



International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 2, Issue 11, November 2014

Review on Stable Relational Pattern Mining in Dynamic Networks

SushamaPatil

P.G.Scholar, Dept. of Computer Engineering, GHRCEM, Wagholi, Pune, India

ABSTRACT: Dynamic networks are used as great means to represent the changing behaviour of the data coming from several complex systems. The relationship occurring between complex entities within these networks over specified time is analysed and stable relational patterns among those entities are discovered. This will help to determine the transition of entity from one state to next and also the external factors responsible for changing the relational patterns. This paper represents overview of all data mining technique which can be used to analyse relations or states between the entities which remain persistent with time in the dynamic network.

KEYWORDS: Dynamic network, stable relational patterns, persistent, external factors, cliques, clustering etc.

I. INTRODUCTION

Dynamic networks are generally modelled through a sequence of time-based graphs. Consecutive snapshots of the network at given intervals are also used. Entities, or nodes, in the network are connected via edges, or links which are also called as relations. Most real-life networks are dynamic in nature because they constantly add and remove nodes and links over time. Because of its flexibility and availability of theoretical and applied tools for efficient analysis dynamic networks have been used as generic model to represent the relations between various entities in diverse applications and to capture the temporal changes and dynamic aspects of the underlying data. Examples of some widely used networks include the popular social networking sites like Facebook, the email network and biological networks [1]. Graphs are considered as the universal data structure to model entities and their relationships because all common data structures, vectors, strings and time series, can be represented as graphs. So, the amount of graph-structured data is ever increasing in a wide range of application domains such as bioinformatics, medicine, large database management and web log analysis.

There is evolution in the amount and variety of data due to storage and exchange of lot of information in various applications. The majority of recent subgraph mining approaches have dedicated on characterizing the topology of static networks. But to model real-world systems, often a temporal component has to be taken into account, as relations between objects here usually occur for a certain period of time only. Therefore, a practical model has to consider that edges will be inserted and/or deleted over time. A new data structure is resulted which is called as a dynamic graph. Dynamic networks are flexible to capture the temporal changes and represent relations between entities in these applications. Numbers of tools are available in dynamic network for efficient analysis. This analysis of networks can provide significant insight about the stable relational patterns and their evolution over time.

II. BACKGROUND

As we see everything around us is connected in some way, forming links, connections, or edges in some theoretical graph. Over the past years many algorithms have been discovered that helps to extract facts out of graphs. Graph pattern mining finds remarkable paths, trees, subgraphs, frequent induced subgraphs [5], and cliques within graph datasets. The information gained from graphs has been useful to solve many challenges in diverse domains such as computational biology, chemical informatics, and communications in networks. Frequent induced subgraph mining is used in the bio-informatics application domain to detect sets of functionally equivalent genetic factor [3]. Tree pattern mining can find out commonly occurring subtrees in a chemical compound dataset. Web access link-path patterns enable internet service providers to predict frequently accessed resources that should be cached. Besides pattern



International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 2, Issue 11, November 2014

mining, clustering is another usual data mining technique that has been applied to graphs. In the social graph, clustering has been used to identify highly consistent distinct groups, e.g. communities. In the chemical and bio-informatics domains, clustering has proven quite useful in the drug discovery process. Clustering is often applied for grouping datasets where the items are graphs, rather than nodes in a large graph.

Most graph mining algorithms are based on the familiar association rule mining, or the frequent itemset problem. Frequent itemset mining algorithms are generally based on the lattice or downward closure property of support. This property states that an itemset cannot be frequent if even one of its subsets is not frequent. Frequent itemset mining algorithms specify all frequent itemsets by effectively pruning the search area. In terms of graph mining, downward closure translates to the fact that a subgraph is frequent only if all of its subgraphs are frequent. Most existing graph mining algorithms generalize up to date frequent itemset mining algorithms to structured data.

III. LITERATURE SURVEY

[1] Kuramochi, M., Karypis, G.

“Over the years, frequent item set discovery algorithms have been used to find interesting patterns in various application areas. However, as data mining techniques are being increasingly applied to non-traditional domains, existing frequent pattern discovery approaches cannot be used. This is because the transaction framework that is assumed by these algorithms cannot be used to effectively model the data sets in these domains. An alternate way of modeling the objects in these data sets is to represent those using graphs. Within that model, one way of formulating the frequent pattern discovery problem is that of discovering sub graphs that occur frequently over the entire set of graphs. They present a computationally efficient algorithm, called FSG, for finding all frequent sub graphs in large graph data sets. They experimentally evaluate the performance of FSG using a variety of real and synthetic data sets. Our results show that despite the underlying complexity associated with frequent sub graph discovery, FSG is effective in finding all frequently occurring sub graphs in data sets containing more than 200,000 graph transactions and scales linearly with respect to the size of the data set.”

[2] A. Chapanond, M. S. Krishnamoorthy, and B. Yener

“Analysis of social networks to identify communities and model their evolution has been an active area of recent research. This paper analyzes the Enron email data set to discover structures within the organization. The analysis is based on constructing an email graph and studying its properties with both graph theoretical and spectral analysis techniques. The graph theoretical analysis includes the computation of several graph metrics such as degree distribution, average distance ratio, clustering coefficient and compactness over the email graph. The spectral analysis shows that the email adjacency matrix has a rank-2 approximation. It is shown that pre-processing of data has significant impact on the results, thus a standard form is needed for establishing a benchmark data.”

[3] L. Cerf, T. Nguyen, and J. Boulicaut

“Several algorithms, namely Cumbeminern, Trias, and Data-Peeler, have been recently proposed to mine closed patterns in ternary relations. They consider here the specific context where a ternary relation denotes the value of a graph adjacency matrix at different timestamps. Then, we discuss the constraint-based extraction of patterns in such dynamic graphs. We formalize the concept of δ -contiguous closed 3-clique and we discuss the availability of a complete algorithm for mining them. It is based on a specialization of the enumeration strategy implemented in Data-Peeler. Indeed, clique relevancy can be specified by means of a conjunction of constraints which can be efficiently exploited. The added-value of our strategy is assessed on a real dataset about a public bicycle renting system. The raw data encode the relationships between the renting stations during one year. The extracted δ -contiguous closed 3-cliques are shown to be consistent with our domain knowledge on the considered city.”

[4] J. Pei, J. Han, B. Mortazavi-Asl, H. Pinto, Q. Chen, U. Dayal, and M. Hsu

“Sequential pattern mining is an important data mining problem with broad applications. It is challenging since one may need to examine a combinatorially explosive number of possible subsequence patterns. Most of the previously developed sequential pattern mining methods follow the methodology of Apriori which may substantially reduce the number of combinations to be examined. However, Apriori still encounters problems when a sequence database is large and/or when sequential patterns to be mined are numerous and/or long. In this paper, they propose a novel sequential



International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 2, Issue 11, November 2014

pattern mining method, called Prefix Span (i.e., Prefix-projected Sequential pattern mining), which explores prefix projection in sequential pattern mining. Prefix Span mines the complete set of patterns but greatly reduces the efforts of candidate subsequence generation. Moreover, prefix-projection substantially reduces the size of projected databases and leads to efficient processing. Their performance study shows that Prefix Span outperforms both the Apriori-based GSP algorithm and another recently proposed method, Free Span, in mining large sequence databases.”

[5] B. Wackersreuther, P. Wackersreuther, A. Oswald, C. B. ohm, and K. Borgwardt

“In many application domains, graphs are utilized to model entities and their relationships, and graph mining is important to detect patterns within these relationships. While the majority of recent data mining techniques deal with static graphs that do not change over time, recent years have witnessed the advent of an increasing number of time series of graphs. In this paper, they define a novel framework to perform frequent sub-graph discovery in dynamic networks. In particular, we are considering dynamic graphs with edge insertions and edge deletions over time. Existing sub-graph mining algorithms can be easily integrated into our framework to make them handle dynamic graphs. Finally, an extensive experimental evaluation on a large real-world case study confirms the practical feasibility of our approach.”

IV. ALGORITHMS

Dynamic graphs are used to represent relationships between entities that evolve over time. Meaningful patterns in such structured data must capture strong interactions as well as their evolution over time. In social networks, such patterns can be seen as dynamic community structures, i.e., sets of individuals who strongly and repeatedly interact with each other.

A. Constraint-based mining approach:

In this approach the temporal evolution of patterns is captured by associating a temporal event type to each identified subgraph to uncover evolving patterns [2]. This uses a new algorithm that finds such subgraphs in a time series of graphs incrementally. This considers the evolving-pattern mining problem in dynamic graph and presents five new pattern types which rely on two things: first, the extraction of dense subgraphs and second, the identification of their evolution. Five basic temporal events used are: The formation, termination, growth, reduction and stability of subgraphs from one time stamp to the next.

This task is divided into a local-to-global framework: Local patterns are first mined in a static graph; then they are combined with the one extracted in the previous graph to form evolving patterns. These patterns are defined by means of constraints that are used by the algorithm EVOLVING-SUBGRAPHS. In this way it efficiently mines all evolving patterns that satisfy the constraints. Constraint-based mining algorithms have advantage of the constraints to prune huge parts of the search space which cannot contain valid patterns.

B. DFS Algorithm:

In many applications, graph mining is important to detect patterns within the relationships. While the majority of recent data mining techniques deal with static graphs that do not change over time. Now days there is increasing number of time series of graphs. A novel framework is presented to perform frequent subgraph discovery in dynamic networks[4].

Dynamic frequent subgraphdiscovery can be performed in two steps. First algorithms for finding frequent subgraphs in the union graph of a time series of graphs is used. Second the resulting static frequent subgraphs for frequent dynamic patterns are searched. DFS algorithm is used which provides a canonical representation based on sparse adjacency list data structure and implements a breadth-first enumeration algorithm for discovering frequent subgraphs in dynamic network. It mainly considers dynamic graphs with edge insertions and edge deletions over time.

C. Graph-based relational mining approach:

A new dynamic graph-based relational mining approach has been also developed to learn structural patterns in biological networks as they change over time [6]. The analysis of dynamic networks is important not only to



International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 2, Issue 11, November 2014

understand life at the system-level, but also to discover original patterns in other structural data. Most current graph-based data mining approaches focus on only static graphs. This approach analyses a sequence of graphs and discovers rules that capture the changes that occur between pairs of graphs in the sequence and helps to understand how the biosystems change over time. Two step algorithm is used: Learning Graph Rewriting Rules and Learning Transformation Rules. First algorithm learns graph rewriting rules in a dynamic graph and represent how two sequential graphs are different. Second algorithm learns the repeated transformation rules in the learned graph rewriting rules and describes how the graph changes over time, where the changes are actually represented as a sequence of revised graphs. For both algorithms a previously developed method for finding the best compressing subgraph in a set of graphs is used. For the first algorithm, repeated application of this method allows to find the set of all subgraphs common to a pair of consecutive graphs. For the second algorithm this method allows to find the subgraphs repeatedly added and removed in the dynamic graph. A frequent subgraph miner is used for this purpose. As the quality and quantity of network and interaction data increases rapidly, the problem of effectively analyzing this data becomes significant. We can understand biological processes at a modular level by providing a framework for understanding cellular organization, functional hierarchy, and evolutionary conservation and using molecular interaction data.

D. MULE Algorithm:

Also an innovative new algorithm, **Mining Unique-Labeled Edgesets (MULE)** is used for detecting frequently occurring patterns and modules in biological networks [10]. The algorithm considers the problems computationally tractable and scalable to large numbers of networks by using an innovative graph simplification technique which is based on ortholog contraction, ideally suited to biological networks. MULE is based on frequent itemset extraction to discover frequent subgraphs among the graphs. It takes into account the nature of molecular interaction data. Existing formulations of isomorphism based frequent subgraph extraction suffer from exponential explosion of problem size due to NP-hardness of both mining and subgraph isomorphism problems. In contrast to such extant approaches, MULE avoids the repeated solution of NP-hard subgraph isomorphism problem and also preserves the biological relevance of identified patterns. This algorithm makes use of a recursive subroutine to extend frequent edgesets.

E. Finding periodic or near periodic subgraph:

It is infrequent and hard to detect interaction patterns within a social network because social interactions occur regularly. To identify such regular behavior, a new mining problem of finding periodic or near periodic subgraphs in dynamic social networks is used [8]. This analyzes the computational complexity of the problem, showing that, unlike any of the related subgraph mining problems, it is polynomial. A practical, efficient and scalable algorithm is used to find such subgraph that takes imperfect periodicity into account. An efficient, single-pass algorithm for mining all closed periodic subgraphs in a dynamic network is given. This algorithm is polynomial time and space algorithm.

The basis of the algorithm is a *pattern tree*. This pattern tree maintains information about all patterns that are either currently periodic or could become periodic at a future timestamp. As each new timestamp is read, the pattern tree is traversed and updated with the information. Any patterns that are no longer periodic are flushed, and new periodic patterns are possibly created. The algorithm maintains two data structures: the pattern tree and a *subgraph hash map*.

There are some disadvantages of existing techniques such as follows:

1. The existing techniques can only detect the frequent patterns in a dynamic network or follow related patterns over time.
2. They are not designed to identify stable relational patterns.
3. They do not focus on tracking the changes of preserved relational patterns over time.

Applications of stable pattern mining are as follows:

1. It is useful in mining biochemical structures.
2. Program control flow analysis can be determined using stable pattern mining.
3. Mining XML structures or Web communities is also possible.



International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 2, Issue 11, November 2014

4. Social networking sites, email networks can be analyzed.

Table 1: Existing Graph mining algorithms

Algorithm	Function
Subdue	It is based on recursively finding a subgraph that provides the best compression based on the Minimum Description Length (MDL) principle. At each step of the algorithm, the subgraph that provides maximal compression, hence is most frequent, is discovered via a beam search heuristic and replaced by a single node.
AGM	It adapts the well-known a-priori algorithm to mining vertex sets that induce frequent subgraphs in a graph database. The main feature of this algorithm is that it provides a canonical labeling for graphs based on an adjacency-matrix representation. This might be computationally infeasible for applications involving large graphs
DFS	It provides a canonical representation based on sparse adjacency list data structure and adopts a breadth-first enumeration algorithm for discovering frequent subgraphs
gSpan	It reduces the overhead introduced by the problems through a DFS-based canonical representation of graphs and enumerates the search space in a depth-first manner to achieve significant speed-up over earlier algorithms such as FSG.
CloseGraph	It is an extension of gSpan designed to discover only those subgraphs that do not have a supergraph of same support to avoid redundancy in the output.
FFSM	It improves upon gSpan by reducing redundant candidate generation through a vertical search scheme based on an algebraic graph framework.
SPIN	It speeds up graph mining by splitting the process into two independent tasks of mining subtrees and extending these subtrees to frequent subgraphs.
Gaston	It relies on the same idea to generate frequent substructures hierarchically by starting from paths, extending frequent paths to trees, and further extending frequent trees to graphs.
MULE	It is used for detecting frequently occurring patterns and modules in biological networks. MULE is based on frequent itemset extraction to discover frequent subgraphs among the graphs. It preserves the biological relevance of identified patterns.

V. CONCLUSION AND FUTURE WORK

This paper gives an overview of the methods used in the mining of stable relational patterns in dynamic networks. It also compares methods based on static graph mining with those using the time component in the network. In many cases, both types of algorithms resulted in similar time complexity. Many newer algorithms appear to build upon older algorithms in dynamic network mining. Two things that helped in producing effective algorithms for certain mining tasks were the use of certain data structures to speed up operations and extreme pruning of the search space. We can also determine conserved relational states in dynamic network to understand change of entity from one state to another state.

REFERENCES

1. D. Boyd and N. Ellison, "Social network sites: Definition, history, and scholarship," *Journal of Computer-Mediated Communication*, vol. 13, no. 11, October 2007.
2. C. Robardet, Constraint-based pattern mining in dynamic graphs, in *IEEE ICDM*, pp.950955, 2009.
3. "Finding frequent patterns in a large sparse graph," *Data Mining and Knowledge Discovery*, vol. 11, no. 3, pp. 243–271, 2005.
4. M. Kuramochi and G. Karypis, An efficient algorithm for discovering frequent subgraphs, *IEEE TKDE*, vol. 16, no. 9, pp. 10381051, 2004.



International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 2, Issue 11, November 2014

5. K M.Borgwardt, H.-P. Kriegel, and P. Wackersreuther, "Pattern mining in frequent dynamic subgraphs," IEEE ICDM. Washington, DC, pp. 818–822, 2006.
6. B. Wackersreuther, P. Wackersreuther, A. Oswald, C. B. Ohm, and K. Borgwardt, Frequent subgraph discovery in dynamic networks, Proc. of the 8th Workshop on Mining and Learning with Graphs, pp.155162,2010.
7. M. Lahiri and T. Berger-Wolf, Mining periodic behavior in dynamic social networks, IEEE ICDM, pp.373382, 2008.
8. M. Koyuturk, Y. Kim, S. Subramaniam, W. Szpankowski, and A. Grama, Detecting conserved interaction patterns in biological networks, Journal of Computational Biology, vol. 13, no. 7, pp. 12991322, 2006.
9. M. Deshpande, M. Kuramochi, N. Wale, and G. Karypis, "Frequent substructure-based approaches for classifying chemical compounds," IEEE TKDE, vol. 17, no. 8, pp. 1036–1050, 2005.
10. A. Chapanond, M. S. Krishnamoorthy, and B. Yener, "Graph theoretic and spectral analysis of enron email data," Comput. Math. Organ. Theory, vol. 11, no. 3, pp. 265–281, 2005.
11. L. Cerf, T. Nguyen, and J. Boulicaut, "Discovering relevant cross-graph cliques in dynamic networks," Foundations of Intelligent Systems, pp. 513–522, 2009.
12. J. Pei, J. Han, B. Mortazavi-Asl, H. Pinto, Q. Chen, U. Dayal, and M. Hsu, "Prefixspan: Mining sequential patterns by prefixprojected growth," ICDE, pp. 215–224. [Online], 2001.
13. B. Wackersreuther, P. Wackersreuther, A. Oswald, C. B. Ohm, and K. Borgwardt, "Frequent subgraph discovery in dynamic networks," Proc. of the 8th Workshop on Mining and Learning with Graphs, pp. 155–162, 2010.

BIOGRAPHY

Sushama R. Patilis is a P. G. Scholar in the Computer Engineering Department, G. H. Raisoni College of Engineering and Management, Wagholi, Pune. She has received Bachelor of Engineering degree (BE) in Computer Science and Engineering in 2012 from Dr. J.J. Magdum College of Engineering, Jaysingpur, India. Her research interests are Web mining, Wireless Networks, Database management etc.