24th European Conference on Artificial Intelligence - ECAI 2020
Santiago de Compostela, Spain

The Challenge of Optimal Paths in Graphs with Item Sets

Adi Botea[1] and **Akihiro Kishimoto**[2] and **Radu Marinescu**[2] and **Elizabeth M. Daly**[2] and **Oznur Alkan**[2]

1 Introduction

Modelling a problem as finding a shortest path in a graph has been successful in many AI problems, including various types of planning. Search algorithms such as A* owe their efficiency in part to the fact that, when an intermediate node can be reached along multiple paths, reaching it through one optimal path can be sufficient. However, the classical shortest path framework might not be suitable for domains such as planning a career pathway optimized to the skills needed along the way. Career recommendations, including longer-term goals and pathways, are an important emerging application [1].

In career pathway planning, the task is to plan a sequence of job roles that would take a user from a current role to a goal role. Each job role has associated a set of skills, and the user needs to have the corresponding skills when being in a given role. Skills can overlap from one job to another. Variants of the career pathway planning problem include: minimizing the number of new skills to acquire on a path towards the goal role; and reaching a given target set of skills in a shortest sequence of job roles. Applications go beyond career planning. For example, an agent on a map may need to broadcast some information to a given set of target listeners. Various locations on the map have a given subset of subscribers (listeners) capable to receive information broadcast at that location. The agent would need to minimize that distance travelled and reach a number of broadcast locations whose union of subscribers cover all the target listeners.

Motivated by such applications, we introduce graphs with item sets, where each node has a set of items from a finite, global set. We define two problems, for the two career planning cases outlined earlier. We show that our problems are NP-complete, unlike the traditional shortest path problem, which is solved in polynomial time.

Work related to item sets includes analysing greedy best-first search [4], and computing optimal delete-relaxed plans. Differences include that, in the latter, limiting the applicable actions to at most 2 per state results in an easy problem. In contrast, our Theorem 2 and its proof show an NP-hardness result even with such a small branching factor in use. A deeper comparison is beyond this paper's focus.

2 Item Sets in Graphs

Definition 1 (Graph with item sets)**.** *Given a finite set A, called an item universe, a graph with item sets, or an IS graph, is a graph $G = (V, E)$, where each node v has associated a set of items It$(v) \subseteq A$.*

Definition 2 (Item set for a path)**.** *Given a path $\pi = (v_1, v_2, \ldots, v_n)$ in an IS graph, the item set of the path is It$(\pi) = \bigcup_{i=1}^{n}$ It(v_i).*

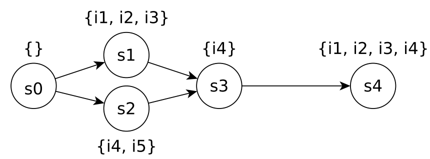

Figure 1. A case where the prefix optimality does not hold.

Definition 3 (Item-set cost of a path)**.** *Given a path π in an IS graph, its item-set cost is defined as cost$_{IS}(\pi) = |$It$(\pi)|$.*

Definition 4 (Prefix optimality [3])**.** *Given a cost function for paths in a graph, we say that the cost function satisfies the prefix optimality if, for any optimal path $(v_1, v_2, \ldots, v_n)$, every prefix $(v_1, v_2, \ldots, v_i)$, with $i \leq n$, is an optimal path from v_1 to v_i.*

Many search problems from the literature satisfy the prefix optimality. See [2] for a notable exception. We show that:

Proposition 1. *The item-set cost lacks the prefix optimality property.*

Proof. Figure 1 gives a counter-example. We show two paths from s_0 to s_4, namely $\pi_1 = s_0, s_1, s_3, s_4$ and $\pi_2 = s_0, s_2, s_3, s_4$. Their item costs are cost$_{IS}(\pi_1) = |\emptyset \cup \{i_1, i_2, i_3\} \cup \{i_4\} \cup \{i_1, i_2, i_3, i_4\}| = |\{i_1, i_2, i_3, i_4\}| = 4$; and cost$_{IS}(\pi_2) = |\emptyset \cup \{i_4, i_5\} \cup \{i_4\} \cup \{i_1, i_2, i_3, i_4\}| = |\{i_1, i_2, i_3, i_4, i_5\}| = 5$. It follows that π_1 is optimal. However, its prefix s_0, s_1, s_3 is not an optimal path from s_0 to s_3. Its cost is cost$_{IS}(s_0, s_1, s_3) = |\emptyset \cup \{i_1, i_2, i_3\} \cup \{i_4\}| = |\{i_1, i_2, i_3, i_4\}| = 4$. Path s_0, s_2, s_3 has a better cost, as cost$_{IS}(s_0, s_2, s_3) = |\emptyset \cup \{i_4, i_5\} \cup \{i_4\}| = |\{i_4, i_5\}| = 2$. □

3 Paths to a Goal Node

In this section we focus on reaching a goal node (e.g., a goal role in career planning), while minimizing the path item-set cost (e.g., the number of new skills to learn along the way).

Definition 5 (Path-to-Node Problem – PtNP)**.** *Input: An IS graph G; an initial node s_0; a goal node s_g; an integer k. The question is if a path from s_0 to s_g exists with the item-set cost no larger than k.*

Theorem 1. *PtNP is NP-complete.*

Proof. The problem belongs to NP, as we can verify a solution in polynomial time. The NP-hardness is shown with a reduction from the minimum k-union problem, an NP-hard problem [6]. In this problem, we have a universe $U = \{1, 2, \ldots, n\}$, a set of subsets of U

[1] Eaton
[2] IBM Research, Ireland
Work performed when the first author was affiliated with IBM Research. We thank the reviewers and Malte Helmert for their feedback.

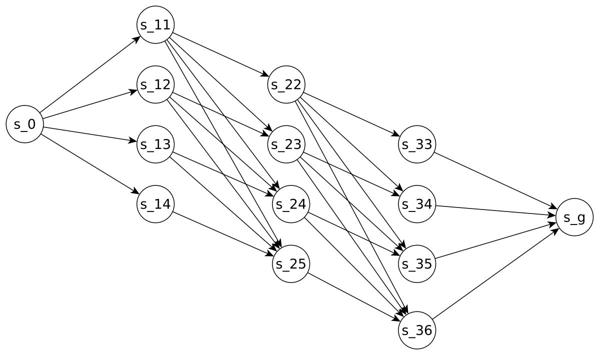

Figure 2. Graph for the PtNP instance built in the proof to Theorem 1.

$S = \{S_0, S_1, \ldots, S_{p-1}\}$, with $S_i \subset U, \forall i$, an integer k, and an integer t. The question is whether a collection of k elements from S exists such that the union of these elements has at most t elements.

Given an arbitrary instance of the minimum k-union problem, we construct a PtNP instance in polynomial time. The graph G of the instance is illustrated in Figure 2. The figure shows a toy example, for $p = 6$ and $k = 3$, to avoid clutter. More generally, there is a node s_0 and a node s_g. We also define k layers of nodes, with each layer having $p - k + 1$ nodes. At layer $i \in \{1, \ldots, k\}$, we have nodes $s_{i,i}, s_{i,i+1}, \ldots, s_{i,i+p-k}$. We define edges from s_0 to every node in layer 1, and from every node in layer k to s_g. Furthermore, we define edges from nodes $s_{i,m}$ in layer i to nodes $s_{i+1,n}$ in layer $i + 1$, for $m < n$. See Figure 2 for an illustration. Nodes s_0 and s_g have an empty item set each. A node $s_{i,m}$ has its item set equal to S_m.

We claim that the instance of the minimum k-union problem has a solution if and only if our PtNP instance has a path from s_0 to s_g with an item cost no larger than t. Assume that the instance of the minimum k-union problem has a solution, $S_{j_1}, S_{j_2}, \ldots, S_{j_k}$, where the indexes are ordered as $j_1 < j_2 < \cdots < j_k$. Consider the path $s_0 \rightarrow s_{1,j_1} \rightarrow s_{2,j_2} \rightarrow \cdots \rightarrow s_{k,j_k} \rightarrow s_g$. By construction, the item set of this path is $S_{j_1} \cup S_{j_2} \cup \cdots \cup S_{j_k}$ and thus it has at most t elements. Assume now that there is a path $s_0 \rightarrow s_{1,j_1} \rightarrow \cdots \rightarrow s_{k,j_k} \rightarrow s_g$ with an item set of at most t elements. Then the collection $S_{j_1}, \ldots, S_{j_k}$ is a solution to the minimum k-union problem with no more than t elements in total. $\square$

4 Paths to a Target Item Set

Consider the case when we want to achieve a set of items, not necessarily to reach a given node in the graph. For example, in career pathway planning, a user may want to achieve a set of target skills.

Definition 6 (Path-to-item-set problem – PtISP). *Input: An IS graph G; an initial node s_0; a set of items I_g; and an integer k. The question is if a path π exists such that: it originates in s_0, it has at most k edges, and its item set is a superset of I_g: $I_g \subseteq It(\pi)$.*

Theorem 2. *PtISP is NP-complete.*

Proof. The problem clearly belongs to NP, as we can verify a solution in polynomial time. The NP-hardness is shown with a reduction from the set cover problem, an NP-hard problem [5]. In the set-cover problem, we have a universe $U = \{1, 2, \ldots, n\}$, a set of subsets of U $S = \{S_0, S_1, \ldots, S_{p-1}\}$, with $S_i \subset U, \forall i$, and an integer k. The question is whether a collection of at most k elements from S exists such that the union of these elements is U (i.e., they cover U).

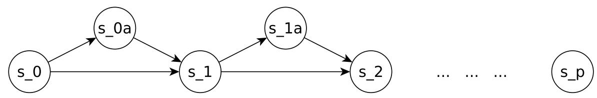

Figure 3. Graph for the PtISP instance built in the proof to Theorem 2.

Given an arbitrary instance of the set cover problem, we construct a PtISP instance in polynomial time. Figure 3 illustrates the IS graph G of the PtISP instance. Nodes are defined as follows. For every subset S_i in S (in the set cover problem instance) we define two nodes, s_i and s_{ia}. In addition, there is one node s_p at the right. Nodes s_i, with $i < p$, have an empty set of items each. Node s_{ia} has the item set S_i. Node s_p has the item set $\{\alpha\}$, where α is a new symbol, not contained in U. Edges in the graph G are defined as follows: (s_i, s_{ia}), (s_i, s_{i+1}), and (s_{ia}, s_{i+1}). See Figure 3 for an illustration. In the PtISP instance, define $I_g = U \cup \{\alpha\}$.

We claim that the PtISP instance has a solution of at most $p + k$ edges if and only if the set cover problem has a solution (cover) of at most k subsets.

Consider that the PtISP instance has a solution of at most $p + k$ edges. Observe that every solution must contain node s_p, as the symbol α is required (i.e., defined in I_g), and no other node contains this symbol in its item set. Thus, every solution is a path from s_0 to s_p, and therefore every solution has at least p steps (edges). This further implies that our solution with at most $p + k$ edges contains at most k "detours" of the type $s_i \rightarrow s_{ia} \rightarrow s_{i+1}$. The item sets of the nodes of the type s_{ia} contained in the solution form a cover of U. Thus, U has a cover of at most k elements.

Consider now that U has a cover set of at most k elements. In the PtISP instance, construct a path with a detour $s_i \rightarrow s_{ia} \rightarrow s_{i+1}$ for each S_i contained in the cover of U, and a "shortcut" $s_i \rightarrow s_{i+1}$ in all other cases. The resulting path has at most $p + k$ elements (since there are at most k detours), and it is a solution to our instance (i.e., we cover $I_g = U \cup \{\alpha\}$). $\square$

5 Conclusion

Motivated by long-term career pathway planning, we discussed two problems for optimal paths in graphs with item sets, and showed their NP completeness. Future work includes efficient problem representations, optimal and approximate algorithms, and effective heuristics.

REFERENCES

[1] O. Alkan, E. M. Daly, A. Botea, A. N. Valente, and P. Pedemonte, 'Where can my career take me?: harnessing dialogue for interactive career goal recommendations', in *Proceedings of the 24th International Conference on Intelligent User Interfaces, IUI*, pp. 603–613, (2019).

[2] G. V. Batz and P. Sanders, 'Time-dependent route planning with generalized objective functions', in *Algorithms – ESA*, pp. 169–180, (2012).

[3] S. Edelkamp, S. Schroedl, and S. Koenig, *Heuristic Search: Theory and Applications*, Morgan Kaufmann Publishers Inc., 2010.

[4] M. Heusner, T. Keller, and M. Helmert, 'Best-case and worst-case behavior of greedy best-first search', in *Proc. of the International Joint Conference on Artificial Intelligence, IJCAI*, pp. 1463–1470, (2018).

[5] R. Karp, 'Reducibility among combinatorial problems', in *Complexity of Computer Computations*, 85–103, Plenum Press, (1972).

[6] S. Vinterbo. A note on the hardness of the k-ambiguity problem, 2002.