

International Journal of Innovative Research in Science, Engineering and Technology

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Visit: www.ijirset.com

Vol. 8, Issue 3, March 2019

Graph Representation for Accurate Mapping and Reconstruction: Node Listed Adjacency Matrix

M.Muthu Kumaran¹, M.Roopesh², S.S.Yogeshwaraa³, Elangovan⁴

Student, Department of Computer Science and Engineering, Velammal Institute of Technology, Panchetti,
Tamil Nadu, India¹

Student, Department of Computer Science and Engineering, Velammal Institute of Technology, Panchetti,
Tamil Nadu, India²

Student, Department of Computer Science and Engineering, Velammal Institute of Technology, Panchetti,
Tamil Nadu, India³

Assistant Professor, Department of Computer Science and Engineering, Velammal Institute of Technology, Panchetti,
Tamil Nadu, India⁴

ABSTRACT: The goal of this research work is to conceive and develop a graph representation model for accurate representation and reconstruction of graphs. An accurate framework to represent graphs, from which the original graph can be reconstructed without any information loss, is essential for a lot of graph based analysis problems including sub-graph mining. In this research undirected unweighted graphs with known number of node types are taken into consideration. Multiple nodes of the same node type can occur in the graphs. This type of graph finds a great deal of application is analyzing various models in fields such as organic chemistry, social media analytics and financial network patterns.

I. INTRODUCTION

Graph theory is a very old branch of mathematics that finds immense applications in a lot of real life problems. Sub-Graph analysis and graph reconstruction from graph representation are widely used in a lot of areas including molecular analysis, demographic studies and disease analysis. For some problems of graph analysis, the existing methods of graph representation are not suitable for precisely counting sub-graph instances. Here some of the previous works done related to this area are discussed.

Fengcai Qiao et al proposed SSIGRAM (Spark based Single Graph Mining), a Spark based parallel frequent subgraph mining algorithm in a single large graph. Aiming to approach the two computational challenges of FSM, we conduct the subgraph extension and support evaluation parallel across all the distributed cluster worker nodes. In addition, we also employ a heuristic search strategy and three novel optimizations: load balancing, pre-search pruning and top-down pruning in the support evaluation process, which significantly improve the performance.

ShawanaJamil et al. focused on investigation of mining frequent sub-graph patterns in DBLP uncertain graph data using an approximation based method. The frequent sub-graph pattern mining problem is formalized by using the expected support measure. Here approximate mining algorithm based Weighted MUSE, is proposed to discover possible frequent sub-graph patterns from uncertain graph data.

International Journal of Innovative Research in Science, Engineering and Technology

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Visit: www.ijirset.com

Vol. 8, Issue 3, March 2019

Jose Cadena et al. published a research work detecting hotspots and anomalies are a recurring problem with a wide range of applications, such as social network analysis, epidemiology, finance, and bio-surveillance, among others. Networks are a common abstraction in these applications for representing complex relationships. Typically, these networks are dynamic-, i.e., they evolve over time. A number of methods have been proposed for anomaly detection in such dynamic network data sets, which are primarily based on changes in network properties.

Javier O.Garcia et al. published a research work where the human brain can be represented as a graph in which neural units such as cells or small volumes of tissue are heterogeneously connected to one another through structural or functional links. Brain graphs are parsimonious representations of neural systems that have begun to offer fundamental insights into healthy human cognition, as well as its alteration in disease. A critical open question in network neuroscience lies in how neural units cluster into densely interconnected groups that can provide the coordinated activity that is characteristic of perception, action, and adaptive behaviors.

Georgios B. Giannakis et al worked on identifying graph topologies as well as processes evolving over graphs emerge in various applications involving gene-regulatory, brain, power, and social networks, to name a few. Key graph-aware learning tasks include regression, classification, subspace clustering, anomaly identification, interpolation, extrapolation, and dimensionality reduction. Scalable approaches to deal with such high-dimensional tasks experience a paradigm shift to address the unique modeling and computational challenges associated with data-driven sciences.

II. ANALYSIS OF EXISTING MODES OF GRAPH REPRESENTATION FOR ACCURATE MAPPING AND RECONSTRUCTION

There are three popular methods for graph representation: Adjacency Matrix, Adjacency List and Incidence Matrix. In this paper, a small graph with characteristics as explained in the abstract is taken into account for analysis. The graph is represented in all the above three methods and based on the output, the accurate reconstruction of graph is attempted.

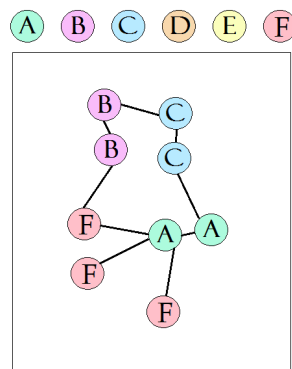


Fig.1: The Graph taken for analysis

Adjacency Matrix:

Adjacency Matrixes represent a given graph in terms of graph with node types arrayed on the row and the column. The matrix created will be filled with 1s and 0s based on whether an edge exists between the two nodes or not.

Presented below is the adjacency matrix for the graph taken for analysis.

International Journal of Innovative Research in Science, Engineering and Technology

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Visit: www.ijirset.com

Vol. 8, Issue 3, March 2019

Adjacency Matrix				
	A	B	C	F
A	1	0	1	1
B	0	1	1	1
C	1	1	1	0
F	1	1	0	0

As it can be observed from above, it is not possible to reconstruct the original graph from the Adjacency Matrix. The most obvious reason being the multiple nodes of the same node type not being represented.

Adjacency List:

Adjacency List has a total number of linked lists as the number of nodes in a given graph. Every linked list that belongs to a given node lists the other nodes that are connected to it by edges.

Adjacency List:			
A	A	C	F
B	B	C	
C	B	C	
F	A	B	

Again in this representation, multiple nodes of the same note type are not represented.

Incidence Matrix:

Incidence Matrix arrays the vertices on the row and edges on the column. Here every node type is listed on the row side. In this representation method, all the edges are mapped but all the nodes are not mapped.

International Journal of Innovative Research in Science, Engineering and Technology

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Visit: www.ijirset.com

Vol. 8, Issue 3, March 2019

Incidence Matrix:									
	E1	E2	E3	E4	E5	E6	E7	E8	E9
A					1	1	1	1	1
B	1	1		1					
C	1		1		1				
F				1			1	1	1

As it can be noted above, reconstruction of the original graph is not possible since multiple nodes of the same node type are not represented in the matrix.

III. PROBLEM STATEMENT

The problem identified here is to convert the given small graph with known number of node types and repeating nodes into a representation matrix wherein, the graph can be reconstructed from. With the data from the matrix, one should also be able to construct the original graph. The conventional adjacency matrix will not have sufficient data to accurately count the occurrences of sub-graphs or reconstruct the graph. This is a pragmatic problem for some graph based analysis where in the adjacency matrix should have sufficient data for accurately counting sub-graphs and also reconstructing the original graph.

IV. PROPOSED GRAPH REPRESENTATION METHOD – NODE LISTED ADJACENCY MATRIX

The solution to solve the above problem is to represent the graph as a kind of adjacency matrix where all the repeating nodes are listed. In the case of incidence matrix, the edges are listed and represented exhaustively. In the matrix proposed herewith, the repeating nodes are numbered and listed exhaustively. This adjacency matrix is called 'Node Listed Adjacency Matrix'. Node Listed Adjacency Matrix is abbreviated as NLAM. NLAM will be very useful for analyzing frequent sub-graph problems with multiple occurrences of the same node type. One such example is analyzing *molecular structure in organic chemistry*.

The NLAM for the above mentioned graph is presented below. It can be observed that a computer program can be developed for converting NLAM into a pictorial graph representation without losing any information of the original graph. The same cannot be done by the conventional Adjacency Matrix, Adjacency List and Incidence Matrix.

Node Listed Adjacency Matrix									
	A1	A2	B1	B2	C1	C2	F1	F2	F3
A1	0	1	0	0	0	0	1	1	1
A2	1	0	0	0	0	1	0	0	0
B1	0	0	0	1	1	0	0	0	0
B2	0	0	1	0	0	0	1	0	0
C1	0	0	1	0	0	1	0	0	0
C2	0	1	0	0	1	0	0	0	0
F1	1	0	0	1	0	0	0	0	0
F2	1	0	0	0	0	0	0	0	0
F3	1	0	0	0	0	0	0	0	0

The NLAM is converted into Sub-Graph Count Matrix (SGCM). SGCM gives the count of instances of single-edged sub-graphs occurring for the given graph. The above NLAM is converted into SGCM and presented below:

International Journal of Innovative Research in Science, Engineering and Technology

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Visit: www.ijirset.com

Vol. 8, Issue 3, March 2019

Sub-Graph Count Matrix				
	A	B	C	F
A	1	0	1	3
B	0	1	1	1
C	1	1	1	0
F	3	1	0	0

In the above presented SGCM, the number of occurrences of each and every single-edged sub-graph is clearly presented. This approach allows the graphs to be analyzed at a very characteristic level which is required for a lot of emerging application areas of graph theory.

V. RESULT

The result is a graph representation for graph analysis and for accurate mapping and graph reconstruction.

VI. CONCLUSION AND FUTURE WORK

Thus the goal to conceive and develop a graph representation model for accurate representation and reconstruction of graphs is achieved. The future enhancement will be the predictive analysis of the graph edges using machine learning (Linear regression). The data is obtained from the graph representation and plotted on the coordinate axis to find the best-fit line (Regression line) which does the predictive analysis.

REFERENCES

- [1] ShawanaJamil, Azam Khan1, ZahidHalim and A. RaufBaig, Weighted MUSE for Frequent Sub-graph Pattern Finding in Uncertain DBLP Data.
- [2] Jose Cadena ,Feng Chen, and Anil Vullikanti, Graph Anomaly Detection Based on Steiner Connectivity and Density.
- [3] Fengcai Qiao 1,*,† ID , Xin Zhang 1,†, Pei Li 1,†, Zhaoyun Ding 1,†, ShanshanJia 2,† and Hui Wang 1,†, A Parallel approach for Frequent Subgraph Mining in a Single Large Graph Using Spark.
- [4] Javier O. Garcia , Arian As hour van, Sarah F. Mu l doon, Je a n M. Vettel , and Daniel le S. Bassett, Applications of Community Detection Techniques to Brain Graphs: Algorithmic Considerations and Implications for Neural Function.
- [5] Georgios B. Giannakis ,Fellow IEEE, Yanning Shen, Student Member IEEE, and Georgios Vasileios Karanikolas, Student Member IEEE, Topology Identification and Learning Over Graphs: Accounting for Nonlinearities and Dynamics
- [6] B. Zhou and F. Chen, Graph-structured sparse optimization for connected subgraph detection, in *Proc. IEEE Int. Conf. Data Mining*.
- [7] Zou, Z.; Li, J.; Gao, H.; Zhang, S. Frequent subgraph pattern mining on uncertain graph data.
- [8] Zou, Z.; Gao, H.; Li, J. Discovering frequent subgraphs over uncertain graph databases under probabilistic semantics.
- [9] David Jensen and Jennifer Neville Data Mining in Social Networks.
- [10] C. C. Aggarwal, Y. Li, J. Wang, and J. Wang, Frequent pattern mining with uncertain data.
- [11] X. Yan and J. Han. gSpan: Graph-based substructure pattern mining.
- [12] W. Eberle and L. Holder, Graph-based approaches to insider threat detection.