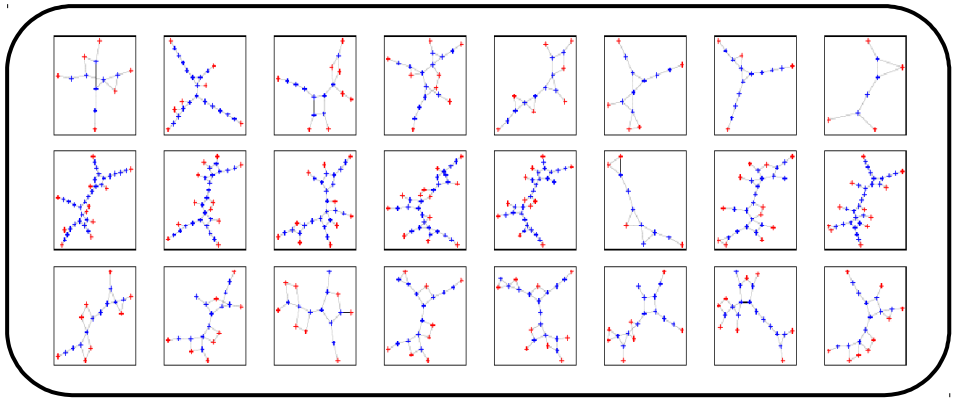
A New Multi-graph Transformation Method for Frequent Approximate Subgraph Mining

N. Acosta-Mendoza, A. Gago-Alonso, J.A. Carrasco-Ochoa, J.Fco. Martínez-Trinidad, J.E. Medina-Pagola

Introduction

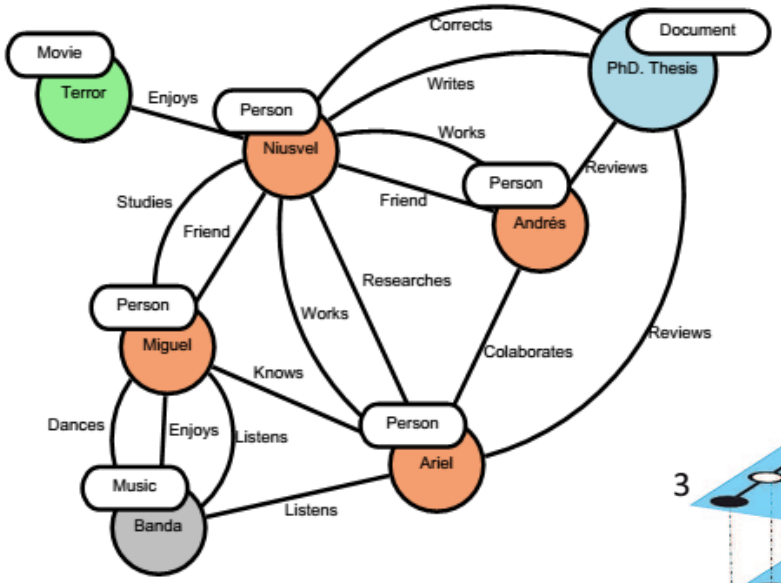
- FAS mining algorithms have become important tools in several applications, such as:

  - Analysis of biochemical structures.
  - Genetic networks analysis.
  - Circuits, links and social networks analysis.
  - Image classification.
  - ...

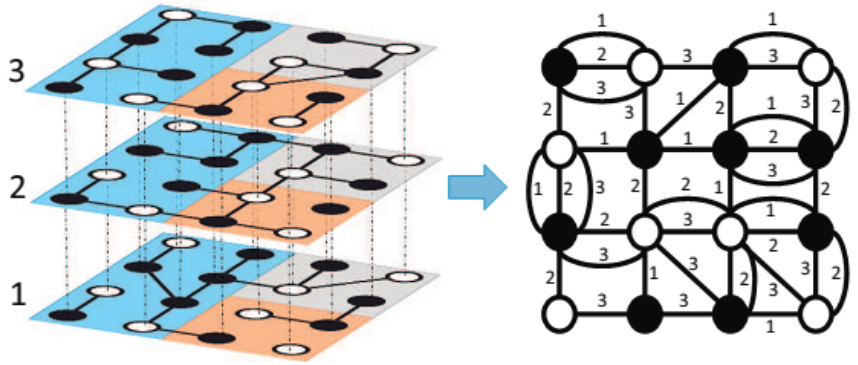


Introduction

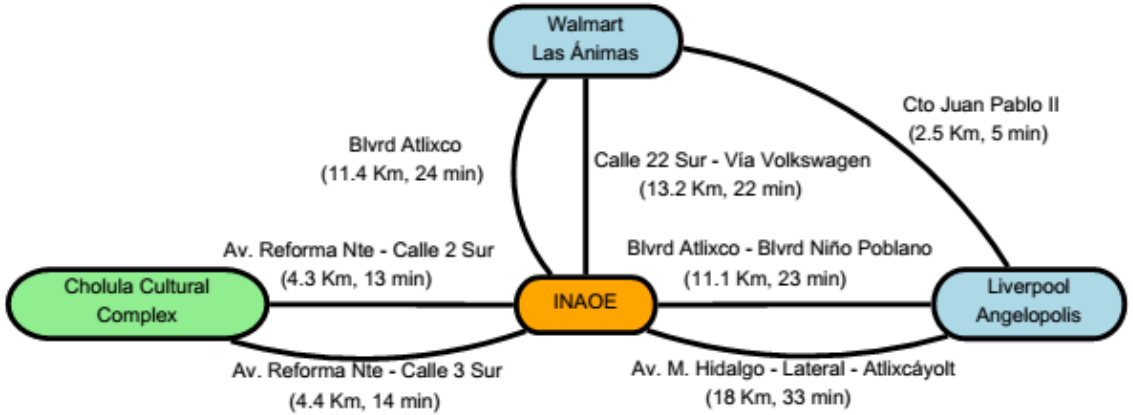
- However, there are real applications based on multi-graphs where traditional FAS miners cannot be applied because they were not designed to deal with this type of graph.



b) A social network with more than one interaction between some entities.



c) An image with three different viewpoints.



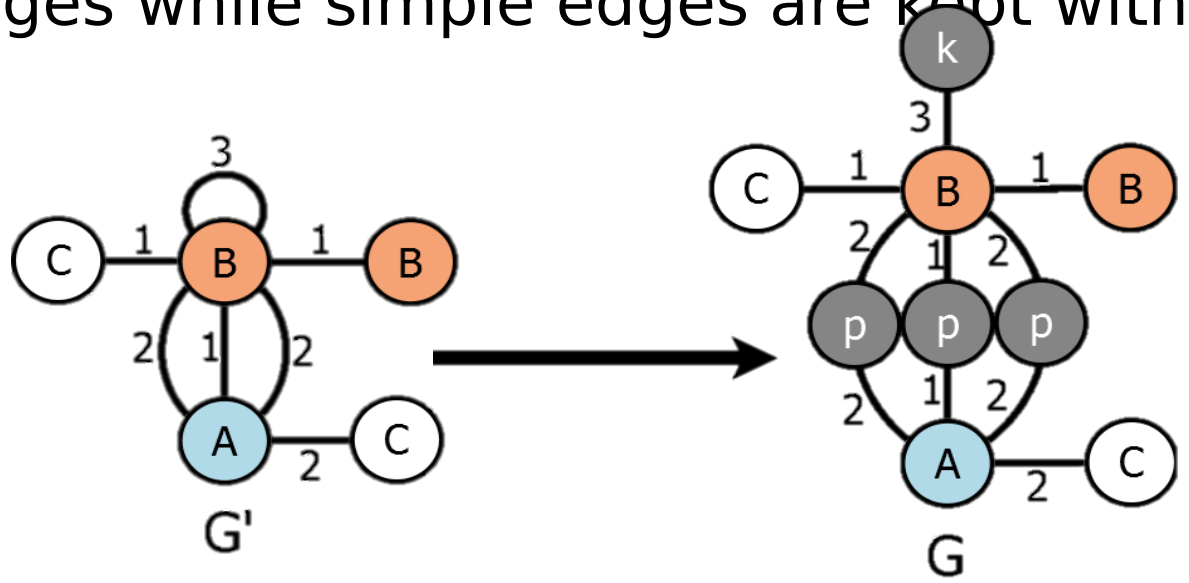
b) A transportation network with more than one path between places.

Introduction

- There is only one method based on graph transformation, which allows the use of traditional simple-graph FAS miners on multi-graph collections.
- With the aim of accelerate the mining process, we propose a more efficient method for transforming multi-graphs into simple graphs and vice versa without losing topological or semantic information that allows using traditional FAS mining algorithms and returning the mined patterns to the multi-graph space.

Proposal Based on Graph Transformations

- The transformation process consists in only transforming loops and multi-edges while simple edges are kept without changes.

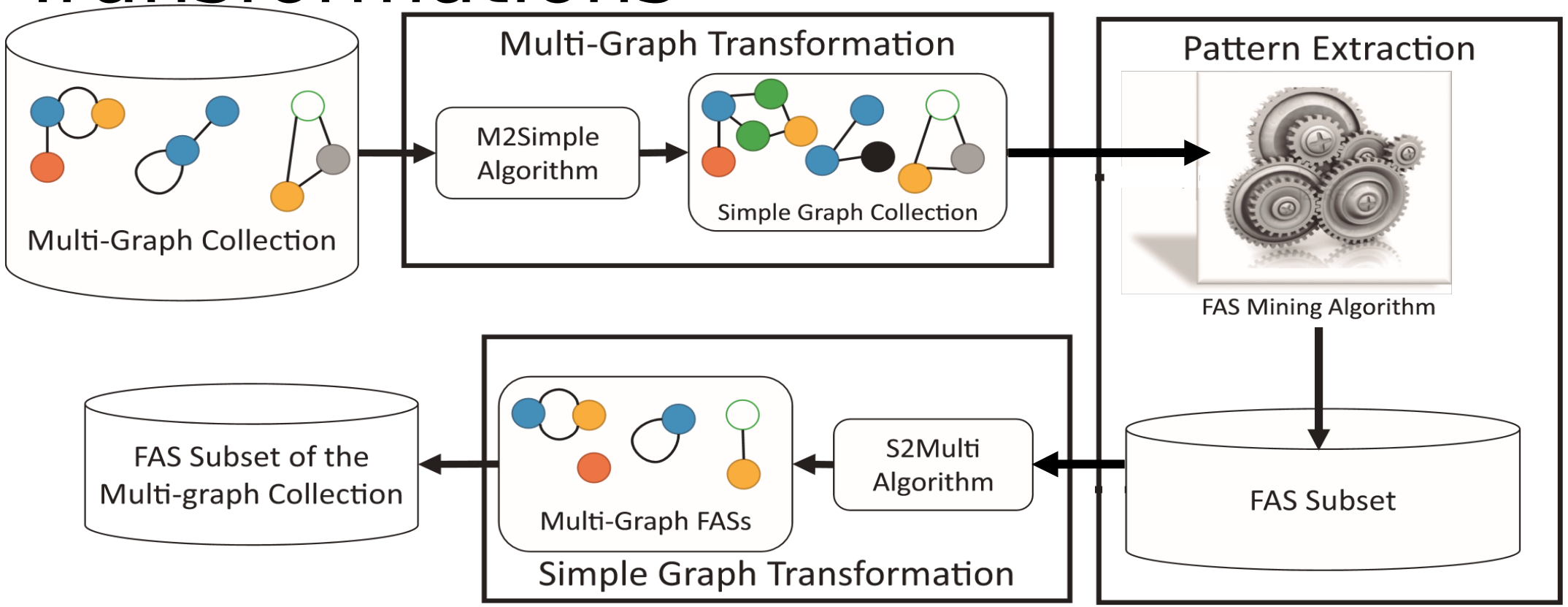


- In this way, a simple edge does not have occurrences in multi-edges and vice versa and this is an important characteristic of the proposal to be into account when it is applied.

Proposal Based on Graph Transformations

- Applying the transformation process over each graph in a given multi-graph collection, we can transform it into a simple graph collection.
  - The computational complexity of this process is $O(qd)$, where q is the average number of edges in the multi-graphs of the collection, and d is the number of multi-graphs in the collection.
  - This complexity is obtained considering that, for each multi-graph, all its edges should be visited.
- Given a multi-graph collection, through the process above described, we can get a transformed simple graph collection. Then, we can apply a conventional FAS miner, and transforming the returnable FASs into multi-graphs.
  - The process of transforming a simple graph FAS into a multi-graph (*S2Multi*) has a computational complexity $O(r)$, where r is the number of edges of the input FAS.
  - When this process is applied over a FAS set C, it has a computational complexity $O(sc)$, where c is the number of FASs in C and s is the average number of edges in the FAS in C.

Proposal Based on Graph Transformations



- Workflow for FAS mining by applying the proposed graph transformation method.

Experiments

- Several synthetic multi-graph collections are used for evaluating the performance of the proposed method.
  - We use multi-graph collections generated varying only one parameter at a time.
  - First, we fix $|D|=1000$ and $|E| = 200$, varying $|V|$ from 200 to 1000, with increments of 200.
  - Next, we fix $|V|=200$, maintaining $|D| = 1000$ and varying $|E|$ from 200 to 1000, with increments of 200.
  - Finally, we vary $|D|$ from 1000 to 5000, with increments of 1000, keeping $|V|=|E| =200$.

- The usefulness of the FASs computed by the proposed transformation method from real images for image classification is shown.
  - The two real image collections were used: COIL and ETH, which contain images of real objects taken from different viewpoints.

Used real-world Multi-graph Collections

- In COIL, we use 25 objects.
  - This collection is split into 198 (11%) images for training and 1602 (89%) for testing.
  - This collection has 144 as average graph size, 19 as average of multi-edges per graphs and 25 classes.

- In ETH, we use 6 categories (apples, cars, cows, cups, horses and tomatoes).
  - This collection is split into 615 (25%) images for training and 1845 (75%) for testing.
  - This collection has 179 as average graph size, 25 as average of multi-edges per graphs and 6 classes.

Disscusing Performance Results

- According to the achieved results, the runtime of the transformation process grows with the increment of $|D|$, $|V|$ and $|E|$, however, when the amount of edges increases, this process grows faster than by increasing the number of vertices and the number of graphs.
- The number of graphs in the collection is an important variable to take into account, because it affects the performance of the graph transformation process when it grows.
- Finally, the proposal allows mining patterns in less time than the method reported in the state-of-the-art.

Disscusing Performance Results

- According to the classification results reported in the literature by classifiers non-based on FASs, our proposal, using a simple pattern based classifier, obtained better results over the COIL collection.
- In the case of the ETH collection, we did not improve upon the results reported.
  - In spite of these results, this experiment shows the usefulness of onlyMulti, which allows transforming a multi-graph collection into a simple graph collection for applying traditional FAS miners.
- Although onlyMulti can be applied in different contexts where data are represented as multi-graphs in order to find out interesting patterns which could be useful for solving different problems.

Conclusions

- In this work, a new method (onlyMulti) for FAS mining in multi-graph collections by transforming multi-graphs into simple graphs and vice versa is proposed.
- From the experiments reported in this work, we can conclude that onlyMulti is able to mine FASs from multi-graph collections in a shorter time, producing smaller simple graphs than the only alternative option reported in the literature.
  - This is very important in order to reduce the cost of the FAS mining step.
  - Based on the experiments we can conclude that the time required for mining multi-graphs using onlyMulti is smaller than applying the closest state-of-the-art transformation method.
- In addition, the usefulness of the FASs computed over multi-graph collections by applying onlyMulti in an image classification problem was shown, where in some cases the results obtained by the patterns computed by using the proposed method outperform the results obtained by state-of-the-art classifiers non-based on FASs.