Approximation and Hardness Results for the Maximum Edges in Transitive Closure Problem

Anna Adamaszek[1], Guillaume Blin[2,3], and Alexandru Popa[4(✉)]

[1] Max-Planck-Institut Für Informatik, Saarbrücken, Germany
anna@mpi-inf.mpg.de
[2] LaBRI, UMR 5800, University of Bordeaux, 33400 Talence, France
[3] CNRS, LaBRI, UMR 5800, 33400 Talence, France
guillaume.blin@labri.fr
[4] Faculty of Informatics, Masaryk University, Brno, Czech Republic
popa@fi.muni.cz

Abstract. In this paper we study the following problem, named Maximum Edges in Transitive Closure, which has applications in computational biology. Given a simple, undirected graph $G = (V, E)$ and a coloring of the vertices, remove a collection of edges from the graph such that each connected component is *colorful* (i.e., it does not contain two identically colored vertices) and the number of edges in the transitive closure of the graph is maximized.

The problem is known to be APX-hard, and no approximation algorithms are known for it. We improve the hardness result by showing that the problem is NP-hard to approximate within a factor of $|V|^{1/3-\varepsilon}$, for any constant $\varepsilon > 0$. Additionally, we show that the problem is APX-hard already for the case when the number of vertex colors equals 3. We complement these results by showing the first approximation algorithm for the problem, with approximation factor $\sqrt{2 \cdot \text{OPT}}$.

1 Introduction

The Maximum Edges in Transitive Closure problem we consider in this paper belongs to the framework of colorful components problems.

COLORFUL COMPONENTS FRAMEWORK: Given a simple, undirected graph $G = (V, E)$ and a coloring $\sigma : V \to C$ of the vertices with colors from a given set C, remove a collection of edges $E' \subseteq E$ from G such that each connected component in the resulting graph $G' = (V, E \backslash E')$ is a *colorful component* (i.e., it does not contain two identically colored vertices). We want the graph G' to be optimal according to some fixed *optimization measure*.

In our problem, the optimization measure is the number of edges in the transitive closure. For a graph consisting of k connected components, each containing respectively $a_1, a_2, \ldots, a_k$ vertices, the number of edges in the transitive closure of the graph is

$$\sum_{i=1}^{k} \frac{a_i \cdot (a_i - 1)}{2} \quad .$$

MAXIMUM EDGES IN TRANSITIVE CLOSURE (MEC): Given a simple, undirected graph $G = (V, E)$ and a coloring $\sigma : V \rightarrow C$ of the vertices, remove a collection of edges $E' \subseteq E$ from G such that each connected component in the resulting graph $G' = (V, E \backslash E')$ is colorful, and the number of edges in the transitive closure of G' is maximum.

Motivation. The colorful components framework is motivated by applications in comparative genomics [8,10], which is a fundamental branch of bioinformatics studying the relationship of the genome structure between different biological species. One of the key problems in this area, the multiple alignment of gene orders, can be captured as a graph theoretical problem, using the colorful components framework, where the colorful graphs represent similarity relationships between genes from different homologous gene families. A partition into colorful components corresponds then to a partition of genes into orthology sets, where any two genes from the same genome belong to different orthology sets. We refer the reader to [10] for a more detailed description of the connection between the multiple alignment of gene orders and the graph theoretic framework considered.

The understanding of orthologous genes of two different genomes as originating from a single gene in the most recent common ancestor of the two species leads to transitivity as a property of the orthology relation. This motivates the study of MEC (see [10] for more details, and for a discussion why MEC yields good results in practice).

Related Work. The Maximum Edges in Transitive Closure problem has been introduced by Zheng et al. [10]. They present heuristic algorithms for the problem, without giving any worst-case approximation guarantee. They also conjecture the problem to be NP-hard. Adamaszek and Popa [1] prove that MEC is APX-hard, even in the case of 4 vertex colors.

The colorful components framework appeared first in the paper by Zheng et al. [10] and has been formally defined by Adamaszek and Popa [1], although problems which fit into this framework have already been studied earlier. We now summarize known results for these problems.

In the problem named either Colorful Components [3,4] or Minimum Orthogonal Partition [5,10], the objective function is to minimize the number of edges removed from G to obtain a graph in which all connected components are colorful. Bruckner et al. [4] show that the problem is NP-hard for three or more colors and they study fixed-parameter algorithms for the problem. Their NP-hardness reduction can be modified slightly to show the APX-hardness of the problem (see [1]). Zheng et al. [10] and Bruckner et al. [3] study heuristic approaches for the problem, and He et al. [5] present an approximation algorithm for some special case of the problem. As the general problem is a special case of the Minimum Multi-Multiway Cut problem, it admits a $O(\log |C|)$ approximation algorithm [2].

Zheng et al. [10] introduce the Minimum Singleton Vertices problem (MSV), where the goal is to minimize the number of isolated vertices in the resulting graph. Zheng et al. [10] present heuristic algorithms for the problem, without giving any worst-case approximation guarantee. They also conjectured that the problem is NP-hard. Tremblay-Savard and Swenson [9] consider a Maximum Orthogonal Edge Cover problem (MAX-OREC), which is a dual problem to MSV. There, the goal is to cover a maximum number of vertices of a graph using vertex-disjoint, non-singleton connected colorful subgraphs. In [9], a 2/3-approximation algorithm for MAX-OREC is presented. Adamaszek and Popa [1] prove that MSV (and therefore also MAX-OREC) can be solved exactly in polynomial time, thus disproving the conjecture in [10].

Adamaszek and Popa [1] introduce another problem, termed Minimum Colorful Components, in which the goal is to delete a subset of edges such that the resulting graph has only colorful components and the number of connected components is minimized. They show that this problem cannot be approximated within a factor of $|V|^{1/14-\varepsilon}$ unless $P = NP$, and within a factor $|V|^{1/2-\varepsilon}$ unless $ZPP = NP$.

Our Results. In this paper we improve the hardness results for the MEC problem, and we present the first approximation algorithm.

First, we show that MEC is APX-hard even for the case when $|C| = 3$. This settles the complexity of the problem when the number of colors is a constant, as for $|C| = 2$ the MEC problem can be solved exactly in polynomial time by using a maximum matching algorithm. Our proof is via a reduction from the Maximum Bounded 3-Dimensional Matching problem (Max 3-DM-3).

For the general case, when the number of colors is arbitrary, we show that MEC is NP-hard to approximate within a factor of $|V|^{1/3-\varepsilon}$ for any constant $\varepsilon > 0$. This result holds even if the input graph is a tree and each color appears at most twice in the graph. We use the same reduction from the Independent Set as Rizzi and Sikora for proving hardness of approximation of the Graph Motif problem [7].

We also show the first polynomial-time approximation algorithm for MEC, which has a ratio of $\sqrt{2 \cdot \text{OPT}}$. We use the exact polynomial time algorithm for the Minimum Singleton Vertices problem [1] to obtain a partition into colorful components and then we show that this partition has a big enough number of edges in the transitive closure.

2   APX-hardness of MEC for $|C| = 3$

In this section, we prove that the MEC problem restricted to instances using only 3 colors is APX-hard. The proof is via a reduction from the Maximum Bounded 3-Dimensional Matching problem. This result strengthens the one presented in [1], which holds for problem instances using 4 colors.

Before we give the reduction, we first state the definition of Max 3-DM-3 and the known hardness result for it.

MAXIMUM BOUNDED 3-DIMENSIONAL MATCHING (MAX 3-DM-3): The input consists of pairwise disjoint sets X, Y, Z and a collection $T \subseteq X \times Y \times Z$ of

triples such that each element from X, Y and Z occurs in at least one and at most three triples in T. The aim is to find a feasible subset of triples $T' \subseteq T$ (i.e., no two elements of T' agree on any coordinate) of maximum cardinality.

Theorem 1 (Theorem 4.4 in [6], Rephrased). *There exists a constant $\varepsilon > 0$ such that it is NP-hard to distinguish between the instances of Max 3-DM-3 with the following properties:*

1. *There is a feasible collection of triples $T' \subseteq T$ such that every element of X, Y and Z belongs to some triple in T'.*
2. *For every feasible collection of triples $T' \subseteq T$ less than $(1 - \varepsilon)$ fraction of elements from $X \cup Y \cup Z$ belong to some triple of T'.*

Without loss of generality we can assume that $|X| = |Y| = |Z| = n$, since if $|X|$, $|Y|$ and $|Z|$ are different, then the case 1 of Theorem 1 cannot hold. Also, define $N = |T|$. It holds that $N \leq 3n$, since each element of $X \cup Y \cup Z$ appears in at most three triples. In the rest of the section, we use $\text{OPT}_{3\text{DM}}$ to denote the size of an optimal solution of a Max 3-DM-3 instance (the instance we refer to will always be clear from the context), and OPT_{MEC} to denote the value of an optimal solution (i.e., the number of edges in the transitive closure of the graph) of the MEC instance obtained via the reduction.

Reduction. Given an instance (X, Y, Z, T) of Max 3-DM-3, we create an instance $(G = (V, E), \sigma)$ of the MEC problem in the following way. See Fig. 1 for a partial illustration. We create the set of vertices V as follows.

1. For each triple $t_j \in T$, we add six vertices $\{t_j^X, t_j^Y, t_j^Z, t_j^{XY}, t_j^{XZ}, t_j^{YZ}\}$.
2. For each element $x_i \in X$ (resp. $y_i \in Y$ and $z_i \in Z$), we add a corresponding vertex x_i (resp. y_i and z_i).

We have that $|V| = 6 \cdot |T| + |X| + |Y| + |Z| = 6N + 3n$. Let us now define the coloring $\sigma : V \to C$ of the vertices using the set of colors $C = \{1, 2, 3\}$.

1. For any $1 \leq i \leq n$ and $1 \leq j \leq N$, $\sigma(x_i) = \sigma(t_j^{XY}) = \sigma(t_j^Z) = 1$.
2. For any $1 \leq i \leq n$ and $1 \leq j \leq N$, $\sigma(y_i) = \sigma(t_j^{YZ}) = \sigma(t_j^X) = 2$.
3. For any $1 \leq i \leq n$ and $1 \leq j \leq N$, $\sigma(z_i) = \sigma(t_j^{XZ}) = \sigma(t_j^Y) = 3$.

Finally, let us define the collection of edges E.

1. For each $1 \leq j \leq N$, each of $\{t_j^X, t_j^{XY}, t_j^{XZ}\}, \{t_j^Y, t_j^{XY}, t_j^{YZ}\}, \{t_j^Z, t_j^{XZ}, t_j^{YZ}\}$ forms a clique of size three.
2. For each $1 \leq i \leq n$ and $1 \leq j \leq N$, if x_i (resp. y_i and z_i) appears in t_j, connect x_i (resp. y_i and z_i) to t_j^X (resp. t_j^Y and t_j^Z).

Analysis. Informally, we show that an instance of Max 3-DM-3 where all the vertices $X \cup Y \cup Z$ can be covered by a feasible collection of triples T' corresponds to an instance of MEC with a large optimal value, i.e., the graph can be partitioned into colorful components inducing a large transitive closure. On the other hand, we show that an instance of Max 3-DM-3 where no more than $(1 - \varepsilon)$ fraction of the vertices $X \cup Y \cup Z$ can be covered by any feasible set of triples corresponds to an instance of MEC with a much smaller optimal value. We now analyze both cases.
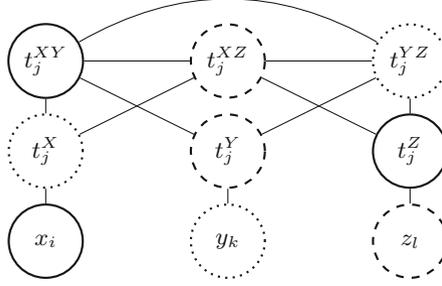
Fig. 1. A subgraph corresponding to a triple $t_j = (x_i, y_k, z_l)$. Colors of the vertices are denoted using the line styles: solid, dotted and dashed lines respectively corresponds to colors 1, 2 and 3.

Lemma 1. *Let (X, Y, Z, T) be an instance of Max 3-DM-3 where $OPT_{3DM} = n$, i.e., where all the vertices of $X \cup Y \cup Z$ can be covered by a feasible collection of triples. Then for the corresponding instance of MEC, we have $OPT_{MEC} \geq 6N + 3n$.*

Proof. The colorful components of the MEC instance are constructed as follows. For each triple $t_j \in T'$ (there are n of them), we add three colorful components, each component consisting of three vertices. Given a triple $t_j = (x_i, y_k, z_l)$, the colorful components are $\{x_i, t_j^X, t_j^{XZ}\}$, $\{y_k, t_j^Y, t_j^{XY}\}$ and $\{z_l, t_j^Z, t_j^{YZ}\}$. For each triple $t_{j'} \in T \setminus T'$ (there are $N - n$ of them), we create two colorful components, each consisting of three vertices: $\{t_{j'}^X, t_{j'}^{XZ}, t_{j'}^Z\}$ and $\{t_{j'}^{XY}, t_{j'}^Y, t_{j'}^{YZ}\}$. See Fig. 2 for an illustration.

As T' is a feasible collection of triples, that is a set of triples such that no two elements agree on any coordinate, we obtain a feasible partition of the graph into colorful components. Clearly, the total number of edges in the transitive closure equals $9n + 6(N - n) = 6N + 3n$, since each of the n triples in T' induces three colorful components of size three and each of the $N - n$ other triples induces two colorful components of size three. □

Lemma 2. *Let (X, Y, Z, T) be an instance of Max 3-DM-3 where $OPT_{3DM} < (1 - \varepsilon)n$, i.e., where every feasible collection of triples covers less than a $(1 - \varepsilon)$ w of vertices $X \cup Y \cup Z$. Then, for the corresponding instance of MEC, we have $OPT_{MEC} < 6N + 3n(1 - \varepsilon/2)$.*

Proof. Let $(G = (V, E), \sigma)$ be the instance of the MEC problem corresponding to an instance of Max 3-DM-3 as defined in the lemma statement. For any triple $t_j = (x_i, y_k, z_l) \in T$, let G_{t_j} be a subgraph of G induced by the following set of vertices $\{x_i, y_k, z_l, t_j^X, t_j^Y, t_j^Z, t_j^{XY}, t_j^{XZ}, t_j^{YZ}\}$, as shown in Fig. 1.

Let us fix an optimal solution S for the MEC problem for (G, σ). This solution defines a partition Γ of G into colorful components. First, notice that each colorful component is contained within some subgraph G_t. Indeed, by construction, the only vertices which belong to multiple subgraphs G_{t_j} are the vertices
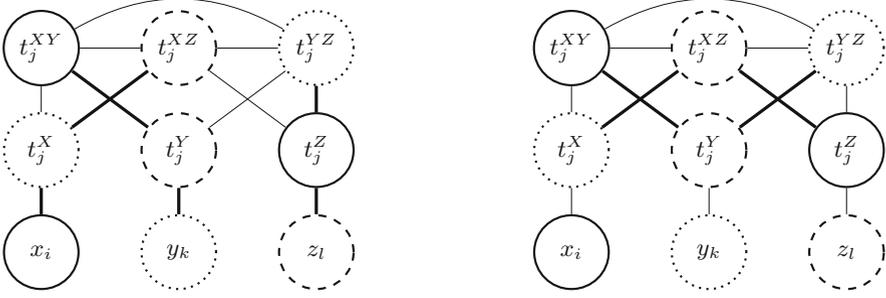
Fig. 2. Colorful components (defined by bold edges) for a triple from T' (left) and a triple from $T \setminus T'$ (right).

$\{x_i, y_i, z_i \mid 1 \leq i \leq n\}$. Moreover, for $1 \leq i \leq n$, each vertex x_i (resp. y_i and z_i) has only neighbours of color 2 (resp. 3 and 1) in G, and therefore is a leaf in its colorful component. Finally, there are no edges between vertices (excluding $\{x_i, y_i, z_i \mid 1 \leq i \leq n\}$) belonging to different subgraphs G_{t_j} and $G_{t_{j'}}$, with $1 \leq j, j' \leq N$.

We will now partition G into vertex-disjoint subgraphs G'_t for $t \in T$, such that G'_t is a subgraph of G_t (possibly G_t itself), and each colorful component from our fixed partition Γ is contained within a single graph G'_t. We proceed as follows. Each vertex from $\{t_j^X, t_j^Y, t_j^Z, t_j^{XY}, t_j^{XZ}, t_j^{YZ}\}$ belongs to a single subgraph G_{t_j}, therefore it belongs also to the subgraph G'_{t_j}. We assign each vertex x_i (resp. y_k and z_l) to one subgraph G'_{t_j}, in such a way that we do not split any colorful component. That means that if x_i (resp. y_k and z_l) is in the same colorful component as t_j^X (resp. t_j^Y and t_j^Z), we assign it to G'_{t_j}. If the colorful component containing x_i (resp. y_k and z_l) is a singleton, we assign the vertex arbitrarily.

Now each subgraph G'_{t_j}, with $t_j = (x_i, y_k, z_l)$, contains 6, 7, 8 or 9 vertices ($\{t_j^X, t_j^Y, t_j^Z, t_j^{XY}, t_j^{XZ}, t_j^{YZ}\}$ and a possibly empty subset of $\{x_i, y_k, z_l\}$). Let us denote by g_6, g_7, g_8 and g_9 the number of subgraphs G'_t containing 6, 7, 8 and 9 vertices, respectively. We have $g_6 + g_7 + g_8 + g_9 = N$. Moreover, $g_7 + 2 \cdot g_8 + 3 \cdot g_9 = 3n$, since each vertex x_i, y_k and z_l belongs to exactly one subgraph G'_t.

As we observed earlier, all colorful components from the partition Γ are contained within the above-defined subgraphs G'_t. Therefore the solution S to the MEC problem on (G, σ) is a union of solutions S_t for the MEC problem for the graphs $(G'_t, \sigma|_{G'_t})$. Now observe the following.

1. The value of an optimal MEC solution for $(G'_t, \sigma|_{G'_t})$ for a subgraph G'_t on 6 vertices is at most 6 (value 6 is achieved when the solution consists of two colorful components of 3 vertices each).
2. The value of an optimal MEC solution for $(G'_t, \sigma|_{G'_t})$ for a subgraph G'_t on 7 vertices is at most 6 (value 6 is achieved when the solution consists of two colorful components of 3 vertices each, and one singleton).

3. The value of an optimal MEC solution for $(G'_t, \sigma|_{G'_t})$ for a subgraph G'_t on 8 vertices is at most 7 (value 7 is achieved when the solution consists of two colorful components of 3 vertices and one component of 2 vertices).
4. The value of an optimal MEC solution for $(G'_t, \sigma|_{G'_t})$ for a subgraph G'_t on 9 vertices is at most 9 (value 9 is achieved when the solution consists of three colorful components of 3 vertices each).

Now, assume toward a contradiction that $\mathrm{OPT}_{\mathrm{MEC}} \geq 6N + 3n(1 - \varepsilon/2)$, i.e., the value of S is at least $6N + 3n(1 - \varepsilon/2)$. We just mentioned that each subgraph G'_t (there are N of them in total) of size 6 (7,8,9) contributes at most a value of 6 (6,7,9, respectively) towards the value of S. Thus, we must have

$$g_8 + 3 \cdot g_9 \geq 3n(1 - \varepsilon/2). \tag{1}$$

We already know that
$$2 \cdot g_8 + 3 \cdot g_9 \leq 3n. \tag{2}$$

Multiplying inequality (1) by 2 and inequality (2) by -1 and adding them yields the inequality $g_9 \geq n(1 - \varepsilon)$.

Notice that a subgraph G'_t on 9 vertices corresponds to a triple from T. Moreover, any two such triples are disjoint. As $g_9 \geq n(1-\varepsilon)$, in the corresponding Max 3-DM-3 instance, we can cover at least a $(1-\varepsilon)$ fraction of elements $X \cup Y \cup Z$ by disjoint triples, which contradicts the lemma statement. Therefore, we must have $\mathrm{OPT}_{\mathrm{MEC}} < 6N + 3n(1 - \varepsilon/2)$. □

Let us now derive the APX-hardness of MEC for instances where vertices are colored using 3 colors from the above lemmas.

Theorem 2. *The Maximum Edges in Transitive Closure problem is APX-hard, even for $|C| = 3$.*

Proof. From Theorem 1, we know that it is NP-hard to distinguish between instances of Max 3-DM-3 such that $\mathrm{OPT}_{\mathrm{3DM}} = n$ or $\mathrm{OPT}_{\mathrm{3DM}} < (1 - \varepsilon)n$. From Lemmas 1 and 2, it is NP-hard to distinguish between instances of MEC for which $\mathrm{OPT}_{\mathrm{MEC}} \geq 6N + 3n$ and instances for which $\mathrm{OPT}_{\mathrm{MEC}} < 6N + 3n(1 - \varepsilon/2)$.

Since $N \leq 3n$, we get that it is NP-hard to approximate MEC within some constant factor, i.e., MEC is APX-hard. As all MEC instances considered here use only 3 colors, MEC is APX-hard already for $|C| = 3$. □

3 Approximation of MEC for an Unbounded Number of Colors

3.1 A Positive Result

In this section we show that the MEC problem for an unbounded number of colors admits approximation within a factor of $\sqrt{2 \cdot \mathrm{OPT}}$. The algorithm is an

exact polynomial time algorithm for the Minimum Singleton Vertices (MSV) problem from [1]. Let us first restate the definition of the MSV problem.

MINIMUM SINGLETON VERTICES: Given a simple, undirected graph $G = (V, E)$ and a coloring $\sigma : V \to C$ of the vertices, remove a collection of edges $E' \subseteq E$ from the graph such that each connected component in $G' = (V, E \setminus E')$ is colorful and the number of isolated vertices is minimum.

Theorem 3 ([1]). *The MSV problem can be solved exactly in polynomial time.*

We are now ready to prove our result.

Theorem 4. *The MEC problem admits a polynomial-time $\sqrt{2 \cdot OPT}$ approximation algorithm.*

Proof. We show that the exact MSV algorithm is a $\sqrt{2 \cdot \text{OPT}}$-approximation algorithm for MEC. Let $G = (V, E)$ be the input graph and let OPT be the value of an optimal solution (i.e., the number of edges in the transitive closure) of the MEC problem on G.

Let G_{MSV} be the graph obtained by running the exact MSV algorithm on G. Clearly, as each connected component of G_{MSV} is colorful, G_{MSV} is a feasible solution for the MEC problem.

Let I_{MEC} be the number of isolated vertices in an optimal solution of the MEC problem, and let I_{MSV} be the number of isolated vertices in G_{MSV}. We have $I_{\text{MSV}} \leq I_{\text{MEC}}$.

We have $\text{OPT} \leq \binom{|V| - I_{\text{MEC}}}{2}$, since the largest possible value of OPT is achieved when all the vertices that are not isolated are in the same connected component.

Define Val_{MSV} to be the number of edges in the transitive closure of G_{MSV}. We get that $\text{Val}_{\text{MSV}} \geq (|V| - I_{\text{MSV}})/2$. Thus, we have

$$\frac{\text{OPT}}{\text{Val}_{\text{MSV}}} \leq \frac{\sqrt{\text{OPT}} \cdot \frac{1}{\sqrt{2}}(|V| - I_{\text{MEC}})}{\frac{1}{2}(|V| - I_{\text{MSV}})} \leq \sqrt{2 \cdot \text{OPT}} \ ,$$

as $|V| - I_{\text{MEC}} \leq |V| - I_{\text{MSV}}$.   □

3.2   A Negative Result

In this section, we show that the MEC problem is NP-hard to approximate within a factor of $|V|^{1/3-\varepsilon}$ for any constant $\varepsilon > 0$. This result holds even if the input graph is a tree and each color appears at most twice in the graph. We use the same reduction as Rizzi and Sikora for proving hardness of approximation of the Graph Motif problem [7].

Reduction. We make a reduction from the Maximum Independent Set problem (MIS). Let $G = (V, E)$ be a MIS instance, and let $n = |V|$. We create an instance $G' = (V', E')$ of MEC in the following way. See Fig. 3 for an illustration.

The set of vertices V' consists of the following vertices:
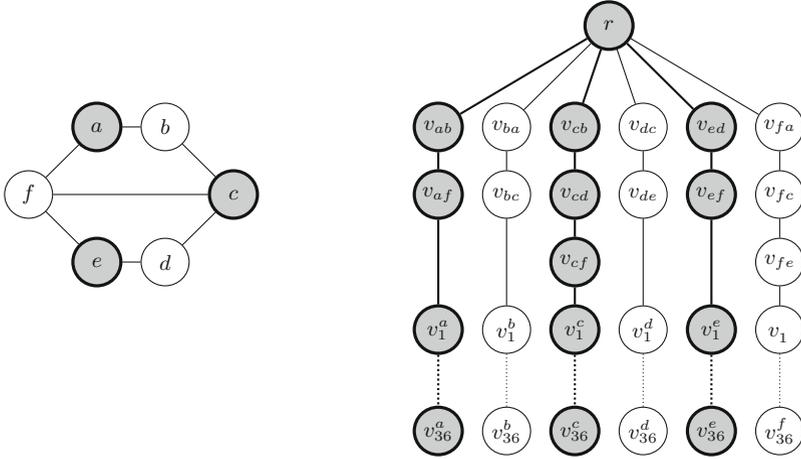
Fig. 3. Reduction from an instance G of MIS (left) to an instance G' of MEC (right). The only pairs of vertices in G' sharing the same color are vertices v_{uw} and v_{wu} for $u, w \in \{a, b, c, d, e, f\}$. An independent set in G corresponds to a colorful component in G' containing the root vertex r (gray vertices).

1. a special vertex r colored with a unique color c_r,
2. for each edge $uw \in E$, vertices v_{uw} and v_{wu} colored with the same color c_{uw},
3. for each vertex $u \in V$, a collection of n^2 vertices $v_1^u, v_2^u, \ldots, v_{n^2}^u$ colored with unique colors $c_1^u, c_2^u, \ldots, c_{n^2}^u$.

The resulting graph G' will be a tree on the set of vertices V', rooted at r. For each vertex $u \in V$, we add to G' a path starting at r which visits all vertices v_{uw} (in an arbitrary order), and then all vertices $v_1^u, v_2^u, \ldots, v_{n^2}^u$.

Analysis

Lemma 3. *If $G = (V, E)$ has an independent set of size α, then there is a solution for the corresponding instance G' of the MEC problem with value at least $\binom{\alpha n^2}{2}$.*

Proof. Let $V_I \subseteq V$ be an independent set in G consisting of α vertices. We will show that there is a colorful component in G' consisting of at least $\alpha \cdot n^2$ vertices.

We construct the set V_C' (see Fig. 3) in the following way. It consists of the root vertex r, together with all vertices lying on the paths corresponding to the vertices $u \in V_I$ (i.e., the vertices v_{uw} where $w \in V$ and $uw \in E$, and the vertices v_i^u for $i = 1, \ldots, n^2$). The subgraph of G' induced by V_C' is connected and consists of at least $\alpha \cdot n^2$ vertices. The subgraph is colorful, as from the construction of G' if two vertices lying on two paths of G' have the same color, then the vertices of G corresponding to these paths are connected by an edge, and therefore they cannot belong to an independent set.

The decomposition of G' into a component induced by V'_C and singletons is a feasible partition into colorful components, and its transitive closure has at least $\binom{\alpha n^2}{2}$ edges.   □

Lemma 4. *If there is a solution for the instance G' of the MEC problem of value at least $n^5/2 + \alpha^2 \cdot n^4$, then G has an independent set of size at least α.*

Proof. First, notice that any colorful component which does not contain the root vertex r consists of less than $n^2 + n$ vertices. Now, consider the colorful component V'_C containing r. Let $V_I \subseteq V$ be the subset of vertices u of G for which $v_1^u \in V'_C$. From the construction of the graph G', V_I is an independent set in G. We will, now, show a lower bound on $|V_I|$.

For any vertex $u \in V$, if $u \notin V_I$, then V'_C contains at most n vertices from the path of G' corresponding to u. If $u \in V_I$, then V'_C contains at most $n + n^2$ vertices from this path. We get that $|V'_C| \le n^2 + |V_I|n^2$. As $|V'| < n^3 + n^2$, we have that $OPT_{MEC} < (n^3 + n^2) \cdot (n^2 + n)/2 + |V_I|^2 \cdot n^4/2$. We get that $|V_I| \ge \alpha$.   □

Theorem 5. *It is NP-hard to approximate the MEC problem within a factor of $|V|^{1/3-\varepsilon}$.*

Proof. The MIS problem is NP-hard to approximate within a factor of $n^{1-\varepsilon}$ for any constant $\varepsilon > 0$ [11]. In particular, it is NP-hard to distinguish whether a graph G has a maximum independent set of size at most n^ε, or at least $n^{1-\varepsilon}$.

In the first case, by Lemma 4, $OPT_{MEC} \le n^5/2 + n^{4+2\varepsilon}$. In the second case, by Lemma 3, $OPT_{MEC} \ge n^{6-2\varepsilon}/2$. As the number of vertices of G' is in $\Theta(n^3)$, we get that approximating MEC within a factor of $|V|^{1/3-\varepsilon}$ is NP-hard.   □

4   Conclusions and Future Work

In this paper we show several approximation and hardness results for the Maximum Edges in Transitive Closure Problem. First we prove that the problem is NP-hard to approximate within a factor of $|V|^{1/3-\varepsilon}$, for any constant $\varepsilon > 0$. Additionally, we show that the problem is APX-hard already for the case when the number of vertex colors equals 3. We complement these results by showing the first approximation algorithm for the problem, with approximation factor $\sqrt{2 \cdot OPT}$.

There are several directions for future work. First, it would be interesting to close the gap between the approximation upper and lower bounds by showing an approximation algorithm with a better ratio or improving the hardness result. Another way to extend the current set of results would be to also consider the problem of maximizing the number of edges in the connected components and to identify and highlight similarities and differences between the two variants. Maximising the number of edges in the components has the same complexity as problem which asks to delete the minimum number of edges (it is the dual problem which is known to be NP-hard). Nevertheless, from the approximation point of view, these two problems are different.

Another direction would be to consider approximation guarantees that take the number of colors into account. However, we believe that there is not much room for improvement in this direction. The presented approximation algorithm has a ratio of $(C-1)/4$, where C is the number of colors and, due the hardness results, we cannot hope for a much better approximation.

References

1. Adamaszek, A., Popa, A.: Algorithmic and hardness results for the colorful components problems. In: Pardo, A., Viola, A. (eds.) LATIN 2014. LNCS, vol. 8392, pp. 683–694. Springer, Heidelberg (2014)
2. Avidor, A., Langberg, M.: The multi-multiway cut problem. Theoret. Comput. Sci. **377**(1–3), 35–42 (2007)
3. Bruckner, S., Hüffner, F., Komusiewicz, C., Niedermeier, R.: Evaluation of ILP-based approaches for partitioning into colorful components. In: Demetrescu, C., Marchetti-Spaccamela, A., Bonifaci, V. (eds.) SEA 2013. LNCS, vol. 7933, pp. 176–187. Springer, Heidelberg (2013)
4. Bruckner, S., Hüffner, F., Komusiewicz, C., Niedermeier, R., Thiel, S., Uhlmann, J.: Partitioning into colorful components by minimum edge deletions. In: Kärkkäinen, J., Stoye, J. (eds.) CPM 2012. LNCS, vol. 7354, pp. 56–69. Springer, Heidelberg (2012)
5. He, G., Liu, J., Zhao, C.: Approximation algorithms for some graph partitioning problems. J. Graph Algorithms Appl. **4**(2), 1–11 (2000)
6. Petrank, E.: The hardness of approximation: gap location. Comput. Complex. **4**(2), 133–157 (1994)
7. Rizzi, R., Sikora, F.: Some results on more flexible versions of graph motif. In: Hirsch, E.A., Karhumäki, J., Lepistö, A., Prilutskii, M. (eds.) CSR 2012. LNCS, vol. 7353, pp. 278–289. Springer, Heidelberg (2012)
8. Sankoff, D.: OMG! orthologs for multiple genomes - competing formulations. In: Chen, J., Wang, J., Zelikovsky, A. (eds.) ISBRA 2011. LNCS, vol. 6674, pp. 2–3. Springer, Heidelberg (2011)
9. Savard, O.T., Swenson, K.M.: A graph-theoretic approach for inparalog detection. BMC Bioinform. **13**(S–19), S16 (2012)
10. Zheng, C., Swenson, K., Lyons, E., Sankoff, D.: OMG! orthologs in multiple genomes – competing graph-theoretical formulations. In: Przytycka, T.M., Sagot, M.-F. (eds.) WABI 2011. LNCS, vol. 6833, pp. 364–375. Springer, Heidelberg (2011)
11. Zuckerman, D.: Linear degree extractors and the inapproximability of max clique and chromatic number. Theory Comput. **3**(1), 103–128 (2007)