Theoretical Computer Science ••• (••••) •••–•••

Contents lists available at ScienceDirect

Theoretical Computer Science

www.elsevier.com/locate/tcs

Minimal equivalent subgraphs containing a given set of arcs

Arne C. Reimers [a,b,c], Alexandra-M. Reimers [a,c,*], Yaron Goldstein [a,d]

[a] Freie Universität Berlin, Arnimallee 6, Room 101, 14195 Berlin, Germany
[b] Berlin Mathematical School, Berlin, Germany
[c] International Max Planck Research School for Computational Biology and Scientific Computing, Max Planck Institute for Molecular Genetics, Ihnestr 63-73, 14195 Berlin, Germany
[d] Forschungszentrum Matheon, Technische Universität Berlin, Straße des 17. Juni 136, 10623 Berlin, Germany

ARTICLE INFO

Article history:
Received 8 July 2015
Received in revised form 5 October 2016
Accepted 16 February 2017
Available online xxxx
Communicated by T. Erlebach

Keywords:
Transitive reduction
Transitive closure
Minimal equivalent subgraph

ABSTRACT

Transitive reductions and minimal equivalent subgraphs have proven to be a powerful concept to simplify networks and to measure their redundancy. Here we consider a generalization of the minimal equivalent subgraph problem where a set of arcs is already given. For two digraphs $D = (V, A)$, $D' = (V, A')$ with $A' \subseteq A$, we ask for the minimal set of edges of D that have to be added to D' such that the transitive closure of D equals the transitive closure of D'.

We present a method to compute such an extension and show that if D is transitively closed, this problem can be solved in the same asymptotic time as computing a transitive reduction of D.

1. Introduction

In bioinformatics, but also in many other fields, we are interested in measuring the amount of information obtained by some data analysis method. For example in the case of *flux coupling analysis* (FCA) [1], this data is provided as a preordered set. A *preordered set* $(V, \rightarrow)$ is a set V with a binary relation $\rightarrow$ that satisfies

- $a \rightarrow a$ for all $a \in V$ (reflexive)
- if $a \rightarrow b, b \rightarrow c$ for $a, b, c \in V$, then $a \rightarrow c$ (transitive).

We can model $(V, \rightarrow)$ also with a *directed graph* (digraph) $D = (V, A)$ that satisfies for all $a, b \in V$

$$a \rightarrow b \Leftrightarrow \text{there exists a directed path from } a \text{ to } b \text{ in } D.$$

We call D *transitively closed*, if

$$a \rightarrow b \Leftrightarrow (a, b) \in A \qquad \forall a, b \in V, a \neq b.$$

The *transitive closure* $\langle D \rangle = (V, B)$ of a digraph $D = (V, A)$ is the smallest transitively closed digraph that models the same preordered set as D.

* Corresponding author at: Freie Universität Berlin, Arnimallee 6, Room 101, 14195 Berlin, Germany.
E-mail addresses: arne.c.reimers@gmail.com (A.C. Reimers), alexandra.reimers@fu-berlin.de (A.-M. Reimers), goldsteiny.de@gmail.com (Y. Goldstein).

http://dx.doi.org/10.1016/j.tcs.2017.02.025
0304-3975/

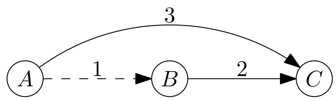

Fig. 1. Example why $f(D, D') = f(D, \emptyset) - f(D', \emptyset)$ does not always hold. The set D' is drawn with continuous edges, while D also contains the dashed edge.

A trivial method to measure the information (not in an information theoretic sense) in a preordered set would be to count the number of pairs that are related, i.e. $|A|$ for a transitively closed digraph $D = (V, A)$ that models the preordered set. However, this has the disadvantage that, as soon as one piece of additional information (one arc) is added, due to the transitivity of the relation many other additional arcs are also induced.

A more robust way to measure the information in a preordered set is to compute a transitive reduction resp. a minimal equivalent subgraph.

A *transitive reduction* of a digraph $D = (V, A)$ is defined to be a smallest digraph $D^{\min} = (V, A^{\min})$ that models the same preordered set as D [2]. Similarly, a *minimal equivalent subgraph* of a digraph $D = (V, A)$ is defined to be a smallest digraph $D^{\min} = (V, A^{\min})$ with $A^{\min} \subseteq A$ that models the same preordered set as D [3]. We observe that these two notions are equivalent if D is transitively closed.

It is easy to see that if one additional arc (one piece of information) is added, the size of $A^{\min}$ will increase by at most 1. In this paper, we now address the question of how the additional information of a digraph D compared to a smaller digraph D' can be measured and computed.

Therefore, we define in Sec. 2 the minimal extension from D' to D as the minimal number of arcs from D that need to be added to D' such that the transitive closure of D' is the same as the transitive closure of D. We then discuss in Sec. 3 how the problem can be reduced to the case where D is acyclic and to the case where D is strongly connected. We further show structural properties that are then used in Sec. 4 to present an efficient algorithm to compute minimal extensions when D is transitively closed.

We write $V^2 := V \times V$ to denote the set of ordered 2-tuples of a set V. Furthermore, we write $D|_X$ for digraphs $D = (V, A)$ to denote the induced subgraph $(X, X^2 \cap A)$. We define the complete digraph on V as $G = (V, A)$ with $A = V^2 \setminus \{(v, v) : v \in V\}$. In the following we always assume that D is simple, i.e. that it does not contain loops and parallel edges must have different orientation.

2. Minimal extensions

For digraphs $D = (V, A)$, $D' = (V, A')$ with $A' \subseteq A$ we define $f(D, D')$ as the minimum number of arcs from D that have to be added to D' so that the transitive closures become the same. Formally,

$$f(D, D') := \min_{E \subseteq A} \left\{ |E| : \langle D \rangle = \big\langle (V, A' \cup E) \big\rangle \right\}.$$

We call a minimizer $E \subseteq A$ a *minimal extension*. A set $E \subseteq A$ with $\langle D \rangle = \big\langle (V, A' \cup E) \big\rangle$ is called an *extension*.

We observe that for a digraph D, $f(D, \emptyset)$ gives the size of a minimum equivalent subgraph. Computing $f(D, \emptyset)$ is NP-hard by reduction from the Hamilton-Cycle problem [3]. If, however, D is transitively closed, this corresponds to computing the transitive reduction of D, which can be done in polynomial time [2].

We note that in general $f(D, D') \neq f(D, \emptyset) - f(D', \emptyset)$, see Fig. 1 for an example. There, $D = (\{A, B, C\}, \{1, 2, 3\})$ and $D' = (\{A, B, C\}, \{2, 3\})$. Then $f(D, \emptyset) = 2$ (the set $\{1, 2\}$), $f(D', \emptyset) = 2$ (the set $\{2, 3\}$), but $f(D, D') = 1$ (the set $\{1\}$).

Since computing $f(D, D')$ is NP-hard in general, we discuss in Sec. 4 algorithms to efficiently compute $f(\langle D \rangle, D')$. This is analogous to computing the transitive reduction instead of the minimal equivalent subgraph of D.

3. Structural properties of minimal extensions

We now discuss structural properties that will help us compute $f(D, D')$ for digraphs $D = (V, A)$, $D' = (V, A')$ with $A' \subseteq A$. Let $\mathcal{C} \subseteq 2^V$ be the set of maximal strongly connected components (represented as the corresponding sets of vertices) of D.

A main simplification is obtained by contracting strongly connected components. This leaves us with a directed acyclic graph (DAG) on which the minimal extension problem is easy (see Observation 1). We therefore define the component graphs $\tilde{D}, \tilde{D}'$, where the strongly connected components of D are contracted to single nodes (see Fig. 2 for an example):

$$(\mathcal{C}, \tilde{A}) = \tilde{D} := D/\mathcal{C}$$
$$(\mathcal{C}, \tilde{A}') = \tilde{D}' := D'/\mathcal{C},$$

where for graphs $G = (W, B)$ and partitions $\mathcal{P}$ of W we define