Three and a Half Degrees of Separation in Twitter

(By Haoxiang Wang)

Introduction

In 1967, Harvard psychology professor Stanley Milgram wanted to describe a web of human connections that connects people to communities. He did a chain letter experiment and found the phenomenon of "six degrees of separation". His summary of this phenomenon is: "You and any stranger will not be separated by more than six people, that is to say, you can know any stranger through at most six people."[1]

After the Milgram experiment, with the popularization of information technology and Internet applications, many social network media appeared, such as Messenger, Facebook, twitter, etc. The popularization and application of these social network media formed an online network of friends. Does the Six Degrees of Separation Theory still apply in the network structure of these online friendships? Facebook's team studied the data of 1.59 billion users, and it was a huge challenge to calculate the number of people who can find a connection between two people through such a huge amount of data. On February 4, 2016, the results of the research titled "Three and a Half Degrees of Separation" were published on the website FACEBOOK research and found that the network diameter in Facebook was 3.57, which means that the distance between each person and others is 3.57 [2]. This paper proposes an approximation algorithm to find the shortest distance between any two people in the world and find a similar conclusion in Twitter is 3.43.

The Algorithm to Calculate the Distance Between Any Two Twitters in the World

Since directly analyzing the profiles of billions of users and finding their connections is a near-impossible task for individuals, I propose an approximate algorithm for calculating the distance between any two people in the world. The algorithm is based on the funnel effect proposed by Milgram, where most of the delivery is done by those very few stars. The algorithm steps are as follows:

1. Take the 1000 most popular twitter accounts in the world (with the most followers).

2. Build a social network among them.

3. Randomly select two active Twitter accounts (number of friends + number of followers>=20), start from these two endpoints, perform breadth-first search, and find the 3 nearest popular accounts respectively.

4. After completing the second step, we can get nine paths between these two tweets, and select the length of the shortest path between them as the distance between the two twitters.
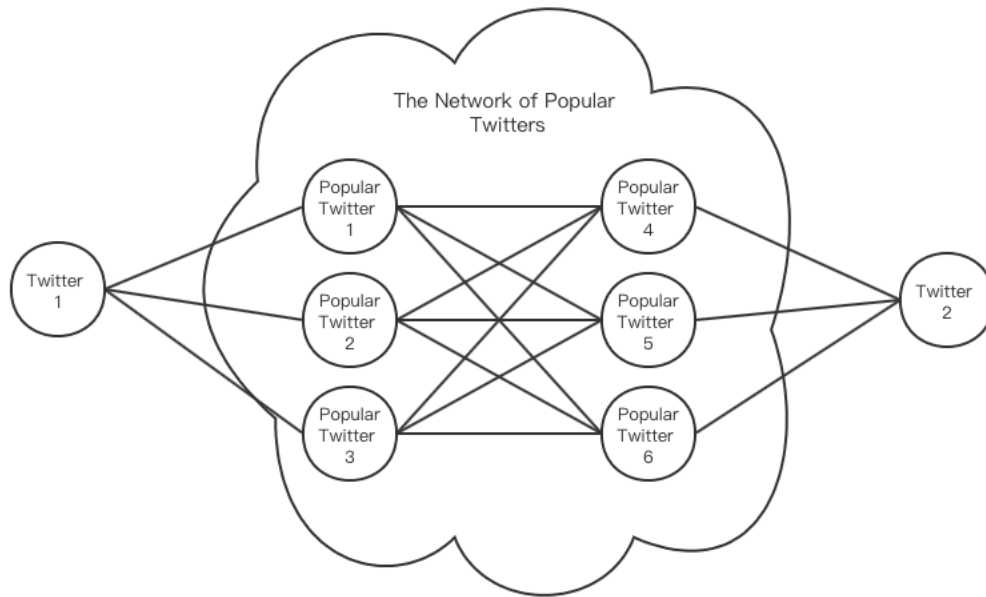

Figure 1 Take the shortest one of the nine paths as the distance between two points

Algorithm Implementation and Results

First of all, I found the usernames of the 1000 most popular twitter accounts on the web (https://www.trackalytics.com/the-most-followed-twitter-profiles/page/1/). Then I used the Twitter API to find their relationships. Every Twitter account has millions of followers but few friends (friend means following). It's worth noting that we only need to get the friends of these popular twitters to find out all their relationships, because if there is a relationship between two people, then one must be a friend of the other. Excluding some accounts that couldn't get their friends list due to permission settings, I got a total of 883 popular twitter and their popular twitter friends (stored in "popular_twitter_friends.txt").

Then I build a network among these 883 popular twitters. After removing some points of degree 0, I obtained a connected graph of 871 nodes that make up the network of popular accounts (stored in "popular_network_nodes.txt"). I used the Floyd–Warshall algorithm to calculate the shortest distance between any two nodes in this network (stored in "popular_network_shortest_distance_matrix.txt"). The results are shown in Figure 1. **The diameter of the network of popular twitters is 6 and the average distance is 2.19.** Most pairs of the popular twitters have a distance of 2.
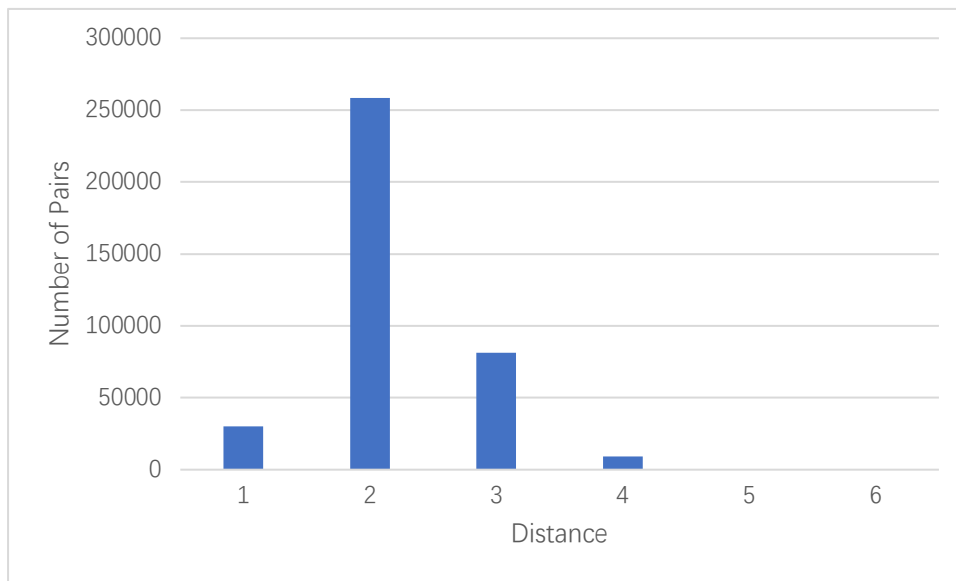
Figure 1 The distance between popular Twitter accounts

Finally, I randomly selected 300 pairs of twitter accounts and calculated the distance between them. On the official website of twitter api, we can see such a description, "Twitter IDs are unique 64-bit unsigned integers, which are based on time, instead of being sequential." So we can randomly pick two twitters by randomly picking the twitter id. For each pair of Twitter accounts, we perform a breadth-first search starting with these two accounts, each finding the three most recent popular accounts. Take the shortest of the nine paths as the distance between them. The results are shown in Figure 2 (Detailed data is stored in "user_distance.txt"). **The results show the average distance between any two people is 3.43 in Twitter.** Most pairs of the twitters have a distance of 3 or 4.
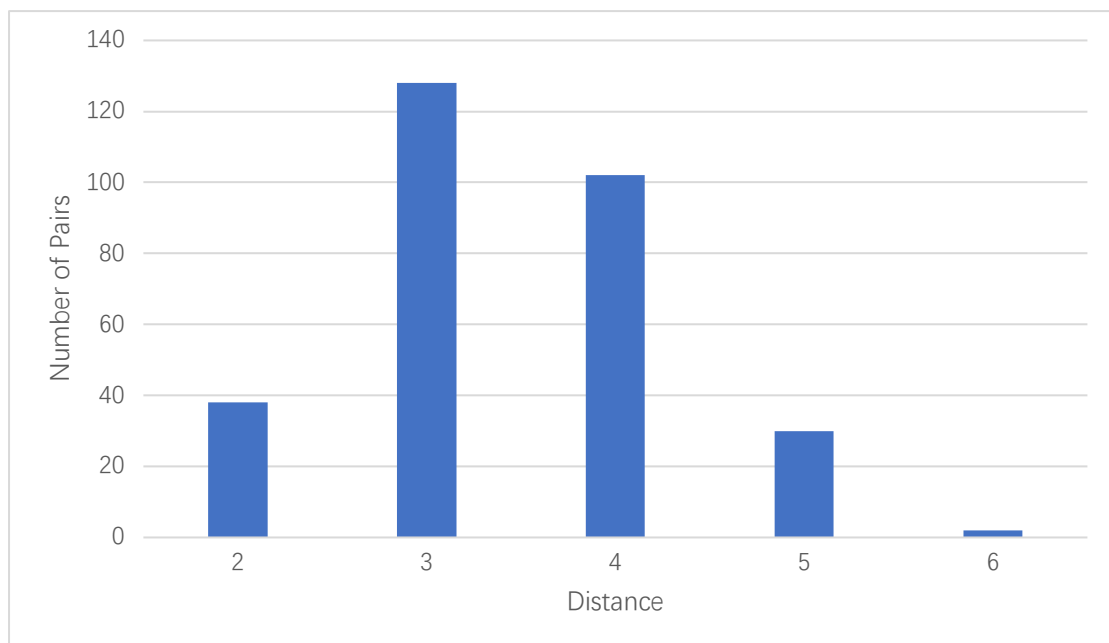

Figure 2 The distance between random 300 pairs of Twitter accounts

Reference

[1] Small-world experiment. (2022). *In Wikipedia*.
https://en.wikipedia.org/wiki/Small-world_experiment

[2] Bhagat, S., Burke, M., Diuk, C., Edunov, S., & Filiz, I.O. (2016). *Three and a half degrees of separation*. https://research.facebook.com/blog/2016/2/three-and-a-half-degrees-of-separation/