



Asynchronous Parallel Algorithm in Gene Networks Analysis: A Review

Monishi Dutta Choudhury¹, Somorita Baishya², Manabendra Dutta Choudhury^{2*}

¹School of Computer Science and Engineering, VIT University, India

²Bioinformatics Centre, Assam University, India

***Corresponding Author:** Manabendra Dutta Choudhury, Bioinformatics Centre, Assam University, India. E-mail: drmdc@bioinfoaus.ac.in

Received: December 17 2017; **Accepted:** January 29 2018; **Published:** February 10, 2018

Copyright: © 2017 Choudhury M D. Asynchronous Parallel Algorithm in Gene Networks Analysis: A Review. Int Res Comp Sci; 1(1): 1-4.

Abstract

The high-throughput (HTP) techniques generate big data at various biological levels. Such big biological data could be employed for designing holistic models such as interactive networks of various biological components like genes or proteins. Mapping of such gene networks following reverse engineering algorithms requires time and if the datasets are large, efficiency and accuracy of the gene networks are compromised. However, following asynchronous parallel algorithm could provide fundamental insights into the biology concerned with better efficiency and accuracy and in less time.

Keywords: Big Biological Data; Gene Networks; Asynchronous Communication; Parallel Algorithm; Master-Slave Mode

Introduction

Life scientists are grappling with the explosion of big biological data that has generated with the advent of high-throughput (HTP) techniques [1]. The HTP techniques generate data at various biological levels—genomic, transcriptomic, proteomic, metabolomic, molecular pathway, molecular imaging, medical records etc and can reach a scale past petabyte (PB) and sometimes even exabyte (EB) [2]. Such big biological data could be employed for designing holistic models such as interactive networks of various biological components like DNA, RNA, genes, proteins, cells and other small molecules. Mapping of such networks would provide fundamental insights into the biology concerned [3]. The biological networks are created by merging thousands of genes/molecules following different sets of algorithms. However, low accuracy of the algorithms and high computational cost poses a difficulty in deciphering the biological networks. Therefore, new engineering techniques are being developed to devise new computational algorithm like parallel algorithm that could bridge the problem [4]. In this review attempt has been made to describe the utility of asynchronous parallel algorithm in gene network analysis.

Parallel Algorithm

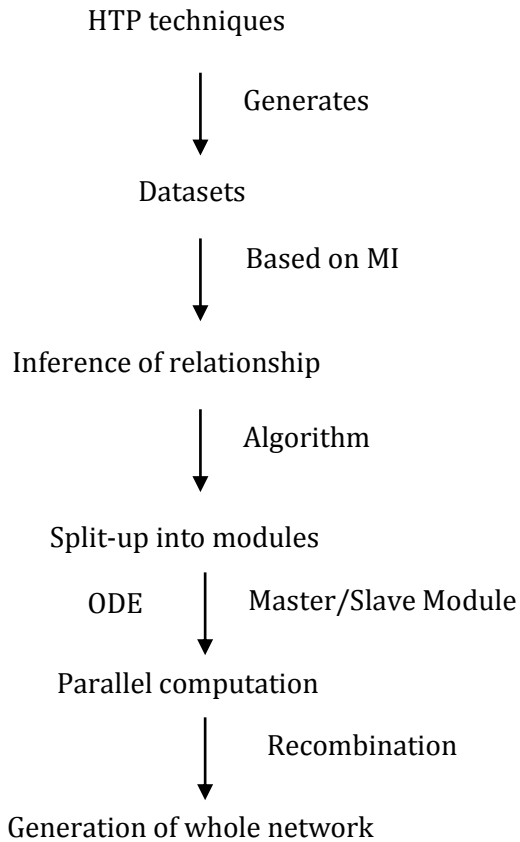
Parallel algorithm is an algorithm which can execute an input simultaneously on many different processing devices as message or module, and at the end combine them together again to get the correct result. For managing big data in quick time such algorithms are used. The interface at which the discretely generated messages are synchronized and exchanged is called message passing interface (MPI). MPI usually has two modes: Peer-to-peer mode and Master/Slaves mode. In the context of generating biological networks, the Slaves process one or several modules and always under the control of Master.

However, before establishing communication between different slaves it is important for the Master to split and select edges (links) between different modules. This will also help in generating more accurate networks. The mutual information (MI) of different modules with respect to one another are calculated and ranked according to scores. Of all the MI values only 5-10% is considered because of the sparsity of big biological networks. Thus, all the slaves get linked to the edges. Each Slave will generate ordinary differential equations (ODE) and Multiple Linear Regression (MLR) models. In such kind of parallel frame-work design different Slaves may transmit signals at different clock rate. Such communications are called asynchronous communication. Once all the Slaves have transmitted their signals the Master collects them all and recombines them to form a single network.

The performance of such asynchronous parallel algorithm i.e. accuracy, stability, run-time and speed of the algorithm was evaluated based on some popular indices like true positive rate, false positive rate, positive predictive value and accuracy and compared with some widely used algorithms which also worked on high-performance computing environments. The results were remarkable. When the datasets were small asynchronous parallel algorithm showed best results and even on increasing network size it showed prominent results. The performance of the algorithm was also stable irrespective of the size of the datasets. On comparing the asynchronous parallel algorithm model with two well-known reverse engineering models for verifying the efficiency aspect, it was seen that on increasing the network size the speed-up value increased exponentially thus, indicating that parallel algorithm is more efficient in handling large datasets. Therefore, the parallel algorithm model definitely promises to construct large biological interaction network with high accuracy and in a short time period, thereby

opening the gates to better understanding of the fundamentals of various biological pathways [4].

Figure (1): Model based on steps followed in Asynchronous Parallel Algorithm



Gene Network

Interaction between genes and proteins is crucial for carrying out the various functional processes in a biological system. These interactions could be regulatory, inducible, repressible, metabolic or functional. All these interactions could be mapped as gene networks (GN) that could demonstrate the functioning of the system more explicitly [5]. Gene network analysis (GNA) can not only illustrate the relationship between genes and different phenotypes but can also find novel genes that are likely to cause a disease or community hazards like antibiotic resistance [6]. GNA also has the potential to discover undescribed local

perturbation within known pathways by assembling reported interactions from literature and various databases [5]. Therefore, uncovering GNs could help in strengthening the concept of cell as a system and also elucidate the impact of various genes on biological functions [6].

Utility of Asynchronous Parallel Algorithm in Gene Network Analysis

The advent of HTP techniques has flooded the biological world with genomic and proteomic data [2]. These data could be used to create GNs and their analysis could present expert interpretation of the coherent biological questions. However, it is difficult to handle process and manage such huge data. Databases like STRING (Search Tool for the Retrieval of Interacting Genes/Proteins) (<http://string-db.org/>) and GeneMANIA (<http://www.genemania.org>) has indexed about 2.5 million proteins from 630 organisms and 747 data sets from 6 organisms respectively [7]. For creation of GNs majority of the protein-protein interaction data are retrieved from these databases. Following reverse engineering algorithms for the creation of GNs may be time taking and the efficiency and accuracy may be low. However, applying asynchronous parallel algorithm could improve the efficiency and accuracy as well as decrease the time required for generating gene networks [4]. Therefore, it is evident that asynchronous parallel algorithm have huge impact in generating huge gene networks.

Conclusion

From the above review, we can understand that asynchronous parallel algorithm has the capability to manage big biological data and generate large gene networks which can unveil many biological questions.

References

1. Marx V (2013) Biology: The big challenges of big data. *Nature*; 498(7453): 255-260.
2. Li Y, Chen L, (2014) Big biological data: challenges and opportunities. *Genomics proteomics & bioinformatics*; 12(5): 187-189.
3. Stolovitzky G, Monroe D, Califano A (2007) Dialogue on Reverse-Engineering Assessment and Methods. *Annals of the New York Academy of Sciences*; 1115(1): 1-22.
4. Xiao X, Zhang W, Zou X (2015) A new asynchronous parallel algorithm for inferring large-scale gene regulatory networks. *Plos one*; 10(3): e0119294.
5. Bag S, Anbarasu A (2015) Revealing the strong functional association of adipor2 and cdh13 with adipoq: a gene network study. *Cell biochemistry and biophysics*; 71(3): 1445-1456.
6. Anitha P, Anbarasu A, Ramaiah S (2016) Gene network analysis reveals the association of important functional partners involved in antibiotic resistance: A report on an important pathogenic bacterium *Staphylococcus aureus*. *Gene*; 575(2): 253-263.
7. Jensen LJ, Kuhn M, Stark M, et al. (2008) STRING 8—a global view on proteins and their functional interactions in 630 organisms. *Nucleic acids res*; 37(suppl_1): D412-D416.
8. Barabasi AL, Oltvai ZN (2004) Network biology: understanding the cell's functional organization. *Nature reviews. Genetics*; 5(2): 101-113.
9. Ideker T, Krogan N J (2012) Differential network biology. *Molecular systems biology*; 8(1): 565.
10. Warde-Farley D, Donaldson SL, Comes O, et al. (2010) The GeneMANIA prediction server: biological network integration for gene prioritization and predicting gene function. *Nucleic acids res*; 38(suppl_2): W214-W220.