different measures resulted in different numbers of pathways. With $p = 0.01$, CLI has generated 1388 pathways, whereas RI has generated 998, TOM has generated 1340 and the control network has generated 1535. With $p = 1.0$, CLI has generated 1285 pathways, whereas RI has generated 1210 pathways, TOM has generated 1251, and the control network has generated 1535 pathways. The resulting list of pathways for all the 8 cases can be downloaded together with all the scripts that are used to apply the methods above.²

4 Discussion

We have proposed a method for enumerating dominant pathways in biological networks that can be represented as directed graphs consisting of activation and repression edges. Our analysis combines methods from static graph theoretical considerations in the literature with those for dynamical systems, based on simulations. Our method emphasises the inherent stochasticity in biological processes as well as the notion that cellular signals are relayed predominantly through highly connected nodes and pathways of least biochemical resistance.

The stochastic simulations in our examples result in individual simulation trajectories that expose the noise in the system. The notion of stochastic flux, delivered by these simulations, provides a direct quantification of information flow, for any time interval, including the transient states. However, averaging over many simulations as in the examples above dampens the stochastic noise. If a deterministic notion of information flow can be characterised, linear noise approximation simulations [2] or deterministic ODE simulations can be considered for the analysis of the systems where the stochastic noise is less of a concern.

² ozan-k.com/pathways.zip.

Table 6. Mean normalised fluxes obtained from 10 simulations for each CRN in Table 1 instantiated with rates given by TOM, RI, CLI and 1.0, and with inhibition constants of $p = 10^{-4}$, $p = 10^{-2}$ and 1.0. The table thus summarises 120 simulations.

| Fluxes / p | TOM | | | RI | | | CLI | | | 1.0 | | |
|--------------|-----------|-----------|------|-----------|-----------|------|-----------|-----------|------|-----------|-----------|------|
| | 10^{-4} | 10^{-2} | 1.0 | 10^{-4} | 10^{-2} | 1.0 | 10^{-4} | 10^{-2} | 1.0 | 10^{-4} | 10^{-2} | 1.0 |
| 0 → 1 | 0.78 | 0.82 | 0.82 | 0.76 | 0.8 | 0.8 | 0.79 | 0.83 | 0.83 | 0.69 | 0.73 | 0.73 |
| 0 → 2 | 0.52 | 0.55 | 0.55 | 0.62 | 0.66 | 0.66 | 0.69 | 0.72 | 0.72 | 0.69 | 0.73 | 0.73 |
| 0 → 3 | 0.78 | 0.82 | 0.83 | 0.76 | 0.8 | 0.8 | 0.79 | 0.83 | 0.83 | 0.69 | 0.73 | 0.73 |
| 1 → 4 | 0.78 | 0.82 | 0.82 | 0.76 | 0.8 | 0.8 | 0.78 | 0.82 | 0.82 | 0.69 | 0.73 | 0.73 |
| 2 → 15 | 0.07 | 0.19 | 0.2 | 0.11 | 0.23 | 0.23 | 0.11 | 0.23 | 0.24 | 0.08 | 0.19 | 0.2 |
| 2 → 18 | 0.44 | 0.36 | 0.35 | 0.51 | 0.43 | 0.42 | 0.58 | 0.49 | 0.48 | 0.61 | 0.54 | 0.54 |
| 3 → 19 | 0.78 | 0.82 | 0.83 | 0.76 | 0.8 | 0.8 | 0.78 | 0.82 | 0.82 | 0.69 | 0.73 | 0.73 |
| 4 → 5 | 0.39 | 0.41 | 0.41 | 0.38 | 0.4 | 0.4 | 0.38 | 0.4 | 0.4 | 0.35 | 0.37 | 0.37 |
| 4 → 7 | 0.39 | 0.41 | 0.41 | 0.38 | 0.4 | 0.4 | 0.38 | 0.4 | 0.4 | 0.35 | 0.36 | 0.37 |
| 5 → 8 | 1.0 | 1.0 | 1.0 | 1.0 | 1.0 | 1.0 | 1.0 | 1.0 | 1.0 | 1.0 | 1.0 | 1.0 |
| 7 → 17 | 1.0 | 1.0 | 1.0 | 1.0 | 0.99 | 1.0 | 0.98 | 0.99 | 0.99 | 1.0 | 1.0 | 1.0 |
| 8 → 9 | 0.99 | 0.99 | 1.0 | 0.98 | 0.98 | 0.98 | 0.91 | 0.91 | 0.91 | 1.0 | 1.0 | 1.0 |
| 9 → 10 | 0.98 | 0.98 | 0.98 | 0.94 | 0.95 | 0.95 | 0.77 | 0.78 | 0.78 | 1.0 | 1.0 | 1.0 |
| 10 → 11 | 0.95 | 0.95 | 0.95 | 0.88 | 0.89 | 0.89 | 0.61 | 0.62 | 0.62 | 1.0 | 1.0 | 1.0 |
| 13 → 14 | 0.88 | 0.77 | 0.76 | 0.77 | 0.65 | 0.65 | 0.57 | 0.45 | 0.45 | 0.92 | 0.81 | 0.8 |
| 13 → 15 | 0.07 | 0.19 | 0.2 | 0.11 | 0.23 | 0.23 | 0.11 | 0.23 | 0.24 | 0.08 | 0.19 | 0.2 |
| 14 → 16 | 0.83 | 0.72 | 0.71 | 0.69 | 0.57 | 0.57 | 0.42 | 0.31 | 0.3 | 0.92 | 0.81 | 0.8 |
| 17 → 13 | 0.98 | 0.98 | 0.98 | 0.94 | 0.94 | 0.95 | 0.84 | 0.84 | 0.85 | 1.0 | 1.0 | 1.0 |
| 18 → 5 | 0.22 | 0.18 | 0.17 | 0.26 | 0.21 | 0.21 | 0.29 | 0.24 | 0.24 | 0.31 | 0.27 | 0.27 |
| 18 → 7 | 0.22 | 0.18 | 0.17 | 0.26 | 0.21 | 0.21 | 0.29 | 0.24 | 0.24 | 0.31 | 0.27 | 0.27 |
| 19 → 5 | 0.39 | 0.41 | 0.41 | 0.38 | 0.4 | 0.4 | 0.38 | 0.4 | 0.4 | 0.35 | 0.36 | 0.37 |
| 19 → 7 | 0.39 | 0.41 | 0.41 | 0.38 | 0.4 | 0.4 | 0.38 | 0.4 | 0.4 | 0.35 | 0.37 | 0.37 |

As evidenced by our case study on *E. coli* network, the measures, TOM, RI, and CLI have a significant effect on determining the dominant pathways. In this regard, a more extensive evaluation of these measures as well as others in the literature is a topic of further investigation. Moreover, the ranking of the pathways is subject to parameters such as pathway length and flux strengths at various segments, which can change the ranking. An evaluation of these parameters in the context of biological evidence for the *E. coli* network and in applications to other large networks are topics of future work.

Table 7. The frequency of transcription factors (TF) in the E. coli network.

| Freq. | TF | Freq. | TF | Freq. | TF | Freq. | TF | Freq. | TF | Freq. | TF |
|-------|-------|-------|------|-------|-------|-------|-----------|-------|-----------|-------|-----------|
| 585 | CRP | 259 | IHF | 84 | NsrR | 66 | NarP | 46 | ModE | 16 | RcsAB |
| 322 | FNR | 195 | H-NS | 82 | FlhDC | 56 | NtrC | 34 | SlyA | 16 | rbsB-BglJ |
| 13 | HypT | 12 | GntR | 11 | NanR | 9 | HU | 9 | AllR | 8 | MatA |
| 13 | BasR | 12 | HprR | 9 | NrdR | 9 | RcdA | 9 | GadE-RcsB | 8 | PgrR |
| 8 | DicA | 7 | UlaR | 7 | DeoR | 6 | GatR | 6 | Zur | 5 | AscG |
| 7 | TdcR | 7 | CueR | 6 | GlrR | 6 | SdiA | 5 | CsiR | 5 | MntR |
| 5 | HipAB | 4 | BluR | 4 | KdpE | 4 | NadR | 3 | DinJ-YafQ | 3 | MazE-MazF |
| 5 | BirA | 4 | McbR | 4 | AtoC | 3 | RclR | 3 | relB-RelE | 2 | RqhC |
| 2 | FabR | 2 | EnvR | 2 | KdgR | 2 | RtcR | 2 | XapR | 1 | YpdB |
| 2 | EnvY | 2 | EbgR | 2 | BCCP | 2 | HigB-HigA | 2 | YefM-YoeB | 1 | ZntR |
| 1 | UhpA | 1 | BtsR | 1 | YhaJ | | | | | | |

References

1. Albert, R.: Scale-free networks in cell biology. *J. Cell Sci.* **118**, 4947–4957 (2005)
2. Cardelli, L., Kwiatkowska, M., Laurenti, L.: Stochastic analysis of chemical reaction networks using linear noise approximation. *Biosystems* **149**, 26–33 (2016)
3. Erhard, F., Friedel, C.C., Zimmer, R.: FERN: a Java framework for stochastic simulation and evaluation of reaction networks. *BMC Bioinform.* **9**, 356 (2008)
4. Gama-Castro, S., et al.: RegulonDB version 9.0: high-level integration of gene regulation, coexpression, motif clustering and beyond. *Nucleic Acids Res.* **44**(D1), D133–D143 (2016)
5. Gillespie, D.T.: Exact stochastic simulation of coupled chemical reactions. *J. Phys. Chem.* **81**(25), 2340–2361 (1977)
6. Gillespie, D.T.: Approximate accelerated stochastic simulation of chemically reacting systems. *J. Chem. Phys.* **115**(4), 1716 (2001)
7. Kahramanoğulları, O.: On linear logic planning and concurrency. *Inf. Comput.* **207**, 1229–1258 (2009)
8. Kahramanoğulları, O.: Quantifying information flow in chemical reaction networks. In: Figueiredo, D., Martín-Vide, C., Pratas, D., Vega-Rodríguez, M.A. (eds.) *AlCoB 2017*. LNCS, vol. 10252, pp. 155–166. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-58163-7_11
9. Kahramanoğulları, O., Lynch, J.: Stochastic flux analysis of chemical reaction networks. *BMC Syst. Biol.* **7**, 133 (2013)
10. Khatri, P., Sirota, M., Butte, A.J.: Ten years of pathway analysis: current approaches and outstanding challenges. *PLoS Comput. Biol.* **8**(2), e1002375 (2012)
11. Kuwahara, H., Mura, I.: An efficient and exact stochastic simulation method to analyze rare events in biochemical systems. *J. Chem. Phys.* **129**(16), 10B619 (2008)
12. Ma, S., Jiang, T., Jiang, R.: Differential regulation enrichment analysis via the integration of transcriptional regulatory network and gene expression data. *Bioinformatics* **31**(4), 563–571 (2015)
13. Ma’ayan, A., et al.: Formation of regulatory patterns during signal propagation in a Mammalian cellular network. *Science* **309**(5737), 1078–83 (2005)

14. Nielsen, M., Plotkin, G., Winskel, G.: Event structures and domains, part 1. *Theor. Comput. Sci.* **5**(3), 223–256 (1981)
15. Persson, O.: Identifying research themes with weighted direct citation links. *J. Informetr.* **4**(3), 415–422 (2010)
16. Planes, F.J., Beasley, J.E.: A critical examination of stoichiometric and path-finding approaches to metabolic pathways. *Brief. Bioinform.* **9**(5), 422–436 (2008)
17. Randić, M.: Characterization of molecular branching. *J. Am. Chem. Soc.* **97**(23), 6609–6615 (1975)
18. Ravasz, E., et al.: Hierarchical organization of modularity in metabolic networks. *Science* **297**, 1551–1555 (2002)
19. Shannon, P., et al.: Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res.* **13**(11), 2498–504 (2003)
20. Zubarev, R.A., et al.: Identification of dominant signaling pathways from proteomics expression data. *J. Proteomics* **71**(1), 89–96 (2008)