Genome-scale Reconstruction of Metabolic Pathways in Cancer to Investigate Algorithms of Community Detection Based on Stoichiometry Matrix Sparsity to Discover Protective Role of Over-expressed Pathways in Cancer Development

BY

MOHAMMAD KASHKOULI
ELAHE MOHAMMADI
SANAZ RAMEZANZADE

FINAL REPORT OF PROJECT OF OPTIMIZATION METHODS IN METABOLIC NETWORKS

DEPARTMENT OF MATHEMATICAL SCIENCE

SHARIF UNIVERSITY OF TECHNOLOGY

FALL 2020

Project Title	Genome-scale reconstruction of metabolic pathways in cancer to investigate algorithms of community detection based on stoichiometry matrix sparsity and flux inhibition analysis to discover protective role of overexpressed pathways in cancer development
By	Mohammad Kashkouli
	Elahe Mohammadi
	Sanaz Ramezanzade
Field of Study	Mathematics, Computer science, Biology
Project Advisor	Dr. Mojtaba Tefagh
Project Co-Advisor	Dr. Mohammadhossien Motealehi
	Dr. Ehsan Zangane
Academic Years	2020

Keywords: : Metabolic network, Clustering, Hyper-graph, Biological network analysing, Glioblastoma

Availability of codes and materials: https://github.com/mohammad-kashkooli/metabolomics

About data:: in this study we work on metabolic network with 548 metabolite and 625 reaction of genome scaled reconstructed model of Glioblastoma cancer in which primary data is taken from ”https://www.frontiersin.org/articles/10.3389/fnins.2016.00156/full”

TABLE OF CONTENTS

1 Introduction

A metabolic network is the complete set of metabolic and physical processes that determine the physiological and biochemical properties of a cell. Metabolic networks can be considered as particular cases of reaction networks. Metabolic pathways are represented in the form of graphs in which nodes represent molecular entities and edges represent reactions or processes relating the molecules involved in the reaction. However, since a reaction may have more than one substrate, and more than one product, the pathway is better represented by a hyper-network in which hyper-edges represent reactions and nodes represent molecular entities.[7]Tools from graph theory have previously been applied to the analysis of structural properties of not only metabolic networks but also many other ones.One of these tools is community detection or clustering. Investigating sub-communities would be helpful to find meaningful gene pathways in some diseases. Therefore this would be a gate to deign medicine. In this project we struggle to cluster a metabolic network based on flux analysing to find some meaningful results.Cancer is one of the leading causes of mortality all over the world[8]. Developing better medications is dependent to better understanding of omics data through different analytical approaches. Systems biology tools and high throughput technologies let us integrate and analyze different types of omics data. In this study, genome-scale reconstruction of metabolomics network will be used because analyses in metabolomics level has the advantage that we can see effect of genes expression on phenotype of tumor cells.it is proposed to develop algorithms of subnetwork detection based on stoichiometric matrix and investigate whether they show meaningful biological metabolic subnetwork in tumor cells or not. Therefore, these subnetworks help us fill gaps of model and interrupt their cross talks.

2 SPECTRAL CLUSTERING OF HYPER-GRAPH

2.1 Preliminaries

Suppose a hyper-graph $G = (V, E, w)$ that H is the incident matrix of hyper-graph means that $H(u, e) = 1$ if vertex $u \in e$. Also let $\delta(e) = |e|$ and $d(v) = \sum_{e \in E, v \in e} w(e)$ be degree of edges and vertices, respectively. Let D_v and D_e diagonal matrices with degrees of vertices and edges. Also let A be the adjacency matrix of network.

Definition .1. *A cut of a hyper-graph $G = (V, E, w)$ is a partition of V into two parts S and S^c : we say that a hyper edge e is cut if it is incident with the vertices in S and S^c simultaneously.*

Definition .2. *Given a vertex subset $S \in V$ the boundary of S is $\partial S = e \in E : e \cap S, e \cap S^c$ and define the volume of S to be the sum of the degrees of the vertices in S, that is,$vol(S) := \sum_{v \in S} d(v)$. Moreover, define the volume of ∂S by*

$$vol(\partial S) := \sum_{e \in \partial S} w(e) \frac{|e \cap S||e \cap S^c|}{\delta(e)}$$

The ideal partitioning is that the connection among vertices in the same cluster is dense as the boundary is sparse. To show this as an optimization problem to cluster hyper-graph problem [1] should be solved:

$$minimize \ \ (vol(\partial S))(\frac{1}{vol(S)} + \frac{1}{vol(S^c)})$$

The hyper-graph normalized cut can be explained in terms of random walks. Given the current position $u \in V$ choose a hyper-edge $e$over all hyper edges incident with u with the probability proportional to $w(e)$; and then choose a vertex $v \in V$ uniformly at random. Obviously, it generalizes the natural random walk defined on simple graphs. For more details refer to [1]. The important result of reference [2] leads us to solve the problem.

Theorem .3. *Combinatorial optimization problem* [1] *is Np-hard.*

Proof. reference [2] □

2.2 Sub-graph centrality and clustering coefficient

Lemma .4. *If the hyper-graph G is connected, then the symmetric and non-negative matrix A is irreducible.*

Proposition .5. *The main eigenvalue has a positive eigen-vector of multiplicity one.*

Lemma .6. *The number of walks of length k in G, from v_i to v_j, is the entry in position (i, j) of the matrix A^k.*

Lemma .7. *The number of walks of length k in G is given by:*

$$W_k = \sum_{i,j} (A^k)_{i,j}$$

Lemma .8. *the number of closed-walks of length k in G is given by:*

$$CW_k = \sum_{i,j} (A^k)_{i,i}$$

Definition .9. *clustering coefficient for a hyper-graph:*

$$C_2(H) = \frac{6 * number\ of\ hyper - triangles}{number\ of\ 2 - paths}$$

3 Algorithms

3.1 Clustering algorithm

Based on analysis of reference [1] we do develop an algorithm for hyper-graph clustering for metabolic network.

inputs: incident matrix, degree matrices, and weight matrix

step 1: compute normalized laplacian

$$L = I - \Phi$$

$$\Phi = \sqrt{D_{\mathrm{v}}^{-1}} H W D_{\mathrm{e}}^{-1} H^T \sqrt{D_{\mathrm{v}}^{-1}}$$

step 2: compute spectral decomposition of normalized laplacian.

step 3: form matrix $X = [\phi_1, ...\phi_k]$ that ϕ_i's are the eigen-vectors of normalized laplacian corresponding to $\lambda_2 \leq \lambda_3 \leq ... \leq \lambda_{k+1}$

step 4: run k-means on rows of X

3.2 Calculate Clustering coefficient algorithm

Based on analysis of reference [6] we do develop an algorithm to calculate clustering coefficient for metabolic network.

inputs: adjacency matrix A

step 1: compute spectral decomposition of $A = UDU^T$

step 2: compute closed walks of lengths one and two

$CW2 = trace(D^2)$ and $CW3 = trace(D^3)$

step 3: compute walks of length two

$$W2 = \sum_{s=1}^{n}(\sum_{i=1}^{n} u_{is})^2 \lambda_{\mathrm{S}}^2$$

step 4: compute clustering coefficient:

$$C_2(G) = \frac{CW3 - 6t}{W_2(G) - CW2 - 6t}$$

4 REVIEW OF LITERATURE

4.1 Hyper-graph Clustering

Like in most mathematical models, hyper-graphs has many applications. Researchers use hyper-graphs to model different networks such as social networks, metabolic networks, protein–protein interaction networks, and neural networks. One of the ways in computer science for analysing a network is ”clustering”. In the machine learning community, there has been increasing interest around this problem.[3]

One one the approaches for hyper-graph clustering is spectral methods which we have used it too. These methods are based on Laplacian matrix and cut the network with the eigen-vector corresponding to the second smallest eigenvalue. This approach can be generalized to weighted or directed hyper-graphs too. Spectral methods also called partition-based methods. One of the well known and beautiful algorithms is of course page rank. In conference of ACM 2020 an article published ”Hyper-graph Clustering Based on Page Rank”

One of the other approaches is based on game theory. Let us refer to [3] as one of the articles with game theory approach. In the latest two weeks ago an article [4]

has been published. this article develop an algorithm for Microbial network one have amazing result.

5 MAIN RESULT

We have data of stoichiometric matrix of hyper-graph of metabolic network. First we do Flux Balanced Analysis to have the flux of each hyper-edge.

In the second part we implement the algorithm of section 3.1 on this problem with the assumption that: the wight of the hyper-edges are fluxes.**As matter of fact we try to cluster the network based on the flux**.

This study involves of some important metabolic pathways which have been found based on their hyper-edge weights. Reconstructed genome scale metabolic model was used as our clustering input and metabolites clustered in 5 clusters including Warburg effect, Gluconeogenesis, Citric Acid Cycle, Glycolysis, Glutamine metabolism, Lipid metabolisms. Interestingly biomass reaction components which were lipid ,ATP , and protein were clustered in related modules which shows flow of those metabolites toward them. Our results were consistent with other Glioblastoma models in lipid metabolism and Gluconeogenesis aspect as Metformin was previously suggested as one of repurposed drugs for inhibition of these axes in Glioblastoma.

We also implement the algorithm of section 3.2 to calculate clustering coefficient which is a parameter tests how the clustering is good.

6 CONCLUSION AND DISCUSSION

In this study, we do Flux balanced analysis on reconstructed metabolic network of Glioblastoma cancer as one of the aggressive type of cancers. We haves studied and reviewed hyper-graph clustering and implement a spectral clustering algorithm based on fluxes. This helps us to detect communities and find meaningful results to suggest medicine.

Appendix A diagrams

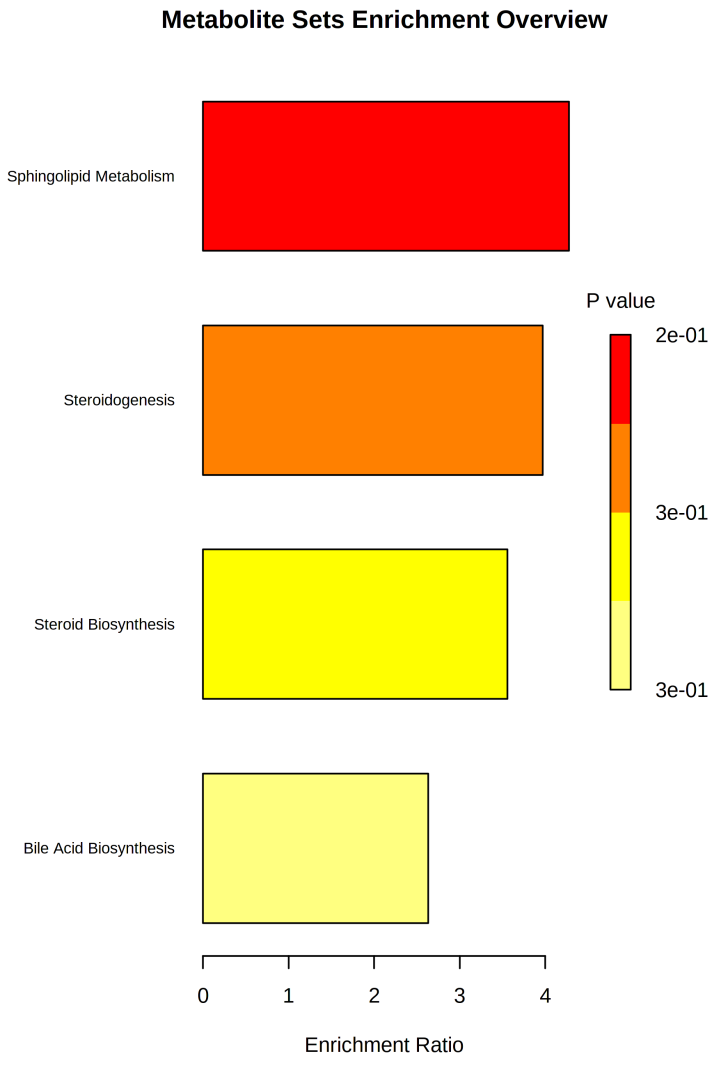

Figure 1: cluster4

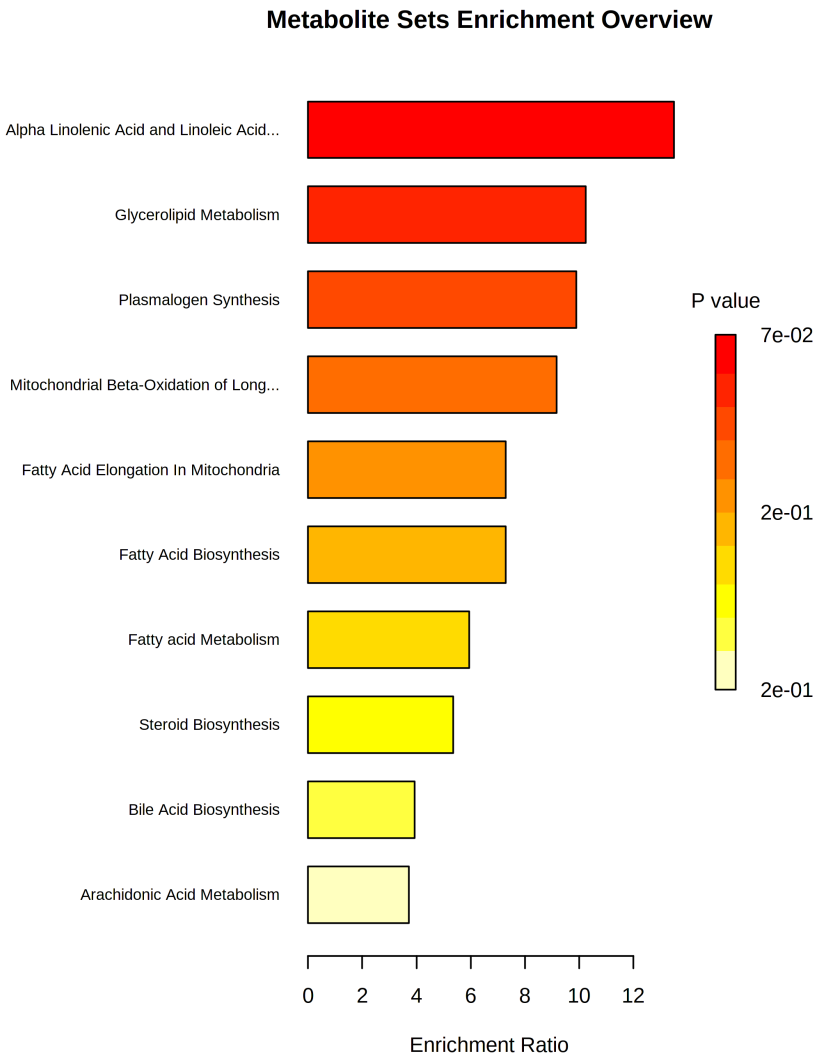

Figure 2: cluster2,3

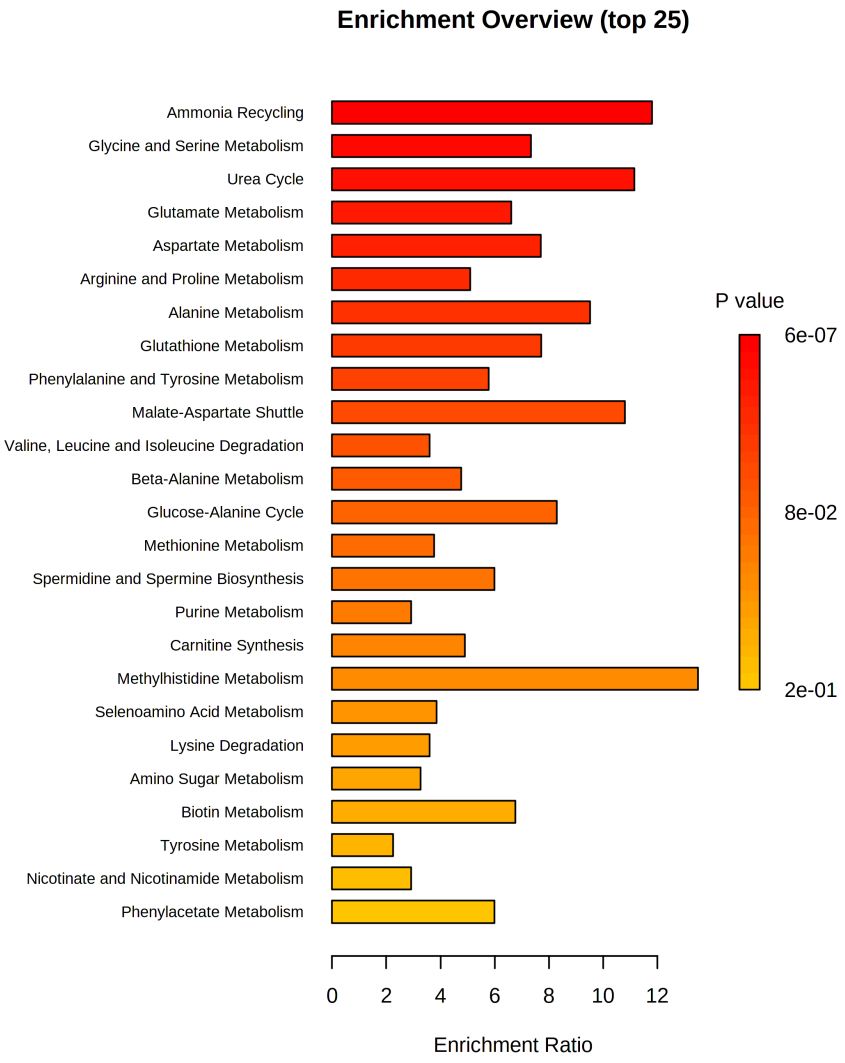

Figure 3: cluster1

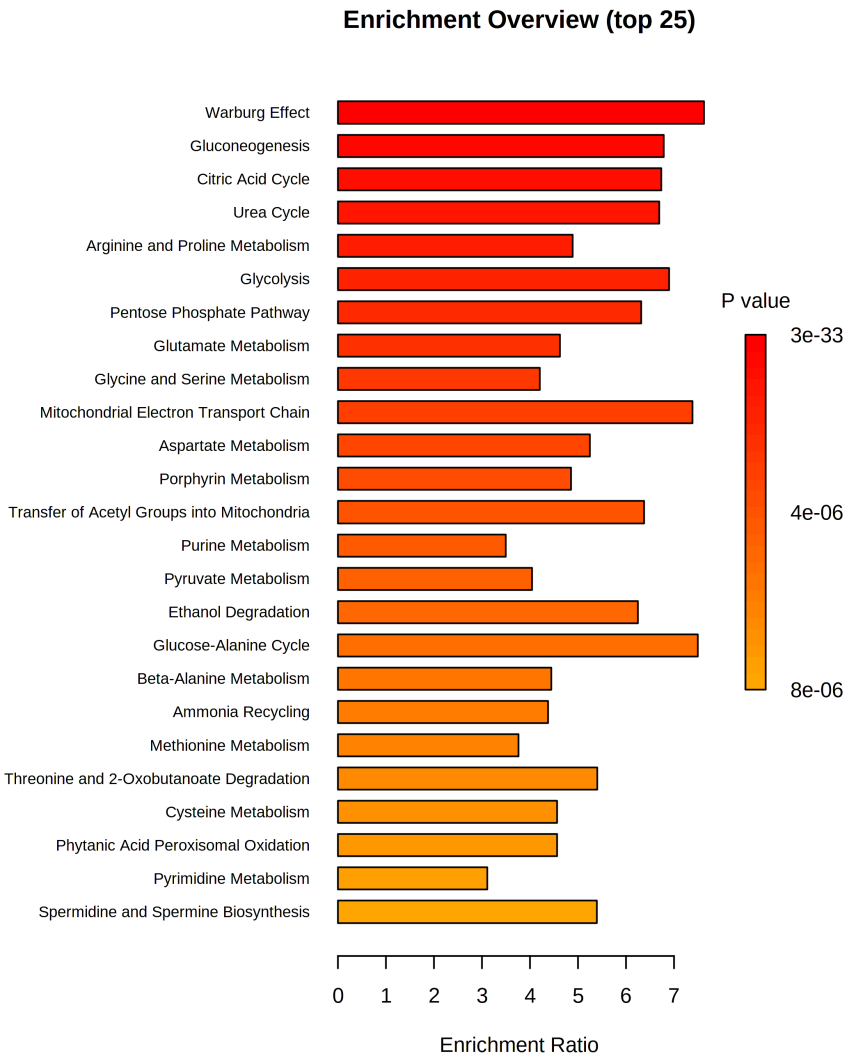

Figure 4: cluster0

References

[1] D. Zhouy, J. Huangz, B. SchÄolkopfx; “Learning with Hypergraphs: Clustering, Classification, and Embedding”; advances in neural information processing systems; 2006

[2] J. Shi, J. Malik; ”Normalized cuts and image segmentation”; IEEE Tran. on Pattern Analysis and Machine Intelligence; 22(8):888905; 2000

[3] S. R. Bulo, M. Pelillo; “A Game-Theoretic Approach to Hypergraph Clustering”; ieee transactions on pattern analysis and machine intelligence; vol 35; issue 6; pages 1312-1327; 2013

[4] L. Yu, X. Shen, J. Yang1, K. Wei, D. Zhong, R. Xiang; ”Hypergraph Clustering Based on Game-Theory for Mining Microbial High-Order Interaction Module”; Evolutionary Bioinformatics; Volume 16: 1–8; 2020

[5] Y. Takai, A. Miyauchi, M. Ikeda, Y. Yoshida; “Hypergraph Clustering Based on PageRank”; proceedings of the 26th ACM SIGKDD in international conference on knowledge discovery and data mining; pages 1970-1978; 2020

[6] E. Estradaa, J. A. Rodrı´guez-Vela´zquezb; ”Subgraph centrality and clustering in complex hyper-networks”; Physica A: Statistical Mechanics and its Applications; Volume 364; Pages 581-594; 2006

[7] L. Krishnamurthy, J. Nadeau, G. Ozsoyoglu, M. Ozsoyoglu, G. Schaeffer, M. Tasan, W. Xu; ”Pathways database system: an integrated system for biological pathways” Bioinformatics 19; 2003

[8] Organization, W.H. Cancer. 12 September 2018; Available from: https://www.who.int/news-room/fact-sheets/detail/cancer