**A Real Time Three Dimensional Representation of the Social Network ‘Reddit’ for use in a Virtual Reality Environment.**

Adam Edward D’Arcy

12523233

Final Year Project – 2017

B.Sc. Hons Computer Science and Software Engineering

****

Department of Computer Science

Maynooth University

Maynooth, Co. Kildare

Ireland

A thesis submitted in partial fulfilment of the requirements for the B.Sc. Hons in Computer Science and Software Engineering

Supervisor: Kevin Casey

## 

## Declaration

I hereby certify that this material, which I now submit for assessment on the program of study as part of B.Sc. Honours Computer Science and Software Engineering qualification, is *entirely* my own work and has not been taken from the work of others - save and to the extent that such work has been cited and acknowledged within the text of my work.

I hereby acknowledge and accept that this thesis may be distributed to future final year students, as an example of the standard expected of final year projects.

Signed: Adam D’Arcy Date: 3rd/March/2017

**Chapter One: Introduction**

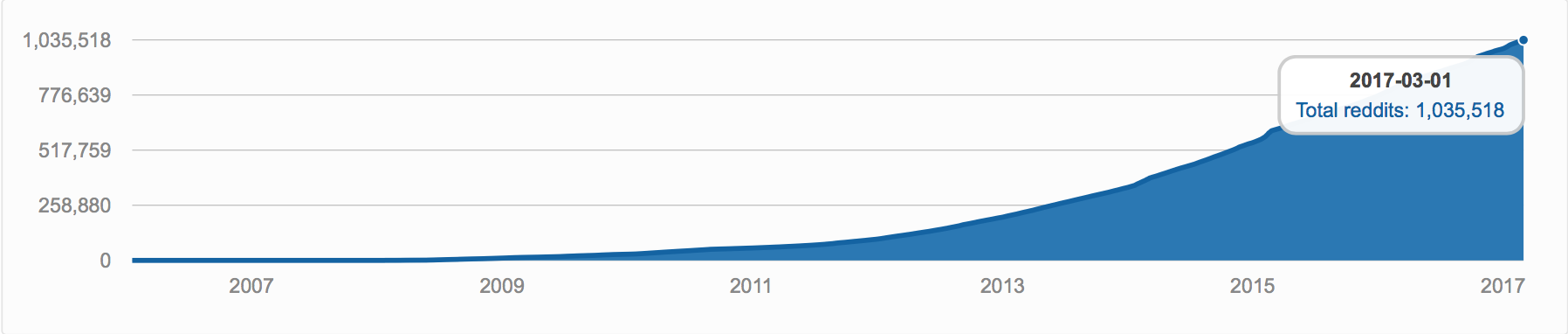
**Summary**

In this chapter I will discuss the main topics addressed in this project, the motivations behind the project, the high level abstract and technical problem the project solves and finally how I approached solving these problems and what I achieved in doing so.

* 1. **Topic addressed in this project**

The main topic addressed in this project was to create an immersive, real time representation of the social network Reddit. In order to achieve a modern, useful and truly immersive experience the development of this representation was geared toward use in a Virtual Reality environment.

There is no greater way to provide an immersive experience at the moment than through that provided by Virtual Reality. Headsets such as the Oculus Rift and HTC Vive and mobile alternatives like Googles Cardboard and Samsung’s Gear VR are providing new avenues to train Professional Sports Stars [1] and treat Posttraumatic stress disorder [2] among many other applications of the technology.



**Figure 1.0: Growth of Number of Subreddits on Reddit, 2006/01/17 – 2017/03/01 [3]**

Reddit is often described as the “front page of the internet”, but the front page is just the tip of the iceberg. As of 01-03-2017, 1,035,518 Subreddits existed on Reddit according to redditmetrics.com [3], the same day approximately 16.7 Subreddits a minute where created on the site. The topic of this project is to develop a way of representing as much of this information as possible into a real time, interactive experience to provide a glimpse of a larger portion of Reddit than isn’t obvious from interacting through Reddit’s text based UI and in doing so give some insight into how these Subreddits are connected and finally to show something about Reddit not visible in the usual boring two-dimensional screen based world we live in.

* 1. **Motivation**

There is something for everyone on Reddit. There are Subreddits for everything from ‘r/FUNNY’ to ‘r/PERSONALFINANCE’ and communities aimed at programmers (r/programming), cat people (r/CatsStandingUp), freaks (r/Freaks), geeks (r/Nerd\_Corner) and those who just want to watch the world burn (r/ooerintensifies) among more than one million others. But what’s the difference between r/programming and r/coding, and how are they connected? Or are they even connected? And how would it be possible to represent this information in an intuitive manner without overwhelming the end user? These are the questions I attempt to address with this project.

With this project I attempted to create a representation of Reddit that encapsulates more than individual Subreddits, individual user preferences or even a subsection of Reddit as a whole. I strived to provide the end user with an experience that immerses them deep within Reddit’s molecular build up, to develop a representation of Reddit, as much of Reddit as possible, that encapsulates how Subreddits are connected, which Subreddits are connected, which Subreddits aren’t connected and to provide a way of instantly delving deep into the soul of Reddit. //add the abstract kinda problem in more detail

* 1. **Problem Statement**

Reddit is a social network compromised of more than one million Subreddits, how would it be possible to develop an intuitive and functional yet enticing representation of this complex network without overwhelming the end user? And would it be possible to show how Subreddits are connected to one another? The main technical problem that has to be solved during the development of this project is to develop an algorithm capable of interpolating Reddit data in real-time and producing useful results that can be used to represent Reddit in a three-dimensional Virtual Reality environment.

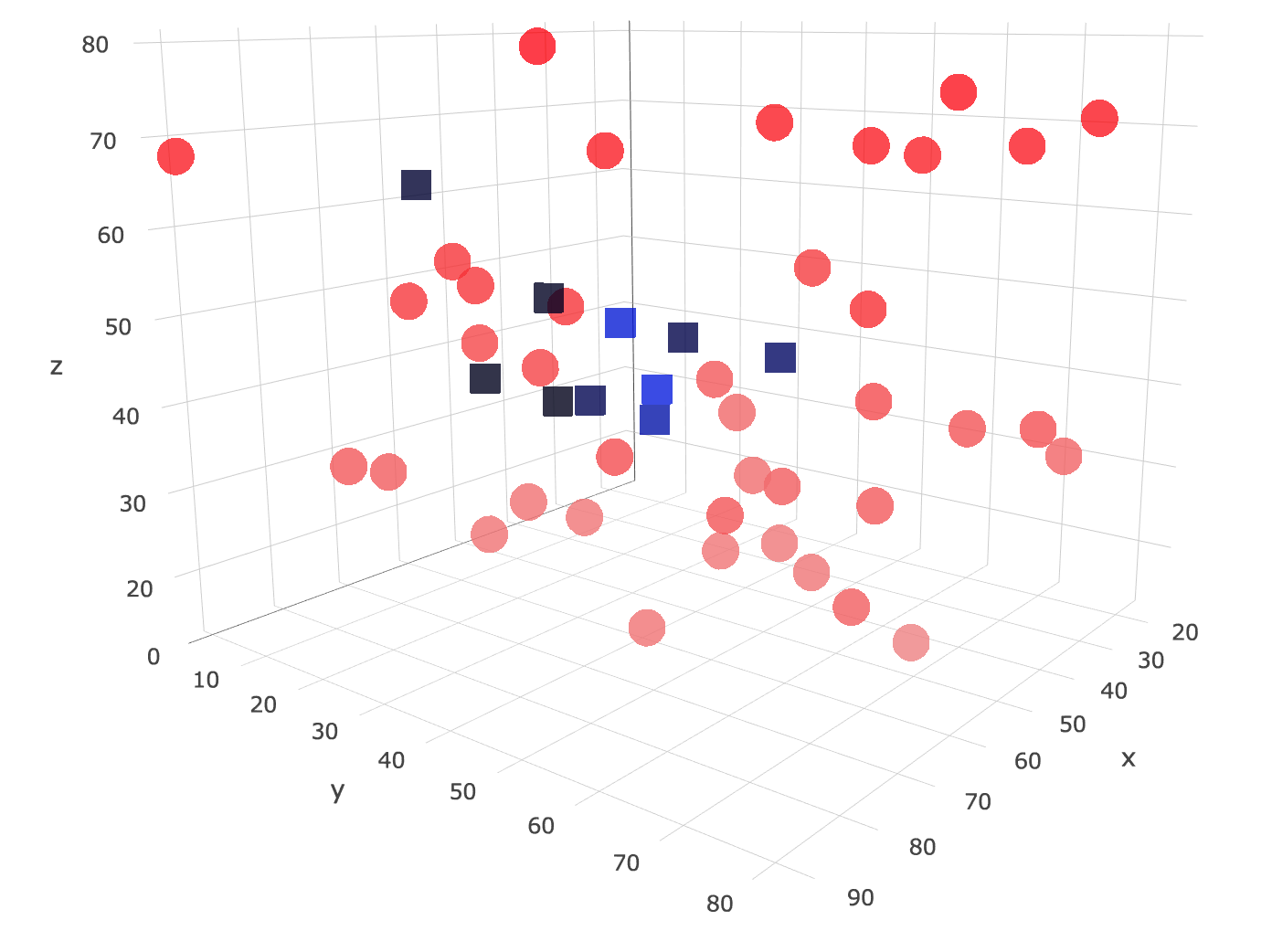
This problem provides many challenges such as;

Speed: //

* 1. **Approach**

//more analysis – online algorithm

To solve the problem at hand I designed a system to graph Reddit in Real time. In the graph Subreddits would be represented as Nodes and the system would update Node positions based on how they are connected too other Nodes. The system will find connections between Nodes by investigating User connections between Subreddits.



**Figure 1.1: An early model of the Reddit Graph – Subreddits are represents as red spheres, Users are represented as blue cubes.**

To decide how closely connected Subreddits are and to position them accordingly the system will investigate user connections between Subreddits. If User A comments on Subreddit0 and Subreddit1, then Subreddit0 and Subreddit1 are connected and should be updated accordingly.

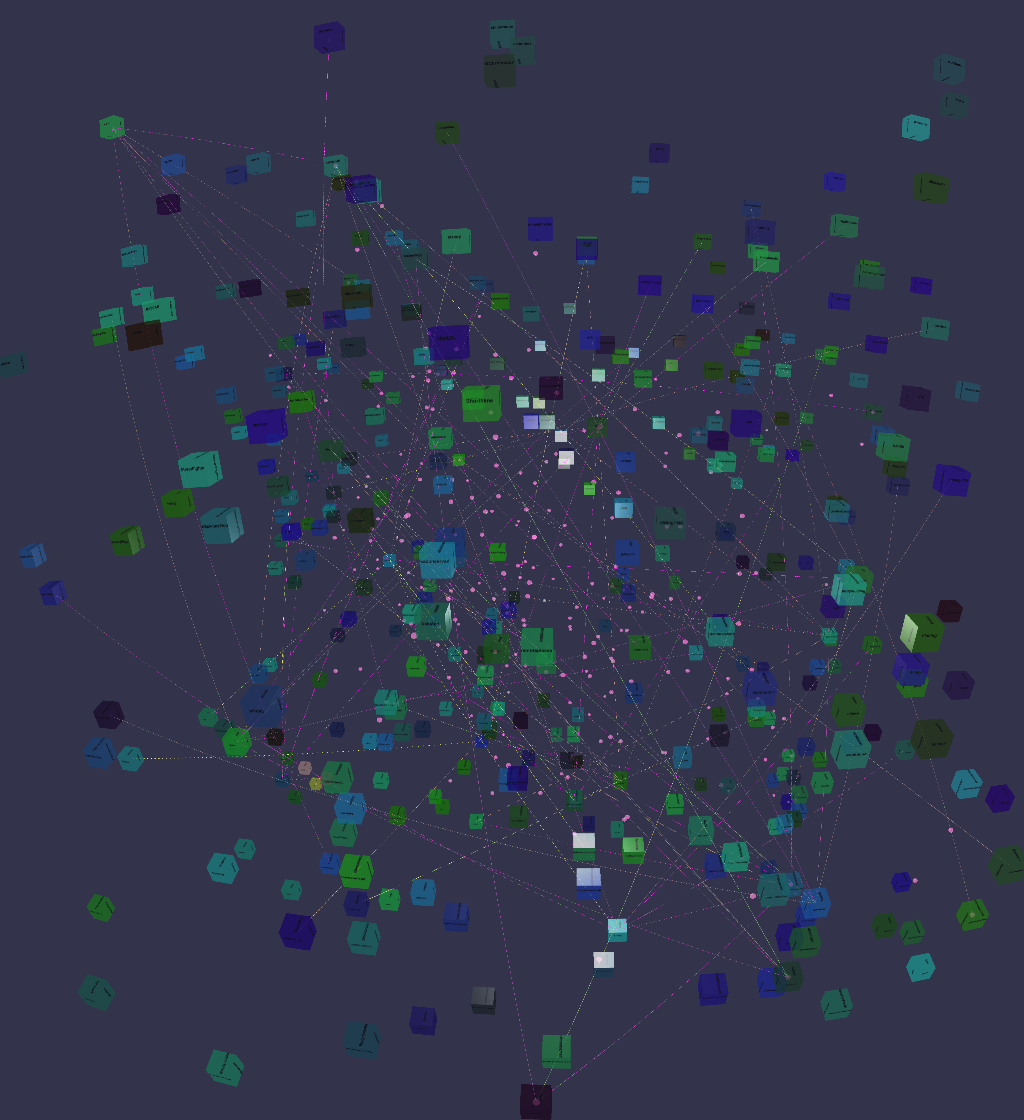
In evaluating this solution the graph should provide a visual representation directly comparable to the raw data. (provide example, control: step-size of positioning algorithm)

Initially the system acted on a large of set of Reddit data to provide a base for future real time updates. Using this base the system will search for new Subreddits and add them to the graph, updating older Nodes positions as it does so.

* 1. **Project**

Throughout this project there where three main achievements;

1. Developing a system to position and update nodes to graph Reddit: Developed an online algorithm to interpolate Reddit data in real time and return three-dimensional coordinates associated with Subreddits and users in order to position them as nodes in a graph. The algorithm updates the positions of these nodes in real time.
2. Designing and implementing a three-dimensional visual representation of Reddit: Using the position data from the above algorithm a three-dimensional representation of Reddit was designed that updates in real time and allows for user interaction with nodes.



**Figure 1.2: Screenshot of three-dimensional representation of Reddit – Subreddits are represented as colored cubes, Users are represented as pink spheres.**

1. Implementing the representation in Virtual Reality…..

**Chapter Two: Technical Background**

**Summary**

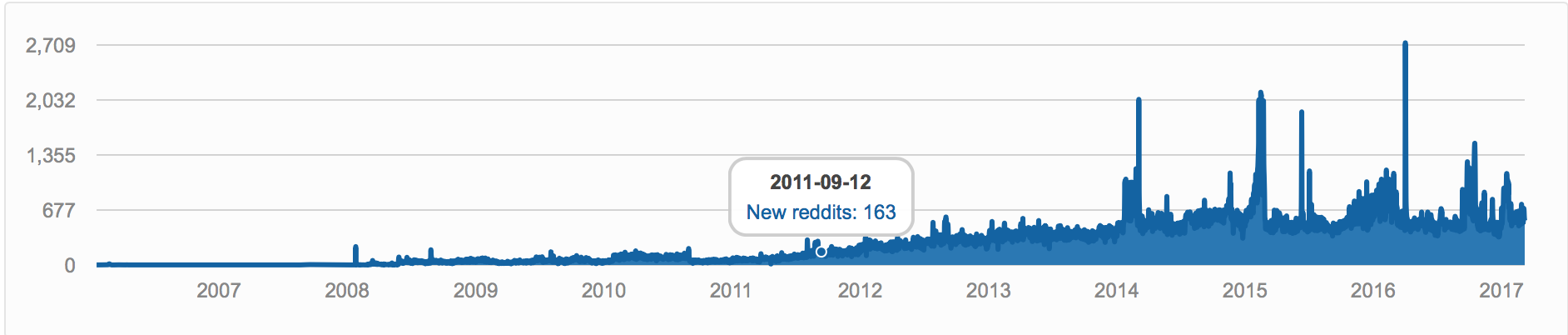
In the following chapter I will discuss the material addressed in the project, both technical and non-technical. For example, what work has already been completed in this area and tools used throughout the development of the project.

**2.1 Topic Material**

There have been a variety of applications developed using the Reddit API, both real time and interactive.

The site redditviz [4] used Reddit data to create an interactive map of Reddit using the Gephi OpenOrd Layout [5] algorithm to decide the layout of the map and has been described by fastcodedesign.com as “Reddit is the “front page of the internet” but this handy tool can take you straight to page 52” [6].

Secondly, the site redditmetrics.com [7] tracks Reddit and Subreddit metrics such as “New Subreddits by date” and “Fastest growing”.



**Figure 1.3: Screen shot from redditmetrics.com of a “New Subreddits by date” graph.**

Across the web there are a wide variety of three dimensional representations of data that demonstrate how data can be not only as a source of information but also as art. Some of the best examples are of this “information art” are provided below.

Metrogram3D [8] is a real time simulation of the Tokyo metro built with WebGL and is a premium example of how real time data three dimensional representations can be used to convey complex network information in interesting, effective and enticing ways that a two-dimensional static simulation could never hope to achieve.

Finally, armsglobe [9] a three-dimensional “interactive visualization of government-authorized small arms and ammunition transfers from 1992 to 2010” produced by Google as part of the Google Ideas INFO Summit with the Peace Research Institute Oslo small arms database is a stunning example of how databases can be represented to convey the information at their heart.

**2.2 Technical Material**

Throughout the development of this project I used a variety of different technical tools and leant on many different resources in order to help implement them.

The tools used throughout the project are as follows:

* HTML5
* CSS5
* JQuery
* JavaScript
* TypeScript
* Python
* PRAW
* Reddit API
* Firebase
* BabylonJS
* // leap motion tools

At the beginning of the process I undertook the task of learning TypeScript and Baylonjs to develop the visualization for the project.

I also began to learn PRAW in order to obtain the data needed from Reddit with Python. PRAW was chosen due to its ease of use compared to the raw Reddit API, and my previous experience with manipulating data in Python.

This process was completed in the early stages of the project, however as the project progressed I continued to refer to a variety of resources and material such as Stack Overflow and tutorial sites.

Other technical material viewed/used included:

* Reddit API [9] this was the starting point for interacting with Reddit.
* PRAW “Getting Started”. [11]
* The Python.org python-firebase [12] webpage to learn from information regarding Firebase setup in Python setup and for interacting with Firebase in Python.
* A “Getting started” tutorial on the Firebase website [13] regarding using Firebase in a web app.
* Information on how to resolve JQuery with TypeScript via Stack Overflow [14]
* Tutorials on Babylonjs [15]
* Information regarding the Leap Motion SDK [16] and getting started with the Leap Motion in Virtual Reality [17].

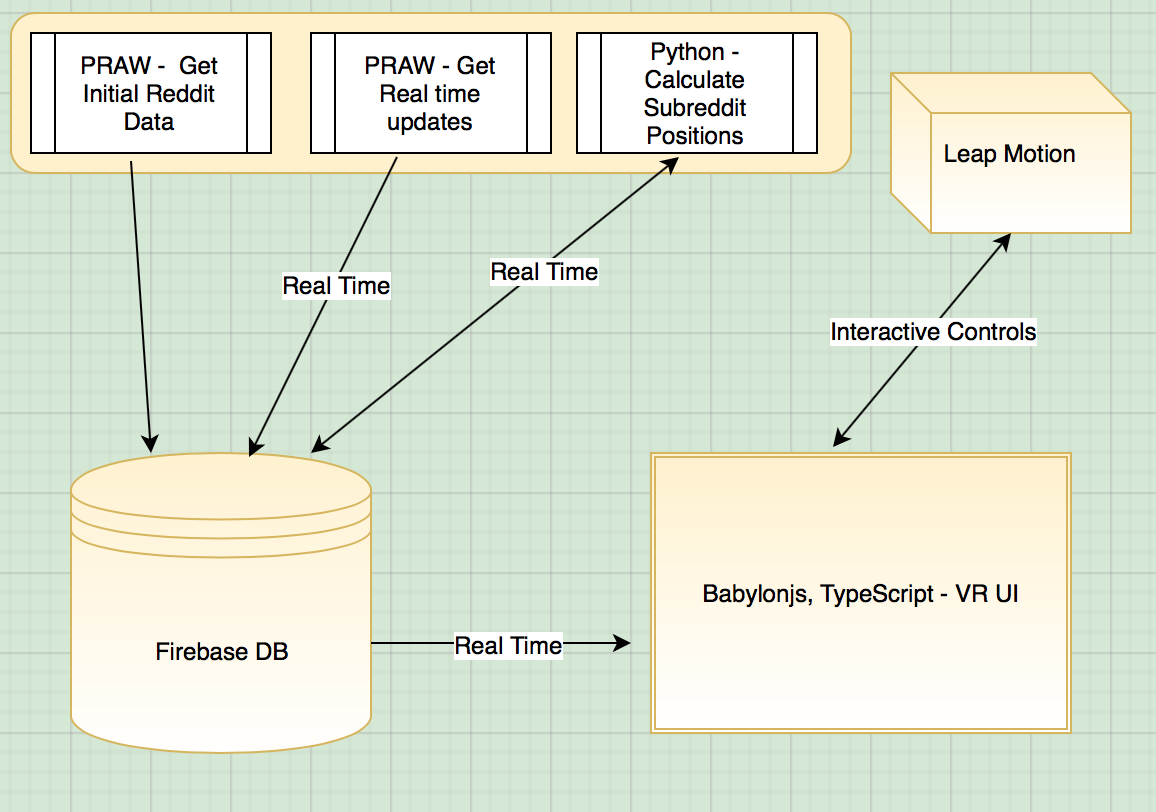
All of this material proved useful at certain stages throughout the project. In section 4.1 I will demonstrate in more detail how some of above technology was used in key areas throughout the development stage.

**Chapter Three: The Problem**

**Summary:**

In the following chapter the Project documentation and analysis of the Project Problem will be discussed. The section includes documentation regarding the development of the project and discussion surrounding both the functional and non-functional requirements of the Project.

**3.1 Project UML Documentation:**

****

**Figure 1.4: High Level Description of Project Architecture**

**3.2 Project Analysis:**

The main aim of this project is to develop an interactive three-dimensional visualization that in some way accurately represents Reddit, ideally a three-dimensional map. To create an accurate representation of Reddit I need some information about the site. To do this I need to gather data about Reddit as it is in real time. Using this data I need to create a program that will output some information about Reddit, such as three-dimensional coordinates corresponding to Subreddits, for use in the visualization. These outputs will need to be an accurate representation of Reddit in some way and updateable in real time. I also need some way of creating a representation of the data for viewing by the user.

Ideally the outputs, f(x), will represent how strongly connected two Subreddits are to one another. For example, if two Subreddits are connected they should move toward one another and the more strongly connected the Subreddits the greater the step toward one another should be. We want to maximize the step size as the distance between two connected nodes decreases, see below;

The problem can be separated into several parts:

1. Obtaining real time Reddit data.
2. Create a system using the data to output and update coordinates for Subreddit Nodes to be used in a three-dimensional map of Reddit.
3. Using the system outputs mentioned above develop, a visualization of Reddit in three-dimensions for use in a Virtual Reality environment.

The functional requirements of the project are as follows:

* Output coordinate information for Subreddits.
* Develop a system that moves nodes to represent how they are connected to other nodes.
* Update coordinate information as nodes are moved.
* Develop a functional and useful Virtual Reality ready visualization of the outputs.

The non-functional requirements are:

* Develop a visually enticing representation of Reddit.
* Develop a system that can be used across a variety of platforms and devices.

**Chapter Four: The Solution**

**Summary:**

**4.1 Choice of Tools**

A number of tools and languages where needed throughout the development of this project. The main languages used are Python and TypeScript.

I decided to use Python as it was the language I was most proficient in and I had previous experience using the language to manipulate data structures and interact with NoSQL databases, also PRAW (discussed in section 4.1.1) is an excellent Python Wrapper for the Reddit API which was used throughout the project and made Python the ideal environment for dealing with the Reddit data.

TypeScript was used during the project as it is the language of choice for the framework BabylonJS (section 4.1.3). TypeScript is a typed superset of JavaScript that compiles to plain JavaScript [18]. The TypeScript type system can reveal many common errors as will they will be flagged by the transpiler revealing many obvious mistakes in your code which can make debugging much easier, lending itself as an ideal language for web development.

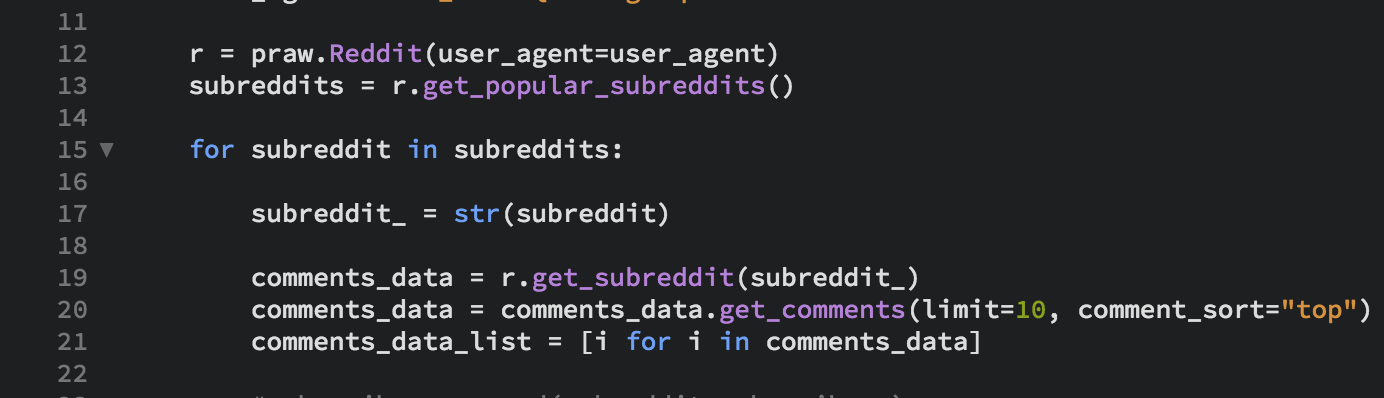
In the following sections the technologies used throughout the project will be discussed in further detail.



**Figure 1.5: An example of compiling a BabylonJS .ts file**

**4.1.1 PRAW**

PRAW (Python Reddit API Wrapper) is a Python package that allows for simple access to Reddit’s API. It allows you to access Reddit through calling PRAW methods on a Reddit instance. This allows for quick and easy access to Reddit data without needing to build Reddit URLs or worry about endpoints. PRAW also provides a built in delay functionality so the application will not elapse Reddit’s API call limit of one request every two seconds.

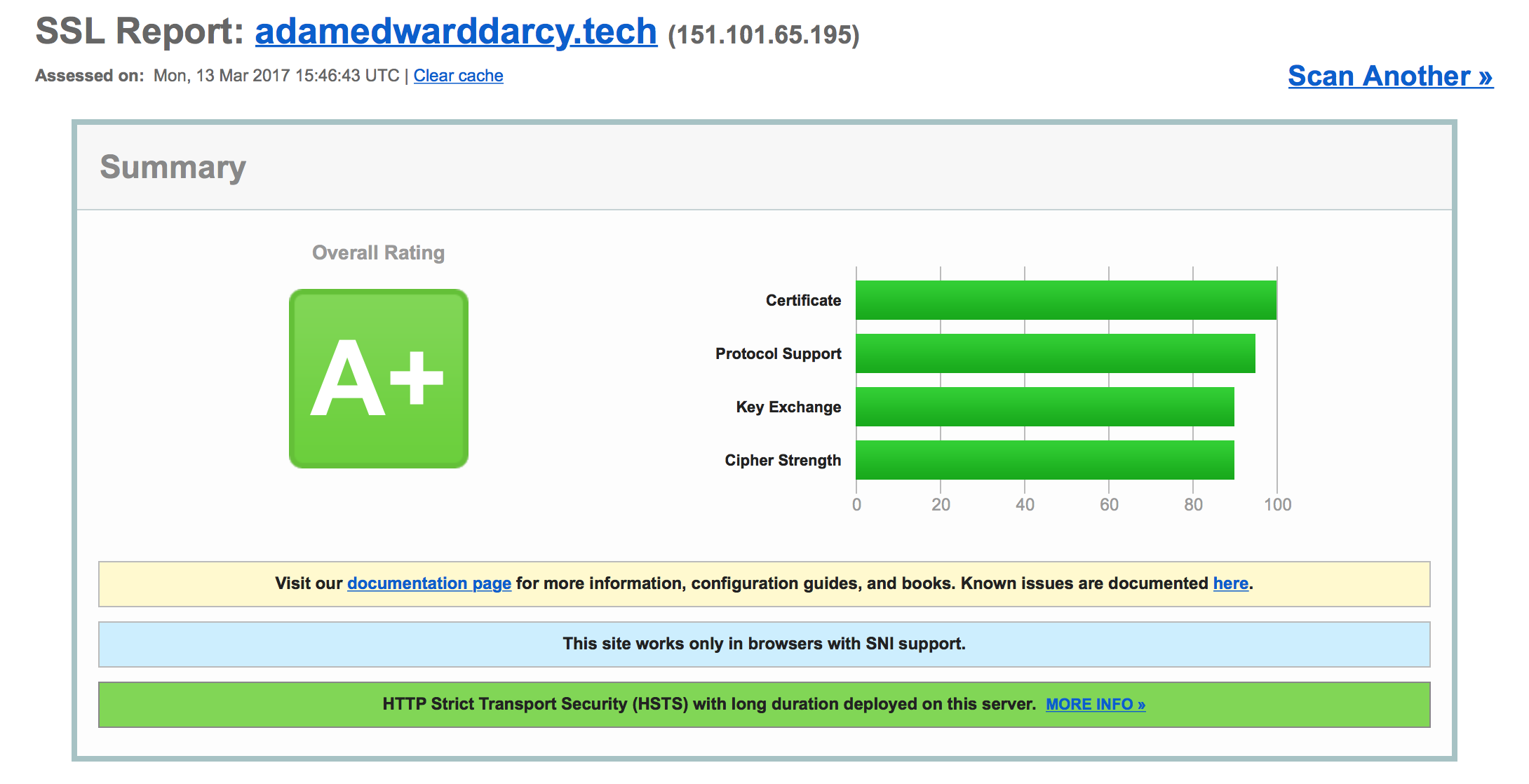


**Figure 1.6: Example of PRAW code from project**

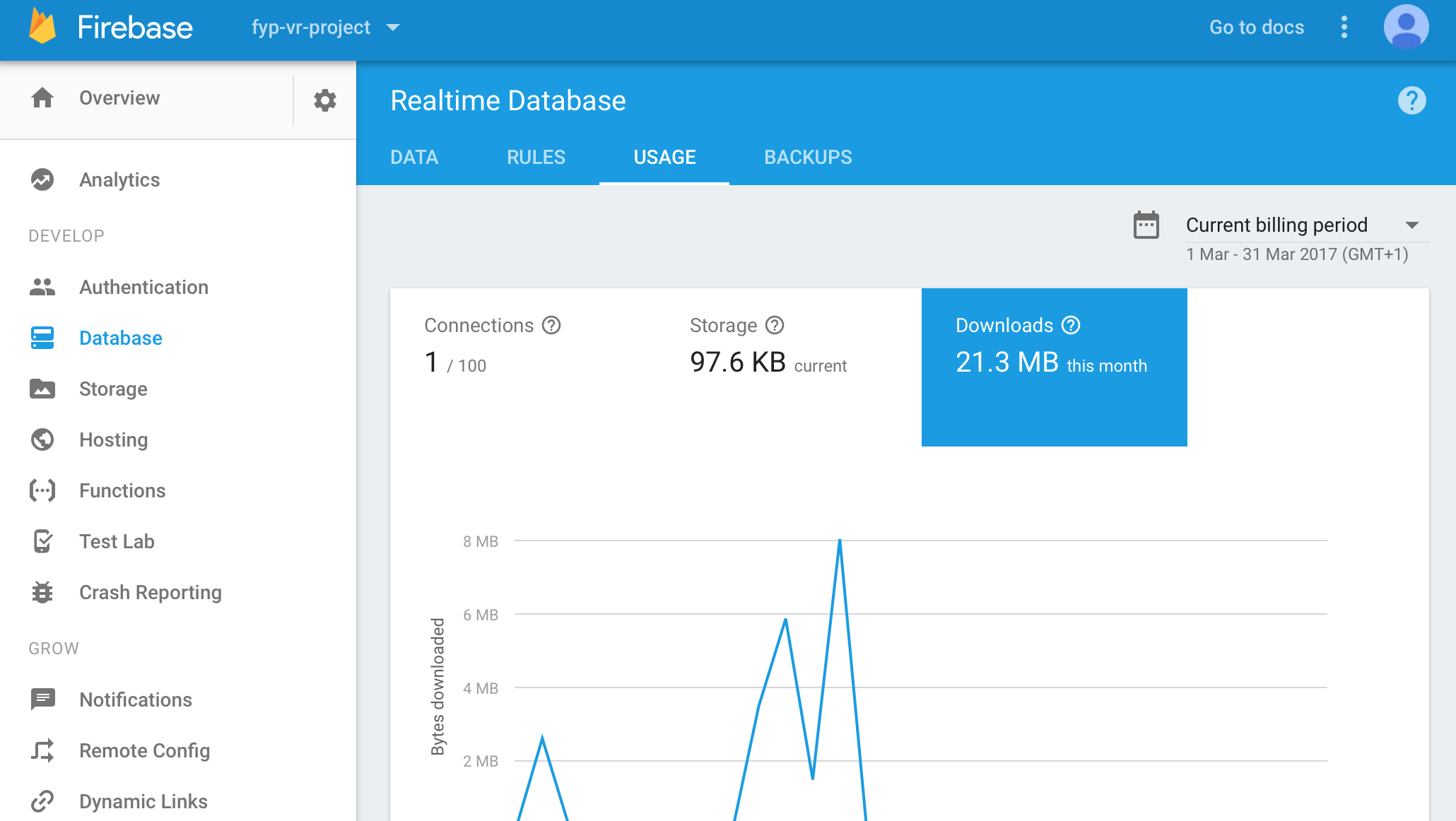
**4.1.2 Firebase**

Firebase is a web application platform that provides a real time database and backend as a service. Firebase also provides hosting for the application complete with HTTPS and SSL encryption (see Figure 1.6).

Firebase was chosen for the project due to its ease of interfacing with both Python and TypeScript making it an ideal choice for every aspect of the design.



**Figure 1.7: SSL report for the project site (** [**https://adamedwarddarcy.tech/**](https://adamedwarddarcy.tech/) **) [19]**

****

**Figure 1.8: The Firebase console**

**4.1.3 BabylonJS**

BabylonJS is a TypeScript framework for building three-dimensional applications with HTML5, WebGL and Web Audio. BabylonJS allows users to develop three-dimensional environments using TypeScript and basic HTML5 elements such as the canvas element.

The built in Virtual Reality cameras and device orientation tracking sets BabylonJS apart as the perfect framework for building a Virtual Reality application.

**4.2 Implementation**

In the following section the implementation of the project is discussed alongside the associated files which can be found in the appendix. After deciding on the final goal of the project it became apparent there would be three main areas to the implementation of the design:

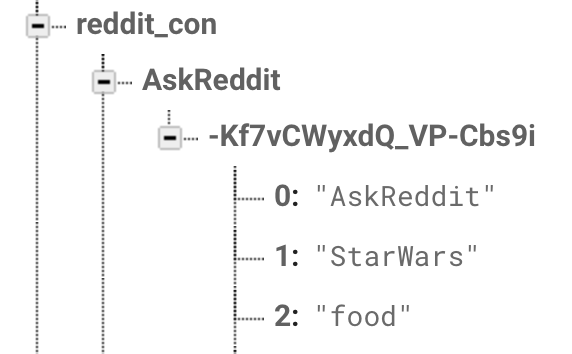
1. To develop a system to obtain real time data from Reddit.
2. To develop an algorithm that can interpolate the Reddit data in real time.
3. To develop a three-dimensional visualization of the Reddit data.

**4.2.1 Getting Started**

In order to begin mapping Reddit data in real time I began by obtaining a large set of batch data. To achieve this I obtained data from the 25 most popular Subreddits. Of these Subreddits I obtained the most recent 10 commenters (commenter = user) and of these 250 users;

I obtained the last 10 Subreddits that each user had interacted with. This process gave me a data set containing a maximum of 2,500 Subreddits.

The data was stored in two tables in the Firebase database, a key, value table where the Keys are the Subreddits initially obtained (i.e the 25 most popular Subreddits) and the values are the corresponding Subreddits obtained from User comment data [table name = reddit\_con];

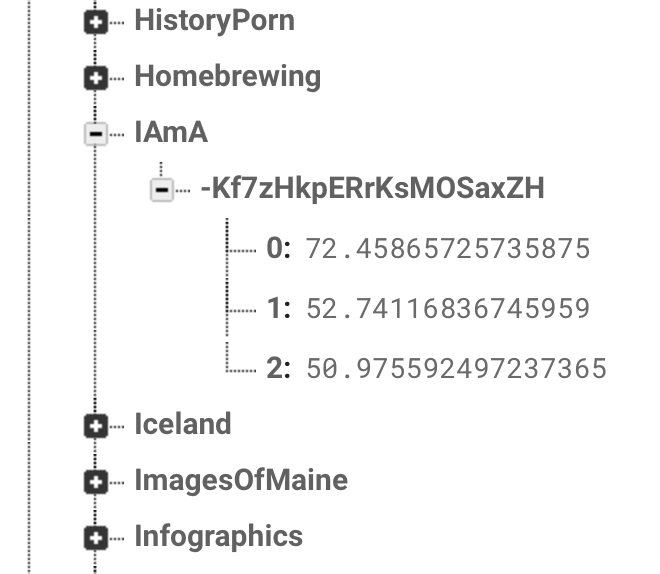


**Figure 1.9: Structure of reddit\_con table**

and a table containing each Subreddit with corresponding three dimensional coordinates [table name = reddits].

*Associated File: get\_data.py.*

How these coordinates are obtained is discussed below.



**Figure 2.0: Structure of reddits table**

Initially positioning Subreddits:

Subreddit nodes are initially positioned randomly using Pythons random uniform method to achieve a uniform distribution, these positions are later updated as will be discussed in the following section.

*Associated File: init\_nodes.py*

**4.2.2 Updating Subreddit positions:**

In order to achieve a useful data representation, Subreddits need to be positioned in a manner which reflects how the nodes relate to one another and the environment as a whole. Subreddit nodes also need to be updateable in an online real time manner. To achieve this functionality the reddit\_con table was utilized.

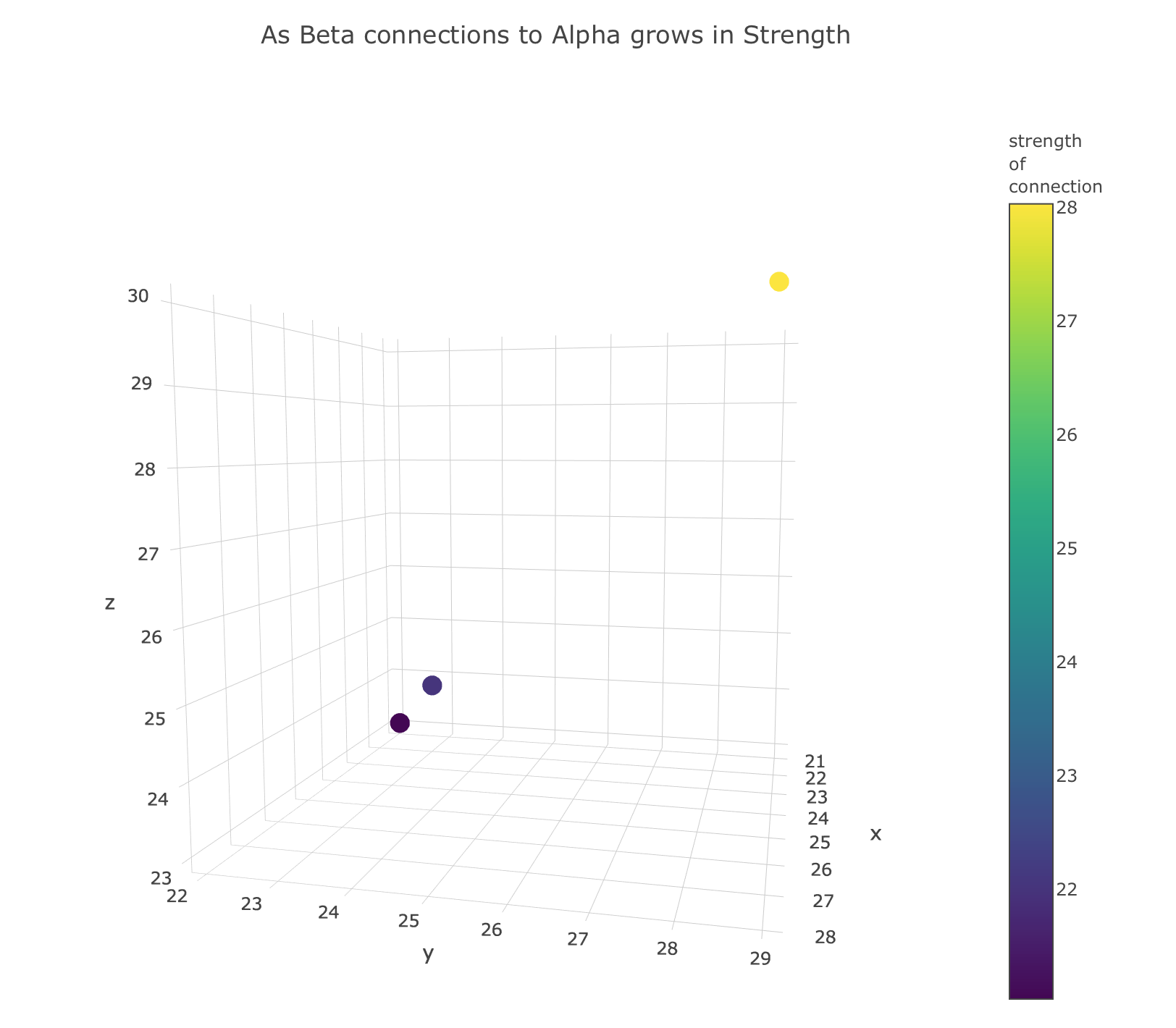
To update Subreddits key value pairs in the reddit\_con table are used, where keys are ‘Alphas’, α, and values are lists of ‘Betas’, β, see below; there corresponding positions can be found in the reddits table which will be updated.

The following method is then used:

Method A:

*α and β are considered connected if a β appears in an α keys value list, and will be updated accordingly. If a β appears in an α value list the β will take a step toward the α, the closer (more strongly connected) the α and β Nodes the larger percentage step the β will take, such that as the distance between an α and β decreases, the connection between them increases. Step size is calculated by the following formula;*

*By the above, as the connection between two nodes grows stronger the percentage step size between the nodes increases up until a threshold of the connection (distance) between two nodes being less than 1. This is demonstrated in the below graph:*

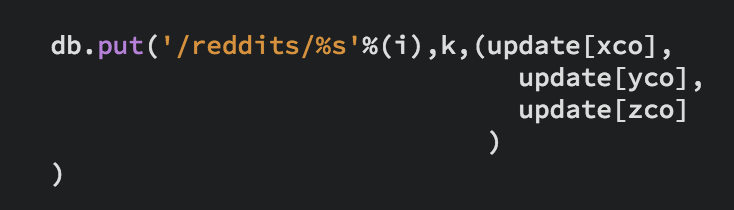
**

***Figure 2.1: A graph demonstrating that; as the connection between two nodes increases, the potential step size also increases.***

*Following the calculation of the step size the updated position of the β node is calculated and returned, as below:*

*The returned updated value is a β that has taken a percentage step toward its associated α, this step will never be larger than:*

*The update is then used to put an update to the database and the reddits table as shown in the code snippet below.*



**Figure 2.2: A code snippet (source: update\_nodes.py) demonstrating an update to the reddits table.**

*Associated Files: fx.py, update\_nodes.py*

**4.2.3 Real time updates**

Thus far only a small portion of Reddit has been mapped. The remainder of the Reddit network will be mapped in real time online as opposed to the batch method used thus far. To achieve this the reddits table is compared to the reddit\_con table and for all reddits[keys] not in reddit\_con[keys] a similar method to that used in section 4.2.1 is used, where the last 10 users to interact with the subreddit are obtained, and the last 10 subreddits each user has interacted with are obtained to create a key value of the following form;

Method A, from section 4.2.2 is then implemented. If a β does not have an associated position, it is assigned a position in the same manner seen in the ‘Initially positioning Subreddits’ portion of section 4.2.1. Once Method A has been implemented, the key, α, is added to the reddit\_con table and the process continues.

*Associated Files: fx.py, realtime.py*

**4.2.4 Three-Dimensional Representation**

At this point in the process the data has been obtained (or in the case of the real time updates, will continue to be obtained in the background), and a three-dimensional representation can begin to be formed.

**References**

1. Stephanie Loh, The San Diego Union-Tribune: <http://www.sandiegouniontribune.com/sports/sdut-virtual-reality-quarterback-trainer-derek-belch-2015may21-story.html>, 2015
2. Mary Aiken & Mike J. Berry: Posttraumatic stress disorder: possibilities for olfaction and virtual reality exposure therapy, *Virtual Reality,* June 2015, Volume 19, [Issue 2](http://rd.springer.com/journal/10055/19/2/page/1), pp 95–109
3. <http://redditmetrics.com/history/month>
4. <http://rhiever.github.io/redditviz/>
5. <https://marketplace.gephi.org/plugin/openord-layout/>
6. Mark Wilson, fastcodedesign.com: <https://www.fastcodesign.com/3021066/infographic-of-the-day/this-brilliant-visualization-could-build-a-better-reddit> , 2013
7. <http://redditmetrics.com>
8. <http://nulldesign.jp/metrogram3d/>
9. <https://armsglobe.chromeexperiments.com>
10. <https://www.reddit.com/dev/api/>
11. <https://praw.readthedocs.io/en/v3.6.0/pages/getting_started.html>
12. <https://pypi.python.org/pypi/python-firebase/1.2>
13. <https://firebase.google.com/docs/database/web/start>
14. David Sherret: <http://stackoverflow.com/questions/32050645/how-to-use-jquery-with-typescript> , 2015
15. <https://doc.babylonjs.com/tutorials>
16. <https://developer.leapmotion.com/documentation/v2/javascript/index.html>
17. <https://developer.leapmotion.com/get-started/>
18. <https://www.typescriptlang.org>
19. <https://www.ssllabs.com/ssltest/analyze.html?d=adamedwarddarcy.tech&s=151.101.65.195>