



**HAL**  
open science

# On the complexity of finding $k$ shortest dissimilar paths in a graph

Ali Al Zoobi, David Coudert, Nicolas Nisse

► **To cite this version:**

Ali Al Zoobi, David Coudert, Nicolas Nisse. On the complexity of finding  $k$  shortest dissimilar paths in a graph. [Research Report] Inria; CNRS; I3S; Université Côte d'Azur. 2021, pp.9. hal-03187276v2

**HAL Id: hal-03187276**

**<https://hal.science/hal-03187276v2>**

Submitted on 29 Sep 2021

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# On the complexity of finding $k$ shortest dissimilar paths in a graph\*

Ali Al-Zoobi<sup>1</sup>, David Coudert<sup>1</sup>, and Nicolas Nisse<sup>1</sup>

<sup>1</sup>Université Côte d’Azur, Inria, CNRS, I3S, France

September 29, 2021

## Abstract

The similarity between two paths can be measured according to the proportion of arcs they share. We study the complexity of several variants of the problem of computing “dissimilar” paths (whose measure of similarity does not exceed a certain threshold) between two given vertices of a weighted directed graph. For four of the most studied measures in the literature, we give a unified and simple proof of the fact that finding  $k$  shortest dissimilar paths is NP-COMplete.

We then consider the problem of finding an alternative to one or more given paths. We show that finding a path that is dissimilar to another given path can be done in polynomial time for one of the four considered measures while it is NP-COMplete for the three remaining measures. In addition, we show that if  $k = 2$  paths are given, finding a new path that is dissimilar to the given ones is NP-COMplete even on DAGs for the four considered measures. Moreover, for the four considered measures, we show that if a path  $P$  is given, finding a shortest path among those that are dissimilar to  $P$  is NP-COMplete in DAGs.

## 1 Introduction

The  $k$  shortest simple paths problem aims at finding a shortest path, a second shortest path, *etc.*, a  $k^{th}$  shortest simple paths between a pair of source and destination node in a digraph. This problem has numerous applications in various kinds of networks (road and transportation networks, communications networks, social networks, *etc.*) and is also used as a building block for solving many optimization problems. Let  $D = (V, A)$  be a digraph, an

---

\*Funding: This work was supported by the UCA<sup>JEDI</sup> Investments in the Future project managed by the National Research Agency (ANR-15-IDEX-01); the project MULTIMOD (ANR-17-CE22-0016); the project Digraphs (ANR-19-CE48-0013); the Région Sud PACA.

$s$ - $t$  path is a sequence  $(s = v_0, v_1, \dots, v_l = t)$  of vertices starting with  $s$  and ending with  $t$ , such that  $(v_i, v_{i+1}) \in A$  for all  $0 \leq i < l$ . A path  $P$  is called *simple* if all of its vertices are distinct, i.e, for every  $i \neq j$ ,  $v_i \neq v_j$ .

Let  $\ell : A \rightarrow \mathbb{R}^+$  be a length function over the arcs. For any path  $P$ , its *length*  $\ell(P) = \sum_{e \in A(P)} \ell(e)$  is the sum of the lengths of its arcs.

However, the  $k$  shortest simple paths are often quite “similar”. Roughly, they often share a “large” proportion of their arcs. This is undesirable in many applications. For instance, in transportation networks, users may expect to have several options offering more diversity: a user prefers a shortest paths, another user wants to avoid a traffic jam, a third one prefers to travel along the coast *etc.*

To deal with this issue, the problem of computing “dissimilar” (shortest) paths has been investigated. Several definitions of the similarity between two paths (including the Jaccard and the Max measures defined below [4]) were first proposed by Erkut and Verter [5], motivated by the transportation of hazardous materials where it is recommended to avoid residential areas and crowded routes.

Akgün *et al.* [2] proposed and analyzed a first basic solution, consisting in computing a huge set of shortest paths and then choosing a subset of these paths that are mutually dissimilar. In their experiments, this method scaled only on small transportation networks (about 300 vertices). The first scalable solutions were proposed by Abraham *et al.* [1] where a shortest path  $P$  is fixed, and “locally shortest” paths with limited intersection with  $P$  are requested (this corresponds to the Asymmetric measure defined below). However, except for the initial path  $P$ , this definition does not guaranty any mutual dissimilarity between the computed paths. A noticeable heuristic proposed in [1] is the penalty based approach. This heuristic adds a penalty on the arcs of the already chosen paths in order to limit the chances of falling back on the same paths.

Chondrogiannis *et al.* [3] offer both theoretical and empirical study of the problem. They formally proved that finding  $k$  shortest dissimilar paths is weakly NP-COMplete for both the Asy measure and a new dissimilarity measure that they define (referred to as Min measure below). For these two measures, they proposed an exact pseudo-polynomial time algorithm, with several pruning techniques, that allows to find 4 dissimilar paths in a road network with 3,000 vertices in less than one second. They also proposed advanced heuristics enabling to scale on a road network with one million vertices while computing paths that are close to shortest ones in practice.

In this paper, we further study the computational complexity of computing (shortest) dissimilar paths for four of the main measures studied so far. More formally, let  $P, P'$  be two  $s$ - $t$  simple paths in  $D$  and let  $X = \sum_{e \in A(P) \cap A(P')} \ell(e)$ , i.e., the total length of the intersection of  $P$  and  $P'$ . The four considered measures are defined as follows.

Name ( $Z$ )	$S_Z(P, P')$
Jaccard [5]	$\frac{X}{\ell(P \cup P')}$
Asy [1]	$\frac{X}{\ell(P)}$
Min [4]	$\frac{X}{\min\{\ell(P), \ell(P')\}}$
Max [5]	$\frac{X}{\max\{\ell(P), \ell(P')\}}$

Table 1: Four similarity measures

Let  $\mathcal{S} = \{Asy, Jaccard, Min, Max\}$ . Given one of the measures  $Z \in \mathcal{S}$  and a threshold value  $0 \leq \theta \leq 1$ , two paths  $P$  and  $P'$  are said  $\theta$ -dissimilar (or  $P'$  is said  $\theta$ -dissimilar to  $P$  in the case of Asymmetric similarity) for the measure  $Z$  if  $S_Z(P, P') \leq \theta$ .

**Our contributions.** In Section 2, we study the problem of finding  $k$  shortest pairwise dissimilar paths. We give a unified and simple proof of the NP-completeness of this problem for each of the four similarity measures defined above. Then, in Section 3, we study the problem of finding a (shortest) path that is dissimilar to a given set of paths. In particular, we show that if only one path  $P$  is initially given, computing a second path that is dissimilar to  $P$  for the Asy measure can be done in polynomial time while it is NP-COMplete for the remaining measures (Min, Max and Jaccard). Then, we prove that finding a path dissimilar (for each of the considered four measures) to a given set of  $k \geq 2$  paths is NP-COMplete on DAGs. Finally, for each of these four measures, we show that computing a shortest path among those dissimilar to a given path is NP-COMplete on DAGs.

## 2 Finding $k$ shortest dissimilar paths

In this section, we show that the problem of finding  $k$  shortest dissimilar paths is NP-COMplete for all the considered similarity measures.

More formally, given a digraph  $D = (V, A)$  with length function  $\ell : A \rightarrow \mathbb{R}^+$ , a pair of vertices  $(s, t) \in V \times V$ , an integer  $k \geq 2$ , a threshold value  $0 \leq \theta \leq 1$ ,  $k$  constants  $L_1, L_2, \dots, L_k$  and a similarity measure  $Z \in \mathcal{S}$ , the problem  $k$ -SHORTESTDISS( $Z$ ) of finding  $k$  shortest dissimilar paths asks to decide whether there exists  $k$  paths from  $s$  to  $t$  that are mutually  $\theta$ -dissimilar with respect to  $Z$  and such that  $\ell(P_i) \leq L_i$  for  $1 \leq i \leq k$ .

Note that, for the extreme case where  $\theta = 0$ , the problem of finding  $k$  dissimilar paths (not necessarily the shortest) is the problem of finding  $k$  arc-disjoint paths, and it can be solved in polynomial time using a min cost

flow algorithm.

Finding  $k$  shortest dissimilar paths has already been proved NP-COMplete for the Asy and Min measures [3]. Here we propose a simple and unified proof (for all considered measures).

**Theorem 1.** *For every  $k \geq 2$  and  $Z \in \mathcal{S}$ , the  $k$ -SHORTESTDISS( $Z$ ) problem is NP-COMplete in the class of DAGs with a single source and a single sink.*

*Proof.* Let us first consider the case  $k = 2$ .

For every  $Z \in \mathcal{S}$ , the problem is clearly in NP. We prove the NP-hardness by a reduction from the MIN-MINDP problem. Given a graph  $G = (V, E)$  with length function  $\ell : E \rightarrow \mathbb{R}^+$ , two terminals  $s, t \in V$  and a real number  $\delta \in \mathbb{R}^+$  as inputs, the MIN-MINDP problem asks whether there exists two edge disjoint paths  $P$  and  $P'$  with  $\ell(P) \leq \delta$ . This problem is NP-COMplete [7].

Let  $I = (D = (V, A), \ell, s, t, \delta)$  be an instance of the MIN-MINDP problem and let  $I' = (D, s, t, k = 2, \theta = 0, \ell, L_1 = \delta, L_2 = n \cdot \max_{e \in A} \ell(e))$  be an instance of the  $k$ -SHORTESTDISS( $Z$ ) problem.

- If  $I$  is a positive MIN-MINDP instance, it means that there are two arc disjoint  $s$ - $t$  paths  $P$  and  $P'$  such that  $\ell(P) \leq \delta$ . Let  $P_1 = P$  and  $P_2 = P'$ . For every similarity measure  $Z \in \mathcal{S}$ , we have  $S_Z(P_1, P_2) = 0$  since  $A(P_1) \cap A(P_2) = \emptyset$ , and so  $\sum_{e \in A(P_1) \cap A(P_2)} \ell(e) = 0$ . In addition,  $\ell(P_1) \leq \delta = L_1$  and  $\ell(P_2) \leq L_2$ . So,  $I'$  is a positive  $k$ -SHORTESTDISS( $Z$ ) instance
- If  $I'$  is a positive  $k$ -SHORTESTDISS( $Z$ ) instance, it means that there are two  $s$ - $t$  paths  $P_1$  and  $P_2$  such that  $\ell(P_1) \leq L_1$  and  $S_Z(P_1, P_2) = 0$  for every similarity measure  $Z \in \mathcal{S}$ . In another word,  $P_1$  and  $P_2$  are arc-disjoint. Let  $P = P_1$  and  $P' = P_2$ .  $P$  and  $P'$  are two arc-disjoint  $s$ - $t$  paths. In addition  $\ell(P) \leq L_1 = \delta$ , so  $I$  is a positive MIN-MINDP instance.

We conclude that the 2-SHORTESTDISS( $Z$ ) problem is NP-HARD.

To extend the result to any  $k \geq 2$ , it is sufficient to add, to the digraph  $D$  in  $I'$ ,  $k - 2$  arc-disjoint  $s$ - $t$  paths  $P_3, \dots, P_k$ , each with length  $L_2$  and to set  $L_2 = L_i$  for all  $2 \leq i \leq k$ .  $\square$

### 3 Finding a path dissimilar to several given paths

In this section, we present our main results. First, we show that the problem of finding a path dissimilar to another given path can be solved in polynomial time for the Asy measure. Then, we prove that the problem of finding a path dissimilar to two given paths is NP-COMplete. Finally, we show that finding a shortest path dissimilar to one given path is also NP-COMplete.

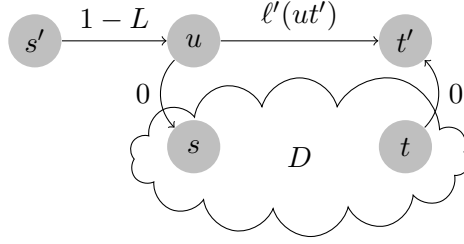


Figure 1: Digraph  $D'$  defined from  $D$  (Theorem 2) with  $\ell'(ut') = 1$  for the Min measure and  $\ell'(ut') = 0$  for the Max and Jaccard measures.

### 3.1 Finding a path dissimilar to another given path

First, let us start with the easiest variant of the problem that is the problem of finding a path dissimilar to another for the Asy measure. Given a digraph  $D = (V, A)$  with  $\ell : A \rightarrow \mathbb{R}^+$ , two vertices  $s, t \in V$ , a threshold value  $0 \leq \theta \leq 1$ , a  $s$ - $t$  simple path  $P$  and a similarity measure  $Z \in \mathcal{S}$ ,  $\text{DISS}(1, Z)$  is the problem of finding a  $s$ - $t$  path  $Q$  that is  $\theta$ -dissimilar to  $P$  using the measure  $Z$ .

**Proposition 1.**  $\text{DISS}(1, \text{Asy})$  can be solved in same time as any shortest-path algorithm.

*Proof.* Let  $\ell' : A \rightarrow \mathbb{R}^+$  be defined such that, for every  $e \in A$ ,  $\ell'(e) = \ell(e)$  if  $e \in A(P)$ , and  $\ell'(e) = 0$  otherwise. Hence, a shortest  $s$ - $t$  simple path  $Q$  is a solution of the  $\text{DISS}(1, \text{Asy})$  problem if and only if  $\ell'(Q) = \sum_{e \in A(P) \cap A(Q)} \ell(e) \leq \theta \cdot \ell(P)$ .  $\square$

**Theorem 2.**  $\text{DISS}(1, Z)$  is NP-COMplete for all  $Z \in \{\text{Jaccard}, \text{Min}, \text{Max}\}$

*Proof.* For every  $Z \in \{\text{Jaccard}, \text{Min}, \text{Max}\}$ , the problem is clearly in NP, so we only prove the NP-hardness by a reduction from the LONG-PATH problem. Given a digraph  $D = (V, A)$  with length function  $\ell : A \rightarrow \mathbb{R}^+$ , two terminals  $s, t \in V$  and a real number  $L \in \mathbb{R}^+$  as inputs, the LONG-PATH problem asks whether there exists an  $s$ - $t$  simple path  $Q$  with  $\ell(Q) \geq L$ . This problem is NP-COMplete [8]. Moreover, it remains NP-COMplete when  $L \leq 1$  (by dividing the length of each arc by  $M = \sum_{a \in A} w(a)$ ).

Let  $I = (D = (V, A), \ell, s, t, L)$  be a LONG-PATH instance with  $0 < L \leq 1$ . Let  $D' = (V \cup \{s', u, t'\}, A \cup \{s'u, ut', us, tt'\})$  with lengths  $\ell'(a) = \ell(a)$  for every  $a \in A$ ,  $\ell'(s'u) = 1 - L$ , the value of  $\ell'(u, t')$  depends on the considered measure and will be specified later,  $\ell'(us) = 0$  and  $\ell'(tt') = 0$  (see Figure 1). Let also  $I' = (D', \ell', P, \theta = 1 - L)$ , with  $P = \{s', u, t'\}$ , be an instance of the  $\text{DISS}(1, Z)$  problem.

First, let us consider the *Min* measure and let  $\ell'(ut') = L$ .

- If  $I$  is a positive LONG-PATH instance, then there is an  $s$ - $t$  path  $R = (s, \dots, t)$  of length at least  $L$  in  $D$ . Let  $Q = (s', u, s, \dots, t, t')$  be the concatenation of  $s', u, R$  and  $t'$ . Note that  $\ell'(P) = 1$  and  $\ell'(Q) \geq 1$ . We have  $S_{Min}(P, Q) = \frac{\ell'(s'u)}{\min\{\ell'(P), \ell'(Q)\}} = \frac{1-L}{1} = \theta$ , and so  $I'$  is a positive DISS(1, *Min*) instance.
- If  $I'$  is a positive DISS(1, *Min*) instance, then there is an  $s$ - $t$  path  $Q$  s.t.  $S_{Min}(P, Q) = \frac{\ell'(P \cap Q)}{\min\{\ell'(P), \ell'(Q)\}} \leq \theta$ . Since  $\ell'(P \cap Q) = \ell'(s'u) = 1 - L = \theta$ , we have  $\min\{\ell'(P), \ell'(Q)\} \geq 1$  and so  $\ell'(Q) \geq 1$  (since  $\ell'(P) = 1$ ). Let  $R$  be the subpath of  $Q$  starting from  $s$  and ending at  $t$ , i.e.,  $R = (s, \dots, t)$  and  $\ell(R) = \ell'(R) = \ell'(Q) - (1 - L) \geq L$  (since  $R$  is a simple path and  $\ell'(Q) \geq 1$ ). Therefore,  $I$  is a positive LONG-PATH instance.

We conclude that DISS(1, *Min*) is NP-Hard.

Using a similar construction, the NP-hardness of DISS(1, *Max*) and DISS(1, *Jaccard*) can be proved.

Precisely, for both the *Jaccard* and *Max* measure, it is sufficient to keep the same reduction as before but setting  $\ell'(ut') = 0$ .

- If  $I$  is a positive LONG-PATH instance, then there is an  $s$ - $t$  path  $R = (s, \dots, t)$  of length at least  $L$  in  $D$ . Let  $Q = (s', u, s, \dots, t, t')$  be the concatenation of  $s', u, R$  and  $t'$ . Note that  $\ell'(P) = 1 - L$  and  $\ell'(Q) \geq 1$ .

In the case of the *Max* measure, we get that  $S_{Max}(P, Q) = \frac{\ell'(P \cap Q)}{\max\{\ell'(P), \ell'(Q)\}} \leq \frac{1-L}{1} = \theta$ , and so  $I'$  is a positive DISS(1, *Max*) instance.

In the case of the *Jaccard* measure, we get that  $S_{Jaccard}(P, Q) = \frac{\ell(P \cap Q)}{\ell(P \cup Q)} = \frac{1-L}{1-L+\ell(R)} \geq 1 - L = \theta$  (since  $\ell(R) \geq L$ ), and so  $I'$  is a positive DISS(1, *Jaccard*) instance.

- If  $I'$  is a positive DISS(1, *Max*) instance, then there is an  $s$ - $t$  path  $Q$  s.t.  $S_{Max}(P, Q) = \frac{\ell'(P \cap Q)}{\max\{\ell'(P), \ell'(Q)\}} \leq \theta$ . Since  $\ell'(ut') = 0$ , we have that  $\ell'(P) = \ell'(su) = \ell'(P \cap Q) = 1 - L = \theta$  and  $\ell'(Q) \geq \ell'(P)$ . Since  $S_{Max}(P, Q) \leq \theta$ ,  $\ell'(Q) \geq 1$ . Let  $R$  be the subpath of  $Q$  starting from  $s$  and ending at  $t$ , i.e.,  $R = (s, \dots, t)$  and  $\ell(R) = \ell'(R) = \ell'(Q) - (1 - L) \geq L$  (since  $Q$  is a simple path). Therefore, there is a path from  $s$  to  $t$  in  $D$  of length at least  $L$  and  $I$  is a positive LONG-PATH instance.
- If  $I'$  is a positive DISS(1, *Jaccard*) instance, then there is an  $s$ - $t$  path  $Q$  s.t.  $S_{Jaccard}(P, Q) = \frac{\ell'(P \cap Q)}{\ell'(P \cup Q)} \leq \theta$ . By construction,  $\ell'(P \cup Q) = \ell'(Q)$ . Since, moreover  $\ell'(P \cap Q) = \ell'(s'u) = 1 - L = \theta$ , then  $\ell'(Q) \geq 1$ . Let  $R$  be the subpath of  $Q$  starting from  $s$  and ending at  $t$ , i.e.,  $R = (s, \dots, t)$  and  $\ell(R) = \ell'(R) = \ell'(Q) - \theta \geq 1 - (1 - L) = L$  (since  $Q$  is a simple path). Therefore, there is a path from  $s$  to  $t$  in  $D$  of length at least  $L$  and  $I$  is a positive LONG-PATH instance.

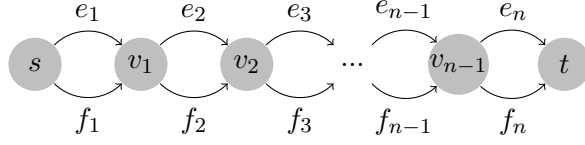


Figure 2: Digraph  $D_S = (V, A)$  defined from  $\mathcal{S} = \{x_1, \dots, x_n\}$ . For all  $1 \leq i \leq n$ ,  $\ell(e_i) = x_i$ . For all  $1 \leq i \leq n$ , we have  $\ell(f_i) = x_i$  in the proof of Theorem 3 and  $\ell(f_i) = M \cdot x_i$  with  $M > 1$  in the proof of Theorem 4.

We conclude that  $\text{DISS}(1, \text{Max})$  and  $\text{DISS}(1, \text{Jaccard})$  are NP-Hard.  $\square$

### 3.2 Finding a path dissimilar to several given paths

Given a digraph  $D = (V, A)$  with  $\ell : A \rightarrow \mathbb{R}^+$ , two vertices  $s, t \in V$ , a threshold value  $0 \leq \theta \leq 1$ ,  $k$   $s$ - $t$  simple paths  $P_1, \dots, P_k$  and a similarity measure  $Z \in \mathcal{S}$ ,  $\text{DISS}(k, Z)$  is the problem of finding a  $s$ - $t$  path  $Q$  that is  $\theta$ -dissimilar to  $P_i$  for all  $i \leq k$  using the measure  $Z$ .

**Theorem 3.** *For every  $k \geq 2$  and  $Z \in \mathcal{S}$ , the  $\text{DISS}(k, Z)$  problem is NP-COMplete even if  $D$  is a Directed Acyclic Graph (DAG) with a single source and a single sink.*

*Proof.* Let  $Z \in \mathcal{S}$ . Let us first consider the case  $k = 2$ . We use a reduction from the PARTITION problem. Recall that the PARTITION problem takes as input a multiset  $S = \{x_1, \dots, x_n\}$  of positive integers and asks whether there exists a partition  $(X, Y)$  of  $S$  such that  $\sum_{x \in X} x = \sum_{x \in Y} x = h$  where  $2h = \sum_{x \in S} x$  (so  $\sum_{x \in S} x$  is even). The PARTITION problem is weakly NP-COMplete [6].

Let  $D_S = (V, A)$  be the DAG defined such that  $V = \{s = v_0, v_1, \dots, v_{n-1}, v_n = t\}$  and, for every  $1 \leq i \leq n$ , let us add arcs  $e_i = v_{i-1}v_i$  and  $f_i = v_{i-1}v_i$  with length  $\ell(e_i) = \ell(f_i) = x_i$  (see Figure 2). Let  $P_1$  be induced by  $\{e_i \mid 1 \leq i \leq n\}$ ,  $P_2$  be induced by  $\{f_i \mid 1 \leq i \leq n\}$  (note that  $\ell(P_1) = \ell(P_2) = 2h$ ) and let  $\theta = 1/2$ .

Note that there is a one-to-one mapping between the  $s$ - $t$  simple paths and the bipartitions of  $\{1, \dots, n\}$ . Indeed, let  $P$  be any such path. Then, for every  $1 \leq i \leq n$ , path  $P$  goes through exactly one of  $e_i$  or  $f_i$ . Let  $X_P = \{1 \leq i < n \mid e_i \in A(P)\}$  and  $Y_P = \{1 \leq i < n \mid f_i \in A(P)\}$ . Clearly,  $(X_P, Y_P)$  is a partition of  $\{1, \dots, n\}$ . Reciprocally, let  $(X, Y)$  be any partition of  $\{1, \dots, n\}$ . Then, let  $P_{XY}$  be the path induced by  $\{e_i \mid i \in X\} \cup \{f_i \mid i \in Y\}$ . Clearly,  $P_{XY}$  is a  $s$ - $t$  simple path.

First, we consider only the three similarity measures  $Asy$ ,  $Min$  and  $Max$ . Note that every  $s$ - $t$  simple path has length  $2h$  and therefore, for every  $s$ - $t$  simple paths  $P$  and  $R$ ,  $S_{Asy}(P, R) = S_{Asy}(R, P) = S_{Min}(P, R) = S_{Max}(P, R)$ . Hence, all similarity measures in  $\{Asy, Min, Max\}$  are equivalent.



By construction, for every bipartition  $(X, Y)$  of  $\{1, \dots, n\}$  (equivalently, for every  $s$ - $t$  simple path  $P_{XY}$ ),  $\ell(P_1 \cap P_{XY}) = \sum_{i \in X} x_i$  and  $\ell(P_2 \cap P_{XY}) = \sum_{i \in Y} x_i$ . Since  $\ell(P_1 \cap P_{XY}) = S_Z(P_1, P) \cdot 2h$  and  $\ell(P_2 \cap P_{XY}) = S_Z(P_2, P) \cdot 2h$ , it follows that  $(D_S, \ell, s, t, \frac{1}{2}, P_1, P_2)$  admits a  $s$ - $t$  simple path  $P$  with  $S_Z(P_1, P) \leq \frac{1}{2}$  and  $S_Z(P_2, P) \leq \frac{1}{2}$  if and only if  $S$  admits a balanced partition. So the  $\text{DISS}(2, Z)$  problem is NP-Hard for all  $Z \in \{Asy, Min, Max\}$ .

Concerning the *Jaccard* measure, i.e, the  $\text{DISS}(2, \text{Jaccard})$  problem, using the same construction proposed above but with  $\theta = \frac{1}{3}$  one can prove that the described reduction is valid.

Finally, to extend the result to any  $k \geq 2$ , it is sufficient to add, to  $D_S$ ,  $k - 2$  arc-disjoint  $s$ - $t$  paths  $P_3, \dots, P_k$  with length  $= 2h$ .  $\square$

### 3.3 Shortest path dissimilar to one given path

We now study the problem of finding a path of bounded length that is dissimilar to a set of  $k$  given paths. By Theorem 3, this problem is NP-COMplete (without bounding the length) whenever  $k \geq 2$ . So, let us study the case for  $k = 1$ . By Theorem 2, the problem is NP-COMplete for  $Z \in \{Min, Max, \text{Jaccard}\}$ . So, the only remaining case is the *Asy* measure. In contrast with Proposition 1, we prove that  $\text{SDISS}(Asy)$  is NP-COMplete. Moreover, this result hold on DAGs and for every  $Z \in \mathcal{S}$ .

Precisely, the  $\text{SDISS}(Z)$  problem takes as input a tuple  $(D, \ell, s, t, \theta, L, P)$  where  $D = (V, A)$  is a directed graph with  $\ell : A \rightarrow \mathbb{R}^+$ ,  $s, t \in V$ ,  $0 \leq \theta \leq 1$ ,  $L \in \mathbb{R}^+$ , and  $P$  is an  $s$ - $t$  simple path. It aims at deciding whether there exists an  $s$ - $t$  simple path  $Q$  that is  $\theta$ -dissimilar to  $P$  and  $\ell(Q) \leq L$ .

**Theorem 4.** *Let  $Z \in \mathcal{S}$ , the  $\text{SDISS}(Z)$  problem is NP-COMplete in the class of DAGs with a single source and a single sink.*

*Proof.* The problems is clearly in NP, so we prove its NP-hardness by a reduction from the PARTITION problem.

Let  $S = \{x_1, \dots, x_n\}$  be an instance of the PARTITION problem and  $2h = \sum_{i=1}^n x_i$ . Let  $M > 1$ . Let  $D_S = (V, A)$  be the DAG defined such that  $V = \{s = v_0, v_1, \dots, v_{n-1}, v_n = t\}$  and, for every  $1 \leq i \leq n$ , let us add arcs  $e_i = v_{i-1}v_i$  and  $f_i = v_{i-1}v_i$  with length  $\ell(e_i) = x_i$  and  $\ell(f_i) = M \cdot x_i$  respectively (see Figure 2).

Let  $P$  be the  $s$ - $t$  simple path that consists of arcs  $e_1, \dots, e_n$  and so  $\ell(P) = 2h$ . Note that, since  $M > 1$ ,  $\ell(P) \leq \ell(P')$  for every  $s$ - $t$  simple path  $P'$ . Finally, let  $L = h(M + 1)$  and let  $\theta = 1/2$  for *Asy* and *Min* measures,  $\theta = \frac{1}{M+1}$  for the *Max* measure and  $\theta = \frac{1}{M+2}$  for the *Jaccard* measure.

As in the proof of Theorem 3, it can be shown that there is a  $s$ - $t$  simple path  $Q$  with  $\ell(Q) \leq L$  and  $Q$  is  $\theta$ -dissimilar from  $P$  if and only if  $S$  is a positive instance of the PARTITION problem.  $\square$

## 4 Conclusion

In this paper, we studied several versions of the problem of finding (shortest) dissimilar paths in a digraph considering four similarity measures. An interesting question is whether there is a similarity measure for which the problem of finding  $k$  dissimilar paths can be solved in polynomial time. Another interesting question regards the accuracy of these similarity measures for real life applications.

## References

- [1] ABRAHAM, I., DELLING, D., GOLDBERG, A., AND WERNECK, R. Alternative routes in road networks. *ACM JEA* 18 (2013), 1–3.
- [2] AKGÜN, V., ERKUT, E., AND BATTA, R. On finding dissimilar paths. *EJOR* 121, 2 (2000), 232–246.
- [3] CHONDROGIANNIS, T., BOUROS, P., GAMPER, J., LESER, U., AND BLUMENTHAL, D. Finding k-shortest paths with limited overlap. *The VLDB Journal* 29, 5 (2020), 1023–1047.
- [4] CHONDROGIANNIS, T., BOUROS, P., GAMPER, J., LESER, U., AND BLUMENTHAL, D. B. Finding k-dissimilar paths with minimum collective length. In *Proceedings of the 26th ACM SIGSPATIAL* (2018), pp. 404–407.
- [5] ERKUT, E., AND VERTER, V. Modeling of transport risk for hazardous materials. *OR* 46, 5 (1998), 625–642.
- [6] GAREY, M. R., AND JOHNSON, D. S. *Computers and Intractability; A Guide to the Theory of NP-Completeness*. 1990.
- [7] GUO, L., AND SHEN, H. On finding min-min disjoint paths. *Algorithmica* 66, 3 (2013), 641–653.
- [8] SCHRIJVER, A. *Combinatorial optimization: polyhedra and efficiency*. Springer, Berlin, 2003.