ADV MATH SCI JOURNAL
Advances in Mathematics: Scientific Journal **9** (2020), no.5, 2397–2404
ISSN: 1857-8365 (printed); 1857-8438 (electronic)
https://doi.org/10.37418/amsj.9.5.3

CONSTRUCTION OF COLOR GRAPHS

G. PRINCESS RATHINABAI[1] AND G. JEYAKUMAR

ABSTRACT. To study and to propose methods to classify the nature of problems in biological tissues, particularly in cancer tissues of humans, the affected tissues have to be converted into graphs called color graphs. Construction of color graphs itself is a very challenging problem. In this paper, we study the existing method of constructing color graphs and introduce a new and novel form of constructing color graphs which has several advantages over the existing method.

1. INTRODUCTION

A graph G is an ordered pair $(V(G), E(G))$ where $V(G)$ is a non empty set and $E(G)$ is a collection of unordered pairs of distinct elements of $V(G)$. Elements of $V(G)$ are called vertices or nodes and that of $E(G)$ are called edges. If $e = (u, v) \in E(G)$, we say that the edge e joins the vertices u and v and denote the edge e by uv.

Graph theory plays a vital role in the study of biological structures since their characteristics are studied by converting them in to a mathematical object. For example, Protein-Protein Interaction Networks [12], Regulatory Networks, Signal Transduction Networks, Metabolic and Biochemical Networks, etc., are some biological networks modeled and studied in the light of Graph Theory.

[1]*corresponding author*

2010 *Mathematics Subject Classification.* 05C15.

Key words and phrases. color graphs, tissues, graph theory.

Apart from these, several graphs have been generated in order to quantify the tissue images. These graphs are widely responsible in encoding cancer tissue images and cancer grading [4]. The graphs for tissue images are constructed mainly through the location of cells (or cell clusters) [4] and tissue components [8, 9]. Some of these graphs are constructed for tissues that belong to brain tumor (glioma) [5, 6] and colon cancer [9] classes. Since graphs are efficient tools in defining the binary relationship between nodes, these graphs are significant tools in quantifying the tissue images through connectivity information [4]. Cell graphs [5], Augmented cell graphs (ACG) [6] and color graphs [9] are some of the graphs generated to represent a tissue image.

Cell-graphs are graphs in which cells or cell-clusters of a tissue image are nodes and edges are defined between pair of nodes under some biological foundation [4]. Using the technique of cell-graph, the tissues are classified as cancerous, healthy and inflamed tissues with accuracy [6]. The tissue images are classified using some topological properties like average degree, average clustering coefficient, average eccentricity, ratio of the giant connected component, percentage of end nodes, percentage of isolated nodes, spectral radius and eigen exponent [6, 7]. Similar to cell graph, ACG is an alternative computational method in which the nodes are cell clusters and edges are defined between pair of nodes under certain probability. Here, the nodes and edges are assigned weights to obtain more information on the topology of tissues. Average degree, eccentricity, average node weight, edge weight, graph spectrum, etc., are some topological properties involved in identifying the nature of the tissue images. These cell graph construction depends on the location of cells (or clusters). Numerous cell-graph construction methods are explained in [3]. Graphs for tissue images can also be generated based on its components. For example, in the construction of color graphs [9], the centroids of the tissue components are assumed as nodes and edges are established using Delaunay triangulation [9]. Assignment of colors to these edges depends on the end points. For colon tissue images, this technique has been introduced and resulted in an ample accuracy in the differentiation of tissue images. The color graph technique is explained in the following section in detail. An attributed graph [10] $G = \{V, E, \mu\}$, where V is a set of nodes, $E \subseteq V \times V$ is a set of edges, and $\mu : V \rightarrow A$, where, A is a set of attributes, is a mapping that maps each node $v_i \in V$ into an attribute label $\alpha_i \in A$, is designed to represent a tissue image based on the

location of tissue components. The attributed graphs are used to represent the normal gland structure of a tissue image which in turn compared with query graphs [10] to classify the tissue images. Graphs also been modeled for segmenting histopathological images through which the spatial relations of tissue components are quantified [1]. Another graph called "object graph" [7] in which the tissue image is decomposed into set of object that represents the tissue components, is modeled to view the gland structures. Among these numerous graph models, this paper prioritize the color graph technique.

2. COLOR GRAPHS

Several approaches are used to quantify a tissue image based on the spatial location of cell nuclei [9] but characterizing a tissue image using its different components gives better insight in cancer study. For colon adenocarcinomatous tissue image, the color graphs [9] are constructed considering the different tissue components like nucleus, stroma and lumina. The initial step of color graph construction is "identification of nodes". Here the nodes correspond to nuclear, stromal and luminal tissue components. Due to the complex nature of tissue images, the exact boundary of these components cannot be determined. So the boundaries are approximated and the tissue components are identified as circular primitives with three color groups using "k - means clustering algorithm". The central point of these circular components are considered as nodes. Nuclear, stromal and luminal are the three types of circular primitives used to represent a tissue image in this approach. The colors purple, pink and white correspond to nuclear, stromal and luminal components respectively [9]. Edge Assignment is the next step of this construction. This is done by implementing the notion of Delaunay triangulation on central points (nodes) and the edges are colored considering the end points. Luminal - Luminal edge, Stromal - Stromal edge, Nuclear - Nuclear edge, Luminal - Stromal edge, Luminal - Nuclear edge and Stromal - Nuclear edge are the six possible edge combinations and hence six different colors are used to color these edges. Average degree, average clustering coefficient and diameter are some graph properties used to quantify the tissue images. In this paper, we propose some mathematical constraints to construct the color graph.

3. CONSTRUCTION OF COLOR GRAPHS

For the construction of color graph, we assume there are k - different components of a tissue. The two steps involved in the color graph construction are "Node Identification" and Edge Assignment. For identification of nodes, we follow the exact method given in [9]. Let the k - different components be $s_1, s_2, \ldots, s_k$. For each component s_i there corresponds a set of nodes n where $1 \leq i \leq k$ and let $n = n_1, n_2, \ldots, n_k$, the total number of nodes.

For a node u, we can assign its neighbor in any one of the following three methods.

Closed Disc Neighbors (CDN): Fix a real number $r > 0$. If the distance between a node v and u is less than or equal to r, we declare u and v are adjacent. (i.e)., $N(u) = \{v\ d(u, v) \leq r\}$.

The graph thus constructed is denoted by G.

Mutually Visible Neighbors (MVN): If a node v is visible from u, then we declare that u and v are adjacent and this graph is denoted by G_M.

Relatively Prime Neighbors (RPN): Two nodes $u = (a, b)$ and $v = (m, n)$, where a, b, m and n denote the integral part of the pixel values so that a, b, m and n are integers, are mutually visible if and only if $(a - m)$ and $(b - n)$ are relatively prime [13] and this graph is denoted by G_R.

We observe that in the CDN approach, for sufficiently small $r > 0$, the resulting graph is a null graph and for sufficiently large r, the resulting graph is complete. These extreme cases are ignored. The challenge is to find a suitable r which gives a graph G that fits for cancer grading. The graph G may be dense or sparse depending on the value of r.

In the MVN approach, edges are established between pair of nodes if one is visible from the other. Due to the complexity nature of tissue images, the graph G_M obtained through this approach is definitely not complete. Thus in this case, the resulting graph can be taken into account for cancer study. Inthe RPN approach, two nodes $u = (a, b)$ and $v = (m, n)$ are adjacent if and only if $(a-m)$ and $(b - n)$ are relatively prime. Note that the graph G_R obtained through RPN method is a subgraph of the graph G_M constructed through MVN method. Also in the CDN method, finding a suitable r is itself a great task.

Moreover, the assignment of edges in the existing color graph method has been done through the Delaunay triangulation technique. Due to the planarity

property of Delaunay triangulation the information on the relation between pair of nodes are obtained scarcely whereas in the MVN method it is entirely acquired. Thus we deduce that the graph G_M obtained through MVN approach is more suitable to comprehend the nature of a tissue and the MVN method is optimal to study the tissue images.

Assignment of colors to edges: Once the edges of the graph is established, the next step is to color the edges. This is done based on the tissue components. Since the adjacency is possible between homogeneous and heterogeneous components, the number of colors required to color the edges is $\frac{k(k+1)}{2}$, we denote this number by 2. The colors that indicate the components of the nodes of the edges are pre-fixed and they are assigned to the edges accordingly.

4. METRICS FOR COLOR GRAPH

We define the metrics for color graph in two ways:

I Ignoring the colors of the edges
II Considering the colors of the edges.

4.1. I. Metrics - ignoring the colors of the edges.

(1) **Average Degree:** The degree [11] of a node u in a graph G, denoted d_u, is the number of lines (edges) incident with u. The average degree of a graph G is computed over the degree of all nodes.
(2) **Average eccentricity:** The eccentricity [11] of a node u in a connected graph G is $\max d(u,v), \forall u \in V(G)$, where $d(u,v)$ is the distance between u and v. The average eccentricity of a graph G is computed over the eccentricity of all nodes. Also the maximum eccentricity of the nodes is the diameter $d(G)$ of a graph and the minimum eccentricity of the nodes is the radius $r(G)$.

4.2. **II. Metrics - considering the colors of the edges:** Edges of graphs give the binary relationship between pairs of nodes. In the color graph technique, the edges define the relationship between various components of tissues. In other words, assignment of color edges among different components conveys how the components are clustered within them. Here we define average degree, average

color intensity and average eccentricity with respect to π colors to grasp the distribution of color edges in a color graph.

(1) **Average Degree:** For a single node, there are π different degrees since there are edges of π colors which may incident with a node u, we define the degree of a node u with respect to the color i as d_{ui} = Number of edges of color i that incident with u. We observe that $d_u = \sum_{i=1}^{\pi} d_{ui}$. In this way, for the π colors the node degrees are calculated. Taking the average of node degrees with respect to each color, we get π average degrees.

(2) **Average Color Intensity:** The color intensity $C_u^{(i)}$ of the node u with respect to the color i where $C_u^{(i)} = \frac{E_u^i}{d_u}$ denotes the number of existing edges of color i incident with u and d_u is the number of neighbors of the node u. Taking the average of the color intensity of each color, gives a set of π - average color intensities for a color graph. For each node u we thus obtain the color intensity vector $C_u^{(i)} = \{C_u^{(1)}, C_u^{(2)}, \ldots, C_u^{(\pi)}\}$ of π - tuples. Further, we observe that $\sum_{i=1}^{\pi} C_u^{(i)} = 1$. The presence of 0 in the j^{th} co-ordinate indicates the non occurrence of j^{th} color edges at the vertex u. Further the linear combination, $\sum d_u C_u$ is a vector of π- tuples whose j^{th} co-ordinate represents the number of edges of color j. Also $\frac{1}{\pi}\sum_u C_u$ is a π - tuple vector whose j^{th} co-ordinate is the probability of a randomly selected edge of the color graph is of color j.

(3) **Average eccentricity:** For a color i, consider the subgraph H induced by the edges of color i. The eccentricity of a node u in H, denoted by e_{ui}, is the $\max d(u,v), \forall v \in V(H)$. Taking the average of the eccentricities with respect to π, we get π average eccentricities. The maximum eccentricity of the nodes of H with respect to the color i is the diameter $D_i(G)$ of the color graph and the minimum eccentricity of the nodes of H with respect to the color i is the radius $r_i(G)$ of the color graph. Also we observe that the computation of eccentricity, radius and diameter is possible only when H is connected and thus if the underlying subgraph H is connected for all colors, the eccentricity vector, the diameter vector and the radius vector can be computed.

The metrics thus defined, set a stage to classify the tissue images through the notion of color graphs.

The following are some parameters that can be studied through the color graph technique:

1. Maximum number of colors represented [2] at a vertex.
2. Minimum number of colors represented at a vertex.
3. Number of colors represented in every node. Such colors convey the relations between the corresponding components.
4. Dominance of certain color indicates that the corresponding components are the major factors in the behavior of a tissue.
5. Absence of certain colors gives the inferior governance of those components throughout the tissue.
6. Cycles of edges of same color, indicates a particular set of nodes n_j for some j, describe the close bond of the respective component.
7. Paths of edges of same color gives the connectivity information of the respective component.
8. Number of cliques of same / different color.
9. Number of trees of same / different color.
10. Number of K_n's of same / different color.
11. Number of $K_{1,n}$'s of same / different color.
12. Ratio of the total number of dominating color edges to the total number of edges of a color graph may help to fix a threshold in cancer grading.

These parameters will serve as an immense assistance in the field of Oncology to understand the organization of the components of the tissues.

5. Conclusion

In this paper, we proposed few methods to construct color graphs and some metrics to understand the nature of tissues. Also some parameters are suggested that can be used in cancer detection. The distribution of the π - different metrics that are derived for color graphs of various tissues can be done as a future research work.

References

[1] A. B. Tosun, C. G. Demir: *Graph Run-length Matrices for Histopathological Image Segmentation*, IEEE Transactions on Medical Imaging, **30**(3) (2011), 721–732.

[2] J. A. BONDY, U. S. R. MURTY: *Graph Theory with Applications*,The Macmillan Press Ltd, 1976.
[3] C. G. DEMIR: *Mathematical Modeling of the Malignancy of Cancer Using Graph Evolution*, Mathematical Biosciences, **209**(2007), 514 – 527.
[4] C. DEMIR, B. YENER, S. H. GULTEKIN: *The Cell Graph of Cancer*, Bioinformatics, **20**(1) (2004), i145 – i151.
[5] C. DEMIR, S. H. GULTEKIN, B. YENER: *Augmented Cell-Graphs for Automated Cancer Diagnosis*, Bioinformatics, **21**(2) (2005), ii7 – ii12.
[6] C. DEMIR, S. H. GULTEKIN, B. YENER: *Learning the Topological Properties of Brain Tumors*, IEEE/ACM Transactions on Computational Biology and Bioinformatics, **2**(3) (2005), 262 – 270.
[7] C. DEMIR, S. H. GULTEKIN, B. YENER: *Spectral Analysis of Cell - Graphs of Cancer*, Technical report, Rensselaer Polytechnic Institute, Department of Computer Science, TR–04–17.
[8] C. G. DEMIR, M. KANDEMIR, A. B. TOSUN, C. SOKMENSUER: *Automatic segmentation of colon glands using object-graphs*, Medical Image Analysis, **14**(2010), 1 – 12.
[9] D. ALTUNBAY, C. CIGIR, C. SOKMENSUER, C. G. DEMIR: *Color Graphs for Automated Cancer Diagnosis and Grading*, IEEE transactions on Biomedical Engineering, **57**(3) (2010), 665–674.
[10] E. OZDEMIR, C. G. DEMIR: *A Hybrid Classification Model for Digital Pathology Using Structural and Statistical Pattern Recognition*, IEEE Transaction on Medical Imaging, **32**(2) (2013), 474–483.
[11] F. HARARY: *Graph Theory*, Rajkamal Electric Press, Delhi, 2001.
[12] G. A. PAVLOPOULOS, M. SECRIER, C. N. MOSCHOPOULOS, T. G. SOLDATOS, S. KOSSIDA, J. AERTS, R. SCHNEIDER, P. G. BAGOS: *Using Graph Theory to analyze biological networks*, BioData Mining, **4**(10) (2011).
[13] T. M. APOSTOL: *Introduction to Analytic Number Theory*, Springer, 1976.

DEPARTMENT OF MATHEMATICS,
ST. JOHN'S COLLEGE,
PALAYAMKOTTAI-627 002, TIRUNELVELI
E-mail address: rathinag95@gmail.com

DEPARTMENT OF MATHEMATICS,
ST. JOHN'S COLLEGE,
PALAYAMKOTTAI-627 002, TIRUNELVELI
E-mail address: gjkumar.maths@gmail.com