

Virtual Reality for the visualization of high-dimensional relationships in bioinformatics

Subtitle

Álvaro Martínez Fernández

INF-3990 Master's thesis in Computer Science November 2020



Contents

List of Figures	iii
1 Introduction	1
1.1 Background	1
1.2 Context	1
1.2.1 Software and frameworks for VR development	2
1.2.2 Clustering analysis	3
1.3 Scope and Research Problem	3
1.4 Method	3
1.5 Assumptions and Limitations	3
1.6 Significance and Contribution	3

List of Figures

1.1 Network view of the MIxT application where nodes represent genes and the modules are represented by colors. Relationships are represented by grey lines that connect a gene with another one.	2
---	---

/ 1

Introduction

1.1 Background

With the increase of biological data, new ways of analysing the produced data have been necessary in order to discover interesting patterns and make the most out of it. No matter how content-rich or expensively obtained the data is if we don't obtain anything valuable.

Humans have a great ability to discover patterns, however we are biologically optimized to see the world and the patterns in 3 dimensions. Virtual Reality (VR) is one of the best ways for better discovery in spatial dimensions. It has been demonstrated that VR help scientists work more effectively in fields like medicine [3][4] and biology[5], to mention some examples.

1.2 Context

When the amount of data is considerably big, sometimes it can be difficult to explore it, especially when we are using a 2-dimensional space like a computer screen. One example is the MIxT web application (Matched Interaction Across Tissues)[2][1]. MIxT provides a tool to identify genes and pathways in the primary breast tumor that are tightly linked to genes and pathways in the patient blood cells[1]. One of the its features is a network view where genes are represented with nodes and the edges represent statistically significant

correlation in expression between the two end-points. However exploring this network and understanding the data and its relationships is hard because there is too much information.

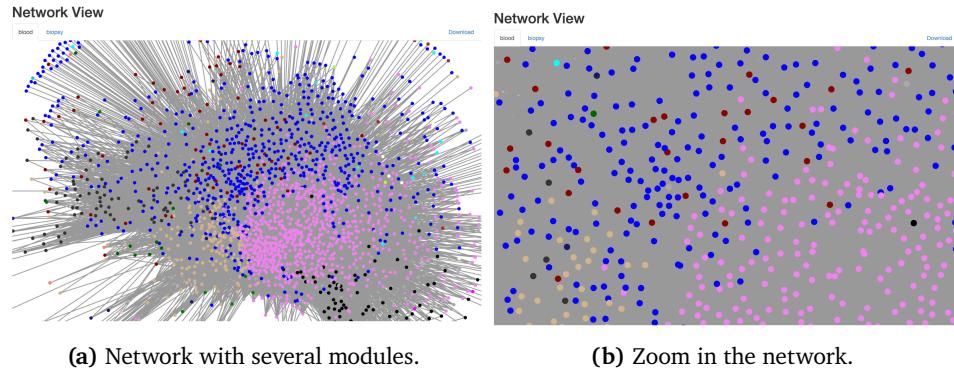


Figure 1.1: Network view of the MIxT application where nodes represent genes and the modules are represented by colors. Relationships are represented by grey lines that connect a gene with another one.

Figure 1.1 shows an example of the network visualization from MIxT. As we can see in Figure 1.1a, there are many nodes and relationships among them and when we zoom in in the network, it becomes too difficult to understand the data and the relationships as shown in Figure 1.1b.

1.2.1 Software and frameworks for VR development

Unity3D¹ and Unreal Engine² are two popular programs for development of videogames and also virtual reality games and applications. They offer integrations for Oculus Quest and other VR devices in the market. In addition, Oculus quest offers a development mode that can be activated once the glasses are connected to the PC. In this way the VR application can be tested directly on the VR device.

Virtual Reality development can also be done for the browser. WebVR³ is an open specification that makes it possible to experience VR in the browser, no matter what VR device is used. We can find many web frameworks to build VR applications for the web that are based on WebVR. Some of these frameworks are A-frame⁴, React360⁵ and three.js⁶.

1. <https://unity.com>
2. <https://www.unrealengine.com>
3. <https://webvr.info>
4. <https://aframe.io/>
5. <https://facebook.github.io/react-360/>
6. <https://threejs.org>

1.2.2 Clustering analysis

Cluster analysis is used to classify objects or cases into relative groups called clusters. Unlike supervised machine learning techniques, in cluster analysis, there is no prior information about the group or cluster membership for any of the objects. We can find many clustering approaches, two of the most commonly used ones are k-means and DBSCAN.

The k-means algorithm starts by choosing k random centers which can be manually set. Then the data points are assigned to the closest center based on their Euclidean distance.

DBSCAN (Density-Based Spatial Clustering of Applications with Noise) is another algorithm that is based on the density of the data points. The algorithm identifies clusters and expands them by scanning neighborhoods. If it cannot find any points to add, it simply moves on to a new point hoping it will find a new cluster.

1.3 Scope and Research Problem

This project focus mainly on solving the problem of visualization of high dimensional data from the MiXT project by using virtual reality. Furthermore the application allows the user to interact with the network created from the data in the virtual environment. It also allows the user compare the blood and biopsy networks at the same time in order to find relationships, which wasn't possible in the MiXT web application as this only allows the user to visualize one network at a time.

1.4 Method

1.5 Assumptions and Limitations

1.6 Significance and Contribution

This project contributes in the exploration of the possibilities that Virtual Reality offers for visualization of big data in bioinformatics.

Bibliography

- [1] Vanessa Dumeaux et al. “Interactions between the tumor and the blood systemic response of breast cancer patients.” In: *PLOS Computational Biology* 13.9 (2017). DOI: 10.1371/journal.pcbi.1005680.
- [2] Bjørn Fjukstad et al. “Building Applications for Interactive Data Exploration in Systems Biology.” In: *Proceedings of the 8th ACM International Conference on Bioinformatics, Computational Biology, and Health Informatics - ACM-BCB 17* (2017). DOI: 10.1145/3107411.3107481.
- [3] KE. Laver et al. “Virtual reality for stroke rehabilitation.” In: *Cochrane Database of Systematic Reviews* 11 (2017). ISSN: 1465-1858. DOI: 10.1002/14651858.CD008349.pub4. URL: <https://doi.org/10.1002/14651858.CD008349.pub4>.
- [4] James Xia et al. “Three-dimensional virtual-reality surgical planning and soft-tissue prediction for orthognathic surgery.” In: *IEEE Transactions on Information Technology in Biomedicine* 5.2 (2001), 97–107. DOI: 10.1109/4233.924800.
- [5] Yuting Yang et al. “Integration of metabolic networks and gene expression in virtual reality.” In: *Bioinformatics* 21.18 (July 2005), pp. 3645–3650. ISSN: 1367-4803. DOI: 10.1093/bioinformatics/bti581. eprint: <http://oup.prod.sis.lan/bioinformatics/article-pdf/21/18/3645/520673/bti581.pdf>. URL: <https://doi.org/10.1093/bioinformatics/bti581>.

