Parallel Algorithms for Mining Large-scale Time-varying (Dynamic) Graphs

Shaikh Arifuzzaman, Naw Safrin Sattar, and Md Abdul Motaleb Faysal
Dept. of Computer Science, University of New Orleans, New Orleans, LA 70148 USA
Email: {smarifuz, nsattar, mfaysal}@uno.edu

***Abstract*—Real complex systems are inherently time-varying and are modeled as temporal graphs (networks). Examples include social, transportation, and many forms of biological networks. Due to the advancement of data and computing technology, the modern world generates a huge volume of network data with detailed information on the time of occurrence and duration of each link. However, standard graph metrics introduced so far in complex network theory are mainly suited for static graphs, i.e., graphs in which the links do not change over time. It is only recently that the problem of mining temporal graphs has drawn considerable attraction.**

In this paper, we present a brief overview of our ongoing work on designing scalable parallel algorithms for mining large time-varying networks. We have been working on devising high-performance computing (HPC) techniques for this data-intensive computation using diverse platforms, e.g., shared-, distributed-memory, and hybrid systems.

***Index Terms*—Parallel algorithms; large graphs; dynamic graph; data-intensive computing; Scalable methods**

Motivation, Methods, and Contribution

Temporal graphs are powerful representation of various social, biological and technological dynamic systems [1], [2]. A complex system is inherently time-varying (dynamic) [3]. Social interactions and human activities, appearance and disappearance of links in the Web, patterns of interactions among genes and in functional brain networks are all dynamic. *Despite this fact, most of the classic studies in complex networks theory are based on the analysis of the topological properties of static graphs.* Analyzing such graphs with detailed information on the time of occurrence and duration of each link help us understand the properties and dynamical process on them. Some key applications are as follows: (*i*) *Diffusion and propagation in complex and social networks:* modeling spread of viruses through a community, (*ii*) *Understanding communication networks:* false news propagation, (*iii*) *Improving transporation systems:* route-planning algorithms depending on the traffic with varied time, (*iv*) *Neuron (brain) network analysis:* locating key neurons in cortical networks, etc. All of the above applications require defining and computing various temporal network metrics. The emergence of network big data from numerous scientific disciplines poses significant challenges for network mining and analysis. Online social networks such as Twitter and Facebook have millions to billions of users. Such massive networks often do not fit in the main memory of a single machine, and the existing sequential methods take a prohibitively large runtime.

In our project on large-scale computation on temporal graphs, we aim at designing scalable algorithms for a particular group of metrics– computing path, centrality, and communities of termporal graphs. These computations have important applications in graph mining– for example, community reveals an organizational unit in social networks or a functional unit in biological networks [2].

Note that, the literature of algorithms for temporal graphs is very nascent. When it comes to mining temporal networks and computing network metrics, very few work currently exist [3], [4]. All of these work are sequential and do not scale to large networks. We want to contribute to fill in this void. Our prior work on scalable parallel algorithms [1] for static graphs provide us important insights. Our current efforts include designing efficient load balancing and communication schemes, data reduction (e.g., graph sparsification and approximation), and efficient formalization of temporal metrics. Fig. 1 shows the performance of our initial version of multithreaded algorithm for computing shortest path in dynamic graphs.

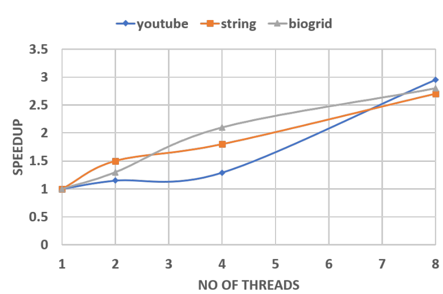

Fig. 1. Speedup factors of our multithreaded shortest path algorithm on three dynamic graphs. We hope to continue improving our algorithms based on current insights. Efficient algorithms for temporal paths will lead to efficient computation of other temporal metrics.

References

[1] S. Arifuzzaman, M. Khan, and M. Marathe, "A fast parallel algorithm for counting triangles in graphs using dynamic load balancing," in *2015 IEEE BigData Conference*, 2015.

[2] J. Leskovec, K. Lang, and M. Mahoney, "Empirical comparison of algorithms for network community detection," in *19th International Conference on World Wide Web*, 2010.

[3] M. M. G. R. C. M. Vincenzo Nicosia, John Tang and V. Latora, "Components in time-varying graphs," *Chaos: An Interdisciplinary Journal of Nonlinear Science*, vol. 22, no. 2, 2012.

[4] K. K. J. K. Lauri Kovanen, Marton Karsai and J. Saramaki, "Temporal motifs in time-dependent networks," *Journal of Statistical Mechanics: Theory and Experiment*, vol. 2011, no. 11, p. P11005, 2011. [Online]. Available: http://stacks.iop.org/1742-5468/2011/i=11/a=P11005