Discrete Applied Mathematics ■ (■■■■) ■■■–■■■

Contents lists available at ScienceDirect

Discrete Applied Mathematics

journal homepage: www.elsevier.com/locate/dam

Deterministic parameterized algorithms for the Graph Motif problem☆

Ron Y. Pinter, Hadas Shachnai, Meirav Zehavi *

Department of Computer Science, Technion - Israel Institute of Technology, Haifa 3200003, Israel

ARTICLE INFO

Article history:
Received 30 July 2014
Received in revised form 16 April 2016
Accepted 21 April 2016
Available online xxxx

Keywords:
Parameterized algorithm
Graph motif
Representative family
Guiding tree

ABSTRACT

We study the classic Graph Motif problem. Given a graph $G = (V, E)$ with a set of colors for each node, and a multiset M of colors, we seek a subtree $T \subseteq G$, and a coloring assigning to each node in T a color from its set, such that T carries exactly (also with respect to multiplicity) the colors in M. Graph Motif plays a central role in the study of pattern matching problems, primarily motivated from the analysis of complex biological networks.

Previous algorithms for Graph Motif and its variants either rely on techniques for developing randomized algorithms that — if derandomized — render them inefficient, or the algebraic narrow sieves technique for which there is no known derandomization. In this paper, we present fast *deterministic* parameterized algorithms for Graph Motif and its variants. Specifically, we give such an algorithm for the more general Graph Motif with Deletions problem, followed by faster algorithms for Graph Motif and other well-studied special cases. Our algorithms make non-trivial use of *representative families*, and a novel tool that we call *guiding trees*, together enabling the efficient construction of the output tree.

1. Introduction

With the advent of network biology and complex network analysis in general, the study of pattern matching problems in graphs has become of major importance [10,15]. Indeed, the term "graph motif" plays a central role in this context, with different node colors used to model different functionalities of the network (see, e.g., [16,6]). Due to the generic nature of Graph Motif (GM) (also known as the Topology-Free Network Query problem), the so called *motif analysis* approach has become useful also in the study of social networks (see, e.g., [21] and the references therein).

GM is a natural variant of classic pattern matching problems, where the topology of the pattern M is unknown or of lesser importance. Given a graph $G = (V, E)$ with a set of colors for each node, and a multiset M of colors, we seek a subtree $T \subseteq G$, and a coloring assigning each node in T a color from its set, such that T carries exactly (also with respect to multiplicity) the colors in M. We call T an *occurrence* of M in G. To allow more flexibility in the definition of an occurrence, and since biological network data often contains noise, a generalized version of GM allows *deleting* colors from M (see below).

Parameterized algorithms solve NP-hard problems by confining the combinatorial explosion to a parameter k. More precisely, a problem is *fixed-parameter tractable (FPT)* with respect to a parameter k if it can be solved in time $O^*(f(k))$ for

Abbreviations: GM, Graph Motif; RGM, Restricted GM.

☆ A preliminary version of this paper appeared in the proceedings of the 39th *International Symposium on Mathematical Foundations of Computer Science (MFCS'14)*.

* Corresponding author. Tel.: +972 4 8295530.
E-mail addresses: pinter@cs.technion.ac.il (R.Y. Pinter), hadas@cs.technion.ac.il (H. Shachnai), meizeh@cs.technion.ac.il (M. Zehavi).

http://dx.doi.org/10.1016/j.dam.2016.04.026

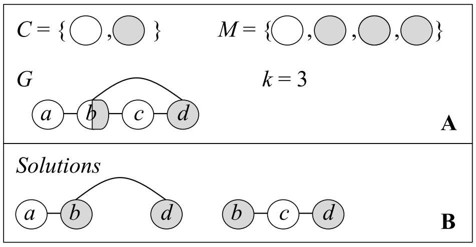

Fig. 1. (A) An input for GM_D, where a and c are associated with white, b is associated with both white and gray, and d is associated with gray. (B) Two possible solutions: Each solution consists of a subtree of G on $k = 3$ nodes, where each node is given one of its associated colors, and the number of occurrences of each color in each solution is bounded by its number of occurrences in M.

some function f, where O^* hides factors polynomial in the input size. Since GM is NP-complete [16], there is a growing body of literature studying its parameterized complexity (see the comprehensive survey in [24]). In this paper, we present fast *deterministic* parameterized algorithms for GM and its variants.

1.1. Problem statement

The most general variant considered in this paper is Graph Motif with Deletions (GM_D): the input is a set of colors C, a multiset M of colors from C, and an undirected graph $G = (V, E)$. The nodes in V are associated with colors via a (set-)coloring $Col : V \to 2^C$. We are also given a parameter $k \leq |M|$.

We need to decide if there exist a subtree $T = (V_T, E_T)$ of G on k nodes,[1] and a coloring $col : V_T \to C$ that assigns a color from $Col(v)$ to each node $v \in V_T$, such that

$$\forall c \in C : |\{v \in V_T : col(v) = c\}| \leq occ(c), \quad (1)$$

where $occ(c)$ is the number of occurrences of a color c in M (see Fig. 1).

Special cases: Restricted GM_D (RGM_D) is the special case of GM_D where for any node $v \in V$, $|Col(v)| = 1$. Also, GM and RGM are the special cases of GM_D and RGM_D, respectively, where deletions are not allowed (i.e., the inequality in (1) is replaced by equality, and $k = |M|$).

We briefly note that given an instance $(C, M, G = (V, E), Col, k)$ of GM_D, it is possible to construct an equivalent instance $(C', M', G' = (V', E'), Col', k')$ of GM as follows. Let $c \notin C$ be a new color, and define $C' = C \cup \{c\}$ and $M' = M \cup \{c\}$, i.e., c has one occurrence in M'. For each vertex $v \in V$, we generate $|M| - k + 1$ vertices v_i, $0 \leq i \leq |M| - k$, connected to v as a path whose first vertex is v_0. Formally, let $V' = V \cup \{v_i : v \in V, i \in \{0, 1, \ldots, |M| - k\}\}$ and $E' = E \cup \{\{v, v_0\} : v \in V\} \cup \{\{v_{i-1}, v_i\} : v \in V, i \in \{1, 2, \ldots, |M| - k\}\}$. Now, let $Col'(v) = Col(v)$, $Col(v_0) = \{c\}$ and $Col(v_i) = C$ for all $v \in V$ and $i \in \{1, 2, \ldots, |M| - k\}$. Finally, define $k' = |M'|$ (i.e., $k' = k + (|M| - k) + 1$). It is easy to verify that any valid solution for the GM_D instance induces a valid solution for the GM instance, and vice versa. However, the parameter k' can be significantly larger than k.

1.2. Known results and our contribution

GM_D and its variants have received considerable attention since GM was introduced by Lacroix et al. [16] (the use of deletions was introduced by Bruckner et al. [6]). The paper [16] also shows that RGM is NP-hard when M is a set and G is a tree. Even seemingly simpler cases of RGM are known to be NP-hard (see [9,2,8]). Moreover, a natural optimization version of RGM_D, minimizing the number of deletions from M, is hard to approximate within factor $|V|^{\frac{1}{3}-\epsilon}$ [22].

On the positive side, using techniques for developing randomized parameterized algorithms, many such algorithms have been obtained for GM_D and its variants [3,5–7,13,14,19,20]. Some of these algorithms can be derandomized, resulting, however, in inefficient algorithms. In particular, Fellows et al. [9] gave a deterministic algorithm for RGM that runs in time $O^*(174^k)$, based on a derandomization of the color coding technique [1]. Currently, the best randomized algorithm for GM_D, due to Björklund et al. [5], runs in time $O^*(2^k)$. This algorithm is based on the narrow sieves technique [4], for which there is no known derandomization. Thus, prior to our study, the existence of a *fast* deterministic parameterized algorithm for GM_D was open.

In this paper, we present fast deterministic parameterized algorithms for GM_D and its variants. In particular, we develop an $O^*(6.86^k)$ time algorithm for GM_D, an $O^*(5.22^k)$ time algorithm for GM, and an $O^*(5.18^k)$ time algorithm for RGM_D.

[1] In an alternative definition for GM_D, one seeks a connected subgraph S of G. This is equivalent to our definition (simply consider some spanning tree T of S).

Please cite this article in press as: R.Y. Pinter, et al., Deterministic parameterized algorithms for the Graph Motif problem, Discrete Applied Mathematics (2016), http://dx.doi.org/10.1016/j.dam.2016.04.026