Check for updates

SOFTWARE TOOL ARTICLE

REVISED Simulation and visualization of multiple KEGG pathways using BioNSi [version 2; peer review: 2 approved, 1 approved with reservations]

Adva Yeheskel¹, Adam Reiter², Metsada Pasmanik-Chor¹, Amir Rubinstein ²

¹Bioinformatics unit, Faculty of Life Science, Tel Aviv University, Tel Aviv, Israel

²School of Computer Science, Tel Aviv University, Tel Aviv, Israel

v2 First published: 11 Dec 2017, 6:2120
<https://doi.org/10.12688/f1000research.13254.1>

Latest published: 14 May 2018, 6:2120
<https://doi.org/10.12688/f1000research.13254.2>

Abstract

Motivation: Many biologists are discouraged from using network simulation tools because these require manual, often tedious network construction. This situation calls for building new tools or extending existing ones with the ability to import biological pathways previously deposited in databases and analyze them, in order to produce novel biological insights at the pathway level.

Results: We have extended a network simulation tool (BioNSi), which now allows merging of multiple pathways from the KEGG pathway database into a single, coherent network, and visualizing its properties. Furthermore, the enhanced tool enables loading experimental expression data into the network and simulating its dynamics under various biological conditions or perturbations. As a proof of concept, we tested two sets of published experimental data, one related to inflammatory bowel disease condition and the other to breast cancer treatment. We predict some of the major observations obtained following these laboratory experiments, and provide new insights that may shed additional light on these results.

Tool requirements: Cytoscape 3.x, JAVA 8

Availability: The tool is freely available at <http://bionsi.wix.com/bionsi>, where a complete user guide and a step-by-step manual can also be found.

Keywords

biological network, simulation, gene expression, KEGG, Cytoscape, BioNSi

Open Peer Review

Approval Status

	1	2	3
version 2			
(revision)
14 May 2018	view		view

version 1
11 Dec 2017	view	view	

1. **Aurélien Naldi** , École Normale Supérieure (ENS), Paris, France
 2. **Giovanni Scardoni**, University of Verona, Verona, Italy
 3. **Ester Feldmesser** , Weizmann Institute of Science, Rehovot, Israel
- Bareket Dassa**, Weizmann Institute of Science, Rehovot, Israel

Any reports and responses or comments on the article can be found at the end of the article.

Corresponding author: Amir Rubinstein (amirr@tau.ac.il)

Author roles: **Yeheskel A:** Conceptualization, Data Curation, Investigation, Methodology, Writing – Review & Editing; **Reiter A:** Software, Writing – Review & Editing; **Pasmanik-Chor M:** Conceptualization, Data Curation, Investigation, Methodology, Writing – Original Draft Preparation, Writing – Review & Editing; **Rubinstein A:** Conceptualization, Methodology, Writing – Original Draft Preparation, Writing – Review & Editing

Competing interests: No competing interests were disclosed.

Grant information: The author(s) declared that no grants were involved in supporting this work.

Copyright: © 2018 Yeheskel A *et al.* This is an open access article distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited. Data associated with the article are available under the terms of the [Creative Commons Zero "No rights reserved" data waiver](#) (CC0 1.0 Public domain dedication).

How to cite this article: Yeheskel A, Reiter A, Pasmanik-Chor M and Rubinstein A. **Simulation and visualization of multiple KEGG pathways using BioNSi [version 2; peer review: 2 approved, 1 approved with reservations]** F1000Research 2018, 6:2120 <https://doi.org/10.12688/f1000research.13254.2>

First published: 11 Dec 2017, 6:2120 <https://doi.org/10.12688/f1000research.13254.1>

REVISED Amendments from Version 1

Some details were clarified and emphasized in response to the referees' reports. The following changes were made in the 2nd version:

- Additional references added to the Introduction.
- The following was emphasized: BioNSi preserves all the alternative names for a KEGG entity. When merging pathways, even if two nodes have different alternative names, they will be fused into a single node. Furthermore, names of entities in the expression level file are matched against any of the alternative names of a node in the network.
- An explanation about the scaling of nodes' states to the range 0–9 was added.
- An explanation about the meaning of steady states in the 2 test cases was added to the Methods section (*Simulation*).
- Typo fixed: BooleSim instead of Boolsim.
- We changed "asynchronous" to "non-simultaneous", and added this column to [Table 1](#).
- We added a brief description to [Box 1](#) about additional regulation types in BioNSi.
- We deleted the redundant paragraph from the Results.

See referee reports

Introduction

Modeling and simulation of biological regulatory networks is becoming an integral part of biological research nowadays¹. Such networks consist of nodes, representing, *e.g.*, proteins, RNA, genes, nutrients, signals, etc. Edges represent the interactions between nodes, such as activation, inhibition, phosphorylation, etc. Simulation tools allow the analysis and visualization of the dynamics of these networks. Biological data, such as protein interactions and gene or protein expression are gathered and provide the required input for such tools. Indeed, in recent years many bioinformatics tools aiming at the simulation of biological networks have been published. The purpose of such tools is to elucidate the relationships between genes, proteins, pathways, or other biological entities involved, to shed new light on the experimental results, and to suggest possible directions for future research.

Simulation tools come in various flavors. Some focus on transcriptional regulatory networks, others on signal propagation or metabolic pathways. There are tools that require programming skills, while others provide a complete graphical user interface (GUI). The underlying mathematical models differ a lot too. Perhaps the most fundamental distinction is between continuous, Boolean and discrete models. Continuous network models typically apply differential equations, using real (as opposed to discrete) numbers to represent the system's variables (*e.g.*, 2), while in Boolean network models variables may assume only one of two values, namely 0 ("OFF", or *inactive*) and 1 ("ON", or *active*) (*e.g.*, 3). Between the two "extremes" are discrete models, in which values are taken from a range of integers (*e.g.*, 0, 1, 2, ...9). This is often a good compromise between the expressiveness of the model and its simplicity, and

it is especially useful when only partial or imprecise biological knowledge of these systems is available⁴.

[Table 1](#) provides a summary of several network modeling and simulation tools properties. These tools are relatively new ones (most of them published within the last four years). In addition, we focused on biologist-friendly tools that require very little computational background and in particular no script writing skills. The last tool in [Table 1](#) is BioNSi - Biological Network Simulator⁵, whose extensions are the focus of this paper, as will be described later.

In [Table 1](#) we compare the following features:

- Web/App: whether the tool is web-based (web) or a stand-alone software or application (app)
- Nodes' levels: the granularity of the model variables, that is, whether nodes levels are Boolean (B), discrete (D) or continuous (C) variables
- Asynchronous update: whether the node level transition rules can be applied asynchronously, rather than simultaneously (for example, by specifying delays on edges, or defining priorities on edges)
- Batch simulation mode: whether the tool supports a batch mode of simulation, in which multiple initial conditions are explored simultaneously and global (emergent) properties such as the system steady states (attractors) are identified
- Edge weights: whether the edges in the network have weights to represent strength (amplitude) of interaction, in contrast to merely +1 (activation) and -1 (inhibition)
- Network generation: which format is supported for network generation - graphical (G), in which the network is drawn on the canvas, or textual (T), in which the user generates a table of nodes and edges
- Merge pathways: whether the tool supports merging of multiple pathways of interest into a single network.
- Upload expression data: whether expression data from *e.g.*, laboratory experiments can be loaded into the network automatically.

Network simulation tools typically require manual network construction based on biological data. For networks beyond a very small scale of a few nodes, this process becomes tedious, time consuming, and error prone. Furthermore, conducting numerous simulations of a network under different biological conditions may require a tiresome and monotonous process of modifying network parameters for each set of conditions. This contributes to the discouragement of many biologists from integrating network simulation tools in their daily laboratory routine. This situation is extremely unfortunate, since a lot of network data based on lab experiments already exists in various common databases. Therefore, the integration of network simulation tools with biological data previously deposited into those databased provides novel opportunities for more comprehensive

Table 1. A representative list of free, network simulation, tools for non-programming biologists. A representative list of seven network simulation tools was obtained from [OmicsTools](#). A summary of some important features is shown. V: denotes that a specific feature exists in the tool. Web/App: a web tool or downloadable software. Nodes levels: Boolean (B)/discrete (D)/continuous (C). Non-simultaneous update: defining delays on edges or priorities on nodes, which inflict order on the update of nodes. Asynchronous update: non-deterministic behavior which enables several successors to a single network state. Batch simulation mode: running a collection of simulation with various parameters to reveal emergent network properties. Edges weights: weights represent strength of interaction. Network generation: textual (T)/graphical (G). A star (*) represents features implemented in the extended BioNSi tool (but not in the original version).

Tool (ref)	Web/App	Nodes levels	Non simultaneous update	Asynchronous update	Batch simulation mode	Edges weights	Network generation	Merge pathways	Upload expression data
SimBoolNet ⁶	app	C			V	V	T		
GINsim ⁷	app	D	V	V	V	V	G		
SQUAD ⁸	app	B+C			V		T+G		
ViSiBool ⁹	app	B	V		V		T+G		
ANIMO ¹⁰	app+web	D	V				T+G		
BooleSim ¹¹	web	B					T+G		
BioNSi ⁵	app	D	V		V	V	T+G	*	*

studies of these networks. Initiatives aiming at such integration may support biological understanding of previous laboratory results, and direct researchers to design new experiments.

There are several existing tools for the generation of mathematical models from KEGG pathway database information¹² (e.g., Büchel *et al.*¹³, Wrzodek *et al.*¹⁴). We aim at integrating a recently published Cytoscape application, BioNSi (Biological Network Simulator⁵) with KEGG, exploiting this convenient simulation environment. For the sake of self-containment of this paper, [Box 1](#) describes BioNSi at a high level. BioNSi was originally used for visualizing biological networks and simulating their time course behavior. A full user guide and a step by step manual can be found on the tool's website (<http://bionsi.wix.com/bionsi>).

Our new developments enable merging of multiple KEGG pathways into a single coherent network under the BioNSi framework. Although the files downloaded from KEGG may omit some protein-protein interactions details, they were found to be useful in presenting networks based on the researcher's interest. Furthermore, uploading experimental expression data enables conducting a simulation of the network. The initial expression levels of network nodes (genes, proteins, etc.) may be obtained either from in-house experiments or from various expression databases (e.g., GEO¹⁵ or ArrayExpress¹⁶). This allows easy simulation of the resulting network under different experimental conditions (expression values). For example, one can simulate the effect of inhibitors, activators, mutations, abnormal cell conditions or the effect of a new drug on the network's behavior, suggesting further experimental exploration. BioNSi is a Cytoscape application¹⁷. Cytoscape is an open source bioinformatics software platform, which hosts various applications for visualization and analysis of networks. The benefit of

this is that the BioNSi simulation results can be further extended with additional analyses based on other applications. The typical workflow with the new version of BioNSi is presented in [Figure 1](#).

We first describe the details of our new extensions to BioNSi. Then, as a proof of concept, we demonstrate the advantages and usability of our tool by presenting the simulation and analysis of previously published experimental data from two sets of experiments: (1) the characterization of inflammatory bowel disease (IBD) conditions, and (2) treatment of MCF7 breast cancer cells with heregulin (HRG). These two test cases may provide insights into how our tool can be used to elucidate existing biological data, raise new hypotheses and suggest new lab experiments to test them.

Methods

Implementation

Cytoscape is an open source software platform for complex network analyses and visualization¹⁷. It enables “plugging in” in the form of dedicated software called applications (Apps), which are developed by users. BioNSi⁵ is one such application that provides a user friendly tool for simulating the dynamics of biological regulatory networks. It is implemented under Cytoscape version 3. Installation is easy and explained in the tool's website at <http://bionsi.wix.com/bionsi>. We have extended BioNSi's capabilities with two key features presented in the next sections.

Operation

Importing and merging multiple KEGG pathways into BioNSi. The original version of BioNSi allowed creating networks “*de novo*”, by adding nodes and edges manually, according to the biological knowledge of the researcher. This

Box 1. BioNSi - A high level description

BioNSi is a biological network simulation tool⁵. It allows constructing a regulatory network and simulating its behavior. The network consists of *nodes* (interactors) and *edges* (interactions). For example, nodes may represent genes, RNA or transcription factors, while edges represent interactions such as transcription, translation, phosphorylation, etc.

For each node, the user manually defines an initial expression level, termed *initial state*, an integer typically taken from the range between 0 and 9 (0 – no activity, 9 – high activity). Time in the model is also discrete, represented as clock “tics” or time steps ($t = 0, 1, 2, \dots$). Nodes’ states are updated in each time step by a transition rule (see below).

Each edge is assigned a weight which reflects the strength of the interaction between 2 nodes, and is either a positive integer (activation, translation, etc.) or a negative one (repression, inhibition, degradation, etc.). Degradation usually appears as a self-negative edge. The effect of node i on node j is the product of node i ’s state and the weight of the edge $i \rightarrow j$. Thus, the strength of this effect is proportional to both the weight of the edge and the relative level of activity of the source node. The “net” effect on a node is the sum of effects of all nodes on it. Figure A shows a “toy” network with 4 nodes (A, B, C, signal; colored rectangles), whose initial states are shown below their names. Edges are labeled with their weights and colored accordingly (green = activation; red = inhibition).

Figure A: a “toy” network in BioNSi

Figure B: graphical presentation of simulation results

A simulation starts with a given configuration of initial states, and consists of the repeated application of a transition rule, simultaneously to all nodes: the state of a node will increase, decrease, or remain unchanged when the “net” effect on it is, respectively, positive, negative or 0. The simulation ends when either a steady state is reached (two consecutive identical configurations of nodes’ states), or a loop of configurations is detected.

The tool presents the course of simulation graphically (see Figure B). The network exhibits an infinite loop in which the states of nodes A, B and C repeatedly go up and down, while the signal node is constant at state 1. BioNSi also enables a batch mode, which is used to gather statistics on a set of simulations, to study global properties of the network (e.g., the distribution of steady states, average time to get to each steady state, etc.).

BioNSi supports various additional features, which are described in the manual on the website. These include (i) delay on edges for non-simultaneous update of nodes, (ii) blocking edges, where nodes X1, X2, ... can block the effect of some node Y on node Z at a specific time step. The blocking nodes can act together or each of them separately. And (iii) nodes that represent external signals (hormone injection, sunlight, etc.) with a pre-defined expression pattern whose behavior is pre-set by the user, etc.

may be a tedious task, when the networks are beyond a very small size. The new feature allows easily importing multiple pathway files that were exported from the KEGG database. The pathways are simply downloaded from KEGG as xml files, and merged into a single, integrated network under BioNSi. The user guide on BioNSi’s website explains the details for this operation. Briefly, nodes contained in several entities often have alternative pathways are fused into a single node, with all their incoming and outgoing edges. KEGG names are all preserved for each node to allow matching of the same entity with different KEGG name.

The user can specify the desired weights for each KEGG type of edge, such as activation, inhibition, phosphorylation, etc., or rely on the default values provided. As biological materials are routinely being degraded, “self-inhibition” loops are automatically added to each node. Subsequently, the resulting network can be “tuned” by adding or removing specific nodes and edges, or by changing specific parameters (e.g., the weight of a specific edge). Thus, constructing large-scale integrative networks from known pathways is done with little effort, and enables inspecting biological pathways at a broader biological context.

Figure 1. The extended BioNSi simulation tool. Gene expression data and selected KEGG pathway data-files are uploaded to BioNSi's Cytoscape Application and analyzed. The resulting network can be modified and tuned. Simulation is performed for gene expression data resulting from various conditions, possibly with addition of artificial perturbations. Differences in the results of simulation under various conditions provide novel biological insights.

Importing nodes' initial states. The second new feature allows the initial nodes' states obtained either from in-house experiments or from various expression databases to be loaded from a simple text file (a two-column comma separated values (csv) text file, in which the left column contains nodes' names and the right one contains the desired initial levels). Such a file can be easily generated from most existing gene expression databases, using *e.g.*, Microsoft Excel. Since existing expression data sources may have different scales (*e.g.*, logarithmic vs. linear), nodes' initial states are read from the input file and normalized to the range 0–9 for consistency and convenience. Note that the initial states are scaled to 0–9 for the whole network based on the maximal expression level in the network, rather than for the range of each gene separately. Since the scaling depends of the maximal expression level of a node in the network, an extreme outlier may restrict the levels of other nodes to a very limited range. Therefore, the use of logarithmic expression data is recommended.

In addition, we extended BioNSi with a graphical representation of the nodes' levels: nodes' background colors indicate their initial states (higher levels are darker gray). Once a simulation ends, nodes colors are automatically updated to reflect their final state, providing a visual image of the incline or decline in nodes' expression. This new feature allows easy comparison between the network's behavior under different experimental conditions, by loading alternative sets of initial states and conducting a new simulation each time.

Simulation. We explored two use-cases as a proof of concept, as described in the next section. In both use cases, KEGG pathways were integrated to form a BioNSi network, and gene expression analysis was used as initial nodes' states. In both use cases, the states of nodes changed throughout the simulation process, terminating in a steady state. This change in states may reflect one of two situations: (i) the expression data itself does not represent a biological steady state, and therefore the changes *in-silico* reflect changes *in-vivo*, and (ii) data is missing from the network, either structural (missing regulators or connections between existing nodes) or quantitative (weights on edges, initial states, etc.). In the latter case the conclusion is that the model needs further improvement.

In the two test cases described in the *Use Cases* section, we conducted simulations for 100 time-steps. All simulations reached a steady state within this number of steps. In order to compare between simulation results, we had to define when a node is considered to exhibit different behavior in the two simulations. We used a simple Euclidian distance measure between the two simulation trajectories of each node. Furthermore, we filtered out nodes whose difference in the two simulations is merely a result of a different initial state. We did that by considering only nodes for which the difference between the simulations, at some time step, was larger than the difference in the initial state. In other words, the gap between the node's states in the two compared simulations should increase at some point beyond the initial gap, in order for it to be considered differential

between the simulations. An implementation of this distance function can be found as a Python script (diff.py) that can be downloaded from the tool's website.

Use cases

Case study 1: Characterization of inflammatory bowel disease (IBD) conditions

IBD are a series of chronic inflammations of the intestine. Pouch surgery is a useful treatment of IBD patients, but pouch inflammation is a very common outcome¹⁸. Crohn's-like disease of the pouch (CLDP) is the most severe inflammation condition. Activation of NFKB was identified as one of the key regulators inducing IBD inflammation by promoting pro-inflammatory cytokines^{19,20}. Parthenolide, a strong NFKB inhibitor, has recently been demonstrated to be a promising therapeutic agent in IBD inflammation, promoting apoptosis of cancer cells²¹. We aim at extending the original study analyzing differentially expressed genes and their functional enrichment, to suggest new targets for treatment.

Expression data source. Affymetrix GeneChip (Human Gene 1.0 ST arrays) was used for gene expression analysis of normal and CLDP donors, as previously described¹⁸.

Pathway data. As reported in 18, 74 genes were found to be up-regulated in CLDP vs normal expression (pFDR<0.05 and fold-change difference=5). WebGestalt KEGG pathway enrichment²² revealed enriched 16 pathways (pBonf<0.05), of which the top 5 were selected for analysis (Table 2). In addition, IBD pathway was added, although not specifically enriched in the analysis (total of 6 pathways).

NetworkAnalyzer out-degree analysis. Network analysis was performed using NetworkAnalyzer tool, which is another built-in Cytoscape application²³. We used it in order to visualize

the most out-connected (notably high out-degree) genes in the network. The most outstanding such node is NFKB1 (19 outgoing edges, see Table 3). Therefore, an additional NFKB signaling pathway (hsa04064) was added from KEGG. Altogether, 7 KEGG pathways were analyzed, composed of 379 nodes (mostly genes).

In the resulting merged network (Figure 2), in addition to genes, compounds and complexes were also detected (presented as triangles and hexagons, respectively). These were automatically assigned expression value 0 as no information could be found for them in the expression data.

Edge weights. We used edge weights as following: inhibition -1, activation +2. As biological materials are being degraded over time, self-inhibition edges (weight -1) were automatically added for each node in the network. Accordingly, dephosphorylation (-p) edges are set to -1. Activation edges, including phosphorylation (+p) were set to +2, to compensate for the degradation edges automatically added to all nodes.

Results. We aim at comparing between two simulations: 1) Control gene expression; 2) CLDP gene expression. The network consists of 379 nodes. The comparison revealed 49 nodes that exhibited different simulation trajectories (genes, compounds and complexes; 12.9% of all nodes in network). 87.8% of these (43 nodes) were downstream in the network, namely, had no outgoing edges (except for self-loops that all nodes have). To visualize this we used Cytoscape's hierarchical layout, as shown in Figure 2. The nodes that differed between the two simulations are colored red. The hierarchical layout positions these nodes at the top of the network. NFKB1 gene (highlighted in turquoise) is highly connected within the network, but not to any of the selected (red) genes.

Table 2. Five top pathways enriched (adjP) in differentially expressed gene analysis¹⁶, in addition to inflammatory bowel diseases (IBD) pathway. Top human KEGG pathways were obtained and enrichment p-values presented. Links to KEGG xml file download are provided.

KEGG enrichment	KEGG (hsa)	adjP	Link to xml files
Cytokine-cytokine receptor interaction	04060	1.50e-11	http://www.genome.jp/kegg-bin/show_pathway?hsa04060
Leishmaniasis	05140	6.52e-10	http://www.genome.jp/kegg-bin/show_pathway?hsa05140
Rheumatoid arthritis	05323	3.49e-09	http://www.genome.jp/kegg-bin/show_pathway?hsa05323
Chemokine signaling pathway	04062	1.98e-08	http://www.genome.jp/kegg-bin/show_pathway?hsa04062
Hematopoietic cell lineage	04640	7.13e-06	http://www.genome.jp/kegg-bin/show_pathway?hsa04640
IBD	05321	Not enriched	http://www.genome.jp/kegg-bin/show_pathway?hsa05321

Table 3. Node distribution by out degree score. Cytoscape's NetworkAnalyzer tool was used to present out-degree score. NFKB1 has an outstanding out-degree score.

Out-degree	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
Number of nodes	0	138	155	51	18	11	2	3	0	0	0	0	0	0	0	0	0	0	0	1 (NFKB1)

Figure 2. The main component of the CLDP test case network. We show only 241 out of 379 genes, which belong to the main connected subnetwork resulting from 7 KEGG pathways (excluding subnetworks of 8 nodes or less). We used Cytoscape's hierarchical layout, in which nodes with no outgoing edges appear at the top. Gene's fill color is according to control expression values (grayscale), except nodes that are different between the simulation of CLDP and control expression, which are marked in red, and NFKB1 which is colored turquoise. Edges are colored according to type of interaction (activation-green; inhibition-red). For convenience, genes or proteins are represented as rectangles, compounds as triangles and complexes (dimers) are hexagons.

A large fraction of the network's genes is not directly connected to the main network, but forms small independent sub-networks (138 nodes with 8 nodes or less). This is a result of selecting only 7 KEGG pathways for the analysis. Naturally, the analysis is highly dependent on the KEGG pathways selected. Disconnected sub-networks may be connected to the main network following the inclusion of additional, less enriched KEGG pathways. However, these subnetworks follow the same pattern seen in the main network, and all of the nodes differing in the 2 simulations are at the top of the hierarchy (Supplement 1; http://www.cs.tau.ac.il/~amirr/files/supp/test-case1_files.zip). For clarity, Figure 2 shows only the main connected subnetwork (241 nodes; 138 nodes were removed from Figure 2, all harboring connected subnetworks with 8 or less nodes, presented in Supplement 1).

In terms of biological function, most of the 49 nodes that exhibit differences between the simulations are of genes that are either immune system receptors (receptors for chemokines, cytokines, interferons, interleukins, PDGF, growth factors (MET) etc.), or various kinases (BTK, MAPK14, MAP3K7, IKBKB, FLT3 and FLT4). Many therapeutic agents are offered based on such molecules²⁴. Several cytokines (that may be targeted by JAK inhibition), were found within the BioNSi resulting nodes that exhibited different simulation trajectories: IL6ST, IL6R and OSMR (gp130 family member), IL10RB (IL10 members), IL2RB, IL7R (γ c family) and CSF1R (β c family), and IFNAR1, IFNAR2 (interferon related genes)²⁵. This analysis may suggest interesting new targets for drug response.

Case study 2: Treatment of MCF7 breast cancer cells with heregulin (HRG)

The purpose of the following example is to demonstrate the effect of HRG treatment on MCF7 breast cancer cells, based on published GEO experimental results (GSE6462;^{24,25}). The simulation results suggest that BioNSi is capable of recapitulating the

major changes in the network after drug treatment, as observed by *in vitro* experiment.

Expression data source. MCF7 breast cancer cells were treated with a growth hormone, HRG, at four different doses and different time course (GSE6462). We used expression data after maximal treatment of 10 nM HRG for 90 minutes. In the experiment, the mRNA expression levels were measured for control (GSM148517) and HRG treated (GSM148572) cells (using Affymetrix Human Genome U133A 2.0 Arrays).

Pathway data. KEGG-xml files for three selected human pathways were exported from the KEGG database¹²: ERBB signaling pathway (hsa04012); MAPK signaling pathway (hsa04010); EGFR cytosine kinase inhibitor resistance pathway (hsa01521). These three representative pathways were reported to be crucial for the HRG treatment response²⁶. The merged network based on these three KEGG pathways consisted of 190 nodes.

Setting network parameters and post-import modification. We used edge weights as follows: inhibition -1, activation +2. As biological materials are being degraded over time, self-inhibition edges (weight -1) were automatically added for each node in the network. Accordingly, dephosphorylation (-p) edges are set to -1. Activation edges (including phosphorylation; +p) were set to +2, to compensate for the degradation edges automatically added to all nodes. A node representing HRG treatment was added manually (diamond shape), either with initial state 0 (without drug), or with initial state 9 (active drug). The HRG node was directly connected to its known target nodes, ErbB3 and ErbB4, in addition to their homodimers nodes, with strong inhibition edges (weight -10 each). Protein homodimers (resulting from KEGG database) were given expression values similar to their monomer molecules (from the microarray mRNA expression results). Heterodimers were given initial state 0, as we have no data concerning their actual expression.

Simulations. A useful application of the BioNSi tool is to compare between the simulation results under different nodes' initial states (reflecting different biological conditions). As illustrated in Figure 3, we compared simulations of control (A) vs HRG treatment (B) gene expression, in addition to simulation of artificial HRG effect on control gene expression (C).

In order to simulate artificial HRG treatment effects on the network, we used control expression values as initial states, and the HRG node at initial state 9. Network's simulation under these conditions resulted in a complete and rapid decline of ErbB3, ErbB4 and their homodimer expression values.

Results. We aimed at comparing between two simulation sets, based on the same network created: (1) Simulation that compares between HRG gene expression vs. control gene expression. The comparison between these simulations revealed 132 different genes (69.5% of all genes in the network (190)) whose expression pattern differs between the simulations of HRG vs Control. (2) Simulation of HRG artificial treatment effect on the same control gene expression, vs. control expression. This comparison revealed 33 genes (17% of the total 190 in network) that were found to be different. Figure 4 shows a Venn diagram that demonstrates these findings. 28 genes are common to both analyses. Artificial HRG treatment results (28 genes) represent ~21% of the experimental results (132 genes). 5 nodes were found to be uniquely changed by the prediction analysis, including the obviously changed (ERBB3, ERBB4 and ERBB4-ERBB4, which are directly inhibited by HRG simulation prediction, in addition to STAT5A).

Interesting observations concerning the resulting genes can be observed in the BioNSi screenshot of the network in Figure 5. All the 33 genes (framed pink and orange), resulting from

Figure 3. Schematic illustration of BioNSi simulations performed. Expression data from GSE6462 experiment was obtained for Control MCF7 breast cancer cells, without (A) or with (B) HRG treatment. (C) BioNSi simulation of artificial HRG effect performed on control expression.

Figure 4. Venn diagram comparing simulation of HRG expression microarray results with simulation of artificial HRG treatment prediction. 28 genes were common to both analyses (pink). 5 genes were unique to simulation (orange), and 104 genes were unique to experimental analysis (blue).

the artificial HRG treatment simulation analysis compared to control, are highly connected, especially downstream of the network. This suggests an interesting signal path of the HRG drug response, which may offer new targets for treatment. We note that the PI3K/AKT/mTOR and STAT pathway genes are predominantly changed following HRG artificial treatment prediction, pathways that may affect cell proliferation and apoptosis, as recently suggested^{5,26}.

Conclusions

The bio-technological revolution we are witnessing in recent decades enables generating unprecedented amounts of biological data of various types. This includes expression levels of RNA and proteins, and the interactions between them and other substances, such as nutrients, hormones, external stimuli etc. While these data are accessible to researchers, their analyses at the system level are less so. Extracting specific and local information is easy, but performing large scale network analyses requires computational approaches, such as simulation tools. However, existing network simulation tools are often meant for manual construction by the researcher of the network under study, and therefore remain inefficient for studying networks beyond a small scale. This is extremely unfortunate in an era when computational approaches to study biological systems are proliferating. Network simulations are one such important computational approach, which has proven efficient in providing new insights, shedding light on existing results, and suggesting new directions for exploration, both in the lab and on the computer. In order for simulation tools to become an integral part of the researchers' lab routine, tools that enable simulation of large scale network data are called for. Such tools must avoid a tedious network construction process, and allow a high level representation of networks to be generated from existing data semi-automatically with little effort and no computational skills.

Figure 5. HRG network view. A network of 190 genes consisting of 3 KEGG pathways (described in the text), using control expression values is shown. Gene frames color code: common genes to the two analyses (pink), unique to prediction of artificial HRG treatment vs control (orange), unique to simulation resulting from HRG treatment vs control expression analysis (blue) or unchanged (black). The network shows genes with nodes fill color grayscale according to expression values and edge colors according to type of interaction (activation—green; inhibition—red). For convenience, genes or proteins are represented as rectangles, compounds as triangles and complexes (dimers) are hexagons. HRG treatment is presented as orange-framed diamond.

In this paper we suggest an extension to an existing network simulation tool, BioNSi, that allows researchers to import multiple pathway data from the KEGG database into a single network representation. The tool then allows for the study of the network via simulations, under different biological conditions. We believe such an approach will foster the use of computational tools in the lab by researchers with little or no computational background.

Software and data availability

Tool requirements: Cytoscape 3.x, JAVA 8

Tool download: <http://bionsi.wix.com/bionsi> (a full user guide and a step by step manual can be found at the tool's website)

Tool source code: <http://doi.org/10.5281/zenodo.1065352>²⁷

Archived source code as at time of publication: <http://doi.org/10.5281/zenodo.1065352>²⁷

License: CC BY 4.0

Data source for test case 1: <https://www.ncbi.nlm.nih.gov/pubmed/24108111>

Data source for test case 2: ftp://ftp.ncbi.nlm.nih.gov/geo/series/GSE6nnn/GSE6462/matrix/GSE6462_series_matrix.txt.gz

Competing interests

No competing interests were disclosed.

Grant information

The author(s) declared that no grants were involved in supporting this work.

Acknowledgements

We thank Varda Wexler from the Multimedia Unit in the Faculty of Life Sciences, Tel Aviv University, for creating the BioNSi logo, and Benny Chor from the School of Computer Science at Tel Aviv University for useful comments during manuscript preparation.

Supplementary material

Test case 1 files: http://www.cs.tau.ac.il/~amirr/files/supp/testcase1_files.zip

Test case 2 files: http://www.cs.tau.ac.il/~amirr/files/supp/testcase2_files.zip

References

- Barabási AL, Oltvai ZN: **Network biology: understanding the cell's functional organization.** *Nat Rev Genet.* 2004; 5(2): 101–13.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Mendes P, Hoops S, Sahle S, et al.: **Computational modeling of biochemical networks using COPASI.** In *Methods Mol Biol.* 2009; 500: 17–59.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Li F, Long T, Lu Y, et al.: **The yeast cell-cycle network is robustly designed.** *Proc Natl Acad Sci U S A.* 2004; 101(14): 4781–6.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Rubinstein A, Gurevich V, Kasulin-Boneh Z, et al.: **Faithful modeling of transient expression and its application to elucidating negative feedback regulation.** *Proc Natl Acad Sci U S A.* 2007; 104(15): 6241–6.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Rubinstein A, Bracha N, Rudner L, et al.: **BioNSi: A Discrete Biological Network Simulator Tool.** *J Proteome Res.* 2016; 15(8): 2871–80.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Zheng J, Zhang D, Przytycki PF, et al.: **SimBoolNet—a Cytoscape plugin for dynamic simulation of signaling networks.** *Bioinformatics.* 2010; 26(1): 141–2.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Naldi A, Berenguier D, Fauré A, et al.: **Logical modelling of regulatory networks with GINsim 2.3.** *Biosystems.* 2009; 97(2): 134–9.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Di Cara A, Garg A, De Micheli G, et al.: **Dynamic simulation of regulatory networks using SQUAD.** *BMC Bioinformatics.* 2007; 8(1): 462.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Schwab J, Burkovski A, Siegle L, et al.: **ViSiBool—visualization and simulation of Boolean networks with temporal constraints.** *Bioinformatics.* 2017; 33(4): 601–604.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Schivo S, Scholma J, van der Vet PE, et al.: **Modelling with ANIMO: between fuzzy logic and differential equations.** *BMC Syst Biol.* 2016; 10(1): 56.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Bock M, Scharp T, Talnikar C, et al.: **BooleSim: an interactive Boolean network simulator.** *Bioinformatics.* 2014; 30(1): 131–2.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Kanehisa M, Furumichi M, Tanabe M, et al.: **KEGG: new perspectives on genomes, pathways, diseases and drugs.** *Nucleic Acids Res.* 2017; 45(D1): D353–D361.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Büchel F, Rodriguez N, Swainston N, et al.: **Path2Models: large-scale generation of computational models from biochemical pathway maps.** *BMC Syst Biol.* 2013; 7: 116.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Wrzodek C, Büchel F, Ruff M, et al.: **Precise generation of systems biology models from KEGG pathways.** *BMC Syst Biol.* 2013; 7: 15.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Barrett T, Wilhite SE, Ledoux P, et al.: **NCBI GEO: archive for functional genomics data sets—update.** *Nucleic Acids Res.* 2013; 41(Database issue): D991–5.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Kolesnikov N, Hastings E, Keays M, et al.: **ArrayExpress update—simplifying data submissions.** *Nucleic Acids Res.* 2015; 43(Database issue): D1113–6.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Su G, Morris JH, Demchak B, et al.: **Biological network exploration with Cytoscape 3.** *Curr Protoc Bioinformatics.* 2014; 47: 8.13.1–24.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Ben-Shachar S, Yanai H, Baram L, et al.: **Gene expression profiles of ileal inflammatory bowel disease correlate with disease phenotype and advance understanding of its immunopathogenesis.** *Inflamm Bowel Dis.* 2013; 19(12): 2509–21.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Atreya R, Neurath MF: **New therapeutic strategies for treatment of inflammatory bowel disease.** *Mucosal Immunol.* 2008; 1(3): 175–82.
[PubMed Abstract](#) | [Publisher Full Text](#)

20. Ahmed S, Dewan MZ, Xu R: **Nuclear factor-kappaB in inflammatory bowel disease and colorectal cancer.** *Am J Dig Dis.* 2014; **1**(2): 84–96.
[Reference Source](#)
21. Kim Y, Kim TK, Kim Y, *et al.*: **Principal network analysis: identification of subnetworks representing major dynamics using gene expression data.** *Bioinformatics.* 2011; **27**(3): 391–8.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
22. Wang J, Duncan D, Shi Z, *et al.*: **WEB-based GEne SeT AnaLysis Toolkit (WebGestalt): update 2013.** *Nucleic Acids Res.* 2013; **41**(Web Server issue): W77–83.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
23. Doncheva NT, Assenov Y, Domingues FS, *et al.*: **Topological analysis and interactive visualization of biological networks and protein structures.** *Nat Protoc.* 2012; **7**(4): 670–85.
[PubMed Abstract](#) | [Publisher Full Text](#)
24. Neurath M: **Current and emerging therapeutic targets for IBD.** *Nat Rev Gastroenterol Hepatol.* 2017; **14**(11): 688.
[PubMed Abstract](#) | [Publisher Full Text](#)
25. De Vries LCS, Wildenberg ME, De Jonge WJ, *et al.*: **The Future of Janus Kinase Inhibitors in Inflammatory Bowel Disease.** *J Crohns Colitis.* 2017; **11**(7): 885–93.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
26. Nagashima T, Shimodaira H, Ide K, *et al.*: **Quantitative transcriptional control of ErbB receptor signaling undergoes graded to biphasic response for cell differentiation.** *J Biol Chem.* 2007; **282**(6): 4045–56.
[PubMed Abstract](#) | [Publisher Full Text](#)
27. Rubinstein A: **BioNSI - Biological Network Simulation Tool (Version 1.2).** *Zenodo.* 2017.
[Data Source](#)

Open Peer Review

Current Peer Review Status:

Version 2

Reviewer Report 14 June 2018

<https://doi.org/10.5256/f1000research.16110.r34005>

© 2018 Feldmesser E et al. This is an open access peer review report distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Ester Feldmesser

Life Sciences Core Facilities , Weizmann Institute of Science, Rehovot, Israel

Bareket Dassa

Life Sciences Core Facilities, Weizmann Institute of Science, Rehovot, Israel

We believe that the extension tools for BioNSi, which are presented in the paper, provide useful network tools for biologists. Still, we suggest several improvements that are listed below.

1. The feature to compare between two networks is useful for the user, however more detailed explanations on how to work with it and how to visualize the differences between networks are missing. We would like to suggest that this feature will be integrated in the BioNSi application.
2. We would like to suggest that manual changes in the node states in the visualized network will affect the resulting node color in the simulation.
3. Uploading an expression data file with negative expression values (log2) is currently not applicable and should be corrected.
4. The names of the edges in the "Edge Table" panel are not informative and it is difficult to relate between the graphics and the table. In the example "toy1" network, there are two edges with the name "Edge_0" and one with "Edge_1".
5. The option to "Destroy Network" exists in two places in the top menu: in the "BioNSi" tab and in the "Edit" tab. Please note that it worked properly only when activated from the "BioNSi" tab.

Is the rationale for developing the new software tool clearly explained?

Yes

Is the description of the software tool technically sound?

Yes

Are sufficient details of the code, methods and analysis (if applicable) provided to allow replication of the software development and its use by others?

Yes

Is sufficient information provided to allow interpretation of the expected output datasets and any results generated using the tool?

Yes

Are the conclusions about the tool and its performance adequately supported by the findings presented in the article?

Partly

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: Bioinformatics

We confirm that we have read this submission and believe that we have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.

Reviewer Report 11 June 2018

<https://doi.org/10.5256/f1000research.16110.r33996>

© 2018 Naldi A. This is an open access peer review report distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Aurélien Naldi

École Normale Supérieure (ENS), Paris, France

The revised article answered most of my reservations with the first version, some minor clarifications are still needed.

The sentence explaining the merging process of multiple pathways is not clear and seems to merge two different sentences: "nodes contained in several entities often have alternative pathways are fused into single node". What happens if gene A is associated to genes B and C in one pathway and to genes C and D in another? Will we obtain:

- a single node merging A,B,C, and D?
- a node merging A and C, and separate nodes for B and D?
- 4 different nodes?

I assume the authors selected the later solution, which facilitates the mapping of external data (as each node corresponds to a single biological entity), but increases the complexity of the network.

Is that correct?

Regarding the mapping of external data, I understand from the answer of the authors that the provided csv file will map official gene symbols, HGNC IDS, as well as KEGG-IDs. It is clearly stated in the documentation, but could deserve a comment in the paper, along with the fact that unmapped nodes will be assigned the value 0.

While the misleading use of asynchronous in table 1 has been fixed, the text referring to this table in the introduction has not been updated.

A closing parenthesis is missing on page 4: "(e.g. GEO or ArrayExpress".

Is the rationale for developing the new software tool clearly explained?

Partly

Is the description of the software tool technically sound?

Partly

Are sufficient details of the code, methods and analysis (if applicable) provided to allow replication of the software development and its use by others?

Partly

Is sufficient information provided to allow interpretation of the expected output datasets and any results generated using the tool?

Partly

Are the conclusions about the tool and its performance adequately supported by the findings presented in the article?

Partly

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: Qualitative modeling of regulatory networks

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.

Version 1

Reviewer Report 21 February 2018

<https://doi.org/10.5256/f1000research.14379.r29782>

© 2018 Scardoni G. This is an open access peer review report distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Giovanni Scardoni

Center for Biomedical Computing (CBMC), University of Verona, Verona, Italy

I tried to simulate the toy example from the BioNSi website, following the step by step guide. It doesn't work. Tried some other examples (cellcyclesB, Testcase1_network) and no results. Maybe there is a bug (I am using Cytoscape 3.6.0) or there is something I'm missing. In the second case the step by step guide should be improved.

This doesn't allow me to provide a complete evaluation of the tool and the paper that seems interesting and well written. I will be happy to reconsider my review once such problems are solved.

Is the rationale for developing the new software tool clearly explained?

Yes

Is the description of the software tool technically sound?

Yes

Are sufficient details of the code, methods and analysis (if applicable) provided to allow replication of the software development and its use by others?

Partly

Is sufficient information provided to allow interpretation of the expected output datasets and any results generated using the tool?

Partly

Are the conclusions about the tool and its performance adequately supported by the findings presented in the article?

Partly

Competing Interests: No competing interests were disclosed.

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard, however I have significant reservations, as outlined above.

Author Response 06 Mar 2018

Amir Rubinstein, Tel Aviv University, Tel Aviv, Israel

Dear Giovanni,

Thank you for the prompt feedback.

After carefully reviewing the data again, we found a bug in opening the toy.cys network on some combinations of Cytoscape, Java and operating systems. We are working on fixing it. In the meantime, we double checked and the csv format should work properly on all versions.

Therefore, we temporarily removed the cys files from the examples tab, leaving only the csv files. In addition, we slightly changed the step-by-step pdf manual to match these changes. We hope this will allow a complete evaluation of the new version of BioNsi.

Thank you,
Amir

Competing Interests: No competing interests were disclosed.

Reviewer Report 05 February 2018

<https://doi.org/10.5256/f1000research.14379.r29573>

© 2018 Naldi A. This is an open access peer review report distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Aurélien Naldi

École Normale Supérieure (ENS), Paris, France

This paper presents an extension of the BioNSi modeling software with two aims: (i) kickstart model construction by integrating KEGG pathways and (ii) use experimental expression data as initial states for the simulation. These aims make sense: facilitating the initial construction and evaluation of dynamical models is an important challenge in systems biology. However, I feel that details on the inner working and discussion of the limitations of this tool are missing from the paper. Some related work should also be taken into account. In particular Wrzodek et al.¹ and Büchel et al.² which generate qualitative models in SBML qual format³ from KEGG pathways. My main concerns are detailed in the following text, none of them would be blocking taken separately, but my overall interpretation is that this work requires deep changes to become convincing.

I installed the BioNSi tool in Cytoscape 3.4. The documentation states that building a network from KEGG pathways requires the selection of a folder containing one or several KEGG xml files and a CSV file with expression data. I tried to provide a folder containing two KEGG pathways and an empty CSV file, but it did not work (error message: "number of expression files is not 1"). I had the same problem with Cytoscape 3.6.

KEGG pathways often "merge" multiple biological entities in a single node, the paper does not say how this case is handled, the documentation only states that ellipsis dots are removed from the

name. Does it mean that these groups are preserved? How does this affect merging multiple KEGG pathways? It can be a concern as the same entity could be involved in different groupings in different KEGG pathways.

More generally, it would make sense to separate the transformation of KEGG pathways into discrete models (taking into account previous work^{1,2}), and the merging of multiple discrete models, which is a more general problem, and could be used to integrate several existing models with new models generated from pathway maps.

The import of initial states from experimental data, and in particular the discretization step deserve further discussion. For a generic tool, I suspect that it uses a simple method where the full range of values for all genes is scaled to 0-9, defining 9 evenly-spaced thresholds. Is it the case or are the thresholds based on the range of each gene separately? As mapping experimental data on the model is one of the two core aims of this paper, it should at least review the alternatives, give proper details on how this is handled and state that alternative mapping can be integrated as a preprocessing step.

Mapping entities in the model to entities in the experimental data is not discussed either, does the tool convert KEGG identifiers to official gene names, does it require the user to do the mapping in advance, does it provide a GUI to fine-tune the mapping afterwards?

The paper does not state what is the objective of using an experimentally measured values as initial state. The experiments presented here seem to represent "steady states" of the biological system, the simulation leads to a steady state which can be compared to the experimentally measured one. Differences between the experimental and computed states can be of two types: (i) inconsistency showing that the model needs further improvement, and (ii) effect that is not measured in the experiment: for example change of phosphorylation which is not measured by expression data. As this is the very core of the paper, a discussion on the possible biological meaning of the results seems required.

The BioNSi tool, which this work extends, supports simulations of qualitative (discrete) models. Its capabilities are summarized in this paper, and it is briefly compared to other tools for qualitative modelling in Table 1. A minor typo in this table replaced BooleSim with BoolSim (there is also a tool called BoolSim in this field: it is the command line tool handling the Boolean part of SQUAD). The paper describes BioNSi (and some other tools) as doing synchronous and asynchronous simulations, but misuses the term "asynchronous". Here asynchronous refers to synchronous with delays or priorities (i.e. all transitions do not happen at the same speed, but are still using the same clock), however it usually means independent trajectories, leading to non-deterministic behaviour. Using the classical asynchronous updating of qualitative models, a state can have several separate successors, each corresponding to the update of a single component. This allows to study systems where the relative speeds of transitions is not known, but introduces numerous concurrent trajectories, making the analysis much harder. GINsim is the only tool in the provided list implementing asynchronous updating, it also supports priorities, but not delays. BoolNet⁴, PyBoolNet⁵, and BoolSim⁶ support asynchronous updating. To handle this complexity, some tools such as BooleanNet⁷ or MaBoSS⁸ perform random walk in the asynchronous dynamics.

In BioNSi, inhibitions have negative weights, while activations have positive weights. These weights are multiplied by the level of their source nodes, the resulting weighted sum is then used

to define the evolution of the target element. While this allows to quickly define a model, I am concerned by the lack of flexibility: some inhibitors only block the effect of one of the activators, some regulators have a cooperative effect, some act independently. I agree that the simple weight system used in BioNSi is convenient when these fine-grained mechanisms are unknown: it provides a good first model. However, the KEGG pathways mostly describe processes which have been deeply studied, and for which such an approximation could be avoided. There is some value in a simple system based on weights, especially as it fits better for integration into cytoscape, but building more precise functions through the use of intermediate nodes representing cooperative effects should at least be discussed as a common refinement.

Some minor points:

- Missing closing of 2 parenthesis on page 2, another on page 3
- Page 6: "138 nodes with 8 nodes or less" is confusing. I suspect it means 138 nodes within sub-networks of 8 nodes or less?
- The first paragraph on page 6 ("simulation of different experimental conditions") repeats the last paragraph of the methods.

References

1. Wrzodek C, Büchel F, Ruff M, Dräger A, et al.: Precise generation of systems biology models from KEGG pathways.*BMC Syst Biol.* 2013; **7**: 15 [PubMed Abstract](#) | [Publisher Full Text](#)
2. Büchel F, Rodriguez N, Swainston N, Wrzodek C, et al.: Path2Models: large-scale generation of computational models from biochemical pathway maps.*BMC Syst Biol.* 2013; **7**: 116 [PubMed Abstract](#) | [Publisher Full Text](#)
3. Chaouiya C, Bérengruer D, Keating SM, Naldi A, et al.: SBML qualitative models: a model representation format and infrastructure to foster interactions between qualitative modelling formalisms and tools.*BMC Syst Biol.* 2013; **7**: 135 [PubMed Abstract](#) | [Publisher Full Text](#)
4. Müssel C, Hopfensitz M, Kestler HA: BoolNet--an R package for generation, reconstruction and analysis of Boolean networks.*Bioinformatics.* 2010; **26** (10): 1378-80 [PubMed Abstract](#) | [Publisher Full Text](#)
5. Klarner H, Streck A, Siebert H: PyBoolNet: a python package for the generation, analysis and visualization of boolean networks.*Bioinformatics.* 2017; **33** (5): 770-772 [PubMed Abstract](#) | [Publisher Full Text](#)
6. Garg A, Di Cara A, Xenarios I, Mendoza L, et al.: Synchronous versus asynchronous modeling of gene regulatory networks.*Bioinformatics.* 2008; **24** (17): 1917-25 [PubMed Abstract](#) | [Publisher Full Text](#)
7. Albert I, Thakar J, Li S, Zhang R, et al.: Boolean network simulations for life scientists.*Source Code Biol Med.* 2008; **3**: 16 [PubMed Abstract](#) | [Publisher Full Text](#)
8. Stoll G, Caron B, Viara E, Dugourd A, et al.: MaBoSS 2.0: an environment for stochastic Boolean modeling.*Bioinformatics.* 2017; **33** (14): 2226-2228 [PubMed Abstract](#) | [Publisher Full Text](#)

Is the rationale for developing the new software tool clearly explained?

Partly

Is the description of the software tool technically sound?

Partly

Are sufficient details of the code, methods and analysis (if applicable) provided to allow

replication of the software development and its use by others?

Partly

Is sufficient information provided to allow interpretation of the expected output datasets and any results generated using the tool?

Yes

Are the conclusions about the tool and its performance adequately supported by the findings presented in the article?

Partly

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: Qualitative modeling of regulatory networks

I confirm that I have read this submission and believe that I have an appropriate level of expertise to state that I do not consider it to be of an acceptable scientific standard, for reasons outlined above.

Author Response 04 May 2018

Amir Rubinstein, Tel Aviv University, Tel Aviv, Israel

Dear Aurélien,

We greatly appreciate the detailed report. Below please find our point-by-point reply. Changes were made to the manuscript in its 2nd version. To ease reading of the responses below, we copy-pasted the comments in the report as well.

Some related work should also be taken into account. In particular Wrzodek et al.¹ and Büchel et al.² which generate qualitative models in SBML qual format³ from KEGG pathways. My main concerns are detailed in the following text, none of them would be blocking taken separately, but my overall interpretation is that this work requires deep changes to become convincing.

- Regarding the additional resources, they were added to the Introduction. Our focus is at integrating a Cytoscape application (BioNSi) with KEGG, exploiting this convenient simulation environment.

I installed the BioNSi tool in Cytoscape 3.4. The documentation states that building a network from KEGG pathways requires the selection of a folder containing one or several KEGG xml files and a CSV file with expression data. I tried to provide a folder containing two KEGG pathways and an empty CSV file, but it did not work (error message: "number of expression files is not 1"). I had the same problem with Cytoscape 3.6.

- In the current version of the tool the csv expression file should not be empty. This is

obviously something to fix in the next version. In the meantime, we provide such a file with arbitrary initial levels of 9. The file can be downloaded from the website and from <http://www.cs.tau.ac.il/~amirr/files/expression.csv>.

- Both Cytoscape 3.4 and 3.6 should work fine.

KEGG pathways often "merge" multiple biological entities in a single node, the paper does not say how this case is handled, the documentation only states that ellipsis dots are removed from the name. Does it mean that these groups are preserved? How does this affect merging multiple KEGG pathways? It can be a concern as the same entity could be involved in different groupings in different KEGG pathways.

- BioNSi preserves all the alternative names for a KEGG entity. When merging pathways, even if two nodes have different alternative names, they will be fused into a single node. Furthermore, names of entities in the expression level file are matched against any of the alternative names of a node in the network. This information was emphasized in the new version of the user manual and paper.

More generally, it would make sense to separate the transformation of KEGG pathways into discrete models (taking into account previous work^{1,2}), and the merging of multiple discrete models, which is a more general problem, and could be used to integrate several existing models with new models generated from pathway maps.

- Such separation of functionality indeed makes a lot of sense. It may be implemented as part of a later version of the tool.

The import of initial states from experimental data, and in particular the discretization step deserve further discussion. For a generic tool, I suspect that it uses a simple method where the full range of values for all genes is scaled to 0-9, defining 9 evenly-spaced thresholds. Is it the case or are the thresholds based on the range of each gene separately? As mapping experimental data on the model is one of the two core aims of this paper, it should at least review the alternatives, give proper details on how this is handled and state that alternative mapping can be integrated as a preprocessing step.

- The initial levels are scaled to 0-9 for all the nodes. The following was added to the paper: "Note that the initial states are scales to 0-9 for the whole network based on the maximal expression level in the network, rather than for the range of each gene separately. Since the scaling depends of the maximal expression level of a node in the network, an extreme outlier may restrict the levels of other nodes to a very limited range. Therefore, the use of logarithmic expression data is recommended, and when the raw data in not logarithmic a simple transformation is required as a preprocessing step."

Mapping entities in the model to entities in the experimental data is not discussed either, does the tool convert KEGG identifiers to official gene names, does it require the user to do the mapping in advance, does it provide a GUI to fine-tune the mapping afterwards?

- We use KEGG xml files which include the complete array of gene\protein names available, including official gene symbol, HGNC id, KEGG id and others. In BioNSi, we use all gene symbols for each entry and therefore, identifiers are not being lost. For the graphical presentation, the first identifier is being used.

The paper does not state what is the objective of using an experimentally measured values as initial state. The experiments presented here seem to represent "steady states" of the biological system, the simulation leads to a steady state which can be compared to the experimentally measured one. Differences between the experimental and computed states can be of two types: (i) inconsistency showing that the model needs further improvement, and (ii) effect that is not measured in the experiment: for example change of phosphorylation which is not measured by expression data. As this is the very core of the paper, a discussion on the possible biological meaning of the results seems required.

- We added such a discussion to the Methods section (*Simulation*): "We explored two use-cases as a proof of concept, as described in the next section. In both use cases, KEGG pathways were integrated to form a BioNSi network, and gene expression analysis was used as initial nodes' states. In both use cases, the states of nodes changed throughout the simulation process, terminating in a steady state. This change in states may reflect one of two situations: (i) the expression data itself does not represent a biological steady state, and therefore the changes in-silico reflect changes in-vivo, and (ii) data is missing from the network, either structural (missing regulators or connections between existing nodes) or quantitative (weights on edges, initial states, etc.). In the latter case the conclusion is that the model needs further improvement. "

The BioNSi tool, which this work extends, supports simulations of qualitative (discrete) models. Its capabilities are summarized in this paper, and it is briefly compared to other tools for qualitative modelling in Table 1. A minor typo in this table replaced BooleSim with BoolSim (there is also a tool called BoolSim in this field: it is the command line tool handling the Boolean part of SQUAD).

- Fixed.

The paper describes BioNSi (and some other tools) as doing synchronous and asynchronous simulations, but misuses the term "asynchronous". Here asynchronous refers to synchronous with delays or priorities (i.e. all transitions do not happen at the same speed, but are still using the same clock), however it usually means independent trajectories, leading to non-deterministic behaviour. Using the classical asynchronous updating of qualitative models, a state can have several separate successors, each corresponding to the update of a single component. This allows to study systems where the relative speeds of transitions is not known, but introduces numerous concurrent trajectories, making the analysis much harder. GINsim is the only tool in the provided list implementing asynchronous updating, it also supports priorities, but not delays. BoolNet⁴, PyBoolNet⁵, and BoolSim⁶ support asynchronous updating. To handle this complexity, some tools such as BooleanNet⁷ or MaBoSS⁸ perform random walk in the asynchronous dynamics.

- We changed "asynchronous" to "non-simultaneous", and added this column to Table 1.

In BioNSi, inhibitions have negative weights, while activations have positive weights. These weights are multiplied by the level of their source nodes, the resulting weighted sum is then used to define the evolution of the target element. While this allows to quickly define a model, I am concerned by the lack of flexibility: some inhibitors only block the effect of one of the activators, some regulators have a cooperative effect, some act independently. I agree that the simple weight system used in BioNSi is convenient when these fine-grained mechanisms are unknown: it provides a good first model. However, the KEGG pathways mostly describe processes which have been deeply studied, and for which such an approximation could be avoided. There is some value in a simple system based on weights, especially as it fits better for integration into cytoscape, but building more precise functions through the use of intermediate nodes representing cooperative effects should at least be discussed as a common refinement.

- BioNSi (already in the original version) does support additional regulation effects in the form of blocking and enabling edges and external signal nodes. These are only mentioned without details in the end of Box 1. We added a brief description of these in box 1.

Some minor points:

Missing closing of 2 parenthesis on page 2, another on page 3

- We cannot find these missing parenthesis, we would appreciate a clarification.

Page 6: "138 nodes with 8 nodes or less" is confusing. I suspect it means 138 nodes within sub-networks of 8 nodes or less?

- Fixed.

The first paragraph on page 6 ("simulation of different experimental conditions") repeats the last paragraph of the methods.

- We deleted the redundant paragraph from the results.

Competing Interests: No competing interests were disclosed.

The benefits of publishing with F1000Research:

- Your article is published within days, with no editorial bias
- You can publish traditional articles, null/negative results, case reports, data notes and more
- The peer review process is transparent and collaborative
- Your article is indexed in PubMed after passing peer review
- Dedicated customer support at every stage

For pre-submission enquiries, contact research@f1000.com

F1000Research