Accepted Manuscript

Scalable parallel simulation of dynamical processes on large stochastic Kronecker graphs

Klavdiya Bochenina, Sergey Kesarev, Alexander Boukhanovsky

PII: S0167-739X(17)30150-4
DOI: http://dx.doi.org/10.1016/j.future.2017.07.021
Reference: FUTURE 3551

To appear in: *Future Generation Computer Systems*

Received date : 25 January 2017
Revised date : 4 May 2017
Accepted date : 8 July 2017

Scalable parallel simulation of dynamical processes on large stochastic Kronecker graphs

Klavdiya Bochenina, Sergey Kesarev, Alexander Boukhanovsky
e-Science Research Institute, ITMO University
49, Kronverksky Pr., St. Petersburg, 197101, Russia
E-mail: k.bochenina@gmail.com, kesarevs@gmail.com, avb_mail@mail.ru

Abstract

Complex networks are widely recognized today as a unified framework to model the dynamical processes in socio-technical systems at the level of interacting elements. A stochastic Kronecker graph (SKG) is a network generative model that allows reproducing real-world networks while keeping their important topological properties. When sizes of SKGs reach dozens of millions of nodes, there is a need to apply parallel computations to simulate processes on networks stored in a distributed manner. In general, parallel simulation of a dynamical process on a complex network implies all-to-all communication between subnetworks at each iteration. In this paper, we study the efficiency of different SKG partitioning algorithms and different data interchange algorithms for dynamical process simulation on large SKGs. We compare the theoretical efficiency given by parallel performance models with experimental results for different communication patterns. An experimental part of the study was carried out for sparse SKGs with a size up to one billion nodes using Lomonosov supercomputer (Moscow State University, Russian Federation). The results show that: (i) proposed algorithm of SKG partitioning provides highly balanced results, (ii) observed parallel performance is well agreed with presented theoretical models, (iii) the scheme with all-to-all-communications between subnetworks is the most efficient up to approximately one hundred cores, (iv) master-slave scheme with a single master per node outperforms all-to-all scheme for a large size of a communicator (for our experiments, it has achieved near-linear speedup for up to several hundred processes).

Keywords: complex networks, parallel simulation, stochastic Kronecker graphs, dynamical processes, load balancing, parallel efficiency

1 Introduction

Complex networks are a powerful tool to describe and to study a wide range of natural, artificial and socio-technical systems. Representing a system as a set of interacting entities, they bring together such different fields like gene regulation, social interactions, transportation and world wide web. The growing interest in this area from the beginning of the 2000s is partially explained by the progress in computing that made possible aggregation and processing of massive data sets, and extraction of huge networks from the data. Modern networks (e.g. online social networks) are so large that they cannot be entirely collected at one time, resulting in the effect of partial observability [1]. The effect can also be caused by the excessively high rate of updates (both in node states and in a network structure) or fundamental impossibility to gather data (e.g. in criminal networks).

To tackle with the unavailability of full-size samples, different network models can be used. One can create a graph from a degree sequence using configuration model [2], adjust some initial approximation to desired topological characteristics using metaheuristic algorithms [3] or let it emerge