Verify Forbidden Triad Theory in Social Media using Parallel Computing

Yu Liang, Master Student in CSE

Jiabo Liang, Master Student in CSE

**Abstract *—*** *Forbidden Triad is a theory in social networks that states that when A is associated with B, and A is associated with C, there is no connection between B and C that is very unlikely to occur. In this paper, we use youtube's social network data to validate the Forbidden Triad theory. In the calculation, we use CUDA parallel computing and CPU multi-threaded computing to calculate in all the triangular relationship the number of closed relationship. In the test, the CPU computes faster than GPU parallel computing. The results show that in the youtube social network, the non-closed triangular relationship is only 99.98%, that is, B and C without any contact is a high probability event. This contradicts the Forbidden Triad theory. We think youtube social networking and daily relationships are still very different, so that the theory of daily relationships can not be simply applied to the Internet social.*

**Index Terms — forbidden triad, CUDA, parallel computing, weak tie, strong tie, social media, the strength of weak tie**

**I. INTRODUCTION**

The application of large data in social networks has become one of the most important technologies in the information society. For the large amount of data generated by the social network, how to carry out complex analysis of them has been the center of scholars around the world. In the study of social networks, how to judge the link between users is to understand the social network behind the important aspects of user relations. Mark granovetter, in his 1974 issue of the weak power of the famous essay, described the weak relationship in social networks can also play a significant role. Weak relationship refers to the relationship between daily life is not close. To give an example, when people are looking for work, people who tend to be weak are briefing the work to them. He presented a hypothesis of forbidden triad in the paper. This hypothesis is that for a triangular relationship, if A and B are strongly correlated, A and C are also strongly related, then there is no connection between B and C is very unlikely. To give an example, you and your mother is strong relationship, you and your wife is also strong relationship, then your mother and your wife do not know each other is very impossible.

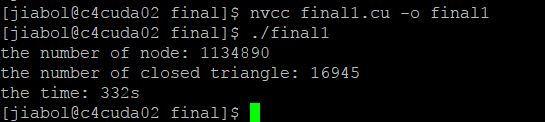
With the development of GPU-based parallel computing in recent years, GPU parallel computing has become a fantastic choice when the CPU can not handle large amounts of social networking data quickly. Through the GPU's large-scale processing power, can quickly verify the strength of the triangular relationship between users. In this study, we use CUDA parallel computing to verify the forbiddent triad hypothesis that if user A is concerned with user B and user A is concerned with user C, then user B and user C are not concerned with each other.

II. METHOD

In order to verify the forbiddent triad hypothesis, we use youtube's attention to network data, that is, if user A is concerned with user B, user A is concerned with user C, then user B and user C are not concerned with each other is very unlikely. The experimental data comes from the network. Each row of data is represented by two node IDs, which means that the two IDs are directly related to each other. We will use CUDA parallel computing to calculate when users A and B, A and C are strong contact, how many B and C without any contact.

Youtube is a video sharing site that contains social networks. In the social network of youtube, users can establish contacts and create groups to link people with the same interests.

The data in this paper is shown in Figure 1, and each row represents an edge, and the number represents the ID of the node. There are 1,134,890 nodes and 2,987,624 edges in the data.



|  |  |
| --- | --- |
| Node 1 | Node 2 |
| 1 | 2 |
| 1 | 3 |
| 1 | 5 |
| 1 | 6 |
| 1 | 7 |
| 1 | 10 |
| 1 | 12 |
| 1 | 13 |
| 2 | 21 |
| 2 | 24 |
| 3 | 15 |

Fig 1. data structure

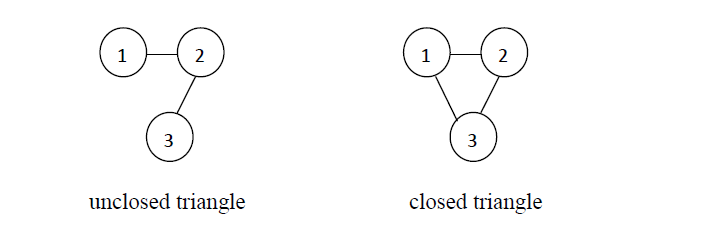
In the data preparation phase, we convert the node to an array with the node ID. And then calculate the number of closed triangular relationships: we in each node, select the two nodes are connected with this node. And then judge whether the two nodes are also linked. If so, then the three nodes constitute a closed triangular relationship. If not, between them are non-closed triangular relationship. For example, in the list of nodes in node 1, we find that node 2 and node 3 are connected to node 1. Then we go to check the node 2 node list to determine whether the node 3 in which, if so, then the node 2 and node 3 are also connected. So node 1, node 2, node 3 to form a closed triangular relationship.

Fig 2. Triangle relationship

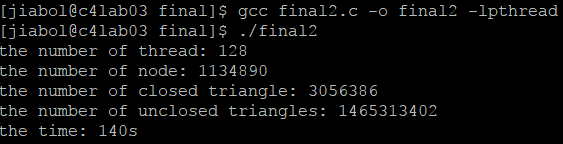
III. RESULT

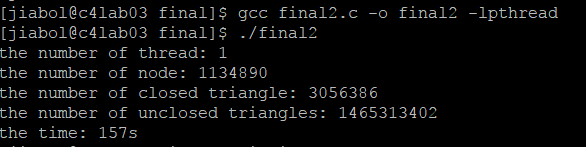
First, we compute in CUDA. But after 20 minutes the program is still running. We check code many times and make sure the code is right. Then we try to only compute one node instead of all node (1134890). The program is immediately finished. Then I compute five nodes, and find it still need over 300 seconds time (Figure 3).

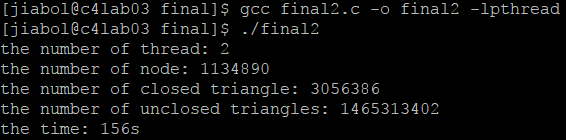
Fig 3. Step 1 result

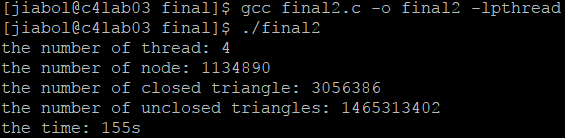
So, we are thinking the CUDA’s property. The CUDA is efficient in the large repeated computations. But in this algorithm, we don’t have much repeated computation. In this program, it need search the array again and again. The array’s size is over one million. So, we guess it’s one reason that influences the GPU’s performance. Then, we rewrite the code to run on CPU in multithreading. We find it’s really fast. It only spends around 2 to 3minutes. The consequence is in the following table.

|  |  |
| --- | --- |
| thread number | cost time |
| 1 | 157s |
| 2 | 156s |
| 4 | 155s |
| 8 | 154s |
| 16 | 151s |
| 32 | 148s |
| 64 | 141s |
| 128 | 140s |
| 256 | 135s |
| 512 | 124s |
| 1024 | 115s |

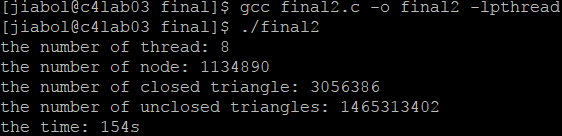
From the table, we can find the cost time decreases with the increase of the number of thread. Multithreading is really increase the performance. Finally, we get the number of closed triangle: 3056386. The number of unclosed triangle is 1465313402. Then the fraction of closed triangles is 3056386/(1465313402+3056386)=0.0020814. So, we can find the possibility that there is a tie between Node 2 and Node 3 is only 0.2 percent.

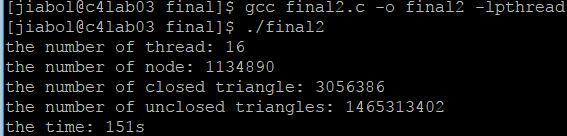


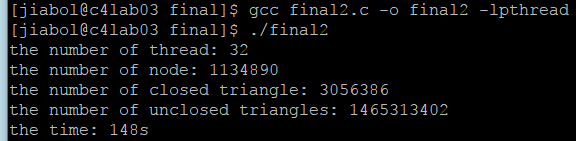
the number of thread:1

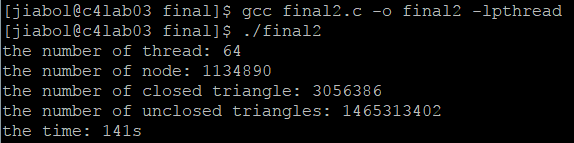
the number of thread:2

the number of thread:4

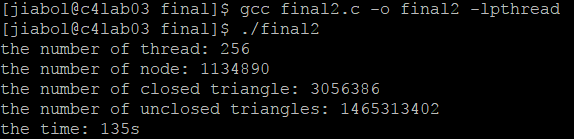


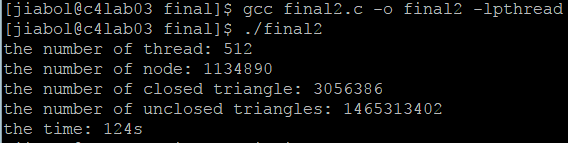
the number of thread:8

the number of thread:16

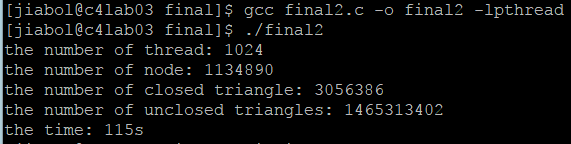
the number of thread:32

the number of thread:64

 the number of thread:128

the number of thread:256

the number of thread:512



the number of thread:1024

III. CONCLUSION

We found that not all of the algorithm GPU parallel computing can bring good results. And social networking on the computer's Internet and people's reality in the social network there is a big difference.

Our research will further validate the impact of the forbidden triad hypothesis on social networks. This will allow people to further study the social network. At the same time, we also explored the use of CUDA parallel computing to solve the problem of triangular relationship, in order to future parallel computing technology used in social network research to explore

III. REFERENCE

[1] J. Aarstad, "Do Cognitive Networks Precede the Development of Real Structures or Vice Versa? A Study on Triad Census Distributions," (in English), *2013 Ieee International Conference on Industrial Technology (Icit),* pp. 1461-1465, 2013.

[2] A. Chatterjee, S. Radhakrishnan, and C. N. Sekharan, "Connecting the dots: Triangle completion and related problems on large data sets using GPUs," (in English), *2014 Ieee International Conference on Big Data (Big Data),* 2014.

[3] S. B. Dutta, R. McLeod, and M. Friesen, "GPU accelerated Nature Inspired Methods for Modelling Large Scale Bi-Directional Pedestrian Movement," (in English), *Proceedings of 2014 Ieee International Parallel & Distributed Processing Symposium Workshops (Ipdpsw),* pp. 448-456, 2014.

[4] A. R. Galvao, L. C. M. de Paula, C. J. Coelho, T. W. Limab, and A. D. Soares, "CUDA parallel programming for simulation of epidemiological models based on individuals," (in English), *Mathematical Methods in the Applied Sciences,* vol. 39, no. 3, pp. 405-411, Feb 2016.

[5] Z. F. Gou and Z. Z. Ji, "Accelerating Triangle Counting on Undirected Graph Using GPU," (in English), *2nd International Conference on Sensors, Instrument and Information Technology (ICSIIT 2015),* pp. 13-17, 2015.

[6] M. S. Granovetter, "The Strength of Weak Ties," (in English), *American Journal of Sociology,* vol. 78, no. 6, pp. 1360-1380, 1973.

[7] S. C. Peng, G. J. Wang, and D. Q. Xie, "Social Influence Analysis in Social Networking Big Data: Opportunities and Challenges," (in English), *Ieee Network,* vol. 31, no. 1, pp. 11-17, Jan-Feb 2017.

[8] T. R. Stovall, S. Kockara, and R. Avci, "GPUSCAN: GPU-Based Parallel Structural Clustering Algorithm for Networks," (in English), *Ieee Transactions on Parallel and Distributed Systems,* vol. 26, no. 12, pp. 3381-3393, Dec 2015.

[9] F. Wang, E. A. Mack, and R. Maciewjewski, "Analyzing Entrepreneurial Social Networks with Big Data," (in English), *Annals of the American Association of Geographers,* vol. 107, no. 1, pp. 130-150, Jan 2017.

[10] Y. Zhang, D. Yi, B. G. Wei, and Y. T. Zhuang, "A GPU-accelerated non-negative sparse latent semantic analysis algorithm for social tagging data," (in English), *Information Sciences,* vol. 281, pp. 687-702, Oct 10 2014.