Efficiency Issues of Computing Graph Properties of Social Networks*

Péter Englert, Márton Balassi, Balázs Kósa, Attila Kiss

Eötvös Loránd University
{enpraai,bamrabi,balhal,kiss}@inf.elte.hu

Abstract

In this paper we consider different properties of social networks, namely the degree and edge betweenness centralities, the diameter, the number of different triangles, the sizes of the strongly connected components [1] and Linerank, which is offered as a substitute for edge betweenness and can be effciently calculated on graphs with more than 10^9 nodes [2]. The computation methods of these features can be grouped into three different algorithm families using the categorization of [2]. We implemented these methods in three different ways: sequentially, using the widely spread MapReduce model – HADOOP in our case – and its newly emerging rival Giraph [3]. We ran these implementations on several generated and real-world networks of different sizes firstly in order to assess the degree of the benefit of parallelization offered by the last two technologies for the algorithm families. Secondly, to see for which magnitude of the size of the networks it is more beneficial to choose one the aforementioned models instead of the sequential approach.

Keywords: Social Networks, MapReduce, Giraph

MSC: 68Q85

References

[1] Newman, M. Networks: An Introduction, *Oxford University Press, Inc.* (2010).

[2] Kang, U., Papadimitriou S., Sun J., Tong H. Centralities in Large Networks: Algorithms and Observations, *SDM*, (2011), 119–130.

[3] Malewicz, G., Austern, M., Bik, A., Dehnert, J., Horn, I., Leiser, N., Czajkowski, G. Pregel: A System for Large-scale Graph Processing, *Proceedings of the 2010 ACM SIGMOD International Conference on Management of Data*, (2010), 135–146.

*This work was partially supported by the European Union and the European Social Fund through project FuturICT.hu (grant no.: TAMOP-4.2.2.C-11/1/KONV-2012-0013).