Reviewer's report

Title: Mimoza: Web-Based Semantic Zooming and Navigation in Metabolic Networks

Version:1**Date:**16 September 2014

Reviewer number:4

Reviewer's report:

Major Compulsory Revisions

* Background

1) "Generalization identifies and groups similar metabolites and similar reactions in the network."

What means similar?

* Metabolic network reconstruction and infrastructure

2) "... while a human can understand at best networks that have hundreds of reactions."

How did the authors come to this conclusion? A reference would be helpful.

* Existing visualisation approaches

3) "that zooming in increases, but does not change or add"

It's not clear what "increases" means in this context. Should it be "resizes"?

* Implementation

* Choosing zoom levels

4) "The most abstract level represents compartmentalization of the model, and focuses on such questions as: Are all the compartments present? Are they well connected by transport reactions? This level shows the compartments of the model, the transport reactions between them, and other reactions happening inside the cytoplasm."

The assumption seems to be that compartments are always defined. What happens if compartment are not defined at all and only the default compartment exists?

5) "Are all the essential processes present?

What are essential processes? How do the authors decide?

6) "Is the structure of each process correct?"

What is the structure of a process?

* Model generalisation

7) The generalization approach has been already published. Nevertheless it would be good to provide some details necessary for understanding this article.

8) In addition to the textual description of the generalization approach it would be good to have an example illustration which shows how the generalization works.

9) What is the ratio between metabolites and reactions which can be generalized and which can't be generalized?

10) What happens if no generalization is possible at all?

11) "... metabolites that do not participate in any pair of similar reactions are not grouped together ..."

"Reactions that share the same generalized reactants and the same generalized products, are considered equivalent and are factored together into a generalized reaction."

This seems to be a race condition.

It reads like metabolites are grouped if they take part in reactions which are grouped, and reactions are grouped based on metabolites which are grouped.

12) "We use the ChEBI ontology, as it is the de facto standard for metabolite annotation in metabolic networks."

Since a name for a metabolite can have any form, what is done if no information is available from ChEBI?

13) "Grouping metabolites increases the number of reactions they participate in ..."

Why does this increase the number of reactions they take part in?

Wouldn't this change the model?

* Layers Layout

14) How are SBML models (the respective SBML files) handled which provide layout information using the SBML layout extension? Please explain.

* Generalized model layout

15) "... the sub-models corresponding ..."

Are these really sub-models or more sub-networks?

16) "We divide the compartment graph into connected components ..."

How is this done? Please, explain.

17) "... an appropriate layout algorithm ..."

What is an appropriate layout algorithm? Please, explain.

18) "Depending on the nature of the connected component subgraph ..."

How is the nature (structure) detected? Please, explain.

19) "... connected component subgraph ..."

What is the connected component subgraph?

20) "* Hierarchical Layout for the components that contain no cycles,

² Circular Layout for the components with less than 100 nodes and less than 3 cycles,

² Force-Directed Layout for all the other components."

Why did the authors choose these algorithms and these graph structures? Please, explain.

21) "To avoid clutter we duplicate all the ubiquitous metabolites before applying the layout algorithms, so that there is a copy of an ubiquitous metabolite for each reaction it participates in."

What happens if the network is split into sub-networks during duplication? Please, explain.

* Generalization-based full model layout

22) "... place similar metabolites or reactions next to each other inside the space used by the corresponding generalized metabolites or reactions

in the generalized model."

How is this done in terms of algorithms?

Does this influence the size of the generalized metabolites and reactions which in turn would influence the generalized layout? Which raises the question, how is the size for the generalized elements determined? Please, explain.

* Relative positions of the compartments

23) How are transport reactions between compartments handled? Please, explain.

* Results and Discussion

24) "... the main processes ..."

What are main processes, how are they determined? Please, explain.

25) "... making the elements at the most general level very small and hard to analyze."

But these elements don't have to be analyzed at the most general level. Using your approach they are even not provided at the most general level. Please, discuss in more detail.

26) "Mimoza especially targets draft models during curation, allowing one to visualize them fully automatically and help to analyze them in a top-down manner, starting from the general structure and going down to the details."

A modeling use case is missing to really show how Mimoza can help in the modeling process. It would also be helpful to discuss/review the modeling workflow and to show how and where Mimoza can be integrated in this workflow.

27) "... (it also tries to annotate them automatically if no annotations are present)."

Either explain in more detail how this is done or remove this sentence.

* References

28) The reference section needs to be reviewed.

For a couple of references information such as journal, issue or page numbers is missing, incomplete or not correct (e.g., ref [1], [4], [5], [6], [14], [15], [19], [20], [23], [25], [26], [29], [30], [32], [39], [41]). Journal names should be written in a consistent manner (Journal of integrative bioinformatics -> Journal of Integrative Bioinformatics, BMC bioinformatics -> BMC Bioinformatics, BMC systems biology -> BMC Systems Biology, Nucleic acids research -> Nucleic Acids Research).

* General comments

29) The order of figures should be reviewed. Figures are referenced Figure 2, Figure 3, Figure 1.

30) The authors should make a decision to use either BE or AE (e.g., generalisation vs generalization).

31) The authors should consider to add a section the describe system requirements for the software.

* Software

I have tried to use the on-line version of the software but there seem to be a couple of problems.

32) I have used this model https://www.ebi.ac.uk/biomodels-main/download?mid=BMID000000101169 but it didn't work (Internet Explorer 11).

When I clicked on "Go!", the message "We are currently processing your model..." appeared, but I ended up on the page with the "Go!" button again and the link to the result page didn't work at all.

I have also tried to use this model https://www.ebi.ac.uk/biomodels-main/BIOMD0000000012. It didn't work as well.

Then I have tried to view the example from the article (http://mimoza.bordeaux.inria.fr/yeast4/comp.html?id=C_8).

Here I could only see a compartment "Peroxi...", the next zoom level was a generalized version, then I got another generalized version on the next zoom level.

On the next zoom level I could see reactions and metabolites but I couldn't read labels (needs mouse over).

Then there is another zoom level with reactions and metabolites but again I couldn't read labels (needs mouse over).

33) Why are there five zoom levels? In the article the authors talk about three levels.

34) Shouldn't the link in the article be http://mimoza.bordeaux.inria.fr/yeast4/comp.html?id=C_1 to reproduce the images from the article? But then it's not possible to zoom into a compartment (even when I click on a compartment).

35) Internet Explorer 11 utilises one core completely on my machine, needs a lot of memory (more than 400MB), and finally isn't usable for a couple of seconds.

Sometimes it even happens that I got the message "Script on inria.fr doesn't respond".

36) On the start page of Mimoza there are two views of models. Why are there two views?

37) The location of pop-ups (on mouse over) is always the bottom right? Pop-ups for nodes at the right or at the bottom (or both) are not visible then.

38) The view at all is very small and uses only a part of the screen. It might be good if the size can be increased/changed by the user.

39) Is the zoom point always the centre of the view? It might be better to use the location of the mouse cursor.

40) The file open dialogue should be adapted to allow sbml (maybe also xml) files only.

Minor Essential Revisions

* Metabolic network reconstruction and infrastructure

1) "CEllML"

CellML

2) "representated"

represented

* Existing visualisation approaches

3) "VANTED[22], Cytoscape[23]"

VANTED[22], and Cytoscape[23]

4) "the ER Stress response[25] map, was created manually in CellDesigner."

According to sbgn.org this map was not the winner of the competition.

5) "force-directed algorithm."

force-directed layout algorithm
Please, add a reference for the force-directed layout algorithm.

6) "pathway information in present in the model"

pathway information is present in the model

7) "SUBSYSTEM"

subsystem

* Implementation

* Choosing zoom levels

8) "Decomposition is performed by splitting the model into compartments and after with the metabolic model generalisation method, so the most appropriate is to adopt 3 levels of semantic zooming: ..."

It seems that something is missing in this sentence.

* Model generalisation

9) "... reactions operating with them in the network ..."

Please rephrase, to operate is probably not the right verb here.

10) "... stoichiometry would be broken ..."

Please rephrase, to break is probably not the right verb here.

* Pipeline

11) "web-cite"

web page / website

Discretionary Revisions

* Metabolic network reconstruction and infrastructure

1) "Genome-scale metabolic models include thousands of reactions that may participate in organism’s metabolism, ..."

It would be good to list examples for reference.

2) "KEGG[14] provides an extensive collection of pathways."

Maybe cite Reactome and PANTHER Pathways as well.

3) "A model representated in these formats can be further enriched with the knowledge from biological databases and ontologies, e.g. ChEBI[18], Uniprot[19], by annotating elements of the models (such as metabolites, reactions) with appropriate identifiers. To keep the identifiers representation unique and machine readable such standardisation efforts as Identifiers.org[20] emerge."

A model represented in these formats can be further enriched with knowledge from biological databases and ontologies, e.g. ChEBI[18], and Uniprot[19], by annotating elements of the models (such as metabolites, reactions) with appropriate identifiers. To keep the identifiers representation unique and machine readable standardisation efforts such as Identifiers.org[20] emerge.

* Existing visualisation approaches

4) "An approach different to a simple graph layout algorithm is needed."

Therefore an approach different to a simple graph layout algorithm is necessary.

* Implementation

* Generalized model layout

5) "... and then apply ..."

and apply

* Results and Discussion

6) "The first level shows ..."

"... the second level shows ..."

"... the most detailed level represents ..."

The first level (bottom) shows

the second level (middle) shows

the most detailed level (top) represents

* General comments

7) It might be good to adapt the structure of the implementation section and discuss the model generalization first followed by choosing zoom levels.

Level of interest: An article of importance in its field

Quality of written English: Acceptable

Statistical review: No, the manuscript does not need to be seen by a statistician.