

AN EFFICIENT IMPLEMENTATION OF GRAPH CLUSTERING TO CREATE KNOWLEDGE BASE FOR INTERACTOMICS ANALYSIS

K.Venkatasubramanian¹, Dr. S.K.Srivatsa², Dr. C. Parthasarathy³

¹Research Scholar, SCSVMV University, Kanchipuram

¹Associate Professor, RCE, Eluru

²Retd. Prof. Anna University, Chennai - 602 025.

³Assistant Professor, Dept. of IT, SCSVMV University Kanchipuram

ABSTRACT

This paper proposes an efficient implementation of graph clustering in the interactomics analysis of the organism, which helps to molecular processes within the cell structures in the human body. The major objective of this paper is to develop the knowledge base for the interactomics through Interest Ontology (IO), Graph Clustering (GC) and Hybrid Filtering (HF). The HF includes Collaborative Filtering (CF) and Content Based Filtering (CBF). The interaction between the cells within human body can be periodically changed due the systematic process of the molecules. The enzymes secreted in the glands are transferred between each cell in our human body. This process is called Protein-Protein Interaction (PPI). The valid knowledge base is needed to diagnose these types of interactions. The knowledge base describes the set of cells are going to interact with each other. The set of cells are formed as a cluster and differentiate them with non performing cells by clustering graph followed by the content based filtering.

Index Terms: Interactomics, Protein-Protein Interaction (PPI), Interest Ontology (IO), Graph Clustering (GC), Content Based Filtering (CBF), Collaborative Filtering (CF), Binary Tree, and Hybrid Filtering (HF).

INTRODUCTION

The concept of graph clustering is the major application for the Protein – Protein Interaction (PPI). The blood cells within human body have communicating each other based on the molecular processes. The blood cells are communicating with other blood cells through the intermediate blood cells. The intermediate blood cells are act as a communication medium. The mode of communication is classified as mesh and nodal topology.

Each and every mesh is formed as a cluster and communicates with another cluster through intermediate nodes [1]. To implement these types of structures, the graph clustering is proposed. The hybrid filtering [2], which has the combination of content based filtering [3] and collaborative filtering [4] is also implemented as a filtering technique for protein structures. The graph clustering is implemented in the blood cells and the output of the filtering module and clustering module are processed by binary tree to obtain the result.

1. PROPOSED METHODOLOGY

At first, the user separates the protein structures and the blood cells. The interest ontology is implemented in the protein structures and the graph clustering is implemented in the blood cells. The result of the interest ontology is filtered by hybrid filtering technique. The obtained output of graph clustering and the hybrid filtering are combined with the help of binary tree [5] to get the suitable diagnosis result. The figure 1 shows the block diagram of the proposed method.

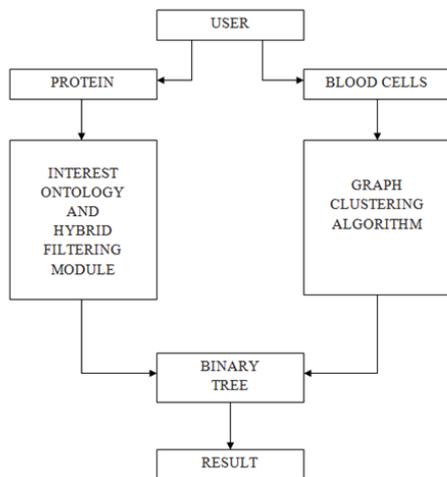


Figure 1 Block diagram of the proposed method

A. Interest Ontology (IO)

The interest graph ontology is implemented in the protein structures by separating the topic interest. The topic interests are the organism, hemopoietic stem cell differentiation, embryonic hemopoiesis and gene (COX2). The inmodal organism obtained from the hemopoietic stem cell differentiation is given to the organism for the further molecular processes. The figure 2 shows the block diagram of the process of interest ontology.

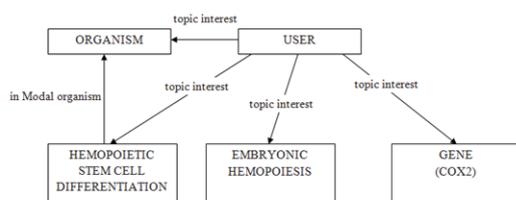


Figure 2 Block diagram of the process of interest ontology

B. Hybrid Filtering

The hybrid filtering technique is a combination of the collaborative filtering and content based filtering. The output from the interest ontology is given to both collaborative filter and content based filter. These two

blocks give the recommendations for the input to get the preferred output [6]. The combiner makes the fusion process of the recommendations from the filters and produces the recommended output [7] as the process of hybrid filtering. The figure 3 shows the block diagram of the process of interest ontology.

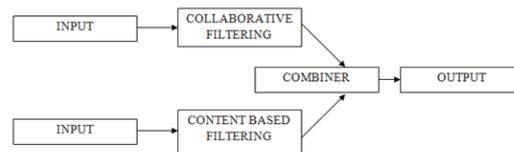


Figure 3 Block diagram of the process of hybrid filtering

C. Graph Clustering

The blood cells are connected and processed with the help of graph clustering. The figure 4 shows the successive stages of the graph clustering.

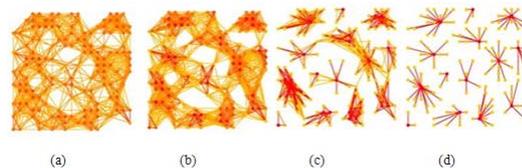


Figure 4 Successive stages of the graph clustering

The R code for the graph clustering by using Highly Connected Subgraphs (HCS) is following,

```

library (GraphClusterAnalysis)
library (RBGL)
library (igraph)
library (graph)
data (HCS_Example)
G = graph.data.frame (HCS_Example, directed=FALSE)
HCSClustering (G, kappa=2)
    
```

2. EXPERIMENTAL OUTCOMES

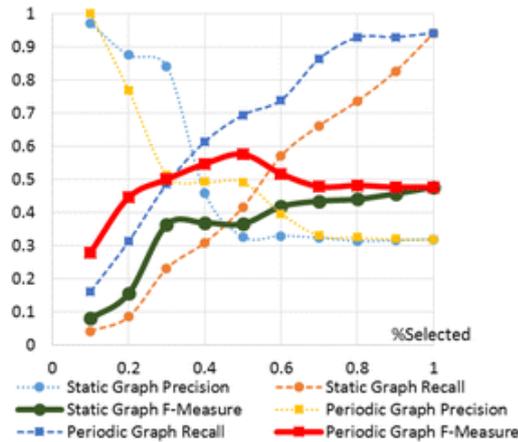


Figure 5 Clustering graph for selecting the corresponding cells

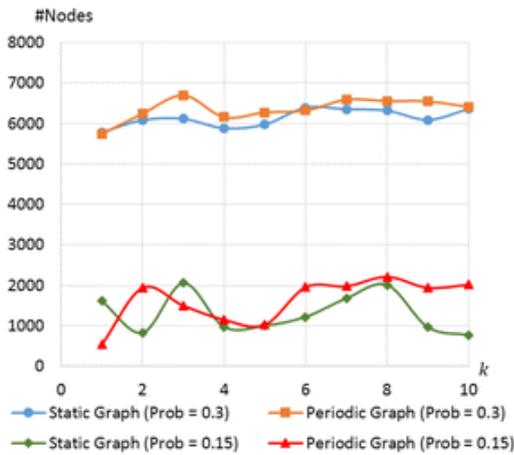


Figure 6 Clustering graph of interaction between the nodes

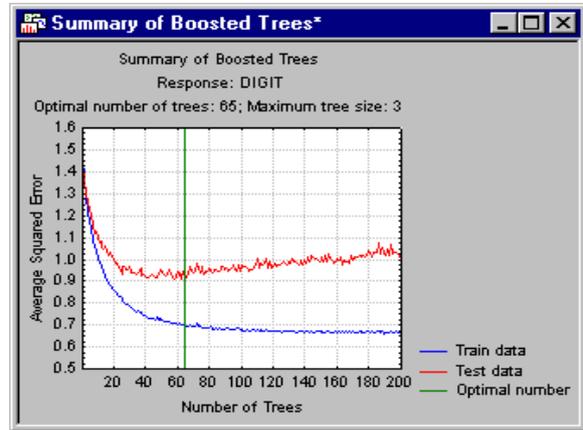


Figure 7 Summary of the binary tree

3. CONCLUSION

Thus the output of the graph clustering and the hybrid filtering module are combined by the binary tree to obtain the preferred output. This can be achieved by the process of the proposed methodology. The obtained results are more efficient than the existing methods of graph clustering.

REFERENCES

- [1]. L Martin, A Loukas and P Vandergheynst, (2017), "Fast Approximate Spectral Clustering for Dynamic Networks", arXiv preprint arXiv:1706.03591.
- [2]. H Meyerhenke, P Sanders and C Schulz, (2017), "Parallel graph partitioning for complex networks", IEEE Transactions on Parallel and Distributed Systems 28 (9), pp.2625-2638
- [3]. H Meyerhenke, P Sanders and C Schulz, (2016), "Partitioning (hierarchically clustered) complex networks via size-constrained graph clustering", Journal of Heuristics 22 (5), pp.759-782.
- [4]. R Peng, H Sun and L Zanetti, (2015), "Partitioning well-clustered graphs: Spectral clustering works", Conference on Learning Theory, pp.1423-1455
- [5]. D Tamascelli and L Zanetti, (2014), "A quantum-walk-inspired adiabatic algorithm for solving graph isomorphism problems", Journal of Physics A: Mathematical and Theoretical 47 (32), 325302

- [6]. H Sun and L Zanetti, (2016), "Distributed graph clustering by load balancing", arXiv preprint arXiv:1607.04984
- [7]. H Sun and L Zanetti, (2017), "Distributed Graph Clustering and Sparsification", arXiv preprint arXiv:1711.01262
- [8]. M. Anastos, (2017), "Randomly coloring simple hypergraphs with fewer colors", Information Processing Letters 126, pp.39-42.
- [9]. W. Pegden, (2017), "Travelling in randomly embedded random graphs", Extended abstract in RANDOM 2017.
- [10]. T. Johansson, (2017), "Minimum-cost matching in a random graph with random costs", SIAM Journal on Discrete Mathematics 31, pp.489-510.
- [11]. T. Johansson, (2017), "On random k-out subgraphs of large graphs", Random Structures and Algorithms 50, pp.143-157.
- [12]. A. Dudek, A. Rucinski and M. Siliakis, (2017), "Embedding the Erdos-Renyi Hypergraph into the Random Regular Hypergraph and Hamiltonicity", Journal of Combinatorial Theore B, 122C, pp.719-740.
- [13]. Yunxiao Chen, Xiaou Li, Jingchen Liu, Gongjun Xu and Zhiliang Ying, (2017), "Exploratory Item Classification Via Spectral Graph Clustering", Volume: 41 issue: 8, pp. 579-599
- [14]. Paul Swoboda, Carsten Rother, Hassan Abu Alhaija, Dagmar Kainmüller and Bogdan Savchynskyy, (2017), "A study of Lagrangean decompositions and dual ascent solvers for graph matching", Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), vol.5, 2017/7/1.

