



SOFTWARE TOOL ARTICLE

REVISED Biological network analysis with CentiScaPe: centralities and experimental dataset integration [version 2; referees: 2 approved]

Giovanni Scardoni¹, Gabriele Tosadori¹, Mohammed Faizan², Fausto Spoto³, Franco Fabbri¹, Carlo Laudanna^{1,4}

¹Center for Biomedical Computing, University of Verona, Verona, 37134, Italy

²Birla Institute of Technology & Science, Goa, 403726, India

³Department of Computer Science, University of Verona, Verona, 37134, Italy

⁴Department of Pathology and Diagnostics, University of Verona, Verona, 37134, Italy

v2 First published: 01 Jul 2014, 3:139 (doi: 10.12688/f1000research.4477.1)
 Latest published: 07 Jul 2015, 3:139 (doi: 10.12688/f1000research.4477.2)

Abstract

The growing dimension and complexity of the available experimental data generating biological networks have increased the need for tools that help in categorizing nodes by their topological relevance. Here we present CentiScaPe, a Cytoscape app specifically designed to calculate centrality indexes used for the identification of the most important nodes in a network. CentiScaPe is a comprehensive suite of algorithms dedicated to network nodes centrality analysis, computing several centralities for undirected, directed and weighted networks. The results of the topological analysis can be integrated with data set from lab experiments, like expression or phosphorylation levels for each protein represented in the network. Our app opens new perspectives in the analysis of biological networks, since the integration of topological analysis with lab experimental data enhance the predictive power of the bioinformatics analysis.



This article is included in the [Cytoscape apps](#) channel.

Open Peer Review

Referee Status:

	Invited Referees	
	1	2
REVISED version 2 published 07 Jul 2015	<input checked="" type="checkbox"/> report	<input checked="" type="checkbox"/> report
	↑	↑
version 1 published 01 Jul 2014	<input type="checkbox"/> report	<input type="checkbox"/> report

1 Ferenc Jordan, University of Trento Italy

2 Ankush Sharma, High Performance Computing and Networking Institute, National Research Council Italy

Discuss this article

Comments (0)

Corresponding author: Giovanni Scardoni (giovanni.scardoni@gmail.com)

How to cite this article: Scardoni G, Tosadori G, Faizan M *et al.* **Biological network analysis with CentiScaPe: centralities and experimental dataset integration [version 2; referees: 2 approved]** *F1000Research* 2015, **3**:139 (doi: [10.12688/f1000research.4477.2](https://doi.org/10.12688/f1000research.4477.2))

Copyright: © 2015 Scardoni G *et al.* This is an open access article distributed under the terms of the [Creative Commons Attribution Licence](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Grant information: This work was supported by: Italian Association for Cancer Research (AIRC, IG 8690) (C.L.); Fondazione Cariverona; Nanomedicine project University of Verona and Fondazione Cariverona (C.L.). Directed centralities have been developed thanks to the GSoC2013 program.

Competing interests: No competing interests were disclosed.

First published: 01 Jul 2014, **3**:139 (doi: [10.12688/f1000research.4477.1](https://doi.org/10.12688/f1000research.4477.1))

REVISED Amendments from Version 1

The first version of this paper was improved mainly by following the suggestions received from our reviewers. In this sense we updated the FastCentiScaPe web-tool, that is now working and that does not require the registration of the user's email anymore. The procedure simply requires a valid email to which the results will be sent as soon as they are ready. We also improved [Figure 1](#) by changing the layout. In the figure we highlighted three nodes, with a black arrow, that are used to show different topological characteristics. These three examples permit to describe the role of each node by adding few biological information and by proposing some directions for further investigations. The figure shows the results from both directed and undirected analysis and gives a comprehensive view of the data we obtained, related to the Stress centrality. Lastly we modified the description about the edge's weighting by adding a less concise and more clear explanation of its biological and topological meaning.

See referee reports

Introduction

Biological processes can be abstracted as networks where the nodes represent biological entities and the edges represent interactions between them. Several kind of biological networks have been described: metabolic networks, gene networks, signal transduction networks and protein-protein interaction networks¹. Such networks are a static representation of the dynamics of complex biological processes, where molecular interactions give rise to cascades of reactions called pathways, determining the life processes of living organisms. Even if the scientific community is far from the capability of simulating the dynamical behaviour of such pathways, several important information can be extracted from the topological analysis of biological networks^{2,3} since the structure of a network can affect its function⁴. In this context, different global parameters are commonly used to describe the properties of the whole network: centralities⁵ are indexes that permit the identification of the most important actors by characterizing the nodes that, more than others, are good candidates as regulators of the underlying biological process.

In this paper we present CentiScaPe 2.1, a Cytoscape app^{6,7} for network centralities analysis. While the built-in NetworkAnalyzer tool is oriented in characterizing the global behaviour of the network, provided with several global network statistics, CentiScaPe is designed to identify the most relevant nodes and provides a more comprehensive set of centralities. The new version introduces the computation of centralities for directed and weighted networks, two features that are not available in any other Cytoscape app. As in the previous version it computes Average Distance, Diameter, Degree, Stress, Betweenness, Radiality, Closeness, Centroid Value and Eccentricity. New indexes as Eigenvector, Bridging centrality and Edge Betweenness have been added, making it the most complete app for network centrality analysis. Some hypothesis about the centralities interpretation and application in a real biological context are presented in the CentralitiesTutorial (see [Supplementary materials](#) section).

A web version of CentiScaPe, i.e. FastCentiScaPe, is also available (<http://www.cbmc.it/fastcent/>). It performs very fast computations for large networks by sending the network to a multiprocessor server. The centrality analysis results are sent to the user by e-mail in a Cytoscape readable, i.e. *.xgmml*, format.

The main aim of CentiScaPe is to produce results that could drive further lab experiments since the high score nodes identified by the computation should be considered as potential targets for drugs and could be the starting point for further investigations and for a deeper understanding of the underlying biological process.

Methods and implementation

The main usage of CentiScaPe is to rank the nodes of a network depending on their topological and experimental relevance. The numerical results are saved as node, edge or network attributes in the Cytoscape attributes browser, depending on the kind of parameters, so all the Cytoscape features for managing attributes are supported; after the computation the centralities are treated as normal Cytoscape attributes. CentiScaPe can be used in undirected networks⁸, in directed networks and in weighted networks. Centralities for directed networks (see [Supplementary materials: CentralitiesTutorial](#)) are useful in the case of metabolic networks in which the direction describes the interaction between the substrates and reactants and the products of the chemical reactions and also in signal transduction networks, in which the direction depends on the flux of information. Considering the direction in the computation of centralities can lead to different results and interpretations than the undirected version.

As an example, in [Figure 1](#) we show the computation of the directed and undirected Stress applied to a network of Oncogenes (see [Supplementary materials: Oncogenes.txt](#) and [Oncogenes_edge_directions.txt](#)). Results of both the computations are shown and discussed.

The image, obtained using Cytoscape's graphical tool, represents the different Stress values by using the colour and the size of the nodes. The size describes the value obtained by using the directed Stress: the bigger the node the higher the value; the colour describes the value obtained by using the undirected Stress: red is used for the highest value, blue for the lowest value. For example a large blue node requires particular attention because it is showing a node with a high centrality value if the network is considered as directed but with a very low value if the network is considered as undirected.

By analyzing the Oncogenes network we saw that the large red node, i.e. AKT1, shows how its Stress values are high using both algorithms. It plays a central role in different cell processes like metabolism, proliferation, cell survival, growth and angiogenesis. This role may highlight its high Stress value but, on the other hand, the high values suggest us to deeply investigate its characteristics; it is also involved in two different kind of cancer: breast and colorectal (see <http://www.uniprot.org/uniprot/P31749>). This evidence suggest that the node could be involved in cancer related processes but this assertion needs validation from several lab experiments.

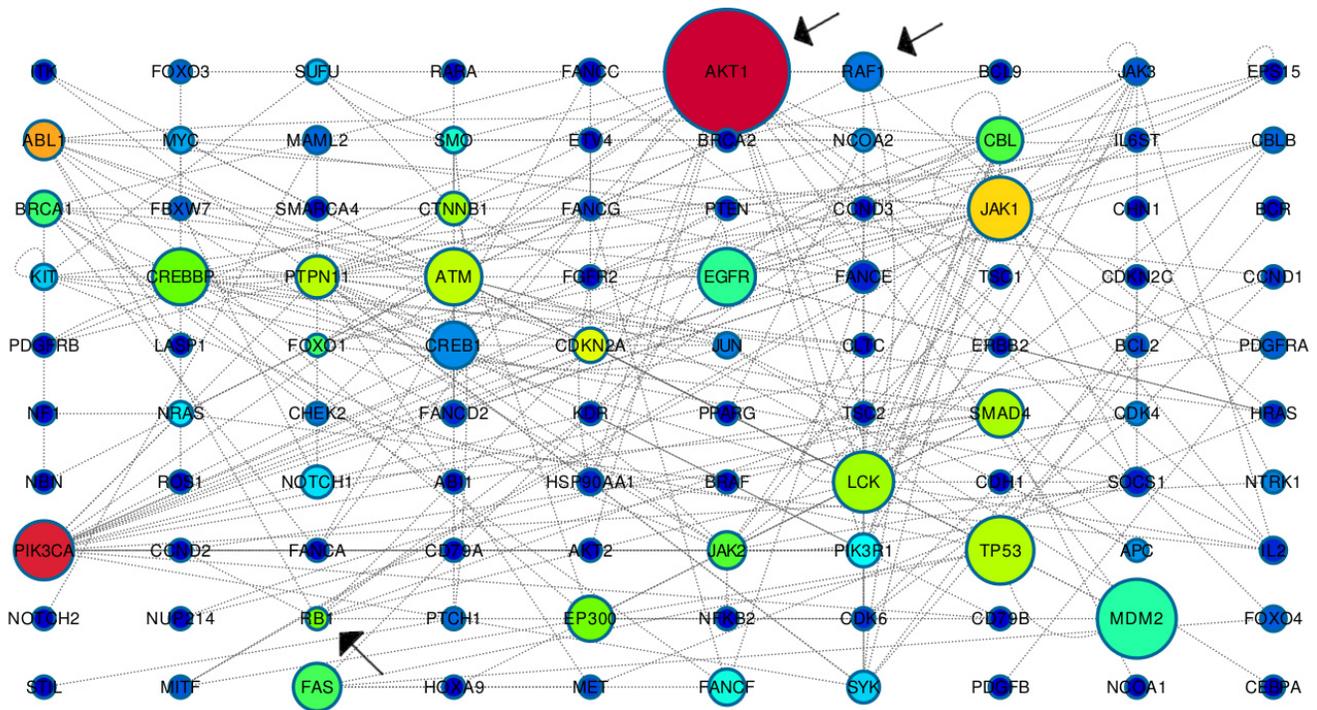


Figure 1. Stress computed on oncogenes. Size represents directed Stress, colour undirected Stress (blue=low, red=high).

Another interesting result is shown by the blue medium sized node RAF1. It shows how, using undirected Stress we obtained a low value, but by using the new algorithm we obtained an interesting Stress value. RAF1 was identified as a proto-oncogene with different and fundamental cellular functions (see <http://www.omim.org/entry/164760>). The results could be interpreted by saying that the directed network gives us a better understanding about how the gene, and its product, are involved in the development of cancer and could highlight that the use of the direction enhance our ability in describing a complex biological process.

The opposite situation is found in the third highlighted node, the small green node, RB1 in the right bottom corner. In this case the value computed with undirected Stress is not very high, but the value computed with the directed Stress is very low. RB1 is a gene involved in coding a protein involved in the retinoblastoma and other type of cancer like bladder cancer and osteogenic sarcoma. It was the first tumor suppressor gene found (see <http://www.ncbi.nlm.nih.gov/gene/5925>). As already said for RAF1, if the directed analysis is considered more reliable than the undirected one, then RB1 seems to be marginally involved in the Oncogenes network otherwise the undirected network is a better description. As already stated an experimental validation should be carried out in order to improve the results from the topological analysis and to better understand the role of each highlighted node.

All the results shown and described must be considered as a possible direction for further lab experiments. The main goal of this kind of analysis is to give us a comprehensive view that could be useful in describing the role of each node involved in a specific biological process and to drive future insights and investigations.

Second important features of the new version of CentiScaPe is the possibility of computing centralities for weighted networks, that are networks in which the edges are provided with an attribute that can be interpreted as a distance between the two connected nodes.

In the network in **Figure 2** we have a distance (*dist*) attribute for each edge. The values are $dist(A, B) = 2$, $dist(B, C) = 3$ and $dist(A, C) = 7$. Since A and C are connected by a single edge, in an unweighted computation, the distance from A to C is equal to one. But if the attributes of the edges are considered as distances, the shortest path between A and C is the one passing through B ($= 2 + 3 = 5$) since it is shorter, or *lighter*, than the one connecting A directly to C ($= 7$). The computation of weighted shortest paths will result in completely different values from the case wherein the weight is not considered. The user should consider that the weight is used in the sense that close nodes are more important than distant nodes. Therefore depending on the meaning of the attributes, one can use the real value or its reciprocal. For example if the attribute represents the speed of a reaction instead of a distance, the reciprocal

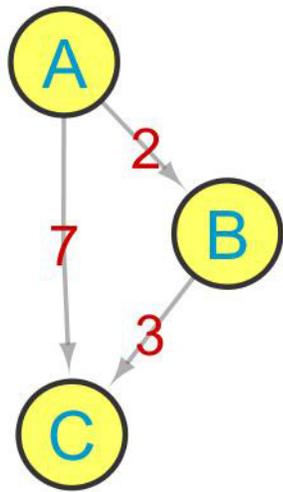


Figure 2. A weighted network. If the weight of the edge is not considered the shortest path from A to C is the directed edge, and the distance is one. If the weights are considered the shortest path from A to C is the one passing through B, and the distance is equal to 5.

should be used. This is because the higher the value of the speed, the nearer the nodes are: an increasing speed determines a decreasing reciprocal and the distance decrease by consequence.

An example of usage of weighted networks centralities analysis, can be found in Holly *et al.*⁹ where an euclidean distance is given to each edge depending on the difference between the phosphorylation level of the proteins connected by that edge.

All the graphical features of the previous version of CentiScaPe, as the plot-by-node, the plot-by-centrality and the boolean-based result panel have been maintained in the new version. A complete guide can be found in Scardoni *et al.*⁸ or in the CentiScaPe userguide available from the website.

Conclusions

CentiScaPe 2.1 have been enriched with new centrality parameters as Eigenvector, Bridging centrality and Edge Betweenness centrality and with the possibility to analyze directed and weighted networks. It allows integration centrality-based network analysis with experimental data by using nodes, edges or network attributes. The

results of the computation can be used directly and are easily portable as Cytoscape attributes allowing the user to exploit all the other features of Cytoscape and its apps. Compared to the Cytoscape's built-in tool, NetworkAnalyzer, CentiScaPe is an excellent integrative tool that allows the identification of potential target nodes from both the topological and the experimental point of view, and can be considered as an essential instrument for the characterizations of nodes in order to drive further experiments.

Software availability

Software available from the Cytoscape App Store

<http://apps.cytoscape.org/apps/centiscape>

Latest source code

<https://bitbucket.org/gioanniscardoni/centiscapepublic/>

Source code as at the time of publication

<https://bitbucket.org/F1000Research/centiscapepublic-archive>

Archived source code as at the time of publication

[http://dx.doi.org/10.5281/zenodo.10652¹⁰](http://dx.doi.org/10.5281/zenodo.10652<sup>10</sup)

License

Lesser GNU Public License 3.0: <https://www.gnu.org/licenses/lgpl.html>

Author contributions

GS is the main designer and developer of CentiScaPe. CL contributed to the design of CentiScaPe and performed the experiments. GT contributed to the definition of the directed centralities and performed the examples of usage in the [Supplementary files](#). MF developed the directed centralities. FS contributed to the development of the last version of CentiScaPe. FF is the main developer of the web version of CentiScaPe.

Competing interests

No competing interests were disclosed.

Grant information

This work was supported by: Italian Association for Cancer Research (AIRC, IG 8690) (C.L.); Fondazione Cariverona; Nanomedicine project University of Verona and Fondazione Cariverona (C.L.). Directed centralities have been developed thanks to the GSoC2013 program.

Supplementary materials

Supplementary files available from: <https://f1000researchdata.s3.amazonaws.com/supplementary/4477/dd4541d4-bdfe-461a-be92-f2e2cf9f2d57.zip>

References

1. Caldarelli G: **Scale-Free Networks: Complex Webs in Nature and Technology (Oxford Finance)**. Oxford University Press, USA. 2007.
[Publisher Full Text](#)
2. Jeong H, Tombor B, Albert R, *et al.*: **The large-scale organization of metabolic networks**. *Nature*. 2000; **407**(6804): 651–654.
[PubMed Abstract](#) | [Publisher Full Text](#)
3. Barabási AL, Oltvai ZN: **Network biology: understanding the cell's functional organization**. *Nat Rev Genet*. 2004; **5**(2): 101–113.
[PubMed Abstract](#) | [Publisher Full Text](#)
4. Strogatz SH: **Exploring complex networks**. *Nature*. 2001; **410**(6825): 268–276.
[PubMed Abstract](#) | [Publisher Full Text](#)
5. Koschützki D, Lehmann KA, Peeters L, *et al.*: **Centrality indices**. In Ulrik Brandes and Thomas Erlebach, editors, *Network Analysis: Methodological Foundations*, Springer. 2005; 16–61.
[Publisher Full Text](#)
6. Cline MS, Smoot M, Cerami E, *et al.*: **Integration of biological networks and gene expression data using Cytoscape**. *Nat Protoc*. 2007; **2**(10): 2366–2382.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
7. Saito R, Smoot ME, Ono K, *et al.*: **A travel guide to Cytoscape plugins**. *Nat Methods*. 2012; **9**(11): 1069–76.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
8. Scardoni G, Petterlini M, Laudanna C: **Analyzing biological network parameters with CentiScaPe**. *Bioinformatics*. 2009; **25**(21): 2857–2859.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
9. Currie HN, Vrana JA, Han AA, *et al.*: **An approach to investigate intracellular protein network responses**. *Chem Res Toxicol*. 2014; **27**(1): 17–26.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
10. Scardoni G, Tosadori G, Faizan M, *et al.*: **F1000Research-centiscapepublic-archive**.
[Zenodo](#). 2014.
[Data Source](#)
11. Newman ME: **Modularity and community structure in networks**. *Proc Natl Acad Sci U S A*. 2006; **103**(23): 8577–8582.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
12. Bhalla US, Iyengar R: **Emergent properties of networks of biological signaling pathways**. *Science*. 1999; **283**(5400): 381–7.
[PubMed Abstract](#) | [Publisher Full Text](#)
13. Scardoni G, Laudanna C: **Centralities based analysis of complex networks**. In Yagang Zhang, editor, *New Frontiers in Graph Theory*. InTech. 2012.
[Publisher Full Text](#)
14. Dijkstra EW: **A note on two problems in connexion with graphs**. *Numerische Mathematik*. 1959; **1**(1): 269–271.
[Publisher Full Text](#)
15. Scardoni G, Montresor A, Tosadori G, *et al.*: **Node interference and robustness: performing virtual knock-out experiments on biological networks: the case of leukocyte integrin activation network**. *PLoS One*. 2014; **9**(2): e88938.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
16. Gilbert D: **JFreeChart**.
[Reference Source](#)
17. Liu YY, Slotine JJ, Barabási AL: **Controllability of complex networks**. *Nature*. 2011; **473**(7346): 167–173.
[Publisher Full Text](#)
18. Chuang HY, Hofree M, Ideker T: **A decade of systems biology**. *Annu Rev Cell Dev Biol*. 2010; **26**(1): 721–744.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
19. Ivanciuc O, Balaban TS, Balaban AT: **Design of topological indices. Part 4. Reciprocal distance matrix, related local vertex invariants and topological indices**. *J Math Chem*. 1993; **12**(1): 309–318.
[Publisher Full Text](#)

Open Peer Review

Current Referee Status:



Version 2

Referee Report 23 July 2015

doi:[10.5256/f1000research.6690.r9389](https://doi.org/10.5256/f1000research.6690.r9389)



Ankush Sharma

Laboratory for Genomics, Transcriptomics and Proteomics, High Performance Computing and Networking Institute, National Research Council, Naples, Italy

The Authors have addressed the concerns raised by reviewers and made necessary changes in manuscript. The manuscript is fully approved for indexing.

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

Competing Interests: No competing interests were disclosed.

Referee Report 08 July 2015

doi:[10.5256/f1000research.6690.r5339](https://doi.org/10.5256/f1000research.6690.r5339)



Ferenc Jordan

Centre for Computational and Systems Biology, University of Trento, Trento, Italy

The Authors have made the necessary changes so it is now fully approved by me.

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

Competing Interests: No competing interests were disclosed.

Version 1

Referee Report 11 July 2014

doi:[10.5256/f1000research.4788.r5341](https://doi.org/10.5256/f1000research.4788.r5341)



Ankush Sharma

Laboratory for Genomics, Transcriptomics and Proteomics, High Performance Computing and Networking Institute, National Research Council, Naples, Italy

In this research article, the authors describe the CentiScaPe app for Cytoscape (and its web version FastCentiScaPe) for computing the several centrality parameters in networks, analyzing the elemental level importance of each node in the network, based on network topology. The addition of centralities parameters, especially Bridging Centrality and Eigenvector centrality, appears to be useful to improve metrics, to infer the informational flow across overlapping modules, and as a measure of the influence of a node in a network. The article is well written with an appropriate title and an abstract which provides sufficient details. The description of the purpose, implementation and usage of CentiScaPe app are informative and detailed for the audience.

Minor comments:

1. The example of calculating stress in the network of oncogenes in the manuscript should comply with experimental functional relevance, e.g. a decrease in bridging centrality is observed for key intermodular nodes upon heat shock response or oxidative stress, partially disintegrating overlapping modules into local communities which alters the centrality statistics of the nodes ^{1,2}. It would be good to see such implementation in the example showing applicability of CentiScaPe.
2. I suggest the authors have a small example of the analysis comparing connected and disconnected networks, to support computation of centralities for networks with a disconnected component.
3. It would be better to implement a user-friendly exploration of results and user registration feature instead of email authorization, in order to avoid delays in using the web version FastCentiScaPe.
4. Correction in the name of built-in tool i.e. NetworkAnalyzer , instead of "Analyze Network" tool.
5. Usage of appropriate words such as "more comprehensive" in place of "more complete" in introduction section.

References

1. Lehtinen S, Marsellach FX, Codlin S, Schmidt A, Clément-Ziza M, Beyer A, Bahler J, Oregno C, Pancaldi V: Stress induces remodelling of yeast interaction and co-expression networks. *Mol Biosyst.* 2013; **9** (7): 1697-1707 [PubMed Abstract](#) | [Publisher Full Text](#) | [Reference Source](#)
2. Mihalik A, Csermely P: Heat shock partially dissociates the overlapping modules of the yeast protein-protein interaction network: a systems level model of adaptation. *PLoS Comput Biol.* 2011; **7** (10). [PubMed Abstract](#) | [Free Full Text](#) | [Publisher Full Text](#) | [Reference Source](#)

I have read this submission. I 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.

Competing Interests: No competing interests were disclosed.

Referee Report 07 July 2014

doi:10.5256/f1000research.4788.r5358



Ferenc Jordan

Centre for Computational and Systems Biology, University of Trento, Trento, Italy

Analysing several centrality measures in parallel is important and there is a need for this, however, it would be even nicer to compare them (see <http://www.sciencedirect.com/science/article/pii/S0304380007001184>) and discuss better their biological relevance (i.e. what is Eigenvector centrality good for?). Central nodes can be of key regulator function but this is not a result, this is still a hypothesis. It should be mentioned briefly that either simulations or lab experiments should reinforce these findings. For infinite distances, using the reciprocal distance matrix has already been suggested as an elegant solution (see <http://link.springer.com/article/10.1007/BF01164642#page-1>). It should be clarified a bit more that a large weight can be considered as a short or as a long path, depending on its biological meaning. In Figure 1, the grid layout algorithm could be replaced by some better one, I think. Finally, a quick English check would be welcome.

I have read this submission. I 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.

Competing Interests: No competing interests were disclosed.
