Industry Report

Wei Zang\_30491754

TABLE OF CONTENT

[1.0 Introduction, background and ethics 2](#_Toc3123)

[3.0 Findings and methodology 4](#_Toc23216)

[3.1 Methodology 4](#_Toc14508)

[3.2 Findings 7](#_Toc20050)

[3.2.1 Social network analysis 7](#_Toc6851)

[3.2.2 Sentiment analysis 11](#_Toc30900)

[3.2.3 Google search analysis 14](#_Toc4805)

[4.0 Recommendation and conclusion 16](#_Toc28627)

[References 18](#_Toc16917)

# 1.0 Introduction, background and ethics

The purpose of this report is mainly use social network analysis to analyze two business figures that have a huge influence in both the technology circle and the business circle on Twitter: the new era technology icon: Elon Musk, and the godfather of Silicon Valley: Paul Graham.

Why should they be compared?

Well, because Graham was an entrepreneur who created one of the first mass scales e-commerce platforms called Viaweb which sold to Yahoo and became Yahoo Stores. His entrepreneurial background directly resulted in his ability to help founders at the earliest possible stage. His second startup in a lot of ways was the investment firm Y Combinator, which itself has helped thousands of founders and is also the most successful incubator ever created with more than $100B in market value. Most VCs work with founders well after the startup has taken off, but YC helps founder’s years before that happens, often when it is just a few people and an idea.

Musk is great too but has not been nearly as active or focused on helping early stage founders by any measure. Musk was born in Pretoria, the administrative capital of South Africa. He has dual citizenship in Canada and the United States. He is an entrepreneur, engineer, and philanthropist. As well as the CEO and CTO, CEO and product architect of Tesla, Chairman of the Board of Directors of SolarCity.

They are both top entrepreneurs and entrepreneurs, but they have very different styles: both are technical entrepreneurs, but the former is more conservative (compared to musk), and the latter is more adventurous and radical.

They are all very active on Twitter. It will be very interesting to compare their interactions with audiences (technical circles, business circles) and users, and they have learned a lot of marketing and communication strategies.

Now is the age when the founder is the brand itself. The company brand and audience can be directly and seamlessly connected through Twitter/Facebook/Instagram. Considering their millions of followers (Graham) and tens of millions of followers (Musk), commercial promotion and brand establishment have already been socialized. The network is deeply tied together. Social networks have removed some middlemen, and made some ups and downs and uninteresting publicity obsolete. Now only attractive and viable founders can capture the audience’s attention

Ethical principles build a strong foundation for driving ethical technologies. Principles alone can be elusive and impractical for application. Ethical frameworks based upon these principles provide a structure to guide technologists when implementing data-driven solutions.

We mainly use Networkx to draw the relationship diagram, meanwhile we use Matplotlib and Numpy are used for data analysis and preprocessing. They use the python built-in library for data cleaning and simple pipe connection and processing. The data download is obtained through twitter API and my own crawler program.

**2.0 Background of twitter account**

****

*Figure 1: Musk's Twitter*



*Figure 2: Graham's Twitter*

# 3.0 Findings and methodology

## 3.1 Methodology

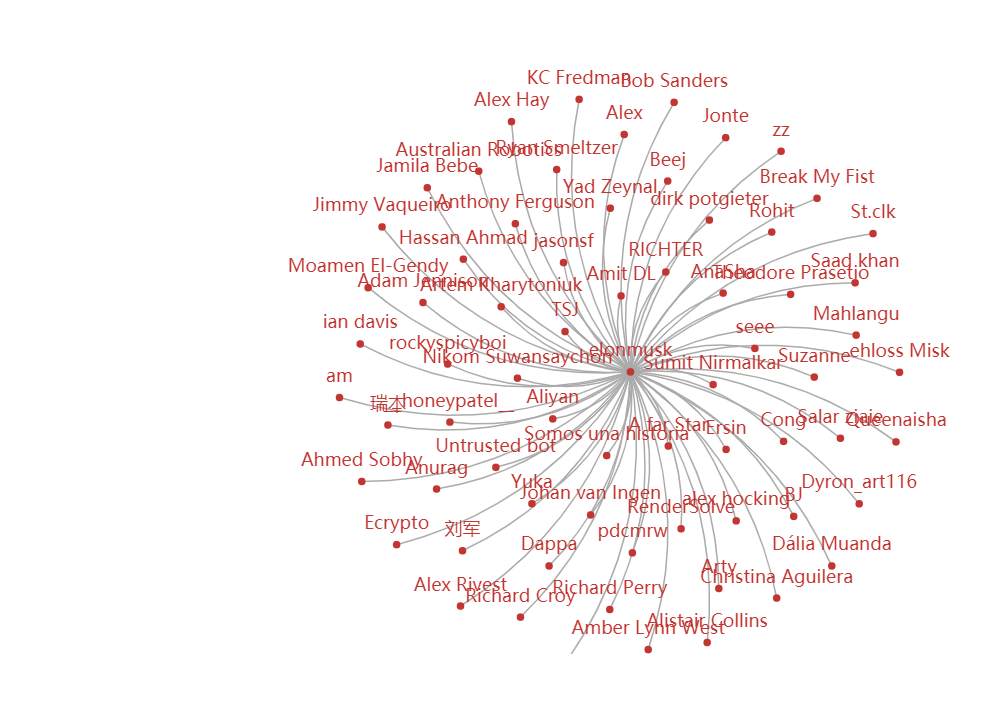
**Gephi**

In Gephi, there are generally two data for network diagrams: edge files and point files.

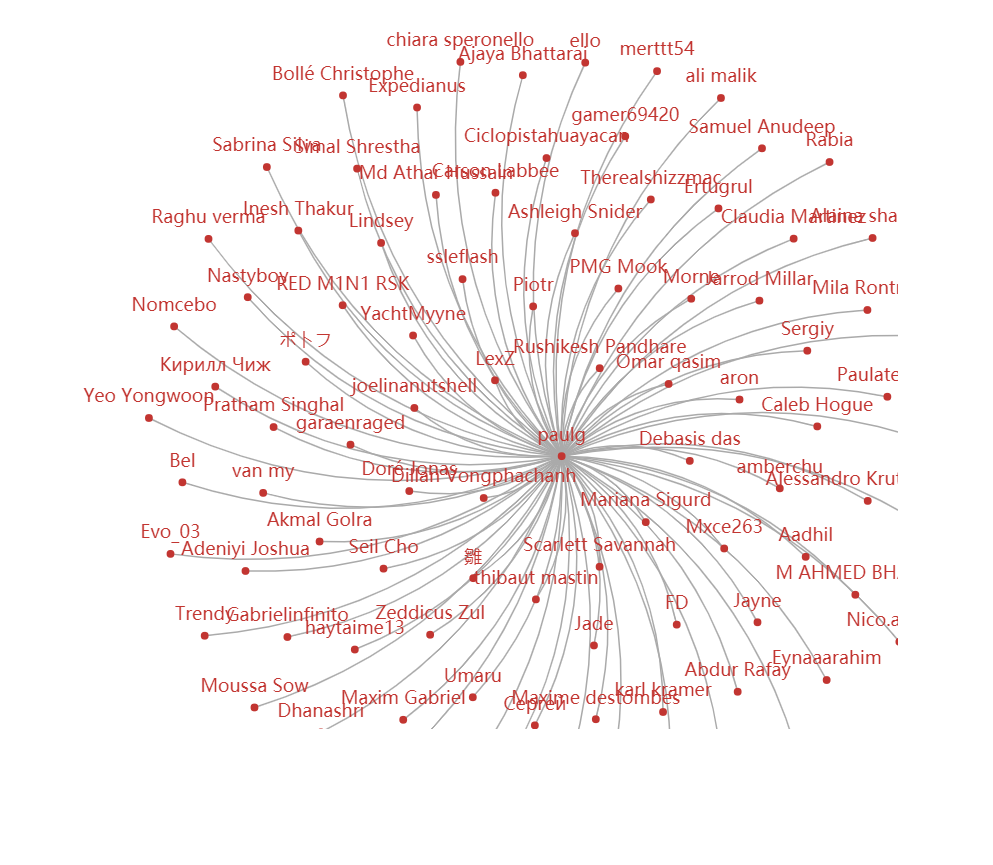
The edge file, as shown in the figure below, is mainly composed of a pair of node IDs. The two points of Sourse and Target are connected to form a line (edge). The edge file can also add edge description information, such as edge id, function type, weight Wait.

The color and size adjustment of nodes and edges can use the "Appearance" area on the upper left. For example, for nodes, you can choose the four adjustments marked, namely node color, node size, label color, and label size. Gephi can not only change the node and label to the same color and size, but also set the color and size of the gradient according to the degree of the node. It also supports displaying different colors according to other attributes of the imported node. The edge can also be similarly set.

We use Gephi to describe the relationship between elon/paulG and their followers, to view simple relationships , and then we will use networkx to draw more complex network models (thousands of edges) based on forwarding and interaction.



The relationship of elonmusk and followers



The relationship of paulG and his followers

After we downloaded the tweets data of the followers, we continued to use gephi to describe the interaction patterns between the followers and visualize them. The pictures are as follows:

Networkx

**图像**

As an open source package of Python, Networkx is convenient for users to create, operate and learn complex networks. Using Networkx, we can store networks in standardized and non-standardized data formats, generate a variety of random networks and classic networks, analyze network structures, establish network models, design new network algorithms, and perform network drawing, etc. In Networkx, there are the following 4 basic graph types:

1. Graph: Refers to the undirected graph, that is, the direction of the edge between two nodes is ignored.
2. DiGraph: Refers to a directed graph (directed Graph), which considers the directionality of the edges.
3. MultiGraph: Refers to multiple undirected graphs, that is, the number of edges between two nodes is more than one, and the vertex is allowed to be related to itself through the same edge.
4. MultiDiGraph: a directed version of the multigraph

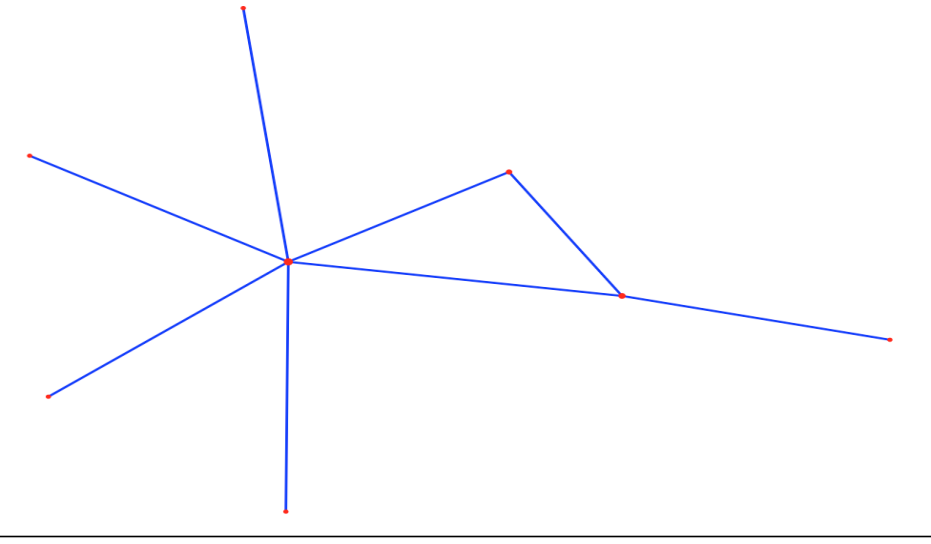
All graph classes allow Hashable objects as nodes. Hashable objects include strings, tuples, integers, etc. Any edge attributes (such as weights and labels) can be associated with edges.

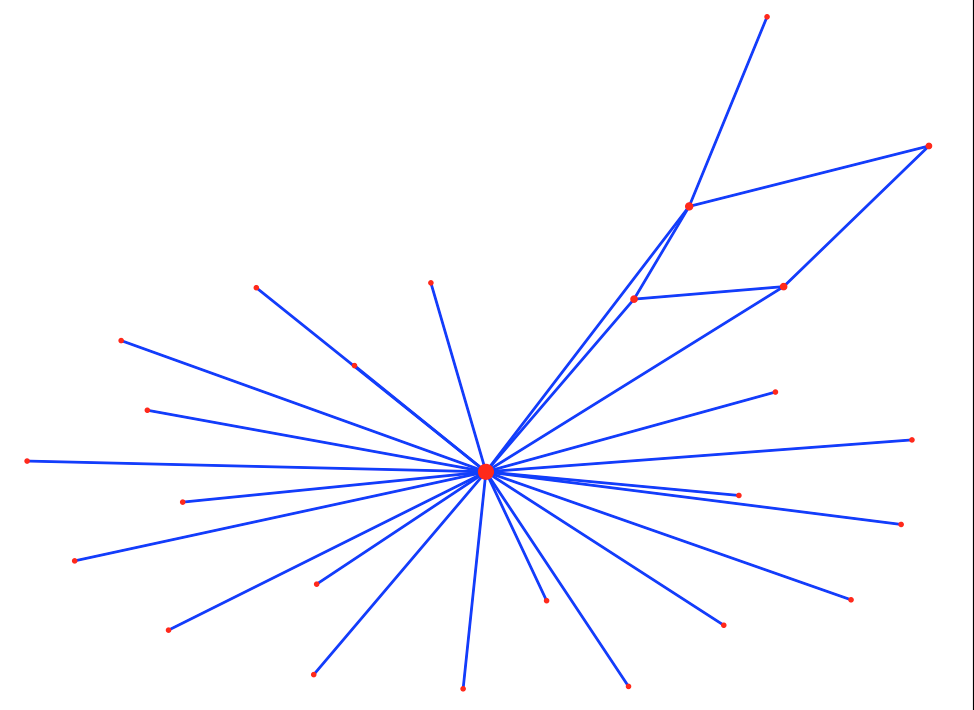
The internal data structure of the graph is based on the adjacency list representation, and the dictionary data structure is implemented using python. The graph adjacency structure is implemented as a python dictionary; the external dictionary is keyed by nodes to values, these values themselves are dictionaries, and the adjacent nodes are keyed to the edge attributes associated with the edge. This "Dict of Dicts" structure allows quick addition, deletion and search of nodes and neighbors in large graphs. Directly access the underlying data structure through the methods in the class definition (programming interface "API"). On the other hand, all functions only use these API methods instead of directly acting on data structures to manipulate graphics-like objects. This design allows replacing the Dicts-based "Dicts of Dicts" data structure with alternative data structures that implement the same method.

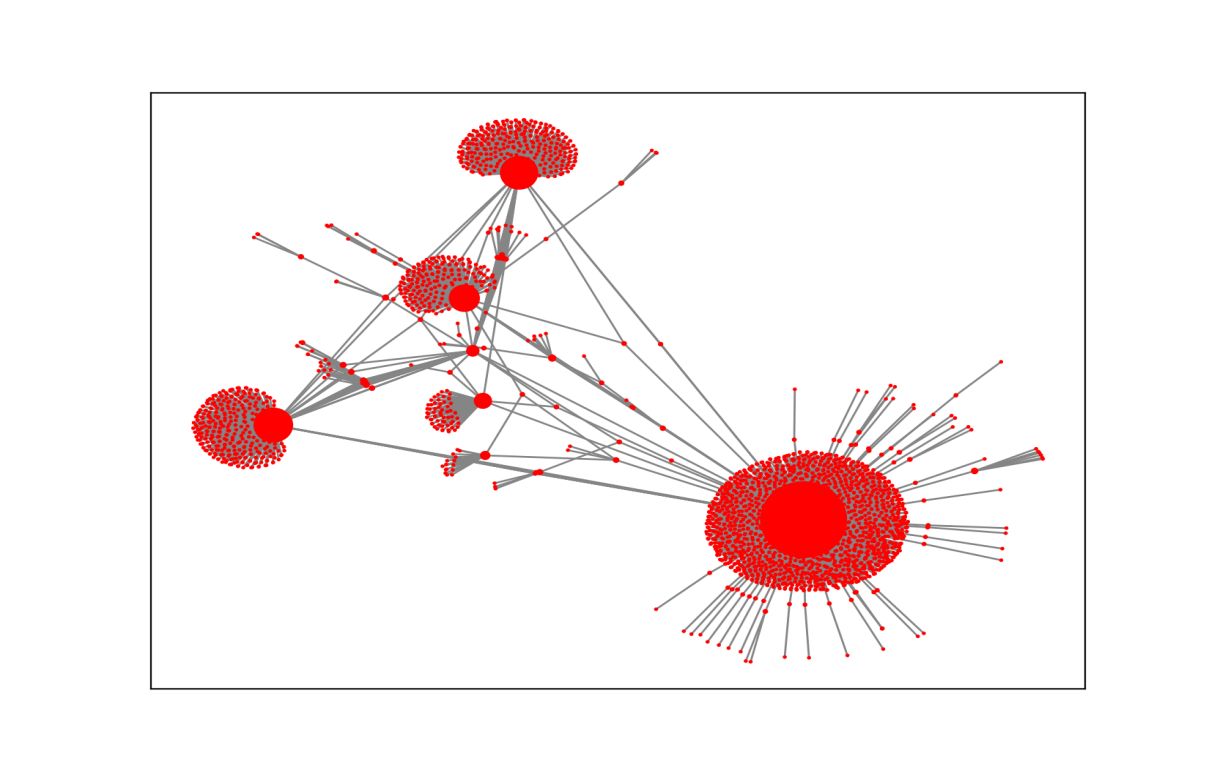
3.2 Findings

### 3.2.1 Social network analysis

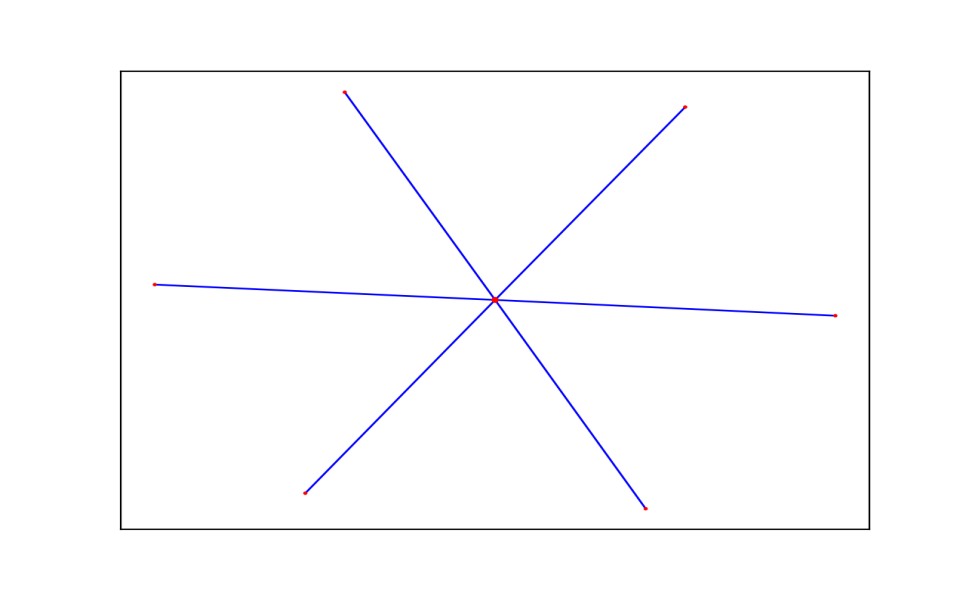
First let us see Musk’s week/month/month social network analysis

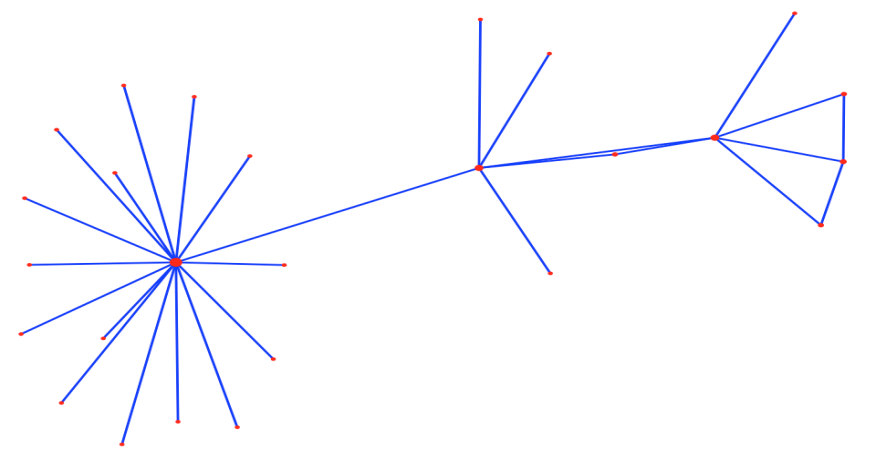
****

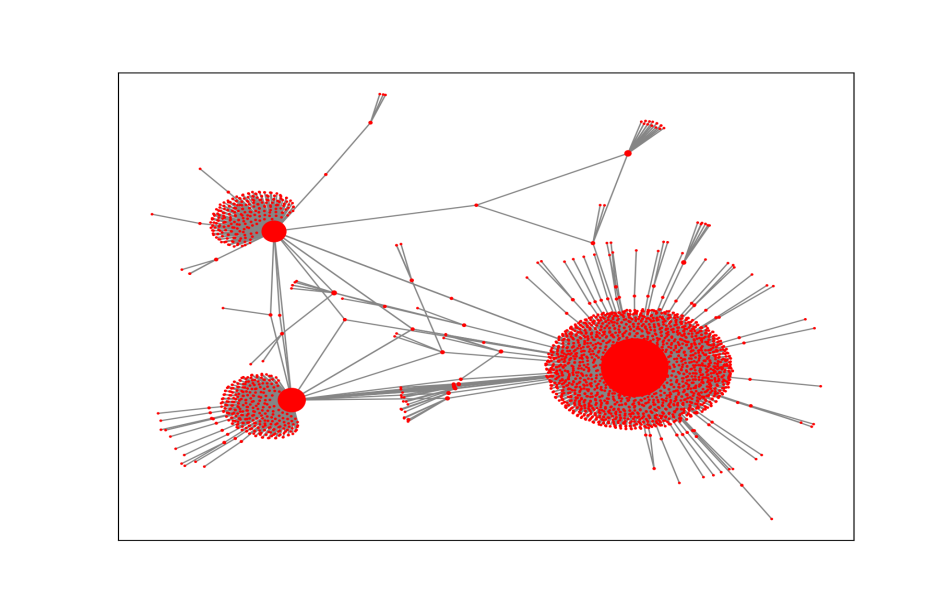
****



Then let use see Paul Graham ’s week/month/month social network analysis



****

****

With the advent of the era of big data, people's awareness of accumulating data has become stronger and stronger, so how to obtain the information we want from the accumulated data has attracted more and more attention. Public comments on social networks, Twitter or Facebook, company e-mail records, disease development, website visits, etc. More and more data that can be connected to the network is saved, so learning to analyze these data can help us find the information we need in the messy connection relationship

We first use Networkx to grab the data of Twitter to obtain the key data of 2 on Twitter. The results show that the network density is 0.0904, indicating that the actors in the network have a certain degree of relevance. There are a total of all nodes in the network. There are 3287 lines, the longest distance between nodes is 6, and that is, the longest distance between any two nodes is 6 lines. The distance reflects the length of the interval between actors, and the smaller the distance, it means that each other the closer the connection. In other words, on Twitter, the media organizations can be connected to each other through up to 5 nodes, which is less than the average distance between any two people in the world that is only 5 nodes apart from the "six degrees of separation" theory emphasizes. In addition, the average distance between the network nodes is 2.937, which means that any two media agencies only need to pass through the intermediary of 1 to 2 nodes to become the interconnected nodes in the network.

The Graham’s network’s more then 8000 edges, meanwhile the Musk’s network’s 5000 edges. The dataset we use is about year of 2020 till now.

If we regard Social Network Analysis as an application as a research method, sometimes it can also be used as a research perspective (perspective). Of course, some middle-level theories have also been produced. The more common ones are Granovetter's weak connection theory, Burt's structural whole theory, Watts' small world model, and Barabasi's Power Law.

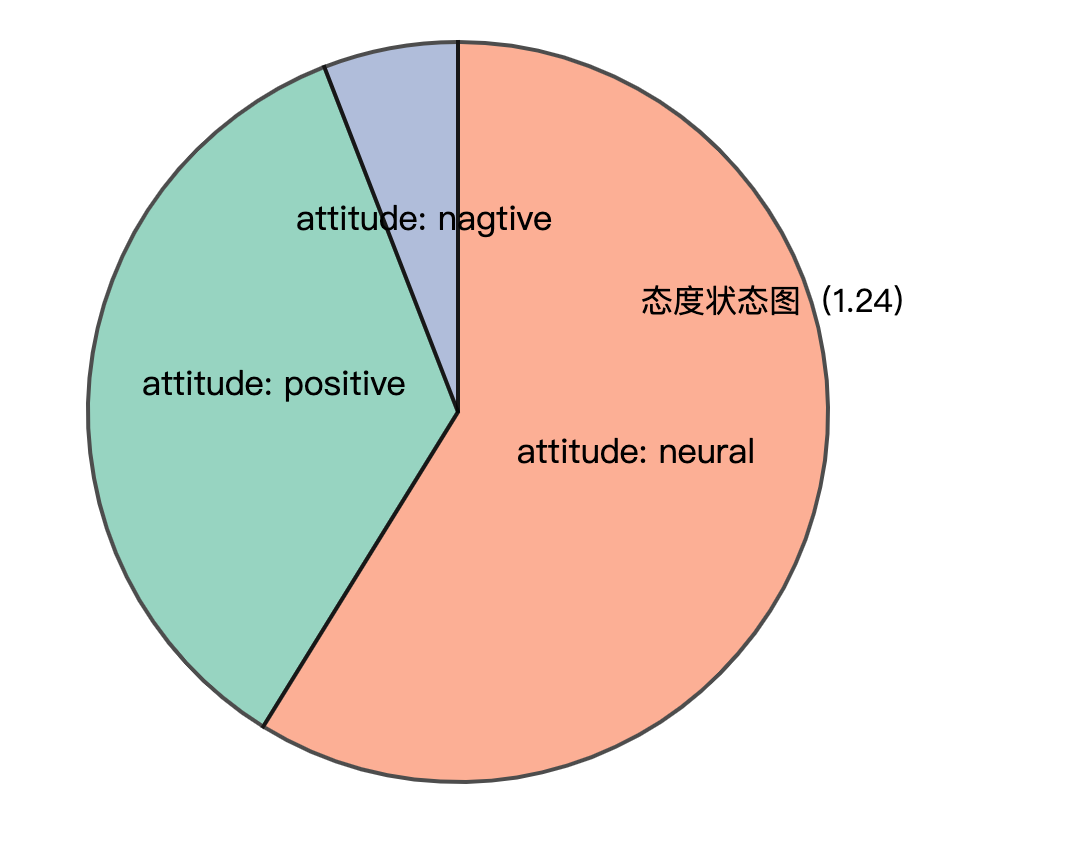
The previous social sciences often focused on the characteristics of individuals (or actors, such as companies, individuals), and ignored the relationships between individuals. The study of social networks is precisely the method and perspective of studying relationships. The biggest feature is that it considers the interdependence between individuals, which is closer to the real society. Displaying these relationships with the pictures shown in the title, you can intuitively see the position of each actor in the network and the overall structure of the network looks very dazzling and very advanced.

A network can be defined from the following three aspects:

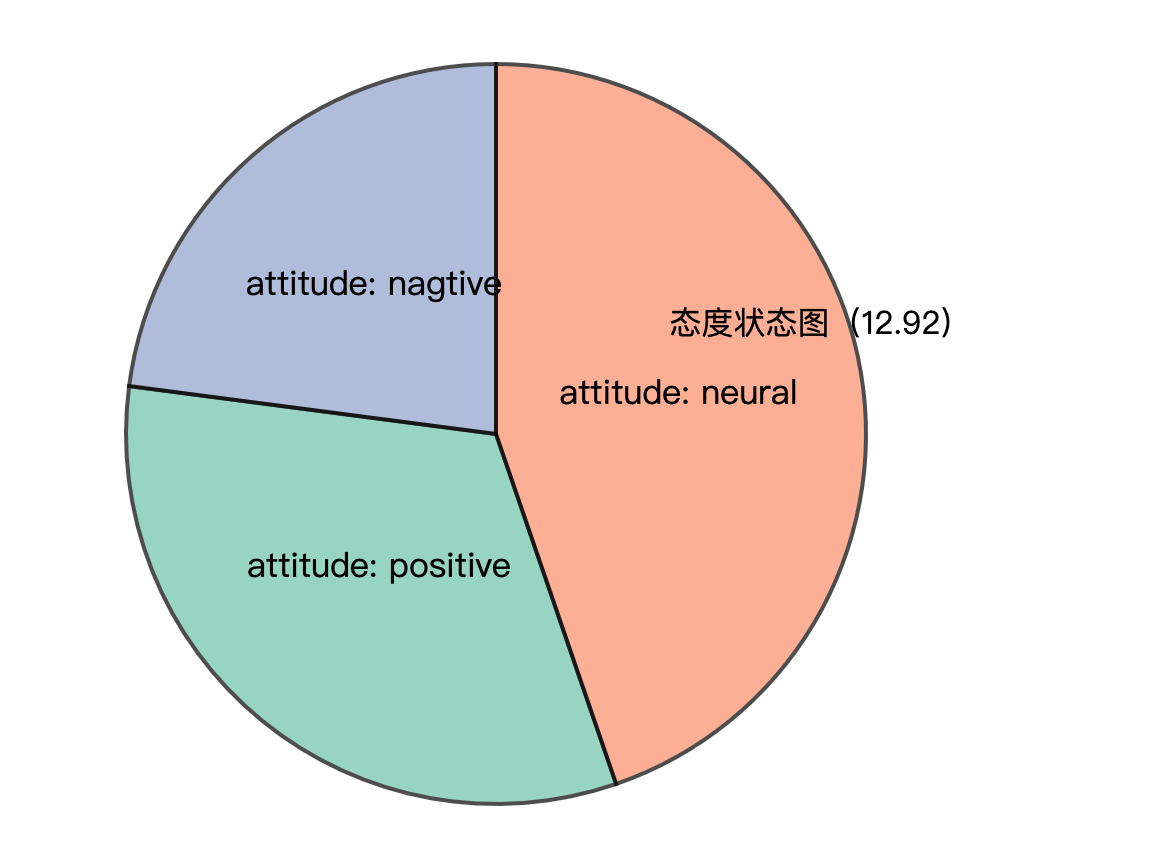
1. The actors. Musk communicates and interact with more brands: like Testla, SpaceX and bitcoin origination. Meanwhile the Graham is more focused on Investor organizations and COVID-19 related organizations.
2. Contact (ties). Edges are often used in graphics. For example, the friendship between people and the bloc chain community relationship, they following and communicate more with Musk. Cooperation between enterprises and enterprises. The hostile relationship between state and state seems communicate with Graham's more often.
3. Boundary. These actors and sides need to define a boundary, such as users in block chain or Astronomy enthusiast /tesla buyer, they are more liked to interact with Musk. Meanwhile Investors, entrepreneurs who want to get investment, communicate and interact more with the Graham.

### 3.2.2 Sentiment analysis

Figure 3 below is the sentiment analysis statistics of Musk’s sent tweets and reposts, and Figure 4 is the sentiment analysis statistics of Graham's sent tweets and reposts

****

*Figure 3:Sentiment analysis statistics of Musk’s sent tweets and reposts*

****

*Figure 4:Sentiment analysis statistics of Graham's sent tweets and reposts*

The explosive development of social network services such as Twitter has also brought great opportunities for researchers. Researchers can analyze the public's emotional changes by analyzing a large amount of emotional data, and have an impact on government management, economics, and entertainment. From the perspective of the government and managers, the United Nations has developed an application for global emotional fluctuation monitoring, Global Pulse, and Beihang’s research team launched the first online emotional system for Chinese Weibo, MoodLens; Romney and Obama during the 2012 US general election Intense propaganda was launched on Twitter to influence ordinary people and journalists, becoming a typical case of Internet participation in the presidential election. In terms of financial applications, many research institutions apply sentiment analysis technology to stock analysis and forecasting systems. For example, Stock Sonar displays daily positive and negative sentiment information for the stock next to the price of each stock, providing investors with instant information For reference, the Twitter sentiment analysis developed by UIC is used to predict and track the rise and fall of the stock market. In the entertainment field, Alibaba Cloud’s artificial intelligence system "Little Ai" successfully predicted Coco Lee's victory in the "I Am a Singer" program. It also relied on live data and comment data on social networks to analyze and predict text sentiment analysis technology. It can be seen that the monitoring, analysis and prediction of social media's emotional big data continue to influence government decision-making and public choice.

The time when the text is published can usually be obtained using simple rules. Therefore, the goal of sentiment analysis is usually to automatically analyze the three elements of Holder, Target, and Polarity from unstructured text. Holder is the issuer of the opinion; Target is the object of the opinion evaluation (such as the entity or entity's attributes, or topic); Polarity is the emotion type expressed. Due to different tasks, the emotion category system will be different, usually including praise and criticism. Joy, Anger, Sorrow, Joy, Sorrow, and emotional scoring (such as 1-5 points) classification system emotions in the text are divided into explicit emotions and implicit emotions. Explicit emotions refer to emotional texts that contain obvious emotional words (such as happy and beautiful), and implicit emotions refer to emotional texts that do not contain emotional words, such as "this There is a layer of ash on the table." Since implicit sentiment analysis is more difficult and relies on background knowledge and common sense knowledge, many current works focus on the research of revealed sentiment analysis. In this report, we are just use revealed sentiment analysis of Polarity.

We do it at the sentence level, it determines whether its sentiment is positive, negative or neutral for each sentence of the document. Unlike the previous document level, some descriptive sentences actually have no emotion, so there is a neutral category here, which means there is no emotion. This question has something to do with subjectivity classification-subjectivity classification is to judge whether a sentence is subjective or objective. Usually neutral sentences are objective, while positive or negative sentiment sentences are subjective. But they are not exactly the same. For example, "We bought the car last month and the windshield wiper has fallen off" is an objective sentence, but it describes an undesirable thing, so it implies negative emotions. Although the sentence "I think he went home after lunch" is subjective, it has no positive or negative emotions

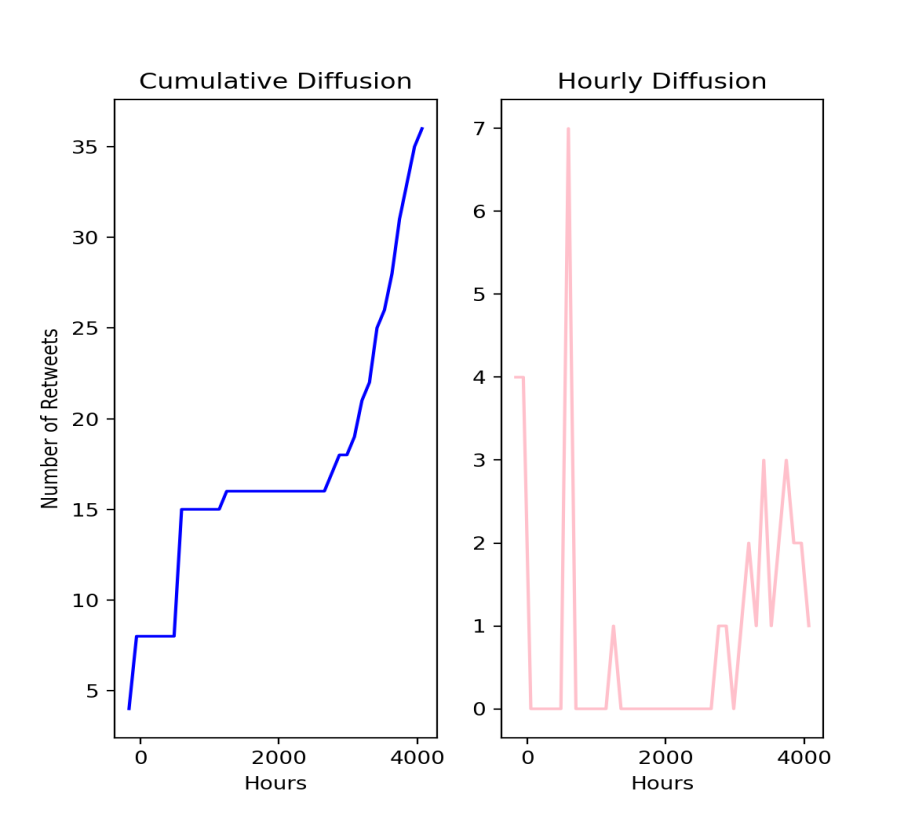
We can see:

Musk has or retweet more positive tweets than Graham, but in common, they are on the Internet, and most topics (more than 1/2) are neutral.

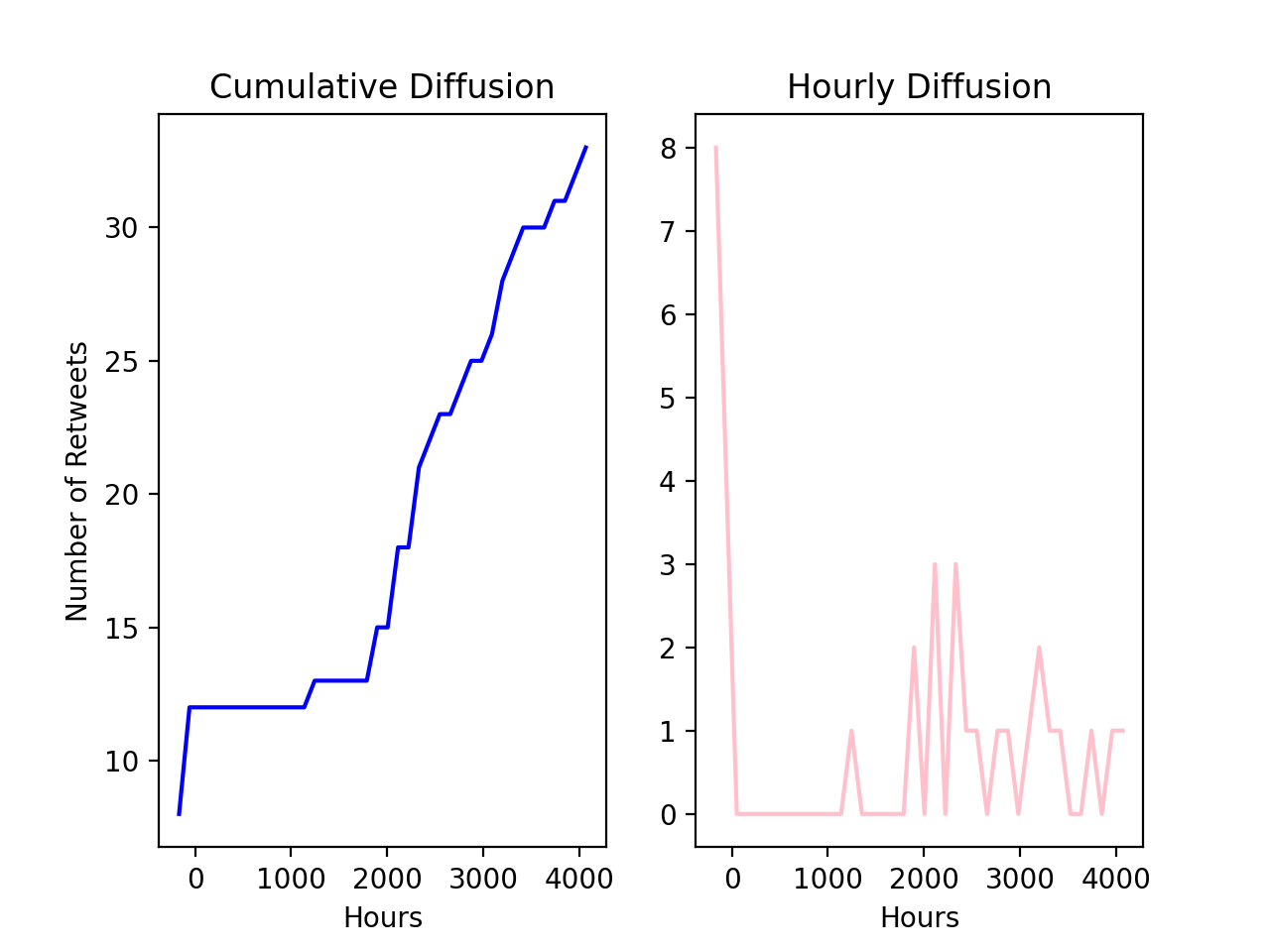
### 3.2.3 Google search analysis

The “Audie

Report in Google Analytics can view the details of visitors. This information can help us tailor content to attract audiences. If we use paid advertising, we can use this report to optimize our targeting .First we analysis their retweet rules from the time and frequency. Then we count the most send / retweet topic. We then search topics on google trends to find the most relevant.

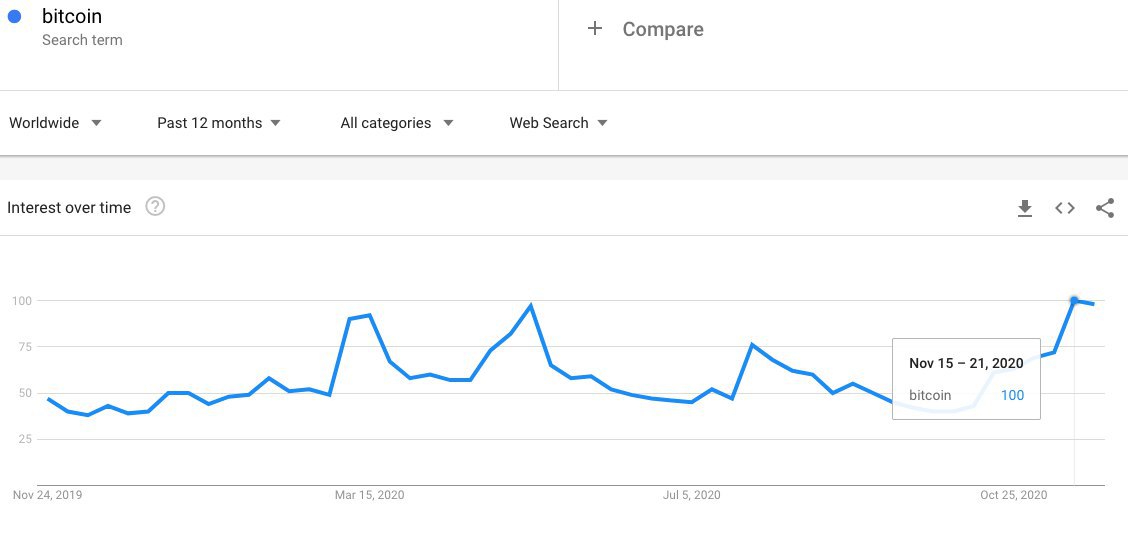


*Figure 5: Musk's frequency of send*

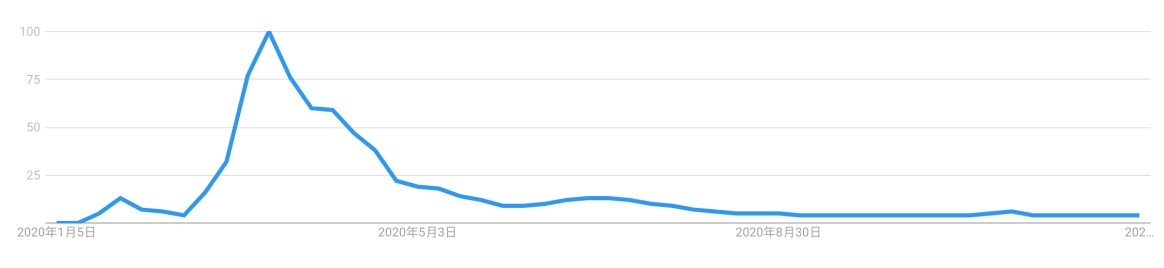


*Figure 6:Graham's frquency of send*

Figure 7 below is the Mush’s most relevant google search trend and Figure 8 is the the Graham’s most relevant google search trend.

****

*Figure 7:Musk's most relevant google search trend*

****

*Figure 8: Graham's most relevant google search trend*

Musk’s tweet interaction rules are more random, but there are no loosening rules for a few months in the middle, and the popularity is not affected; and Graham uses tweets to interact with the audience with similar probability throughout the year.

Through the Google search correlation analysis, it can be seen that the main audience of Musk is most concerned about the bitcoin price and block chain-related news.

Graham’s audience is most relevant to some news and discussions about the new crown

# 4.0 Recommendation and conclusion

To sum it up, the 21st century is a century in which mankind is highly dependent on data and deeply integrated into the information society. In the information society, online social networks have built a huge functional platform. People express opinions and make friends and interact on social networks, and hundreds of millions of information are generated every day. Online social interaction is changing people's behavior patterns and social patterns. Online social network data is also becoming the most mature big data. Through research and analyzing this technology, people are expected to have an unprecedented depth of understanding of user behavior and social phenomena behind online social network big data.

This analysis is very enlightening

Common place:

1. Stay hot and interact with your audience frequently. Of course, the most important thing is to be sincere and discuss the issues that we really care about (for Musk is the future of mankind, journey to Mars, Tesla, reduce global warming), for Graham, he is most concerned about how to help the initial entrepreneurship People, build alliances for them, news about the new crown pneumonia and how to help everyone tide over the difficulties
2. Both individuals have maintained close communication with their most important partners. It can be seen from the visualization of social networks that there are basically 3 and 4 communication objects and groups, and not all topics and groups are involved. It is important to choose our own groups and interest groups
3. Be objective, sincere, and enthusiastic. Through sentiment analysis, it can be found that the two people still maintain a neutral attitude on most topics, and post meaningful opinions and

# References

Christopher M.Bishop (2006). *Pattern Recognition and Machine Learning*. Springer-Verlag

Cover T M, Hart P E (1967). Nearest neighbor pattern classification. IEEE Trans Inf Theory IT-13(1):21-27[J]. *IEEE Transactions on Information Theory*, 1967, 13(1):21-27.

Daral N (2005). *Histograms of Oriented Gradients for Human Detection*[J]. Proc. of CVPR, 2005.

David J. Hand and Robert J. Till( 2001). *A Simple Generalization of the Area Under the ROC Curve for Multiple Class Classification Problems* . Machine Learning , 45(2), 171- 186 .

Hang L.,(2012). Statistical Learning Methods, *Tsinghua University Press*. (Chinese)

Dalai,N,B.Triggs (2005). Histograms of Oriented Gradients for Human Detection. Computer Vision and Pattern Recognition, 2005.CVPR 2005.*IEEE Computer Society Conference on*

Kazemi V, Sullivan J (2014). One Millisecond Face Alignment with an Ensemble of Regression Trees[C] *Computer Vision and Pattern Recognition*. IEEE:1867-1874.

Qcon (2017) *Feature engineering by Gabriel Moreira*

Rumelhart D E, Hinton G E, Williams R J (1986). *Learning representations by back-propagating errors.*[J]. 323(6088):399-421.

Scikit-learn, (2021) Available at: https://scikit-learn.org/stable/index.html [accessed on 06-Jun-2021]

Thomas M.Cover, JoyA. Thomas (2006). *Elementsof InformationTheory.*

Zhou Z, H.,(2016). Machine Learning, *Tsinghua University Press*(Chinese)