Misinformation Containment in OSNs leveraging Community Structure

Arnab Kumar Ghoshal
Department of Computer Science
Asutosh College
Kolkata, India
mr.arnabghoshal@gmail.com

Nabanita Das
Advanced Computing and Microelectronics Unit
Indian Statistical Institute
Kolkata, India
ndas@isical.ac.in

Soham Das
Microsoft Corporation
Redmond, U.S.A.
88soham@gmail.com

***Abstract*—With the emergence of Online Social Networks (OSNs) as a major platform of communication, its abuse to spread misinformation has become a major threat to our society. In this paper, we study the misinformation containment problem in OSN. Given a snapshot of the OSN with a set of misinformed nodes, and a budget in terms of maximum number of seed nodes, the problem is to select the seed nodes, referred here as the *beacon nodes*, to plant the correct information, to minimize and eventually eradicate the misinformation at the earliest. We leverage the community structure of the OSN to select the beacon nodes, prioritizing the *Community Boundary Nodes*. To the best of our knowledge, this is the first work to exploit the topology of the OSN to combat misinformation spread. A modified form of *Independent Cascade Model* is followed to study the adversarial propagation of both misinformation and the correct information. Simulation on real data set shows that the proposed algorithm outperforms earlier algorithm [1] significantly in terms of *maximum (average) infected time* and the *point of decline*.**

***Index Terms*—Online Social Networks (OSN); Information Diffusion; Community Structure; Misinformation Propagation; Infected Time; Point of Decline.**

I. Introduction

In the last couple of decades, online social networks (OSNs) have emerged as the most popular platform of connectivity for communication, information sharing, advertising, surveys and reviews [2]–[7]. Statistics show that 62% of adults worldwide use social media and spend 22% of online time on social networks [8]. This provides OSNs the ability to reach out to a large population in very little time. OSNs also allow users to share their personal opinions on any issue or topic, creating a word-of-mouth effect. Moreover, a user's opinion is influenced by his/her friends or colleagues [9]. This has significant impact on the adoption rate of new products [10]–[12], new opinions etc. But at the same time, in OSN's, there is very little scope today to validate the information before being released to the network. As a result, OSNs easily fall prey to rumors and misinformation. We often find users ranging from media, news channels, websites to individuals exploiting this and spreading rumors in the network. If not controlled properly in the nick of time, this may have widespread adverse effects on the entire population. A recent rumor that went viral on social media and caused havoc is the fake news of former Sri Lankan cricketer Sanath Jayasuriya's death for example. It disturbed the cricketing community and fans to a large extent and generated a lot of traffic on Twitter [13].

A lot of research has been conducted to study rumor spread in social networks [14]–[21], the most popular and well received are the SIS model, SIR model [14], Daley Kendall (DK) SIS model, Maki-Thompson (MT) model [15], Voter model [16], [17], linear threshold model, independent cascade model (ICM) [18], multi-cascade diffusion model [21] etc. The ICM being the basic one has drawn much attention in the research community for its simplicity to represent information diffusion in non-adversarial scenario. In this paper we have applied a modified form of the independent cascade model of information diffusion, first introduced in [1], to capture the simultaneous propagation of misinformation and its correct version. To reiterate Tripathy et al., we use this model because it is able to ensure the spread of the rumor through the entire network in a finite time, making it a robust adversary. Hence if we can establish that our algorithm can eradicate misinformation following this diffusion model then other weaker models should be more easily contained. However none of the previous research works leveraged community structure of the social network which may have a significant role to contain and diffuse information in the network.

With the emergence of very large scale networks, like online social networks, biological networks etc., identification and analysis of community structures are becoming important. Community structure in networks generally means natural division of the nodes into densely connected partitions, such that interconnections among partitions are sparse, compared to the connections within the same partition. In fact, the community structure of a network reveals a lot of information hidden in the topology and plays a crucial role in characterizing the properties of the network [22], [23]. Hence, lot of research has been conducted recently to identify the community structures embedded in a network [23]–[26]. There is also a particular demand in identifying the nodes responsible for information flow between communities. For example, in temporal Twitter networks edges between communities play a key role in propagating spikes of activity when the connectivity between communities is sparse and few edges exist between different clusters of nodes [27]. In this paper, for combating the spread of misinformation, we propose to utilize the community structure of the graph, to place *beacon* nodes with correct information.

GP-GPU Computing for Analysis of Large Scale networks

Arnab Kumar Ghoshal*
University of Calcutta
Kolkata, India
mr.arnabghoshal@gmail.com

Nabanita Das
Indian Statistical Institute
Kolkata, India
ndas@isical.ac.in

ABSTRACT

Over last few years the interest in large scale network processing has gained momentum due to the increased importance of efficient processing of such networks for the analysis and problem solving in Social networks, Internet of Things(IoT), Data mining, Biological networks etc.. The steep increase in volume of data being produced, necessitates the development of parallel and distributed graph processing algorithms that demand extensive computation. Now a days, because of high-computing-power of GP-GPUs (General Purpose Graphic Processor Unit) and its cost-effectiveness, GP-GPUs are being widely used in almost all areas of computing. In this paper, we address the problem of designing efficient parallel graph processing algorithms for GP-GPU platform which is the upcoming trend in high end machines. Presently, the problem of community detection and its influence on misinformation containment in online social networks have been investigated thoroughly, and parallel CUDA algorithms have been proposed. Simulation studies show a significant speed-up compared to the existing sequential/parallel techniques.

CCS CONCEPTS

• **Computing methodologies** → **Parallel computing methodologies**.

ACM Reference Format:
Arnab Kumar Ghoshal and Nabanita Das. 2020. GP-GPU Computing for Analysis of Large Scale networks. In *ICDCN '20: 21st International Conference on Distributed Computing and Networking, January 04–07, 2020, Kolkata, India*. ACM, New York, NY, USA, 1 page. https://doi.org/0000001.0000001

1 GRAPH PROCESSING ON GP-GPU

Since the parallel graph processing algorithms often exhibit irregular data access patterns, the applications may not reach the peak performance in GP-GPU. In addition, GP-GPUs also have less memory compared to CPU, thus moving data from the host memory to the GP-GPU memory causes the extra overhead and the conditional branches in parallel graph computations do not fully exploit the high degree of parallelism offered by the SIMD executions. In [1], authors have shown that a wide range of fundamental graph problems can be solved using theoretically efficient parallel algorithms, even on a single commodity shared-memory machine. Using our expertise in GP-GPU, we proposed parallel algorithms for community detection in large graphs on GP-GPU platform [2], and have shown how this topological information of large graphs may help us to eradicate misinformation in online social networks [3]. Our study shows that even with wide load imbalance among heavily-connected nodes and sparsely-connected nodes, parallel algorithms on GP-GPU may achieve significant speed-up.

2 EXPERIMENT

The execution time of our parallel and sequential implementations are measured on NVIDIA Quadro $P4000$ GPU and Intel Xeon CPU respectively. In case of community detection and misinformation containment problems, we obtain speed-up up to a factor of 4.27 and 4.07 respectively. To eradicate misinformation, our strategy outperforms earlier algorithm [4] by 20% reduction in the *maximum infected time* with up to 24% and 10% reductions in *average infected time* and *point of decline* respectively. In Fig. ii, $|V_b(0)|$ is used to denote the initial set of beacon nodes.

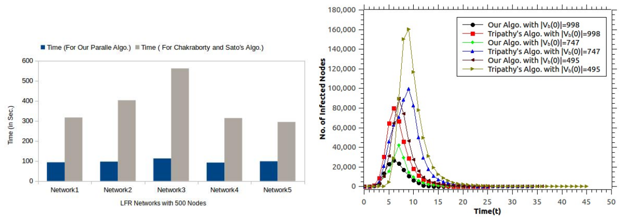

Fig.i) Execution time for community detection algorithms ii) Performance comparison for misinformation restriction

3 FUTURE WORK

Efficient parallel algorithms are to be developed to analyze the correlation among the topological structure of graphs and its properties.

REFERENCES

[1] L. Dhulipala, G. E. Blelloch, and J. Shun, "Theoretically efficient parallel graph algorithms can be fast and scalable," in *30th SPAA*. ACM, 2018, pp. 393–404.

[2] A. K. Ghoshal, N. Das, S. Bhattacharjee, and G. Chakraborty, "A fast parallel genetic algorithm based approach for community detection in large networks," in *11th IEEE COMSNETS*, 2019, pp. 95–101.

[3] A. K. Ghoshal, N. Das, and S. Das, "Misinformation containment in osns leveraging community structure," in *10th IEEE iCAST*, 2019.

[4] R. M. Tripathy, A. Bagchi, and S. Mehta, "A study of rumor control strategies on social networks," in *Proceedings of the 19th ACM ICIKM*, 2010.

*student author

Conference'17, July 2017, Washington, DC, USA

ACM ISBN 978-x-xxxx-xxxx-x/YY/MM. . . $15.00
https://doi.org/0000001.0000001