Top-k Reliable Edge Colors in Uncertain Graphs

Arijit Khan*, Francesco Gullo#, Thomas Wohler*, Francesco Bonchi#
*Systems Group, ETH Zurich, Switzerland #Yahoo! Labs, Barcelona, Spain
arijit.khan@inf.ethz.ch, {gullo, bonchi}@yahoo-inc.com, twohler@student.ethz.ch

ABSTRACT

We study the fundamental problem of finding the set of top-k edge colors that maximizes the reliability between a source node and a destination node in an uncertain and edge-colored graph. Our top-k reliable color set problem naturally arises in a variety of real-world applications including pathway finding in biological networks, topic-aware influence maximization, and team formation in social networks, among many others. In addition to the $\#\mathbf{P}$-completeness of the classical reliability finding problem between a source and a destination node over an uncertain graph, we prove that our problem is also $\mathbf{NP}$-hard, and neither sub-modular, nor super-modular. To this end, we aim at designing effective and scalable solutions for the top-k reliable color set problem. We first introduce two baselines following the idea of repetitive inclusion of the next best edge colors, and we later develop a more efficient and effective algorithm that directly finds the highly-reliable paths while maintaining the budget on the number of edge-colors. An extensive empirical evaluation on various large-scale and real-world graph datasets illustrates that our proposed techniques are both scalable and highly accurate.

Categories and Subject Descriptors

H.3.3 [**Information Search and Retrieval**]: Search process; H.2.8 [**Database Applications**]: Uncertain networks

General Terms

Algorithms, Performance

Keywords

Uncertain Graphs; Reliability; Edge Colors; Top-K

1. INTRODUCTION

Uncertainty is inherent in graph data due to a variety of reasons, such as noisy measurements, inference and prediction models, or explicit manipulation, e.g., for privacy purposes. In these cases, data is represented as an uncertain graph, that is, a graph whose arcs are accompanied with a probability of existence. A fundamental problem in uncertain graphs is *reliability query*, which asks to estimate the probability that a given destination node is reachable from a given source node. The reliability estimation problem has been widely studied in device networks [1], social networks [9], as well as in biological networks [10].

Nevertheless, most of these reliability queries over uncertain graphs are performed without considering any edge attributes, which we simply refer to as edge colors. Since complex networks, such as biological, social, and information networks usually exhibit diverse types of relationships among the entities, it is often meaningful to define reliability via a constrained set of edge colors [3]. To this end, we study the following novel and critical problem: *given a source and a destination node in an edge-colored, uncertain graph and a small positive integer k, find the edge-color-set of size k that maximizes the reliability from the source to the destination.*

Application. The top-k reliable color set problem naturally arises in a variety of real-world scenarios as follows.

Pathway Formation in Biological Networks: In order to understand the metabolic chain reactions in cellular systems, biologists utilize metabolic networks, where each vertex represents a compound, and a directed edge between two compounds indicates that one compound can be transformed into another through a certain chemical reaction [8]. The edge colors record the enzymes which control these reactions. In addition, uncertainty arises in metabolic network edges due to noisy measurements, experimental errors, inference, and prediction models. One of the basic questions on such networks is finding the top-k set of enzymes which create pathways of very high probabilities between two given compounds.

Topic-Aware Information Cascade: Marketing companies are gradually turning to social networks such as Facebook, Twitter, and LinkedIn for campaigning of their products. However, the influence of an individual over another in a social network often changes drastically based on advertisement contents [2]. Therefore, one can formulate the topic-aware information cascade problem with an uncertain graph model, where the probabilities on the edges vary based on advertisement features. In this setting, it is critical for the marketing companies to identify the top-k advertisement features such that the information cascade from an early adopter to a group of target customers could be maximized.

Challenges. Our top-k reliable color set problem is a non-trivial one — in fact, the simplest reliability computation problem over uncertain graphs is a $\#\mathbf{P}$-complete problem [1]. Due to the large size of networks, most work in this regard has resorted to Monte-Carlo (MC) sampling methods [6], as well as other sampling techniques improving upon the efficiency of MC methods (e.g., RHT-sampling [9]). These sampling-based approaches, in reality, estimate the reliability between two nodes very well, and they usually require only polynomial time in the size of the network.

CIKM'15, October 19-23, 2015, Melbourne, VIC, Australia

http://dx.doi.org/10.1145/2806416.2806619.

However, even considering polynomial-time sampling techniques to estimate reliability, the top-k reliable color set problem remains **NP**-hard. More importantly, unlike the classical max-k cover problem, our problem is neither sub-modular, nor super-modular. Therefore, an iterative hill-climbing algorithm that maximally increases the marginal gain at every iteration, and which has been widely used for solving the max-k cover problem, can no longer be employed in our case for deriving similar approximation guarantees.

Our contribution. We propose two baselines to solve our top-k reliable color set problem, and we also design a more efficient and effective algorithm that directly finds the highly-reliable paths while maintaining the budget on the number of edge-colors. Our experimental results over three real-world large-scale graph datasets attest the effectiveness and efficiency of our approach.

2. PRELIMINARIES

2.1 Problem Formulation

An edge-labeled, uncertain graph $\mathcal{G}$ is a quadruple (V, E, C, P), where V is a set of n nodes, $E \subseteq V \times V$ is a set of m directed edges, C is the set of all edge-colors in $\mathcal{G}$, whereas $C(e) \subseteq C$ is a set of edge-colors assigned to the edge $e \in E$. Finally, $P : E \times C \rightarrow (0, 1)$ assigns a conditional probability on an edge given a specific color, i.e., $P(e|c) \in (0, 1)$.

Edge Existence Probability. In this work, we assume that the conditional probability of an edge $e \in E$ given some color $c \in C(e)$, that is, $P(e|c)$ is independent [3] of the other colors in $C(e)$. Thus, the edge-existence probability of e given the edge-colors $c_1, c_2, \ldots, c_r \in C$ is: $P(e|c_1 c_2 \ldots c_r) = 1 - \prod_{i=1}^{r}(1 - P(e|c_i))$. Given a predefined edge-color set $C_1 \subseteq C$, one can compute all the edge-existence probabilities in the uncertain graph $\mathcal{G}$. If the edge-color set C_1 is predefined, we simply write the edge-existence probabilities as $P(e|C_1)$.

Possible World Semantics. The bulk of the literature on uncertain graphs and device-network-reliability assumes the existence of the edges in the graph independent from one another and interprets uncertain graphs according to the well-known possible-world semantics [6, 8, 9]. More precisely, given a pre-defined edge-color set C_1, a possible graph $G \sqsubseteq \langle \mathcal{G}, C_1 \rangle$ is a pair (V, E_G), where $E_G \subseteq E$, and its sampling probability is:

$$Pr(G|C_1) = \prod_{e \in E_G} P(e|C_1) \prod_{e \in E \setminus E_G} (1 - P(e|C_1)) \quad (1)$$

For a possible deterministic graph $G \sqsubseteq \langle \mathcal{G}, C_1 \rangle$, we define an indicator function $I_G(s, t)$ to be 1 if there is a path in G from a source node $s \in V$ to a target node $t \in V$, and 0 otherwise. Finally, the probability that t is reachable from s in the uncertain graph $\mathcal{G}$ and via a pre-defined edge-color set C_1 is defined as the *edge-color-constrained reliability* from s to t, and it is denoted by $R_{C_1}(s, t)$. The edge-color-constrained reliability is computed as follows.

$$R_{C_1}(s, t) = \sum_{G \sqsubseteq \langle \mathcal{G}, C_1 \rangle} [I_G(s, t) \times Pr(G|C_1)] \quad (2)$$

The number of possible worlds $G \sqsubseteq \langle \mathcal{G}, C_1 \rangle$ is exponential in the number of edges, which makes the exact reliability computation a #**P**-complete problem; and hence, almost infeasible even for moderately-sized graphs.

Problem Statement. We are now ready to define our problem statement.

PROBLEM 1 (TOP-k RELIABLE COLOR SET). *Given a source node $s \in V$ and a destination node $t \in V$ in an edge-colored, uncertain graph $\mathcal{G} = (V, E, C, P)$, and a small positive integer k, find the edge-color-set C_1 of size k that maximizes the edge-color-constrained reliability $R_{C_1}(s, t)$ from s to t. Formally,*

$$\arg\max_{C_1 \subseteq C} R_{C_1}(s, t) \quad \textit{such that} \quad |C_1| = k \quad (3)$$

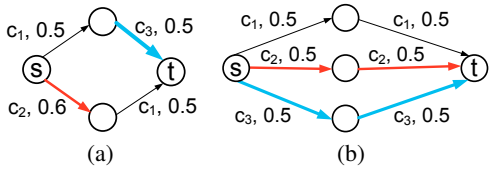

Figure 1: Example for Non-Sub-Modularity and Non-Super-Modularity

Intuitively, the top-k reliable edge-colors create multiple paths of high probabilities from source to destination node.

2.2 Hardness Results

Our problem, however, is non-trivial. Theorem 1 shows that Problem 1 is **NP**-hard, even when one considers polynomial-time reliability estimation approaches (e.g., MC sampling).

THEOREM 1. *The top-k reliable color set problem is* **NP***-hard.*

PROOF. We prove **NP**-hardness by a reduction from the max-k cover problem. In max-k cover problem, we are given a universe U, and a set of h subsets of U, i.e., $\mathcal{S} = \{S_1, S_2, \ldots, S_h\}$, where $S_i \subseteq U$ for all $i \in [1 \ldots h]$. The objective is to find a subset $\mathcal{S}^*$ of $\mathcal{S}$ of size k such that the number of elements covered by $\mathcal{S}^*$ is maximized, i.e., so as to maximize $|\cup_{S \in \mathcal{S}^*} S|$. Given an instance of the max-k cover problem, we construct in polynomial time an instance of our top-k reliable color set problem.

We put in our edge-colored, uncertain graph $\mathcal{G}$ a source node s and a destination node t. Next, we include a set of nodes $u_1, u_2, \ldots, u_Z$, one for each element in $U (Z = |U|)$, and connect each of these nodes u to the destination node t with a directed edge (u, t). Each such edge (u, t) has color c, and we assign a probability $P((u, t)|c) = p$, with $p < 1$. We then add a set of nodes $x_1, x_2, \ldots, x_Z$, one for each element in $U (Z = |U|)$, and connect each of these nodes x to the source node t with a directed edge (s, x). Each such edge (s, x) also has color c, and we assign a probability $P((s, x)|c) = p$, with $p < 1$. Finally, if some element $u_i \in U$ is covered by at least one of the subsets in $\mathcal{S}$, we add a directed edge (x_i, u_i) in $\mathcal{G}$. For each of $S_j \in \mathcal{S}$ that covers the item u_i, we assign a color c_j on the edge (x_i, u_i), and then, we also assign a probability $P((x_i, u_i)|c_j) = 1$.

Now, we ask for a solution of our problem on the graph constructed this way by using $k + 1$ colors. One may observe that every solution to our problem necessarily takes color c, because otherwise there would be no way to connect s to t. Also, the reliability is maximized by properly selecting colors that make each of the edges (x_i, u_i) exist with probability 1. However, in order for each edges (x_i, u_i) to exist with a probability 1, it suffices to have selected only one of the colors between such a pair of nodes. Thus, we can see that maximizing reliability with $k+1$ colors corresponds to maximizing coverage of elements in U with k sets in $\mathcal{S}$. Hence, the theorem. □

In this paper, we leave the following question open whether Problem 1 can be approximately solved within a constant factor in polynomial time or not. However, we show that unlike the max-k cover problem, our top-k reliable color set problem is neither sub-modular, nor super-modular; therefore, making it difficult to design an approximate solution with provable performance guarantees.

CLAIM 1. *The top-k reliable color set problem is not sub-modular.*

A function $f()$ is sub-modular if it satisfies the following property: $f(A \cup x) - f(A) \geq f(B \cup x) - f(B)$, for all elements x and all pairs of sets $A \subseteq B$. We show non-sub-modularity of our problem with an example in Figure 1(a). More specifically, let $C_1 = \{c_2\}$, $C_2 = \{c_1, c_2\}$. It is easy to verify that $R_{C_1}(s,t) = 0$, $R_{C_1 \cup \{c_3\}}(s,t) = 0$, $R_{C_2}(s,t) = 0.3$, and $R_{C_2 \cup \{c_3\}}$=0.475. Clearly, the sub-modularity property does not hold in this example.

CLAIM 2. *The top-k reliable color set problem is not super-modular.*

A function $f()$ is super-modular if it satisfies the following property: $f(A \cup x) - f(A) \leq f(B \cup x) - f(B)$, for all elements x and all pairs of sets $A \subseteq B$. We show non-super-modularity of our problem with an example in Figure 1(b). Let $C_1 = \{c_1\}, C_2 = \{c_1, c_3\}$. One may verify that $R_{C_1}(s,t) = 0.25$, $R_{C_1 \cup \{c_2\}}(s,t) = 0.438$, $R_{C_2}(s,t) = 0.438$, and $R_{C_2 \cup \{c_2\}} = 0.578$. Hence, the super-modularity property does not hold in this example.

3. ALGORITHMS FOR TOP-K COLOR RELIABILITY

As the top-k reliable color set problem is **NP**-hard, we develop two greedy baselines, as well as a more effective and efficient heuristic solution that provides a good approximation to our problem.

3.1 Individual Top-k: First Baseline

Our individual top-k algorithm estimates the reliability between the source and the destination nodes attained by each edge-color individually. In other words, we compute $R_{\{c\}}(s,t)$ for every edge-color $c \in L$. We report the top-k edge colors that achieve the highest reliability individually.

Time Complexity. For each color, we can estimate reliability by applying the MC sampling technique. If we require total K iterations of MC sampling in order to get a good estimate, then the time complexity to compute the reliability for each color is given by: $\mathcal{O}(K(n+e))$. Here, n and e are the number of nodes and edges in the uncertain graph, respectively. Therefore, the overall complexity of our individual top-k baseline algorithm is $\mathcal{O}(|C|K(n+e) + |C| \log k)$, the last term is due to finding the top-k colors based on individual reliability values.

Difficulties. The individual top-k algorithm suffers from several shortcomings, which are both accuracy and efficiency-driven.

- This baseline algorithm is unable to capture the contribution of the paths that consist of multiple edge-colors. For example, in Figure 1(a), the individual reliability attained by each of the three colors is 0; and therefore, if we are to select the top-2 color-set, it will be a random selection by our first baseline. However, in reality, the top-2 color set is $\{c_1, c_2\}$.
- For large-scale graph datasets, the MC sampling itself is very inefficient [9]; and performing such sampling for $|C|$ times, that is, one for each edge-color causes scalability bottleneck.

3.2 Iterative Hill-Climbing: Second Baseline

Our iterative hill-climbing baseline approach attempts at solving the accuracy bottleneck of the individual top-k algorithm. At each iteration of our hill-climbing algorithm, we add the color c^* to C_1 that maximizes the marginal gain in terms of reliability given the partial set C_1, which was already computed in the previous iterations. Formally,

$$c^* = \arg\max_{c \in C \setminus C_1} [R_{C_1 \cup \{c\}}(s,t) - R_{C_1}(s,t)] \quad (4)$$

We perform k iterations to identify the top-k reliable color set.

Time Complexity. The time complexity of each iteration of our hill-climbing algorithm is $\mathcal{O}(|C|K(n+e))$. Since, we require total k iterations, the overall complexity of our second baseline is $\mathcal{O}(|C|kK(n+e))$.

Difficulties. The iterative hill-climbing method also suffers from both accuracy and efficiency issues.

- Our second baseline performs MC sampling over the entire graph for $|C|k$ times. Hence, this is even slower than our first baseline method.
- Although the iterative hill-climbing algorithm partially solves the accuracy issue of our first baseline, the issue is still present in the initial phases of the algorithm. For example, in Figure 1(a), the individual reliability attained by each of the three colors is 0. Therefore, in the first iteration of our hill-climbing method, it will perform a random selection. If our algorithm selects c_3 as the first color in C_1, then the second selected color would be c_1. One may note that the top-2 reliable color set is $\{c_1, c_2\}$, while the iterative hill-climbing may find the set $\{c_1, c_3\}$, which is a sub-optimal choice. We refer to this issue as the *"cold-start"* problem.

3.3 Most-Reliable-Path based Heuristic

We finally introduce our most-reliable-path based heuristic approach that eliminates the efficiency bottleneck of the two baselines. We follow a two-step approach as discussed below.

Most Reliable Paths Selection. Given an uncertain, edge-colored graph $\mathcal{G} = (V, E, L, P)$, a source $s \in V$, and a destination $t \in V$, we first convert $\mathcal{G}$ into an edge-colored, uncertain, multi-graph $\mathcal{G}'$ as follows. For each edge (u,v) in $\mathcal{G}$, if the edge-color set $C(u,v) = \{c_1, c_2, \ldots, c_i\}$ has total i colors, we add i edges $\{e_1, e_2, \ldots, e_i\}$, with colors $c_1, c_2, \ldots, c_i$, respectively, between u and v in the multi-graph $\mathcal{G}'$. Each newly constructed edge e_i is assigned a probability: $P(e_i) = P((u,v)|c_i)$. One may note that $\mathcal{G}$ and $\mathcal{G}'$ are equivalent in terms of our problem. Next, we select the top-r most reliable paths from source s to destination t in $\mathcal{G}'$, where the reliability of a path is defined as the product of the edge-probabilities along that path. The main intuition behind selecting the top-r most reliable paths is that the reliability between two nodes can often be approximated well by a collection of the top-r most-reliable paths between those two nodes [4]. The value of the parameter r is determined empirically, such that the inclusion of the top-$(r+1)$-th reliable path does not significantly increase the reliability from s to t that was already achieved via the subgraph induced by the top-r most reliable paths.

The top-r most reliable paths from s to t can be obtained by first converting the uncertain, multi-graph $\mathcal{G}'$ into an edge-weighted, multi-graph $\mathcal{G}''$ as follows. Each edge e with probability $P(e)$ in $\mathcal{G}'$ is assigned a weight $\{-\log P(e)\}$ in $\mathcal{G}''$. Therefore, the top-r shortest paths in $\mathcal{G}''$ will be the top-r most reliable paths in $\mathcal{G}'$. We next apply the fastest known algorithm by Eppstein et. al. [5] in order to find the top-r shortest paths (with cycles) in $\mathcal{G}'$, which has time complexity $\mathcal{O}(|C|e + nlogn + r)$. Here, $|C|e$ denotes the maximum possible number of edges in the multi-graph G'.

Iterative Path Inclusion. We formally define our iterative path inclusion problem as follows.

PROBLEM 2 (ITERATIVE PATH INCLUSION). *Given a set $\mathcal{P}$ of the top-r most reliable paths from s to t in $\mathcal{G}'$, find the subset $\mathcal{P}_1 \subseteq \mathcal{P}$, such that the reliability $Rel_{\mathcal{P}_1}(s,t)$ from s to t, via the*

Algorithm 1 Iterative Path Inclusion Algorithm

Require: Top-r most-reliable path set $\mathcal{P}$ between s to t in $\mathcal{G}'$, budget k on the number of colors

Ensure: A subset of paths $\mathcal{P}_1 \subseteq \mathcal{P}$ that maximizes $Rel_{\mathcal{P}_1}(s,t)$, while total no. of edge-colors in $\mathcal{P}_1$ less than k

1: $\mathcal{P}_1 = \phi$
2: **while** total no. of edge-colors in $\mathcal{P}_1$ less than k **do**
3: $P^* = \arg\max_{P \in \mathcal{P} \setminus \mathcal{P}_1} Rel_{\mathcal{P}_1 \cup \{P\}}(s,t)$, such that total no. of edge-colors in $\mathcal{P}_1$ and P^* less than k
4: $\mathcal{P}_1 = \mathcal{P}_1 \cup \{P^*\}$
5: **end while**
6: output $\mathcal{P}_1$

Table 1: Graph Dataset Characteristics

Data Set	# Node	# Edge	# Color	Avg. # Color per Edge	Edge Prob: Mean, SD, Quartiles
Freebase	28 483 132	46 708 421	5 428	1	0.50, 0.24, {0.250, 0.500, 0.750}
BioMine	1 045 414	6 742 943	20	1	0.27, 0.17, {0.116, 0.216, 0.363}
Flixter	29 357	280 517	10	4	0.17, 0.26, {0.003, 0.056,0.212}

Table 2: Avg. Reliability and Efficiency over Datasets; Top-k=5

	Reliability			Running Time (Sec)		
Datasets	Base1	Base2	Rel-Path	Base1	Base2	Rel-Path
Freebase	0.21	0.21	**0.22**	104.9	1278.0	**0.6**
BioMine	0.21	**0.38**	0.35	4240.7	341960.5	**27.5**
Flixter	0.29	**0.62**	0.53	1.3	15355.6	**0.99**

subgraph induced by the paths in $\mathcal{P}_1$, is maximized; while the total number of colors on the edges of paths in $\mathcal{P}_1$ does not exceed k.

$$\arg\max_{\mathcal{P}_1 \subseteq \mathcal{P}} Rel_{\mathcal{P}_1}(s,t)$$
$$\textit{such that} \quad |\cup_{e \in \mathcal{P}_1} C(e)| \leq k \tag{5}$$

Unfortunately, our iterative path inclusion problem is **NP**-hard; and it is neither sub-modular, nor super-modular with respect to the inclusion of paths. We omit the details of the proof due to limitation of space. Next, we design an efficient heuristic algorithm (Algorithm 1) for the iterative path inclusion problem.

Our heuristic procedure works in successive iterations. At each iteration, we add the path P^* to $\mathcal{P}_1$ that maximizes the marginal gain in terms of reliability given the partial set $\mathcal{P}_1$, that was already computed in the previous iterations. While selecting the path P^* in the current iteration, we also ensure that the total number of colors used in the paths $\mathcal{P}_1 \cup \{P^*\}$ is no more than k. Finally, we terminate our algorithm either when there is no path left in the top-r most reliable path set $\mathcal{P}$, or we cannot include any more paths without violating the overall edge-color budget k. We report the edge-colors present in $\mathcal{P}_1$ as our final solution. Also, if the total number of edge-colors present in $\mathcal{P}_1$ is $k' < k$, then we select uniformly at random additional $k - k'$ colors that are not in $\mathcal{P}_1$. We report all of them, along with the edge-colors in $\mathcal{P}_1$, as our output for the top-k reliable color set problem.

Time Complexity. Let us denote by n' and e' the number of nodes and edges in the subgraph induced by the top-r most-reliable path set $\mathcal{P}$. Our iterative path selection algorithm can have at most r iterations. At each iteration, we perform MC sampling for $\mathcal{O}(r)$ times over the subgraph induced by the selected paths. If K is the number of samples used in each MC sampling, the overall time-complexity of our iterative path selection algorithm is $\mathcal{O}(r^2(n' + e')K)$. We here emphasize that the subgraph induced by the top-r most reliable paths is much smaller than the input uncertain graph $\mathcal{G}$. Thus, our iterative path selection technique is more efficient than the two baselines introduced earlier.

4. EXPERIMENTS

Datasets: We summarize our data sets in Table 1. While *BioMine* [10] and *Flixter* [2] have both edge-probabilities and edge-colors, *Freebase* [7] contains only edge-colors. Thus, we assign edge-probabilities in *Freebase* with uniform distribution from $(0, 1)$.

Accuracy and Efficiency: We compare the accuracy and efficiency of our reliable-path based heuristic with two baselines in Tables 2, 3, and 4. Each result is reported as an average over 500 uniformly selected source-destination pairs. The number of top-r paths for our reliable-path based heuristic is set as 20, as increasing it more than that value does not significantly increase the reliability between the source-destination pair. The number of Monte Carlo samples is fixed as 1000 [9]. In all our experiments, we find that the reliable-path based method is several orders of magnitude faster compared to the second baseline, while it still achieves similar reliability from the source to the destination node. Also, the reliability achieved by reliability-path based method and by the second baseline is much higher than that of the first baseline.

Table 3: Avg. Reliability and Efficiency with Varying Top-k, *Freebase*

	Reliability			Running Time (Sec)		
Top-k	Base1	Base2	Rel-Path	Base1	Base2	Rel-Path
5	0.21	0.21	**0.22**	104.9	1278.0	**0.6**
10	0.21	0.21	**0.23**	116.6	2560.2	**0.6**
15	0.21	0.21	**0.23**	120.0	3835.0	**0.6**
20	0.21	0.21	**0.23**	139.9	5112.0	**0.7**

Table 4: Avg. Reliability and Efficiency with Varying Distance from Source to Destination, *BioMine*, Top-k=5

Distance	Reliability			Running Time (Sec)		
(# hop)	Base1	Base2	Rel-Path	Base1	Base2	Rel-Path
2	0.30	**0.53**	0.46	3615.0	283621.0	**19.1**
4	0.12	**0.23**	**0.23**	4865.0	400300.0	**35.8**

5. CONCLUSIONS

We study the novel problem of finding the top-k edge-colors set that maximizes the reliability from a source to a destination node in an uncertain graph. Our proposed reliable-path based heuristic is several orders of magnitude faster than various naïve baselines, while it also achieves comparable accuracy to that of our most effective baseline method. In future work, we shall consider the problem with a set of source and destination nodes.

6. REFERENCES

[1] M. O. Ball. Computational Complexity of Network Reliability Analysis: An Overview. *IEEE Tran. on Reliability*, 1986.

[2] N. Barbieri, F. Bonchi, and G. Manco. Topic-Aware Social Influence Propagation Models. In *ICDM*, 2012.

[3] M. Chen, Y. Gu, Y. Bao, and G. Yu. Label and Distance-Constraint Reachability Queries in Uncertain Graphs. In *DASFAA*, 2014.

[4] W. Chen, C. Wang, and Y. Wang. Scalable Influence Maximization for Prevalent Viral Marketing in Large-Scale Social Networks. In *KDD*, 2010.

[5] D. Eppstein. Finding the k Shortest Paths. *SIAM J. Comput*, 28(2):652–673, 1998.

[6] G. S. Fishman. A Comparison of Four Monte Carlo Methods for Estimating the Probability of s-t Connectedness. *IEEE Tran. Rel.*, 1986.

[7] N. Jayaram, A. Khan, C. Li, X. Yan, and R. Elmasri. Querying Knowledge Graphs by Example Entity Tuples. *TKDE*, 2015.

[8] R. Jin, H. Hong, H. Wang, N. Ruan, and Y. Xiang. Computing Label-Constraint Reachability in Graph Databases. In *SIGMOD*, 2010.

[9] R. Jin, L. Liu, B. Ding, and H. Wang. Distance-Constraint Reachability Computation in Uncertain Graphs. *PVLDB*, 2011.

[10] P. Sevon, L. Eronen, P. Hintsanen, K. Kulovesi, and H. Toivonen. Link Discovery in Graphs Derived from Biological Databases. In *DILS*, 2006.