

A software tool for network topology analysis under a Metabolic Engineering perspective

José P. Pinto
CCTC and CEB/IBB
Campus Gualtar, Univ. Minho
Braga, Portugal
tel: +351253604421
josepedr@di.uminho.pt

Isabel Rocha
CEB/IBB
Campus Gualtar, Univ. Minho
Braga, Portugal
tel: +351253604408
irocha@deb.uminho.pt

Miguel Rocha
CCTC
Campus Gualtar, Univ. Minho Braga,
Portugal
tel: +351253604435
mrocha@di.uminho.pt

ABSTRACT

In this work, we present a software application that runs as a plug-in over the OptFlux Metabolic Engineering platform allowing the topological analysis of metabolic networks. The major aim of this tool is to allow the interconnection between phenotype simulation tasks (using algorithms such as Flux Balance Analysis) and topological analysis of the same networks. The provided methods include node degree and degree distributions, shortest path analysis, clustering coefficients and several node rankers (betweenness and closeness centrality, hubs and authorities, etc). Also, it allows the creation of sub-networks through several filters, including some based on the results of phenotype simulation.

Categories and Subject Descriptors

J.3 [Computer Applications]: Life and Medical Sciences - *Biology and Genetics*.

General Terms

Algorithms, design.

Keywords

Metabolic networks, Network topology analysis, Metabolic engineering, Graph algorithms, Open-source software.

1. INTRODUCTION

Over the last few years, the field of Metabolic Engineering (ME), dealing with the design of strains with enhanced production capabilities of desired compounds, has grown considerably given the growing demands of the white Biotechnology industry [1]. In this area, there have been recent important advances in the development of genome-scale mathematical models of metabolism for a growing number of microorganisms. These have

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, to republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

ACM-BCB 2010, Niagara Falls, NY, USA.
Copyright 2010 ACM ISBN 978-1-4503-0192-3... \$10.00.

been used for phenotype simulation (using methods such as Flux Balance Analysis) [2]) and also to address strain optimization tasks [3].

To make these methods usable by a larger audience, recently the authors' research group has proposed OptFlux (<http://www.optflux.org>) [4], an open-source software platform that uses mathematical models to simulate and optimize the behavior of microorganisms. Its main objective is to be a reference platform for research in ME. It incorporates strain optimization methods, allowing to find *in silico* mutant strains of microorganisms with enhanced capabilities, regarding an user defined objective function, involving the production of a compound of industrial interest. It also enables other functionalities such as several phenotype simulation methods, both for wild type and mutant strains (e.g. Flux Balance Analysis), Metabolic Flux Analysis, flux variability analysis and Elementary Modes analysis. All these features are freely provided in a user-friendly environment that can also be easily extended, given its modular plug-in based architecture.

A distinct approach to the analysis of metabolism takes stoichiometric models, which are composed by the metabolites, reactions and their stoichiometry and reversibility, and represents these biological systems as networks or graphs. This path has led to the discovery that metabolic networks share a similar architecture with other complex networks, indicating that similar laws may govern most complex networks in nature. This also allows the expertise from large and well-mapped non-biological systems to be used to characterize the intricate interwoven relationships that govern cellular functions [5].

As a result from this effort, and among many other results, metabolic networks have been characterized as highly connected networks that exhibit scale-free degree distributions [6], display the small-world property [7] and exhibit a high degree of modularity [8].

Although some interesting results have been obtained both by the model-based and network-based methods aforementioned, these strategies have largely remained independent. The area of

pathway analysis, including the calculation and analysis of Elementary Flux Modes [9] provides a bridge between network topology and metabolic models, but given its complexity, it cannot be applied to genome-scale models.

In this work, we aim to provide software tools to bridge this gap by providing within the OptFlux ME platform a number of tools that allow the analysis of the topological features of networks created from the same metabolic models that can be used to perform phenotype simulation and strain optimization. Given OptFlux's plug-in based architecture, which allows the extension of its functionality through the inclusion of plug-ins, the logical way to add network creation and analysis capabilities was through the development of a plug-in, named *TopologyNetworkAnalyser for OptFlux*, or briefly *TNA4OptFlux*.

The paper is organized as follows: in the next section we introduce some important concepts on metabolic networks and their topological analysis; next, in section 3 we describe the main functionalities of the proposed tool; in section 4, we present a case study and some of the results obtained; we finish with some conclusions and further work.

2. METABOLIC NETWORKS

2.1 Definitions

The metabolism is essentially a linked series of chemical reactions, which function to synthesize building blocks for usable cellular components and to extract energy and reducing power from the cellular environment [10]. The metabolic reactions are in turn catalyzed by the enzymes produced by the organisms, allowing living beings to adapt to changing conditions, which is one of the sources of life's adaptability.

The system formed by an organism's cellular reactions and the metabolites produced/ consumed by those reactions is organized as a highly complex network called the metabolic network. Due to their high complexity and the fact that often only part of the metabolism is known, it is not unusual for biological networks used in research to only contain part of the metabolism. As it was mentioned before, metabolic networks share a similar architecture with other complex networks, and as such, methods developed for the analysis of complex networks can be used to obtain information about an organism's metabolism.

There are several variants of metabolic networks, being the most important the following:

- (i) directed graphs composed by two types of nodes (called bipartite graphs): metabolites and reactions, where edges represent the consumption or production of metabolites by reactions;
- (ii) directed or undirected graphs where nodes represent the metabolites and edges represent the

reactions that consume or produce those metabolites;

- (iii) directed or undirected graphs where nodes represent reactions and edges stand for metabolites that are shared by two reactions.

2.2 Topological Analysis

In this section, we present some of the main metrics for topological analysis used on several types of complex networks that can also be applied to metabolic networks.

2.2.1 Degree and degree distribution

Of all the network metrics in graph theory, the most basic one is probably the node degree. The degree of a vertex is simply the number of edges that connect with it. This metric is used in different operations of graph analysis and in the calculation of more complex metrics. One of its applications is to evaluate the importance of each node, assuming that more important nodes will have higher degrees. Thus, a ranking can be performed, ordering the nodes by descendent degree.

In directed graphs, a distinction can be made between the incoming degree, or *indegree*, that is the number of edges that end in the vertex and the outgoing degree, or *outdegree*, that is the number of edges that start in the vertex.

Another important metric related to the degree is the degree distribution of a network. This metric is particularly useful since many properties of a network can be determined by the law followed by its degree distribution. For instance, the discovery that metabolic networks follow a power law led to the discovery that these networks are scale-free [6].

2.2.2 Shortest paths

Another important functionality in topological analysis is the calculation of the shortest paths between pairs of nodes in the graph. In metabolic networks, this is simply the number of nodes needed to reach the destiny node from a origin node. It should be noted that in the case a certain node is not reachable from another, the shortest path does not exist (or it is considered as infinity). This case is quite common in directed networks.

For the whole network, it is possible to calculate all shortest paths between all pairs of nodes (sometimes called the shortest path spectrum). It is also possible to calculate the mean over all these values (the mean path length). This last measure can be used to check if the network has the small world property.

2.2.3 Betweenness Centrality

Centrality metrics are very often used for the analysis of social networks, being designed to rank the nodes according to their position in the network and interpreted as the prominence of nodes embedded in a social structure [11]. However, the utility of

centrality measures is not limited to social networks since they have proven to be very useful in the analysis of citation networks, computer networks and also biological networks [12].

The betweenness centrality (BC) is one of the widest used centrality metrics, calculated according with the number of shortest paths between pairs of nodes in the network that pass through the target node, using the following formula to calculate the between centrality of node n :

$$BC(n) = \sum_{s \neq n \neq t \in V} \frac{\sigma_{st}(n)}{\sigma_{st}}$$

where σ_{st} is the number of shortest paths from the vertices s and t and $\sigma_{st}(n)$ is the number of those vertices that pass through the vertex n .

This metric was initially developed to analyze networks where there is an exchange of information between the nodes through the edges. In this case, the BC is, in some sense, a measure of the influence a node has over the spread of information through the network [12].

2.2.4 Closeness Centrality

Similarly to the BC, the closeness centrality (CC) is also a measure of node importance in the network. It specifies which nodes have the shortest paths to all others [13], or in some variations, such as the one used in this work, it evaluates the nodes based in the shortest paths to all reachable nodes. The CC for a node n is calculated as follows:

$$CC(n) = \frac{1}{\sum_{t \in V} d(n, t)}$$

where $d(n, t)$ is the shortest path between the vertices n and t and V is a group composed by all nodes connected to n .

It should be noted that nodes that do not have any neighbors do not have a centrality value since in that case $\sum_{t \in V} d(n, t)$ will equal zero.

2.2.5 HITS

The hubs-and-authorities (also known as Hyperlink-Induced Topic Search or HITS) algorithm differs from the other ranking metrics presented before, in that it calculates two ranks for each node: hubness and authority. HITS was originally developed to rate Web pages based on their content, splitting the pages evaluated into the ones with good content on a given topic, called authorities, and directory-like pages with many hyperlinks to pages on the topic, called hubs [14]. Despite smilingly unrelated to metabolic networks, web page analysis algorithms are sometimes used to analyze biological networks (e.g. PageRank centrality is used in regulatory networks [15]).

HITS is more complex than BC or CC, since the way it calculates the values of hubness and authority cannot be explained by a

simple formula, being instead necessary to understand the algorithm. It is a recursive algorithm in which all nodes initially have the same values of hubness and authority; in each iteration these values are recalculated based in their neighbors until they converge [14].

2.2.6 Clustering Coefficient

In many networks, nodes tend to be organized in strongly connected clusters and metabolic networks are not an exception. A measure of the clustering level of a given node is given by the clustering coefficient. The clustering coefficient for a node $n - C(n)$ - is calculated by the expression:

$$C(n) = \frac{2e}{k_i(k_i - 1)}$$

where e is the number of edges that connect the node to its neighboring nodes and k_i the number of its neighbors.

In a compound-reaction network (bipartite graph), reaction nodes are connected only with compound nodes and compound nodes are only connected with reaction nodes. This means that a node and its neighbors will never share neighbors. The clustering coefficient is, in this case, 0 for all nodes. So, in metabolic networks, this metric only makes sense in reaction-reaction or metabolite-metabolite networks.

3. THE TNA4OPTFLUX PLUGIN

In this work, we present the TNA4OptFlux plug-in for the OptFlux platform. As previously mentioned, the main aim of this plug-in is to make available a number of topological analysis tools, within the OptFlux platform which is devoted to ME tasks.

The plug-in is freely available in the project's web site (<http://www.optflux.org>). Some documentation is also provided in the form of a set of "How To's" that explain how to use the main features of the tool.

In this section, we will firstly mention the topological analysis algorithms provided. Afterwards, we will address some features regarding the integration of topological analysis with some of the ME features from OptFlux, namely phenotype simulation. This section will conclude with some remarks regarding the plug-in implementation.

3.1 Main functionalities

This plug-in creates metabolic networks by extracting the necessary data from the mathematical model contained in an OptFlux project and then organizing the data in a graph. The network created is then associated with the same project from where the model was created. The plug-in allows the creation of any type of network from the ones described in section 2.1.

Once a network is created, it is possible to use the provided analysis tools on it. These functionalities consist on the calculation of all metrics that were presented in Section 2.2: node degrees, degree distribution, node rankers (BC, CC and HITS), shortest paths and clustering coefficients. Each of these features will be detailed next. Also, in Figure 1, a typical workflow of network analysis is depicted, with some screenshots of the application.

Since this plug-in was built over the main OptFlux platform, it shares some of its basic principles regarding the user interaction. Therefore, all data are organized in a clipboard where all results are stored when an operation is run, organized in an hierarchy of data types. To view these results, the user clicks the object and the views are displayed in the main area. Several views might exist for the same object, accessible through different tabs.

3.1.1 Node degrees and degree distribution

The plug-in calculates the node in and out degrees when each network is created. These are available in the nodes view of the network object. Also, it is possible to calculate both the in and the out degree distribution histograms running the respective operation from the menu.

3.1.2 Shortest paths analysis

Shortest path analysis in *TNA4OptFlux* can be launched by running the respective operation from the menu, which calculates the shortest paths between all pairs of nodes in the network. This creates an object with the results, that can be afterwards viewed. Also, some metrics for the whole network are calculated: the shortest path spectrum and the mean path length. Also, this operation display information about the network's connectivity.

When viewing the results, the user can select among a group of four different panels, each one for the visualization of different kinds of data. The first two panels, called shortest path view and reversed path view, both have a list of all nodes in the network and show for each node the number of connected nodes and the average shortest path length distances. They differ, however, since the first shows the average path distance from the nodes listed to the nodes they connect to and the later shows the average path distance using the listed nodes as a destiny. It should be noted that in an undirected graph both views have identical data. The third panel, path view, is used to determine the shortest path between two nodes of the network. Unlike the first two views this one prints the nodes present in a specific shortest path. Finally, the last panel shows all nodes that a selected node is connected to and their distance.

3.1.3 Ranking algorithms

The plug-in allows the nodes to be ranked according to three different algorithms: betweenness centrality, closeness centrality

and HITS. The resulting view shows the ordered ranking of the nodes by the selected criteria.

3.1.4 Clustering

The plug-in also allows the user to calculate the clustering coefficient for all nodes in the network and ranks the nodes according to this criterion. Also, considering the whole network, the user can calculate the average clustering coefficient and also the function $ACC(k)$ that gives the average clustering coefficient for all nodes with degree equal to k .

3.2 Network filters and the integration with OptFlux

There are some situations when it may be useful to separate part of the network and analyze that sub-network as an independent graph. With that in mind, the possibility of creating sub-networks was included in *TNA4OptFlux*. Also, this was the strategy used to make most interconnections between OptFlux and the network topology features implemented in this plug-in.

The simplest criterion for the creation of a sub-network is to identify parts of the network that have no connection between themselves and treat them as separate networks. The plug-in includes an option to identify connected sub-graphs (components) within the whole network. Optionally, the user can create a sub-network based on a selected component.

Other alternatives for the creation of sub-networks are based in the removal of nodes based in their degree or their ranking value calculated using one of the previously proposed algorithms (naturally, the ranking algorithms have to be run for the network in question first). This is a useful functionality since the analysis of metabolic networks is typically conducted over a network where currency metabolites (that are used in many reactions of the network as carriers of electrons and functional groups) are removed.

It may be useful to remove these compounds for the analysis of the network because of three main reasons: (i) these compounds may provide a false robustness to the network. For instance, if an essential reaction is removed, it is expected that the metabolic network will be fragmented; however, the presence of these highly connected nodes means that the network will still be intact despite the fact that, from a biological point of view, that is not meaningful; (ii) some ranking metrics that are used to find the compounds that are more important in the network are dominated by these compounds, and (iii) the great number of edges that these nodes have results in a significant increase of the processing time that it is necessary to analyze the network.

In TNA4OptFlux, the nodes to be removed can also be selected manually by the user from a list. Since metabolic networks tend to have a great number of nodes, this method is only recommended in the case where the number of nodes is small or to make some minor "tweaks" in a sub-network created by other means.

Also, sub-networks are used to create interesting links between topology analysis and phenotype simulation methods. These methods allow the calculation of flux values for each reaction in the network. Thus, the user can define a filter that removes reaction nodes based in their flux value in a selected OptFlux simulation. This allows the comparison among several different results obtained for different environmental conditions (e.g. different carbon sources, aerobic/ anaerobic conditions) or genetic modifications (e.g. gene deletions). This option allows to filter the original network, creating networks where only the reactions active in these simulations are active (flux higher than a user-defined value).

Since, in some situations, it might be necessary to combine several of the filtering options described above, in order to create a desired sub-network, the plug-in has a wizard that allows the combination of several of the previous alternatives in a user friendly way.

3.3 Implementation issues

The whole OptFlux platform, including the *TNA4OptFlux* plug-in, is implemented using the Java programming language. Most of the design options (e.g. the use of the Model-View-Control paradigm), user interactions principles and basic libraries are described in the paper where the OptFlux platform is presented [5]. For the implementation of the topological analysis tools, the JUNG library (<http://jung.sourceforge.net>) was used.

4. CASE STUDY AND RESULTS

4.1 Case study

In order to show the usefulness of the plug-in, a case study was designed. The model selected for this study is a genome-scale metabolic model of the bacterium *Escherichia coli* [16]. This model was used to create four distinct networks. Those were analyzed with all tools of the plug-in. All networks are bipartite graphs that contain both compound and reaction nodes. The objective of this analysis was not to obtain new results but rather to show the capabilities of NetworkAnalyzer.

The first network (net 1) was obtained by creating it directly from the OptFlux project containing the model without using any filter. For the creation of the second network (net 2), the process was the following: firstly, a simulation of the wild type strain (typical growth medium) was run using the Flux Balance Analysis algorithm in OptFlux; afterwards, the sub-network functionalities

were used (over net 1) to create a sub-network where the nodes that corresponded to reactions that had a flux value of zero in the simulation were removed. The rationale underlying this process is that reactions with a zero flux are not used by the organisms in the conditions of the simulation.

The currency metabolites, such as H^+ and ATP were removed from both net 1 and net 2, in this way creating net 3 and net 4. The full list of removed metabolites is provided in Table 1, together with their name on the SBML model used in the study.

Table 1. Compound nodes removed in net 3 e 4

Compound	Node Id
Hydrogen	M_h_b, M_h_c, M_h_e
Water	M_h2o_b, M_h2o_c, M_h2o_e
Phosphate	M_pi_b, M_pi_c, M_pi_e
ATP	M_atp_c
Nicotinamide adenine dinucleotide reduced	M_nadph_c
Nicotinamide adenine dinucleotide	M_nad_b, M_nad_c, M_nad_e
ADP	M_adp_c
Diphosphate	M_ppi_c
Nicotinamide adenine dinucleotide phosphate	M_nadp_c
Nicotinamide adenine dinucleotide phosphate reduced	M_nadph_c
AMP	M_amp_b, M_amp_c, M_amp_e
Coenzyme A	M_coa_c

4.2 Results - Ranking Metrics

4.2.1 Betweenness Centrality

Since the value of BC of a node is directly proportional to the number of shortest paths it is part of in a metabolic network created, there should be two kinds of nodes with a high value:

1. Metabolite nodes that are produced and consumed in a great number of reactions. It should be noted that for a node to have a high BC it has to be both produced and consumed by a number greater than average of reactions.
2. Reaction nodes that produce and consume the compounds mentioned in the previous point. Again reactions have to both consume and produce the high BC compounds. Reversible reactions will have a higher average BC.

As expected in nets 1 and 2 most of the nodes with higher BC either make part of the list presented in Table 1 or are reactions

that produce and consume these metabolites (results for net 1 are shown in Table 2; results for net 2 are not shown since they are very similar). While results of this kind validate our assertion about the BC, they are hardly surprising.

The results obtained with nets 3 and 4 (Tables 3 and 4) were more interesting as they represent compounds or reactions that, although being relevant to the metabolism, are not necessarily involved with currency metabolites. Relevant examples include pyruvate and glutamate that are known to play major roles in metabolism. Further analysis might include checking the importance for the survival of the organism of reactions found to have a high betweenness in these networks.

Table 2. Nodes with higher betweenness centralities in net 1

Type	Node Id	Betweenness
compound	H	2459889
compound	H ₂ O	930881
compound	ATP	429191
compound	Phosphate	354293
compound	H (external)	319376
compound	ADP	180546
compound	Pyruvate	162282
compound	NAD	128833
compound	Diphosphate	126864
reaction	ATPS4r	108972

Table 3. Nodes with higher betweenness centralities in net 3

Type	Node Id	Betweenness
compound	Pyruvate	793787
compound	L-Glutamate	571072
compound	CO ₂	448784
compound	NH ₄	406818
compound	2-oxoglutarate	351565
compound	Acetyl-CoA	324292
compound	Glyceraldehyde-3-phosphate	282511
reaction	ALATA_L	265103
compound	Phosphoenolpyruvate	255559
reaction	ASPTA	226139

Table 4. Nodes with higher betweenness centralities in net 3

Type	Node Id	Betweenness
compound	L_glutamate	95733,9
compound	CO ₂	74757,51
compound	2_oxoglutarate	61596,76
reaction	ASPTA	52741
compound	L_aspartate	49804
compound	Glyceraldehyde_3_phosphate	37129
compound	UDP	36437
reaction	CTPS2	34110
compound	L-Glutamine	33601
reaction	GLNS	33567

4.2.2 Closeness Centrality

Because of the fact that metabolic networks are highly connected most nodes with a high centrality are nodes that have few neighbors which in turn also have very few neighbors. In other words, nodes with high closeness in the networks analyzed tend to be connected to few other nodes and paths to the connected nodes tend to be small.

Nodes with a high closeness in nets 1 and 3 (Tables 5 and 7) are reactions that produce compounds that are not used by other reactions (these may be excretions of *E. coli* or substances that are used in reactions that have not been included in the metabolic networks) or compounds that are only used by a few reactions that do not produce any other compounds or that produce compounds that are not used in any reaction.

Regarding net 2, the results (Table 6) are quite different from the previously mentioned ones. Taking a closer look, it seems that this network does not have any nodes with the properties associated with a high closeness. Indeed, since only the part of the network that has a non-zero flux value under a steady-state condition has been kept in net 2, it makes sense that those poorly connected or even deadends of the network have all been removed. It is well known that network gaps or deadends cannot have a flux under steady-state conditions. The fact that net 4 already shows higher centrality values is probably related with the fact that some additional nodes have been removed after the simulation, which could again create deadends or gaps in the network. These results indicate that closeness centrality could be a good measure of the existence of deadends in a given network and approaches to evaluate a new network, for example, a genome-scale model under development, based on this metric should be further explored.

Table 5. Nodes with higher closeness centralities in net 1

Type	Node Id	Centrality
reaction	UDPGALM	1
compound	Trimethylamine	0,75
compound	1-5-Diaminopentane	0,75
compound	Dimethyl sulfide	0,75
compound	gamma-butyrobetaine	0,75
compound	AMP	0,75

Table 6. Nodes with higher closeness centralities in net 2

Type	Node Id	Centrality
compound	H	0,31316
reaction	ATPS4r	0,30149
compound	ATP	0,29065
compound	H2O	0,28730
reaction	GLUDy	0,28377
reaction	ACCOACr	0,27485

Table 7. Nodes with higher closeness centralities in net 3

Type	Node Id	Centrality
reaction	DHPTDCs	1
compound	Lipopolysaccharide	1
compound	Adenosine-3-5-bisphosphate	1
reaction	SPMDAT2	1
compound	P1-P5-Bis-5-adenosyl-pentaphosphate	1
reaction	TDPDRR	1
compound	Dephospho_CoA	1
reaction	UDPGD	1
reaction	SPMDAT1	1
reaction	PACCOAL	1
compound	L-leucine	1

4.2.3 HITS

HITS is the most complex of the metrics supported by TNA4OptFlux. As such, the analysis of the results of the algorithm is the hardest. For nets 1 and 3 (Tables 8 and 9), all the authorities of the network were compounds and all hubs were reaction, in these networks (net 2 is not shown since the results were similar to net 1). The split between authorities/compounds and hubs/reaction was absolute, as there were no compounds with a significant value of hubbness or reactions with a significant value of authority. After carefully looking at the results, it was

evident that in nets 1, 2 and 3 the compounds with a higher authority were important compounds for the metabolism, such as ATP and hydrogen, and the reactions with higher hubbness were the ones that produced these compounds, especially those that produced multiple ones. In this case, it seems that HITS can be used as a “double” metrics that is capable of evaluating both reactions and compounds differentially.

In net 4 the characteristics of the nodes are reversed, with reactions being the higher authorities of the network and compounds the better hubs. The separation between compound and reaction nodes is also more tenuous than in the other nets, since there are nodes in this network that have a values of authority very close to hubbness values. We are still unsure on the reason of the reversion in the classifications on this network and further analysis will be necessary before drawing any conclusions.

Table 8. Analysis with HITS for net 1

Authorities	Authority	Hubs	Hubbness
H	0,862764	ATPS4r	0,079358
ADP	0,351190	DBTSr	0,073056
Phosphate	0,284655	ACCOACr	0,072651
H (external)	0,130022	PRASCS	0,072413
NADH	0,090845	Biomass	0,072339
Diphosphate	0,074197	PRAGSr	0,072160
ATP	0,055968	ALAALAr	0,071904
AMP	0,043742	CBPS	0,070658
NAD	0,043403	CTPS2	0,070560
H2O	0,040823	PRFGS	0,070426

Table 9. Analysis with HITS for net 3

Authorities	Authority	Hubs	Hubbness
CO2	0.934896	KAS14	0.145308
NH4	0.184248	UGLYCH	0.144774
Pyruvate	0.182570	GLYCL	0.144469
Acyl carrier Protein	0.175768	SADH	0.144388
Ubiquinol 8	0.048900	KAS16	0.143718
L-glutamate	0.045672	C181SN	0.143332
2-Oxoglutarate	0.041802	C140SN	0.143332
L-alanine	0.040432	C161SN	0.143332
Acetoacetyl-ACP	0.034058	C141SN	0.143332
Succinate	0.028756	C160SN	0.143332

Table 10. Analysis with HITS for net 4

Authorities	Authority	Hubs	Hubbness
Biomass	0.925803	L-glutamate	0.316219
ASPTA	0.095323	L-aspartate	0.190471
ALATA_L	0.090570	Acetyl-CoA	0.186644
VALTA	0.089285	L-glutamine	0.172212
PHETA1	0.088518	GTP	0.168417
TYRTA	0.088518	UTP	0.158907
ILETA	0.088518	L-alanine	0.158077
ACGS	0.074688	L-serine	0.155988
GLUDy	0.068496	CTP	0.155107

4.3 Degree Distribution

We calculated the degree distribution of the four networks. To facilitate the visualization of the results we present plots of the histograms both for the indegree and outdegree (blue for indegree, red for outdegree). In Figures 2 and 3 we show the results of net 1 and 4 (nets 2 and 3 behave similarly; not shown). We can check the scale-free nature of the networks, both in their original configuration and when only the active reactions in the wild type are considered.

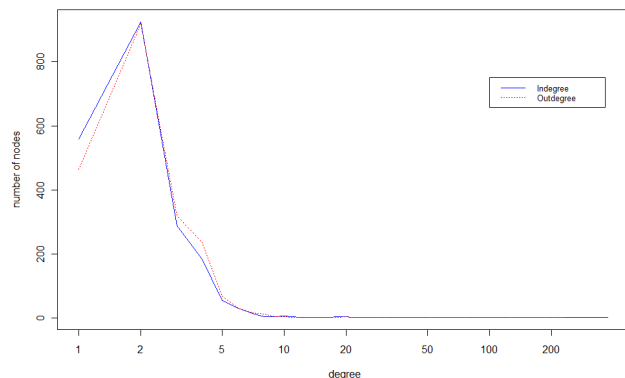


Figure 2. Degree distribution plots for net 1.

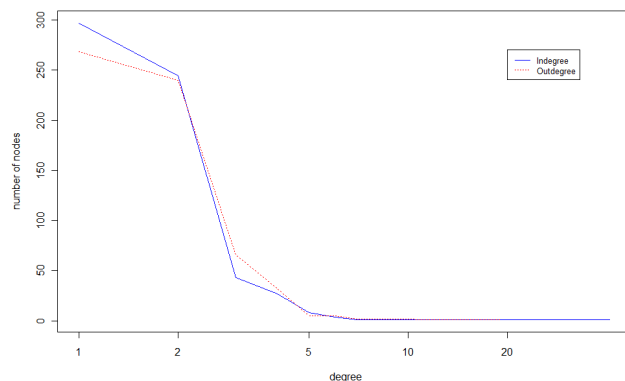


Figure 3. Degree distribution plots for net 4.

4.4 Shortest Path Analyses

In this analysis, the shortest path metric was used to calculate the distance between the node that corresponds to Glucose-6-phosphate (main substrate after its entrance in the cell) and the reaction that produces biomass (one of the final aims of metabolism). To get a better idea of the possible ways that these nodes are connected, the shortest path between the Glucose-6-phosphate node and the nodes that correspond to all compounds consumed by the reaction were also calculated. The purpose is to evaluate the distance between the substrate and the different building blocks that are made from it.

Table 11. Shortest path form the Glucose-6-phosphate to the biomass reaction in nets 1 e 2

Path
Compound: D-Glucose_6_phosphate
Reaction: G6PDH2r
Compound: NADP
Reaction: Biomass

For nets 1 and 2 the shortest path between the nodes had a length of four (Table 11). However, these paths pass through the node that corresponds to NADP, a currency compound that is involved in many reactions (45 in this model, to be exact) which means that this path may not be very significant. The result obtained with the filtered networks (nets 3 and 4) is probably more useful (Table 12).

Table 12. Shortest path form the Glucose-6-phosphate to the biomass reaction in nets 1 e 2

Compound
Compound: D-Glucose-6-phosphate
Reaction: PGMT
Compound: D-Glucose-1-phosphate
Reaction: GALU
Compound: UTP
Reaction: Biomass

5. CONCLUSIONS AND FURTHER WORK

The plug-in described in this work still lacks some of the analysis capabilities that a few network analysis applications possess, such as Pajek [17] or Vison [18]. In this project, we decided de emphasize the communication between the plug-in and the main OptFlux application. With the provided tool, an user can easily run a create a project in OptFlux and then create a network without having to worry with converting the data into a format supported by a networking application. Furthermore, if the user

runs a simulation it can gain further insights into the parts of the model that are used in the simulation through the creation of a sub-network.

As far as the authors are aware, the only application that supports a combination of tools similar to the ones provided by OptFlux combined with TNA4OptFlux is CellNetAnalyzer [19]. However, the topological analysis tools provided by this application are quite limited and the interactions between FBA simulations and topological analysis is not direct. Furthermore, CellNetAnalyzer is based on MatLab, a commercial platform, while our approach is to provide software freely available and fully open-source.

However, the plug-in is still under development and there are some points that should be improved in the future. The current version only supports a few ranking metrics, something we aim to correct in new releases with the inclusion of new algorithms. The connection between the plug-in and OptFlux is probably the main feature. Currently this data exchange consists only in the creation of new networks and when it is used to filter the nodes of the sub-networks. Other interactions will be explored in the future.

Finally we believe that it might be useful to be able to export the networks into formats supported by the most commonly used applications in network visualization.

6. ACKNOWLEDGMENTS

The work of J.P. Pinto is supported by the Portuguese FCT under the PhD grant ref. SFRH/BD/41763/2007.

7. REFERENCES

- [1] Stephanopoulos G, Aristidou A, Nielsen J. Metabolic engineering. San Diego: Academic Press; 1998.
- [2] Edwards JS, Covert M, Palsson B. Metabolic modelling of microbes: the flux-balance approach. *Environmental Microbiology*, 4:133-140, 2002.
- [3] Rocha M, Maia P, Mendes R, Pinto JP, Ferreira EC, Nielsen J, Patil KR, Rocha I: Natural computation meta-heuristics for

- the *in silico* optimization of microbial strains. *BMC Bioinformatics*, 9:499, 2008.
- [4] Rocha I, Maia P, Evangelista P, Vilaça P, Soares S, Pinto JP, Nielsen J, Patil KR, Ferreira EC, Rocha M. OptFlux: an open-source software platform for in silico metabolic engineering. *BMC Systems Biology*, in press, 2010.
 - [5] Barabási, AL and Oltvai, Z. N. 2004. Network biology: understanding the cell's functional organization. *Nature Reviews*, 5:101-114, Feb 2004.
 - [6] Jeong H., Tombor B., Albert R., Oltvai Z., Barabasi AL, The large-scale organization of metabolic networks. *Nature* 407:651-654, 2000.
 - [7] Fell DA, Wagner A. The small world of metabolites. *Nature Biotech* 18(11):1121-1122.
 - [8] Ravasz E., Somera AL, Mongru DA, Oltvai ZN, Barabasi AL. Hierarchical organization of modularity in metabolic networks. *Science* 297:1551-1555, 2002.
 - [9] Schuster S, Fell DA, Dandekar T: A general definition of metabolic pathways useful for systematic organization and analysis of complex metabolic networks. *Nature Biotechnology* 2000. 18:326-332.
 - [10] Tanaka, R.. Scale-rich metabolic networks. *Physical Review Letters* 94, 2005.
 - [11] Brandes, U., A Faster Algorithm for Betweenness Centrality. *Journal of Mathematical Sociology* 25(2):163-177, 2001.
 - [12] Newman, ME. A measure of betweenness centrality based on random walks. *Social Networks* 27:39-54, 2005.
 - [13] Junker, B. H. and Schreiber, F., 2008. Analysis of Biological Networks. Wiley Series in Bioinformatics. WileyBlackwell.
 - [14] Henzinger, MR. Hyperlink Analysis for the Web. *IEEE Internet Computing* 5(1):45, 2001.
 - [15] Koschützki, D. and Schreiber, F., Centrality analysis methods for biological networks and their application to gene regulatory networks. *Gene Regul Syst Bio*, 2:1-9, 2008.
 - [16] Reed JL, Vo TD, Schilling CH, Palsson BO: An expanded genome-scale model of Escherichia coli K-12 (iJR904 GSM/GPR). *Genome Biology*, 4:R54, 2003.
 - [17] Pajek webpage. <http://www.pajel-imfm.si/doku.php>.
 - [18] Vison webpage <http://visone.info/index>.
 - [19] Klamt S, Saez-Rodriguez J, Gilles ED: Structural and functional analysis of cellular networks with CellNetAnalyzer. *BMC Systems Biology*, 1:2, 2007.