

Minimal Equivalent Subgraphs Containing a Given Set of Arcs

Arne C. Reimers^{1,2,3,*}, Alexandra M. Reimers^{1,2}, Yaron Goldstein^{1,4}

March 27, 2014

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 polynomial time.

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 $a, b \in V$

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

¹ Department of Mathematics and Computer Science, Freie Universität Berlin, Arnimallee 6, 14195 Berlin, Germany

² International Max Planck Research School for Computational Biology and Scientific Computing (IMPRS-CBSC), Max Planck Institute for Molecular Genetics, Ihnestr 63-73, D-14195 Berlin, Germany

³ Berlin Mathematical School (BMS), Berlin, Germany

⁴ DFG-Research Center Matheon, Berlin, Germany

* e-mail: arne.c.reimers@gmail.com

We call D *transitively closed*, if

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

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 .

A trivial method to measure the information 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.

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 $(X, X^2 \cap A)$. 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,

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

We call a minimizer $E \subseteq A$ a *minimal extension*. A set $E \subseteq A$ with $\langle D \rangle = \langle (V, A' \cup E) \rangle$ 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 [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 Figure 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 $\{1\}$).

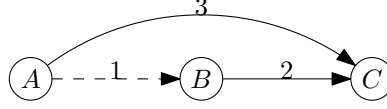


Figure 1: Example why $f(D, D') = f(D, \emptyset) - f(D', \emptyset)$ does not always hold. The set D' is drawn with continuous edges, while D also contains the dashed edge.

3 Computation of Minimal Extensions

We now discuss how to 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 .

We now define digraphs \tilde{D}, \tilde{D}' , where the strongly connected components of D are contracted to single nodes (see Fig. 2 for an example):

$$\begin{aligned} \tilde{D} &:= (\mathcal{C}, \tilde{A}) && \text{with} \\ \tilde{A} &:= \{(C_1, C_2) \in \mathcal{C}^2 : \exists(a, b) \in A, a \in C_1, b \in C_2\} \end{aligned}$$

and

$$\begin{aligned} \tilde{D}' &:= (\mathcal{C}, \tilde{A}') && \text{with} \\ \tilde{A}' &:= \{(C_1, C_2) \in \mathcal{C}^2 : \exists(a, b) \in A', a \in C_1, b \in C_2\}. \end{aligned}$$

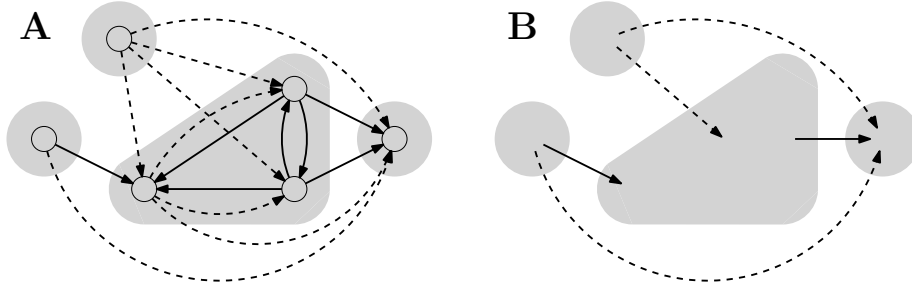


Figure 2: **A)** D is the digraph containing both the continuous and the dashed edges, while D' is the subgraph of D that contains only the continuous edges. The individual connected components of \mathcal{C} are shaded in grey. **B)** \tilde{D} is the digraph drawn with continuous and dashed edges, while \tilde{D}' is the one drawn only with continuous edges.

Proposition 1 \tilde{D} and \tilde{D}' are acyclic.

PROOF If \tilde{D} would contain a cycle, this would contradict the maximality of the components in \mathcal{C} .

Since $\tilde{A}' \subseteq \tilde{A}$, if X is a cycle in \tilde{A}' , then X is also a cycle in \tilde{A} . This contradicts the acyclicity of \tilde{D} proven above. Hence, \tilde{D}' is also acyclic. ■

Proposition 2 *If D is transitively closed, then \tilde{D} is transitively closed.*

PROOF Assume $(C_1, C_2) \in \tilde{A}$ and $(C_2, C_3) \in \tilde{A}$. By definition, there exists $a \in C_1, b \in C_2, b' \in C_2, c \in C_3$ with $(a, b) \in A$ and $(b', c) \in A$. Since $D|_{C_2}$ is strongly connected and transitively closed, $b, b' \in C_2$ implies that $(b, b') \in A$. Therefore, there exists a path (a, b, b', c) from a to c , and, from the transitivity of A , there must exist also an edge $(a, c) \in A$. It then follows that $(C_1, C_3) \in \tilde{A}$. ■

Theorem 1 *It holds that*

$$f(D, D') = f(\tilde{D}, \tilde{D}') + \sum_{C \in \mathcal{C}} f(D|_C, D'|_C).$$

PROOF \geq : Let $E \subseteq A$ minimal s.t. $\langle D \rangle = \langle (V, A' \cup E) \rangle$. Clearly, each edge of E either lies inside a single strongly connected component of D , or connects two different strongly connected components of D .

Claim 3.1 $\langle (C, (A' \cup E) \cap C^2) \rangle = \langle D|_C \rangle$ for all $C \in \mathcal{C}$.

PROOF Let B be the edges of the graph $G := \langle (C, (A' \cup E) \cap C^2) \rangle$. Let $e = (s, t) \in A \cap C^2$ arbitrary but fixed. Since $\langle D \rangle = \langle (V, A' \cup E) \rangle$ it follows that there exists a path P that only uses edges in $A' \cup E$ from s to t . Since C is a strongly connected component, it follows that all the edges in P are contained in C . Hence, the path also exists in G and thus, $e \in B$ and $A \cap C^2 \subseteq B$. Since $(A' \cup E) \cap C^2 \subseteq A \cap C^2$, the claim follows. ■

Claim 3.2 $\langle \tilde{D} \rangle = \langle (C, \tilde{A}' \cup \tilde{E}) \rangle$ with $\tilde{E} := \{(C_1, C_2) \in \mathcal{C}^2 : s \in C_1, t \in C_2, (s, t) \in E\}$.

PROOF Let $G = (C, B) := \langle (C, \tilde{A}' \cup \tilde{E}) \rangle$. Let $e = (C_1, C_2) \in \tilde{A}$ be arbitrary but fixed. Since $\langle D \rangle = \langle (V, A' \cup E) \rangle$ it follows that there exists $s \in C_1, t \in C_2$ with $(s, t) \in A$ and a path $P = (s = p_1, p_2, \dots, p_n = t)$, $n \geq 2$ using only edges in $A' \cup E$ from s to t . Let $Q := (q_1, q_2, \dots, q_n)$ with q_i being the strongly connected component of \mathcal{C} that contains p_i , $i = 1, \dots, n$. Note that $q_1 = C_1$ and $q_n = C_2$. It follows that Q is a path in $(C, \tilde{A}' \cup \tilde{E})$. Hence, $e \in B$ and $\tilde{A} \subseteq B$. Since $\tilde{A}' \cup \tilde{E} \subseteq \tilde{A}$, the claim follows. ■

From the claims it follows that E is an extension for each connected component $C \in \mathcal{C}$ of size $|E \cap C \times C| \geq f(D|_C, D'|_C)$ and it also induces a valid extension for \tilde{D}' to \tilde{D} of size $|\{(s, t) \in E : s \in C_1 \in \mathcal{C}, t \in C_2 \in \mathcal{C}, C_1 \neq C_2\}| \geq |\tilde{E} \cap \mathcal{C} \times \mathcal{C}| \geq f(\tilde{D}, \tilde{D}')$. Since an edge is either inside a connected component or between two connected components, the \geq relation of the formula follows.

\leq : Let $E_C \subseteq A$ be an extension for each connected component $C \in \mathcal{C}$, i.e. $\langle\langle C, (A' \cap C \times C) \cup E_C \rangle\rangle = \langle D|_C \rangle$. Let $\tilde{E} \subseteq \tilde{A}$ be a valid extension from \tilde{D}' to \tilde{D} , i.e. $\langle\langle \tilde{D} \rangle\rangle = \langle\langle \mathcal{C}, \tilde{A}' \cup \tilde{E} \rangle\rangle$.

We now construct a valid extension for D' to D with $\sum_{C \in \mathcal{C}} |E_C| + |\tilde{E}|$ edges. For each $e = (C_1, C_2) \in \tilde{A}'$ choose an arbitrary but fixed $s(e) \in C_1, t(e) \in C_2$ with $(s(e), t(e)) \in A'$, and for each $e = (C_1, C_2) \in \tilde{E}$ choose an arbitrary but fixed $s(e) \in C_1, t(e) \in C_2$ with $(s(e), t(e)) \in A$. Define $\bar{E} := \{(s(e), t(e)) : e \in \tilde{E}\}$ and $E := \bar{E} \cup \bigcup_{C \in \mathcal{C}} E_C$.

Claim 3.3 $\langle D \rangle = \langle\langle V, A' \cup E \rangle\rangle$

PROOF Let $G = (V, B) := \langle\langle V, A' \cup E \rangle\rangle$.

Let $s \in C_1 \in \mathcal{C}, t \in C_2 \in \mathcal{C}$ with $(s, t) \in A$. By definition of \tilde{E} there exists a path $Q = (C_1 = q_1, q_2, \dots, q_n = C_2), n \geq 1$ using edges in \tilde{A}' and \tilde{E} from C_1 to C_2 . Let $P := (s, s(q_1, q_2), t(q_1, q_2), s(q_2, q_3), t(q_2, q_3), \dots, s(q_{n-1}, q_n), t(q_{n-1}, q_n), t)$. By construction there exist edges from $s(q_i, q_{i+1})$ to $t(q_i, q_{i+1}), i = 1, \dots, n-1$ in $\bar{E} \cup A' \subseteq B$.

We observe that $\{s, s(q_1, q_2)\}, \{t(q_{i-1}, q_i), s(q_i, q_{i+1})\}$ for $i = 1, \dots, n-1$, and $\{t(q_{n-1}, q_n), t\}$ are each in the same connected component. Since $E_C \subseteq E$ is a valid extension for each connected component $C \in \mathcal{C}$, there exists for all $a, b \in C$ a path using edges in $A' \cup E_C$ from a to b . Hence, $(a, b) \in B$. Hence, it follows that $(s, s(q_1, q_2)) \in B, (t(q_{i-1}, q_i), s(q_i, q_{i+1})) \in B$ for $i = 1, \dots, n-1$, and $(t(q_{n-1}, q_n), t) \in B$.

It follows that P is a path connecting s to t using only edges in B . Since G is transitively closed, it follows that $(s, t) \in B$. Thus, $B \supseteq A$ and since $A' \cup E \subseteq A$, we have $G = \langle D \rangle$. \blacksquare

We observe that $|E| = |\tilde{E}| + \sum_{C \in \mathcal{C}} |E_C|$ and the theorem follows. \blacksquare

By Thm. 1 it follows that we only need to be able to compute $f(D, D')$ for digraphs that are acyclic and for digraphs where D forms one strongly connected component. We first analyse the case of acyclic digraphs. The following proposition characterizes the arcs in a minimal extension. It should be noted that it does not apply for digraphs in general, since in general there exists no unique minimal extension.

Proposition 3 (Locality) *Assume D, D' are acyclic and let E be minimal s.t. $\langle D \rangle = (V, B) = \langle\langle V, A' \cup E \rangle\rangle$. Then it holds for $e = (a, b) \in A$ that $e \in E$ if and only if there exists no $c \in V$ with $(a, c) \in B$ and $(c, b) \in B$.*

PROOF \Rightarrow : Assume there exists $(a, c), (c, b) \in B$. Due to the acyclicity of D it follows that there exists no path in A that connects a to c and uses (a, b) and there exists no path in A that connects c to b and uses (a, b) . Thus, there exists no path in $A' \cup E$ that connects a to c or c to b and uses (a, b) . It follows that there exists a path from a to b that does not use (a, b) . Hence, $(a, b) \notin E$.

\Leftarrow : Assume there exists no $c \in V$ with $(a, c) \in B$ and $(c, b) \in B$. Since $\langle D \rangle$ is closed under transitivity, it follows that there exists no path from a to b in A that does not use (a, b) . Hence $(a, b) \in E$. \blacksquare

We observe that from the locality property (Prop. 3), we can easily compute a minimal extension and hence $f(\tilde{D}, \tilde{D}')$. We now only have to show how to compute the size of a valid extension for each of the strongly connected components.

Theorem 2 (Full extension) *Let $D = (V, A)$, $V \neq \emptyset$ be the complete digraph and $D' = (V = S \dot{\cup} T, A')$ be a bipartite digraph with edges only going from the sources S to the sinks T and assume each vertex is used by at least one edge. Then $f(D, D') = \max\{|S|, |T|\}$.*

PROOF Since each node in S has at least one ingoing arc and each node in T has at least one outgoing arc, it can be easily seen that $f(D, D') \geq \max\{|S|, |T|\}$. Hence, we only show the other direction.

Define $E \subseteq A'$ by iteratively removing one edge $e = (s, t)$ and its vertices (including incident edges) from $G = D'$ and adding e to E until it is not possible anymore. Let V' be the vertices used by edges in E . We observe that (V', E) consists of $|E|$ weakly connected components (each a single edge). It follows that there exists $|E|$ edges $F \subseteq (V' \cap T) \times (V' \cap S)$ such that $\langle (V', E \cup F) \rangle$ is one strongly connected component.

We further observe that for each $s \in S$ there exists $t \in V'$ with $(s, t) \in A'$ and for each $t \in T$ there exists $s \in V'$ with $(s, t) \in A'$ (otherwise we could add another edge to E .) It follows that in $(V, A' \cup F)$ we can reach from every $s \in S$ every $v \in V$, and every $t \in T$ can be reached from every $v \in V$.

We now iteratively add edges (t, s) to F where $t \in T, s \in S$ are not yet used by any edge in F . We observe that after this operation $|F| = \min\{|S|, |T|\}$ and either only sources with no in-arc or sinks with no out-arc are left. For each of these nodes we now add an edge to F to connect it to a node that has in-arcs and out-arcs. It follows that finally $|F| = \max\{|S|, |T|\}$.

It can be easily seen that now every source can also be reached from at least one sink (and hence from any node) and from each sink we can reach a source and hence, the whole network. ■

Usually, however, the connected components $C \in \mathcal{C}$ do not have the form as in Thm. 2, i.e. $D|_C$ is not bipartite. Hence, we have to bring them into that form first.

Corollary 1 *Let $D = (V, A)$, $V \neq \emptyset$ be a complete digraph and D' a digraph. Then*

$$\begin{aligned} f(D, D') &= 0 & \text{if } \langle D' \rangle &= D \\ f(D, D') &= \max\{S, T\} + K, & \text{otherwise} \end{aligned}$$

where S is the number of sources (nodes without in-arc), T is the number of sinks (nodes with out-arc) and K is the number of isolated (not weakly connected to anything else in D') strongly connected components of D' .

PROOF If $\langle D' \rangle = D$, it is strongly connected and we obviously do not have to add any arcs.

For each strongly connected component of D' , we do not have to add any arcs. Hence, we can contract all strongly connected components to a single node

and assume w.l.o.g. that D' does not contain non-trivial strongly connected components (see Fig. 3 A and B).

We notice that after this reduction we have K isolated nodes.

We observe that every node that is not a source can be reached from a source, and from each node that is not a sink we can reach a sink. It follows that for each $E \subseteq A$ where every source is reachable from every sink and vice versa in $(V, A' \cup E)$ it holds that $(V, A' \cup E)$ is strongly connected. Hence, we can ignore nodes that are neither sources nor sinks (see Fig. 3 C).

Since D' contains more than one strongly connected component (i.e. after the transformations D' contains not only a single node) it follows that we have to add one in-arc and one out-arc for each isolated node in D' which is sink and source at the same time. Hence, we can split up each isolated node into a sink and a source without having to invest more edges to build a valid extension (see Fig. 3 D). Finally, Thm. 2 yields $f(D, D') = \max\{S+K, T+K\} = \max\{S, T\} + K$ (see Fig. 3 E and F). ■

Note that in Thm. 1 and Prop. 3 we do not require that D is transitively closed. This property is however important for Thm. 2 because otherwise a minimal extension cannot be computed in polynomial time. This can be seen by studying $f(D, \emptyset)$, where $\langle D \rangle$ is the complete digraph. This corresponds to finding a Hamiltonian cycle in D [2, 3].

4 Conclusion

We have shown that we can compute the size of a minimal extension $f(D, D')$ for digraphs D, D' by computing minimal extensions for each of the strongly connected components and for the acyclic digraph obtained by contracting all the strongly connected components of D into single nodes.

We showed that, for directed acyclic digraphs, $f(D, D')$ can be computed by simple reachability tests in polynomial time. If D is the complete digraph, we can also efficiently compute the size of an extension by simply counting the number of source and sink nodes in D' .

Alltogether, we presented a polynomial time algorithm for transitively closed D to compute the number $f(D, D')$ of arcs to be added from D to D' such that the transitive closures become the same.

References

- [1] A. P. Burgard, N. E. V., C. H. Schilling, and C. D. Maranas, “Flux coupling analysis of genome-scale metabolic network reconstructions,” *Genome Research*, vol. 14, no. 2, p. 301–312, 2004.
- [2] A. V. Aho, M. R. Garey, and J. D. Ullman, “The transitive reduction of a directed graph,” *SIAM Journal on Computing*, vol. 1, no. 2, pp. 131–137, 1972.

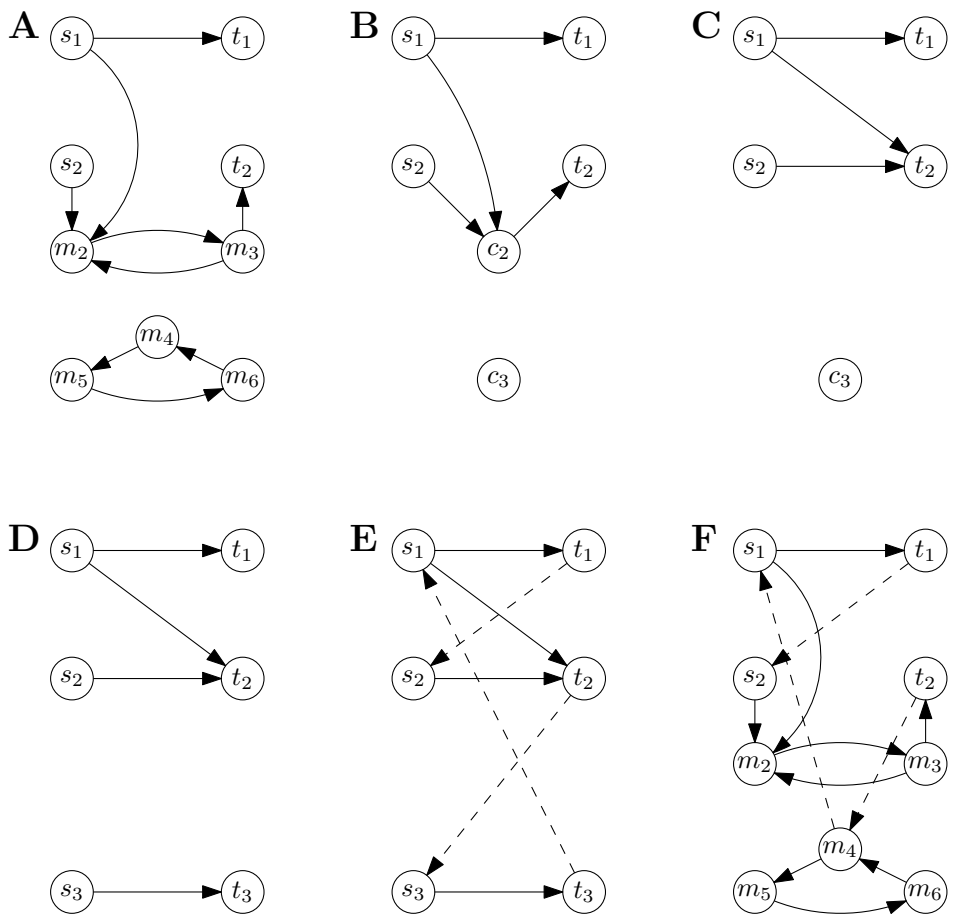


Figure 3: **A)** An example for D' . **B)** Compressed strongly connected components. **C)** Ignored nodes that are neither sources nor sinks. **D)** Splitting isolated nodes. **E)** Extension of D' (in dashed edges). **F)** Extension in the original digraph.

[3] D. M. Moyles and G. L. Thompson, "An algorithm for finding a minimum equivalent graph of a digraph," *Journal of the ACM*, vol. 16, no. 3, pp. 455–460, 1969. [Online]. Available: <http://doi.acm.org/10.1145/321526.321534>