Journal of Computer and System Sciences 82 (2016) 366–379

Contents lists available at ScienceDirect

Journal of Computer and System Sciences

www.elsevier.com/locate/jcss

Linear kernel for ROOTED TRIPLET INCONSISTENCY and other problems based on conflict packing technique ☆

Christophe Paul [a], Anthony Perez [b], Stéphan Thomassé [c]

[a] LIRMM, CNRS, Université de Montpellier, France
[b] Univ. Orléans, INSA Centre Val de Loire, LIFO EA 4022, France
[c] LIP – ENS Lyon, France

ARTICLE INFO

Article history:
Available online 25 September 2015

Keywords:
Kernelization
Fixed parameterized algorithms

ABSTRACT

We develop a technique, that we call *conflict packing*, to obtain (and improve) polynomial kernels for several well-studied editing problems. We first illustrate our technique on FEEDBACK ARC SET IN TOURNAMENTS (k-FAST) yielding an alternative and simple proof of a linear kernel for this problem. The technique is then applied to obtain the first linear kernel for theDENSE ROOTED TRIPLET INCONSISTENCY (k-DENSE-RTI) problem. A linear kernel for BETWEENNESS IN TOURNAMENTS (k-BIT) is also proved. All these problems share common features. First, they can be expressed as modification problems on a dense set of constant-arity constraints. Also the consistency of the set of constraints can be characterized by means of a bounded size obstructions. The conflict packing technique basically consists of computing a maximal set of small obstructions allowing us either to bound the size of the input or to reduce the input.

1. Introduction

The theory of *fixed-parameter algorithms* [17] has been developed to cope with NP-hard problems. The goal is to identify a secondary parameter k, independent from the data-size n, which captures the exponential growth of the time complexity to solve the problem at hand. That is the complexity of a *fixed parameter tractable* algorithm is $f(k) \cdot n^{O(1)}$, where f is an arbitrary computable function.

As one of the main techniques to design efficient fixed parameter algorithms, *kernelization algorithms* [8] have attracted a lot of attention during the last few years. A kernelization algorithm transforms in polynomial time, by means of *reduction rules*, a parameterized instance (I, k) of a parameterized problem into an equivalent instance (I', k'), the so-called *kernel*, with the property that the parameter k' and the size $|I'|$ of the kernel only depend on k. The smaller the size of the kernel is, the faster the problem can be solved. Indeed, the problem can be efficiently tackled on the kernel with any exact algorithm (e.g. bounded search tree or exponential time algorithms). It is well-known that fixed-parameterized tractability of a problem is equivalent to the existence of a kernel, but of exponential size only. The standard question for a fixed-parameter tractable problem is then to provide a polynomial size kernel or give some evidence of non-existence of such a kernel [9–11].

The algorithmic tool-kit to design kernelization algorithms is getting richer and richer over the years (the interested reader should refer to standard books [17,19,26]). In this paper, we systematically apply a kernelization technique, that we called *Conflict Packing*, on four classic parameterized problems:

☆ A preliminary version of this paper appeared as an extended abstract in [27].

E-mail addresses: christophe.paul@lirmm.fr (C. Paul), anthony.perez@univ-orleans.fr (A. Perez), stephan.thomasse@ens-lyon.fr (S. Thomassé).

http://dx.doi.org/10.1016/j.jcss.2015.08.002

- Feedback Arc Set in Tournaments (k-FAST): A *feedback arc set* of a digraph $D = (V, A)$ is a subset $F \subseteq A$ of arcs such that reversing the orientation of the arcs of F yields an acyclic digraph. Given a tournament and a parameter $k \in \mathbb{N}$, the parameterized k-FAST problem consists in computing a feedback arc set of size at most k if one exists.
- Dense Rooted Triplet Inconsistency (k-dense-RTI): A *rooted triplet* is a rooted binary tree over three leaves. A binary tree over a leaf set V is a rooted binary tree such that its leaf set is mapped to V. A collection $\mathcal{R}$ of rooted triplets over a leaf set V is *consistent* if there exists a binary tree T over the leaf set V such that every rooted triplet $\{a, b, c\} \in \mathcal{R}$ is homeomorphic to the subtree of T spanning $\{a, b, c\}$. The set $\mathcal{R}$ is *dense* if it contains a rooted triplet for every distinct triplet $\{a, b, c\} \subseteq V$. Given a *dense* set $\mathcal{R}$ of rooted triplets over V and a parameter $k \in \mathbb{N}$, the parameterized k-dense-RTI problem consists in computing a subset $\mathcal{F} \subseteq \mathcal{R}$ of at most k rooted triplets to edit in order to obtain a consistent set of rooted triplets.
- Betweenness Inconsistency in Tournaments (k-BIT): Let V be a finite set. A *betweenness constraint* is defined by an ordered triple (a, b, c) of distinct elements of V. A set $\mathcal{B}$ of betweenness constraints is consistent if there exists an ordering of V such that for every $(a, b, c) \in \mathcal{B}$, either $a <_\sigma b <_\sigma c$ or $c <_\sigma b <_\sigma a$. A set $\mathcal{B}$ of betweenness constraints over V is a *betweenness tournament* if $\mathcal{B}$ contains a constraint for every triple $\{a, b, c\} \subseteq V$. Given a betweenness tournament $\mathcal{B}$ over V and a parameter $k \in \mathbb{N}$, the parameterized k-BIT problem consists in computing a subset $\mathcal{F} \subseteq \mathcal{B}$ of at most k betweenness constraints to edit in order to obtain a consistent set of betweenness constraints.

Known results The parameterized k-FAST problem is a well-studied problem from the combinatorial [18,30] as well as from the algorithmic viewpoint [3,25]. It is known to be NP-complete [3,15], but fixed-parameter tractable [4,24,29]. The first kernelization algorithms for k-FAST [4,16] yield $O(k^2)$ vertex-kernels. Recently, a kernel with at most $(2+\epsilon)k$ vertices, $\epsilon > 0$, has been proposed [6]. This latter result is strongly based on a PTAS which computes a linear vertex ordering σ with at most $(1+\epsilon)k$ backward arcs (i.e. arcs vu with $u <_\sigma v$), $\epsilon > 0$. All these kernels rely on the fact that a tournament is acyclic if and only if it does not contain any directed circuit of size three. The parameterized k-dense-RTI problem is a classic problem from phylogenetics (see e.g. [14,32]). It is known to be NP-complete [14] but fixed-parameter tractable [21, 22], the fastest algorithm running in time $O(n^4) + 2^{O(k^{1/3} \log k)}$ [22]. Moreover, Guillemot and Mnich [22] provided a quadratic vertex-kernel for k-dense-RTI. Finally, the parameterized k-BIT problem is NP-complete [2] but fixed-parameter tractable [24, 31].

Contributions It is worth noting that the problems we tackle share some similar properties. They all can be formalized as modification problems on a dense set of fixed-arity constraints. For example, in the case of Feedback Arc Set, a constraint is represented by an arc uv to express that u has to occur before v in the sought vertex ordering. Likewise the rooted triplets and the betweenness constraints form arity three constraints. They all address complete or dense instances for which the acyclicity or the consistency can be characterized by means of small bounded obstructions: directed circuit of size 3 for tournaments [folklore]; inconsistent sub-instance of size 4 for dense sets of rooted triplets [5,22] and for dense set of betweenness constraints (see Lemma 4.1).

Approaches similar to Conflict Packing have already been used in [13,33]. In this paper, using Conflict Packing we are able to either improve the best known kernel bound or provide an alternative and/or simpler proof of the best bound. In Section 2, we present a linear-vertex kernel for the parameterized k-FAST problem. Together with the Conflict Packing technique, a key ingredient in our proofs is the notion of *safe partition* introduced in [6]. Next, we use the Conflict Packing technique and adapt the notion of safe partition to the context of trees to obtain a linear vertex-kernel for the k-dense-RTI problem (Section 3). This improves the best known bound of $O(k^2)$ vertices [22]. Finally, we apply the *Conflict Packing* technique together with a sunflower-based reduction rule to the k-BIT problem. This allows us to obtain a linear-vertex kernel with at most $5k$ vertices (Section 4). Based on some of the results we present here, this latter result has recently been improved to a kernel of size at most $(2+\epsilon)k + 4$ vertices [28].

2. Feedback arc set in tournaments

In this section, we consider asymmetric and loopless directed graphs (digraphs for short). Let $D = (V, A)$ be a digraph. An arc oriented from vertex $u \in V$ to $v \in V$ is denoted uv. The *in-neighbourhood* of a vertex x is the set $N^-(x) = \{y \in V \mid yx \in A\}$ and the *out-neighbourhood* of x is $N^+(x) = \{y \in V \mid xy \in A\}$. If $V' \subseteq V$, then $D[V'] = (V', A')$ is the subdigraph *induced* by V', that is $A' = \{uv \in A \mid u \in V', v \in V'\}$. Similarly, if $A' \subseteq A$, then $D[A'] = (V', A')$ denotes the *digraph* where $V' = \{v \in V \mid \exists uv \in A' \text{ or } vw \in A'\}$. The vertices $x_0, x_1 \dots x_\ell$ form a *circuit* of length $\ell + 1$ if for every $0 \leqslant i \leqslant \ell$, $x_i x_{(i+1) \bmod \ell}$ is an arc. We say that D is *acyclic* if it does not contain any circuit and that D is *transitive* if for every triple of vertices $\{u, v, w\} \subseteq V$ such that $uv \in A$ and $vw \in A$, it holds that $uw \in A$. A *tournament* $T = (V, A)$ is a complete asymmetric digraph. We denote by $\mathbf{FAS}(D)$ the size of a minimum feedback arc set of D.

A vertex ordering σ of D is a total order on the vertex set V. We denote by $u <_\sigma v$ the fact that u is smaller than v in σ. It is well-known that every acyclic digraph has a *topological ordering* σ of the vertices: for every arc $uv \in A$, $u <_\sigma v$. An *ordered digraph* is a triple $D_\sigma = (V, A, \sigma)$ where $D = (V, A)$ is a digraph and σ a vertex ordering of D. An arc $uv \in A$ in D_σ is *backward* if $v <_\sigma u$, *forward* otherwise. Let uv be a backward arc, then $span(uv) = \{w \in V \mid v <_\sigma w <_\sigma u\}$. A *certificate* $c(uv)$ in D_σ is the vertex set of a directed path from v to u only composed of forward arcs. Observe that every vertex $w \in c(uv)$ distinct from u and v belongs to $span(uv)$ and that $c(uv)$ induces a circuit in which uv is the unique backward