Advanced Science and Technology Letters
Vol.31 (MulGraB 2013), pp.180-183
http://dx.doi.org/10.14257/astl.2013.31.40

Nearest Neighbor Query in Fuzzy Graph

Gao Jun[1]

[1] Harbin University of Science and Technology, Heilonjiang, China

Abstract. Large uncertain networks arise in various domains. K-nearest-neighbor query is an important query in these networks. In this query, it needs a measure of the distance between any two nodes in the network. To that end, we model uncertain network as fuzzy graph and define the credible shortest path distance and the credible shortest path expected distance between two nodes in the fuzzy graph. Based on credible distance, we put forward fuzzy graph credible k-nearest-neighbor query concepts and the algorithm under the condition of constraint network distance. Theoretical analysis and experimental results show that our approximation algorithm in the fuzzy graph can well solve k-nearest-neighbor query problem under the uncertain network.

Keywords: Uncertain network, Fuzzy graph, Credible distance, Credible k-nearest- neighbor

1 Introduction

Uncertainty of the relationship between entities exists in the reality. The nearest neighbor query result according to the shortest path distance between nodes does not guarantee that is the most effective nearest neighbor. So it needs to consider the nearest neighbor query under the uncertain network.

For this question, there exists literature concerned: [1] analyzes the uncertainty in social network, [2] presents using *k*-nearest-neighbor method to solve the problem of probability path, [3] indicates that the biological domain also have such query require, and [4] gives the method of processing nearest neighbor query in uncertain network using random theory. Those works all use probability to describe the uncertainty of the relationship between entities. But the uncertainty of the relationship between entities sometimes characterized by fuzziness in the reality.

This article explores using fuzzy set theory to deal with the nearest neighbor query under the uncertain network in which the entity is clear and the relationship between entities is fuzziness.

2 Basic Concepts

Definition1. Fuzzy Graph is $\tilde{G}$ =(*V*, *E*, *W*, *Cr*), in which *V*, *E*, *W*, *Cr* respectively represents the vertex set, the edge set, the weight set of edge and the credibility set of

ISSN: 2287-1233 ASTL

edge. The $w(e)$ and the $c(e)$ respectively denote the weight and the credibility of edge e. $c(e)>0$ if and only if $e\in E$.

Definition2. Sample Graph G is an instance of fuzzy graph $\tilde{G}$. E_G is its edge set. Its credibility is $cr(G)=\prod_{e\in E_G} cr(e)$.

Given fuzzy graph $\tilde{G} = (V, E, W, Cr)$, and any pair of nodes $(v_i, v_j)\in V\times V$. G is a sample graph of fuzzy graph $\tilde{G}$.

Definition3. $d_G(v_i, v_j)=d$ denote that shortest-path distance between v_i and v_j in sample graph G is d. E_d is its edge set. Its credibility $cr(d_G(v_i , v_j)=d)=\prod_{e\in E_d} cr(e)$.

Definition4. $d(v_i, v_j)=d$ denote that shortest-path distance between v_i and v_j in fuzzy graph $\tilde{G}$ is d. Its credibility $cr(d(v_i, v_j)=d)=\sum_G cr(d_G(v_i, v_j)=d)$.

Definition5. Credible shortest-path distance between v_i and v_j in fuzzy graph $\tilde{G}$ is $d_C(v_i, v_j)=\arg\max_d cr(d(v_i,v_j) = d)$.

Definition6. Credible shortest-path expectation distance between v_i and v_j in fuzzy graph is $d_{CE}(v_i, v_j)=\sum_d d\times cr(d(v_i, v_j) = d)$.

3 Calculation of credible distance

The method of computing the credible distance is to approximate it using fuzzy theory and fuzzy simulation in combination. Use the following process:

1. Sample r graphs in fuzzy graph. Get the approximation to credibility of the shortest-path distance between v_i and v_j in the fuzzy graph. That is $cr_r(d(v_i, v_j)=d)=\sum_{l=1}^{r} cr(d_{G_l}(v_i,v_j)=d)$.

2. Sample $r+\Delta r$ graphs in fuzzy graph.. Get the approximation to credibility of the shortest-path distance between v_i and v_j in the fuzzy graph. That is $cr_{r+\Delta r}(d(v_i, v_j)=d)=\sum_{l=1}^{r+\Delta r} cr(d_{G_l}(v_i,v_j)=d)$

3. If given $\varepsilon > 0$, such that $|cr_{r+\Delta r}(d(v_i, v_j)=d)-cr_r(d(v_i, v_j)=d)|\ \varepsilon$ holds. Then $cr_{r+\Delta r}(d(v_i, v_j)=d)$ is the approximation to credibility of the shortest-path distance between v_i and v_j in the fuzzy graph. Else $r= r +\Delta r$, go to 2.

4. Compare the credibility of the shortest-path distance between v_i and v_j in the fuzzy graph. The distance with maximum approximation of credibility is an approximation to the shortest-path distance between v_i and v_j in the fuzzy graph.

Theorem: Let the credibility of the shortest-path distance between two nodes is convergent in a fuzzy graph. Then given any $\varepsilon > 0$, as $r \rightarrow$ co such that | $cr_{r+□r}$ $(d(v_i, v_j)=d)$-$cr_r(d(v_i, v_j)=d)$| ε holds.

∑ cr(d (v , v) = d) .

G ij

I

Proof: Let the credibility of the shortest-path distance between two nodes in the fuzzy graph tend to limit c. That is $\lim_{r□□} cr_r(d(v_i, v_j)=d)=c$. Given any $\sum 2 > 0$, as r □ □ such that$_($ $cr_r(d(v_i, v_j)=d)-c| \delta \sum 2$ holds, we have $_{crr+□r}$ $(d(v_i, v_j)=d)-cr_r(d(v_i, v_j)=d)|=|\{$ $_{crr+□r}$ $(dv_i, v_j)=d)-c\}-\{cr_r(d(v_i, v_j)=d)-c\}| \delta | cr_{r+□r} (d(v_i, v_j)=d)-c|+|cr_r(d(v_i, v_j) =d)-c|\delta \sum 2 + \sum 2 = \sum$.

End proof.

Same as the calculation of credible shortest-path distance, the approximant of the credible shortest path expectation distance can also be obtained.

4 Query of the credible *k*-nearest neighbor

Definition10. Query of the credible *k*-nearest neighbor is to find the set of *k* nodes $k\text{-}nn_C=\{ {}_1\ v_{i_2} ,v_{\ i} ,..., v_{\ i_k}\}$ for which the distance $d_C(q, v_{im})$ (or $d_{CE}(q, v_{im})$) is less or equal to the distance $d_C(q, v)$ (or $d_{CE}(q, v)$) for any other $v□V/k\text{-}nn_c$.

Credible *k*-nearest neighbor query algorithm (based on $d_C(q, v)$):

Input: Fuzzy graph $\tilde{G} = (V, E, W, Cr)$, query object $q□V$, the number of sample r, the increment number of sample □ r, the number of nearest neighbor k, distance constraint D, precision $\sum$;

Output: $k\text{-}nn_c$;

```
1. $k\text{-}nn_c$ = □ ;
2.let $d_{C\text{-}max}$ is the maximum of credible shortest-path distance between $q$ and the node in the $k\text{-}nn_c$;
3.start from $q$ perform the Dijkstra algorithm on $\tilde{G}$ until we visit a node whose distance exceeds $D$;
4.for each node $v$ □ $V$ visited do
5.    $QUEUE(v)=<(d(q, v),cr(d(q, v)))>$
6.    sample $r$ graphs;
7.    for each sample graph do
        compute $d(q, v)=d$ $cr(d(q, v)=d)=$ $\underset{e□E_d}{□} cr(e)$ ;
8.        if $d$ is not in $QUEUE(v)$ then en-queue $(d(q, v)=d\ cr(d(q, v)=d))$;
9.        else accumulate $cr(d(q, v)=d)$;
10.sample $r$+ □ $r$ graphs and repeat 7-9;
11. if | $cr_{r+□r}(d(q, v)=d)$- $cr_r(d(q, v)=d)$ |< $\sum$ then $cr(d(q, v)=d)=$ $_{crr+□r}$ $(d(q, v)=d)$;
12. else $r$←$r$+ □ $r$ go 6;
13. $d_C(q,v)=\arg\max_d cr(d(q, v)=d)$;
14. if | $k\text{-}nn_c$ |<$k$ then $k\text{-}nn_c$ = $k\text{-}nn_c$ ∗ {$v$}
15. else if $d_C(q,v)$< $d_{C\text{-}max}$ then
16.            $k\text{-}nn_c$ = $k\text{-}nn_c$ ∗ {v};
17.return $k\text{-}nn_c$;
```

5 Experimental evaluation

We implemented all our methods in Visual C++6.0. All the experiments are run on a Windows XP with 2.0GHz dual-core processors and 1GB of memory. We experimented on two different datasets, one coming from real-world road network, the other being synthetic network. Experiment is divided into two parts: the first testing the characteristic of the credible distance, the second testing the influence of various parameters on the nearest neighbor query algorithm based on the credible distance.

6 Conclusion

To solving the problem of nearest neighbor query in uncertain network, we give a new method. At first the uncertain network is modeled as a fuzzy graph. Then we give the definition of credible shortest-path distance and credible shortest-path expectation distance in fuzzy graph, based on credible distance, put forward the concept of credible nearest neighbor query. Aim for the problem of nearest neighbor query with distance constraint in uncertain network, we give the credible nearest neighbor query algorithm. The approximate result is got by using the sample method, and its precision can be adjusted according to the need of the actuality. The algorithm makes the problem that the time complexity is exponential solved in polynomial time. Theoretical analysis and experimental results show that the algorithm is feasible and has stable performance. The further research is that the randomness and fuzziness of the uncertain network are considered at the same time, in order to improve the ability of describing the uncertain network, so as to make the nearest neighbor query more efficient in uncertain network.

Acknowledgments. We are grateful to Digital Chart of World for the network dataset.

References

1. Adar, E., Re, C.: Managing Uncertainty in Social Networks. J. IEEE Data Eng. Bull. Vol. 30, no. 2, 15--22 (2007)
2. Ghosh, J., Ngo, H. Q., Yoon, S., Qiao, C.: On a Routing Problem within Probabilistic Graphs and Its Application to Intermittently Connected Networks. In: 26th IEEE International Conference on Computer Communications, pp. 1721--1729. Anchorage, AK (2007)
3. Asthana, S., King, O. D., Gibbons, F. D., Roth, F. P.: Predicting Protein Complex Membership Using Probabilistic Network Reliability. J. Genome Research, Vol. 14, 1170--1175(2004)
4. Potamias, M., Bonchi, F., Gionis, A.: Nearest-neighbor Queries in Probabilistic Graphs. http://www.cs.bu.edu