Complex Patterns in Dynamic Attributed Graphs

Rina Singh, Jeffrey A. Graves, Douglas A. Talbert
Department of Computer Science
Tennessee Technological University
Cookeville, TN USA
{rsingh43, jagraves21}@students.tntech.edu, DTalbert@tntech.edu

ABSTRACT

In recent years, there has been huge growth in the amount of graph data generated from various sources. These types of data are often represented by vertices and edges in a graph with real-valued attributes, topological properties, and temporal information associated with the vertices. Until recently, most pattern mining techniques focus solely on vertex attributes, topological properties, or a combination of these in a static sense; mining attribute and topological changes simultaneously over time has largely been overlooked. In this work-in-progress paper, we propose to extend an existing state-of-the-art technique to mine for patterns in dynamic attributed graphs which appear to trigger changes in attribute values.

1. INTRODUCTION

Recently there has been a huge growth in the amount of complex data generated from various sources such as citation databases, social networks, bioinformatics, and sensor networks. More traditional data mining approaches can fail to uncover interesting or useful patterns due to underrepresentation of structural relationships in the data. Data containing implicit or explicit relationships can often be modeled as attributed graphs. Entities within the data are represented as vertices with associated attribute-value pairs, while relationships between entities are represented by edges. Dynamic attributed graphs can be used when the data or relationships between data entities change over time. These graph representations are often better suited to capturing structural relationships between data entities and are becoming more ubiquitous for analyzing complex structural data.

In this work-in-progress paper, we present an approach for uncovering patterns that appear to trigger changes in data entities. Such patterns can be useful in explaining the causes of change or in the prediction of future change. For example, one may wish to explain the downfall of a once popular social media figure or predict the next rising star. Until recently, most dynamic attributed graph mining techniques have focused on attribute values or structural relationships separately, while approaches combing these two have been limited to static graphs. Taking inspiration from Kaytoue et al. [3], we have attempted to leverage both attribute values and structural information to better explain changes in dynamic data.

WWW'16 Companion, April 11–15, 2016, Montréal, Québec, Canada.
ACM 978-1-4503-4144-8/16/04.
http://dx.doi.org/10.1145/2872518.2889374.

Table 1: Approaches to Attributed Graph Mining

Features \ Techniques	TopGraphMiner [4]	MINTAG [2]	TRIGAT [3]
Dynamic		✓	✓
Attributed	✓	✓	✓
Local Structure	✓		✓
Global Structure	✓	✓	

1.1 Related Work

There are several state-of-the-art approaches currently being researched for attributed and dynamic graph mining. TopGraphMiner makes use of vertex attributes and topological properties when mining graph topology from static graphs [4]. MINTAG is capable of uncovering trends in dynamic graphs in the form of subgraphs [2]. While MINTAG makes use of both attributes and topological properties, it only focuses on global topological properties (e.g., graph diameter) and ignores local topological properties (e.g., vertex centrality measures). The TRIGAT mining algorithm makes use of both vertex attributes and local topological measures while searching dynamic graphs for patterns which indicate changes in graph structure [3]. A summary of current approaches in dynamic attributed graph mining can be found in Table 1.

Inspired by the work of Kaytoue et al., we are searching for patterns which indicate change in vertex attribute values. Kaytoue has formulated the problem of finding patterns that indicate structural changes as a sequential pattern mining problem. Sequences are made up of sets of attribute and topological changes from one timestamp to another in the dynamic graph. Kaytoue then searches for specific types of sequences that end in a single topological change (i.e., sequences that end in a singleton set which contains a topological change). In order to find patterns that indicate changes in attribute values, we focus on mining sequences that end in a single attribute change.

1.2 Contributions

In their work, Kaytoue et al. attempt to uncover patterns involving both attributes and topological properties which appear to cause changes in topological properties by reducing this task to that of sequential pattern mining. We extend the work of Kaytoue in two directions. First, we are attempting to uncover patterns which indicate changes in at-

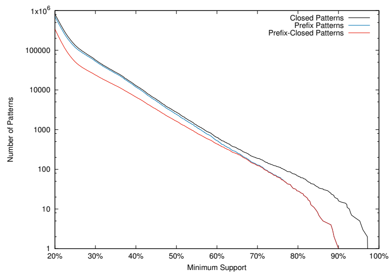

Figure 1: Min. Support vs. Patterns

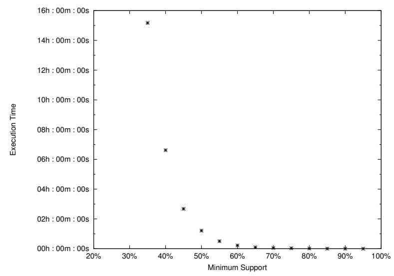

Figure 2: Run Time

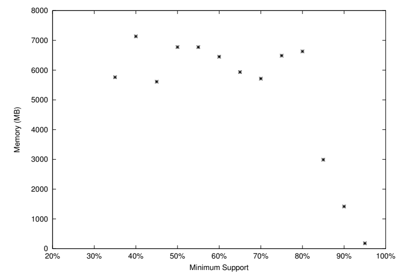

Figure 3: Memory Usage

tribute values (not topological properties). Second, we are considering the use of alternative closed sequential pattern mining approaches for solving this problem.

2. METHODOLOGY

Given a dynamic attributed graph (i.e., a sequence of attributed graphs), a finite sequence of sets can be associated with each vertex. The ith set in the sequence is a collection of symbols which indicates changes in either attributes or topological properties between the $(i-1)$th and the ith timestamp in the dynamic graph. Conceptually, this sequence describes how the attributes and topological properties of a vertex change over time. We have included the following vertex centrality measures as our topological properties of interest: degree, closeness, betweenness, eigenvector, clustering coefficient, PageRank, hub score, and authority score. A detailed description of the construction of the sequences associated with the dynamic vertices can be found in Kaytoue's paper [3].

After constructing the sequences associated with the dynamic vertices, these sequences are then mined to find frequent prefix-closed subsequences that end with a set containing a single attribute change. These prefix-closed subsequences represent a sequence of changes that appear to trigger the associated attribute change. Kaytoue has proven that every prefix-closed subsequences will appear as a subsequence of a closed subsequences. And so, closed sequential pattern mining is used to obtain a collection of closed subsequences, from which the prefix-closed subsequences can be extracted. Kaytoue originally applied the CloSpan algorithm for mining closed sequential patterns. Rather than relying on CloSpan, we are using the BIDE algorithm, which has been shown to outperform CloSpan both in terms of memory usage and runtime [5].

3. EXPERIMENTAL ANALYSIS

We have evaluated the above approach on a real-world DBLP co-authorship network made available by Desmier et al. [1] containing 2,723 authors with publications in 43 journals/conferences. Each author is represented as a vertex in the dynamic graph and the number of publications in a given five year span make up the associated attributes. The dynamic graph consists of 9 timestamps representing a contiguous period of five years from 1990-2010 (consecutive graphs contain a 2 year overlap). Only authors having at least 10 publications between 1990 and 2010 are included.

The experimental results show that as the minimum support increases, the number of closed sequences drop drastically. A large number of closed sequences contain subsequences ending in an attribute change. This is because there are significantly more attribute values than topological measures. Figure 1 shows the number of closed/prefix/prefix-closed patterns meeting a given minimum support. Figure 2 shows the run time of the BIDE algorithm as a function of support. As expected in sequential pattern mining, the run time appears to follow the power law with respect to support. Interestingly, the memory requirements do not appear to be as strongly influenced by the support, as seen in Figure 3.

4. WORK IN PROGRESS

The computational complexity of sequential pattern mining is the major drawback to this approach to finding patterns which indicate changes in attribute values. The time and memory required in sequential pattern mining relies heavily on the length of the sequences (i.e., number of timestamps in the dynamic graph) and the number of symbols found in the sequences (i.e., the granularity in measuring attribute and topological changes). We are currently exploring parallel sequential pattern mining algorithms to help reducing run time. In addition, we are considering introducing gap constraints on the subsequences to further prune the search space and reduce run time. In another vein, we are also considering reformulations of the task into the sequential rule mine problem and sequence prediction problem.

5. REFERENCES

[1] E. Desmier, M. Plantevit, C. Robardet, and J.-F. Boulicaut. Cohesive co-evolution patterns in dynamic attributed graphs. In *Discovery Science*, pages 110–124. Springer, 2012.

[2] E. Desmier, M. Plantevit, C. Robardet, and J.-F. Boulicaut. Trend mining in dynamic attributed graphs. In *Machine Learning and Knowledge Discovery in Databases*, pages 654–669. Springer, 2013.

[3] M. Kaytoue, Y. Pitarch, M. Plantevit, and C. Robardet. Triggering patterns of topology changes in dynamic graphs. In *Advances in Social Networks Analysis and Mining (ASONAM), 2014 IEEE/ACM International Conference on*, pages 158–165. IEEE, 2014.

[4] A. Prado, M. Plantevit, C. Robardet, and J.-F. Boulicaut. Mining graph topological patterns: Finding covariations among vertex descriptors. *Knowledge and Data Engineering, IEEE Transactions on*, 25(9):2090–2104, 2013.

[5] J. Wang and J. Han. Bide: Efficient mining of frequent closed sequences. In *Data Engineering, 2004. Proceedings. 20th International Conference on*, pages 79–90. IEEE, 2004.