The Importance of Health Informatics in Public Health during a Pandemic
J. Mantas et al. (Eds.)

doi:10.3233/SHTI200570

A 3D Surface Fitting Layout for Complex Networks Visualization

Laura BROASCĂ [a,1], Versavia-Maria ANCUȘA [a] and Horia CIOCÂRLIE [a]
[a] *Computer and Information Technology Department, Automation and Computers Faculty, "Politehnica" University of Timisoara,* Timisoara, Romania

Abstract. Large datasets visualization, prevalent in Bioinformatics, is a challenging task. Simple 2D layouts sometimes offer to many details to be instantly helpful, while filtering might remove critical information. Extension to 3D seems like the logical answer, however, in order to capture the complex relationships between data, the third dimension should reflect inherent properties. This paper proposes a novel method to generate 3D parameterized layouts, by using a function to correlate the Oz axis and biospecific features.

Keywords. Complex networks, visualization tools, 3D surface fitting

1. Introduction

Beyond a certain order of magnitude, human brain cannot fully extract useful information out of studied data [1] and data visualization techniques are used to circumvent that. Representation layouts arrange the data according to different criteria [2] in a relevant manner for large data sets. Combined with complex networks usage as a common means of arranging data [3] this approach seems to bring the Bioinformatics data together while emphasizing connections between nodes (e.g.: genes, proteins, compounds), and cluster formation [4] [5] (e.g: related genes, chemical compounds, drug families, etc).

Many network visualization tools have been developed lately: Gephi, Python's NetworkX, R's iGraph, D3.js's D3-3D plugin. Most offer different approaches to representing data from basic metrics (node degree, betweenness/closeness centrality, etc.) to force directed visualization algorithms such as Force Atlas 2 (FA2) [6], Fruchterman-Reingold [7], OpenOrd, etc.

Visualizing 2D genetic / biologic data as networks presents no obvious vertical hierarchy between vertices (genes / compounds) other than their visual size, proportional to a certain graph-specific metric. 3D networks can take advantage of spatially scattering graph elements, highlighting the distance and correlation between them [8]. However, specifically in large biological networks, there are no spatial landmarks to help the researcher interpret the 3D position of a certain node [9].

The aim of this paper is to offer more meaning to 3D network layouts in a biomedical context, i.e. to generate with the help of Octave a new type of layout which uses the *Oz* axis as a true 3rd dimension, giving the viewer a depth perception into data and revealing meaningful or even missing correlations, among genes and proteins.

[1] Corresponding Author, Laura Broasca Automation and Computers Faculty, "Politehnica" University of Timisoara, Timisoara, Romania; E-mail: laura.broasca@cs.upt.ro.

2. Methods

We have defined a 3D parabolic surface and dispersed graph nodes over it, maintaining predefined links between them, and fitting the surface equation. A characteristic emphasized is the nodes hierarchy, by positioning them into concentric circles around the centre vertical axis of the parabola. Once we associate one network metric (degree, centrality, etc.) to the *Oz* axis, then the importance of vertices is directly proportional to their *z* value. In addition, each concentric circle composing the parabola corresponds to a hierarchical level and all vertices considered equals in terms of the chosen metric lie on the same *z*-level and circle.

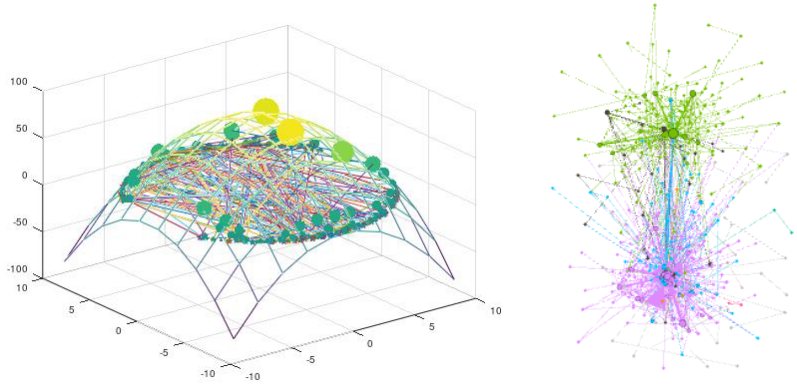
The data set used throughout this experiment consists of a medium-sized set of DNA genes and proteins with role in transcription regulation for E.coli [10]. Some basic network characteristics are: nodes (N): 424, edges (K): 519.

The selected tool for network layout rendering is GNU Octave [11]. Data characteristics of the network can be chosen as dominant (degree) to establish the value of each vertex on the *Oz* axis. As such, we have defined the equation of the 3D space (a parabola) which will then be used to calculate the position of each vertex in the graph:

$$z = a - x^2 - y^2 \tag{1}$$

Assuming the *z* values correspond to the *degree* metric, *Fig. 1. a* shows the distribution of nodes alongside the 3D surface in comparison with a Gephi OpenOrd layout (*Fig. 1. b*). Substituting *z* with degree values, we can calculate the *x* and *y* values such that they satisfy a circle equation $x^2 + y^2 = a - z$ of radius $r^2 = a - z$ and *O(0,0)* is the centre.

3. Results



a) b)

Figure 1. Graph layout rendered in a) Octave with 3D surface fitting b) Gephi

Although the Gephi (semi) 3D layout *Fig.1.b* looks more structured and clustered, the user does not have a spatial differentiation between nodes and cannot for example quantify their importance (in the process of transcription regulation) in comparison with each other aside from node size and colour. On the other hand, in *Fig. 1. a)* there is a true 3D distribution of nodes.

Considering *Fig. 2* it is clear that this method also needs improvement in terms of node clustering. However, there is an important visual indicator showing the rank of each node which is equivalent to the importance of each gene with regards to the whole genetic mechanism. The larger the bubble and the closer it is to the centre, the more the user gets the idea of how essential that vertex is according to the chosen discriminating metric.

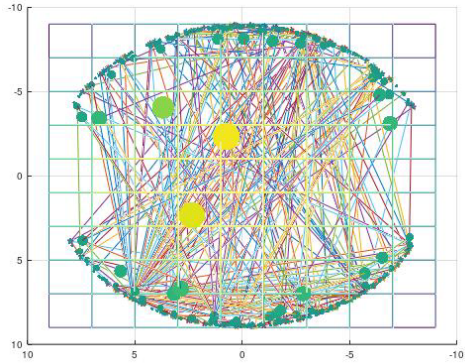


Figure 2. Graph layout rendered in Octave with 3D surface fitting – 2D view from above

4. Discussion

In order to determine how this new approach to data visualization differentiates from other existing ones, we have graded it according to certain subjective criteria. For this purpose we have chosen a basic comparison with Gephi (OpenOrd algorithm) considered to be the current standard in complex networks visualization. [9].

Table 1. Evaluation table

Criteria	Data set 3D implementation	
	Gephi	Octave
Speed	3	2
Visual appeal	4.5	3.5
Relevance/Utility	3.5	4
User friendliness	5	3.5
Customizability	3	*4.5*
ML capability	0	1
No programming skills required	1	0
Interactivity 0 (No)/ 1(Yes)	1	1*
Totals	21	19.5

All criteria have been evaluated according to empirical tests. In terms of speed, medium sized networks behave similarly. With Gephi (FA2, OpenOrd algorithms), average processing time for such a bio network (N < 1000) ranges between $t \in [5,10]$ seconds. With the new Octave 3D implementation however, since for this study we have not used optimized networks frameworks, rendering times are admittedly slower, $t \in [7,11]$ seconds. Visual appeal is subjective, yet based on the direct feedback of a small group of subjects (10 data specialists). Simplistic graphics in Octave do not outrank Gephi's UI and this is directly proportional to user friendliness. Relevance/Utility is the category where the proposed algorithm aims to bring improvements and although not surpassing its competitor by much, it seems that the 3D innovation factor gives it an edge.

Customizability and ML capability are two strong points in Octave's arsenal which is specifically designed for mathematical and programmatic purposes, with the cost of requiring programming skills. Both tools are interactive, yet Gephi excels in this area.

5. Conclusion

In this paper we presented a new approach to 3D biological graphs visualization in a parameterized manner. This comes as an alternative to regular 3D visualization tools which fail to spatially structure the elements of such networks. 3D surface fitting graphs maintain classical attributes such as: node size and color, spatial distance between nodes, grouping nodes according to genetic similarity coefficients, biological role, etc. On larger networks, if the resulting figure risks becoming too crowded, node distribution can be adjusted by defining a composed metric so as to include more genetic specificities (upward/downward regulation role of a gene/protein, activation time, the probability of a gene playing a part in multiple metabolic processes, etc).

The most important advantage of the method proposed here is that it offers a parameterized view into 3D network visualization which helps give more insight into node position and hierarchy according to chosen criteria. Currently, the proposed technique is still under development and can be further improved to better accommodate the representation and customization of more diverse topologies.

References

[1] Broasca L, Ancusa V, Ciocarlie H. Bioinformatics Visualisation Tools: An Unbalanced Picture. Studies in Health Technology and Informatics 2016: 760-764
[2] Bastian M, Heymann S, Jacomy M. Gephi: An Open Source Software for Exploring and Manipulating Networks. Proceedings of the Third International ICWSM Conference 2009.
[3] Attara N, Aliakbary S. Classification of complex networks based on similarity of topological network features. Chaos: An Interdisciplinary Journal of Nonlinear Science 2017.
[4] Batool K, Muaz NA. Modeling the internet of things: a hybrid modeling approach using complex networks and agent-based models. Complex Adaptive Systems Modeling 2017;5(4).
[5] McGee F., Ghoniem M, Melancom G, Otjacques B, Pinaud B. The State of the Art in Multilayer Network Visualization. Computer Graphics Forum 2019;38(6): 125-149.
[6] Jacomy M, Venturini T, Heymann S, Bastian M. ForceAtlas2, a Continuous Graph Layout Algorithm for Handy Network Visualization Designed for the Gephi Software. PLOS ONE 2014.
[7] Fruchterman TMJ, Reingold EM. Graph Drawing by Force-directed Placement. Software - Practice and Experience; 1991.
[8] Zhou G, Xia J. OmicsNet: a web-based tool for creation and visual analysis of biological networks in 3D space. Nucleic Acids Research 2018;46(W1): 514-522.
[9] Broasca L., Ancusa-Maria V., Ciocarlie H. A Qualitative Analysis on Force Directed Network Visualization Tools in the Context of Large Complex Networks. Proceedings of the 23rd International Conference on System Theory, Control and Computing (ICSTCC); 2019.
[10] Latora V, Nicosia V, Russo G. Complex Networks: Principles, Methods and Applications, Cambridge, United Kingdom: Cambridge University Press, 2017.
[11] Eaton JW. GNU Octave Scientific Programming Language 03 2020. [Online]. Available: https://www.gnu.org/software/octave/.