3D Visualization of Heterogeneous User Interactions in a Social Network

by

Wanjun Pei

Thesis submitted to the
Faculty of Graduate and Postdoctoral Studies
In partial fulfillment of the requirements
For the M.A.Sc. degree in
Electrical and Computer Engineering

School of Electrical Engineering and Computer Science
Faculty of Engineering

Abstract

With the popularity of social networks, users can communicate with each other in a more convenient way. However, the increasing amount of data poses new challenges for the analysis of the social activities of the users. In this article, we propose to visualize the heterogeneous information of user interactions in a social network in a three-dimensional way using the concept of solar systems. The target user represents the center of the solar system. We determine physical variables from interaction frequencies between the user and their friends to be visualized. This gives the user a better insight about their relationships on Facebook. To show the interactions between the user and their friends, we choose four variables related to the solar system: the size of each planet, its angular velocity, and the semi-major and the semi-minor axis of its elliptical orbit. Our system measures the interaction frequencies between a user and their friends based on a linear model. Its coefficients are derived from an online survey we performed. The experimental results indicate that the accuracy of our estimated interaction frequency is better than the accuracies for each individual interaction feature. The average accuracy improvement is 15.93%.

Acknowledgements

I finally completed my thesis, after I experienced numerous challenges and difficulties, which gave me a significant meaningful time. Therefore, I am really grateful for all people in my life during this time, especially, I indeed would like to give my most truthfully appreciation, sincerest gratitude and highest respect to my supervisor Prof. Abdulmotaleb El Saddik for his continuous guidance, support, encourage to my work.

Special and sincere thanks go to Dr. Benjamin Guthier for the invaluable assistance, guidance, and feedback he provided throughout my research, as well as his revision of this thesis, and Dr. Shiai Zhu for the support with the data processing in the experiment and machine learning.

I would also like to thank all colleagues in DISCOVER and MCRLab and all my friends for their contributions throughout the research and their help on my life.

Finally, I am grateful to my family. Their inspiration, understanding and support made this thesis possible. This work is dedicated to them.

Table of Contents

List of Tables

List of Figures

Chapter 1

Motivation and Introduction

1.1 Motivation

Recently, big data visualizations have attracted a surge of interest. According to a report from Facebook on their website, there are hundreds of thousands of users logging in Facebook daily, and every moment or second there will be many activities such as status updates, posting of photos, giving likes or comments or chatting with friends. This means that there are large amounts of data being produced. The study in [2] has pointed out that the behaviors of users on social networks are related to their personalities. However, there are other factors that influence users' behaviors on social networks, such as the relationships between the user and their friends. Therefore, we have to admit that these enormous amounts of data are very valuable for humans to investigate and analyze. So far, the method named data mining is widely used but it has not reached its full potential and thus in this thesis, we will push the boundaries of this method to obtain more and better results.

In addition, even if we pay attention to this data, it needs to be represented in an appropriate form. We have had a lot of ways to show our information in the past; we wrote it on paper, we had pictures and we created websites. Later, we found out that we could analyze data in 2 dimensional visualizations, including tables, diagrams, pies, etc. However, presently, our data is far larger than we can express; even though we can arrange and analyze it and only take the useful parts to express with the methods mentioned above, it is still a challenge. This is because the results of all this data is too messy for human vision to distinguish, and the structure of the visualization is not clear.

1.1.1 Choice of 3D visualization

A research study [3] says that half the adults and three-quarters of the teenagers in America use social networking sites(SNS) and Facebook is by far the most popular of these sites. The survey also states that users on Facebook can possibly reach an average of over 150,000 users through their Facebook friends. It also shows that there can be more close friendships or other relationships associated with tagging Facebook friends in photos. We can learn that these interactive data coming from Facebook are very huge but are promising. If we can choose Facebook as the social network where we collect data from, the promising value in Facebook can be explored, which will have a significant effect on the life of human.

There was another problem which came to us, which was why we visualized our results in a 3D visualization rather than a 2D visualization. Scientists can make sense out of intellectually large data collections by using scientific visualization, which can be achieved by using visualization and animation to stimulate cognitive recognition of patterns in the data while extracting information from the human perceptual system. In the social network world, there is a similarity situation existing, which is we have to deal with these exponentially growing data every second. But it is more difficult to access and manage information in a large data space, and the most difficult part is to visualize all the relationships and structures clearly in a certain room. So 3D visualization gives a potentially better solution for this problem compared with 2D visualization. Giving an simple example, if we present our information in two-dimensional visualization with a pie chart, then the area in which we have to present data is the area of a circle with radius r and the result is πr^2. However, if we choose three-dimensions to visualize our data, then it will be represented by using the surface area of the sphere with radius r, and the surface area is $4\pi r^2$ that is far larger than the area of a circle with the same radius. With people's need for information growing, the data for visualization is larger, and the result will be better. It explains why 3D visualization instead of 2D visualization.

1.1.2 Choice of a solar system representation

Facebook is one of the most popular social networks with approximately 1.3 billion active users as of June 2015 [4]. In a survey [5], we can see the amount of information posted on Facebook on an average day as follows:

- 15% Facebook users update their own status
- 22% comment on another's post or status
- 20% comment on another user's photos
- 26% " like " another user's content
- 10% send another user a private message

It is apparent that there are large amounts of data stored on the Facebook servers because there are large number of users with enormous data of interactions. It is a difficult challenge to visualize this data in either a two-dimensional or a three dimensional representation, because it is hard for the human visual system to process such tremendous amounts of data, especially complex relationships. According to the survey [5], the average Facebook user has 229 Facebook friends. They reported that their friend lists contain:

- 22% people from high school
- 12% extended family
- 10% coworkers
- 9% college friends
- 8% immediate family

- 7% people from voluntary groups
- 2% neighbors

It says that there are over 31% of Facebook friends who cannot be classified into these categories. Furthermore, 7% of Facebook friends have never been met by the user in person, and 3% of the users friends on Facebook have met each other only one time. The remainder of the friends are friends of friends and social ties that are not currently active, but these "interaction ties" may become a significant source of information in the future. In the same point, it is difficult to visualize relationships with a large amount of data. However, we can see a resemblance of this situation with the universe since it also holds a large number of stars and planets and therefore is a good model for representing this situation via multi-dimensional physical variables.

We thus borrow the concept of a solar system from the universe to structure the large amount of data from a social network so that we can help users understand their own situation in the network. The model gives the user on Facebook a visualization of their relationship ties.

1.1.3 Coefficients of a linear model

In order to build the linear model, we did an online survey to get the coefficients. In a study in [6], the authors measured how an user's activity on Facebook relates to their personality by the standard Five Factor Model, and they also tested correlations within the users personality and the properties of their Facebook account. The Five Factor Model [7] [8] [9] is currently a very popular and accepted model of personality, whose ability to predict human behavior is well studied. This model was examined in [8] [10] and was shown to subsume most known personality traits. It was thus claimed to represent the "basic structure" underlying human personality. These research studies lead us to understand that the different behaviors of the users on Facebook depend on their personalities, which means that the opposite also holds ground. Based on that finding, we proposed the online survey to estimate the weights for these features on Facebook according to different personalities. There are already similar studies to the online survey made, such as [11] that added more objective criterias such as measurements of the user upload information on Facebook, and [2] that uses self-reports to evaluate the connection between a user's personality and the user's behaviors on Facebook. The results of the study [11] made progress on finding a stronger connection between the user's personality and their behaviors but both studies are relevant to our thesis.

1.2 Objectives

In this thesis, to solve the problems mentioned above, we propose a measurement of the interaction frequency to represent the tie strength between users and their friends on Facebook, based on the linear model [12]. This model will be helpful for users to make decisions on their relationships, once we mine all these data extracted from Facebook. The interaction data collected from Facebook during a fixed time span can demonstrate changes in relationships based on different periods of someone's life, like when someone is busy with his or her work, or with children or family.

The aim of the entire system we proposed is to help the user learn more about their relationships with their friends. In detail, we divided our work into three parts. The first part describes what kind of tools we used and what social networks we chose to collect data from. It also includes how we used the tools to extract the data about online interaction. Because we need to obtain the interactive data between the user and the user's friends, it is necessary to get permission from the user; this also guarantees the security of the user's personal information for the purpose of this thesis. In the second part, the first task is to determine which kind of variables we need in order to build a linear model so that we can measure the tie strength of the interaction frequency between the user and the user's friends. In this thesis, we determined a total of seven features for the linear model, and the seven features were decided through a questionnaire we performed online. We also did an online survey to determine the coefficients of each feature of the linear model. Another task is to normalize each feature by using a statistical method, before drawing the linear model. We then define the analyzed data as four objects in Section 4 for visualization. For the 3D visualization part of the user's interaction on a social network, a system is built by the game engine "Unity3D" and we borrowed the concept of a solar system in order to create a simulation based on the interactions of the user on Facebook. The determined data of the four objects that are explained in detail in the Subsection 4.1 are represented as the semi-major and semi-minor axis of the elliptical orbit, the size of the planet and the speed of the planet in the 3D visualization, respectively. The structure of a solar system is similar to the structure of relationships between people and their friends in a social network, thus why we chose this concept. Here, the tie strength between friends represents the interaction frequency. The solar system also allows for interesting visual effects that are appealing to the viewer.

1.3 Contributions

The thesis makes the following contributions:

- We analyzed seven-dimensional heterogeneous input data from the social network and proposed a linear model to calculate a measure of interaction frequency between friends. According to [12], the interaction frequency between friends is approximately linear. The weight of the coefficients for the linear model are determined in an online survey we conducted about relationships in Facebook.
- Modeling of the relationships on a social network to the relationships in a solar system.
- Design and development of a 3D visualization system. In detail, our visualization illustrates the tie strength between a target user and their friends as physical variables in a solar system. Four physical variables are matched with features to create the three-dimensional visualization.

1.4 Publication

In the process of completing this work, the following publications have been submitted, accepted or published:

- Conference Paper: Wanjun Pei, Benjamin Guthier, Abdulmotableb El Saddik, "3D Visualization of Heterogeneous User Interactions in a Social Network," Proceedings of the 2015 ASPCASE conference on Social Network.

1.5 Thesis Structure

In terms of the rest of this thesis, it is structured as follows. The second chapter of this thesis presents the related work and background of the thesis. Then, chapter three indicates the methodology of the data collection and building the linear model. The implementation of the visualization system is illustrated in the fourth chapter called 3D Visualization as a Solar System. The fifth chapter explains the process and results of the survey we mentioned. Finally, the conclusion and possible future improvement is written in the sixth chapter.

Chapter 2

Related Work and Background

2.1 Related Work

2.1.1 The research on social network analysis

Many related works exist on mining and analyzing social networks. As we know, unprecedented amounts of social data are being generated with the universal use of social media. Social networks conveniently offer an accessible platform for users to interact with each other and share their information. Mining social networks and building them into interactive visual graphics can be beneficial for businesses, users, and consumers. The data coming from social networks are enormous, noisy, unstructured, and dynamic in nature, thus creating challenges. In the study [13], which gives a brief introduction to the field of basic data mining and social media, recurrent research problems in mining data from social medias are explained, using examples to illustrate the application of data mining to social media, these projects of mining social media are then explained for the purpose of humanitarian assistance and disaster relief for real-world applications, which also gives a reason for using data mining instead of using unstructured, noisy and natural data.

The research in [14] mainly describes the processes of mining data in the context of the analysis of the social network Facebook. In detail, the study divided this work of analyzing social networks into two phases. For the first phase, a subgraph made up of high-degree nodes (i.e. users having a large number of friends) was extracted from a Facebook social graph. Secondly, the attributes of these high-degree nodes were analyzed using the social network analysis tool named GEPHI at the 2013 5th International Conference on Computational Intelligence and Communication Networks.

The study also mentioned that they started their work with a dataset collected in April of 2009 through data scraping from Facebook [15] [16].
This dataset consisted of two files

- mhrw-socialgraph-anonymized (1.4GB)
- mhrw-nodeproperties-anonymized (17MB)

MHRW - A sample of 957K unique users obtained Facebook-wide by 28 independent Metropolis-Hastings random walks. For the first one, it included each sampled userID, the number of times the user was sampled and the userIDs of friends.

$$< uid >< \#timessampled >< friend\_uid\_1 >< friend\_uid\_2 > \ldots < friend\_uid\_j >$$

The second one contained extra node properties for each sampled user. For each sampled userID we had the number of times sampled, the total number of friends, the privacy settings and network membership.

$$< uid >< \#timessampled >< \#totalfriends >< privacysettings >< networkID(s) >$$

The subgraph was obtained from this dataset through file handling techniques: First of all, the total number of friends, corresponding to each user id was observed from the second file. Accordingly, a threshold of 920 was set up on the parameter "total friends" and the user ids having more friends than this limiting value were identified. These high-degree nodes (ids) were the primary concern of our study.

Once the high-degree nodes were identified, the first file which is actually the Facebook social graph (consisting of user ids and ids of their friends) was truncated to retain the high degree nodes obtained above and all other nodes(ids) were removed. So, the resulting subgraph was an undirected graph (because friendship is undirected) consisting of only high-degree nodes. The attributes of these high-degree nodes and the relationship between them were analyzed using an SNA tool called GEPHI.

However, instead of simply detecting binary relationship ties between people, the studies [17] [12]give another way to collect data and analyze a user's data. They extracted data from a social network and conducted these data into kernels; then they analyzed the strength of each tie based on those kernels that they had built before. We introduced the detail as below according to the two studies, respectively.

In the study [17], it investigates the challenging problem of Social Strength Modeling (SSM) of users in the social media communities. Especially, because Flickr has rich user-generated content, and is one of the most popular online photo-sharing sites, they chose it as the social media platform for their study. The platform includes user-annotated tags, shared photos, comments, etc., and they can obtain certain useful data for each Flickr user from their contact list of other users formed by friendship, and creating and joining hobby groups where users share photos of common interests and make comments to each other. Besides the explicit mutual linkage between users, the uploaded photos and their associated metadata (e.g., tags, comments, etc.) can also be leveraged to infer the implicit relationship between users. The multimodal information available on Flickr poses opportunities yet challenges for the research on SSM. Figure 2.1 shows the proposed about learning how to rank framework with kernel-based .

There are total four parts in the figure 2.1. The first part introduce the information connected with users, then for second part they built three graphs(only three kinds of graphs presented for illustration). The third part is that they learn the weight called θ by maximally aligning the combination of textual and visual graphs to the friend graph at first, and then they try to learn to rank framework with logistic loss to estimate the tie of social strength. After they learned social strength by last part, they can make a wide variety of applications in the fourth part. [17]

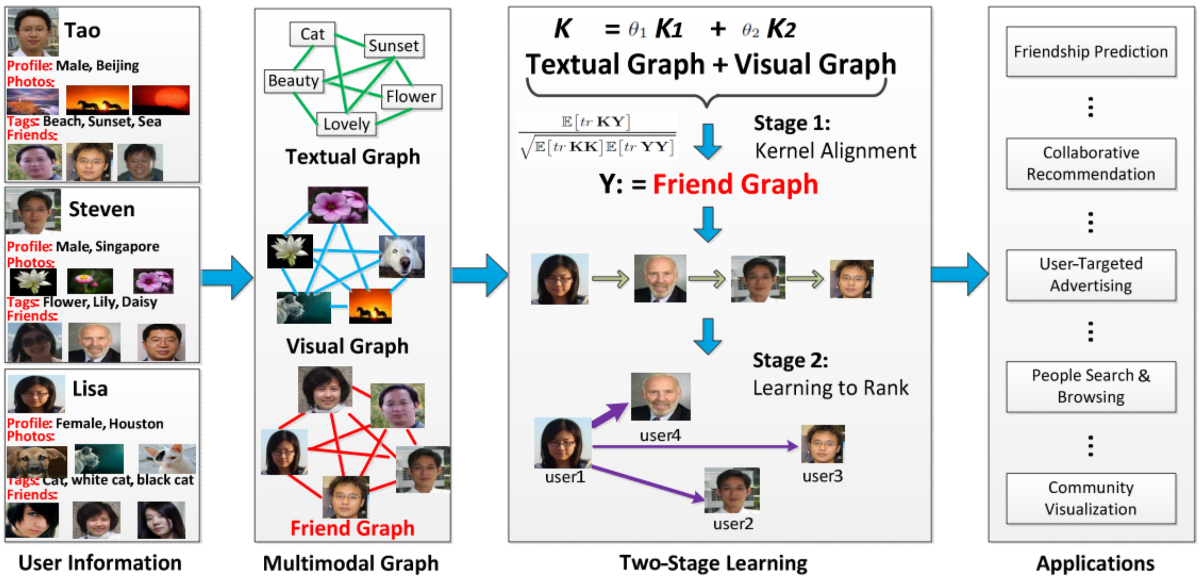


Figure 2.1: Learning to rank framework with kernel-based.

Additionally, the explanation of the algorithm-learning kernels in the form of linear combinations can be found in [12]. In this study, to estimate the continuous tie strength between users in order to recommend friends, the authors constructed a representative model and the heterogeneous data is collected from multiple social media communities. They classify these multi-model data into two classes: similarity data such as common friends, groups, tags, geo and visual or interaction data such as comments, and marking favorite photos.

Then instead of using the conventional symmetrical relationships in the interaction data for tie strength estimation, they propose to use asymmetrical relationships. Furthermore, they found that the social connections between users functions as a linear equation by exploring the behavior of users in a social media community.

2.1.2 3D Visualization

The need for making the results of data mining useful to people has inspired other representative studies in the field of 3D visualization. There is a study on 3D visualization of hierarchical information based on the concept of cone trees [18](Plate one to Plate five), which is shown in Figure 2.2. The aim of the 3D visualization was to use the emerging technology of 3D visualization and interactive animation to provide potential solutions to the problems of managing and accessing large information. Therefore, they proposed the method of cone trees to present the information. In particular, the 3D visualization can maximize effective use of available screen space and enable visualization of the entire structure. Another example is in [19], where the goal of the study is to decrease clutter, and produce cleaner network visualization.

After that, many research studies on 3D visualization are geared towards biological or medical fields. The work in [20] presented a new visualization tool named Arena3D, which shows a new concept of staggered layers in 3D space. The data can be visualized in the tool, such as proteins, chemicals, or pathways, which are classified onto separate layers and organized via layout algorithms, such as Fruchterman-Reingold, distance geometry, and a novel hierarchical layout.

For each layer, the data can be clustered via k-means, affinity propagation, Markov clustering, neighbor joining, tree clustering, or UPGMA ('unweighted pair-group method with arithmetic mean'). Each node can be defined by a simple input format with a name and an URL, and connections or similarity scores can be determined between pairs of nodes. Arena3D is illustrated in Figure 2.3 with datasets related to Huntington's disease.

a It introduces the results of a query starting from Huntington's Disease (HD). HD is related to nine associated genes which are linked to 10 proteins, and the Huntington gene 'htt' which shows two forms: mutant and wild-type. These proteins link to 75 protein structures.

b It shows nine polyQ-related diseases (top layer). On the middle layer, 66 proteins

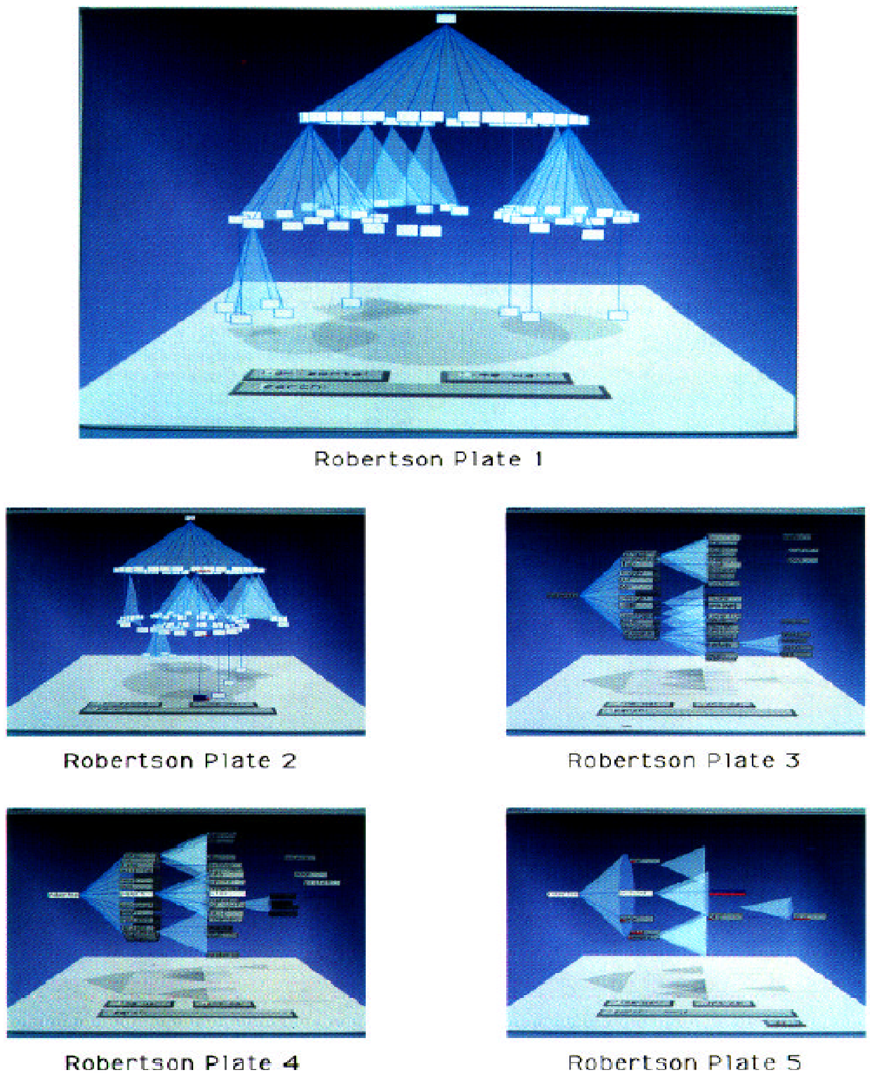


Figure 2.2: The 3D visualization basic on cone trees

known to be associated to these diseases were clustered, and on the bottom layer, 151 domains associated with these 66 proteins are shown. On the middle layer, 6 proteins that are involved in both Huntington and another polyQ disease are highlighted. On the bottom layer the 8 domains present in these six proteins are highlighted. WW and atrophin domains are connected with proteins related to different diseases.

c It shows the proteins related to Huntington (top, red) and their connection to the GO ontology hierarchy.

As we introduced above, we can see that 3D visualization is an emerging type of research, which can give more space to better manage and access data, and also its structure gives more visual reality for humans. The data associated with interactions is growing exponentially as the number of users in social networks increases. It can have significant benefits for us to visualize the relationships or the strength of ties between users on social networks, once the large amounts of data are analyzed. Compared to humans, the earth is infinitely bigger. However, so many planets bigger than the earth can be gathered in a solar system. We can therefore learn about organizing our relationships based on the visualization of a solar system. There is an application which attempts to visualize the universe [21], and in the application, we capture two figures as shown in Figure 2.4 and Figure 2.5. It visualizes the solar system and the system of Jupiter in the universe in two dimensions.

2.1.3 Facebook API

For extracting information from Facebook, there are a lot of APIs provided by Facebook in order to develop the third party support for Facebook applications by developers using GraphAPIs, or the JavaScript SDK. RestFB is one of the most popular APIs, which is a simple and flexible Facebook Graph API, and uses the Old REST API, a client written in Java. It is open source software released under the terms of the MIT License.

The API uses the FQLquery language that is the official language provided by Facebook

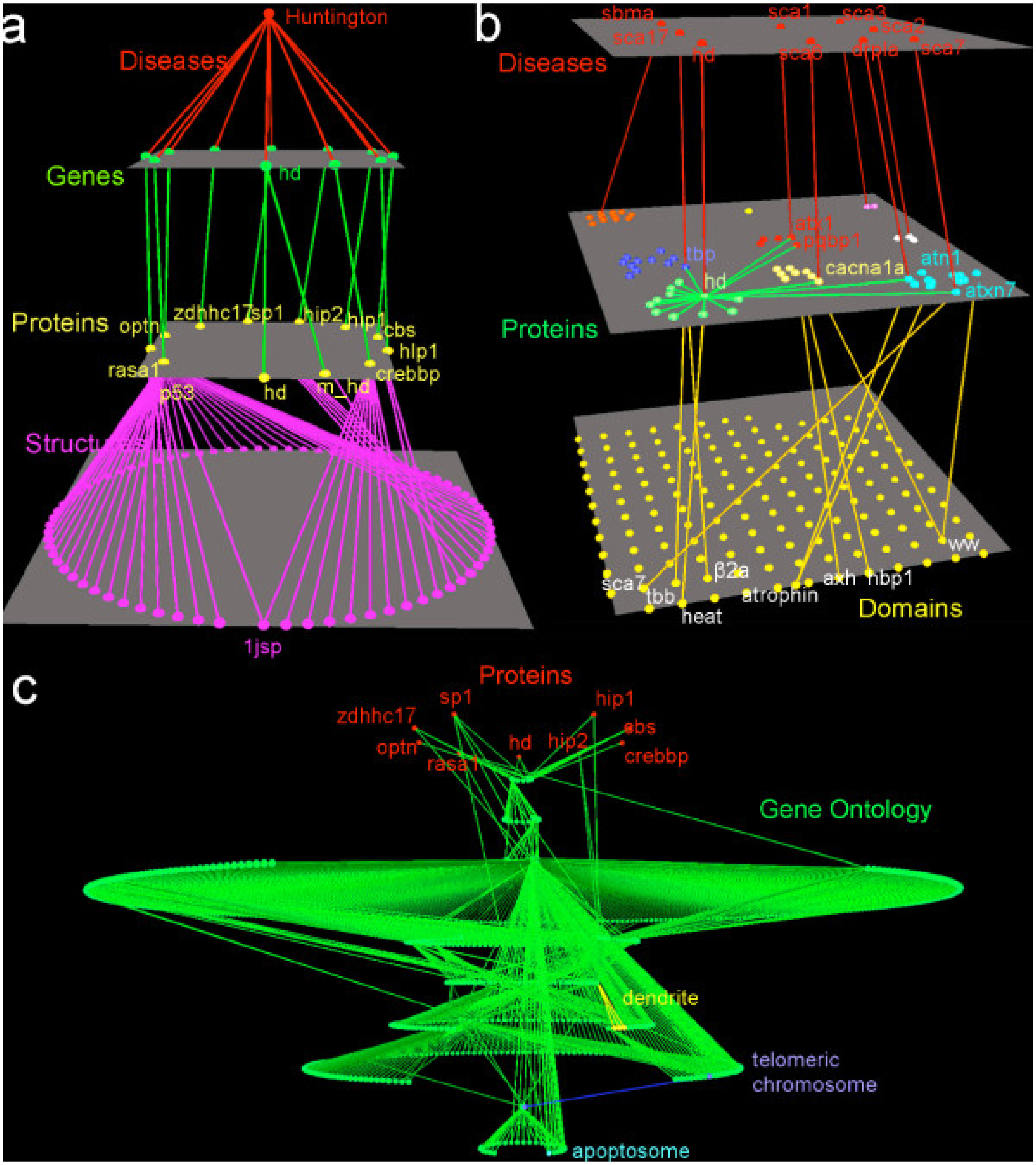


Figure 2.3: Screenshots of Arena3D showing data related to Huntington's disease.

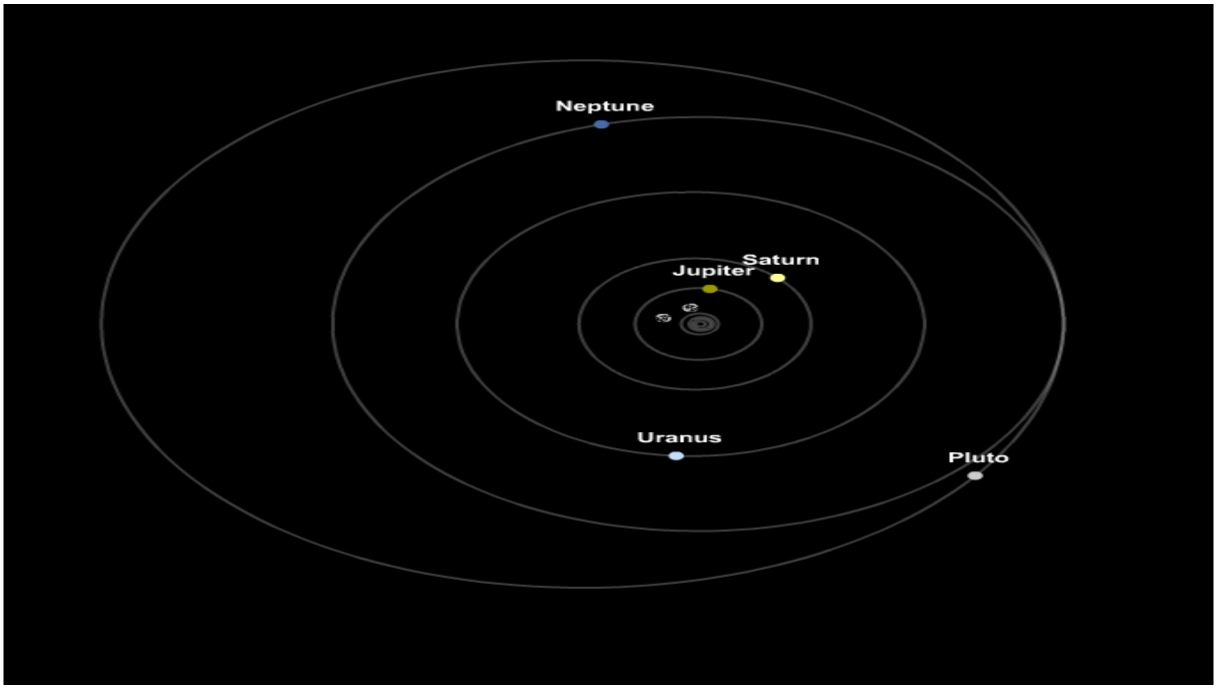


Figure 2.4: The visualization of solar system

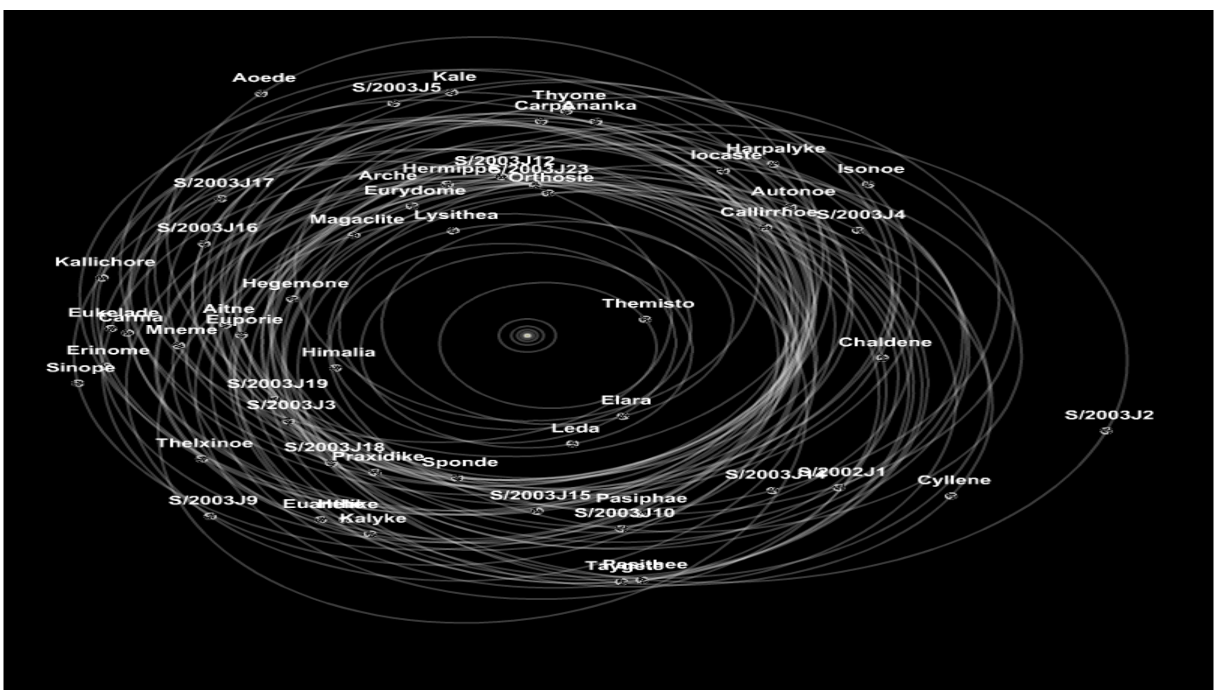


Figure 2.5: The visualization of Jupiter

for developing third party applications. For querying purposes and for fetching data it is being used in the development of a personalized recommender System [22].

Some of them:

- Fetching Information About Post
- Fetching information about Messages
- Fetching information about Photo
- Fetching information friend movies

There is one study [23] showing the process of getting access token and authentication for the application. It introduced a process for the whole structure; the first step is registering the Facebook developer, then constructing a new application and getting verification. Next, it should set up the application, and get APP_ID and APP_secret. Then the study in [24] describes the procedure for obtaining the data from the user on Facebook in detail.

2.1.4 Unity3D

By turning planes into a 3D visualization, it allows a user friendly interaction as seen with animations and games. Among the engine development platforms, Unity 3D has gradually occupied an important position. In online games such as web games and mobile games. Unity 3D is a program with a cross-platform scripting language, supports a variety of plug-ins and has an official forum with many active discussions. It has therefore become the first choice for many developers. [25] The edition 4.2 of the program was published on more than ten different platforms, such as Mac, iOS, Android, Flash, Xbox360, PS3, Wii and so on. In th recent edition 5, more state machine behaviors were added, for better control of script functions.

In the study named "A new method of virtual reality based on Unity3D" [26], it is mentioned that Unity3D has a highly optimized graphics pipeline for both DirectX and

OpenGL. Animated meshes, particle systems, advanced lighting and shadows, all run blazingly fast. It also offers individual operations to create rain, sparks, dust trails, and anything that can be imagined. For the purpose of our work, the following is needed:

- Programming

  To build specific or simple functions of scenes, the already included editor of Unity 3D can be used. Since Visual Studio [27] is currently the most powerful editor, to create more complex scenes, Visual Studio is a better choice, but require the user to download the add-on UnityEngine.dll.

- Shading

  There are 40 shaders ranging from the simple (Diffuse, Glossy, etc.) to the more advanced (Self Illuminated, Bumped Specular, etc.)for the Unity 3D engine. All the built-in shaders integrate perfectly with any type of light, with cookies or without. The developer can also write his/her own shaders in Unity3D's ShaderLab language with Cg and GLSL, if the off-the-shelf shaders cannot satisfy the needs. In the study [26], besides the ordinary shaders, and to meet more difficult requirements, the authors write their own shaders, which lead to more realistic scenario results.

- Showing Information by GUI Control

  As for the geographic entities, the Unity3D not only has attribute information but also spatial information. Developers can use GUI functions to introduce the attribute information get real-time effects and dynamics. Furthermore, the database-MySQL data management system can be used to store all this attribute information.

- Collision Detection

  Another good feature is the full capability of the Ageia PhysX next-gen Physics Engine for Unity3D, and games like Unreal Tournament 2007 and Ghost Recon 3 are both using the engine. Unity3D supports full Rigidbody physics. Rigidbodies is usually used when objects are acting under forces and colliding, but it is also used in joints, with no scripting required. The virtual reality built by Unity3D is more true to daily life. Users can click the mouse to determine any display area; after that, pressing keys on the keyboard will choose where to go by executing scripts on a certain GameObject. During the process of walking around, the collision detection can be handled by the navigation of the person. Then if the navigation person runs into buildings, he automatically returns to the road. The collision detection is to ensure that the route of the road is not affected by the nonlinear structure.

- Integrating
  There is an asset in a Unity3D Project for importing and saving all works we have completed, automatically and immediately. Those works include scenes, scripts, audios, attribute data, pictures, textures etc. Even if users are visiting inside the editor, assets can update the works at any moment. On the other hand, Unity Asset Server can help a developer to optimize the calculations when the project is big.

- Publishing
  Unity 3D provides a Unity Web Player Plug-in (about 3 MB) to help the developer to publish their applications online, and the Unity Web Player Plug-in can be downloaded to a website. Users can browse the application by downloading the small Unity Web Player Plug-in. It auto-installs without a browser restart, and already has an 8-digit distribution. All modern browsers including Internet Explorer, Firefox, Safari, and most Mozilla-based browsers accept the Plug-in.

- Others
  The developers can add sounds, animations or videos in Unity3D by using GameObjects. Thus it can truly enhance the user's immersion in the virtual reality system [28].

2.2 Background

2.2.1 Solar system

As we know, we live on a planet, in the solar system, which is called Earth. The solar system consists of the earth and many other planets. In the study we found in [1], it says that the Solar System [1] comprises the Sun and the objects that orbit it, either directly or indirectly.[2] Of those objects that orbit the Sun directly, the largest eight are the planets[3] as shown in Figure 2.6 that form the planetary system around it, while the remainder are significantly smaller objects, such as dwarf planets and small Solar System bodies (SSSBs) such as comets and asteroids.

The Solar System formed 4.6 billion years ago from the gravitational collapse of a giant interstellar molecular cloud. The vast majority of the system's mass is in the Sun, with most of the remaining mass contained in Jupiter. The four smaller inner planets, Mercury, Venus, Earth and Mars, also called the terrestrial planets, are primarily composed of rock and metal. The four outer planets, the giant planets, are substantially more massive than the terrestrials. The two largest, the gas giants Jupiter and Saturn, are composed mainly of hydrogen and helium; the two outermost planets, the ice giants Uranus and Neptune, are composed largely of substances with relatively high melting points compared with hydrogen and helium, called ices, such as water, ammonia and methane. All planets have almost circular orbits that lie within a nearly flat disc called the ecliptic.

The Solar System also contains smaller objects. The asteroid belt, which lies between Mars and Jupiter, mostly contains objects composed, like the terrestrial planets, of rock and metal. Beyond Neptune's orbit lies the Kuiper belt and scattered disc, populations of trans-Neptunian objects composed mostly of ices, and beyond them a newly discovered population of sednoids. Within these populations are several dozen to possibly tens of thousands of objects large enough to have been rounded by their own gravity. [29] Such objects are categorized as dwarf planets. Identified dwarf planets include the asteroid Ceres

[1] Capitalization of the name varies. The IAU, the authoritative body regarding astronomical nomenclature, specifies capitalizing the names of all individual astronomical objects, but uses mixed "Solar System" and "solar system" in their naming guidelines document. The name is commonly rendered in lower case ("solar system"), as, for example, in the Oxford English Dictionary and Merriam-Webster's 11th Collegiate Dictionary.

[2] The moons orbiting the Solar System's planets are an example of the latter.

[3] Historically, several other bodies were once considered planets, including, from its discovery in 1930 until 2006, Pluto. See Former planets.

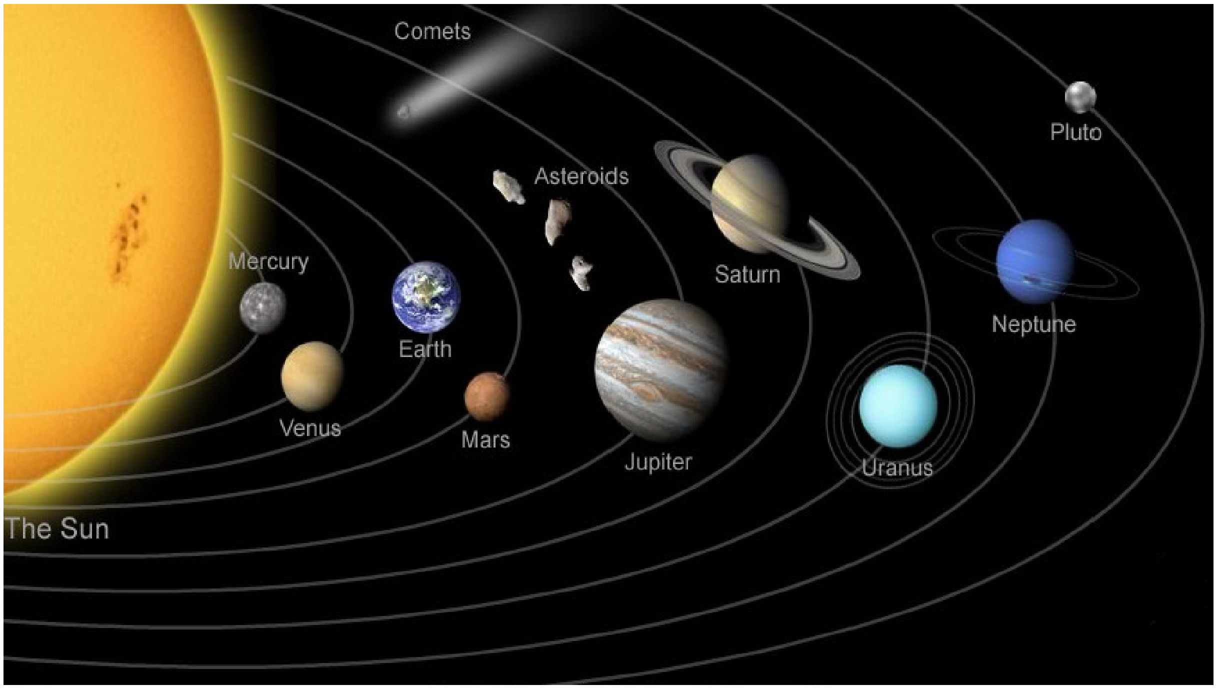


Figure 2.6: The main members of the solar system [1]

and the trans-Neptunian objects Pluto and Eris.[4] In addition to these two regions, various other small-body populations, including comets, centaurs and interplanetary dust, freely travel between regions. Six of the planets, at least three of the dwarf planets, and many of the smaller bodies are orbited by natural satellites,[5] usually termed "moons" after Earth's Moon. Each of the outer planets is encircled by planetary rings of dust and other small objects.

[4] Pluto does not fit the definition. [30] Instead, Pluto is considered to be a dwarf planet, a body orbiting the Sun that is massive enough to be made near-spherical by its own gravity but which has not cleared planetesimals from its neighbourhood and is also not a satellite. [30] In addition to Pluto, the IAU has recognized four other dwarf planets in the Solar System: Ceres, Haumea, Makemake, and Eris. [31] Other objects commonly (but not officially) treated as dwarf planets include 2007 OR10, Sedna, Orcus, and Quaoar. [32] In a reference to Pluto, other dwarf planets orbiting in the trans-Neptunian region are sometimes called "plutoids". [33]

[5] See List of natural satellites of the Solar System for the full list of natural satellites of the eight planets and first five dwarf planets.

2.2.2 Gravitation

The original law to explain the reason for the motion of the planets in the universe is Newton's law of universal gravitation. The law indicates that any two objects in the universe attract each other because of a force, which is directly proportional with the product of their mass and inversely proportional with the square of the distance between the two objects.[6] The general physical law is derived from empirical observations by what Issac Newton called introduction [34].

In modern language, it can be explained as "the law states: Every point mass attracts every single other point mass by a force pointing along the line intersecting both points. The force is proportional to the product of the two masses and inversely proportional to the square of the distance between them". [35] "The first test of Newton's theory of gravitation between masses in the laboratory was the Cavendish experiment conducted by the British scientist Henry Cavendish in 1798". [36]

As we know, the law of gravitation is similar to Coulomb's law of electrical forces, which adapts to calculate the magnitude of electrical force arising between two charged bodies. For both of them, they obey the inverse-square laws, where force is inversely proportional to the square of the distance between the bodies. Coulomb's law also adapts to be directly proportional to the product of two charges' masses as well as the electrostatic constant in place of the gravitational constant.

The Modern Form is as follows and the Figure shows the variables.

$$F = G\frac{m_1 m_2}{r^2} \tag{2.1}$$

[6] It was shown separately that large, spherically symmetrical masses attract and are attracted as if all their mass were concentrated at their centers.

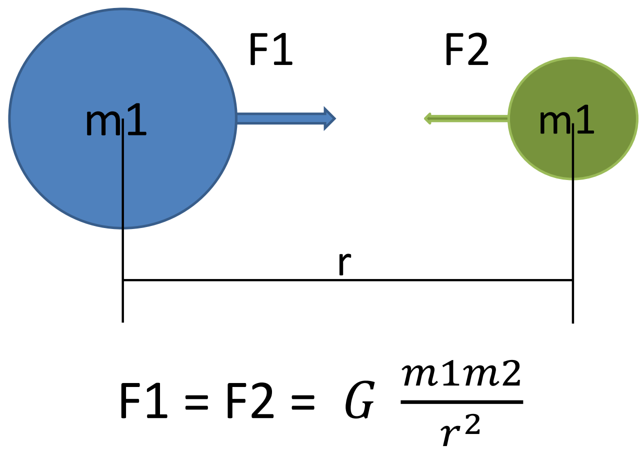


Figure 2.7: The modern form of gravitation

Where: F is the force between the masses[7]; m1 is the first mass; m2 is the second mass; r is the distance between the centers of the masses; G is the gravitational constant(6.673 $\times 10^{-11} N \cdot (m/kg)^2$);

2.2.3 Kepler's law

Before we start to introduce Kepler's Law, it is necessary to introduce Newton's laws. The Newton's laws of motion also give the explanation and investigate the motion of many physical objects and systems including the solar system. [38] The laws include three physical laws that together laid the foundation for classical mechanics. The relationship between a body and the forces moving upon it, and the activity in response to said forces

[7] Assuming SI units, F is measured in newtons (N), m1 and m2 in kilograms (kg), r in meters (m), and the constant G is approximately equal to 6.673 $\times 10^{-11} N \cdot (m/kg)^2$. [37] The value of the constant G was first accurately determined from the results of the Cavendish experiment conducted by the British scientist Henry Cavendish in 1798, although Cavendish did not himself calculate a numerical value for G [36]

is explained by the laws. In the past nearly three centuries, the laws have been described in different ways.[8]

In summary, they can be written as below [39].

- First law: When viewed in an inertial reference frame, an object either remains at rest or continues to move at a constant velocity, unless acted upon by an external force. [40] [41]
- Second law: The vector sum of the forces F on an object is equal to the mass m of that object multiplied by the acceleration vector a of the object: F = ma.
- Third law: When one body exerts a force on a second body, the second body simultaneously exerts a force equal in magnitude and opposite in direction on the first body.

The reason for introducing the laws is that the laws are used to explain Kepler's laws of planetary motion. Then it can be easier for us to understand Kepler's laws. Because Kepler's laws are used to introduce the solar system in detail, Newton's laws can explain more broadly the motion of objects as a result of the gravitational attraction of several particles formed.

Kepler's laws were formulated by Kepler, and mainly concern planetary motion. In 1609 in the journal Science called "New Astronomy", he published two laws on planetary motion, and in 1618, he discovered the third law [42]. The laws include three scientific laws and they describe the motion of the planets around the sun.

In summary, the three laws and Figure 2.8, Figure 2.9, Figure 2.10are shown below.

[8] For explanations of Newton's laws of motion by Newton in the early 18th century, by the physicist William Thomson (Lord Kelvin) in the mid-19th century, and by a modern text of the early 21st century, see:- Newton's "Axioms or Laws of Motion" starting on page 19 of volume 1 of the 1729 translation of the "Principia"; Section 242, Newton's laws of motion in Thomson, W (Lord Kelvin), and Tait, P G, (1867), Treatise on natural philosophy, volume 1; and Benjamin Crowell (2000), Newtonian Physics.

- Kepler's first law: Kepler's first law is also known as the elliptical law or orbital law. Each planet travels along an elliptical orbit around the sun, and the sun is at one focus of the ellipse shown in Figure 2.8.

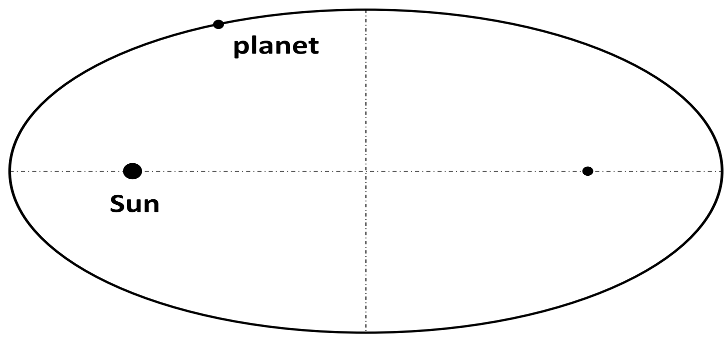


Figure 2.8: Kepler's first law

- Kepler's second law: Kepler's second law, is also known as the equal-area law. A line segment joining a planet and the Sun sweeps out equal areas during equal intervals of time. This law reveals that the angular momentum of the planets surrounding the sun is constant expressed by the formula 2.2.

$$S_{A,B} = S_{C,D} = S_{E,K} \tag{2.2}$$

- Kepler's third law: Kepler's third law, is also known as the law of the period. The square of the orbital period of a planet is proportional to the cube of the semi-major axis of its orbit. It is not difficult to export: the gravity between the planets and the sun is inversely proportional to the square of the elliptical orbital radius, which is an important basis for the law of gravity of Isaac Newton, expressed by the formula 2.3.

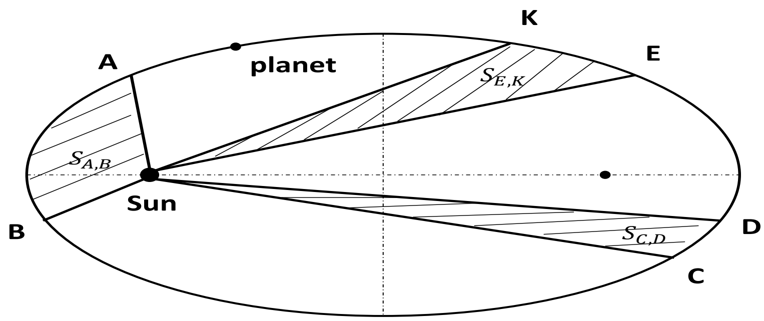


Figure 2.9: Kepler's second law

$$\frac{a^3}{\tau^2} = K \tag{2.3}$$

Where a is a semi-major axis of planetary orbits, τ is a planetary orbital period. K is a constant.

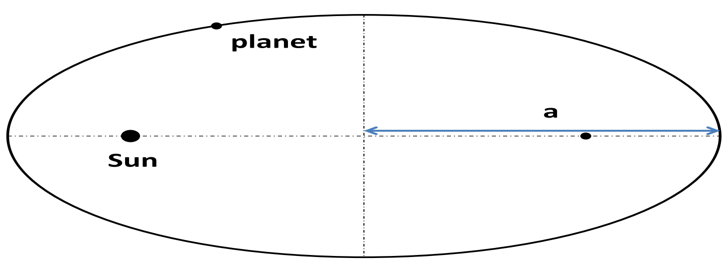


Figure 2.10: Kepler's third law

2.3 Conclusion

In this chapter, we have introduced related works and our background. Related works include other people's studies in 3D visualization, such as 3D visualization based on cone trees in figure 2.2 and Arena3D in figure 2.3. We also explained the theory behind social network analysis, Facebook API and Unity3D. The background subsection introduces the concept of a solar system and laws we will be using, such as Kepler's third law of motion and Newton's law of universal gravitation. in the next chapter, we will need to introduce the methodology of our system.

Chapter 3

Methodology

For the 3D visualization of the solar system, we divided our research into three parts, which was different from the traditional visualization for the social network, in order to show the whole structure of the social network and the relations between users. Our visualization aimed at introducing the interaction frequency of the relationship between the user and the user's friends. The system needed the user's account and password for Facebook to log in to their account, and the process of the system operating is shown as 3.1. To start with, we found an API of Facebook to collect the unstructured data we needed from the website. Secondly, we determined the interaction frequency with a linear model. Finally, we analyzed the result of the model and chose four variables to visualize in order to make unstructured data be structured data, so that the 3D visualization based on a concept of the solar system could be created.

- Data Collection

  This is the first part in that we needed to collect data from the social network using the API of Facebook. To improve the efficiency of the data processing, we designed the algorithm for seven features, according to the questionnaire we did online for calculating the weight of every feature in the linear model as described in Section 3.1.1.

- Determination of Interaction Frequency with User Linear Model

According to the study of relationships on social networks, we organized these unstructured data to be seven kernels which included the number of likes, comments on the user's status, the number of likes and comments for the user's photos, the number of tags from user's friends to user, the number of tags from user to his or her friends, and the number of mutual friends for user and user's friends. After that, we could build a linear model to calculate the interaction frequency that we propose between user and user's friends. This task is described in Section 3.2.

- 3D Visualization with Solar System Concept
  There are four physical variables that can be visualized in the solar system, which are the speed of planets, the size of the planets, the semi major axis, and semi minor axis of the elliptical orbit. We conducted four interaction frequencies from the analyzed data to connect with the physical variables. The whole system regarded the user as the center of planets, which can indicate a clear structure of the relationships and interactions between the user and the user's friends. The visualization is described in chapter four.

3.1 Data Collection

Before we start to collect the data that we need, we compared two tools for getting the data from Facebook, and finally, we decided to use the API called RestFB to obtain a user's data in a certain time that can be decided by us. To begin with, to get the necessary data from the user, we needed to find the access token by browsing the website that is offered by Facebook. The process of extracting the user's access token in detail can be seen in Figure 3.2. The access token is a permission for a third party software to access the data of a user on Facebook. On the other hand, we need to log in to Facebook with the user's Facebook account username and password so that we could have this page of the website shown in Figure 3.2. We then got Figure 3.3 as shown below, which shows the permissions the user should allow to be obtained by the system. After we clicked the button "Get Access Token", we receive the long code combined with letters and symbols. In the next

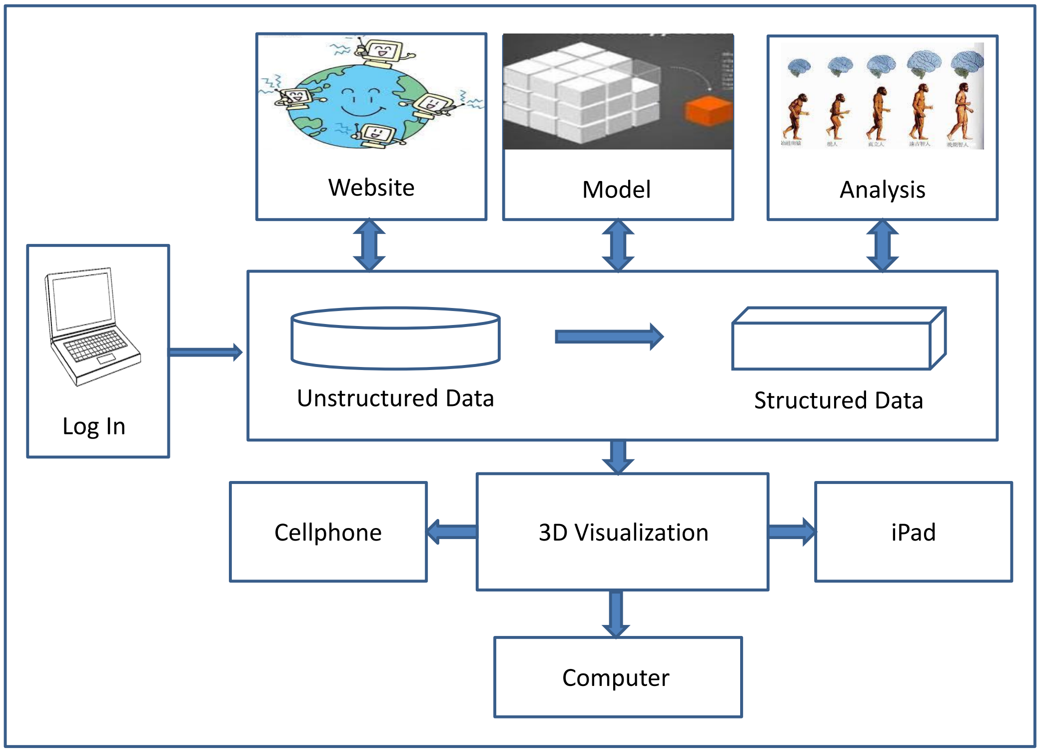


Figure 3.1: Functional structure of the entire system

step, to get all the interaction data, we browse a user's home page with the access token in a certain time, that we defined as six months, but the rank of the time was from the time when the user's account has been activated until now. we then need to decide what kind of data is needed that will be useful to determine the tie strength of interactions.

We did an online questionnaire to help us gather the thoughts people on interactions between them and their friends on Facebook, which is mentioned in detail in the Section 3.1.1. After that, the results of the questionnaire helps us determine what kind of data we should collect. Two types of situation can come from the results of the questionnaire, which come respectively from the user's status and the user's photos. First of all, to improve the efficiency of data processing, we executed the process of extracting and classifying the data at the same time, so that not only do we save time on collecting, but we can also use less processes to save space and resources. Here are the steps we took to gather data from the seven features mentioned in Section 3.1.1. For the first two features, likes and comments

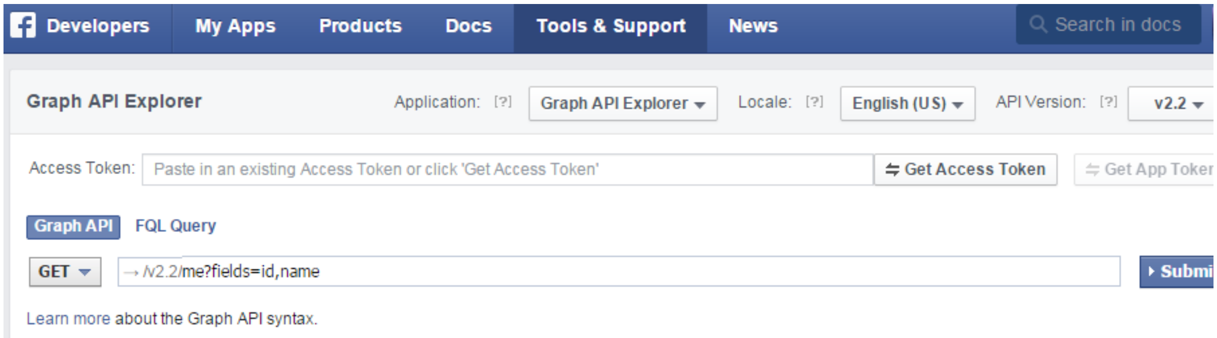


Figure 3.2: The page of getting access token

on the user's statuses, the user's homepage can be visited automatically with the access token that lets us obtain all the user's status during a certain amount of time on Facebook, and every status posted by the user gives links that will show who has ever given likes and comments to the user. Our system can get all the names of those who have given comments or likes by following the links for every user's status. In addition, the number of likes or comments that are given by the user's friends can be recorded by the system. The other features mainly come from the user's photos. To get the information about the number of likes and comments that come from the user's friends, the system links to every photo in the user's albums and it extracts the name of the user's friends who have ever given the user likes or comments on their photos. The number of likes or comments from all of the user's friends who has ever given these interactions can be counted automatically by the system that we built. So far, we can get a total of four arrays of heterogeneous data for analyzing the interaction frequency. As for the tags, every photo has their own ID so we can get the information of every photo and information on the tags the user made and people who have tagged the user. With the help of the ID, those data can be collected by following the ID of the user's photos. A list of the user's friends will be obtained and we will obtain two pieces of information on each person on the list: how many times they have tagged the user in photos and how many times the user has tagged them in photos. Last of all, the seventh feature which is the number of mutual friends between user and each of his or her friend can be counted by following the personal file of the user and each of his or her friends. In detail, we find out how many friends in total does the user have and

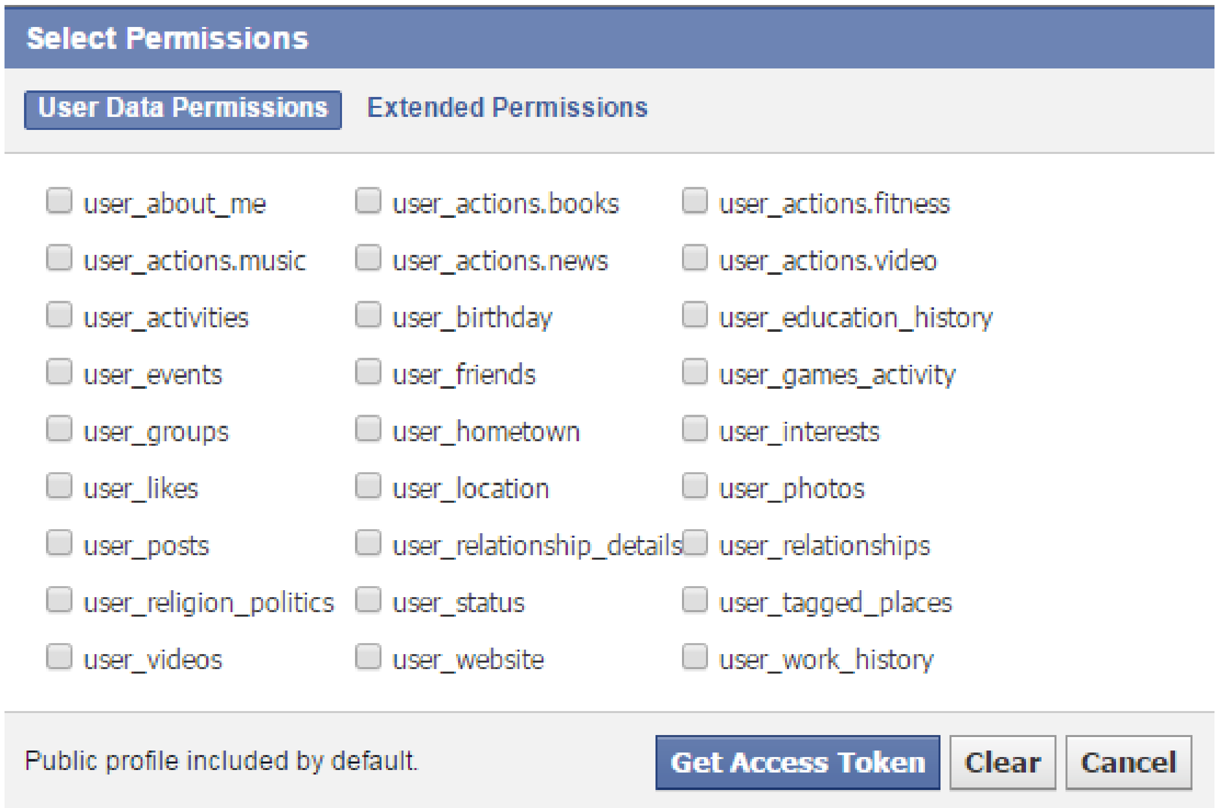


Figure 3.3: The page of getting user's permission

how many friends in total does each of the user's friends have. Then to get the number of mutual friends between the user and each of his or her friends, we obtain that with their ID. The system then collects the number of mutual friends with the help of the Facebook API by using a pair of IDs, the user's and his or her friend's. The API will then return the number of mutual friends they have in common. Every user's friend and user will have a set of three values(the user's total friends, each friend's total friends and the total number of mutual friends between the pair)and the reason for why we used these three values are explained further in detail in Section 3.2. But before we start collecting all these data mentioned above in order to start construction in Chapter 4, another problem we have is deciding on the method of storage of the data and the type of database needed.

To solve the problem mentioned earlier, we compare three kinds of database management software as shown in the Table 3.1, which introduces the advantages and disadvantages of these software programs. Finally, we decide to use the MySQL database as our tool to store data and develop our system by taking into consideration the cost and practicability of each software. The database can be used for connecting to the 3D visualization tool, it is essential for when we will build the 3D visualization system using the concept of solar systems because we need the results of the analyzed data transferred to Unity3D. This means that the MySQL database is used as a connection between the Facebook API and the 3D visualization tool.

3.1.1 Questionnaire

Firstly, we did a questionnaire online to help us find the thoughts on interactions between users and their friends on Facebook. In total, there were 112 users who completed the questionnaire online, which included sixty four males and forty eight females. They gave their thoughts on which interactions that they often like to do with their friends on social networks. The result is shown in Table 3.2. As is shown, there are six features in total that are popular for the user to interact with their friends. They are likes and comments for a user's status, likes and comments for a user's photos, the number of tags in a user's photos that come from the user's friends, and the number of tagging user's friends in user's or friends' photos. And there is also an extra feature called mutual friends between user and user's friends, which can give a factor to illustrate how close is the friend's circle between a user and the user's friends. We can get the result in Table 3.2, which shows that these six features play a significant role in calculating the interaction frequency.

Moreover, We did another small questionnaire. The objectives of this questionnaire is to find out for what purposes the user using the social network Facebook leaves comments on their friend's photos. The table 3.3 indicates that the feature of leaving comments on friends' photos has an influence on the tie strength of interaction frequency between a user and the user's friends. fifteen people have participated in this questionnaire. The questionnaire shows that there are four reasons a user would leave a comment on a friend's photo. They include: liking the photo itself (eight people agree), Supporting their friends (seven people agree), socializing with friends (four people agree) and giving opinions on

MS ACCESS	Advantages	Easy deployment; file using more flexible; can develop database applications on desktop (UI); can be used as front-end development tools with other databases with application development.
	Disadvantages	A small amount of data storage; security is not high; the user-level passwords easy to crack; only support Windows system; poor adaptability to high-intensity operations; access database has some limits; if the data reach 100M, is likely to cause the server iis suspended animation, or consumed cause the server to crash the server's memory.
MS SQL SERVER 2008	Advantages	High security; true client / server architecture; the graphical user interface is more intuitive and simple; a rich programming interface tool for users of the program designed to provide a greater choice; has good scalability; can span multiple platforms; SQL Server also provides data warehousing capabilities; this feature is available only on Oracle and other more expensive DBMS.
	Disadvantages	Operational complexity; only supports Windows operating system.
Oracle	Advantages	Stable; full-featured; powerful performance; advanced technology; the highest level of security certification ISO standards; support for multiple operating systems.
	Disadvantages	Expensive; many tools need to purchase.

MySQL	Advantages	A multi- user multi-tasking document database system; good security; cross multiplatform database management software; support commands and graphical management; fewer system resources, faster, and is open source.
	Disadvantages	Most of the operations are carried out under dos; not easy to use graphical interface.

Table 3.1: The advantages and disadvantages of the three database.

Features	Likes	Comments	Tag friend	Tag user	Likes photos	Comments photos
Often	87.43%	85.505%	78.46%	78.46%	90.895%	92.82%
Rarely	12.57%	14.495%	21.54%	21.54%	9.105%	7.18%

Table 3.2: Average like and rarely like for these interactions.

photos (six people agree).

Reasons	Like the photos	Support friends	To socialize	Opinions of photos
Number of agreement	80%	70%	40%	60%

Table 3.3: The proportion for different reason of giving comments to friends' photos

3.2 Linear Model of Interaction Frequency

In Section 3.1, we have obtained the heterogeneous data needed. The next significant task is the analysis and calculation of these data. The main task is determining a linear model to estimate interaction frequency and this task is divided into three parts. Firstly, we set the seven features(the heterogeneous data collected) as seven kernels so that they can be used as the elements to create the model of interaction frequency. In our scenario, the seven features which are likes and comments for user's status, the likes and comments for the user's photos, the number of tags that the user is giving to his or her friends, the number of tags that user's friends are giving to the user and the number of mutual friends will be combined into a linear model to analyze the tie strength of interactions between the user and his or her friends on Facebook. In terms of the second part, it is to normalize these seven kernels. This is one of the most important tasks, because if the seven kernels can be combined, they should be on a comparable level. Finally, the third part is creating a specific linear model by calculating the weights of the coefficients of the linear model, and these weights can be obtained by doing a survey, which is mentioned in chapter 5.

Overall, for this section, it will introduce the process of constructing the specific linear model to measure the interaction tie strength of the relationships on social networks; in our situation, we extract the heterogeneous data from Facebook. The results of the model of interaction make people's relationships into a kind of abstract emotion which become specific quantitative values. We will introduce every part as follows.

3.2.1 Interaction features between a user and the user's friends

According to the results of the online questionnaire we mentioned in Section 3.1.1, we extract the data pertaining to the seven features from the social network, and the process of the collection has been introduced in Section 3.1 as well. The questionnaire also shows that these features are mainly the ways the user in the social network interacts with the user's friends. Then we can determine these heterogeneous data as seven kernels to analyze the tie strength of relations in a linear model. The seven features are: likes for the user's status messages, comments for the user's status messages, likes for the user's photos, comments for the user's photos, tags the user is giving his or her friends in the user's photos, tags the user's friends are giving the user in the user's photos and mutual friends between the user and his or her friends. We use the same method on these elements as [43].This method of measuring the interaction, compared to measuring the interaction by dataset, will show the interactions in more detail, because it can collect the actual content of the activities between the user and the user's friends without advertisement and stranger requests, which means it has a lower noise than the way of collecting by datasets.

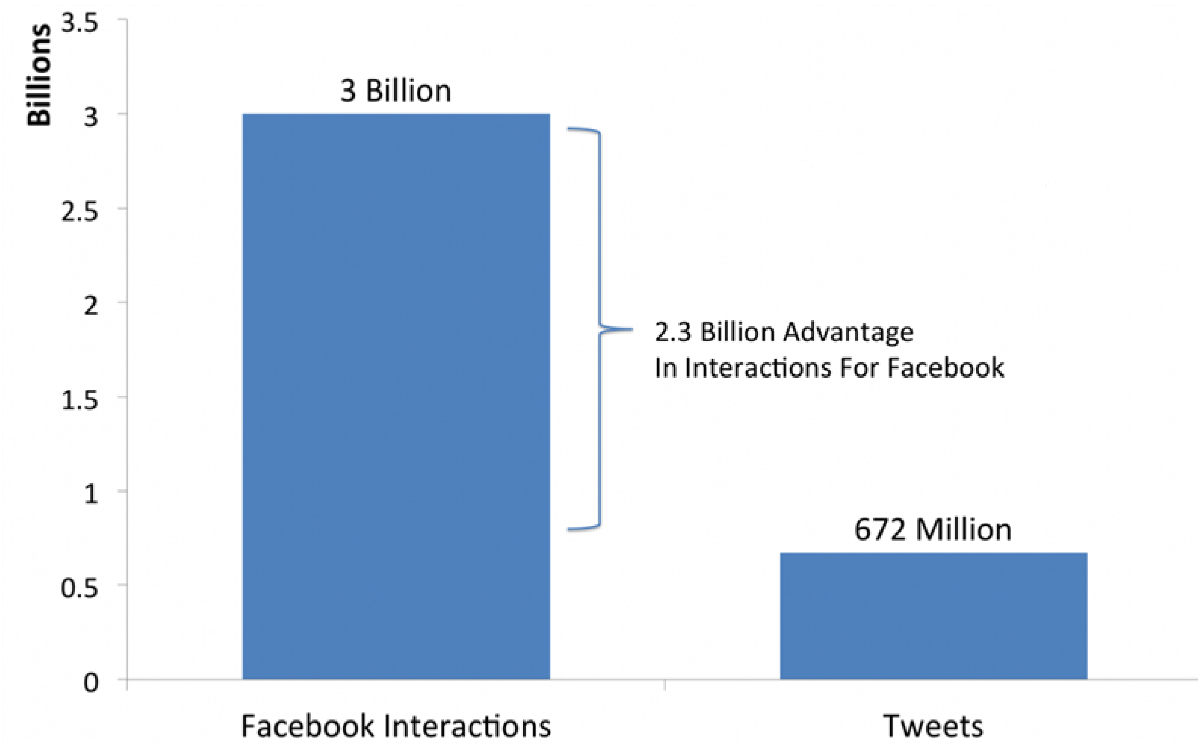


Figure 3.4: The comparison between Facebook and Tweets of the interactions

In future possible works, the detailed content of a user's comments in his or her status or photos can also be extracted so that it is possible to analyze the exact emotion that the friends have towards the user. There is an explanation for every kernel element as shown below.

Interaction by likes

There are a lot of interactions between friends in social networks. However, different social networks behave in different ways. Facebook is a very different social network than Twitter, and we can see the interactions by comparing Facebook to Twitter in Figure 3.4 which comes from website [44].

People prefer to browse their friends' photos, status messages or shared videos on their own home page rather than having private chats with their friends. Giving the status messages or pictures of friends a like becomes one of the interaction ways on the social network. The meaning of giving likes for user's friends is telling the user that they like it, they have paid attention to their photos or whatever, if they likes the photos, they just think that the photos belong to the user who is their good friend and they should give the person support. The original meaning of the like button for the Facebook creator is to represent the emotion of a user's friends toward the user. This kind of emotion that a user's friends has plays a significant role in determining the interaction frequency of tie strength. Gestalt psychology [45] illustrates that human behavior is determined by the relationship between people and the environment, and the behavior refers to external activities governed by the heart. It is evident that giving likes for the user's photos is one preferred type of interaction. We define the interaction of friend w_i with the user as the number of likes which friends w_i give to the user's photos. The equation is given as follows.

$$K_1(w_i) = \#likes\ for\ photos\ given\ by\ w_i$$

On the other hand, the number of likes that user's friends give to the user for the user's status messages can also be another feature to be added into the linear model. The status messages of the user are a way of representing the user's thinking in his or her life or things that happen in his or her life. Therefore, when friends give feedback by giving likes for the status messages, it means that the interaction can represent a type of attention or a interaction of emotion, whether that is negative or optimistic. Then we define the interaction of a friend w_i with the user as the number of likes which friends w_i give to the user's status messages. The equation is defined as follows.

$$K_2(w_i) = \#likes\ for\ status\ messages\ given\ by\ w_i$$

Additionally, a study in [46] indicates that analyzing a person's behaviors on social media can help learn about the person's personality. The number of likes the user receives can also be regarded as a significant element to detecting a person's personality.

Interaction by comments

This part can be divided into two parts which include comments for a user's status messages, and comments for a user's photos. As we know, according to the rule of Facebook, people can only give their comments to their friends' photos or status messages.
The photos and status messages of a user are the main way for a user to show his or her life in the real world, which is one of the most important aims for Facebook, to offer the user a platform. For the user's photos, they can provide a visual performance of the real life of the user, based on the user's personal habits, the user's recent activities, and what the user wants to show to the public. After the user posts the photos to the user's post wall, not every friend of the user will give a comment to the user's photos. The reasons for the user's friends to give a comment to the user are enormous, and depend on the situation of the user's friends. We randomly visited fifteen people in the SITE building of the University of Ottawa, asking the question what is the aim of giving a comment to their friends' photos on Facebook? The result of the questionnaire gives a total of four aims of giving comments to their friends, and more details are shown in Subsection 3.1.1 Table 3.3. The table indicates that the feature has an influence on the tie strength of interaction frequency between a user and the user's friends.
We define the interaction of friend w_i with the user as the number of comments w_i he or she gives to the user's photos.

$$K_3(w_i) = \#comments\ for\ photos\ given\ by\ w_i$$

It also can be found that the interaction frequency can reflect the popularity of the user depending on if they have a lot of likes or not. Giving comments to the user's status is also another significant element to determine the interaction frequency. We define the interaction of a friend w_i with the user as the number of comments w_i given to the user's status messages.

$$K_4(w_i) = \#comments\ for\ status\ messages\ given\ by\ w_i$$

Therefore, for the comments on the user's status messages and photos, we conclude the two features of interaction as two elements of all the elements for determining the linear model. The interaction of comments between a user and the user's friends compares with the interaction of likes, which can reflect more accurate information. Every interaction of like is produced by one incident of clicking the like button, and every interaction of comment is produced by friends thinking in their minds and typing many letters to the user. This set of behaviors represents the degree of attention to the user for the user's friends. There is no doubt that the comments are another more important element to measure the tie strength of interaction frequency on a relationship in a social network.

Interaction by being tagged in photos

As we know, the interaction frequency is the main factor to analyze the tie strength of a relationship between a user and the user's friends in Facebook, which can be obtained by combining these heterogeneous data in a linear model. However, the feature of being tagged by the user in the user's photos is one of these features, and it should be counted in the model. It can be solved with another question on why we tag, that refer to as the [47]. The research on the motivation to provide annotation in mobile and online media elaborates that tagging serves both a personal and a social purpose. Such an action thus shows that the friends not only interact with the user inside the social network, but they also have interactions in their real life, according to the content of the photos. On the other hand, to study social networks, sociologists conduct the relationships in a variety of ways. One important dimension is to examine the user's core relationships: those users discuss important matters with their friends. These core ties can be interactions online, offline, or most likely both. In a study report, and elsewhere, it has been reported that users tend to discuss important matters with more people than just those who do not use the internet [48]and [49]. As part of another report on Social Networking Sites and Our Lives [5], it found that frequent Facebook users tended to have even more close relationships than those who do not use Facebook. It means that the interaction frequency that the user has on Facebook has an important influence on the relationship between the user and their friends in their real lives, and there is no doubt that the user being tagged by the user's friends in photos increases a sense of intimacy or awareness. The number of tags in which the user is tagged by the user's friends can be defined as one of all the features. So we define the number of times a friend w_i tagged the user in a photo as

$$K_5(w_i) = \#times\ friend\ w_i\ tags\ the\ user\ in\ a\ photo$$

Interaction by tagging friends in photos

Generally, if one user tags one of his or her friends in his or her photos, it shows that the user must have emotions for the friend. People prefer to take an interest in their lives or people who are related to them such as their friends, relatives or enemies rather than complete strangers. And for a user, when the user posts a photo or picture on their home page, the user means to represent a kind of emotion in their life or what happened to the user. Then if the user tags one or more of the user's friends in the photos or picture, it can be said that the user would like to ask for attention from the friends that are tagged by the user. In another words, it means that the tagged friends have a vital position in the users heart, and that the user wants to obtain attention from their friends. So it is reasonable to say that the interactions of tagging friends can affect the tie strength of interaction frequency for determining the relationship in the social network.

Otherwise, sometimes the user posts not only photos, but some other pictures with tags of his or her friends. There is just another question on why the user tags his or her friends. In fact, this can be regarded as a kind of connecting with friends without words, so the tagged friends may give feedback or not. The entire process of the interaction can be achieved on the condition of the tie strength of relationship in the real world, which is useful for analyzing further more information about the character of the user, so it can improve the accuracy of detecting the strength of interaction frequency. In addition, if the post can be attractive to both of the user and the tagged friends, it also can show the similarity of habit for both of them. The factor of the similarity also can be an important role in the interaction frequency. According to the analysis of all the above, we define the number of times a user tags a friend w_i as follows.

$$K_6(w_i) = \#times\ user\ tags\ friend\ w_i\ in\ a\ photo$$

Number of mutual friends

The number of mutual friends is also a factor in the relationship between a user and the user's friends. When the sum of the total friend number of user and the totally friend number of a friend is fixed, it means that the more mutual friends the two persons have, the closer are their friend circles between each other. Because the user and the friend are friends, then the more mutual friends they have on Facebook, which will improve the probability of learn more each other by their mutual friends.

Meanwhile when the number of mutual friends that they have increases, it can show that they have a few common factors on making friends, or the similarity of making friends. Then it is not difficult to get the result on that the user and the friend have similarity elements. This is also a kind of tie in the relationship. To combine this factor with the others features, we need to count in the total number of user and user's friends, because when we compare the difference between user's one friend and user's another friend on Facebook, we can not only calculate the total mutual friends number of the user and the friend, but it also should count the affect of the number of friends that the user and the friend respectively has. Then we find out a way to solve the problem by using the percentage of the mutual friend number in the total friends number. However, this method basics on the set in math, and here we need to seek a total set with the two sets according to the Set theory in [50].

We can see the set A, which means the total number of the friend and the set B means the total number of the user. If we would like to get the percentage of the mutual friends in the total number of their friends, we should firstly get the total friends number . For the sum of the two sets, the equation should be built as follows.

$$A \bigcup B = A + B - A \bigcap B \tag{3.1}$$

$$C = A \bigcap B \tag{3.2}$$

Then we can get the sum of the two sets as shown in equation 3.3 by combining the equation 3.1 and the equation 3.2.

$$A \bigcup B = A + B - C \tag{3.3}$$

We obtain the two important values as they are shown in the equation 3.3 and the equation 3.2. We can define the number of the mutual friends as the following ratio:

$$K_7(w_i) = \frac{C}{A \bigcup B} \tag{3.4}$$

$$K_7(w_i) = \frac{T(mutual\ friends)}{T(w_i) + T - T(mutual\ friends)}$$

where $T(mutual\ friends)$ is the total number of mutual friends between the user and w_i, then $T(w_i)$ and T respectively represents the total number of friends of w_i, the total number of friends of the user on Facebook. Finally, the ratio can be obtained by this equation 3.4 to be our last defined feature. Then another issue is that the number of mutual friends can not always be extracted from website because the permission for this part depends on the friend w_i. If the friend makes the information public on Facebook or the user is defined as a close friend of the friend w_i, then we can get the number of mutual friends between the user and the friend w_i. When our system can not collect the part of data, the value of the part will be counted as zero in the calculation of interaction frequency.

3.2.2 Normalize the interaction features

Before we combine the features in the next section, we need a method to normalize the seven features respectively, and the method is called the Normalization method [51]. We have to do this step before we construct the specific model. The reason is that the interaction frequency for every acquaintance of the user should be comparative, which means that the data should be in the same dimensions and specifications.

As a first step, the aim is to find the feature column with the largest value, which results in a total of seven maximum values for seven feature columns. As we know, every friend of the user has seven features; therefore, the system should do the step above for the set of seven features of each user's friend in order to standardize these features. The processing of the procedure can be seen as followed.

$$\{M_{1,w_j},\ M_{2,w_j},\ M_{3,w_j},\ M_{4,w_j},\ M_{5,w_i},\ M_{6,w_i},\ M_{7,w_i}\ \} \ \in \ M(w_j)$$

Where $M(w_j)$ are the seven data dimensions of the friend w_j of the user, which includes seven elements. These seven elements are each chosen from the set of elements of the seven features for each element having the largest value in each feature elements, these elements are respectively represented as: $M_{1,w_j},\ M_{2,w_j},\ M_{3,w_j},\ M_{4,w_j},\ M_{5,w_i},\ M_{6,w_i},\ M_{7,w_i}$.

Secondly, to explain the procedure by using the first column as an example, we divide every element of the first feature column by the largest value that we get with the first step in the same column. The first feature column is now rebuilt with new standard values. Finally, the same procedure has to be done on the remaining six feature columns from K_2 to K_7.

3.2.3 Build a linear model for interaction frequency

In terms of this part, we build a specific linear model with the seven features that we got in the Section 3.2.1. We have defined the seven inputs for the linear equation and the meaning of the output of the linear equation, but there is still a question concerning the coefficients of the equation that we do not know and that needs to be figured out. A survey is posted online to get the coefficients so that the problem can be solved. The details of the survey are introduced in the Section of Experiment results. We analyze and calculate the proportion of each feature in all of the interaction features. The result of the data in the survey is shown as Figure 5.3 and Figure 5.4 in Section 5.1. The survey has been introduced in two editions, where one has been written in English and the second in Chinese respectively.

We then combine the results of the two editions of the survey into one result and calculate the proportion for each feature. The proportions of the seven features that we calculate are regarded as the coefficients of the model we built. Figure 3.5 shows the proportions as below and more explicit proportions of weights are shown in Section 5.1 Table 5.1.

We built a model to visualize the interaction frequency between a user and their friends. According to [52], we use a linear combination of the seven features K_1 to K_7 to calculate the interaction frequency. The interaction value as calculated here gives an insight to the online interaction behavior of the user's friends. We define the interaction frequency Y_i with friend w_i as follows:

$$\begin{aligned} Y_i = K_1(w_i)\beta_1 + K_2(w_i)\beta_2 + K_3(w_i)\beta_3 + K_4(w_i)\beta_4 \\ + K_5(w_i)\beta_5 + K_6(w_i)\beta_6 + K_7(w_i)\beta_7 \end{aligned}$$

$$Y_i = \sum_{n=1}^{7} K_n(w_i)\beta_n \tag{3.5}$$

Where K_n represents the seven dimensions of the data for the features that influence the interaction frequency. The array of β_n are the coefficients representing the proportion of every independent feature in the calculation of the interaction frequency. The goal of the combination of the two editions of the surveys was to investigate the degree to which the people on Facebook like interacting with their close friends by using a certain method of interaction. Figure 3.5 illustrates the relative proportion of each attribute. These

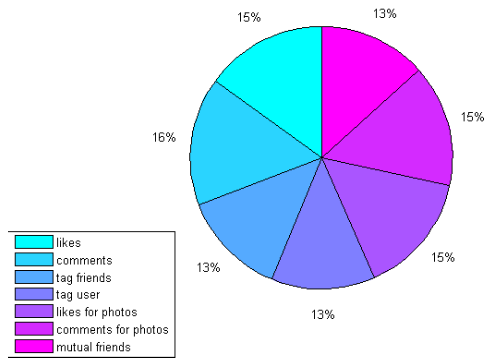


Figure 3.5: Weighted proportion for each method of interaction.

proportions in Section 5.1 Table 5.1 are used as the coefficients of Equation 3.5. Then the linear model of interaction frequency was built.

3.3 Conclusion

In conclusion, we introduced this chapter in two parts. The first part includes what kind of data we should collect and how do we collect these data. The second part includes the process in which we build the linear model of interaction frequency. We have concluded that we need to collect data from seven features form the social network Facebook with he Facebook API RestFB. We then transformed these seven features into seven kernels. We normalized these kernels and did an online survey to find the coefficients in order to build the linear model. In the next chapter, we mainly go through the process of building the 3D visualization system with a solar system concept.

Chapter 4

3D Visualization as a Solar System

First of all, in this section we describe the reason we chose the solar system as the model to create a 3D visualization. The user's friends with the highest interaction frequency with the user are represented as planets surrounding the user, which represents the sun. The various interaction attributes between the user and their friends are visualized by the physical variables: semi-major and semi-minor axes of the elliptical orbits, the size of the planets, and the angular velocity of the planets.

In addition, this section also introduces the mapping process of the interaction attributes with the physical variables, by using the Gravitation laws and Kepler's laws, and the definitions of the different variables elaborate various meanings. Finally, the procedure of building the 3D visualization of the social network with the user as the center of the solar system is introduced in detail. The entire systems aim is to work for the user in the social network, which can help the user learn more about their relationships with their friends.

4.1 Variables of the 3D visualization

We use the result of the data analysis and processing from Section 3.2 to define four dimensions of heterogeneous data for visualization in the model of the solar system. From the results of our survey, we find that the seven features can have influences on measuring the tie strength of the interaction frequency, and that the values of the weight for each feature are close to each other but the values are not exactly the same. So, not only it is valuable to combine all of these features into a linear model to indicate the interaction frequency so that we can predict the relationship between the user and the user's friends, but every single feature based on the interaction frequency is also useful to predict the personalities of the user's friends.

There is a result from a study in [53], and it reveals new insights on two systems used in their research, which are RE [54] and SybilGuard [55] and confirms their hypothesis that studies of social applications should use real indicators of user interactions in social network graphs. Each feature should be extracted by the system during a specific time period, and therefore, it gives a reason for us to visualize the interaction frequency of individual features. Another reason for the visualization of the interaction frequency of each feature is that the interaction of every feature can help predict the friends' personalities [46]. The detailed procedure of conducting four variables is introduced as follows in individual parts.

The first two variables are defined as the interaction frequency of likes and comments by the users friends. We collect the data of the number of likes and comments in the experiment and we set the timeline of the data to the past 6 months, which means that the number of likes and comments is changing as the time changes, and the degree of the change depends on the behavior of the user's friends. According to the research [46], it says that users' activity on Facebook relates to their personality, and the study measures the idea with a standard five factor model [56] called NEO-PI-R. Then, the interaction frequency of likes and comments is visualized in two physical variables. In addition, the different friends of the user can have different personalities that lead to different behaviors on Facebook. In a word, this visualization of the two factors can show the difference between the different friends of the user.
Hence, we break the two factors into two equations as follows:

$$X_1(w_i) = Y_i + K_2(w_i)\beta_2 \tag{4.1}$$

$$X_2(w_i) = Y_i + K_4(w_i)\beta_4 \tag{4.2}$$

Where $X_1(w_i)$ represents the interaction frequency of the friend w_i giving likes to the user's status messages. The equation means that it is a type of interaction frequency of giving likes to the user's status messages, which is a kind of extension based on the total interaction frequency. The $X_2(w_i)$ is the interaction frequency of the friend w_i leaving comments to the user's status messages. It represents the factor in the same way as $X_1(w_i)$. Obviously, there are two variables that should be visualized in the 3D visualization by borrowing the concept of a solar system.

Otherwise, there is another variable that is necessary to be visualized, which is the total interaction frequency Y_i. The Y_i is the only way to tell the user the total interaction frequency of the user's friends, and another advantage is that it can give a list of all the user's friends, according to the degree of the tie strength of the interaction frequency for each of the user's friends. Actually, the tie between weak and strong can lead to prediction of the friendship, which can be found in the study [52]. Therefore, the value of the Y_i is another vital variable needed to be shown as the third variable.

The third variable is defined as follows, and the equation of Y_i is introduced in Subsection 3.2.3. For this part, it directly transforms into the third variable.

$$X_3(w_i) = Y_i \tag{4.3}$$

Lastly, the number of the likes and comments given to the user's photos is used as the fourth variable. For this part, the interactions of giving likes and comments is unstable, depending on the number of photos in the user's albums. It can show the situation of friends' attention to user's photos in real time and the changes can be shown in the visualization, which is the reason for picking it as the fourth variable.

It is defined in the following way:

$$X_4(w_i) = Y_i + K_1(w_i)\beta_1 + K_3(w_i)\beta_3 \tag{4.4}$$

Overall, we define four variables to be visualized in the visualization part , and the four variables are: the interaction frequency for the number of likes given to the user's status messages as the first variable, the interaction frequency of the number of comments given to the user's status messages as the second variable, the total interaction frequency by combining the seven features as the third variable, and the interaction frequency of the total number of the likes and comments given to the user's photos as the fourth variable.

4.2 Mapping the variables onto the physical model

According to subsection 4.1, we defined four variables for the visualization. We chose to match these four variables with the parameters of the solar system as shown in Figure 4.1. It shows that there are four parameters in the solar system to be used for visualization. They include the size of the planets, the speed of the planets, as well as the long axis and short axis of the elliptical orbits. However, when the solar system is running, the four variables of the interaction frequency that we need to visualize also should obey the universe's laws, because these laws are the reasons of the formation of the solar system.

Before we introduce the next part, which is the process in which we map the variables, it is necessary to know a short definition of the solar system. A solar system refers to a star and all the objects that travel in orbit around it. Our solar system consists of the sun - our star - eight planets and their natural satellites (such as our moon); dwarf planets; asteroids and comets. At the same time, we also decided to only choose the top nine friends in the list of the user's friends in the 3D visualization. There are two points to support the choice we made. The first one is that the visualization is done so that the user can understand their relationships in the social network, so it doesnt mean that the more friends there are, the better results there are. Relationships with great value will have the best results in the visualization for the user. The second point is that if there are more friends put together, the whole structure of the visualization will be messy and complex, leading to confusion for the user, and this result can be useless for the user in understanding the relationships.

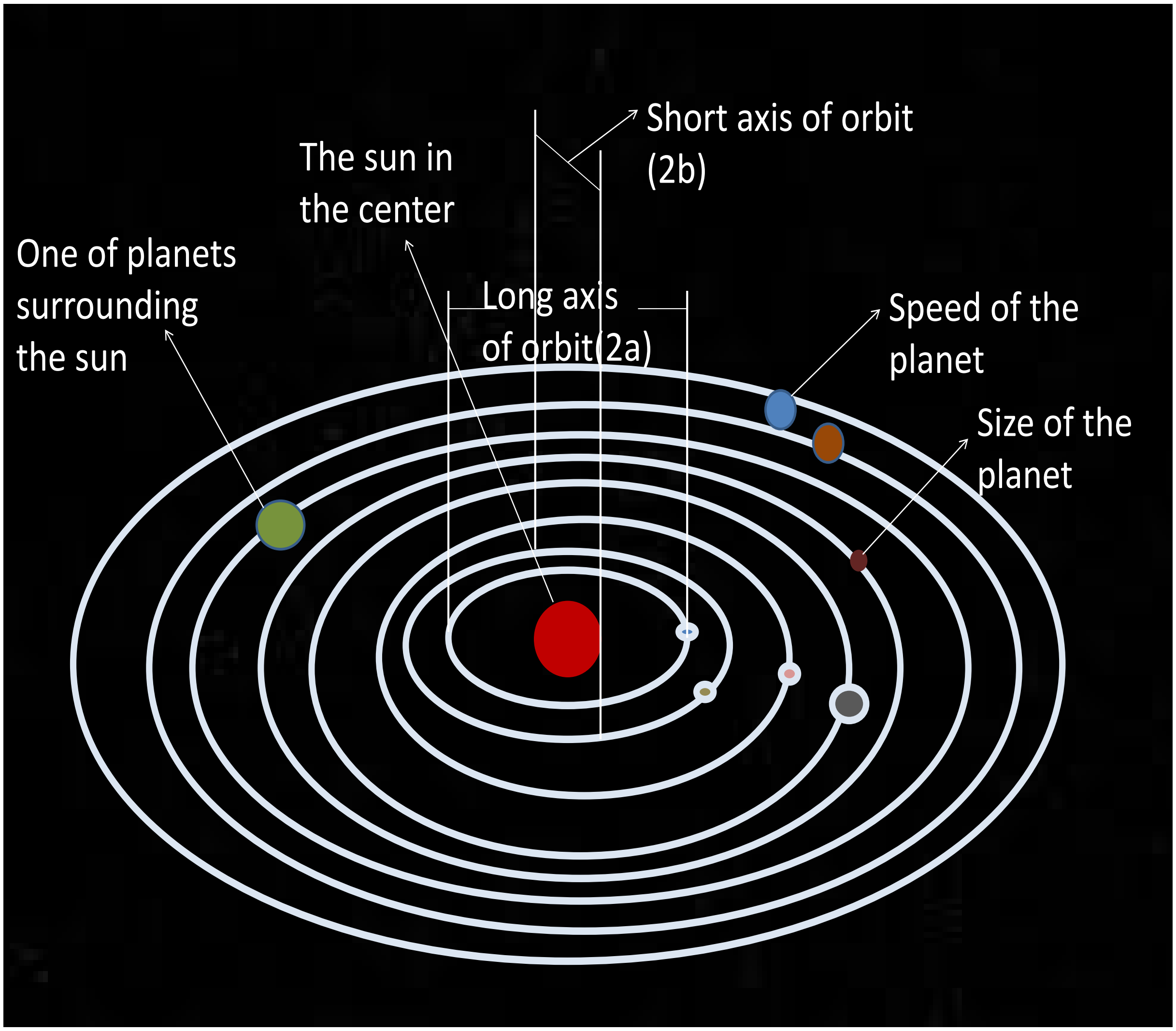


Figure 4.1: The parameters of a solar system

In the part of the mapping process, firstly, Newton's law of universal gravitation explains why planets maintain their orbit. The law states that "every point mass attracts every single other point mass by a force which points along the line of interaction between the two points . The force is the proportional to the product of the two masses and inversely proportional to the square of the distance between them" [35]. The original equation is as follows:

$$F_{gravitaion} = \frac{GMm}{r^2} \tag{4.5}$$

where G is the gravitational constant ($6.674 \times 10^{11}\ N(m/kg)^2$), M is the mass of the object in the center, m is the mass of the orbiting object and v is its velocity. r is the distance between the center of masses.

The equation 4.6 is the original equation of centripetal force, which is directed at right angles to the motion and also along the radius towards the center of the circular path [57] [58]. The mathematical description was derived in 1659 by Dutch physicist Christiana Huygens [59].

$$F_{centerforce} = \frac{mv^2}{r} \tag{4.6}$$

where m is the mass of the object doing circle motion , and v is its velocity. r is the radius of the trajectory.

For a solar system, every planet surrounds the central sun in an elliptical orbit or circle orbit, which means that every planet except the center sun will need a centripetal force to support this work. Therefore, in the Universe, the gravitation is used for the centripetal force [35]. We know the equation 4.5 equals the equation 4.6; then we can get the equation as follows:

$$F_{gravitaion} = F_{centerforce} \tag{4.7}$$

$$\sqrt{\frac{GM}{r}} = v \tag{4.8}$$

We get the equation of the speed v used in the visualization as one of the four variables. On the other hand, the shape of a planet's orbit is elliptical and it follows Kepler's laws of planetary motion. There are three laws introducing the motion for the surrounding planets in Kepler's laws of planetary motion. We can see the laws by seeing Figure 4.2, which indicates the relationship between the center planet and the other planets as well as the laws in a clearer way. The laws are shown below and they are:

- The orbit of a planet is an ellipse with the Sun at one of the two foci [60].
- A line segment joining a planet and the Sun sweeps out equal areas during equal intervals of time [61].
- The square of the orbital period of a planet is proportional to the cube of the semi-major axis of its orbit.

Kepler's work improved the heliocentric theory of Nicolaus Copernicus, explaining how the planets' speeds vary, and using elliptical orbits rather than circular orbits with epicycles [62], and the explanation of how it works uses an original equation for the orbit of planets and is given as follows.

$$r = \frac{ab}{\sqrt{a^2 \sin^2\theta + b^2 \cos^2\theta}} \tag{4.9}$$

$$\theta = V_\theta * t \tag{4.10}$$

where r represents the radius of the ellipse at one point in time and V_θ is the angular velocity of the moving planet. a and b are the semi-major and semi-minor axes, respectively.

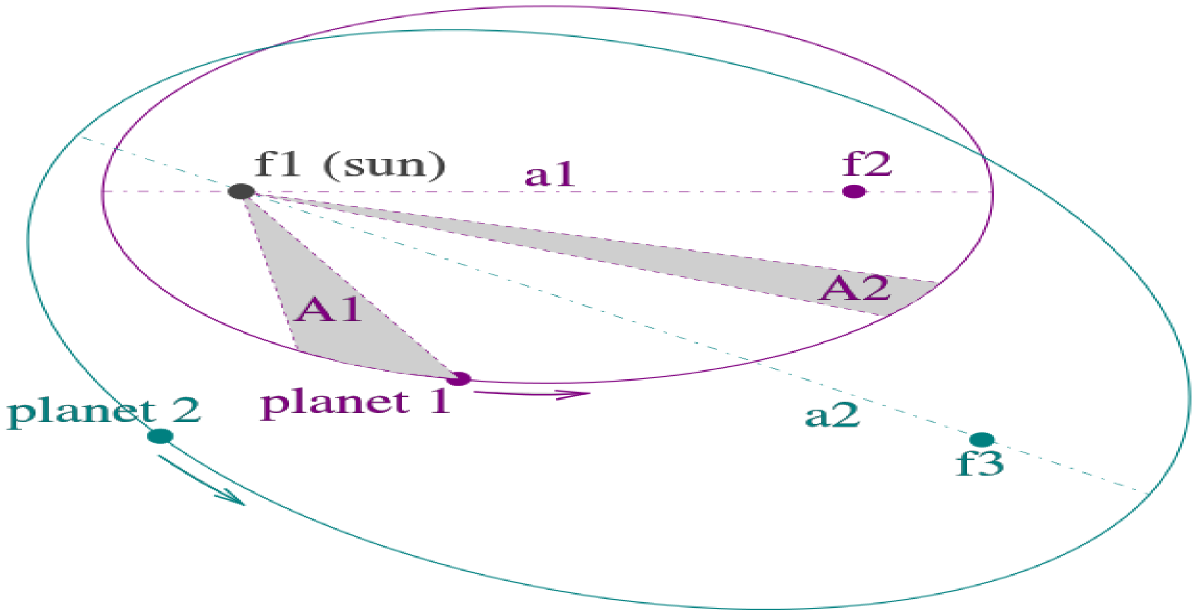


Figure 4.2: Illustration of Kepler's three laws with two planetary orbits.

In the next step, we connect the Equation 4.9 with the data that we defined. We assign our first data variable (see Equation 4.1) to the semi-major axis a, and the second data variable (see Equation 4.2) to the semi-minor axis b, of the elliptical orbit. We then assign the fourth data variable (see Equation 4.4) to the angular velocity V_θ of the planet, so that we can obtain a value for θ by using Equation 4.10.

Lastly, the third variable (see Equation 4.3) is used as the size of the planets. Putting it all together, by combining equations 4.1, 4.2 and 4.9 with $a = X_1 w_i$, $b = X_2 w_i$, then we can get

$$r = \frac{X_1(w_i)X_2(w_i)}{\sqrt{X_1(w_i)^2 \sin^2\theta + X_2(w_i)^2 \cos^2\theta}} \tag{4.11}$$

The fourth data variable in Equation 4.4 is then set to be the angular velocity V_θ of the angle θ, and it is combined with Equation 4.9. In the same way, we can obtain a value for r 4.11. By inserting Equation 4.11 into Equation 4.8, the value of the speed v can be calculated by us as follows:

$$v = \sqrt{\frac{GM\sqrt{X_1(w_i)^2\sin^2\theta + X_2(w_i)^2\cos^2\theta}}{X_1(w_i)X_2(w_i)}} \tag{4.12}$$

Now we obtain v as the speed of a surrounding planet in the 3D visualization, which is directly proportional to the angular velocity $X_4 w_i$, and the speed of v can meet the laws for the planets' running, which is the acceleration for the near center of the movement, and when it is far from the center, its movement decelerates. Overall, we united all four variables of heterogeneous data in the 3D visualization of the solar system in four different physical variables. In summary, the extension of the interaction frequency with the likes and comments for the user's status messages were used as a and b of the elliptical orbits, respectively. According to the Equation 4.12, we have speed v as the speed of the planets with V_θ equals X_i^4. The interaction frequency X_i^3 is represented as the size of the planet.

Finally, for this part, we finished the mapping of the visualized variables with the physical variables with the physical laws coming from Newton and Kepler.

4.3 Building the 3D visualization for a user

Table 4.1 indicates three possible tools that can be used to create 3D visualization and that are the most popular tools for developers to build the 3D models. By comparing the disadvantages and advantages between the three tools, we decide to use the game engine Unity 3D to visualize our solar system.

The tool that we chose allows us to build interactive, three dimensional scenes, and it also supports many different programming languages, target platforms, and a number of data connectivity interfaces, and there are a lot of plugins found on the website. However, for the other two tools, the programs are too big for a small team to create a 3D model and the tools are very expensive. On the other hand, the languages they support are not

Tools	Languages	Advantages	Disadvantages
Unity 3D	Javascript C++	Many plugins can improve efficiency. Application of flexible function. It is convenient for individual to develop	Releasing to the platform of iOS or Android will cost a lot of money.
Maya	Phython	Very high quality frame effects. It is usually used for 3D movies or animation.	When it deal with complex model, it will take a lot time, and it needs a team to work.
3DMAX	Maxscript (similar to C)	Powerful functions. It can execute a 64 bit system and the kinds of plugins are rich. It can match autocad.	

Table 4.1: The comparison of three 3D visualization tools

common, so we need time to learn them. We therefore decided to choose the Unity3D as our implement tool and the reasons for using it as described above.

In the next step, we introduce our final algorithm to the process of visualization, which is given in Algorithm 1. We divide the entire system into three parts as explained in our algorithm, which are: the values of input variables, the formulas for the processing with these variables we calculated, and at last, we can get the results of the formulas.

When processing the visualization this way, the indicated speed, size, length and width of the elliptical orbits are proportional to the four-dimensional heterogeneous data variables. At first, Figure 4.3 shows the first step; we created a cubes rotation simulating the planets in the solar system, then the second figure 4.4 introduces the initial solar system. We can see, in the second figure, that there is a sphere in the center of the system, and that the other planets are surrounding the center planet that can be regarded as the sun. As shown in Figure 4.5, Figure 4.6 and Figure 4.7, three different views of our visualized solar system for a specific user and their friends are demonstrated. For the last figure of the three, it illustrates that there is more than one solar system only when there is more than one user logging into the system, because if only one user logs into the system, we can only collect the data from the user's homepage, not the data of his or her friends unless they are also logged in. In the Figure 4.7, we find five volunteers to join the experiment. Then have the results as shown.

When the interaction frequency is changed, the elliptical orbits of the planets will be changed, and they may become larger or smaller. The speed of the planets and the size of the planets also increase or decrease with the interaction frequency changing at the same time. The top ten list of the user's friends depends on the value of the r of the elliptical orbit. When the value of r is larger, which means their interaction tie strength is stronger, the inverse of the value r will be smaller, and the distance will be closer, which means that the relationship for the user and the friend is closer than others.

In order to make it clearer, we add lights for the orbits of the running planets, so that we can see the shape of the elliptical orbits.

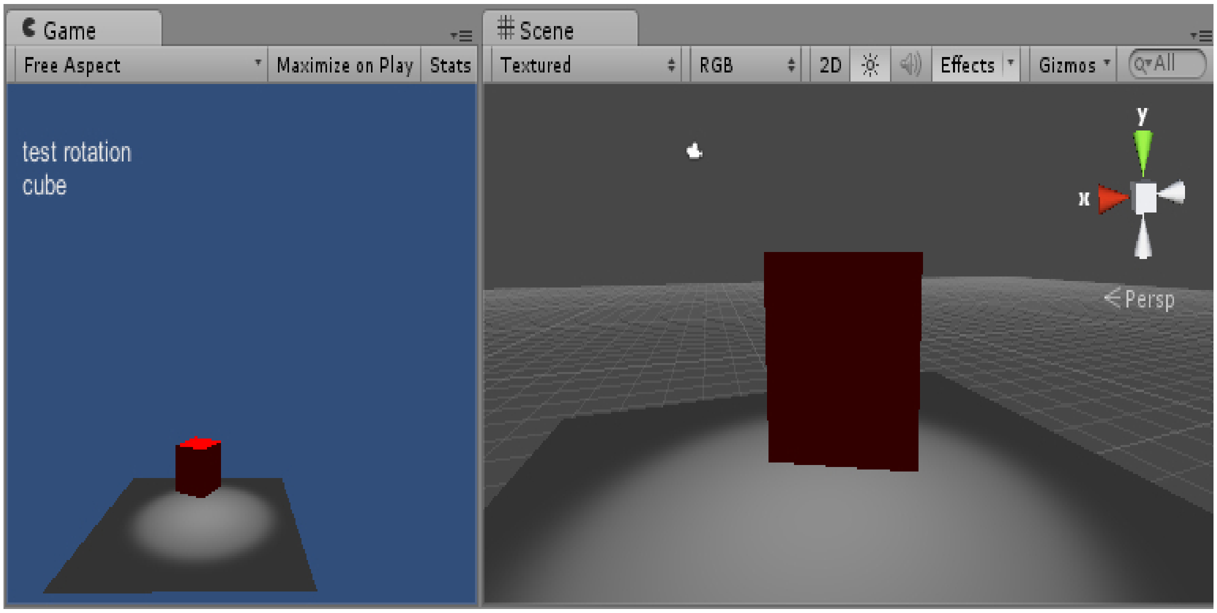


Figure 4.3: The test of rotation cube

We add a slider which helps the user change the speed of all planets at the same time. When we change the speed to zero, then all the planets will be frozen. At this time, when we put the mouse on different planets, it shows the information of each planet that represents a friend of the user and the information includes the name and ID for the friend. So far, we achieve all these goals and visualize all them in our system.

In addition, for the system, we know that every user will have a lot of friends on Facebook, and therefore it is useless to visualize all the friends of the user, and also it is difficult for human vision to distinguish the whole visualization, which will be messy.

algorithm 1 Procedure of creating a solar system

input: $Array\ X_1$, $Array\ X_2$, $Array\ X_3$, $Array\ X_4$, $Array\ ListID$, n, $UserID$

output: $Array\ a$, $Array\ b$, $Array\ m$, $Array\ Position$, M, $Array\ Speed$

function LongaxisofEllipse(X_1, n)
    $i \leftarrow 0$, $j \leftarrow n$
    while $i < j$ **do**
        $a[i] \leftarrow X_1[i]$
        $i++$
        return $a[i]$
    end while
end function
function ShortaxisofEllipse(X_1, n)
    $i \leftarrow 0$, $j \leftarrow n$
    while $i < j$ **do**
        $b[i] \leftarrow X_2[i]$
        $i++$
        return $b[i]$
    end while
end function
function Weight(X_3, n)
    $i \leftarrow 0$, $j \leftarrow n$
    while $i < j$ **do**
        if $ListID[i] == UserID$ **then**
            $M \leftarrow X_3[i]$
            return M
        else $m[i] \leftarrow X_3[i]$
            return $m[i]$
        end if
        $i++$
    end while
end function

function POSITION(X_4, n)
  $i \leftarrow 0$, $j \leftarrow n$
  $x \leftarrow 0$, $y \leftarrow 0$
  while $i < j$ **do**
    $\theta \leftarrow X_i^4 * t$
    $x = a[i]\sin\theta\ ;\ y = b[i]\cos\theta$
    $Position[i] \leftarrow (x, y)$
    $r[i] = \frac{X_1(w_i)X_2(w_i)}{\sqrt{(X_1(w_i))^2 \sin\theta^1 + (X_2(w_i))^2 \sin\theta^2}}$
    $i++$
  end while
  return *Position*
end function
function SPEED(M, r, n)
  $i \leftarrow 0$, $j \leftarrow n$
  while $i < j$ **do**
    $Speed(i) = \sqrt{\frac{GM}{r[i]}}$
    $i++$
  end while
  return *Speed*
end function

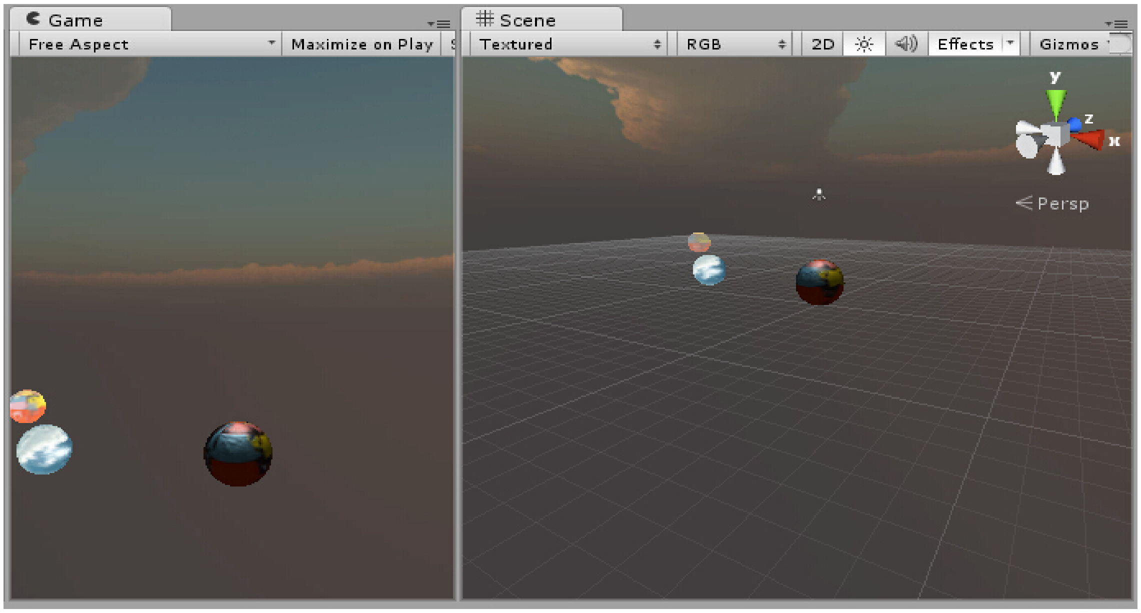


Figure 4.4: The initial solar system

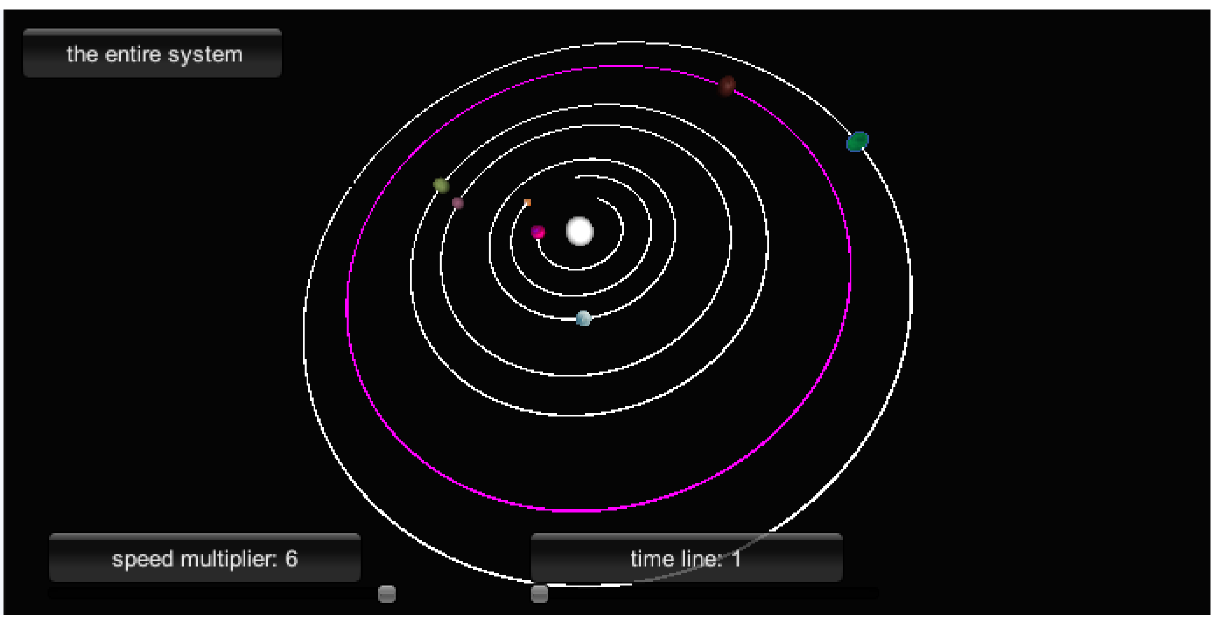


Figure 4.5: The user in the center is surrounded by their friends. The semi-major and the semi-minor axis of every elliptical orbit represent two types of interactions. The different distance between user and his or her friends represent the tie strength of interaction frequency.

Figure 4.6: This alternative view of the solar system shows the different speeds and sizes of the planets. All of them are in the same dimension and this matches the solar system because they are in the same space means that they in the same standardization.

4.4 Conclusion

For this chapter, we introduced this chapter by three steps. The first step describes what kind of variables we should visualize and how do we define these variables. The second step includes the process in which we map the variables onto the physical model. The last step introduces the result of building the 3D visualization for a user. We have concluded that we determine four variables from the seven features form the social network Facebook. We then map these four variables onto four physical variables in the physical model. Finally, we build the entire system and visualize the relationships between the user and his or her friends in a 3D physical model. In the next chapter, we mainly illustrate the process of experiment and the results of the experiment and the online survey we did is introduced in detail.

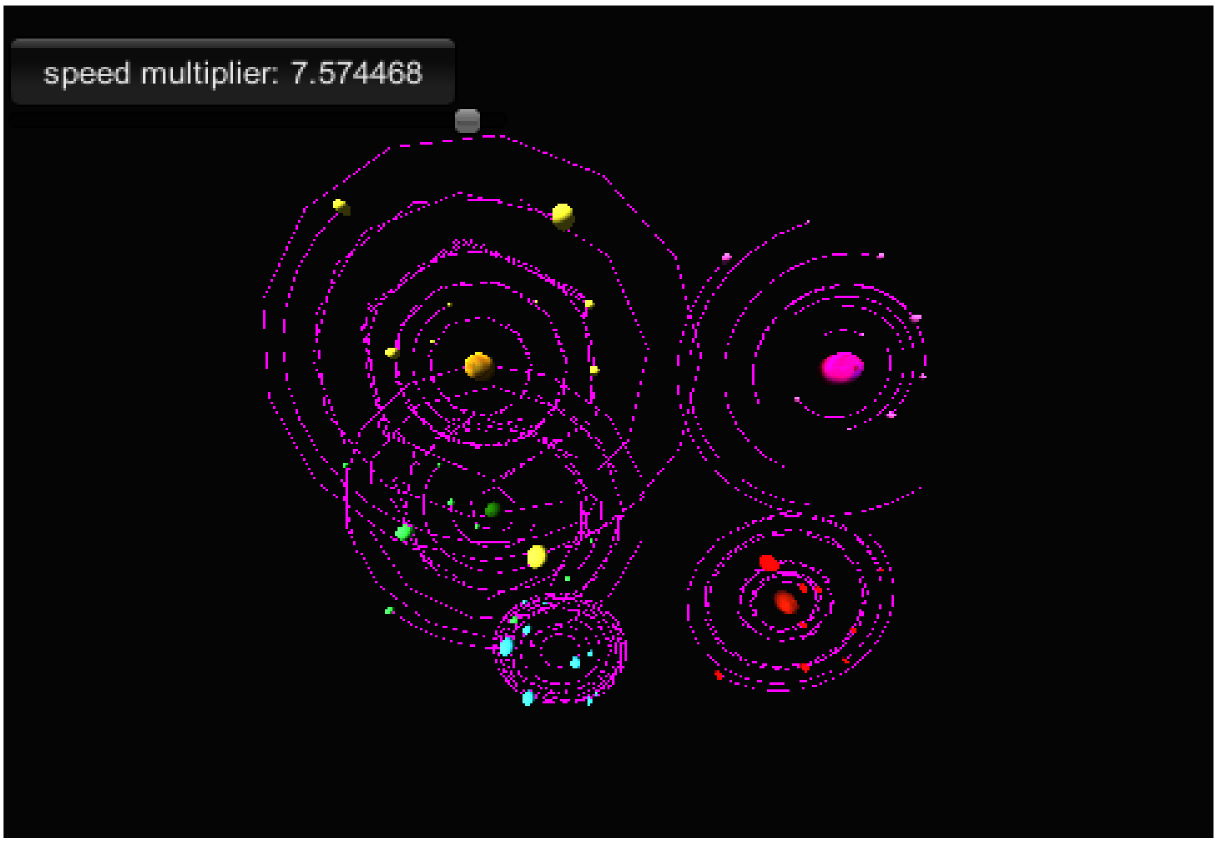


Figure 4.7: This view visualizes the relationship of the five users and their friends with physical variables.

Chapter 5

Survey and Experiment Results

We divide this section into four parts. The first part indicates the propose of the questionnaire and the procedure of collecting information in detail. The second part is about extracting the results of the survey. The third part introduces the comparison method and explains the processing of the comparison between the system friends' rank and the ground truth from the volunteers. The last part shows the final result and evaluations.

5.1 Survey

Firstly, as we known, if the user knows which one kind of behavior he or she thinks they have a tendency to exhibit on Facebook, these behaviors will really be shown on social networks, according to the conclusion that behaviors depend on their personality. Then we designed the survey by using the results of these research studies. We added seven features into our survey and let the volunteer rate each feature, because if one volunteer thinks that one feature is the way that he or she has a tendency to interact with his or her friends on Facebook, this means that this way is more comfortable for the user, or in another words, the chosen way is more suitable to the user's personality. Therefore the rating is also representative of the user's personality. For the rating, the range of the score is from one to ten points. The most likely answer is ten points and the least likely answer is one point.

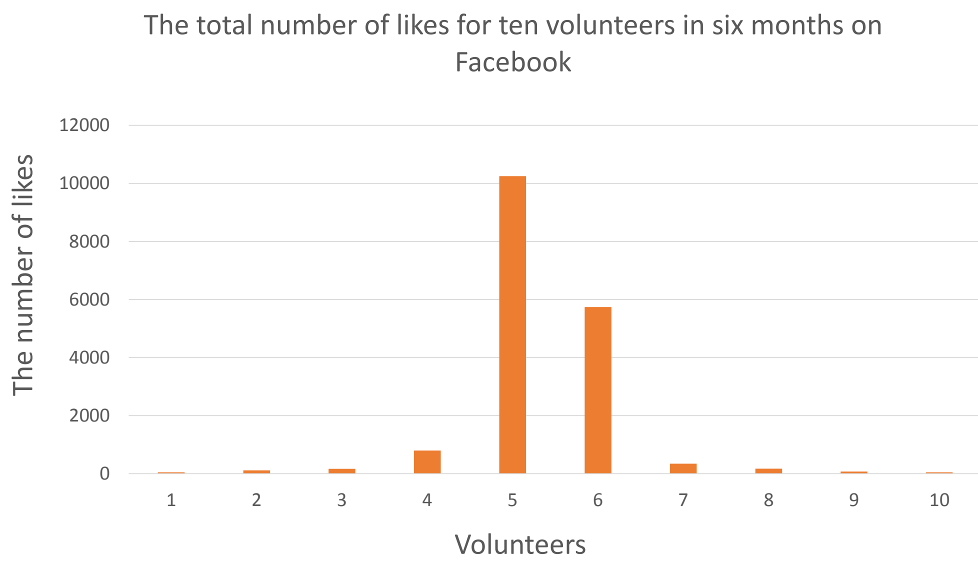


Figure 5.1: The total likes for user's status messages and photos

Then, we can see that figure 5.1 shows the total number of likes that the user's friends give to the user's status messages and photos by inviting ten volunteers to join the data collection for the last six months on Facebook. Figure 5.2, as shown, illustrates the number of comments that user's friends give to the user's status messages. User five has the most comments during this time, however, number seven has the least number of comments. There must be a reason for the distinction between the most and the least number of comments for the user's status. It can be obvious to find that the interaction frequency can reflect the popularity for the user so that the user's friends like to give comments to the user's status or the social situations for the user that mean, if the user has more friends on Facebook than another user, then it is possible for him or her to obtain more comments for a user's status messages. Sometimes, there may also be some of not friendly comments for the user's status, but it actually is still a kind of interaction between a user and the user's friends, which can be regarded as the user's friends' attention on the user.

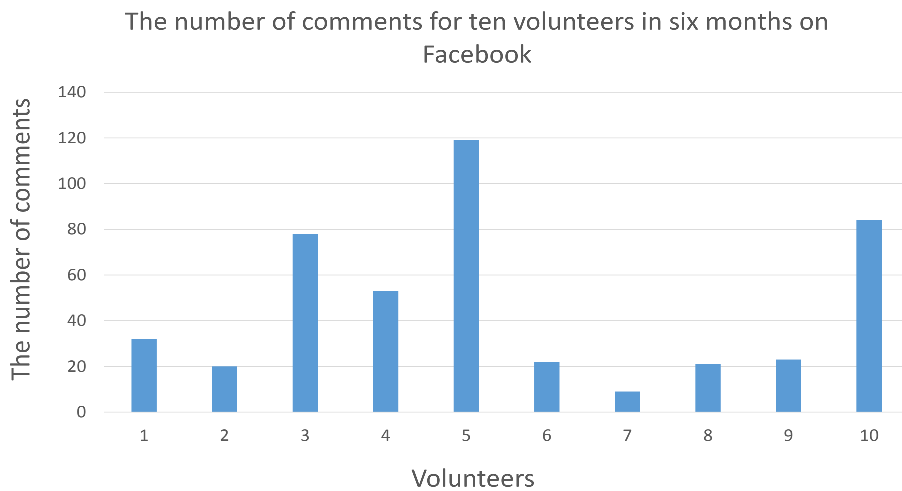


Figure 5.2: The comments for user's status messages

Another problem is to calculate the weights for each feature. The solution of this problem is to analyze the results of the survey. There are a total of eighty five people joining the survey online, and there is a short introduction in Figure 5.4 and Figure 5.3 as shown. According to Figure 5.3, a total of forty four volunteers participate in the survey written in Chinese. The participants can rate each feature on a scale from 1 to 10 depending on their satisfaction with the feature. The diagram shows this, with the ratings on the horizontal axis and the percentage of people voting for each rating on the vertical axis, and each line representing one feature, for a total of seven lines. As seen, the proportion of people voting 10 points for each feature is more than any of the other 9 ratings. On the other hand, for the survey written in English, a total of forty one volunteers were participating in the survey, and the questions in the survey are the same as the survey written in Chinese. The results can be seen in Figure 5.4, which shows that most of the volunteers rated between five and ten points for each of the seven features. There are only a few volunteers that rated either one or two points. To show the results clearly, we add the two figures together, and the result of the whole figure is shown in Figure 5.7, which shows the proportions of

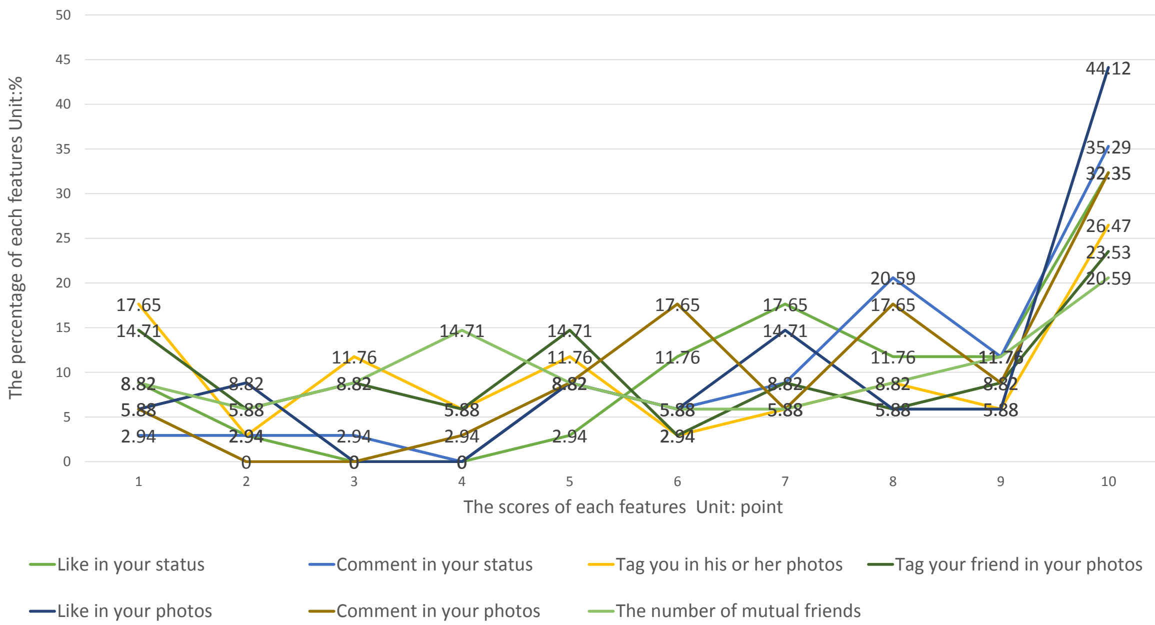


Figure 5.3: The proportion of seven features in Chinese

Features	Likes	Comments	TagFriend	TagUser
Average	7.45	7.85	6.44	6.35
Features	LikesPhotos	CommentsPhotos	MutualFriends	
Average	7.45	7.49	6.59	

Table 5.1: Average preference given to each method of interaction on Facebook.

the ratings for each feature. A total of 85 people participated in the survey. The average score for each feature is shown in Table 5.1.Most people thought that these seven features were the most popular ways for them to interact with their friends on Facebook.

Additionally, to solve the problem of how to combine the two surveys together, we use the method of weighted arithmetic mean [63]. The original equation of the method of weighted arithmetic mean is shown as follows.

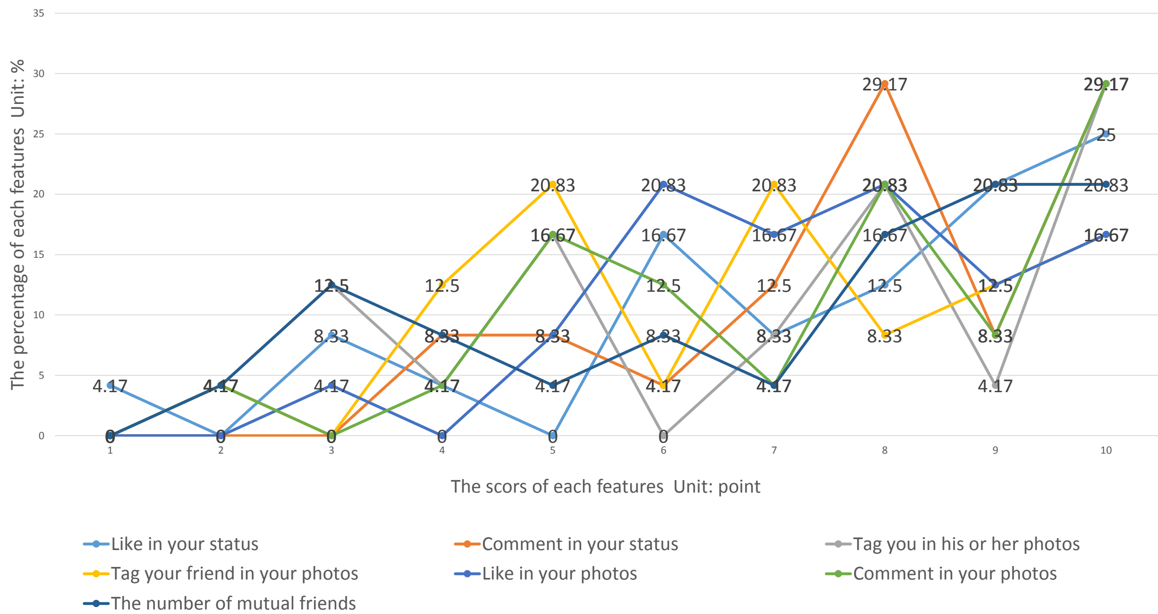


Figure 5.4: The proportion of seven features in English

$$\overline{X} = \frac{w_1 x_1 + w_2 x_2 + \ldots + w_n x_n}{w_1 + w_2 + \ldots + w_n}$$

$$\overline{X} = \frac{\sum_{i=1}^{n} w_i x_i}{\sum_{i=1}^{n} w_i} \tag{5.1}$$

Then, we combine the two surveys by the method of weighted arithmetic mean. Firstly, we can extract four tables from the two surveys and form two pairs of tables from the two surveys. They are shown as below.

First pair for the first survey:

$$\left\{\begin{matrix} x_{1,1} & x_{1,2} & \cdots & x_{1,10} \\ x_{2,1} & x_{2,2} & \cdots & x_{2,10} \\ \vdots & \vdots & \vdots & \vdots \\ x_{7,1} & x_{7,2} & \cdots & x_{7,10} \end{matrix}\right\} \; and \; \left\{\begin{matrix} w_{1,1} & w_{1,2} & \cdots & w_{1,10} \\ w_{2,1} & w_{2,2} & \cdots & w_{2,10} \\ \vdots & \vdots & \vdots & \vdots \\ w_{7,1} & w_{7,2} & \cdots & w_{7,10} \end{matrix}\right\}$$

In the first pair of tables, they are called table one(x) and table two(w). Both of them come from the first survey written in Chinese. In this table, each element is represented by $x_{i,j}$ of the table, with the proportion for the feature i(total features are seven) and the rating j points. In table two, each element of the table shows the frequency of the people that voted for the feature i with the j points.

Second pair for the Second survey:

$$\begin{Bmatrix} y_{1,1} & y_{1,2} & \cdots & y_{1,10} \\ y_{2,1} & y_{2,2} & \cdots & y_{2,10} \\ \vdots & \vdots & \vdots & \vdots \\ y_{7,1} & y_{7,2} & \cdots & y_{7,10} \end{Bmatrix} and \begin{Bmatrix} z_{1,1} & z_{1,2} & \cdots & z_{1,10} \\ z_{2,1} & z_{2,2} & \cdots & z_{2,10} \\ \vdots & \vdots & \vdots & \vdots \\ w_{7,1} & z_{7,2} & \cdots & z_{7,10} \end{Bmatrix}$$

For the second pair of tables, there are table three(y) and table four(z). Table three(y) has the same explanations as table one(x), and table four(z) is the same as table two (w) except that these are done with the English version of the survey.

The next step is to combine the four tables into one table with the theory of the Equation 5.1. Then, we can have an equation to calculate each element in the new table as follows.

$$\overline{X_{i,j}} = \frac{w_{i,j}x_{i,j} + y_{i,j}z_{i,j}}{w_{i,j} + z_{i,j}} \tag{5.2}$$

Where $x_{i,j}$, $w_{i,j}$, $y_{i,j}$ and $z_{i,j}$ are extracted from the first table, second table, third table and fourth table, respectively. i $\in [1, 7]$ represents the seven features and j $\in [1, 10]$ is the rating given between 1 point and ten points.

The final table can be seen in Figure 5.8. The figure introduces the different scoring situations for each feature. The figure is named the map of the percentage accumulating area, and it is used to stress that the percentage of each features value on the horizontal axis changes when the rating points are changed on the horizontal axis. According to the results of the figure, we can see that these features all have a majority of ratings from 8-10

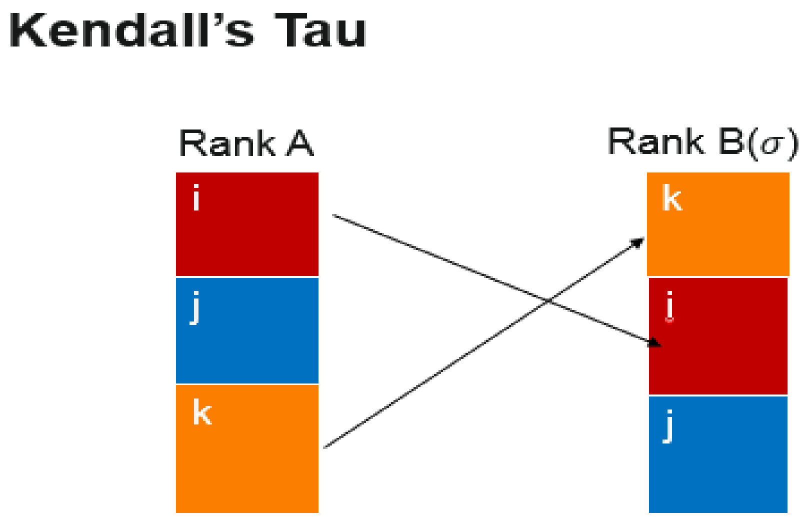


Figure 5.5: The method of Kendall's Tau

points and are therefore vital to the visualization, especially for calculating the weights for the final table.

5.2 Experiment Results

The second part is to compare the real true ranking to the ranking we obtain with the system. We used a method called Kendall's Tau [64]. This method of comparing ranks is realized by the number of combinations possible in the true ranking and then comparing them to the same combinations in the test rank (obtained with the system) and then seeing how many of them are right, thus given us a ratio. when the test combination does not match the combination of the true ranking, we call it an inversion. The Figure 5.5 illustrates the meaning of inversion: A pair of elements i and k such as i > k in a ranking and $\sigma(i) < \sigma(k)$ in another ranking. We can see an example in Figure 5.5 and Figure 5.6, which are a pair of inversions. Then the accuracy of the ranking B compared to ranking A is the total number of inversions in the ranking Bσ, so we have the equation of the method as shown in Equation 5.3.

For the ranking A, the total number of pairs of inversions are shown in the Equation 5.4 where n represents the total number of elements in the ranking A.

$$K_D(\sigma) \;=\; \sum_{i<j} 1_{\sigma(i)\;>\;\sigma(j)} \tag{5.3}$$

$$K_D \;=\; \sum_{i=1}^{n-1} i \tag{5.4}$$

We then get the equation below to calculate the accuracy.

$$R_a cc \;=\; \frac{K_D(\sigma)}{K_D}$$

$$R_a cc \;=\; \frac{\sum_{i<j} 1_{\sigma(i)\;>\;\sigma(j)}}{\sum_{i=1}^{n-1} i} \tag{5.5}$$

Where $R_a cc$ is the accuracy by comparing the ranking A with the Ranking B. According to this method, we regard the true ranking as A, and regard the ranking coming from the system as B (σ). We can calculate an accuracy for each volunteer by using the Equation 5.5. As we know, there are seven features for each user, according to each feature's data; it can also give a ranking of the user's friends for the same feature. This means we can get seven new rankings (one for each feature). The purpose of we doing this is introduced as below.

For our research, it is necessary to consider the influence of each feature on the friends' ranking, and therefore we also should calculate the accuracy of each feature for the user's data by comparing the seven new rankings to their true rankings respectively.

Therefore, in the last part, we found ten volunteers to take part in the experiment. To get each rank of seven features for each volunteer and build our kernels mentioned in Section 3.2, we extracted seven types of interaction information from each of the ten volunteers

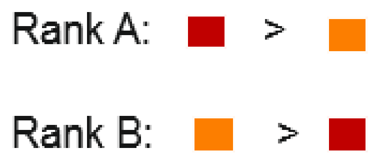


Figure 5.6: The example for the rank

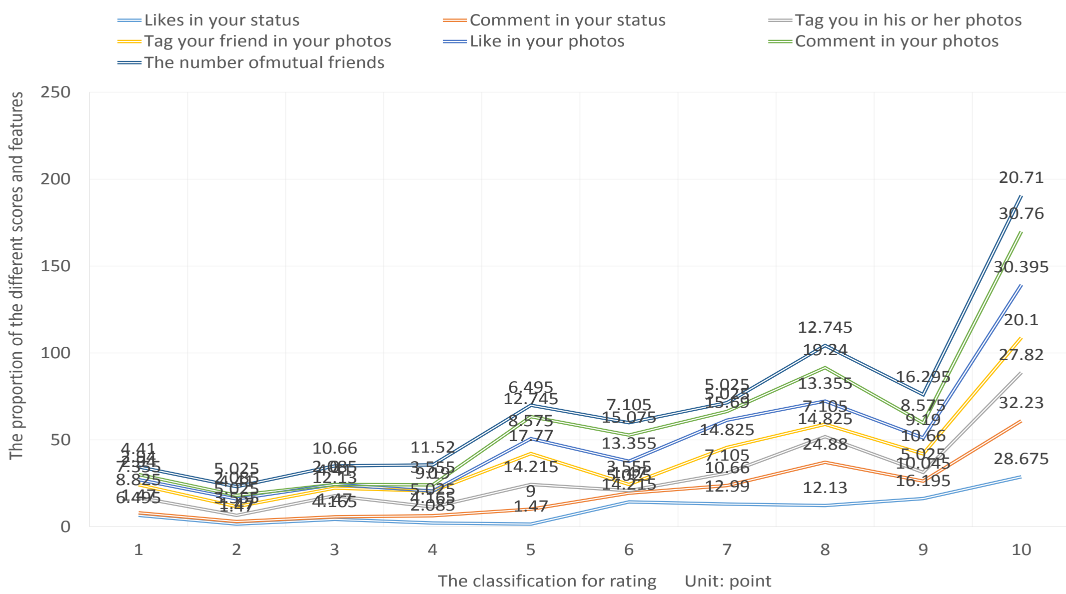
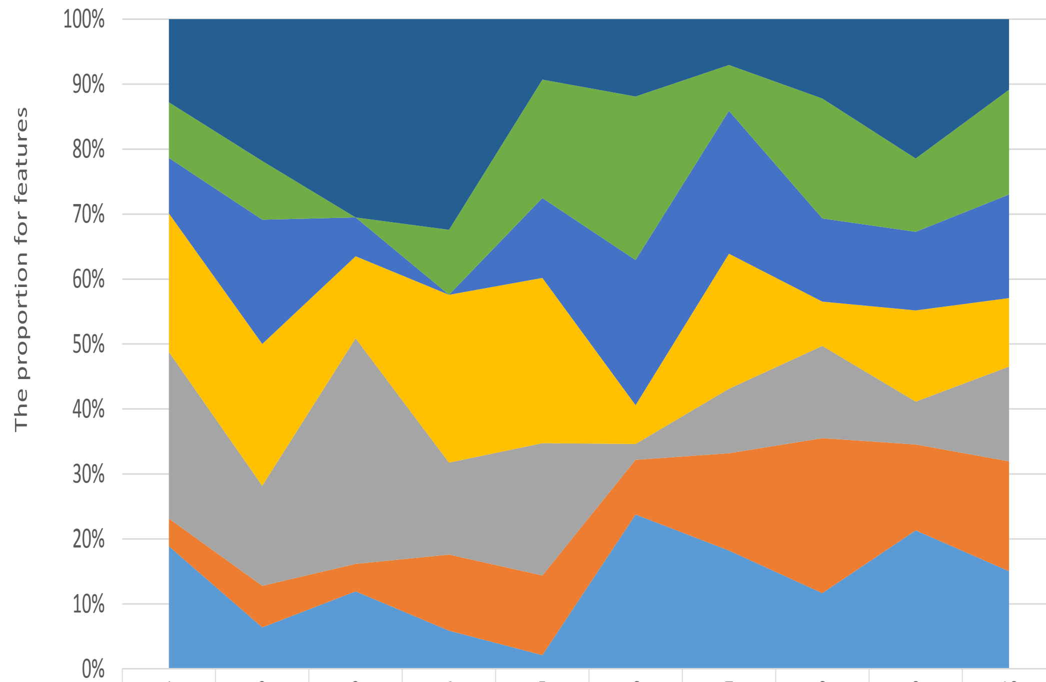


Figure 5.7: The proportion for each feature

on Facebook. We selected six types of features and left out the mutual friends feature in the experiments. The reason for this is that for each user, a few of their friends have mutual friends with the user while the rest do not have any, and therefore have the number zero. If we calculate the accuracy of the ranking with those zeroes in the rank, the result will be biased. However, even though it creates errors in the ranking, it still plays a vital role in the linear model. Table 5.2 contains the six dimensions of the basic heterogeneous data that we collected from the volunteers.

To explain the process of comparing two ranks clearly, we take one volunteer out of the ten volunteers as an example. For this volunteer, we had him or her write out a ranking of his or her top ten friends and we used that as our ground truth. We also got a ranking



	1	2	3	4	5	6	7	8	9	10
The number ofmutual friends	4.41	5.025	10.66	11.52	6.495	7.105	5.025	12.745	16.295	20.71
Comment in your photos	2.94	2.085	0	3.555	12.745	15.075	5.025	19.24	8.575	30.76
Like in your photos	2.94	4.41	2.085	0	8.575	13.355	15.69	13.355	9.19	30.395
Tag your friend in your photos	7.355	5.025	4.41	9.19	17.77	3.555	14.825	7.105	10.66	20.1
Tag you in his or her photos	8.825	3.555	12.13	5.025	14.215	1.47	7.105	14.825	5.025	27.82
Comment in your status	1.47	1.47	1.47	4.165	9	5.025	10.66	24.88	10.045	32.23
Likes in your status	6.495	1.47	4.165	2.085	1.47	14.215	12.99	12.13	16.195	28.675

The scores for features   Unit: point

Figure 5.8: The proportion for each feature

Features	Likes	Comments	TagFriend	TagUser	LikesPhotos	CommentsPhotos
Total Number	1010	465	610	2234	18531	17371

Table 5.2: The six dimensions of heterogeneous data acquired over the past half year.

of the volunteer's top ten friends from the system. To obtain the accuracy of comparing the two ranks, we analyzed the two ranks by the method we mentioned in the beginning of this section called Kendall's Tau.

To analyze the accuracy of each feature with a total of six features, we must choose the correct data from the list of the ground truth given by the volunteers and from the system, by making sure that the features have the same names listed. For example, one of the features from which we have extracted all data was the feature of the number of likes given to the users status messages. We also have the users entire friend list from Facebook as well as the list of ten friends written by the user, regarded as the ground truth. Then we can rearrange the order of the ten friends according to the number of likes for each name in the list in order to get the new list of the feature of giving likes on status messages. By comparing the new list with the ground truth, we will get the accuracy of this feature. Moreover, the data of giving likes and giving comments to the users status messages are collected for six months; however, the other features such as tagging friends and giving a like to the users photos are related to the number of photos they posted and not related to time, which means that without this time limit, we can get larger amounts of data than the two features of comments and likes to the user's status messages, that do have the time limit. We thus can learn that the accuracy of the ranks can be influenced by the amount of data collected

For the rest of the five features, the method to find the accuracy is the same. Figure 5.9 indicates the achieved accuracy when comparing the ranking from each individual feature as well as the ranking from all available data variables to the ground truth ranking given by volunteers. The result of analyzing the ranking accuracy is summarized in Table 5.3. The table shows that the ranking accuracy for the linear model built from the results of the survey as coefficients is better than the accuracy of every single feature.

In detail, the first line introduces the average accuracy for each feature involving the

Features	Likes	Comments	TagFriend
Average Accuracy	0.4233	0.4945	0.5374
Standard Deviation	0.1035	0.1566	0.1801
Improvement	38.74%	18.76%	9.29%
Features	TagUser	LikesPhotos	CommentsPhotos
Average Accuracy	0.5295	0.5299	0.5487
Standard Deviation	0.1374	0.1794	0.1346
Improvement	10.92%	10.83%	7.04%

Table 5.3: The result ranking accuracy

data of ten volunteers in total. The second line indicates the influence of each feature by comparing the average accuracy of each feature with the average accuracy of the linear model to predict the tie strength of the interaction frequency. It clearly shows that the feature of likes to status message has the highest percentage, with 38.74%. On the other hand, we see the smallest percentage for the feature of comments to the user's photos, which is 7.04%, which means that the ranking coming from this feature is closer to the final result. This means that this feature of comments made to the users photos plays a more vital part than the other features in building the tie strength of interaction frequency. However, the number of likes and comments to the user's status messages has the least influence.

Overall, the average improvement for all features is 15.93%, we used the equation as shown.

$$Average = \frac{M_1 + M_2 + ... + M_i}{i}$$

Then we can get

$$Average = \frac{38.74\% + 18.76\% + 9.29\% + 10.92\% + 10.83\% + 7.04\%}{6} \tag{5.6}$$

The final result is shown.

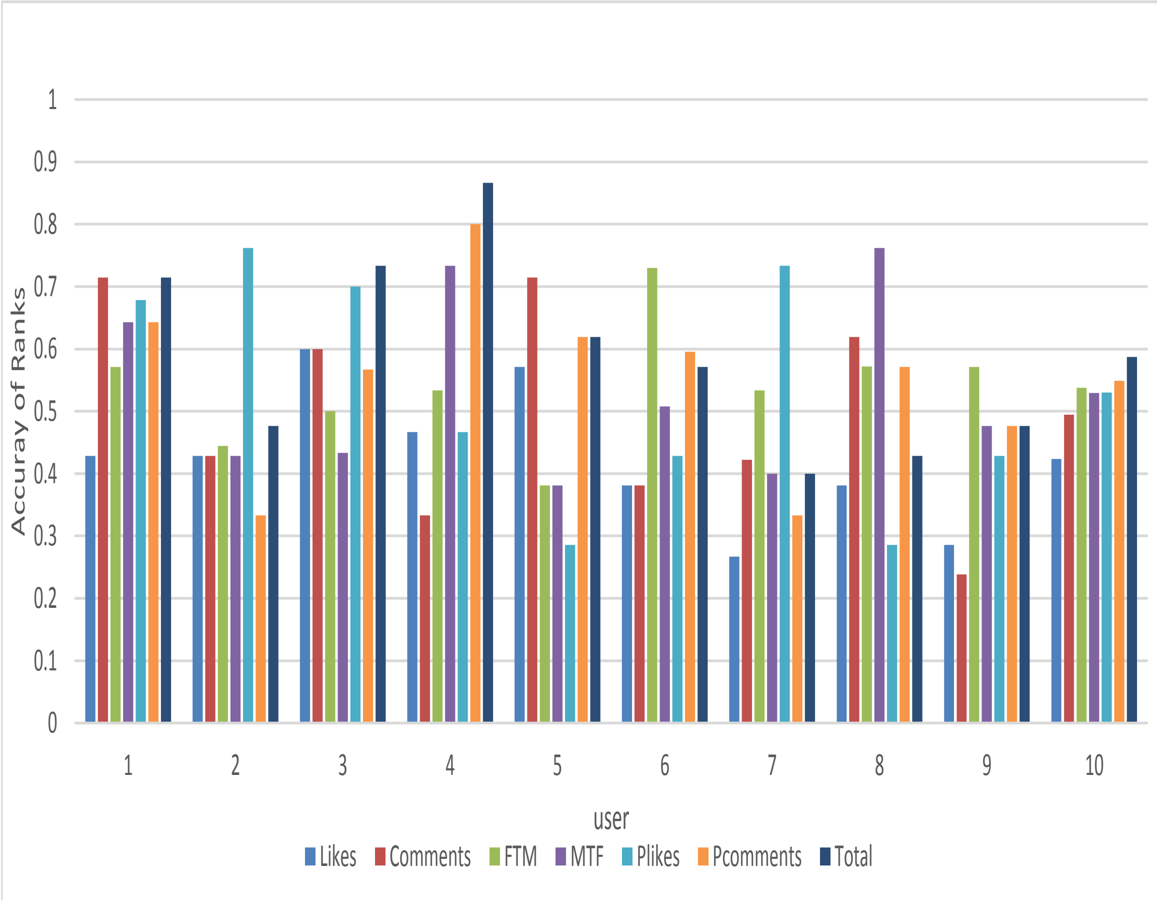


Figure 5.9: The example for the rank

$$Average = \ 15.93\%$$

The final result shows the linear model that is used to measure the interaction frequency can give a more explicit accuracy of the ranking of the user's friends than the accuracy of each single feature.

5.3 Conclusion

In conclusion, we introduced the chapter five in three parts. Firstly, it explain how do we do a survey on line and the process of obtaining the coefficients from the survey. The second part includes the process in which we do a experiment to compare two ranks by the method of Kendall's Tau. At last, we get the result of the experiment, which shows the linear model that is used to measure the interaction frequency can give a more explicit accuracy of the ranking of the user's friends than the accuracy of each single feature. The average improvement for all features is 15.93%.

Chapter 6

Conclusion and Future Work

We have presented our research on estimating and visualizing the interactions between a user and their friends on the social network Facebook. Heterogeneous data was obtained from the social network and categorized into seven dimensions of input data. The seven sets of data were combined into our estimation of interaction frequency by employing a linear model whose coefficients were determined from the results of an online survey. The sets of input data and the estimated interaction frequencies were aggregated into a new set of four variables for visualization. These variables were then mapped to the physical parameters of a solar system. With the physical parameters calculated, a user's view of the social network was then visualized with the concept of a solar system. By comparing our interaction frequency with interaction rankings given by ten volunteers, we found that the use of all available data points increased the ranking accuracy by an average of 15.93% over using each data point individually.

For any future work that may be done on this project, as we know, we have found the strength of the tie of interaction frequency, and the tie is an important factor to predict the user's friends' personality, because there is related research that mentions that the behaviors of the user on a social network can predict the user's personality; therefore we can do more extensive research on this for predicting the personality of the user's friends. Additionally, we can add more information to the 3D visualization so that it can give more types of information in a three dimensional model to the user.

References

[1] solarsystem. In *http://en.wikipedia.org/wiki/Solar-System*.

[2] Craig Ross, Emily S Orr, Mia Sisic, Jaime M Arseneault, Mary G Simmering, and R Robert Orr. Personality and motivations associated with facebook use. *Computers in Human Behavior*, 25(2):578–586, 2009.

[3] Keith N Hampton, Lauren Sessions Goulet, Cameron Marlow, and Lee Rainie. Why most facebook users get more than they give. *Pew Internet & American Life Project*, 3, 2012.

[4] Company info-facebook newsroom. Facebook, February 6, 2015.

[5] Keith N Hampton, Lauren Sessions Goulet, Harrison Rainie, and Kristen Purcell. *Social networking sites and our lives: How people's trust, personal relationships, and civic and political involvement are connected to their use of social networking sites and other technologies*. Pew Internet & American Life Project, 2011.

[6] Yoram Bachrach, Michal Kosinski, Thore Graepel, Pushmeet Kohli, and David Stillwell. Personality and patterns of facebook usage. In *proceedings of the 3rd annual ACM web science conference*, pages 24–32. ACM, 2012.

[7] PT Costa Jr and Robert R McCrae. Neo personality inventory–revised (neo-pi-r) and neo five-factor inventory (neo-ffi) professional manual. *Odessa, FL: Psychological Assessment Resources*, 1992.

[8] Lewis R Goldberg. The structure of phenotypic personality traits. *American psychologist*, 48(1):26, 1993.

[9] Mary T Russell and Darcie L Karol. Administrator manual for the 16pf. *Institute for Personality and Ability Testing, Champaign, IL*, page 18, 1994.

[10] Ernest C Tupes and Raymond E Christal. Recurrent personality factors based on trait ratings. *Journal of personality*, 60(2):225–251, 1992.

[11] Yair Amichai-Hamburger and Gideon Vinitzky. Social network use and personality. *Computers in human behavior*, 26(6):1289–1295, 2010.

[12] Cong Guo, Xinmei Tian, and Tao Mei. User specific friend recommendation in social media community. In *Multimedia and Expo (ICME), 2014 IEEE International Conference on*, pages 1–6. IEEE, 2014.

[13] Pritam Gundecha and Huan Liu. Mining social media: A brief introduction. *Tutorials in Operations Research*, 1(4), 2012.

[14] Nadeem Akhtar, Hira Javed, and Geetanjali Sengar. Analysis of facebook social network. In *Computational Intelligence and Communication Networks (CICN), 2013 5th International Conference on*, pages 451–454. IEEE, 2013.

[15] Minas Gjoka, Maciej Kurant, Carter T Butts, and Athina Markopoulou. Walking in facebook: A case study of unbiased sampling of osns. In *INFOCOM, 2010 Proceedings IEEE*, pages 1–9. IEEE, 2010.

[16] Networking group, wiki page,uci. In *http://odysseas.calit2.uci.edu/doku.php/public:online-social-networks*.

[17] Jinfeng Zhuang, Tao Mei, Steven CH Hoi, Xian-Sheng Hua, and Shipeng Li. Modeling social strength in social media community via kernel-based learning. In *Proceedings of the 19th ACM international conference on Multimedia*, pages 113–122. ACM, 2011.

[18] George G Robertson, Jock D Mackinlay, and Stuart K Card. Cone trees: animated 3d visualizations of hierarchical information. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 189–194. ACM, 1991.

[19] Tarik Crnovrsanin, Chris W Muelder, Robert Faris, Diane Felmlee, and Kwan-Liu Ma. Visualization techniques for categorical analysis of social networks with multiple edge sets. *Social Networks*, 37:56–64, 2014.

[20] Georgios A Pavlopoulos, Seán I O'Donoghue, Venkata P Satagopam, Theodoros G Soldatos, Evangelos Pafilis, and Reinhard Schneider. Arena3d: visualization of biological networks in 3d. *BMC systems biology*, 2(1):104, 2008.

[21] . Michael VanDaniker, Andrew Lund and Douglas Hamilton. Solar system visualizer. In *https://janus.astro.umd.edu/SolarSystems/*.

[22] Mark Allen. Restfb. *RestFB-A Lightweight Java Facebook Graph API and Old REST API Client*, 2012.

[23] Hemant Kumar Kushwaha and J Jeysree. Personalized recommender system. *International Journal of Computer Applications*, 93, 2014.

[24] Diego Terrana, Agnese Augello, and Giovanni Pilato. Facebook users relationships analysis based on sentiment classification. In *Semantic Computing (ICSC), 2014 IEEE International Conference on*, pages 290–296. IEEE, 2014.

[25] T Higgins. Unity-3d game engine, 2010.

[26] Sa Wang, Zhengli Mao, Changhai Zeng, Huili Gong, Shanshan Li, and Beibei Chen. A new method of virtual reality based on unity3d. In *Geoinformatics, 2010 18th International Conference on*, pages 1–5. IEEE, 2010.

[27] Nicoletta Adamo-Villani, Edward Carpenter, and Laura Arns. 3d sign language mathematics in immersive environment. *Proc. of ASM*, pages 2006–2015, 2006.

[28] Brian J Gold, Marc Pomplun, Nichola J Rice, and Robert Sekuler. A new way to quantify the fidelity of imitation: preliminary results with gesture sequences. *Experimental brain research*, 187(1):139–152, 2008.

[29] The PI's Perspective. Today we know of more than a dozen dwarf planets in the solar system.

[30] IAU. The final iau resolution on the definition of "planet" ready for voting. 2007-03-02.

[31] Michael E Brown, Chad Trujillo, and David Rabinowitz. Discovered by.

[32] Ron Ekers. Iau planet definition committee. *International Astronomical Union. Retrieved*, pages 10–13, 2008.

[33] Paris International Astronomical Union. Plutoid chosen as name for solar system objects like pluto. International Astronomical Union, 2008.

[34] Isaac Newton. In experimental philosophy particular propositions are inferred from the phenomena and afterwards rendered general by induction, 1729.

[35] Isaac Newton. *The principia: mathematical principles of natural philosophy.* Univ of California Press.

[36] Laurent Hodges. The michell-cavendish experiment, 1998.

[37] Peter J Mohr, Barry N Taylor, and David B Newell. Codata recommended values of the fundamental physical constants: 2006a). *Journal of Physical and Chemical Reference Data*, 37(3):1187–1284, 2008.

[38] introduction. In *http://gravitee.tripod.com/axioms.htm*.

[39] Isaac Newton. The principia, translated by andrew motte, 1995.

[40] Michael E Browne and Michael E Browne. *Schaum's Outlines: Physics for Engineering and Science.* McGraw-Hill, 2010.

[41] Steven Holzner. *Physics I for dummies.* John Wiley & Sons, 2011.

[42] John L Russell. Kepler's laws of planetary motion: 1609–1666. *The British Journal for the History of Science*, 2(01):1–24, 1964.

[43] Cong Guo, Xinmei Tian, and Tao Mei. User specific friend recommendation in social media community. In *Multimedia and Expo (ICME), 2014 IEEE International Conference on*, pages 1–6. IEEE, 2014.

[44] Comparing facebook and twitter. In *http://blogs.dw.de/innovation/world-cup-special/*.

[45] gestalt psychology. In *http://baike.baidu.com/view/73571.htm.reference-[11]-73571-wrap*.

[46] Alvaro Ortigosa, Rosa M Carro, and José Ignacio Quiroga. Predicting user personality by mining social interactions in facebook. *Journal of Computer and System Sciences*, 80(1):57–71, 2014.

[47] Morgan Ames and Mor Naaman. Why we tag: motivations for annotation in mobile and online media. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 971–980. ACM, 2007.

[48] Keith N Hampton, Lauren F Sessions, and Eun Ja Her. Core networks, social isolation, and new media: How internet and mobile phone use is related to network size and diversity. *Information, Communication & Society*, 14(1):130–155, 2011.

[49] Keith N Hampton, Lauren F Sessions, Eun Ja Her, and Lee Rainie. Social isolation and new technology. *Pew Internet & American Life Project*, 4, 2009.

[50] Karel Hrbacek and Thomas Jech. *Introduction to Set Theory, Revised and Expanded*, volume 220. Crc Press, 1999.

[51] Yadolah Dodge, David Cox, Daniel Commenges, Patricia Joy Solomon, Suzan Wilson, et al. *The Oxford dictionary of statistical terms*. Oxford University Press, 2003.

[52] Eric Gilbert and Karrie Karahalios. Predicting tie strength with social media. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 211–220. ACM, 2009.

[53] Christo Wilson, Bryce Boe, Alessandra Sala, Krishna PN Puttaswamy, and Ben Y Zhao. User interactions in social networks and their implications. In *Proceedings of the 4th ACM European conference on Computer systems*, pages 205–218. Acm, 2009.

[54] Scott Garriss, Michael Kaminsky, Michael J Freedman, Brad Karp, David Mazières, and Haifeng Yu. Re: Reliable email. In *NSDI*, volume 6, pages 22–22, 2006.

[55] Haifeng Yu, Michael Kaminsky, Phillip B Gibbons, and Abraham Flaxman. Sybilguard: defending against sybil attacks via social networks. In *ACM SIGCOMM Computer Communication Review*, volume 36, pages 267–278. ACM, 2006.

[56] Paul T Costa and Robert R McCrae. Neo pi-r professional manual. 1992.

[57] Russelkl C Hibbeler. Equations of motion: Normal and tangential coordinates. In *Engineering Mechanics: Dynamics (12 ed.)*. Prentice Hall. p. 131. ISBN 0-13-607791-9, 2009.

[58] Gene Mosca Paul Allen Tipler. *Physics for scientists and engineers (5th ed.)*. Macmillan. p. 129. ISBN 0-7167-8339-8, 2003.

[59] M. Piau P. Germain and D. Caillerie. Theoretical and applied mechanics. Elsevier. ISBN 9780444600202, 2012.

[60] Harold Hilton. *Plane Algebraic Curves*. Oxford. p. 69, 1920.

[61] Pavlyk Oleksandr Bryant, Jeff. *Kepler's Second Law*. Wolfram Demonstrations Project Retrieved, 2009.

[62] Stephen G Holton, Gerald James; Brush. *the Human Adventure: From Copernicus to Einstein and Beyond (3rd paperback ed.)*. Rutgers University Press. pp. 40?1. ISBN 0-8135-2908-5, 2001.

[63] Frederick James. *Statistical methods in experimental physics*, volume 7. World Scientific Singapore, 2006.

[64] Ravi Kumar and Sergei Vassilvitskii. Generalized distances between rankings. In *Proceedings of the 19th international conference on World wide web*, pages 571–580. ACM, 2010.