Network analysis using cytoscape

Srishti Verma
Amity University
Lakshay Mangla (manglalakshay4902@gmail.com)
Amity University

Research Article

Keywords:

Posted Date: March 7th, 2023

DOI: https://doi.org/10.21203/rs.3.rs-2487773/v1

Additional Declarations: No competing interests reported.

Network Analysis Using Cytoscape

Srishti Verma[1], Lakshay Mangla[2*]
[*1]Btech Bioinformatics with minor degree in Data science and Analytics, Amity Institute of Biotechnology, Amity University, Noida,
E-mail: srishtiverma793@gmail.com
[2]Btech Biotechnology, Amity Institute of Biotechnology, Amity University, Noida,
E-mail: manglalakshay4902@gmail.com

ABSTRACT:
In the following research, we have made cytoscape an easy and compact tool for network analysis. We study the relation between the data sets. Creating edge list and node list for determining relations between the columns and rows hence forming a csv table for the same. Here we have used Cytoscape and Open refine software. Open refine to create the csv file and cytoscape to study the networks.

Introduction

Cytoscape is an open source software project that intends to integrate molecular states, high-throughput expression data, and biomolecular interaction networks into a unified conceptual framework. However, Cytoscape functions best when combined with sizable datasets of the genetic, protein-protein, and protein-DNA interactions that are becoming more available to people and model animals. Cytoscape can be used with any instrument that analyses molecular components and interactions. The Cytoscape Core programme allows for simple formatting and querying, as well as the ability to visually combine the community with expression profiles, phenotypes, and different chemical states and link it to databases with helpful annotations. [1]The Core may be readily expanded with new computational analyses and capabilities thanks to a genuine plug-in design. The numerous Cytoscape plug-ins are reviewed in several case studies. These include a look at protein complexes involved in cell recovery from DNA damage, an inference of a community with mixed physical and practical interaction for the Halo bacterium, and a connection to certain stochastic/kinetic gene regulatory models.

Cytoscape has shown its attractiveness as a platform for community biologic review with about 17,600 international downloads per month, 5,000 start-ups per day, and more than a thousand direct citations each year. [2]Researchers may interactively explore complicated *omics datasets using the analysis and visualisation capabilities provided by Cytoscape, as well as the large and active community of app contributors.

However, interactive use has proven to be insufficient for accurately scaling to big quantities, sharing complex studies, or manufacturing evaluation. Additionally, even while Cytoscape apps offer community biology capabilities that are very effective and useful, the specialised programming knowledge and sometimes lengthy development timeframes they require may make them uneconomical for transferring complicated and evolving workflows.[3] Last but not least, Cytoscape is not positioned to charge for growing workflows that incorporate one or more external statistics collecting and evaluation tools because it is an interactive tool (e.g., Galaxy, Taverna, and libraries supplied in repositories along with PyPI and Bioconductor). Aiming at a specific approach or disease process of interest, models are often constructed by describing recent literature as a system of differential and/or stochastic equations (Gilman and Arkin 2002). However, pathway-

specific patterns are currently being supplemented with global data obtained for a full cellular or organism through the use of supplementary approaches. It is now possible to degree route form systematically thanks to the use of high-throughput displays for protein-protein, protein-DNA, and genetic interactions (Ito et al. 2001; von Mering et al. 2002). (Tong et al. 2001). In order to supplement those facts, a second set of high-throughput approaches is accessible to depict the molecular and mobile states produced by route interactions under unique experimental settings.

Gygi et al. (1999), Zhou et al. (2001), and Griffin et al. (2001) used mass spectrometry, NMR, and other more complex techniques to measure changes in protein abundance, protein phosphorylation state, and metabolite concentrations.[4] De Risi et al. (2007) used DNA microarrays to measure overall changes in gene expression. High-throughput data relating molecular interactions and states are well matched because both data types are global (presenting information for all additives or interactions in an organism), high-level (outlining relationships among pathway additives without specialised information on response rates, binding constants, or diffusion coefficients), and coarse-grained (yielding qualitative results).

Researchers have created a number of software tools to organise and analyse the ensuing vast amounts of data as a result of the explosion in experimental technology for measuring molecular interactions and states.[5]

The information can be prepared and displayed as a two-dimensional community for molecular interactions using common-cause graph viewers like Pajek (Batagelj and Mrvar 1998), Graphlet and daVinci; specialised tools like Osprey. Similar technologies, like GeneCluster, exist for grouping, classifying, and visualising gene expression patterns and different biochemical states1999, Tree-View and GeneSpring.

Software that can integrate molecular interactions and country measurements in a well-known spatial framework and then connect those data with a wide number of version parameters and other organic features is still urgently needed.[6] Additionally, the interaction community may require a flexible and open device to handle commonly used and extensible calculations (Karp 2001). Through these calculations, low-stage physico-chemical models can eventually interface with high-stage interaction data and enhance their pressure properties. To satisfy these requirements, we enhanced Cytoscape, a common-cause modelling framework for integrating bimolecular interaction networks and states.

We begin by providing a quick summary of C1ytoscape's fundamental features for representing and implementing bimolecular community models. We next go over three case studies of active research initiatives where the Cytoscape platform is modified to impose new algorithms and collaborative calculations or utilised to tackle real-world biological challenges.

Objective

Using Cytoscape , you will learn how to design, visualise, and analyse networks in this practical. [7]You will also learn how to export the findings of these analyses.

Work done

Cytoscape interface

From the Windows applications menu, launch Cytoscape. [8]Cytoscape will present you with a welcome screen after the software has loaded. Choose "New/Empty Network"

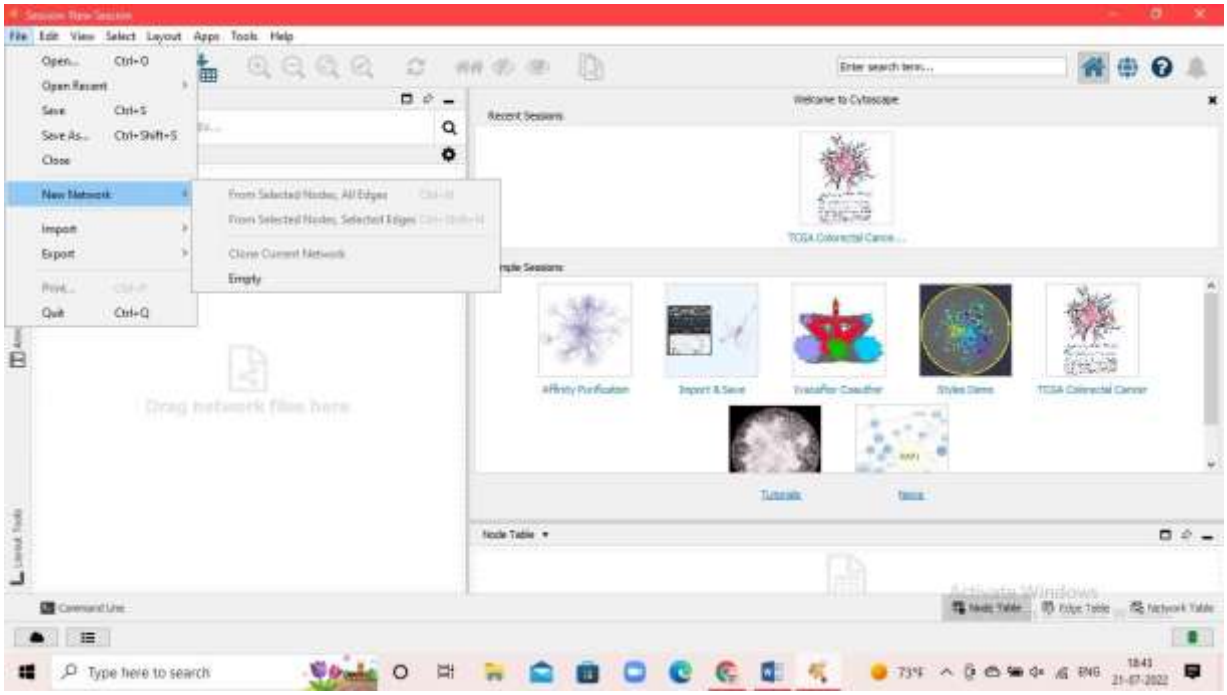

Fig.1 Starting a fresh session in cytoscape

Always start out by saving a fresh session. File > Save As can be found on the top menu. Select the location where you want to store the new session report from the dialogue field that opens, then click the Save button.(Fig 1)

Your Cytoscape window's top will now display the session call. Cytoscape uses its own format for session storage;[9] these files typically end in.cys.

The workspace, table panel, and control panel are the three primary sections of the Cytoscape user interface. The Control Panel, which includes three unique tabs called Network, VizMapper, and Filters, is located on the left side of the screen.

The Network tab is now empty but will eventually contain a list of all the networks you have imported or opened. You can utilise visible patterns for your network and change how each piece is displayed by using the VizMapper tab (see segment 7 below). You can build and modify filters on the Filters tab to make your own selections from the community (see segment 4 below).

Second, you can find the Table Panel at the bottom of the screen. [10]Here, you can browse using selected nodes, edges, networks, and their properties. Now, the desk panel must be left blank. It's important to pay attention to the five buttons on the Table panel's upper left.

The first one allows you to toggle the table mode, the second one lets you turn on and off attributes, the third one lets you select ALL attributes, the fourth lets you turn them off, and the fifth lets you add new attributes. The final one enables attribute deletion.

Third, your workspace is the dark area between the Control Panel and the Table Panel. When you create network windows, this is where they will appear. As you'll see in section 5, you can adjust node location and make choices within network home windows. ***Network input files***

For the input files, we must create an edge list. The edge list essentially informs us of the relationships between each row and the various table columns. We have two options: import an existing file from Github, or build a new one. In order to transform the table into an edge list for this practical, I had to reconstruct it. You may only need one element to develop networks: a list of all edges or arcs connecting all nodes (a part is a line without a route that connects points, an arc on the alternative hand does have a route, e.g. from node A to node B).

As we can see below, these kinds of lists can be pretty basic. We might also need to import records related to certain edges or arcs in addition to this listing.[11] This data is represented by the "attributes" of the arcs or edges. However, as

ofright moment, we still don't have any records on the nodes. We may need a list of every node and its properties in order to import records on nodes. We are able to have a primary examination of this statistics' ability to build networks in this segment. Numerous unique report kinds can be imported by Cytoscape. Spreadsheets made in Microsoft Excel can be used in this practical.

When we received the edge list, it should only have two columns and each row should indicate the relationship between them, with the first column serving as a connecting link to the second. The edge list is then converted to a csv file using the Openrefine programme.

OpenRefine, previously Google Refine, is a useful tool for managing messy data. It may be cleaned, transformed from one layout into another, and expanded with external data and web offers.[12] Until YOU need to share or contribute, OpenRefine continuously keeps your data private for your own computer. Unless you specifically request it, your personal information never leaves your laptop. OpenRefine is available in more than 15 languages. It operates by employing running a small server on your computer, and you use your internet browser to interface with it. A component of Code for Science & Society is OpenRefine.

How to get node lists and edge lists?

There are two steps in the statistics transformation. Create a nodes sheet at the initial stage, where each node is given a completely distinct Id. Create an edges sheet in the second stage where all node family members are expressed as relationships between Ids. [13][14]

Before you begin, make sure that every person on your "Character Interaction" sheet has a totally distinct call, and that this call is consistently used in all of your interactions with that person.

Create the nodes sheet 1 in stage 1. In Excel, make a new worksheet and name it "Nodes" (or some thing of your choice). [15]

2. In the Nodes sheet you just made, copy and paste every character from the "Character Interaction" sheet into a single column. This gives you a list of every character in the brief narrative.
3. In the Nodes sheet, click any non-empty cell. Next, select Remove Duplicates to get rid of duplicate values in the "Name" column.
4. This provides a list of the brief story's distinctive characters.
5. Give each character a distinct Id and place it in the Id column.
6. Give your Nodes sheet a new name and save it as a CSV report.

Create the edge sheet in stage two. 1. In Excel, make a new worksheet and name it "Edges" (or something of your choice).

Once the edges and nodes have been matched, you will have a single list that you may refer to as the edge list(Fig 2). To create a csv file, we will now insert the edge list into openrefine.

Introduction to OpenRefine

1.Investigate Data

You may easily study big data sets with the aid of OpenRefine.

2.Cleaned and Reorganized Data

3.Consolidate and compare data

Your dataset can be linked to and expanded with different webservices using OpenRefine. Some options also allow OpenRefine to add your deleted records to a main database, such as Wikidata. The wiki contains a growing list of extensions and plugins.

We now use open refine to place the edge list that has been converted into a csv file. [16]

When we have the csv file, we can proceed to Cryptospace to begin the data analysis.

Chapter 4: Results

Edge list:

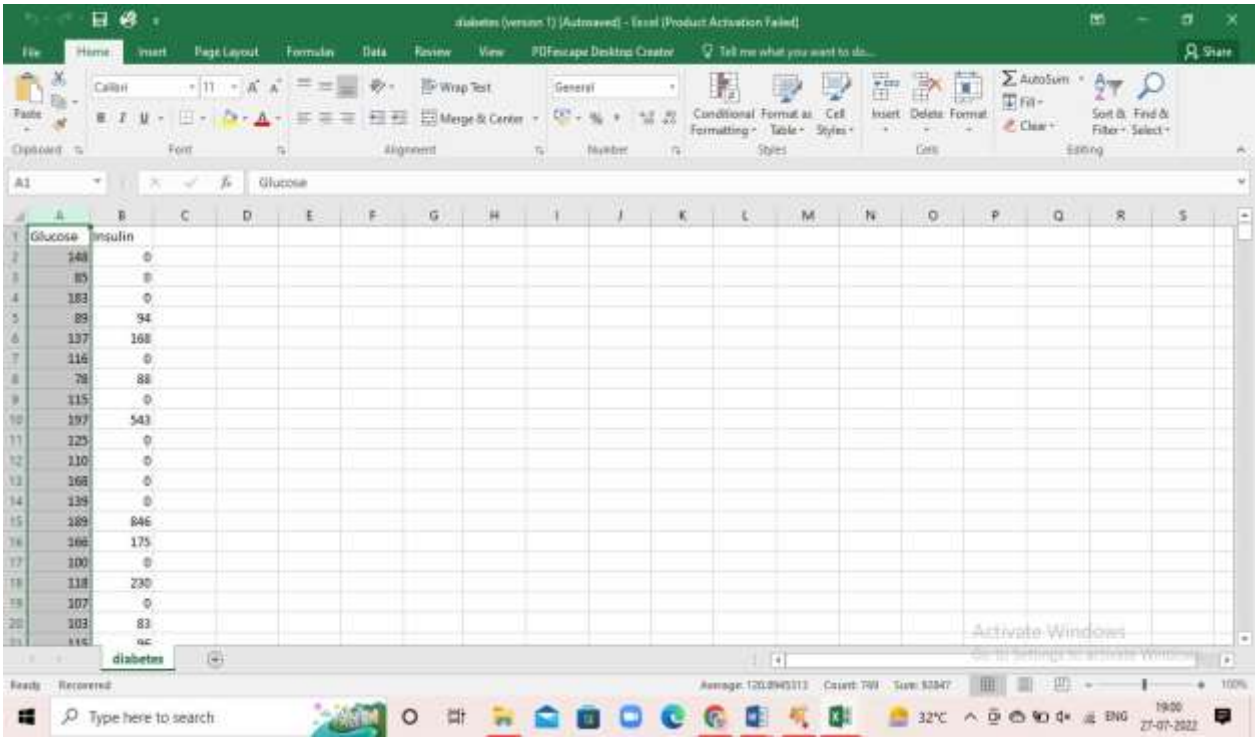

Fig. 2 Edge list is a straightforward way to represent a graph, with each member in the list denoting a connection between two nodes. It consists of a collection of edge-connected pairs of vertices.

Placing the edge list into cytospace: Source node:

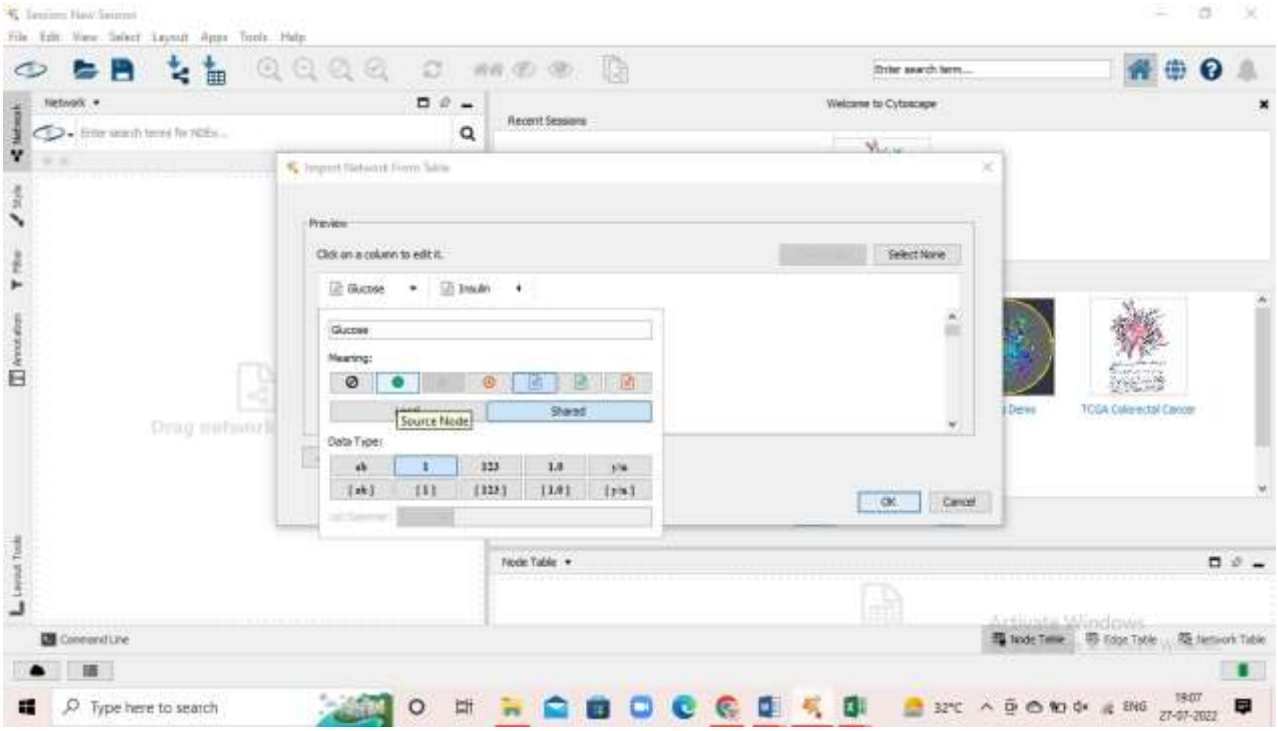

Fig. 3 In a directed graph, a node with no incoming edges is known as a source node. A source node is a node from which no edges arise, to put it another way.

Target node:

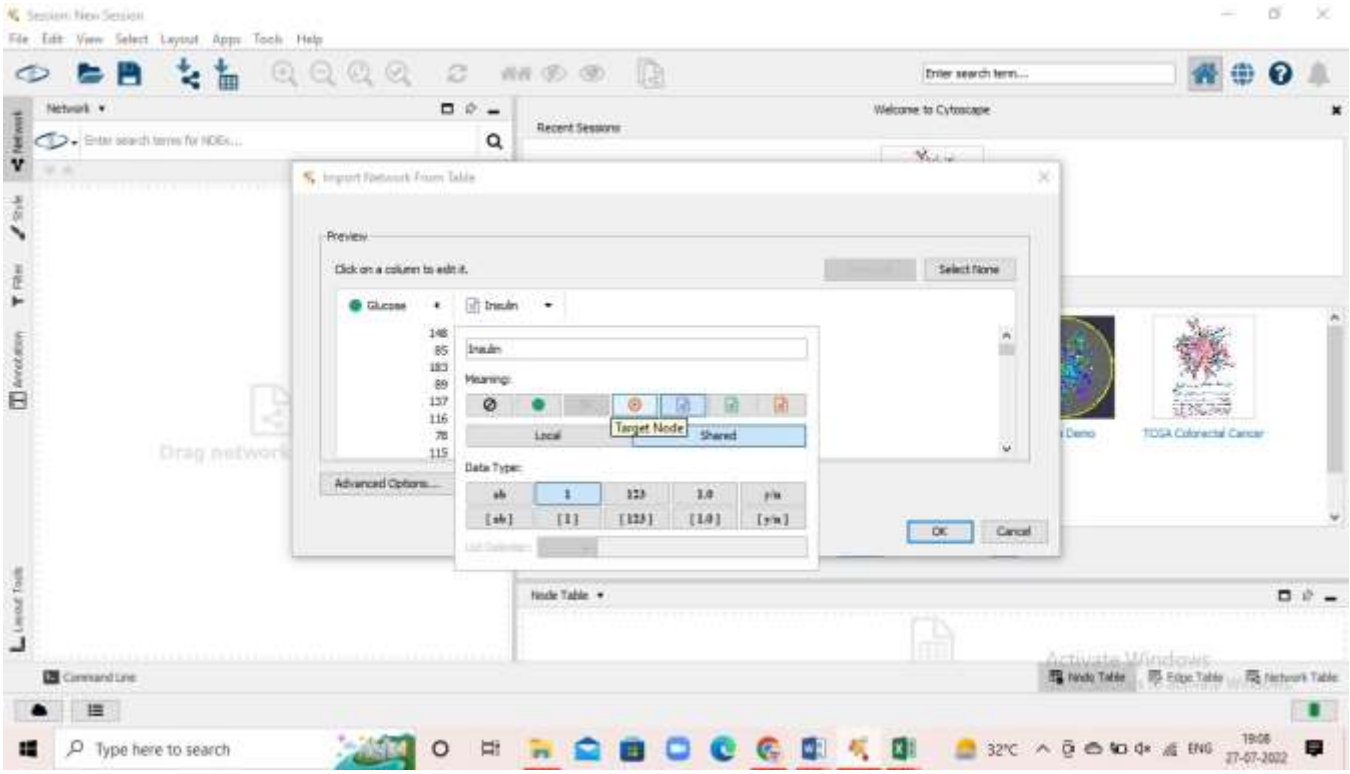

Fig. 4 A target node in graph theory is a node in a directed graph without any outbound edges. A target node is one on which no edges terminate, to put it another way.

Network obtained:

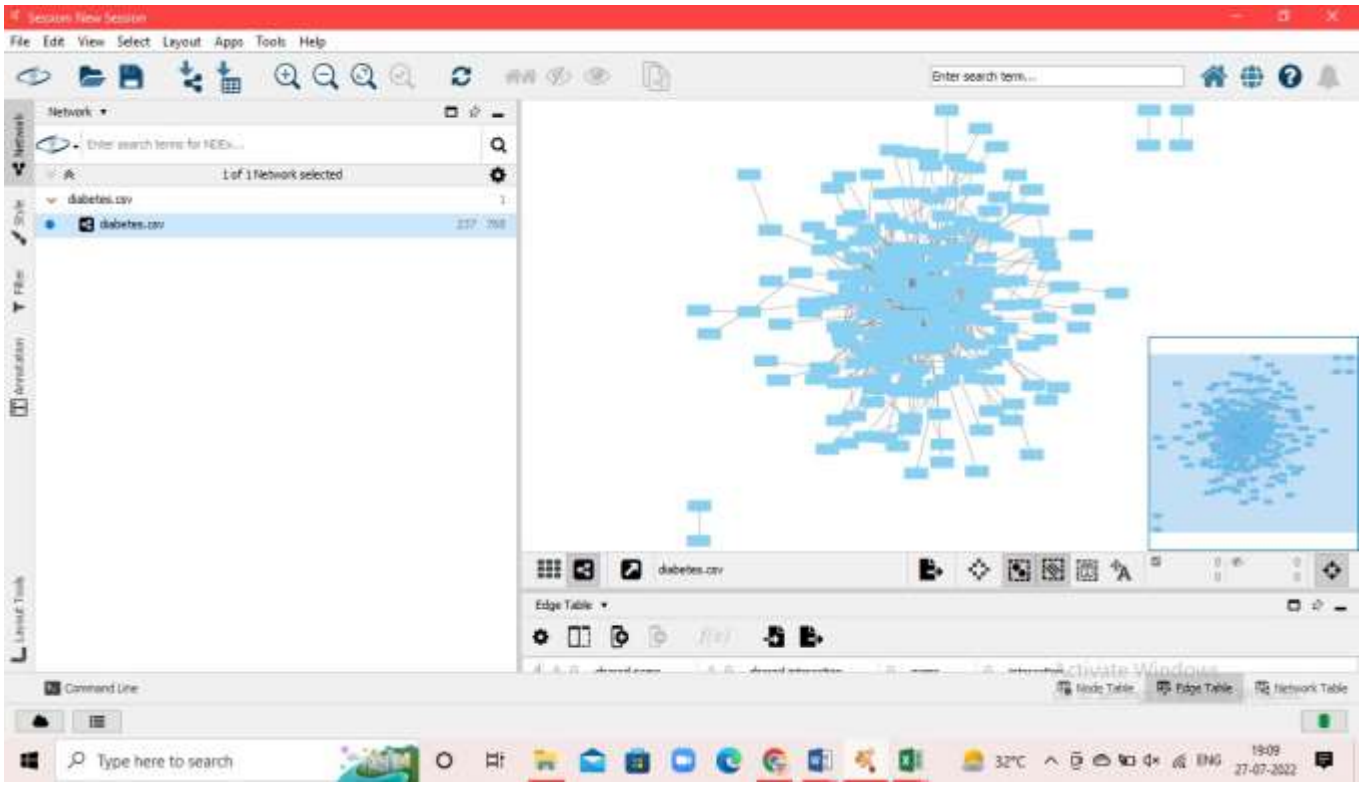

Fig. 5 The network obtained by putting the edge list into cytoscape

Navigation and layout:

Networking Foundations

A zoomable user interface is used by Cytoscape to examine and navigate networks.[17] Zooming and panning are the two navigational methods used by ZUIs. Depending on how much or how little a user wishes to see, zooming magnifies a view more or less. Users can pan a screen to change the area of a display that is in focus. *Zoom:*

Four methods of zooming are available in Cytoscape: the scroll wheel, keyboard shortcuts, menu items, and toolbar buttons.

To zoom in and out of the interaction network visible in the current network view, use the zooming buttons on the toolbar. Icons for zoom are described below:

There are two methods for network panning:

Holding down the left mouse button while dragging the mouse allows you to pan the network view.

Network Overview Box: By left-clicking and dragging the blue box in the overview panel in the lower portion of the display, you can also pan the view. [18]

To pick a node, edge, or annotation, click the left mouse button on it.(Fig 3)

To include a node, edge, or annotation in the selection, left-click it while holding down the Shift or Ctrl key (Command on a Mac). The same action on a selected element deactivates it.

To pick collections of nodes, edges, or annotations, hold down the left mouse button on the canvas

background and drag the mouse while holding down the Shift or Ctrl key (Command on a Mac).

It's important to keep in mind that the selection action (mouse click or drag-selection) only

functions if the Selection Mode is enabled for that element type (e.g., nodes, edges, or annotations). Toggle the matching button at to activate or disable the selection of an element type(Fig 4)

.

Manual Design

By clicking and dragging a node, you can manually arrange a network in the simplest way possible. The chosen nodes are all shifted together.

on them and dragging them to a new location. The selected nodes will move one pixel in the desired direction by clicking one of the arrow keys (←, ↑, →, ↓) while selecting one or more nodes with the mouse. [19]The selected nodes will move 15 pixels in the direction selected if an arrow key is pushed while holding down the Shift key.

The arrow keys on the keyboard can now be used to move nodes in Cytoscape in addition to clicking

Layout Node Tools

The View Show Tool Panel menu option or Layout Node Layout Tools are two ways to access the Tool Panel. [20]

Numerous Node Plan Tools are included, which can be used to automate or fine-tune a layout. [22]

Scale

To alter the length of the edges, move the Scale slider. Not the node sizes but their positions will be scaled.[21] Styles allow for the modification of

node size. Selected (yellow) nodes are scaled to 50% of the default value in the photos below.

Rotate

The network's orientation can be changed for the entire network or just a specific section using the

Rotate function. [23]The networks in the photos below have some nodes rotated by 90 degrees.

Layout algorithms that are automatic

With the help of the Layout menu, you may align and rotate groups of nodes, change the size of the network, and visually organise the network using one of several algorithms.[24][25] There are three different sources for Cytoscape layouts, and the Layout menu reflects this.

All Cytoscape layouts have a Settings... panel where you may modify the algorithm's parameters, and you can choose to act solely on the selected nodes.[26] Before performing the layout, the majority of Cytoscape layouts also partition the graph(Fig 5). Additionally, many of these layouts offer the choice of taking into account node or edge columns. The following are a few layout algorithms:

Grid Layout:

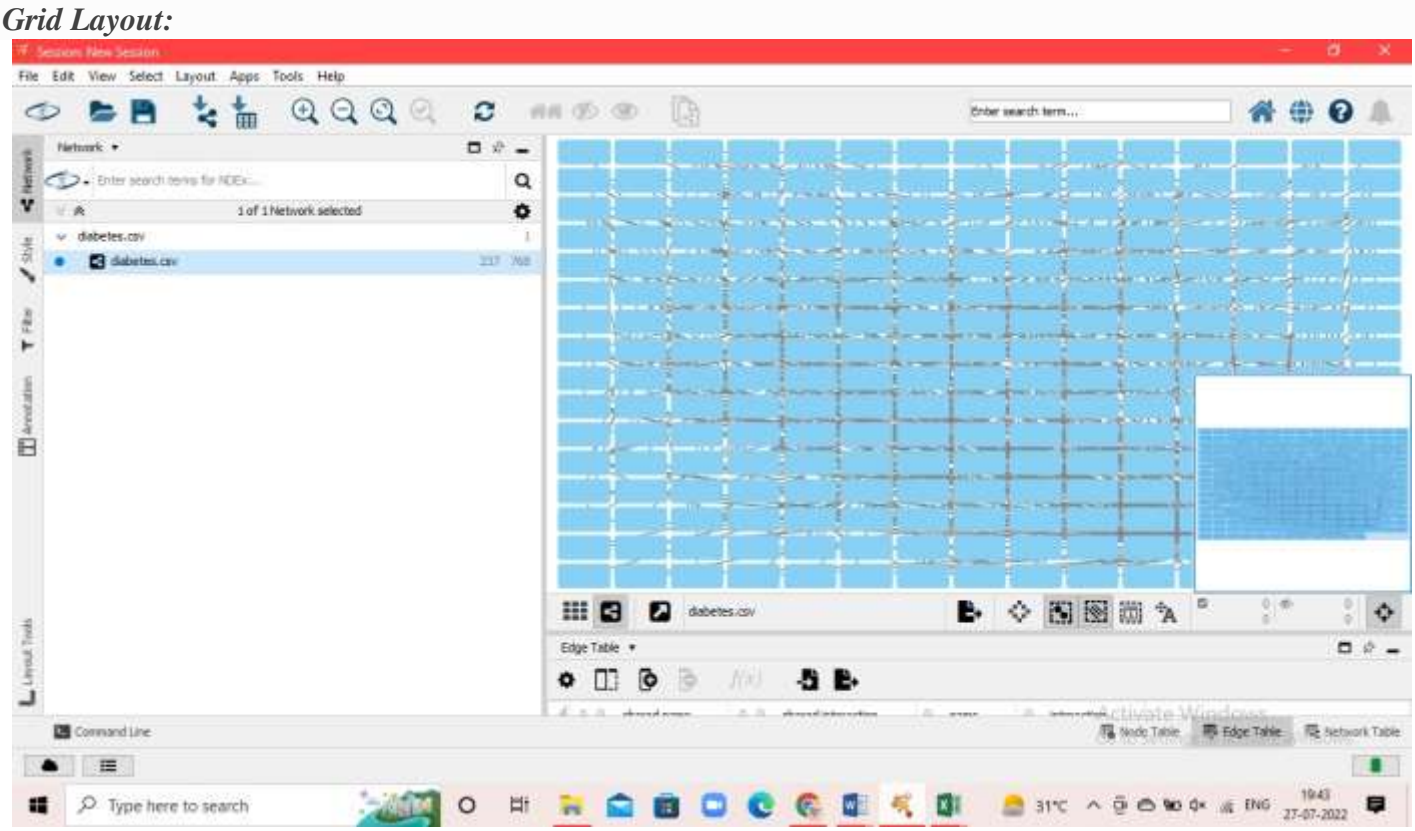

Fig. 6 The grid layout arranges the nodes into a grid with adequate spacing. In the order they are provided to the layout, the nodes are arranged from left to right and top to bottom.

The grid layout (Fig 6) is a straightforward design that places each node in a square grid.[27] The Cytoscape core always includes this layout, which is the default. You can access it by choosing Layout Grid Layout.[28] Above is a screenshot example.

***Circular layout**:*

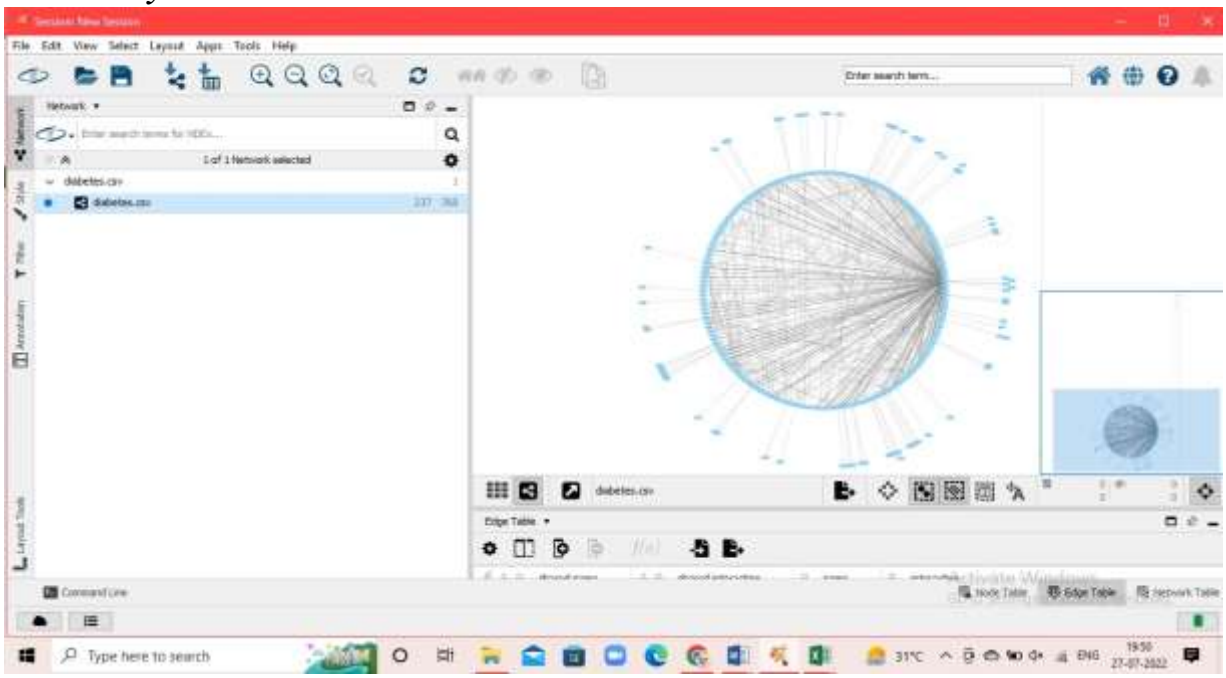

Fig. 7 The nodes are arranged into a circle in the circle pattern. By default, the nodes are arranged in the layout's clockwise direction, starting at the 12 o'clock position.

This technique generates network configurations that highlight group and tree topologies.[29][31] It divides the network into sections by examining its connectivity pattern, and then positions the sections as distinct circles.[30] The actual circles have been organised in a radial tree arrangement. You can access this algorithm by choosing Circular Layout under Layout (Fig 7).

Hierarchical Layout

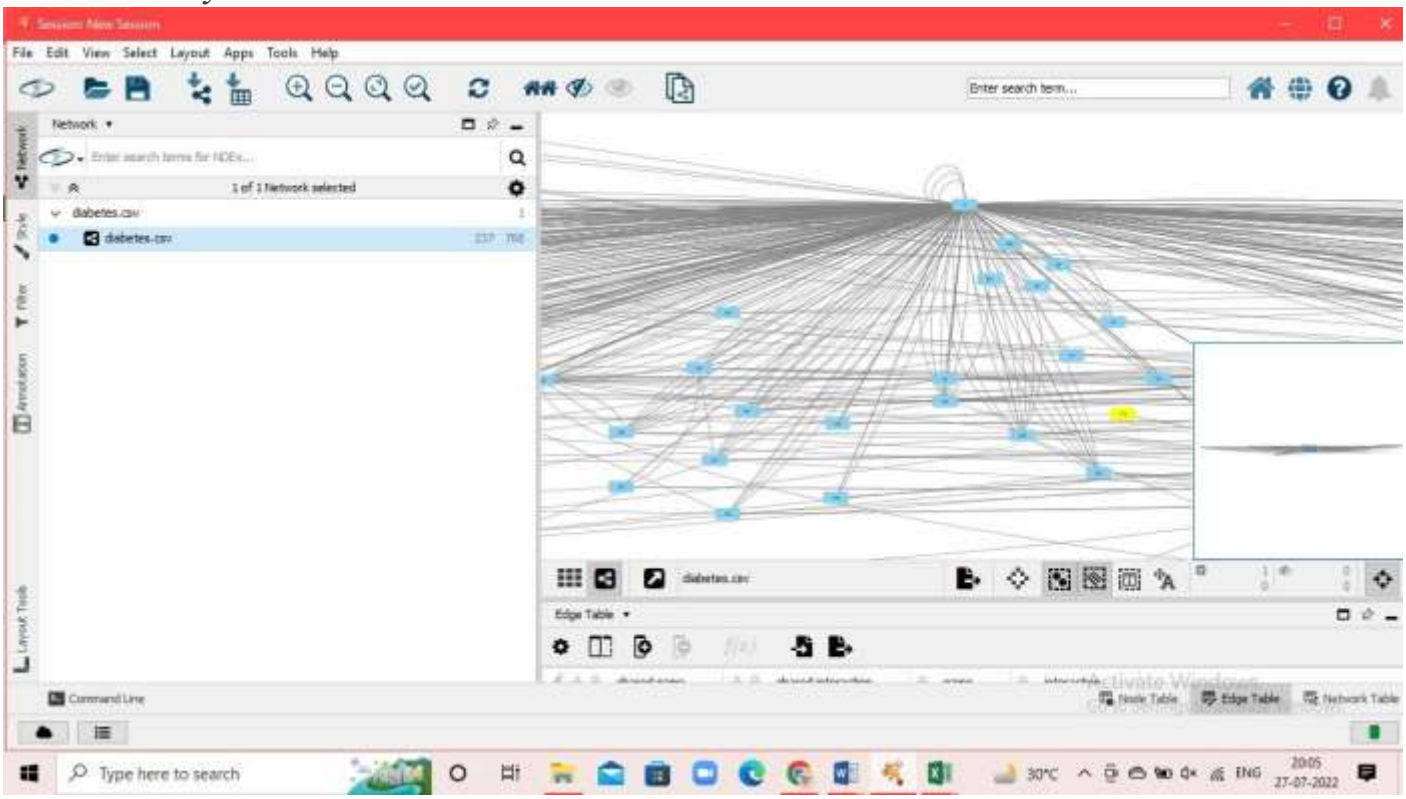

Fig. 8 When the network is naturally tree-structured, hierarchical layout functions best. It also functions reasonably well when the network is largely hierarchical.

The main direction or flow inside a network can be accurately represented by the hierarchical layout algorithm. Nodes are organised into layers that are hierarchically ordered, with the nodes within each layer being arranged in a way that minimises the amount of edge crossings. Choose Layout → Hierarchical Layout to access this algorithm (Fig 8).

Analyze:

For both directed and undirected networks, the analyzer calculates a wide range of topological parameters, including:

- Number of connected components, edges, and nodes.[31]
- The characteristic path length, network radius, diameter, and clustering coefficient.
- Charts for betweenness, proximity, and topological coefficients.

.

Depending on whether the network is directed or undirected, the analyzer will run different statistics. [32][33]Based on the specification of a target arrow style, the programme will attempt to

Fig. 9 Result panel with the analysed results

- Degree distributions, neighbourhood connectivity, average clustering coefficients, shortest path lengths, number of common neighbours, and stress centrality are all considered.

Network Analysis

Analyze Network

To run Analyzer, select **Tools → Analyze Network**

determine the type of network, however as this is not

Results will show up on the Results Panel whenever they are ready (Fig 9).

Subset of Nodes Analyzed

Earlier iterations of this programme let users choose whether to analyse all nodes or just a particular subset of them. The application no longer directly supports this. Instead, to establish

the desired subnetwork for analysis, use the command File New Network From Selected Nodes, All Edges in the Command Panel.

Plot Elements

The Node Table gains a number of new columns after the Analyzer has been executed (and an EdgeBetweenness column is added to the Edge

Table). Right-click on the column header and choose Display Histogram... to plot any of these additional columns as a single parameter

distribution or Plot Scatter... to plot the data as a bivariate plot. [33]It is possible to choose a part of the data within either of these charts and then choose the nodes (edges) in the main graph window that correspond to that region on the chart.

Chapter 5: Conclusion

Hence in this research we started from the basics of cytospace and learned how to analyse network using cytospace.

Declarations

Ethics approval and consent to participate:

Not applicable for this work as no ethical clearance was needed.

Consent for Publication:

Not applicable

Availability of Data:

https://github.com/plotly/datasets/blob/3aa08e58607d1f36159efc4cca9d0d073bbf57bb/diabetes.csv

https://drive.google.com/file/d/1hOBL8x9gmDwcAQh1d85jAMgONpYYq3OW/view?usp=share_link

Competing interest:

Not applicable

Funding:

Not applicable

Author's contribution:

S.V. and L.M. conceived the presented idea. S.V. developed the theory and performed the computations, verifying the analytical methods. L.M. helped in formatting the final manuscript. All authors discussed the results and contributed to the final manuscript.

Author's Information:

Ms. Srishti Verma
Btech Bioinformatics with Minor degree in Data Science and analytics.
Amity Institute of Biotechnology
Amity University, Uttar Pradesh
Noida-201301
Mr. Lakshay Mangla
Btech Biotechnology
Amity Institute of Biotechnology
Amity University, Uttar Pradesh
Noida-201301

Acknowledgement
It is high privilege for me to express my deep sense of gratitude to those entire faculty Members who helped me in the completion of the NETWORK ANALYSIS USING CYTOSCAPE under supervision of Dr. Priyanka Narad.
My special thanks to all other faculty members, batchmates & seniors of Amity institute of biotechnology, Amity University Uttar Pradesh for helping me in the completion of project work and Its report submission.

REFERENCES

1. Bader, G.D., Donaldson, I., Wolting, C., Ouellette, B.F., Pawson, T., and Hogue, C.W. 2001. BIND—The biomolecular interaction network database. *Nucleic Acids Res.* **29:** 242-245.

2. Baliga, N.S., Pan, M., Goo, Y.A., Yi, E.C., Goodlett, D.R., Dimitrov, K., Shannon, P., Aebersold, R., Ng, W.V., and Hood, L. 2002. Coordinate regulation of energy transduction modules in *Halobacterium sp*. analyzed by a global systems approach. *Proc. Natl. Acad. Sci.* **99:** 14913-14918.

3. Batagelj, V. and Mrvar, A. 1998. Pajek—Program for large network analysis. *Connections* **21:** 47-57.

4. Begley, T.J., Rosenbach, A.S., Ideker, T., and Samson, L.D. 2002. Damage recovery pathways in *Saccharomyces cerevisiae* revealed by genomic phenotyping and interactome mapping. *Mol. Cancer Res.* **1:** 103-112.

5. DeRisi, J.L., Iyer, V.R., and Brown, P.O. 1997. Exploring the metabolic and genetic control of gene expression on a genomic scale. *Science* **278:** 680-686.

6. Eades, P. 1984. A heuristic for graph drawing. *Congressus Numerantium* **42:** 142160.

7. Eisen, M.B., Spellman, P.T., Brown, P.O., and Botstein, D. 1998. Cluster analysis and display of genome-wide expression patterns. *Proc. Natl. Acad. Sci.* **95:** 14863-14868.

8. Endy, D. and Brent, R. 2001. Modelling cellular behaviour. *Nature* **409:** 391-395.

9. Enright, A.J., Iliopoulos, I., Kyrpides, N.C., and Ouzounis, C.A. 1999. Protein interaction maps for complete genomes based on gene fusion events. *Nature* **402:** 86-90.

10. Gibson, M.A. and Bruck, J. 2001. *A probabilistic model of a prokaryotic gene and its regulation in computational modeling of genetic and biochemical networks* (eds. J.M. Bouer and H. Boluri). MIT Press, Cambridge.

11. Gilman, A. and Arkin, A.P. 2002. GENETIC "CODE": Representations and dynamical models of genetic components and networks. *Annu. Rev. Genomics Hum. Genet.* **3:** 341-369.

12. GO Consortium. 2001. Creating the gene ontology resource: Design and implementation. *Genome Res.* **11:** 1425-1433.

13. Griffin, J.L., Mann, C.J., Scott, J., Shoulders, C.C., and Nicholson, J.K. 2001. Choline containing metabolites during cell transfection: An insight into magnetic resonance spectroscopy detectable changes. *FEBS Lett.* **509:** 263-266.

14. Gygi, S.P., Rist, B., Gerber, S.A., Turecek, F., Gelb, M.H., and Aebersold, R. 1999. Quantitative analysis of complex protein mixtures using isotope-coded affinity tags. *Nat. Biotechnol.* **17:** 994-999.

15. Hucka, M., Finney, A., Sauro, H.M., Bolouri, H., Doyle, J., and Kitano, H. 2002. The ERATO Systems Biology Workbench: Enabling interaction and exchange between software tools for computational biology. *Pac. Symp. Biocomput.* 450-461.

16. Ideker, T., Ozier, O., Schwikowski, B., and Siegel, A.F. 2002. Discovering regulatory and signalling circuits in molecular interaction networks. *Bioinformatics* **18:** S233-S240.

17. Ito, T., Chiba, T., and Yoshida, M. 2001. Exploring the protein interactome using comprehensive two-hybrid projects. *Trends Biotechnol.* **19:** S23-S27.

18. Kanehisa, M., Goto, S., Kawashima, S., and Nakaya, A. 2002. The KEGG databases at GenomeNet. *Nucleic Acids Res.* **30:** 42-46.

19. Karp, P.D. 2001. Pathway databases: A case study in computational symbolic theories. *Science* **293:** 2040-2044.

20. Lee, T.I., Rinaldi, N.J., Robert, F., Odom, D.T., Bar-Joseph, Z., Gerber, G.K., Hannett, N.M., Harbison, C.T., Thompson, C.M., Simon, I., et al. 2002. Transcriptional regulatory networks in *Saccharomyces cerevisiae*. *Science* **298:** 799-804.

21. Loew, L.M. and Schaff, J.C. 2001. The Virtual Cell: A software environment for computational cell biology. *Trends Biotechnol.* **19:** 401-406.

22. Mendes, P. 1997. Biochemistry by numbers: Simulation of biochemical pathways with Gepasi 3. *Trends Biochem. Sci.* **22:** 361-363.

23. Oesterhelt, D. and Stoeckenius, W. 1973. Functions of a new photoreceptor membrane. *Proc. Natl. Acad. Sci.* **70:** 2853-2857.

24. Pellegrini, M., Marcotte, E.M., Thompson, M.J., Eisenberg, D., and Yeates, T.O. 1999. Assigning protein functions by comparative genome analysis: Protein phylogenetic profiles. *Proc. Natl. Acad. Sci.* **96:** 4285-4288.

25. Ruepp, A. and Soppa, J. 1996. Fermentative arginine degradation in *Halobacterium salinarium* :Genes, gene products, and transcripts of the arcRACB gene cluster. *J. Bacteriol.* **178:** 4942-4947.

26. Tamayo, P., Slonim, D., Mesirov, J., Zhu, Q., Kitareewan, S., Dmitrovsky, E., Lander, E.S., and Golub, T.R. 1999. Interpreting patterns of gene expression with self-organizing maps: Methods and application to hematopoietic differentiation. *Proc. Natl. Acad. Sci.* **96:** 29072912.

27. Tatusov, R.L., Natale, D.A., Garkavtsev, I.V., Tatusova, T.A., Shankavaram, U.T., Rao, B.S., Kiryutin, B., Galperin, M.Y., Fedorova, N.D., and Koonin, E.V. 2001. The COG database: New developments in phylogenetic classification of proteins from complete genomes. *Nucleic Acids Res.* **29:** 22-28.

28. Tollis, I.G., Battista, G.D., Eades, P., and Tamassia, R. 1999. *Graph drawing—Algorithms for the visualization of graphs*. Prentice Hall, Upper Saddle River, NJ.

29. Tomita, M., Hashimoto, K., Takahashi, K., Shimizu, T.S., Matsuzaki, Y., Miyoshi, F., Saito, K., Tanida, S., Yugi, K., Venter, J.C., et al. 1999. E-CELL: Software environment for whole-cell simulation. *Bioinformatics* **15:** 72-84.

30. Tong, A.H., Evangelista, M., Parsons, A.B., Xu, H., Bader, G.D., Page, N., Robinson, M., Raghibizadeh, S., Hogue, C.W., Bussey, H., et al. 2001. Systematic genetic analysis with ordered arrays of yeast deletion mutants. *Science* **294:** 2364-2368.

31. von Mering, C., Krause, R., Snel, B., Cornell, M., Oliver, S.G., Fields, S., and Bork, P. 2002. Comparative assessment of large-scale data sets of protein-protein interactions. *Nature* **417:** 399-403.

32. Wingender, E., Chen, X., Fricke, E., Geffers, R., Hehl, R, Liebich, I., Krull, M., Matys, V., Michael, H., Ohnhauser, R., et al. 2001. The TRANSFAC system on gene expression regulation. *Nucleic Acids Res.* **29:** 281-283.

33. Zhou, H., Watts, J.D., and Aebersold, R. 2001. A systematic approach to the analysis of protein phosphorylation. **19:** 375-378.