NEW APPROACHES FOR VISUALIZING AND ANALYZING METABOLIC PATHWAYS

Tim O. F. Conrad [1]

[1] School of Computer Science and Software Engineering,
Monash University

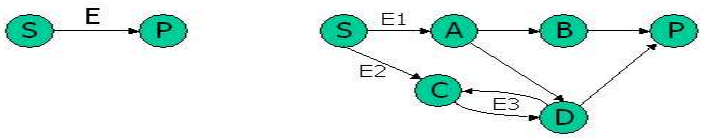
ABSTRACT: Visualizing of metabolic pathways (or networks) has been done by many different approaches. In this work, we implemented and tested existing graph layout algorithms, and present a new approach to lay-out medium size metabolic pathways (500-20,000 vertices) by implementing and combining three well known graph lay-out algorithms (high dimension embedding, spring-embedder preprocessing, spring-embedder), through 3D space density analysis facilitated by the Octree technique. For the analysis of the results of metabolic pathways simulations we present two new techniques: firstly, a powerful technique to visualize pathways simulation data was created to unveil and understand concentration flows through metabolic pathways. This was achieved by mapping the color encoded concentration value of every substance from each time step of the simulation to its graphical representation in the layout. By combining all resulting images (from each time step) and displaying them as a movie, many characteristics such as sub-networks, alternative routes through the network, and differences between a modified pathway and its unmodified version can be revealed. Secondly, a new method to detect co-regulated substances in metabolic pathways and to recognize differences between two versions of a pathway, was established. To do this, we transformed the simulation data into a row-based representation, color-coded these rows, and reordered them with respect to similarity by using a Genetic Algorithm variant. From the arising discrete 2-dimensional matrix consisting of concentration values, a continuous 2-dimensional fourier row function was computed. This function can be used to measure properties, such as similarities in a pathway between time steps, or substances, or to detect and evaluate differences between modified versions of the same pathway.

INTRODUCTION

Metabolism is the collection of several thousand individual chemical reactions or interconversions (that is, generation or degradation of organic substances) carried out by living organisms as they feed, grow and reproduce. This set of reactions forms a highly branched network with a high number of cross-connections due to the majority of reactions involving at least two organic substances, also called compounds, reactants, or metabolites.

In textbooks of molecular biology, the metabolism of biological organisms is commonly divided into parts according to the functional tasks they perform. These parts are called pathways. Figure 1 shows a very small metabolic pathway as a sequence of metabolic reactions.

Figure 1: Left: Metabolic reaction; Right: Metabolic pathway



The general aim of simulations of these metabolic network is to estimate the time evolution of the concentrations of the substances in the system. This is commonly achieved by numerical integration of the rate equations. From these concentration time series many properties of the system can be derived, such as elasticity coefficients, which quantitatively describes the susceptibility of the system to perturbation of selected parameter of the model, or more higher level properties, like detection of functional pathways and cycles, identification of optimal / sub-optimal operating modes, assessing the importance of single reactions, or the analysis of correlated reactions.

Ultimately, computer simulation, visualization, and analyzing tools are hoped to allow the step to prediction of metabolic dynamics and revealing of properties of these networks. Applications include the design of engineered biological systems, the generation of testable hypotheses regarding network structure and function, and the elucidation of properties that can not be described by simple descriptions of individual components (such as product yield, network robustness (that is the ability of the metabolic network to respond to changes in its structure), and correlated reactions).

Project Outline

The aims of this work are twofold: firstly a new approach towards visualizing metabolic pathways is developed and presented. Secondly two new approaches for visualizing the result of simulating a metabolic pathway are presented.

The paper is organized as follows: the following section will give a brief overview of the history of metabolic network analysis and visualization. Subsequently the visualization approach followed by the analyzing techniques is presented. To sum it up, a short conclusion is given.

RELATED WORK

Network-based metabolic pathway analysis has a relatively short history. It started with Bruce Clarke [Cla80] who relied mainly on kinetic information and used convex analysis techniques (a branch of mathematics that can be used to analyze systems of linear inequalities [Roc70]), to study stability in chemical reaction networks. An actual mathematical formalism for analyzing biochemical pathways was then introduced by Seressiotis and Bailey [SB88] who presented an AI based algorithm to generate a set of pathways able to transform a given substrate to a given product by using a database containing enzyme and metabolite information. This algorithm was further advanced by Mavrovouniotis et al. [MM90], to deal with multiple reactants and products. The generated solutions could then be compared to the pathways found in nature.

To create an even more general approach to pathway analysis, researchers began to apply techniques from linear algebra to models of metabolic networks based on the stoichiometric matrix. The linear basis of the stoichiometric matrix was suggested as a set of pathways that would characterize all the possible routes through a metabolic network [Red88]. This approach was applied within the well-studied framework of metabolic control analysis (MCA) [RH77, KB73], and was used to describe kinetic and structural features of metabolic networks [Fel92, Red88].

About ten years ago, the mathematics of convex analysis was rediscovered in this context and used to develop an algorithm that generates a unique set of pathways called elementary modes [SS94]. This algorithm allows the identification of unique basis vectors, and thus the evaluation of invariant systems properties, such as estimating the maximum theoretical yield of a particular product.

These and many other algorithms and analyzing techniques have been implemented and improved over the last years, and are now part of several software frameworks that deal with the analysis of metabolic pathways. Examples are: GEPASI [Men97], Jarnac [M.00], DBSolve [GHS99], Patika [DBD$^+$02, DBD$^+$04], or Simpathica [APP$^+$03, APUM03].

Graph visualization, on the other hand, has been extensively studied in information visualization (see [HMM00] for a survey). However, most of the visualization methods dealing with metabolic pathways (like those described in [DBETT99]) are designed to display the well established textbook view of things: isolated, static two-dimensional images, like in Cytoscape [PS03], JDesigner [SHF$^+$03], and Osprey [BST03]. However, these are not capable of fulfilling the special requirements of biochemists and biologists, such as dynamically generated views of the underlying data, and the need for visual context information [GHM$^+$02, Men00, KZM$^+$04, FPR$^+$02].

Laying-out and subsequently visualizing graphs adequately in general and in the field of metabolic pathways in particular, has been commonly done by hierarchical or force-directed methods. While hierarchical layout techniques (e.g. Sugiyama layout [STT81]) are mainly used for small pathways, force-directed

layout algorithms are capable of producing layouts of bigger graphs, up to a couple of hundred nodes. In this work we will focus on graphs too big to be handled by the hierarchical approach.

"Force-directed" layout (also known as "spring-embedder") algorithms model the graph as a system of springs in between vertices connected by an edge. Thus, connected vertices attract each other, whilst others act repelling. The algorithm now tries to find a minimum energy configuration of this system. First described by Tutte [Tut63], it was adapted and enhanced during the 1990s [FR91, DH96, HK00]. For a survey see [Bra95], who compare "force-directed" methods for constructing straight-line drawings of general undirected graphs. The results show this approach can achieve very successful layouts. However, it also has severe difficulties (in terms of runtime) when dealing with graphs of more than a few hundred nodes, because in each single iteration each node pair is evaluated, which leads to a complexity of order $O(n^2)$. Furthermore, the convergence to a global minimum is not guaranteed. Walshaw [Wal01] developed a multi-level "spring-embedder" approach that was able to lay-out a graph with $75,000$ nodes in less then 12 minutes[1]. However, due to the heuristic and partitioning he used, the algorithm is not well suited for graphs containing vertices of very high degree, which is the case in metabolic pathways.

Another method to layout big graphs is the "High Dimensional Embedding" (HDE) technique first introduced by [HK02]. This technique is capable of displaying a graph having up to 10^5 nodes in nearly real-time (less than 3 seconds[2]) and its complexity is linear in the number of nodes and edges. It is based on the observation that it is much easier to first embed and lay-out a graph in a rather high dimensional space, and then to project this embedding into the desired visualization space (usually two or three dimensions).

The major drawback here (as we will see later) is that the algorithm is not suitable for drawing trees or tree-like graphs. This might be due to the fact that the high dimensional drawing of these graphs spans a "wild" subspace of quite a high dimension, and therefore makes it hard to find a good way to project back to a lower dimensionality.

NETWORK VISUALIZATION

To gain insights of what is going on in the modeled pathway, it is important to understand not only the function of each interaction itself, but also the network as a whole in which the pathway occurs. Thus, a method capable of structurally representing the network in a form that can be easily understood by humans would be of great value.

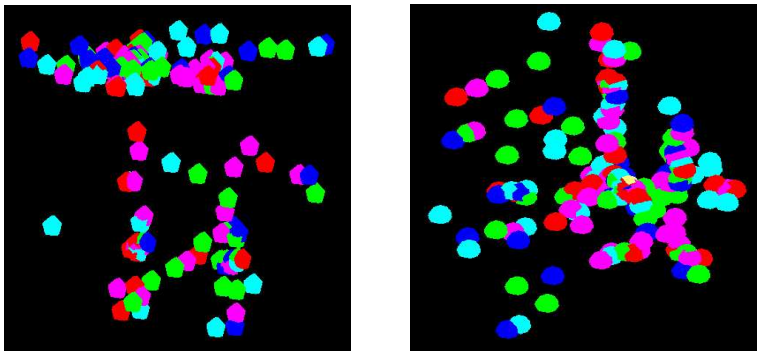
HDE and Octrees

As mentioned in the background chapter, the "High-Dimensional Embedding (HDE)" algorithm by Harel and Koren has severe difficulties when dealing with tree like structures. This is due to the fact that a tree spans a sub-space in the high-dimensional space of quite a few dimensions, and every dimension is as good a choice for projection as any other. As a result, parts of the projection look quite good and reveal a lot of structural information, while others are clustered and very densely packed with nodes. Unfortunately a metabolic pathway, despite being actually a graph, has many tree-like properties, which make the plain HDE algorithm unsuitable for metabolic pathways (see Figure 2).

To overcome this problem, we developed the following technique: We first use HDE to get a first layout of the pathway. Subsequently, we apply an adaptive 3-dimensional mesh, known as "Octree". An Octree (also known as adaptive space partitioning) is a way of subdividing the 3-dimensional space into cubes. It starts with a cube in which the entire data set fits. This cube is then split into eight sub-cubes (hence the name), each of which is subdivided again in eight sub-cubes, and so on, until the cube contains less then a predefined amount of points. Child cubes that would not contain any data points at all, are not created. The cubes containing the actual points are called leaves. In 2001, Quigley et al. [Qui01] presented a similar approach in two dimensions (using quad-trees) and gaining the initial layout by geometric clustering.

[1] on a Sun SPARC Ultra with 333MHz
[2] on a single-CPU Intel 1.7GHz PC

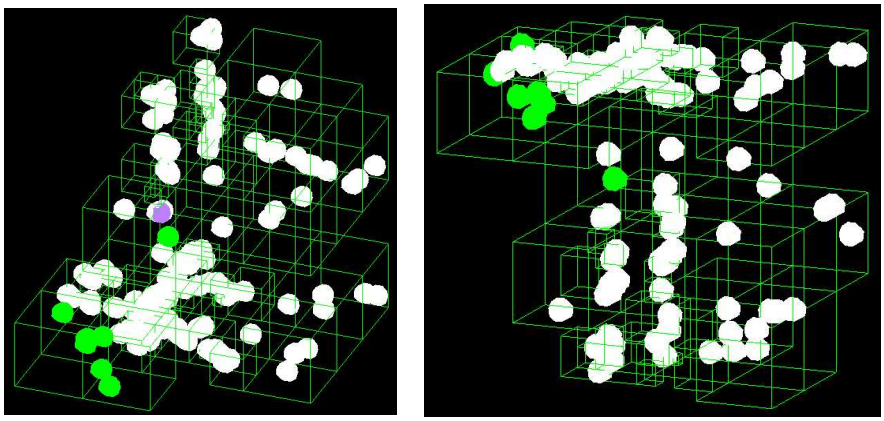
Figure 2: E.coli (1154 vertices, 2877 edges) after applying the HDE method and projecting back to 3-dimensional space (from different points of view)



(a) (b)

Applying the Octree algorithm to the metabolic pathway from Figure 2 yields a separation of the 3-dimensional space which is shown in Figure 3. It can be clearly seen that the density of points in

Figure 3: E.coli (1154 vertices, 2877 edges) after applying the HDE, and creating an Octree (10,10)



(a) (b)

some areas is significantly higher than in others, and some points seem to overlay. The third and final step applies the force-directed spring-embedder approach on the regions of high density identified in the previous step. As mentioned in the background section, this class of algorithms cannot deal with graphs having more then a few hundred nodes, because the runtime increases quadratically with the number of nodes. Thanks to the Octree analysis, we can guarantee an upper bound on the nodes the spring-embedder algorithm has to deal with.

To speed up this last step even more, we preprocess the vertices in each region before the actual application of the spring-embedder ([MR02]) to find an initial lay-out of the nodes in each of the regions, which (on average) leads to fewer invocations of the spring-embedder to produce a stable drawing. Experiments have shown, that this preprocessing step can reduce the overall calculation time by a factor of 10, while obtaining a similar quality of the lay-out. After all regions have been processed, the Octree is calculated again and the process is repeated until a predefined number of iterations have been performed, or no noticeable changes are made.

NETWORK ANALYSIS

This section will describe the algorithms and methods developed and used in this project for the analysis of the simulations of the metabolic networks. As mentioned in the introduction, the dynamical evolution (that is, the change in substance concentration) in a metabolic network is simulated in this work by using the ODE based Jarnac simulator [M.00]. Our framework transmits the model description to the Jarnac program and receives the calculated time series results via the SBW framework [HFS+02].

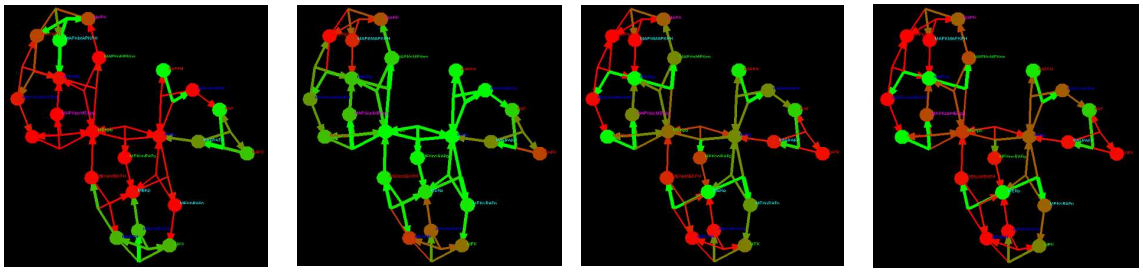
In order to visualize these results, our tool displays them in an XY-plot style in which each curve represents the changes in concentration of each substance. This simple visualization style can be confusing and counterintuitive when it involves many different curves. One possible solution is to display only a small number of curves simultaneously in one diagram. However, by doing this, the relationship between the different curves ("The big picture") gets lost, and only particular sections of the network are visible. We have developed two approaches to preserve the "global view": mapping the concentrations into the actual visual representation of the network (visualizing the flow), and visualizing and analyzing of the results in a so called "Fluvial Topography".

The purpose of both techniques is two-fold: to reveal relationships between the concentration of substances of a single simulation run (detecting clusters), and to detect differences between two simulation runs after changing properties of the simulated network.

Visualizing the flow through the network

The first approach we have taken to preserve the "global view" while still showing the concentration of all substances involved, is very simple but very effective: to map the concentrations of each metabolite to the actual visualization of the network. This can be thought of as a movie, where each frame shows a time step of the simulation. Each substance in every frame is drawn in a color which reflects its current concentration: Dark red represents extremely low concentration, while bright green represents extremely high concentration (see Figure 4 for an example). As can be seen in the Figure, the actual "flow" of the substances throughout the whole network becomes visible. Even different stages of the flow become identifiable.

Figure 4: Results of a simulation (100 time steps) of the MAP-Kinase pathway and displaying them in the actual graphical representation of the network.



(a) Frame 1 (b) Frame 25 (c) Frame 75 (d) Frame 95

Of course, accurate visualization of the metabolic network in the first place is tightly linked to the success of this technique.

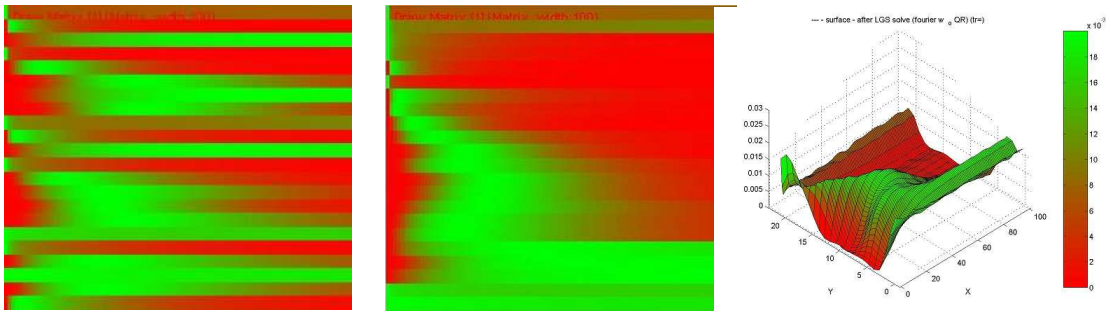
Fluvial Topography

In our second approach, we interpret the outcome of a simulation run as a matrix $S = \{s_{i,j}\}, \quad 1 \leq i \leq n, \quad 1 \leq j \leq m$, where n is the number of substances under scrutiny, and m is the number of time points in this simulation. Therefore, entry $s_{i,j}$ reflects the amount available of substance i at time j in the metabolic network. This matrix can be visualized by encoding the actual concentration in red and green

color nuances, as described previously. A XY-plot can be transformed to the diagram shown in Figure 5 (left), where each row represents one substance.

The key-idea is to then reorder the rows in the initial matrix S in such a way that two neighboring horizontal lines are as equal as possible to each other (see Figure 5). The initial row order is found by randomly choosing a row and then "greedily" placing the most similar row next to it, then finding the most similar to this one, and so forth. Subsequently, a continuous representation of the data (that is, a two-dimensional function which can be subject to further analyses) is sought for. This can reveal interesting properties such as substances with similar concentrations at the same time, or substances which never have a high or low concentration in the simulated time frame, given the initial conditions.

Figure 5: Result matrix before and after applying the GA



(a) Original (b) Sorted (c) Approximated 3D

The following section describes how we implemented this idea.

To be able to define a distance between two rows, firstly a distance measure has to be defined. We have chosen a simple and computationally cheap euclidian based approach (other options are certainly possible):

$$dist(s_1, s_2) = \sum_{i=0}^{N} \big(conc(s_1, i) - conc(s_2, i)\big)^2 \tag{1}$$

where $conc(s, i)$ gives the concentration of substance s at time i, and N is the number of time points in the current simulation. Minimization of equation (1) for all combination of substances, is then used to find a rearrangement of the rows, such that related (from the point of view of their concentration level) substances are placed in nearby locations. Obviously, this problem is NP-hard.

To get a good solution in reasonable time we decided to utilize a metaheuristic. In particular, we chose a genetic algorithm approach combined with a local search after each iteration, known as called "Memetic Algorithms (MA)". Memetic algorithms were first mentioned in 1989 by [Mos89] and are also often referred to as "Hybrid Genetic Algorithms" or "Genetic Local Search", because they basically extend the classical genetic approach by a local search heuristic. The local search property makes them even more applicable for parallel computing environments. For a survey and a comparison to other metaheuristic based approaches see [PMF04].

The implementation we use in this thesis is based on the work of [CMG+03], who proposed an MA suitable for the "Microarray Gene Ordering" problem, another NP-hard problem with strong implications in Medicine and Biology. It consists in ordering a set of genes, and grouping together the ones with similar behavior. We will now give a brief outline of the algorithm; we refer the interested reader to the original article for a more detailed description.

To measure the "fitness" (that is the quality) of a potential solution, a so called fitness function is needed. Let $\pi = \{\pi_1, \pi_2, \ldots, \pi_n\}$ be the order of the n substances of the current solution. The fitness of π is then defined as:

$$fitness(\pi) = \sum_{l=1}^{n} \sum_{i=min(l-s_w,1)}^{max(l+s_w,n)} w(i,l) \cdot dist(\pi_l, \pi_i) \tag{2}$$

where: $\boldsymbol{dist}(\cdot)$ is the function introduced in Equation (1), $\boldsymbol{s_w}$ reflects the number of neighboring rows to both sides[3] involved in each partial distance calculation (hence, the actual "window size" is $2s_w + 1$), and $\boldsymbol{w(i,l) = s_w - abs(l-i) + 1}$ weights the influence of the substance located at position i on the substance located at position l. This is obviously linearly proportional in the distance of the rows.

The representation chosen for this problem takes some ideas from hierarchical clustering. A solution is represented as a binary tree whose leaves are the substances. Thus, besides the actual order of the substances, a solution also contains information about the level of relationship among the substances. This has the advantage of being less disruptive, when producing offspring by using genetical operators like recombination and mutation.

To produce the next offspring generation we use both crossing over and mutation and tailor them to meet the binary tree specifications. After n (usually $100 \leq n \leq 500$) iterations without improvements, a local search is applied. Two local search strategies are used in this approach: (1) is a pairwise interchange, where each pair of adjacent substances is tested to see if a swap results in a better fitness value, and (2) which acts on the binary tree representation and inverts the branches of every sub-tree. Experiments have shown, that this often leads to joining together formerly separated groups that contain similar substances (in terms of concentrations).

As can be seen in Figure 5 the results are quite promising. The result shown was achieved after 25 iterations in less then 5 seconds. To be able to tackle even larger simulation results and larger metabolic networks with many substances, a parallelized "Island model" version of this approach was implemented. Island model GAs are based on the phenomenon of an island chain where natural populations evolve relatively isolated from each other but individuals within some particular (sub)population (or island) occasionally migrate to another. The island model is a natural choice for implementation on non-shared memory systems. This model involves periodic communications between processors and is known for a good performance in terms of efficiency and effectiveness.

After obtaining a good re-ordering of the substances in terms of clustering similar rows, the last step is performed: the calculation of a 2-dimensional function, which approximates the discrete data-points. Recall that the discrete data is stored in a matrix $S = \{s_{i,j}\}, \quad 1 \leq i \leq n, \quad 1 \leq j \leq m$, where n is the number of substances under scrutiny, and m is the number of time points in this simulation.

The basis for this step is the creation of a 2-d gradient vector field from the ordered discrete data. This and the following steps are entirely programmed in the Matlab^{TM} toolkit. The gradient vector field is achieved by calculating two matrices FX, FY from the matrix S. FX corresponds to $\frac{\delta S}{\delta x}$, the differences in the x (column) direction. FY corresponds to $\frac{\delta S}{\delta y}$, the differences in the y (row) direction.

We used a 2-D Fourier-row approach, introduced in [Con03], for the integration of the vector-field data and thus the approximation of the function:

$$f(x,y) = \sum_{k,l=0}^{N} \big(a_k \cdot \sin(k_x) + b_k \cdot \cos(k_x)\big) \cdot \big(c_l \cdot \sin(l_y) + d_l \cdot \cos(l_y)\big) \tag{3}$$

[3] Experiments have shown that taking into account more than just the immediate neighboring rows into the distance calculations leads to a significantly better grouping.

The parameter k, l are given by the x- and y-coordinates of the vector-field grid. N is the degree the Fourier-row is computed to. To determine the coefficients a, b, c and d a linear equation system has to be solved:

$$\frac{\delta F(x,y)}{\delta x} = \delta_x F, \qquad \frac{\delta F(x,y)}{\delta y} = \delta_y F \tag{4}$$

F is the function sought, and $\delta_x F$ and $\delta_y F$ the components of the negative gradient from (5).

$$A \cdot z = b \tag{5}$$

The matrix A consists of all terms from (3) without the coefficients ($a \dots d$), the vector z consists of the coefficients $a \dots d$, and the vector b contains the components of the negative gradients ($\delta_x F$, $\delta_y F$).

To improve the bad condition of the left side, the original equations system (5) was transferred to an equivalent system via QR-decomposition (pre-conditioning, see [PD02]). A is split via QR-decomposition into two triangular matrices and the actual calculation is performed as follows:

$$R \cdot z = Q^T \cdot b, \quad \text{because} \quad A = Q \cdot R \quad \text{and} \quad Q^T \cdot Q = I \tag{6}$$

Thus, the badly conditioned part of the Vandermonde matrix is no longer considered in the actual calculation, and provides an estimation of the error at the same time [PD02].

Having calculated F, we now have a function, which is a continuous representation of the concentrations, or, more interestingly, a system of dependent functions in two dimensions, which can be plotted and/or analyzed. Thus, the function reflects inherent properties of the system, such as dependencies, similarities, and differences. Because it is in two dimensions, it not only reveals properties between substances, but also in the evolution of a substance over time. Furthermore, because of the dependencies, the simulation for "well behaved" substances can be done for less time-points, and the missing parts can be approximated. This approximation is not only a linear interpolation between the values available but, thanks to the properties of fourier rows, it preserves much more of the structure.

With this representation it is also possible to spot and calculate differences between two models (such as the original pathway and a modified version) fairly easily with basic mathematical tools.

CONCLUSION

We have presented a new approach to visualize big metabolic pathways and networks which previously could not have been tackled with common force-directed methods.

Furthermore, we presented two new approaches for the visualization of simulation data. These techniques have proved useful in actually analyzing metabolic pathways. For example, changing parameters of the network and subsequent comparison of the resulting flows through the network or the topologies reveals important properties of the pathway analyzed. Moreover, the ordering of the substances with regards to their concentrations reveals correlations of single substances and functional parts of the networks where they occur. Again, changes in the parameters of the pathways and comparison of the resulting topologies can unveil interesting properties, such as the importance of single reactions, and give hints for further investigations.

ACKNOWLEDGEMENTS

I wish to thank my supervisor Dr. Maria Garcia de la Banda from the "Optimization and Constraint Solving Group" at Monash University for her assistance in this project, and for many helpful discussions, ideas and hints. Furthermore, I would like to thank the three anonymous reviewers for their useful and helpful comments on the first submission.

REFERENCES

[APP+03] M Antoniotti, F Park, A Policriti, N Ugel, and B Mishra. Foundations of a query and simulation system for the modeling of biochemical and biological processes. *Pac Symp Biocomput*, pages 116–27, 2003.

[APUM03] Marco Antoniotti, Alberto Policriti, Nadia Ugel, and Bud Mishra. Model building and model checking for biochemical processes. *Cell Biochem Biophys*, 38(3):271–86, 2003.

[Bra95] F. J. Brandenburg. Designing graph drawings by layout graph grammars. *Lecture Notes in Computer Science*, 894:416–??, 1995.

[BST03] Bobby-Joe Breitkreutz, Chris Stark, and Mike Tyers. Osprey: a network visualization system. *Genome Biol*, 4(3):R22, 2003.

[Cla80] B.L. Clarke. *Stability of Complex Reaction Networks*. Advances in Chemical Physics. John Wiley, New York, 1980.

[CMG+03] C. Cotta, A. Mendes, V. Garcia, P. França, and P. Moscato. Applying memetic algorithms to the analysis of microarray data. In G. Raidl et al., editors, *Applications of Evolutionary Computing*, volume 2611 of *Lecture Notes in Computer Science*, pages 22–32. Springer-Verlag, Berlin, 2003.

[Con03] T. Conrad. Berechnung von freien energie flaechen am beispiel des alanin-dipeptides. *BSc Thesis, Freie Universitaet Berlin, FB Mathematik und Informatik, Institut fuer Mathematik II*, July 2003.

[DBD+02] E Demir, O Babur, U Dogrusoz, A Gursoy, G Nisanci, R Cetin-Atalay, and M Ozturk. Patika: an integrated visual environment for collaborative construction and analysis of cellular pathways. *Bioinformatics*, 18(7):996–1003, Jul 2002.

[DBD+04] E Demir, O Babur, U Dogrusoz, A Gursoy, A Ayaz, G Gulesir, G Nisanci, and R Cetin-Atalay. An ontology for collaborative construction and analysis of cellular pathways. *Bioinformatics*, 20(3):349–56, Feb 2004.

[DBETT99] G. Di Battista, P. Eades, R. Tamassia, and I.G Tollis. *Graph drawing - algorithms for the visualization of graphs*. US Imports & PHIPEs, 1999.

[DH96] Ron Davidson and David Harel. Drawing graphs nicely using simulated annealing. *ACM Trans. Graph.*, 15(4):301–331, 1996.

[Fel92] D. A. Fell. Metabolic control analysis: a survey of its theoretical and experimental development. *Biochem. J.*, (286):313330, 1992.

[FPR+02] Michael Forster, Andreas Pick, Marcus Raitner, Falk Schreiber, and Franz J Brandenburg. The system architecture of the biopath system. *In Silico Biol*, 2(3):415–26, 2002.

[FR91] Thomas M. J. Fruchterman and Edward M. Reingold. Graph drawing by force-directed placement. *Software - Practice and Experience*, 21(11):1129–1164, 1991.

[GHM+02] Alexander Goesmann, Martin Haubrock, Folker Meyer, Joern Kalinowski, and Robert Giegerich. Pathfinder: reconstruction and dynamic visualization of metabolic pathways. *Bioinformatics*, 18(1):124–9, Jan 2002.

[GHS99] I Goryanin, TC Hodgman, and E Selkov. Mathematical simulation and analysis of cellular metabolism and regulation. *Bioinformatics*, 15(9):749–58, Sep 1999.

[HFS+02] M. Hucka, A. Finney, Herbert M. Sauro, H. Bolouri, J. Doyle, and H. Kitano. The erato systems biology workbench: enabling interaction and exchange between software tools for computational biology. In R. Altman et al., editor, *The Proceedings of the Pacific Symposium on Biocomputing 2002*, The Proceedings of the Pacific Symposium on Biocomputing 2002, January 2002.

[HK00] David Harel and Yehuda Koren. A fast multi-scale method for drawing large graphs. In *Proceedings of the working conference on Advanced visual interfaces*, pages 282–285. ACM Press, 2000.

[HK02] David Harel and Yehuda Koren. Graph drawing by high-dimensional embedding. In *Revised Papers from the 10th International Symposium on Graph Drawing*, pages 207–219. Springer-Verlag, 2002.

[HMM00] Herman, G. Melançon, and M. S. Marshall. Graph visualization and navigation in information visualization: A survey. *IEEE Transactions on Visualization and Computer Graphics*, 6(1):24–43, /2000.

[KB73] H. Kacser and J.A. Burns. The control of flux. *Symp. Soc. Exper. Biol.*, (27):65104, 1973.

[KZM+04] Cynthia J Krieger, Peifen Zhang, Lukas A Mueller, Alfred Wang, Suzanne Paley, Martha Arnaud, John Pick, Seung Y Rhee, and Peter D Karp. Metacyc: a multiorganism database of metabolic pathways and enzymes. *Nucleic Acids Res*, 32 Database issue:D438–42, Jan 2004.

[M.00] Sauro H. M. Jarnac: a system for interactive metabolic analysis. In Hofmeyr J-H. S, J. M Rohwer, and Snoep J. L, editors, *Animating the Cellular Map 9th International BioThermoKinetics Meeting*, number 33 in Animating the Cellular Map 9th International BioThermoKinetics Meeting. Stellenbosch University Press, 2000.

[Men97] P Mendes. Biochemistry by numbers: simulation of biochemical pathways with gepasi 3. *Trends Biochem Sci*, 22(9):361–3, Sep 1997.

[Men00] P Mendes. Pathdb: a second generation metabolic database. *9th International BioThermoKinetics Meeting*, 2000.

[MM90] et al. M.L. Mavrovouniotis. Computer-aided synthesis of biochemical pathways. *Biotechnol. Bioeng.*, (36):11191132, 1990.

[Mos89] Pablo Moscato. On evolution, search, optimization, genetic algorithms and martial arts: Towards memetic algorithms. Technical Report C3P 826, Pasadena, CA, 1989.

[MR02] Paul Mutton and Peter Rodgers. Spring Embedder Preprocessing for WWW Visualization. In *Proceedings Information Visualization 2002*. IVS, IEEE, July 2002.

[PD02] A. Hohmann P. Deuflhard. *Numerische Mathematik I: 3., Uberarbeitete Auflage*. Walter De Gruyter Inc, 2002.

[PMF04] Q. Zhang P. Mills, E. Tsang and J. Ford. A survey of ai-based meta-heuristics for dealing with local optima in local search. Technical Report CSM-416, Department of Computer Science, University of Essex, Colchester CO4 3SQ, September 2004.

[PS03] Trey Ideker et al. Paul Shannon, Andrew Markiel. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res*, 13(11):2498–50, Nov 2003.

[Qui01] Aaron J. Quigley. Large scale 3d clustering and abstraction. In Peter Eades and Jesse Jin, editors, *Selected papers from Pan-Sydney Workshop on Visual Information Processing*, volume 2 of *Conferences in Research and Practice in Information Technology*, pages 117–118, Sydney, Australia, 2001. ACS.

[Red88] C. Reder. Metabolic control theory: a structural approach. *J. Theor. Biol.*, (135):175201, 1988.

[RH77] et al. R. Heinrich. Metabolic regulation and mathematical models. *Prog. Biophys. Mol. Biol.*, (32):1–82, 1977.

[Roc70] R. T. Rockafellar. Convex analysis. Princeton University Press, 1970.

[SB88] A. Seressiotis and J.E. Bailey. Mps: an artificially intelligent software system for the analysis and synthesis of metabolic pathways. *Biotechnol. Bioeng.*, (31):587602, 1988.

[SHF+03] Herbert M Sauro, Michael Hucka, Andrew Finney, Cameron Wellock, Hamid Bolouri, John Doyle, and Hiroaki Kitano. Next generation simulation tools: the systems biology workbench and biospice integration. *OMICS*, 7(4):355–72, 2003.

[SS94] C. Hilgetag S. Schuster. On elementary flux modes in biochemical reaction systems at steady state. *J. Biol. Syst.*, (2):165182, 1994.

[STT81] K. Sugiyama, S. Tagawa, and M. Toda. Methods for visual understanding of hierarchical system structures. *IEEE Transactions on Systems, Man and Cybernetics*, SMC-11(2):109–125, February 1981.

[Tut63] W. T. Tutte. How to draw a graph. In *Proceedings London Mathematical Society*, volume 13 of *3*, pages 743–768, 1963.

[Wal01] Chris Walshaw. A multilevel algorithm for force-directed graph drawing. In *Proceedings of the 8th International Symposium on Graph Drawing*, pages 171–182. Springer-Verlag, 2001.