

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
List of Tables	v
1 Introduction	1
1.1 Challenges and research problem	3
1.2 Proposed solution	4
1.3 Significance and contribution	4
1.4 Outline	4
2 BigNet VR	5
2.1 Visualization and interaction with the network	6
2.1.1 Locomotion	7
2.1.2 Translation of the network	9
2.1.3 Zooming in the network	10
2.1.4 Interaction with the nodes	11
2.1.5 Node relationships	11
2.2 Building the network	11
2.3 Filtering information in the network	12
2.4 Switching dataset	14
3 Related work	15
3.1 BioVR	15
3.2 CellexaVR	15
3.3 BigTop	15
4 Evaluation and discussion	17
4.1 Experimental setup	18
4.2 Testing scalability of the system	18
5 Conclusion and future work	19

List of Figures

1.1 Visualization for network biology. a Undirected unweighted graph showing co-expression relationship between genes. b A 2D representation of a yeast protein-protein interaction network visualized in Cytoscape (left) and potential protein complexes 3D identified by the MCL algorithm from that network (right). c A 3D network of genes showing co-expression relationships. d A multilayered network integrating different types of data visualized by Arena3D. e A hive plot view of a network where nodes are mapped to and positioned on radially distributed linear axes. f Visualization of network changes over time. g Static picture showing part of lung cancer pathway. h Navigation of networks using hand gestures. i Integration and control of 3D networks using VR devices. Figure adapted[10].	2
1.2 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.	4
2.1 BigNet VR. Example of the application running on a Oculus Quest.	6
2.2 Mapping of the Oculus Quest controllers for the different actions implemented in BigNet VR: 1. Snap rotation. 2. Filter menu. 3. Scale environment. 4. Translate environment. 5. Select item in menu. 6. Oculus menu. 7. Teleport. Adapted figure from Oculus developer's page[9].	7
2.3 Teleportation technique. The user can use the joystick from the right controller to teleport to a different spot. To choose the spot a parabolic arc will appear.	8
2.4 Translation of the network functionality. The user holds the translation button on the Oculus controller and moves the hand to the direction where he or she wants the network to translate.	9

2.5	Zooming in the network functionality. The user can hold the scaling buttons on the Oculus controller to make the network bigger or smaller. In this example if we strech our hands outside, the network will expand.	10
2.6	Diagram explaining how the network is built in Unity.	12
2.7	Filtering menu in BigNet VR.	13

List of Tables

2.1	Fragment of the dataset with the categories and the genes belonging to each category from the biopsy sample.	13
2.2	Fragment of the dataset used to build the network relationships of the blood sample.	13

/ 1

Introduction

It is very cheap nowadays to produce data and many people are doing it due to technological advancement. Just as an example, in the field of genomics, the sequencing of the first human genome (2002) took around 13 years and cost over \$3 million to complete. Now we can sequence hundreds of genomes in a few days[7]. This leads to the accumulation of vast quantities of genomic data, which can be used for new scientific discoveries, diagnose of rare diseases, etc. However we still need a human expert to visually inspect the data to find new signals and discover interesting patterns or to set a diagnosis. No matter how much resources we use into extracting the data if we don't get anything interesting out of it[14]. Therefore there is a great need for new and better tools to support this task.

Some of the main problems that researchers face when analysing genomic data are information overload, data interconnectivity and high dimensionality. One way to deal with all this data is to invent novel analysis. However we still need visual inspection of the data, so the information overload still remains as an important challenge and this is what we attempt to solve. For this reason it is very important to implement efficient visualization technologies that can lead to find new patterns and the extraction of good conclusions of the data. In the field of system biology there are usually network representations where the nodes or bioentities are connected to each other, where these edges represent associations. Networks can increase dramatically in size and complexity and this is due to computational power to create very large networks, scientific knowledge about large networks and the big amounts of data that we have to

analyze. We need therefore better visualization systems for the analysis and inspection of these networks.

In Figure 1.1 we can see a representation of the evolution for visualization of networks in system biology. Before the computers, networks were represented in 2 dimensions and they were static representations that lacked interactivity, see Figure 1.1 A. With the computer era and the advancement in computer graphics, 3D representations were possible, with the addition of interactivity with the network (See Figure 1.1 C). As computer science progressed, there was a big improvement in visualization and also new technologies emerged like virtual reality, which has a huge potential with regards to the interactivity (Figure 1.1 I).

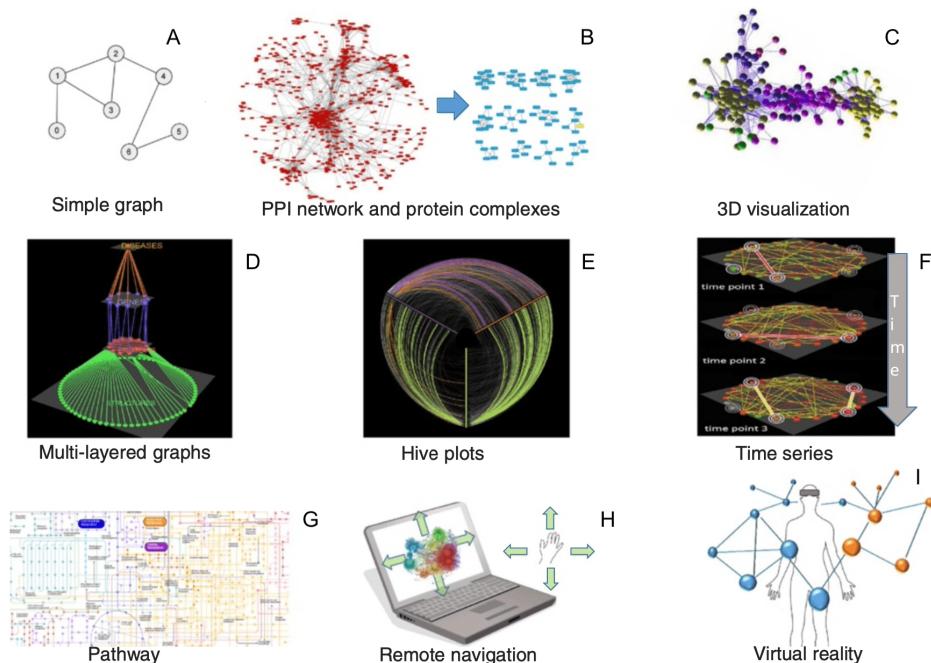


Figure 1.1: Visualization for network biology. a Undirected unweighted graph showing co-expression relationship between genes. b A 2D representation of a yeast protein-protein interaction network visualized in Cytoscape (left) and potential protein complexes 3D identified by the MCL algorithm from that network (right). c A 3D network of genes showing co-expression relationships. d A multilayered network integrating different types of data visualized by Arena3D. e A hive plot view of a network where nodes are mapped to and positioned on radially distributed linear axes. f Visualization of network changes over time. g Static picture showing part of lung cancer pathway. h Navigation of networks using hand gestures. i Integration and control of 3D networks using VR devices. Figure adapted[10].

We believe that virtual reality (VR) can offer new possibilities for visual in-

spection in large networks. Even though VR is still a field under exploration, it has been demonstrated that it helps scientists work more effectively in fields like medicine [6][12][3], biology[13][11] and neuroscience[1][8], to cite some examples. VR can be very powerful because it takes advantage of the way the human being perceives and analyzes things. We have a great ability to discover patterns, however we are biologically optimized to see the world and the patterns in 3 dimensions. Some of the advantages that VR has over non-VR approaches are the following:

1. Visualization of the network in a 3D space.
2. Possibility to move around the virtual environment and visualize the network from different perspectives as in real life.
3. Interaction with the environment by using controllers and our virtual hands in the virtual world.

We have implemented a virtual reality application for the visualization of 3-dimensional networks of data. We have used up-to-date virtual reality techniques that we think improves the visualization of this type of data structures. The techniques that we have used consist in: the exploration of the network by moving around the virtual space, making it easier for the user to see the network from different angles; interaction with the network and nodes to comprehend better the data that is being visualized; and the use of 2-dimensional user interfaces to filter and have more control of the data.

What did we learn... [Write when the evaluation is finished].

Thesis statement: *Virtual reality techniques can improve the visualization and interactivity of data networks.*

1.1 Challenges and research problem

This project focuses 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 to 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.

MixT[5] is a web application for bioinformaticians. Among other tools, it offers

a network visualization of genes which are represented as nodes in the network and where the edges represent statistically significant correlation in expression between two nodes. This tool was used in a study[4] that identifies genes and pathways in the primary tumor that are tightly linked to genes and pathways in the systemic response of a patient with breast cancer. When exploring a network in MiXT, it can be hard to understand the data and its relationships because there is too much data. This problem is easy to occur when there are too many node and edges. In figure 1.2 we can see an example of the network visualization from MiXT. As we can see in Figure 1.2a, there are many nodes and relationships among them and when we zoom in in the network, it becomes very difficult to understand the data and the relationships as shown in in Figure 1.2b.

The network is also in 2-dimensions and what we propose in this project is to use a virtual reality 3d visualization in order to cope better with this problem.



Figure 1.2: Network view of the MiXT application where nodes repsent genes and the modules are repesented by colors. Relationships are represented by grey lines that connect a gene with another one.

1.2 Proposed solution

1.3 Significance and contribution

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

1.4 Outline



BigNet VR

BigNet VR is a virtual reality application for the interactive visualization of large-scale networks in a 3D space. The network is represented using nodes and connections between them. In order to explore and visualize the data in BigNet VR, the user can walk around the 3D environment, scale the network, translate it to other places, filter the nodes using a user interface and also obtain detail information about the data.

BigNet VR works with a dataset that contains the information of the nodes and relationships that make up the network. This dataset needs to be obtained from an external source. Once we have our dataset we can load it to the application and then run it using a HMD (Head Mounted Display) for VR. Finally we can explore the network and interact with it to visualize the dataset in a VR experience.

The implementation of BigNet VR is carried out in Unity, a cross-platform game engine. This software is used for a wide range of applications, especially for the development of videogames in 3D and 2D, VR applications and engineering solutions. The programming language used to develop the application inside Unity is C#. We also used VRTK, a VR toolkit to build VR solutions in Unity. As for the VR hardware, we used an Oculus Quest headset. This type of headset is an all-in-one HMD, which means that it doesn't need to be connected to a PC to run an application, it can be run inside the hardware of the headset itself. However during the development process, the headset needs to be connected to the PC and the application can be run and tested directly from Unity.

For this chapter we will use dataset examples from MiXT to illustrate the concepts that we will talk about. MiXT is a web application that is used for the visualization of bioinformatic data[5][4] and the datasets used here contain genetic information about a woman with breast cancer. There are in total 2 datasets; the first one is from a blood sample and the second one is from the tumor sample. In Figure 2.1 we can see an example of the application running using the blood dataset from MiXT.

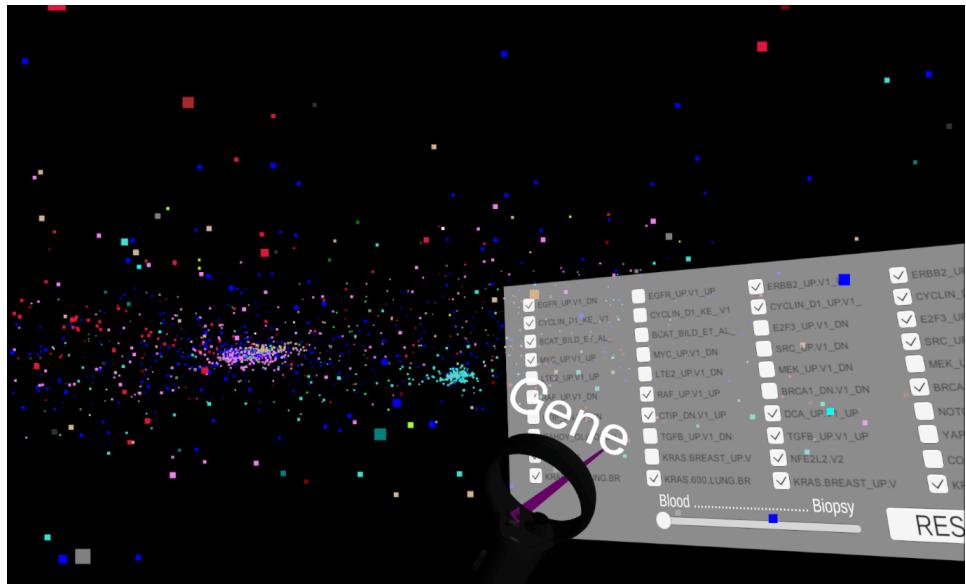


Figure 2.1: BigNet VR. Example of the application running on a Oculus Quest.

2.1 Visualization and interaction with the network

Virtual reality headsets offer a rich immersive experience. It's not only about immersing the user into a 3D environment, but also giving the user the possibility to interact with the environment itself. This makes it possible to build complex VR applications where the user can do almost anything in a virtual world. Some examples of what is possible to do in virtual reality is for example moving around, grabbing objects, interact with the environment using your hands or virtual tools, 2D interfaces and menus, etc. In this section we will explain the techniques that we have implemented to visualize and interact with the network and make the most of VR. These techniques correspond to how the user can move around in the network and how the user can visualize the data.

Oculus Quest is an all-in-one VR headset and so it doesn't need a PC nor wires to run the applications. Apart from the headset, it comes with 2 controllers; one for each hand. These controllers have inputs as buttons, thumbsticks and triggers that can be used to activate actions in the VR application. We have used some of these inputs available in the controllers in BigNet VR and mapped them to different actions that allow the user interact with the network and the environment. In Figure 2.2 we can see which actions correspond to each input from the controllers. We will explain now the different methods that we use in BigNet VR to visualize the data and how we can enable them using the controllers.

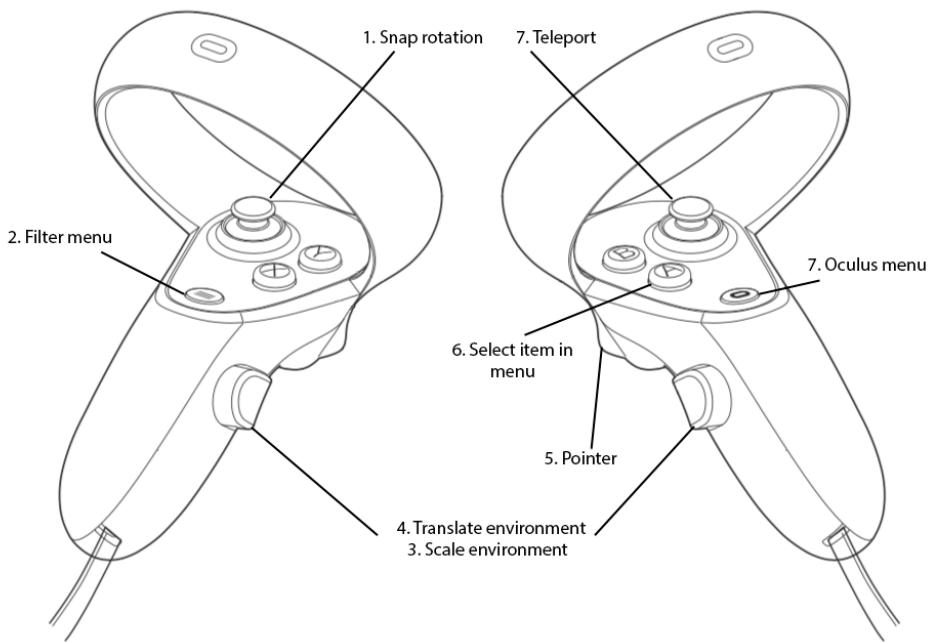


Figure 2.2: Mapping of the Oculus Quest controllers for the different actions implemented in BigNet VR: 1. Snap rotation. 2. Filter menu. 3. Scale environment. 4. Translate environment. 5. Select item in menu. 6. Oculus menu. 7. Teleport. Adapted figure from Oculus developer's page[9].

2.1.1 Locomotion

Locomotion is one of the most important ways of interaction in virtual reality experiences. It can be defined as a self-propelled movement in the virtual world. Even though moving around is not the main goal in most of VR applications, it is an important aspect for the user's perspective in order to move the user's viewpoint in the virtual world and navigate around it.

Locomotion can have an strong influence in the user's experience. A poorly designed locomotion technique can reduce the user's immersion and even introduce motion sickness, which is related to the movement that the technique produces. HMDs like Oculus Quest allow the users to control the position and the orientation of the viewpoint by moving their heads and walking. However large virtual environments such as BigNet VR need a big physical tracked area which cannot be covered by just walking around. It is for this reason that we need to use a locomotion technique that makes it possible to move around without the having to walk[2].

The locomotion technique that we use in BigNet VR is called teleportation. It consists in choosing a spot on the floor were we want to teleport to. To do this the user has to move forward the thumbstick from the right controller (see "7. Teleport" from Figure 2.2). In addition it is possible to choose which direction the user will face once the teleportation is completed. To do this we just need to rotate the same thumbstick to the desired direction. Once the user releases the thumbstick, a black flash will be followed by the new position in the space. This black flash is very important when implementing some techniques because it prevents from producing motion sickness and disorientation.

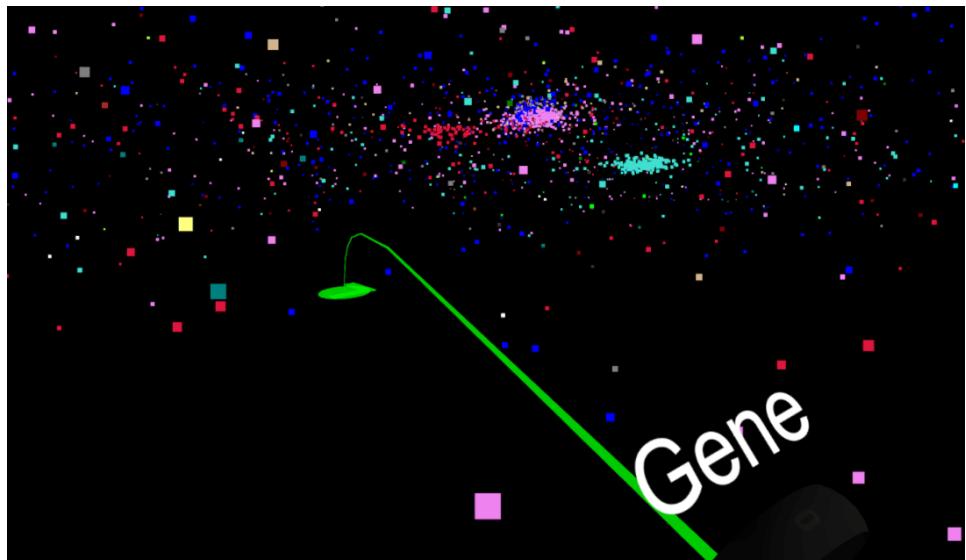


Figure 2.3: Teleportation technique. The user can use the joystick from the right controller to teleport to a different spot. To choose the spot a parabolic arc will appear.

In Figure 2.3 we can see an example of how the teleportation technique is used in BigNet VR. A parabolic arc is created in the 3D space with a circle representing the teleporting place. It can be seen as if we are throwing an object to the spot where we want to teleport to. The green circle includes also an

arrow, indicating the direction that we will face once we are teleported.

In addition to the teleportation, we have also added the possibility to rotate to the left or to the right so that the user doesn't need to rotate the head too much. This action is triggered using the thumbstick on the left hand (See 1. Snap rotation in Figure 2.2). By moving the thumbstick to the left side, the camera will rotate 45° to the left side, and 45° to the right side if the user moves it to the right side. A black transition is also used in this case before the rotation happens to avoid motion sickness.

2.1.2 Translation of the network

By teleporting to different places in the environment we allow the user see the network from different perspectives. However it is also interesting being able to move the network and specially move in a precise way so that the user has more control over what it is being visualized. The user might for instance be able to see the network or a specific node or cluster from above or also from below. To do this we have implemented a functionality to translate the network in the 3D space.

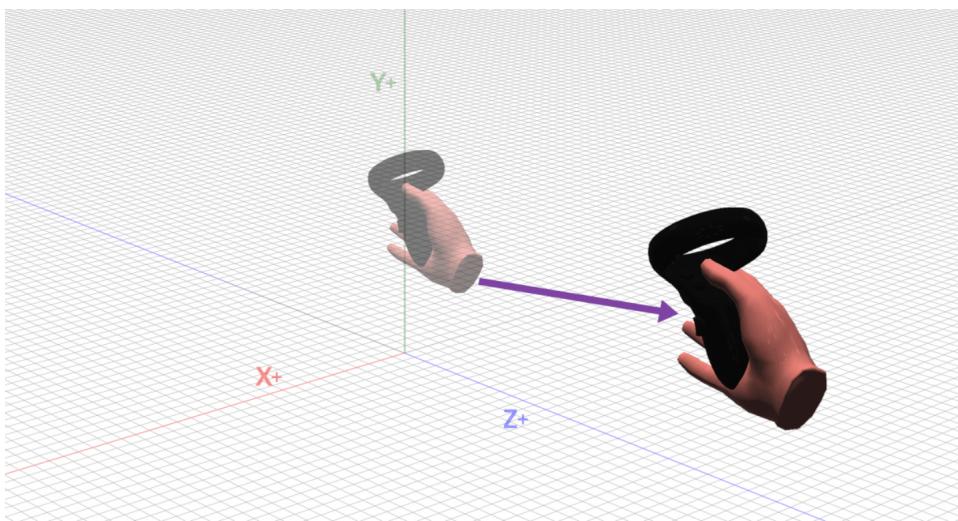


Figure 2.4: Translation of the network functionality. The user holds the translation button on the Oculus controller and moves the hand to the direction where he or she wants the network to translate.

To translate the network in BigNet VR, the user needs to press on the hand trigger from the right controller (see "4. Translate environment" in Figure 2.2). Then the user needs to keep holding this trigger down and move the hand to the direction to which we want the network to move to. This intuitive approach

feels like we are just pulling from a rope tied to the network and we just move it to the direction we want.

2.1.3 Zooming in the network

When exploring a big network with hundreds of nodes and several clusters, sometimes the information can be too crowded. In our example dataset that we use in BigNet VR, there are some clusters of nodes that have too many nodes close to each other and it gets very hard to visualize them properly. A way to cope with this problem is for instance by "zooming" in the part of the network that we want to explore better. We implement then a scaling functionality that makes the network bigger or smaller.

The way we implemented the zooming functionality in BigNet VR is by using the hand triggers with the name "3. Scale environment" (see the reference in Figure 2.2). In the first place the user needs to press and hold these triggers from both controllers and then we need to expand or contract the arms, as if we were stretching out or contracting the network itself. This is also an intuitive action to do since the user might think that we are actually stretching the network with the hands.

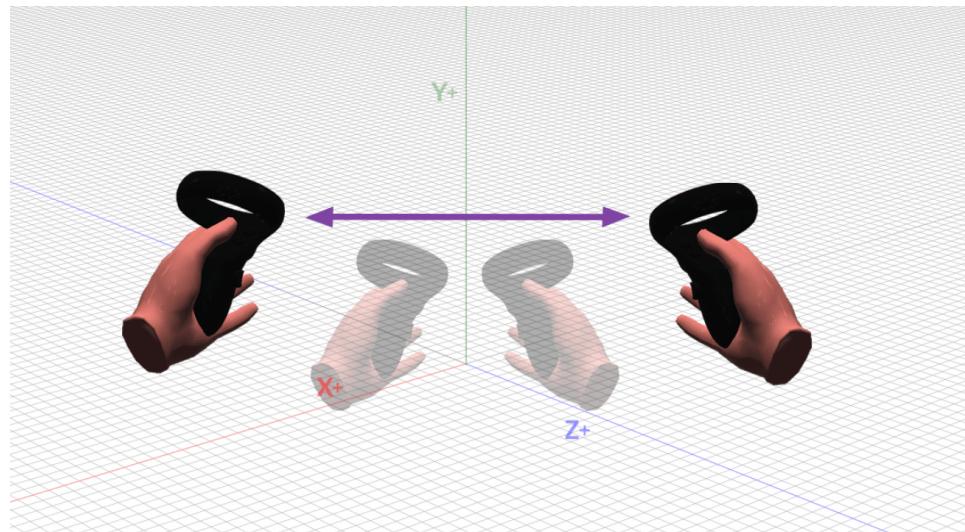


Figure 2.5: Zooming in the network functionality. The user can hold the scaling buttons on the Oculus controller to make the network bigger or smaller. In this example if we stretch our hands outside, the network will expand.

In Figure 2.5 there is a visual example of how the zooming works using the Oculus controllers. In this example the user is stretching the hands out in order to make the network bigger. The user starts in a initial position, then holds

the zooming triggers from both controllers and then moves the hands out. If we wanted to make the network smaller we would do the opposite action, by contracting the hands to the inside.

2.1.4 Interaction with the nodes

BigNet VR provides also detailed information about the data that is being displayed. The user can interact with the nodes of the network to obtain information about each of them. In our example, the nodes represent genes and the user might be interested in knowing which gene name corresponds to a specific node. The action that we need to do to obtain the name of the gene is to point at the node that we are interested in and press the "5. pointer" index trigger on the right controller (see Figure 2.2). When we press this trigger, a laser pointer will appear from the virtual controller inside of the application. By pointing to a specific node with the pointer, we will get the name of that gene node that will be displayed in a rendered text.

[User figure showing how the pointer works and explain it. This part of the implementation needs to be improved.]

2.1.5 Node relationships

Finally, our dataset has information about the relationships between the nodes. BigNet VR is implemented to show also this information. Because there can be many relationships in the dataset we don't show them all at the same time. We can only see those of the nodes that are close to the user. The way that these relationships are represented is with lines between the nodes. These lines disappear once the user moves away from the nodes.

[User figure showing how the relationships are shown in the application and explain the example. This part of the implementation needs to be improved.]

2.2 Building the network

BigNet VR uses datasets that have been created from an external source to build the network. The datasets have to be in CSV format and we need 2 datasets in total to build the network. The first one contains the information about the categories and the nodes that belong to each category and the second dataset has information about the relationships between each of the nodes.

The following diagram in Figure 2.6 explains in a broad way what steps we follow to build the network in Unity. We start with the 2 datasets described before, containing the data for the network. We process the dataset with information about the categories and build a particle system object in Unity, where each particle corresponds to a node. We also add to each particle a color, which is related to the category that the particle belongs to. Finally we apply a cluster algorithm using the data from the dataset with information about the relationships.

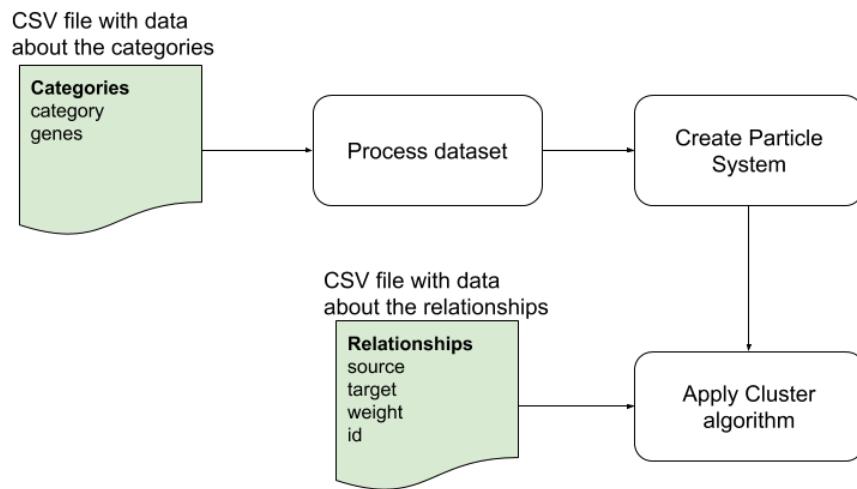


Figure 2.6: Diagram explaining how the network is built in Unity.

Following our bioinformatic network example, in Table 2.1 and Table 2.2 we show an extract of how the datasets look like. Table 2.1 has the data about the nodes or genes in this case and the categories to which each node belongs to. These categories are named by colors and these color names will be used by BigNet VR to color each particle of the particle system. Then in Table 2.2 we have the information about the relationships. This dataset can be very extend because each row of the CSV file corresponds to a relationship between 2 nodes or genes in this case and one gene can be connected to many other nodes.

2.3 Filtering information in the network

Another feature that BigNet VR uses to improve the visualization of networks is a filtering menu. When we have huge amounts of data in large networks, it is sometimes necessary to show less or more data. By filtering the nodes we

category	genes
brown	ARHGAP30 FERMT3 ARHGAP25 CD53 PLEK IRF8 DOCK2
cyan	SAFB MOB3A RAB35 ABR ASCC2 CDC37 ANKFY1 GLTSCR1
darkgrey	RAB40C ZNF213 ZNF263 PIGQ RHDF1 RAB11FIP3
darkorange	TCEB1 MRPL13 ENY2 MTERF3 UBE2W WDYHV1

Table 2.1: Fragment of the dataset with the categories and the genes belonging to each category from the biopsy sample.

source	target	weight	id
AAMP	ARGLU1	0.102486209330144	AAMP-ARGLU1
ACADM	FOXN2	0.107506881676173	ACADM-FOXN2
ACADM	MBNL1	0.12269622045714	ACADM-MBNL1
ACADM	PPM1B	0.103496640767895	ACADM-PPM1B

Table 2.2: Fragment of the dataset used to build the network relationships of the blood sample.

can visualize only the part that we are interesting in.

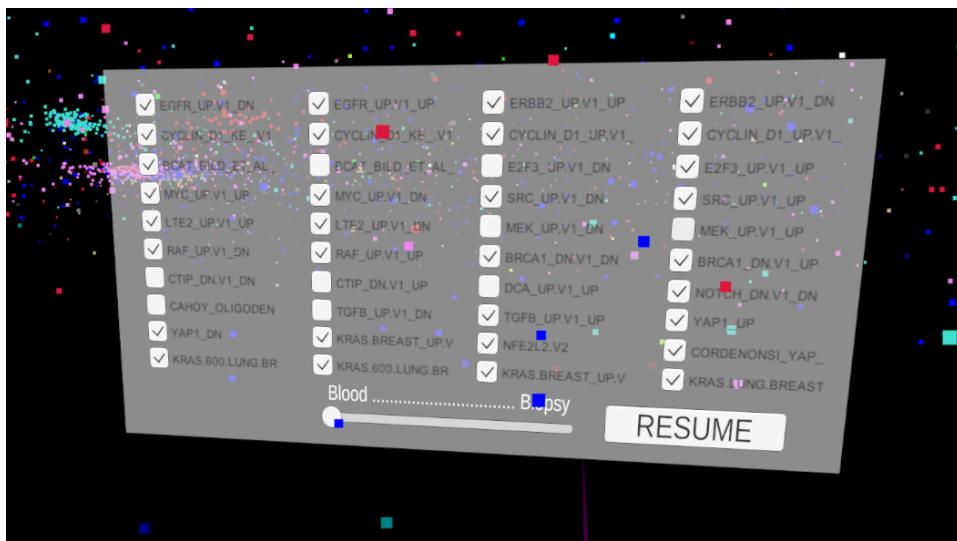


Figure 2.7: Filtering menu in BigNet VR.

We have built a 2 dimensional menu in Unity, see Figure 2.7, to filter the data in our example network. We use checkboxes for the filtering. From a starting point, all the boxes are checked, and if the user wants to hide a part from the visualization it is done by unchecking the box. To show the filtering menu we need to press on the menu button from the left controller, see the "2. Filter menu" in Figure 2.2. The to check or uncheck the boxes we need to use the A button from the right controller, named "6. Select item", see Figure 2.2.

2.4 Switching dataset

Finally BigNet VR has also the possibility to switch between datasets. This can be done in the filtering menu by pressing the menu button from the left controller and there we can see a slider UI element as in Figure 2.7 which we can move to the right or to the left in order to change the dataset that is being visualized. When we switch the dataset, the filters are also reset to their default state, so all of them will be checked in our example. In our bioinformatic usecase we can switch from the blood dataset to the biopsy one.

/3

Related work

I will focus in this chapter on VR applications found in the literature for the visualization of bioinformatic data.

3.1 BioVR

3.2 CellexaVR

3.3 BigTop

/ 4

Evaluation and discussion

Questions to answer in the evaluation chapter:

1. How big can the graph be so that it is comfortable visualizing the network?
What is comfortable? Number of FPS? How can we scale the graph? By adding nodes and spread them around, by adding more interconnexions? Should the experiment split in several parts? Scaling, filtering, moving around, etc. What is the performance by using Oculus Link and the performance using just the Quest hardware? -We can use the Unity GPU Profiler for Oculus Quest and Go in order to see the performance.
See: Getting Started w/ The Unity GPU Profiler for Oculus Quest and Go
2. How is this way of visualizing the graph better by using VR?
We are researchging the technology and the test with actual users is for future work.
3. In what way can the application and the visualization of the graph be improved?
Argue in the discussion part.

4.1 Experimental setup

@TODO Write about the specs of the system being used for the evaluation: the machine, software, etc.

4.2 Testing scalability of the system

@TODO Write about how scalability was tested. A test scene was built in order to test de network and the scalability of it. A random network is built and we test how comfortable it is to navigate through it. This is measured by the FPS rate.

/5

Conclusion and future work

Bibliography

- [1] Corey J. Bohil, Bradly Alicea, and Frank A. Biocca. “Virtual reality in neuroscience research and therapy.” In: *Nature Reviews Neuroscience* 12.12 (2011), 752–762. DOI: 10.1038/nrn3122.
- [2] Evren Bozgeyikli et al. “Point & Teleport Locomotion Technique for Virtual Reality.” In: *Proceedings of the 2016 Annual Symposium on Computer-Human Interaction in Play - CHI PLAY 16* (2016). DOI: 10.1145/2967934.2968105.
- [3] “Commentary on Rose, F.D., Brooks, B.M., & Rizzo, A.A., Virtual Reality in Brain Damage Rehabilitation: Review.” In: *CyberPsychology & Behavior* 8.3 (2005), 263–271. DOI: 10.1089/cpb.2005.8.263.
- [4] 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.
- [5] 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.
- [6] 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>.
- [7] Mike May. “LIFE SCIENCE TECHNOLOGIES: Big biological impacts from big data.” In: *Science* 344.6189 (2014), pp. 1298–1300. ISSN: 0036-8075. DOI: 10.1126/science.344.6189.1298. eprint: <https://science.sciencemag.org/content/344/6189/1298>.
- [8] Matthias Minderer et al. “Virtual reality explored.” In: *Nature* 533.7603 (2016), 324–325. DOI: 10.1038/nature17899.
- [9] *OVRInput*. URL: <https://developer.oculus.com/documentation/unity/unity-ovrinput/>.
- [10] Georgios A. Pavlopoulos et al. “Visualizing genome and systems biology: technologies, tools, implementation techniques and trends, past, present and future.” In: *GigaScience* 4.1 (2015). DOI: 10.1186/s13742-015-0077-2.

- [11] David A. Thorley-Lawson, Karen A. Duca, and Michael Shapiro. “Epstein-Barr virus: a paradigm for persistent infection – for real and in virtual reality.” In: *Trends in Immunology* 29.4 (2008), 195–201. DOI: 10.1016/j.it.2008.01.006.
- [12] 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.
- [13] 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>.
- [14] Jimmy F. Zhang et al. “BioVR: a platform for virtual reality assisted biological data integration and visualization.” In: *BMC Bioinformatics* 20.1 (2019). DOI: 10.1186/s12859-019-2666-z.

