Towards Association Rule Mining on Property Graphs

Yuya Sasaki
sasaki@ist.osaka-u.ac.jp
Osaka University

ABSTRACT

Graph association rule mining is a data mining technique for discovering regularities from graph data. In this paper, we propose a new graph association rule mining, which is called vertex-centric graph association rule mining. Our technique aims to discover regularities between graph patterns and properties of vertices. We discuss its application and define our technique formally. Our rule mining can support typical measures of association rule mining such as support and confidence. Finally, we discuss our future work.

KEYWORDS

Association rule mining, isomorphic matching

ACM Reference Format:
Yuya Sasaki. 2018. Towards Association Rule Mining on Property Graphs. In *Woodstock '18: ACM Symposium on Neural Gaze Detection, June 03–05, 2018, Woodstock, NY*. ACM, New York, NY, USA, 2 pages. https://doi.org/10.1145/1122445.1122456

1 INTRODUCTION

Association rule mining is a fundamental data mining technique for discovering regularities between items in large databases. Association rules have a form $X \Rightarrow Y$, where X and Y are disjoint and called antecedent and consequent, respectively. Association rule mining are recently extended to the context of graph data to capture associations rules with graph patterns [5, 10]. In graph association with graph patterns, X and Y are disjoint graph patterns. Graph association rules have many applications as follows.

Social analysis: Social analysis is important to understand and improve our life. For example, social relationships are affected to our health [8] and happiness [7] The social relationships can be modeled by graphs and graph association rule mining can discover regularities between people's' status and their relationship patterns. In more concretely, we can find association rules that people who feel happiness have specific friendships and profiles.

Recommendation: Recent recommendation systems capture user behavior to products, such as viewing and click [3, 9]. User behavior to products can be modeled by bipartite graphs. Graph association rule mining supports to find potential costumers by finding regularities among user behavior, products, and users' profiles.

Woodstock '18, June 03–05, 2018, Woodstock, NY

ACM ISBN 978-1-4503-XXXX-X/18/06…$15.00
https://doi.org/10.1145/1122445.1122456

Bias checking: Bias in datasets is an emerging problem that we need to solve for developing fair machine learning models. For example, automated systems to review applicants' resumes developed by Amazon inchad a significant gender bias towards male candidates over females due to historical discrimination in the training data [2]. Machine learning models for graph data could be suffered from such discriminatory bias in graphs as same as transactional data. In graph data mining, each model is trained from not only their properties of vertices but also relationships of vertices. Since graph association rules can discover regularities that include discriminatory bias, they can support removing discriminatory bias in the graph data.

There are few works related to graph association rules. Since existing techniques have been developed in their own purposes and assumptions, they have different semantics each other. As far as we know, existing works cannot be applied to the above applications due to their semantics and assumptions (see Sec. 2 in detail). In particular, the above applications focus on regularities that vertices have what properties and are involved what subgraphs. Therefore, to apply graph association rule mining to the above examples, new mining techniques are necessary.

In this paper, we We mainly discuss our idea of a new mining technique called *vertex-centric graph association rule mining* which aims to discover regularities among both properties of vertices and subgraphs that the vertices are involved. We formally define the vertex-centric graph association rule mining. Then, we describe that the association rules can be applied to many existing measures that evaluate importance of rules, such as support and confidence.

2 RELATED WORK

We here review frequent subgraph mining and association rule mining techniques.

Frequent subgraph mining: The frequent subgraph mining problem is to find subgraphs in a data graph and enumerate all subgraphs with support above a given minimum threshold, This problem can be divide into two categories; transactional data graph (i.e., graph databases compromising multiple small graphs) and a single large graph. Frequent subgraph mining in transactional data graph searches for subgraphs that are included transactional data graphs more than the given minimum support. While, grequent subgraph mining in a single large graph searches for subgraphs that appear a single graph more often than the given minimum support.

Algorithms for transactional data graphs have been studied well and can be applied extensions of algorithms for transaction data such as Apriori and pattern growth methods, because anti-monotonic properties are naturally hold. In a single large, anti-monotonic properties are not hold if we simply count the number of matched subgraphs. To hold anti-monotonic properties, several support counting metrics have been proposed; minimum-image-based support (MNI), minimum vertex cover (MVC), MCP, and maximum independent edge set support (MIES). These support counting metrics are applied only to the given absolute support, not to relative support because we cannot count the maximum numbers of subgraphs that possibly appear in the single graph. As far as we know, support measure proposed in [5] can be applied to related support for a single large graphs. They proposed vertex-centric support measure in which how many vertices are involved the subgraphs (i.e., the possible maximum absolute support is the number of vertices). We utilize the same vertex-centric support measure in our association rule mining.

Association rule mining: Association rules first have been proposed by Agrawal et al. [1] for transaction data. Association rules for graph analysis have been studied recently.

Association rules for graph patterns, called GPAR, were introduced in [5]. Their graph patterns use vertex-centric subgraphs in which designated vertices have specific graph patterns. This is similar to our association rule mining, but their association rules focus on specific graph patterns; (1) consequent is just a single edge and (2) antecedent and consequent share vertices. That is, their association rules evaluate an edge type specified by consequent is included in subgrahps specified by antecedent. In addition, their algorithm is developed to find diversified association rules with fixed consequent, instead of finding all association rules. Fan et al. [6] extends GPAR to find quantified graph association rules, which handles potential edges in graphs. Additionally, Fan et. al. extends it to deduce associations [4].

Wang et al. [10] extends GPAR to be more generalized than GPAR in [5]. This work takes different semantics from GPAR [5]; (1) support measure is MSI, (2) antecedent and consequent are both patterns that are connected and include at least one edge, and antecedent and consequent share vertices and have no edge in common. In this work, we can specify graph patterns as consequent, and in addition it aims to enumerate association rules. The drawbacks of this rules are that we cannot use relative support value and find regularities between graph patterns and properties of vertices (e.g., occupation and gender), because it uses MSI as support and both antecedent and consequent must have at least one edge (i.e., we cannot specify property of vertices as consequent).

Our graph association rules keep good properties of the two works; vertex-centric support and both antecedent and consequent are freely specified (graph patterns and properties). The difference from these works is that our association rules share just a designated vertex instead of sharing all vertices. Since the semantics of association rules are different, we need new mining methods for efficiently enumerating the rules.

3 PROPOSAL

We define the vertex-centric graph association rule mining.

We consider graphs $G = (V, E, L, A)$ where V is a finite set of vertices, $E \subset V \times L \times V$ is the set of edges with label $l \in L$, and each $v \in V$ has a tuple $A(v) = (A_1 = a_1, \ldots, A_n = a_n)$ of attributes of a finite arity. A graph pattern is $Q = (V_Q, E_Q, A_Q)$, where V_Q is a set of pattern vertices, E_Q is a set of pattern edges, A_Q assigns attributes $A_Q(v)$ to each vertex $v \in V_Q$.

We now define a vertex-centric graph association rule mining. A pattern Q includes following literals.

- attribute of v;
- edge from v_i to v_j; and
- a single v as pivot.

A vertex-centric graph association rule is define as

$$Q_x \rightarrow Q_y,\ where\ pivots\ of\ Q_x\ and\ Q_y\ are\ shared.$$

$Q_x \not\subset Q_y$ and $Q_y \not\subset Q_x$ but Q_x and Q_y could have the same literals. We denote $||Q_x, G||$ as the number of vertices that are involved Q_x.

Intuitively, a vertex-centric graph association rule mining find vertices that are involved both two subgraphs Q_x but Q_y as pivots.

We then define measures of vertex-centric graph association rules; support and confidence.

The absolute support $ASupp$ of $Q_x \rightarrow Q_y$ for graph G is defined as

$$ASupp(Q_x \rightarrow Q_y)_G = ||(Q_x \cup Q_y, G)||.$$

The relative support $RSupp$ of $Q_x \rightarrow Q_y$ for graph G is defined as

$$RSupp(Q_x \rightarrow Q_y)_G = \frac{||(Q_x \cup Q_y, G)||}{|V|}.$$

The confidence $Conf$ of $Q_x \rightarrow Q_y$ for graph G is defined as

$$Conf(Q_x \rightarrow Q_y)_G = \frac{ASupp(Q_x \cup Q_y, G)}{ASupp(Q_x, p_x, G)}.$$

In the same way, we can define other measures such as lift.

4 CONCLUDING REMARKS

We proposed a new graph association rule mining. The mining technique has different semantics from existing works, which aim several interesting applications. As our future works, we would like to develop efficint algorithms to find vertex-centric association rules and conduct experimental studies to find interesting rules on real-world graphs.

REFERENCES

[1] Rakesh Agrawal, Tomasz Imieliński, and Arun Swami. 1993. Mining association rules between sets of items in large databases. In *SIGMOD*. 207–216.
[2] Jeffrey Dastin. 2018. Rpt-insight-amazon scraps secret ai recruiting tool that showed bias against women. *Reuters* (2018).
[3] James Davidson, Benjamin Liebald, Junning Liu, Palash Nandy, Taylor Van Vleet, Ullas Gargi, Sujoy Gupta, Yu He, Mike Lambert, Blake Livingston, et al. 2010. The YouTube video recommendation system. In *Proceedings of the fourth ACM conference on Recommender systems*. 293–296.
[4] Wenfei Fan, Ruochun Jin, Muyang Liu, Ping Lu, Chao Tian, and Jingren Zhou. 2020. Capturing associations in graphs. *Proceedings of the VLDB Endowment* 13, 12 (2020), 1863–1876.
[5] Wenfei Fan, Xin Wang, Yinghui Wu, and Jingbo Xu. 2015. Association rules with graph patterns. *Proceedings of the VLDB Endowment* 8, 12 (2015), 1502–1513.
[6] Wenfei Fan, Yinghui Wu, and Jingbo Xu. 2016. Adding counting quantifiers to graph patterns. In *SIGMOD*. 1215–1230.
[7] Max Haller and Markus Hadler. 2006. How social relations and structures can produce happiness and unhappiness: An international comparative analysis. *Social indicators research* 75, 2 (2006), 169–216.
[8] James S House, Karl R Landis, and Debra Umberson. 1988. Social relationships and health. *Science* 241, 4865 (1988), 540–545.
[9] Weiyang Lin, Sergio A Alvarez, and Carolina Ruiz. 2002. Efficient adaptive-support association rule mining for recommender systems. *Data mining and knowledge discovery* 6, 1 (2002), 83–105.
[10] Xin Wang, Yang Xu, and Huayi Zhan. 2020. Extending association rules with graph patterns. *Expert Systems with Applications* 141 (2020), 112897.