Journal of Discrete Algorithms ••• (••••) •••–•••

Contents lists available at ScienceDirect

Journal of Discrete Algorithms

www.elsevier.com/locate/jda

Algorithms for topology-free and alignment network queries

Ron Y. Pinter, Meirav Zehavi *

Department of Computer Science, Technion, Haifa 32000, Israel

ARTICLE INFO

Article history:
Received 12 November 2012
Received in revised form 29 November 2013
Accepted 10 March 2014
Available online xxxx

Keywords:
Topology-free network query
Alignment network query
Computational biology
Parameterized algorithm
Subgraph homeomorphism

ABSTRACT

In this article we address two pattern matching problems which have important applications to bioinformatics. First we address the *topology-free network query problem*: Given a set of labels L, a multiset P of labels from L, a graph $H = (V_H, E_H)$ and a function $Label_H : V_H \to 2^L$, we need to find a subtree S of H which is an occurrence of P. We provide a parameterized algorithm with parameter $k = |P|$ that runs in time $O^*(2^k)$ and whose space complexity is polynomial. We also consider three variants of this problem. Then we address the *alignment network query problem*: Given two labeled graphs P and H, we need to find a subgraph S of H whose alignment with P is the best among all such subgraphs. We present two algorithms for cases in which P and H belong to certain families of DAGs. Their running times are polynomial and they are less restrictive than algorithms that are available today for alignment network queries. Topology-free and alignment network queries provide means to study the function and evolution of biological networks, and today, with the increasing amount of knowledge regarding biological networks, they are extremely relevant.

1. Introduction

Performing topology-free and alignment network queries is an important problem in the analysis of biological networks. Such queries provide means to study their function and evolution. In comparison, similar queries for sequences have been studied and used extensively for the past 30 years. Today, with the increasing amount of knowledge we have regarding biological networks, they are extremely relevant for them as well.

In both problems we are given a pattern P and a host graph H, and we need to find a subgraph S of H which resembles P. Furthermore, in both problems P and H have labels which we consider when measuring the similarity between P and the subgraphs of H.

However, there is one significant difference between them: A topology-free network query does not assume we know the topology of P and therefore requires only the connectivity of the solution, while an alignment network query assumes such knowledge and therefore requires resemblance between the topology of P and the solution. Another difference is that in the optimization version of the topology-free network query problem we consider the sum of the weights of the edges of the solution, while in the alignment network query problem we are interested in maximizing the similarity between the nodes of P and the solution. This is mostly a result of the uncertainty concerning the existence of many edges in biological networks that are more suitable for topology-free network queries.

* Corresponding author.
E-mail addresses: pinter@cs.technion.ac.il (R.Y. Pinter), meizeh@cs.technion.ac.il (M. Zehavi).

http://dx.doi.org/10.1016/j.jda.2014.03.002
1570-8667/

Table 1
Known parameterized algorithms with parameter $k = |P|$ for the topology-free network query problem (including its limited case and weighted variant).

Reference	General\limited	Time complexity	Polynomial space?	Considers weights?
Bruckner et al. [9]	General	$O^*(k!3^k)$	No	Yes
Fellows et al. [17]	Limited	$O^*(64^k)$	No	No
Betzler et al. [5]	General	$O^*(10.88^k)$	No	No
	Limited	$O^*(4.32^k)$	No	No
Guillemot et al. [20]	General	$O^*(4^k)$	Yes	No
	General	$O^*(4^k W^2)$	No	Yes
Koutis [24]	Limited	$O^*(2.54^k)$	Yes	No
This paper	General	$O^*(2^k)$	Yes	No
	General	$O^*(2^k W)$	No	Yes

Throughout this paper we use O^* to hide factors polynomial in the input size and $\tilde{O}$ to hide factors polylogarithmic in the input size. Furthermore, given a graph G, we use V_G and E_G to denote its node set and edge set, respectively.

1.1. Topology-free network queries

In the general case of the **topology-free network query problem** we are given:

1. L – A set of labels.
2. P – A multiset of labels from L.
3. H – An undirected graph.
4. $Label_H : V_H \to 2^L$.

Our goal is to find a pair $(S, label)$, where S is a subtree of H and $label$ is a function $label : V_S \to L$, such that the following requirements hold.

1. $\forall v \in V_S$: $label(v) \in Label_H(v)$.
2. $\forall l \in L$: The number of occurrences of l in P is $|\{v \in V_S : label(v) = l\}|$.

In the **limited case** of this problem we add the following restriction on the input: $\forall v \in V_H$: $|Label_H(v)| = 1$.

The topology-free network query problem was introduced by Lacroix et al. [25]. It is motivated mostly by the fact that there are many cases when querying biological networks in which we do not know the topology of P and thus require only the connectivity of the subgraph S. For example, this occurs often when querying protein–protein interaction networks [9].

We consider the general case of the problem. We allow matching a node from H with a label from P that is similar (rather than identical) to its label in order to allow more solutions. Thus each node of H is given a **set** of labels by $Label_H$ (i.e., its label and the labels similar to it). For example, when querying protein–protein interaction networks, a node representing a certain protein can be matched with any label representing a protein homologous to it.

The **weighted variant** of this problem was introduced by Bruckner et al. [9]. In this variant the edges of H have weights and we are also given a weight W. The weights can represent the confidence we have in the existence of the interactions their edges represent. As in [20], we assume that the weights are positive integers. Our goal is to find a pair $(S, label)$ which fulfills the requirements of the unweighted variant and such that the sum of the weights of the edges of S is at most W. Clearly, we can use an algorithm for this variant to solve its corresponding optimization problem (i.e., the problem in which we need to minimize the sum of the weights of the edges of S).

We note that when we refer to the problem without specifying its variant, we consider the unweighted general case.

Lacroix et al. [25] give an exponential time algorithm for the problem which is implemented by a tool called MOTUS. Fellows et al. [17] prove that the limited case of the problem is NP-hard when P is a set and H is a tree of maximum degree three, and when P consists of only two different labels and H is bipartite with maximum degree four. Moreover, the limited case of the problem is NP-hard when H is a tree of maximum degree four in which each label appears at most twice [15], and when P is a set and H has diameter two [2].

The limited case of the problem is polynomial-time solvable when H has maximum degree two, when P consists of a bounded number of different labels and H has a bounded treewidth, and when P is a set and H is a tree in which each label is shared by at most two nodes [17]. It is also polynomial-time solvable when P is a set and H is a caterpillar [2].

The problem is W[1]-hard when parameterized by the number of different labels in P [17]. Betzler et al. [4] prove that the problem is W[1]-hard when parameterized by $|V_H| - |P|$ even if P consists of only two different labels or it is a set. They also prove that if we search for an l-connected or l-edge connected subgraph of H instead of a subtree of H, the problem is W[1]-hard when parameterized by $|P|$.

On the positive side, the problem is fixed-parameter tractable with parameter $k = |P|$. Table 1 presents a summary of known parameterized algorithms for the problem. The algorithms presented in [9,17,5] are based on color coding [1], and

Please cite this article in press as: R.Y. Pinter, M. Zehavi, Algorithms for topology-free and alignment network queries, Journal of Discrete Algorithms (2014), http://dx.doi.org/10.1016/j.jda.2014.03.002