Incremental Graph Pattern based Node Matching

Guohao Sun, Guanfeng Liu, Yan Wang, Mehmet A. Orgun
Department of Computing, Macquarie University, Sydney, NSW 2109, Australia
guohao.sun@students.mq.edu.au; {guanfeng.liu, yan.wang, mehmet.orgun}@mq.edu.au

Background

Graph Pattern based Node Matching (GPNM) is to find all the matches of the nodes in a data graph G_D based on a given pattern graph G_P. GPNM has become increasingly important in many applications, e.g., group finding and expert recommendation.

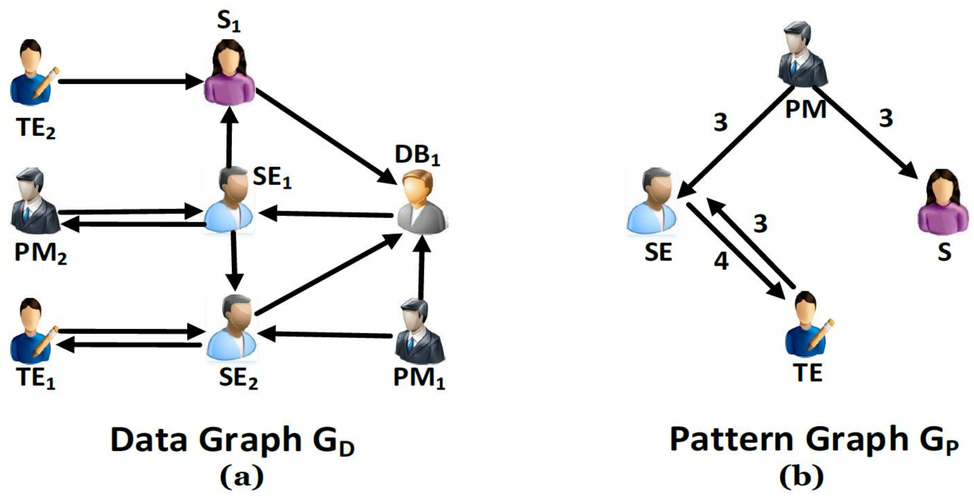

Fig. 1. Graph Pattern based Node Matching.

Problem

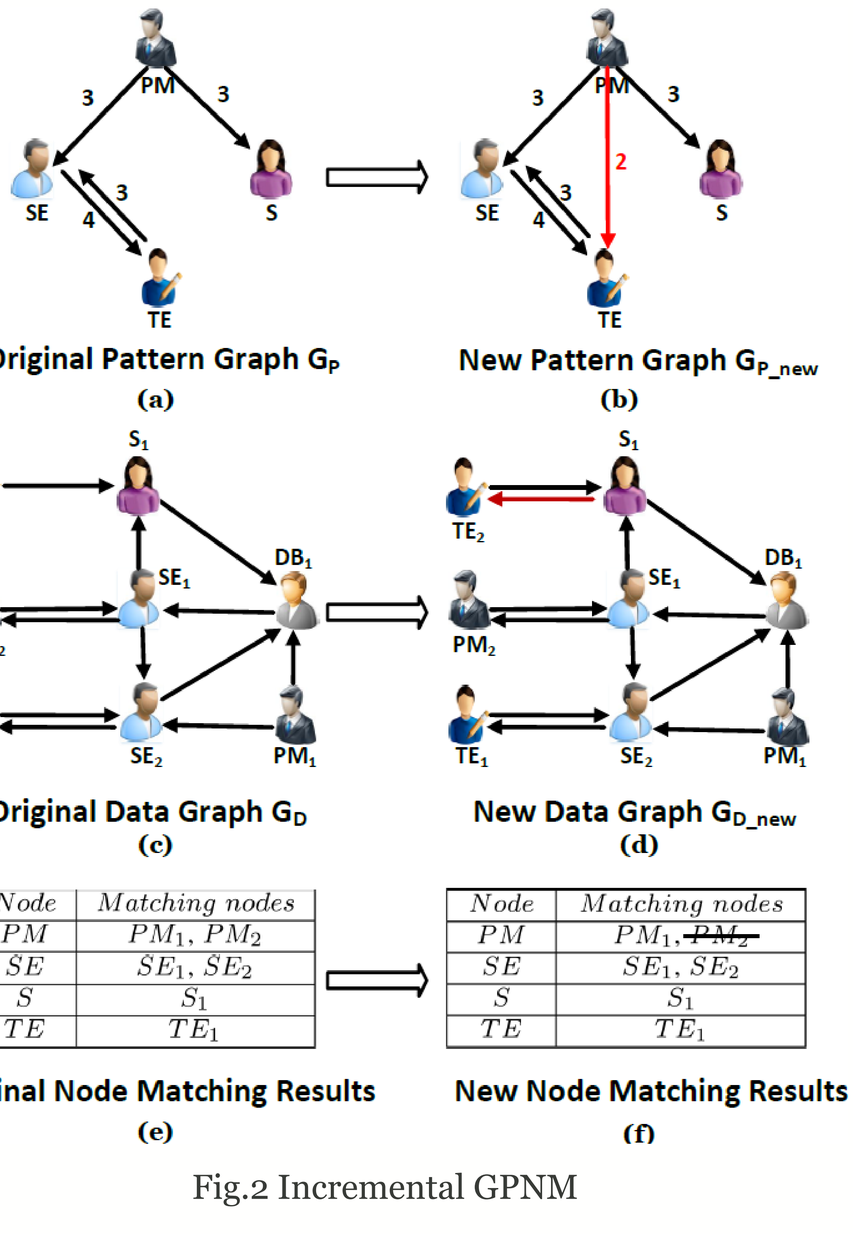

Fig.2 Incremental GPNM

In real scenarios, nodes and edges in both G_P and G_D are usually frequently updated over time.

Research Gap: The existing GPNM methods need to perform a new GPNM procedure from scratch to deliver the node matching results based on the updated G_P and updated G_D, which consumes much time. Therefore, there is a pressing need for a novel method to efficiently deliver the node matching results.

Contribution: We propose a novel INCremental GPNM method called INC-GPNM, based on the index structure and our novel search strategies, INC-GPNM can efficiently deliver node matching results taking the updates of G_P and G_D as input, and can greatly save the query processing time with improved time complexity.

Methodology

Step 1: Input the G_P, G_D and the node matching result N_{ui} .

Step 2: When G_P updates (denoted as $\triangle G_P$), we use the indices we generated to check how the $\triangle G_P$ will influence the N_{ui} . Then, we use PMatch$^+$ to deal with the situation when edges or nodes are inserted into G_P, and use PMatch$^-$ to deal with the situation when edges or nodes are deleted from G_P.

Step 3: When G_D updates (denoted as $\triangle G_D$), we use the indices we we use the indices we generated to check how the $\triangle G_D$ will influence the N_{ui_temp} . Then, we use DMatch$^+$ to deal with the situation when edges or nodes are inserted into G_D, and use DMatch$^-$ to deal with the situation when edges or nodes are deleted from G_D.

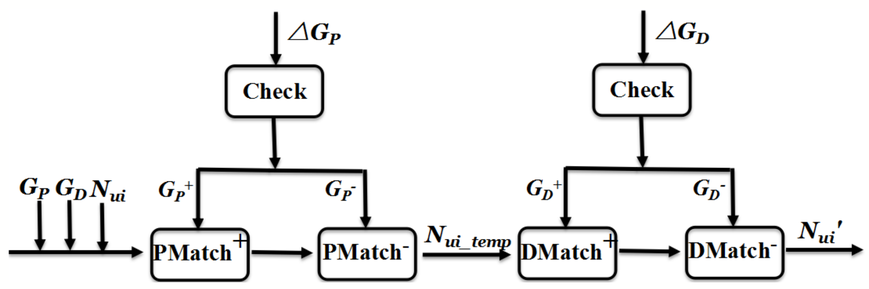

Fig.3 The basic framework of INC-GPNM

Experiment Results

In the experiments, we implement the most promising state-of-the-art graph pattern based nodes matching method as the BaseLine method, and then we compare the query processing time of the BaseLine method with that of our proposed INC-GPNM.

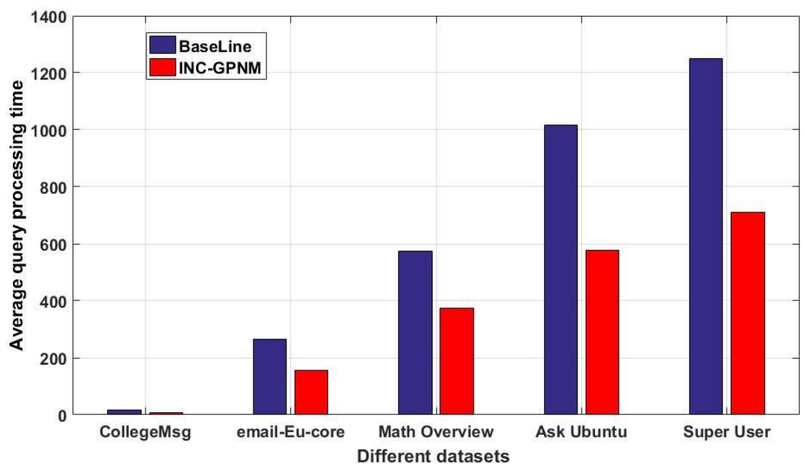

Fig.4 The query time of INC-GPNM and Baseline

Dataset	BaseLine	INC-GPNM	Comparison with BaseLine
CollegeMsg	16.82s	7.01s	**58.31% less**
email-Eu-core	266.59s	154.94	**41.88% less**
Math Overview	575.65s	374.07s	**35.02% less**
Ask Ubuntu	1015.45s	578.67s	**43.01% less**
Super User	1248.95s	710.19s	**43.14% less**

Tab.1 The comparison between INC-GPNM and Baseline

Conclusions

1. We have proposed an INCremental Graph Pattern node Matching method INC-GPNM to deliver the GPNM results based on the updates of both pattern graph and data graph.
2. The experiments on five real-world social graphs have demonstrated that our INC-GPNM significantly outperforms the state-of-the-art GPNM method in efficiency.
3. This paper has been accepted by ICDE 2018 (Rank A*).