GPU-based Massively Parallel Quantum Inspired Genetic Algorithm for Detection of Communities in Complex Networks

Shikha Gupta
Department of Computer Science
University of Delhi
New Delhi, India
shikha.gupta.cs.du@gmail.com

Naveen Kumar
Department of Computer Science
University of Delhi
New Delhi, India
nk.cs.du@gmail.com

ABSTRACT

The paper presents a parallel implementation of a variant of quantum inspired genetic algorithm (QIGA) for the problem of community structure detection in complex networks using NVIDIA® Compute Unified Device Architecture (CUDA®) technology. The paper explores feasibility of the approach in the domain of complex networks. The approach does not require any knowledge of the number of communities beforehand and works well for both directed and undirected networks. Experiments on benchmark networks show that the method is able to successfully reveal community structure with high modularity.

Categories and Subject Descriptors

B.2 [**ARITHMETIC AND LOGIC STRUCTURES**]: Design Style—*Parallel*; D.2.2 [**Design Tools and Techniques**]: [Top-down programming]

Keywords

Parallel Evolutionary Algorithm; Complex Network; Community Detection; Quantum-Inspired Genetic Algorithm; Modularity

1. INTRODUCTION

A network is a system of interconnected entities in a domain and is used to represent relationships between these entities. Many of the complex networks of interest such as online social networks, WWW, and biological networks are too large to be handled by traditional computing architectures. Further, many of these networks can be considered as complex networks characterized by collective behavior that does not follow trivially from the behaviors of the individual entities in the network [6]. One such behavior is hierarchical structure in the network, the networks at the lower level of hierarchy (sub-networks) being popularly termed as *communities*. The complex network analysis is both computationally and memory intensive and is thus well-suited for parallel programming paradigm. In this paper, we present a parallel implementation of a variant of QIGA [4] for community structure detection in complex networks. The proposed approach employs a top-down bi-partitioning strategy to optimize network modularity [8], deploying a QIGA at each level of the hierarchy. The algorithm is implemented on NVIDIA's massively parallel Compute Unified Device Architecture (CUDA) technology, one of the leading general-purpose parallel computing architectures, and employs single-population fine-grained parallel evolutionary approach[1].

The proposed algorithm maintains a population of quantum chromosomes, where the size of the chromosome equals the number of nodes in the graph. An n-qubits chromosome can represent 2^n states at the same time but each qubit collapses to a single state 0 or 1 when measured, resulting in a classical binary chromosome. In each generation of the algorithm, the measurement of each qubit generates a classical bit probabilistically based on its state of superposition and is followed by an application of mutation and quantum rotation gate as variation operators. The lookup table for quantum rotation gate is as given by Han and Kim in [4] with,

$$\theta_1{=}\theta_7{=}0,\ \theta_2{=}\theta_4{=}{-}\theta_6{=}\theta_8{=}0.001,\ \theta_3{=}{-}\theta_5{=}0.038.$$

For example, for a 12-node network, consider the following 12-qubit chromosome,

$$q = \left[\begin{array}{c|c|c|c|c} \frac{\sqrt{5}}{\sqrt{6}} & \frac{\sqrt{3}}{2} & \frac{-1}{\sqrt{2}} & \cdots & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{6}} & \frac{1}{2} & \frac{1}{\sqrt{2}} & \cdots & \frac{-1}{\sqrt{2}} \end{array}\right]$$

where the probability of occurrence of state 0 in the first qubit is 5/6 and of state 1 is 1/6. This chromosome, when measured, may yield a binary chromosome such as

$$\begin{array}{|c|c|c|c|c|c|c|c|c|c|c|c|} \hline 0 & 1 & 1 & 1 & 0 & 0 & 1 & 1 & 0 & 1 & 1 & 0 \\ \hline \end{array}.$$

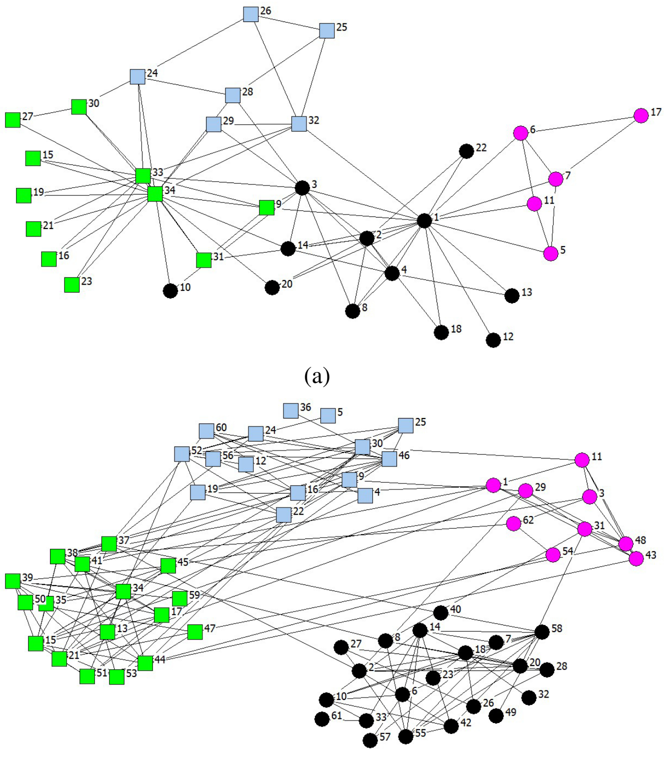
All the nodes in the graph are initially assigned to the same community 0 (assuming that the community IDs to be assigned are 0, 1, . . .) and the first bi-partitioning step divides the network into two communities 0 and 1. In this context, the values 0 and 1 in the above chromosome may be interpreted as community IDs associated with the corresponding nodes. That is, nodes 1, 5, 6, 9, and 12 belong to the community 0 while nodes 2, 3, 4, 7, 8, 10, and 11 belong to the community 1. The existing communities are repeatedly bi-partitioned in a depth first manner. That is, in the example, at the next level in bi-partitioning, nodes in community 0 are split into communities 0 and 2, while nodes in community 1 (assuming that communities 0 and 2 cannot be split further) are split into communities 1 and 3, and may yield a chromosome such as

$$\begin{array}{|c|c|c|c|c|c|c|c|c|c|c|c|} \hline 2 & 3 & 1 & 1 & 0 & 0 & 1 & 3 & 2 & 3 & 1 & 2 \\ \hline \end{array}.$$

The fitness function (network modularity) is computed after every partitioning. The partitioning that increases the modularity is accepted. The process terminates when no further increase in the modularity is possible.

The algorithm is implemented in CUDA C assigning each element

GECCO'14, July 12–16, 2014, Vancouver, BC, Canada.
ACM 978-1-4503-2881-4/14/07.
http://dx.doi.org/10.1145/2598394.2598437.



(a)

(b)

Figure 1: Communities and sub-communities identified in (a) Zachary karate club network and (b) Lusseau dolphins social network using the proposed algorithm. Vertex colour indicates community membership in the final partition. Vertex shape represents the principle division into two communities

.

(qubit) of the chromosome to a separate thread. A chromosome may then be spread over multiple blocks in a grid. Each of the functions of QIGA- initialize, measure, mutate, update, and fitness function (modularity) evaluation, are implemented as GPU kernels.

2. EXPERIMENTAL RESULTS

We have applied the proposed algorithm on real-world benchmark datasets (Figure 1)- Zachary karate club network [12] and Lusseau dolphins network [5]. Table 1 shows the comparison of the modularity value results of the proposed algorithm with those of Newman-Fast [7], GN-Fast [8], CNM [2], Walktrap [10], DA [3], GATB [11], and MOGA-Net [9] algorithms. Karate club network has two groups in the original partition centered on nodes 1 and 33. If we cut the dendrogram generated by the proposed solution strategy at two communities, we obtain a modularity value of 0.3718, higher than 0.36 of the original structure. Only nodes 3 and 10 are misclassified. It is worth noting that node 3 is linked to the principal nodes 1 and 33, and links to equal number of nodes in both communities of known partition. It is therefore possible for node 3 to get placed in any community. Also, node 10 is linked to only two nodes, one being node 3, and is placed in the same community as that of node 3 in both original and found partition. The algorithm finally splits the network into 4 communities with the maximum modularity value of 0.4188. Dolphins network has two groups in the original partition. If we cut the dendrogram generated by the proposed solution strategy at two communities, we obtain a modularity value of 0.4015, higher than 0.4004 of the original structure. Only nodes 40, 54, and 62 are misclassified. It is worth noting that all the three nodes are in the same community in both known and found partitions. The algorithm further splits the two communities found and terminates with a maximum modularity of 0.5268.

Table 1: Comparison of the modularity (Q) values.

Algorithm	Zachary	Dolphin
PQG-CD	.4188	.5268
Newman-Fast	.3807	.4955
GN-Fast	.4013	.519
CNM	.381	.515
Walktrap	.394	.517
DA	.4188	.5264
GATB	.402	.52
MOGA-Net	.416	.505

3. ACKNOWLEDGEMENTS

The work is supported by University of Delhi research grant no. DRCH/R&D/2013-14/4155. The authors acknowledge the support provided by NVIDIA in using the GPU cluster facility at its Pune design center which has been extremely useful in carrying out this research work.

4. REFERENCES

[1] E. Cantú-Paz. A survey of parallel genetic algorithms.

[2] A. Clauset, M. E. J. Newman, and C. Moore. Finding community structure in very large networks. *Phys. Rev. E*, 70:066111, Dec 2004.

[3] J. Duch and A. Arenas. Community detection in complex networks using extremal optimization. *Physical review E*, 72(2):027104, 2005.

[4] K.-H. Han and J.-H. Kim. Quantum-inspired evolutionary algorithm for a class of combinatorial optimization. *IEEE TRANSACTIONS ON EVOLUTIONARY COMPUTATION*, 6(6):580, 2002.

[5] D. Lusseau. The emergent properties of a dolphin social network. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, 270(Suppl 2):S186–S188, 2003.

[6] M. Newman. Complex systems: A survey. *Am. J. Phys.*, 79(arXiv: 1112.1440):800–810, 2011.

[7] M. E. Newman. Fast algorithm for detecting community structure in networks. *Physical review E*, 69(6):066133, 2004.

[8] M. E. J. Newman and M. Girvan. Finding and evaluating community structure in networks. *Physical Review*, E 69(026113), 2004.

[9] C. Pizzuti. A multiobjective genetic algorithm to find communities in complex networks. *Evolutionary Computation, IEEE Transactions on*, 16(3):418–430, 2012.

[10] P. Pons and M. Latapy. Computing communities in large networks using random walks. *Journal of Graph Algorithms and Applications*, 10(2):191–218, 2006.

[11] M. Tasgin and H. Bingol. Community detection in complex networks using genetic algorithm. In *ECCS '06: Proc. of the European Conference on Complex Systems*, Apr. 2006.

[12] W. W. Zachary. An information flow model for conflict and fission in small groups. *Journal of anthropological research*, pages 452–473, 1977.