De la difficulté de trouver des chemins dissimilaires

Ali Al Zoobi, David Coudert, Nicolas Nisse

De la difficulté de trouver des chemins dissimilaires †

Ali Al Zoobi [1] et David Coudert [1] et Nicolas Nisse[1]

[1] *Université Côte d'Azur, Inria, CNRS, I3S, France*

Lorsque l'on demande à son GPS un chemin pour aller à La Rochelle, celui-ci propose généralement plusieurs chemins assez différents les uns des autres : l'un par l'autoroute, l'autre par la côte, un pas cher, *etc.* La notion de (dis)similarité entre deux chemins a été définie, dans la littérature, de différentes manières qui toutes sont liées à un ratio entre la longueur de leur intersection et une certaine fonction de leurs longueurs.

Nous étudions la complexité de plusieurs variantes du problème du calcul de chemins "dissimilaires" (dont la mesure de similarité n'excède pas un certain seuil) entre deux sommets d'un graphe orienté et pondéré. Pour quatre des mesures les plus étudiées dans la littérature, nous donnons une preuve unifiée et simple du fait que trouver k plus courts chemins dissimilaires est NP-complet.

En pratique, ce que l'on cherche est une alternative à un ou des chemins que l'on connait *a priori*. Nos résultats principaux concernent ce type de problème. Plus précisemment, pour chacune des quatre mesures considérées, nous montrons que si $k \geq 2$ chemins sont donnés, en trouver un nouveau qui soit dissimilaire des premiers est NP-complet. Enfin, nous montrons que si un chemin P est donné, trouver un plus court chemin parmi ceux qui sont dissimilaires de P est NP-complet. Ce dernier résultat est à mettre en contraste avec le fait que, pour l'une des mesures, trouver un chemin dissimilaire à un chemin donné peut être résolu très simplement en temps polynomial.

Mots-clefs : k plus courts chemins simples, similarité entre chemins

1 Introduction

The k shortest simple paths problem ($kSSP$) aims at finding the top-k 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 s-t simple path is a sequence $(s=v_0,v_1,\cdots,v_l=t)$ of distinct vertices starting with s and ending with t, such that $(v_i,v_{i+1})\in A$ for all $0\leq i<l$. Let $\ell : A\rightarrow \mathbb{R}^+$ be a length function of the arcs. For any path P in D, let $\ell(P)=\sum_{e\in A(P)}\ell(e)$, i.e, the length of a path 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 path, another user wants to avoid a traffic jam, a third one prefers to travel along the cost, *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 also studied in [CBG+18]) were first proposed by Erkut and Verter [EV98], motivated by the transportation of hazardous materials where it is recommended to avoid residential areas and crowded routes. Akgün *et al.* [AEB00] proposed and analysed the first basic solution, consisting in computing a huge set of shortest paths and then choosing a subset of 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.* [ADGW13] 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

†This work has been supported by the UCA^JEDI Investments in the Future project managed by the National Research Agency (ANR-15-IDEX-01), project MULTIMOD (ANR-17-CE22-0016), project Digraphs (ANR-19-CE48-0013), and by Région Sud PACA.

paths. A noticeable heuristic proposed in [ADGW13] 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. Several studies by Chondrogiannis *et al.* (see [CBG+20]) offer both theoretical and empirical study of the problem. First, they formally proved that finding k shortest dissimilar paths is weakly NP-complete for both the Asymmetric 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 enabling to find 4 dissimilar paths in the road network of Rome (3,000 vertices) in less than one second. They also proposed advanced heuristics enabling to scale on a road networks with one million vertices and two millions arcs while achieving an acceptable paths "quality".

In this paper, we further study the computational complexity of computing (shortest) dissimilar paths for four of the main measures. More formally, let P, P' be two paths of D and let $X = \sum_{e \in A(P) \cap A(P')} \ell(e)$. The four considered measures are defined as follows.

Name (Z)	Jaccard [EV98]	Asymmetric [ADGW13]	Min [CBG+18]	Max [EV98]
$S_Z(P,P') =$	$\dfrac{X}{\ell(P \cup P')}$	$\dfrac{X}{\ell(P)}$	$\dfrac{X}{Min\{\ell(P),\ell(P')\}}$	$\dfrac{X}{Max\{\ell(P),\ell(P')\}}$

Let $\mathcal{S} = \{Asymmetric, Jaccard, Min, Max\}$. Given one of the similarity measures $Z \in \mathcal{S}$ and a threshold $0 \leq \theta \leq 1$, two paths P and P' are said θ-*dissimilar* (or P' is said θ-*dissimilar to* P in the case of asymmetrical similarity) for a 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. Nevertheless, in many practical scenarios, a part of the solution is generally already given (e.g., any shortest path). Therefore, a natural question is whether one can find a (shortest) path dissimilar to a set of given paths. Section 3 is devoted to this problem. More precisely, we first prove that finding a path dissimilar (for each of the considered four measures) to a set of $k \geq 2$ given paths is NP-complete. If only one path P is initially given, we show that computing a second path that is dissimilar to P for the Asymmetric measure can be done in polynomial time. In contrast, for each of the four measures, we show that computing a shortest path among those dissimilar to P is NP-complete.

2 Finding k shortest dissimilar paths

In this section, we show that finding k shortest dissimilar paths is NP-complete for each of the considered similarity measures. More formally, given a digraph $D = (V,A)$ with length function $\ell : A \to \mathbb{R}^+$, a pair of vertices $(s,t) \in V \times V$, an integer $k \geq 2$, a threshold $0 \leq \theta \leq 1$ [‡], k constants $L_1, L_2, \cdots, L_k$ and a similarity measure $Z \in \mathcal{S}$, the problem of finding k shortest dissimilar paths (k-ShortestDiss(Z)) asks to decide whether there exists k paths from s to t that are mutually θ-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 Asymetric and Min measures [CBG+20]. Here we propose a unified (for all considered measures) and simpler proof.

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. We consider the case $k = 2$ that can easily be extended to any $k \geq 2$ by adding $k-2$ dummy arc-disjoint sufficiently long paths.

For every $Z \in \mathcal{S}$, the problems are clearly in NP, so we only prove the NP-hardness by a reduction from the Min-MinDP problem [GS13]. Given a graph $G = (V,E)$ with length function $\ell : E \to \mathbb{R}^+$, 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 [GS13].

‡. In this paper, we suppose that the representation of θ is polynomially bounded by the size of the input digraph

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

- If I is a positive MIN-MINDP instance, it means that there is two arc disjoint s-t paths P and P' s.t. $\ell(P) \leq \delta$. 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 $\sum_{e \in A(P_1) \cap A(P_2)} \ell(e) = 0$. In addition, $\ell(P_1) \leq \delta = L_1$ and $\ell(P_2) \leq M$. So, I' is a positive k-SHORTESTDISS(Z) instance
- If I' is a positive k-SHORTESTDISS(Z) instance, it means that there is two s-t paths P_1 and P_2 s.t. $\ell(P_1) \leq L_1$ 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. □

3 Finding a (shortest) path dissimilar to given paths

In this section, we present our main results. First, we show that finding a path dissimilar to another given path is polynomial for the Asymmetric measure. Then, we prove that the problem of finding a path dissimilar to two given paths is NP-complete. Similarly, we show that finding a shortest path dissimilar to one given path is also NP-complete.

First of all we start with the easiest variant of the problem that is the problem of finding a path dissimilar to another for the Asymmetric measure. Given a digraph $D = (V, A)$ with $\ell : A \to \mathbb{R}^+$, two vertices $s, t \in V$, $0 \leq \theta \leq 1$, and a simple s-t path P, DISS$(1, Asymmetric)$ is the problem of finding an s-t path Q that is θ-dissimilar to P using the Asymmetric measure.

Proposition 1 *DISS$(1, Asymmetric)$ can be solved in the same time as any shortest-path algorithm*

Proof. Let $\ell' : A \to \mathbb{R}^+$ 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 simple s-t path Q is a solution of the DISS$(1, Asymmetric)$ problem if and only if $\ell'(Q) = \sum_{e \in A(Q)} \ell'(e) = \sum_{e \in A(P) \cap A(Q)} \ell(e) \leq \theta \cdot \ell(P)$. □

Now, we consider the generic version of the problem, called DISS(k', Z). Let $Z \in \mathcal{S}$ be a similarity measure and let $k' \in \mathbb{N}^*$. The DISS(k', Z) problem takes a tuple $(D, \ell, s, t, \theta, P_1, \cdots, P_{k'})$ as input where $D = (V, A)$ is a digraph with $\ell : A \to \mathbb{R}^+$, $s, t \in V$, $0 \leq \theta \leq 1$, and k' simple s-t paths $P_1, \cdots, P_{k'}$. It aims at deciding whether there exists a simple s-t path Q such that $S_Z(P_i, Q) \leq \theta$ for every $1 \leq i \leq k'$ (i.e., Q is θ-dissimilar to each of the P_i for the similarity measure Z).

Theorem 2 *For every $k' \geq 2$ and $Z \in \mathcal{S}$, the DISS(k', Z) problem is NP-Complete even if D is a 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 a multiset $S = \{x_1, ..., x_n\}$ of positive integers as input 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 well known to be weakly NP-complete [GJ90].

Let $D_S = (V, A)$ be the DAG defined such that $V = \{s = v_0, v_1, \cdots, v_{n-1}, t = v_n\}$ 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 1). 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 simple s-t paths and the bipartitions of $\{1, \cdots, n\}$. 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 $\{1, \cdots, n\}$. Reciprocally, let (X, Y) be any partition of $\{1, \cdots, n\}$. 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 simple s-t path.

First, we consider only the three similarity measures $Asymmetric, Min$ and Max. Note that every simple s-t path has length $2h$ and therefore, for every simple s-t 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 $\{Asymmetric, Min, Max\}$ are equivalent.

By construction, for every bipartition (X, Y) of $\{1, \cdots, n\}$ (equivalently, for every simple s-t 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 simple s-t path P with $S_Z(P_1, P) \leq \frac{1}{2}$ and $S_Z(P_2, P) \leq$

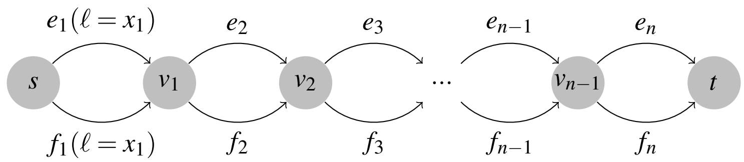

FIGURE 1: Digraph $D_S = (V,A)$ defined from $\mathcal{S} = \{x_1, \cdots, x_n\}$. For all $1 \leq i \leq n$, $\ell(e_i) = x_i$. For all $1 \leq i \leq n$, $\ell(f_i) = x_i$ (Theorem 2).

$\frac{1}{2}$ if and only if S admits a balanced partition. So the DISS$(2,Z)$ problem is NP-Hard for all $Z \in \{Asymmetric, Min, Max\}$.

Concerning the *Jaccard* measure, i.e, the DISS$(2,J)$ 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 $k' > 2$, it is sufficient to add $k'-2$ dummy arc-disjoint s-t paths $P_3, \cdots, P_{k'}$ with length $= 2h$. $\square$

In contrast with above two results, let us focus on the case where $k' = 1$ (only one path P is already given) but where we look for a path dissimilar to P but which is as short as possible. The SHORTESTDISS$(1,Z)$ problem takes a tuple $(D, \ell, s, t, \theta, L, P)$ as input 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 a simple s-t path. It aims at deciding whether there exists a simple s-t path Q such that Q is θ-dissimilar from P (for a measure Z) and $\ell(Q) \leq L$.

Theorem 3 *Let $Z \in \mathcal{S}$. The* SHORTESTDISS(k',Z) *problem is NP-Complete for $k' \geq 1$ in the class of Directed Acyclic Graphs (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 $S = \{x_1, ..., x_n\}$ be an instance of the PARTITION problem and $2h = \sum_{i \leq n} x_i$. Let $M > 1$. Let $D_S = (V,A)$ be the DAG defined such that $V = \{s = v_0, v_1, \cdots, v_{n-1}, t = v_n\}$ 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.

Let P be the simple s-t path that consists of arcs $e_1, \cdots, e_n$ and so $\ell(P) = 2h$. Note that, since $M > 1$, $\ell(P) \leq \ell(P')$ for every simple s-t path P'. Finally, let $\theta = 1/2$ and $L = h(M+1)$.

As in the proof of Theorem 2, it can be shown that there is a path Q with $\ell(Q) \leq L$ and Q is θ-dissimilar from P if and only if S is a positive instance of the PARTITION problem. $\square$

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.

Références

[ADGW13] I. Abraham, D. Delling, A. Goldberg, and R. Werneck. Alternative routes in road networks. *ACM JEA*, 18 :1–3, 2013.

[AEB00] V. Akgün, E. Erkut, and R. Batta. On finding dissimilar paths. *EJOR*, 121(2) :232–246, 2000.

[CBG+18] T. Chondrogiannis, P. Bouros, J. Gamper, U. Leser, and D. B. Blumenthal. Finding k-dissimilar paths with minimum collective length. In *Proceedings of the 26th ACM SIGSPATIAL*, pages 404–407, 2018.

[CBG+20] T. Chondrogiannis, P. Bouros, J. Gamper, U. Leser, and D. Blumenthal. Finding k-shortest paths with limited overlap. *The VLDB Journal*, 29(5) :1023–1047, 2020.

[EV98] E. Erkut and V. Verter. Modeling of transport risk for hazardous materials. *OR*, 46(5) :625–642, 1998.

[GJ90] M. R Garey and D. S Johnson. *Computers and Intractability ; A Guide to the Theory of NP-Completeness*. 1990. ISBN :0716710455.

[GS13] L. Guo and H. Shen. On finding min-min disjoint paths. *Algorithmica*, 66(3) :641–653, 2013.