EE447 Mobile Internet HW 1: The Optimal Connectivity Determination in Uncertain Graphs

Name: Hongjie Fang   Student ID:518030910150   Email: galaxies@sjtu.edu.cn

Determining the source-destination connectivity in uncertain graphs has wide applications in real life, e.g., routing detection, information diffusion control, etc. For an uncertain graph, each edge exists independently with some probability, and the existence of each edge can be unraveled through edge testing with a certain cost. Our goal is to determine whether source s and destination t are connected or not with the minimum expected cost.

Given the following several kinds of uncertain graphs, please provide the corresponding optimal strategy with the minimum expected cost, and prove its optimality, respectively.

1. For the pyramid graph in Figure 1, each edge exists with probability p, and the cost of testing the existence of each edge is 1.
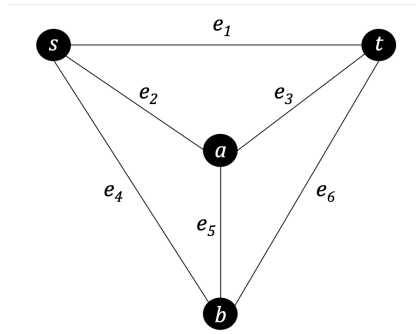

Figure 1: Pyramid Graph

Solution. We number the vertices and edges for convenience of explanations, as Figure 1 shown. Then we make the following clarifications.

(a) To prove s is connected with t, we only need to find one available path.

(b) To prove s is disconnected with t, every possible path between s and t should be unavailable. Specially, in the pyramid graph, every edge may contribute to the connectivity between s and t, so every edge should be checked to determine that every path between s and t is unavailable, $i.e.$, s is disconnected with t.

There are totally 5 paths between s and t, listed in Table 1.

Table 1: The paths between s and t and their probabilities

Path Number	P_1	P_2	P_3	P_4	P_5
Path	$s \xrightarrow{e_1} t$	$s \xrightarrow{e_2} a \xrightarrow{e_3} t$	$s \xrightarrow{e_4} b \xrightarrow{e_6} t$	$s \xrightarrow{e_2} a \xrightarrow{e_5} b \xrightarrow{e_6} t$	$s \xrightarrow{e_4} b \xrightarrow{e_5} a \xrightarrow{e_3} t$

According to the optimal policy which is introduced in the class [1], obeying the following rules makes the strategy optimal.

Rule 1. Test if direct edge exists between s and t;

Rule 2. List all paths connecting s and t with the minimum number of potential edges, and find the set M that contains the minimum potential cut between s and t.

Rule 3. Sharpen Rule 2 by specifying which particular edge in M should be tested, *i.e.*, test any edges in M connecting s and t.

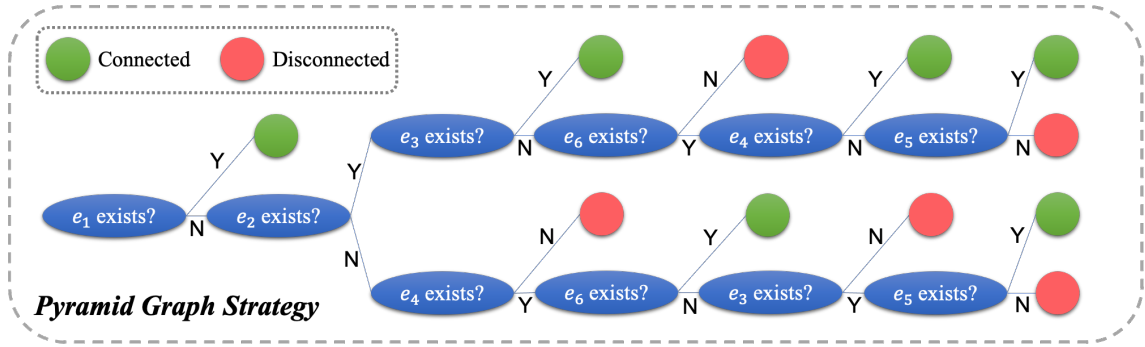To apply those rules in this situation, we can derive the strategy shown in Figure 2.

Figure 2: The Strategy for Pyramid Graph

Optimality Explanation. Our strategy obey the previous rules, according to the theorem of the paper[1], our strategy is optimal with minimum expected cost as follows.

$$E = p + 3(1-p)p^2 + 3(1-p)^3 + 4(1-p)^3p + 4(1-p)^2p^2 + 5(1-p)^2p^3 + 10(1-p)^3p^2 + 5(1-p)^4p$$

$\square$

[1]Luoyi Fu, Xinbing Wang, and P. R. Kumar. "Optimal determination of source-destination connectivity in random graphs." *Proceedings of the 15th ACM international symposium on Mobile ad hoc networking and computing.* 2014.

2. For the graph with m parallel edges between nodes s and t, edges are labeled e_1 through e_m as shown in Figure 3. The probability that edge e_i $(i = 1, 2, \cdots, m)$ exists is p_i. And the test cost of edge e_i is t_i.
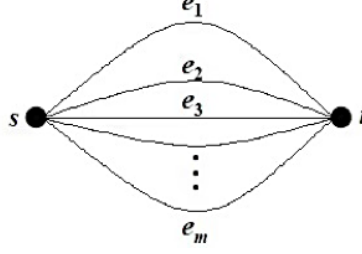

Figure 3: Parallel Graph

Solution. We make the following clarifications.

(a) To prove s is connected with t in the parallel graph, we only need to find one available edge of the parallel graph.

(b) To prove s is disconnected with t in the parallel graph, every edge of the parallel graph should be checked.

Let us define the **revenue** r_i of edge e_i to be p_i/t_i, which indicates that the average percentage of determined connectivity per unit of cost if we check the i-th edge. We present our strategy as follows. Intuitively, we can just sort the edges by their revenue r_i in a descending order, and check them one by one.

Now, let us prove the strategy gives optimal solution with minimum expected cost.

Definition 1 (Static strategy). *A **static strategy** means that we won't change the checking order when new information is collected. Let t_i denotes the order that edge e_i is checked in a strategy. Then, a static strategy can be represented as $\{t_1, t_2, \cdots, t_m\}$.*

Definition 2 (Dynamic strategy). *A **dynamic strategy** means that we may change the checking order when new information is collected.*
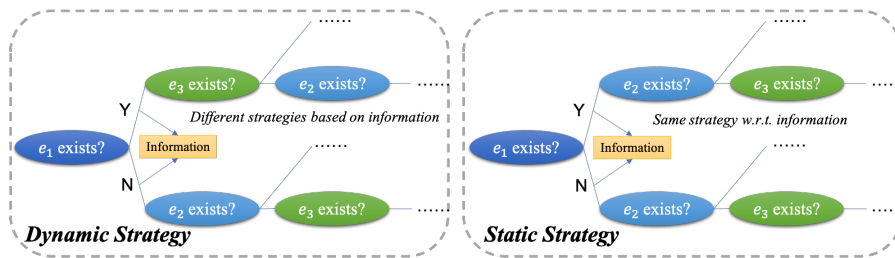

Figure 4: Dynamic Strategy and Static Strategy

Lemma 1. *In parallel graph, there is always a static strategy that is not worse than a given dynamic strategy.*

Proof. According to the clarifications, if our new-collected information is that an edge exists, then the whole process is finished, we don't need to change the following order. If the new-collected information is that an edge does not exist, then we can direct apply the order of dynamic strategy without waiting for the collected "information" and get a static strategy with the same quality. Hence, we can always transform a dynamic strategy to a static strategy that is not worse than dynamic one. □

Lemma 2. *Our strategy is a static strategy.*

Proof. We won't change the checking order during the checking process, hence our strategy is a static strategy. □

Lemma 1 indicates that we only need to focus on the statis strategies, since dynamic strategies can always be replaced with static strategies. Therefore, in the following discussions, the word "strategy" stands for static strategy by default.

Definition 3 (Ordered pair). *Given a strategy, if edge e_i is checked before edge e_j in the strategy and the revenue of e_i is less than the revenue of e_j, i.e., $t_i < t_j$ and $r_i < r_j$, then (e_i, e_j) is an **ordered pair**.*

Lemma 3. *Our strategy has no ordered pair.*

Proof. Since we sort the edges by their revenue r_i in a descending order and check them according to this order, no ordered pair will be produced in this process. □

Definition 4 (Consecutive ordered pair). *Given a strategy, an ordered pair (e_i, e_j) of the strategy is a **consecutive ordered pair** if and only if $t_i + 1 = t_j$.*

Lemma 4. *Any strategy that has an ordered pair must have a consecutive ordered pair.*

Proof. Assume (e_i, e_j) is an ordered pair, *i.e.*, $r_i < r_j$. then $(e_i, e_{i+1}), \cdots, (e_{j-1}, e_j)$ must contain at least one consecutive pair; otherwise, $r_i \geq r_{i+1} \geq \cdots \geq r_j$ will derive a contradiction with the fact that $r_i < r_j$. □

Lemma 5. *Given a strategy with consecutive ordered pairs, we can swap the edge order of a consecutive ordered pair and get a better strategy. In other words, if a strategy $\{t_1, t_2, \cdots, t_{i-1}, t_i, t_{i+1}, \cdots, t_{j-1}, t_j, t_{j+1}, \cdots, t_m\}$ has a consecutive ordered pair (e_i, e_j), then strategy $\{t_1, t_2, \cdots, t_{i-1}, t_j, t_{i+1}, \cdots, t_{j-1}, t_i, t_{j+1}, \cdots, t_m\}$ is better than the original strategy.*

Proof. The expected cost of the original strategy and the new strategy are as follows:

$$E_o = \cdots (t_i + (1 - p_i)(t_j + (1 - p_j)(\cdots)))$$
$$E_n = \cdots (t_j + (1 - p_j)(t_i + (1 - p_i)(\cdots)))$$

Hence,

$$E_n - E_o = t_j + (1 - p_j)t_i - t_i - (1 - p_i)t_j = p_i t_j - p_j t_i$$

Notice that, since (e_i, e_j) is a consecutive ordered pair, then $r_i < r_j$, *i.e.*, $p_i/t_i < p_j/t_j$. Since $t_i, t_j > 0$, we have $E_n - E_o = p_i t_j - p_j t_i < 0$, *i.e.* $E_n < E_o$. Therefore, our new strategy is better than the original strategy in the aspect of expected cost. $\square$

Theorem 6. *The strategy with no ordered pair is the optimal strategy with lowest expected cost.*

Proof. Suppose another strategy with at least one ordered pair is optimal, *i.e.*, better than the strategy with no ordered pair. Then according to Lemma 4, the strategy must include at least one consecutive ordered pair. Then from Lemma 5 we know that, swapping the consecutive ordered pair can form a better strategy, which contradicts with the optimality of the strategy. Hence, the strategy with no ordered pair is the optimal strategy with lowest expected cost. $\square$

Theorem 7. *Our strategy is the optimal strategy with lowest expected cost.*

Proof. From Lemma 3, our strategy has no ordered pair. According to Theorem 6, the strategy with no ordered pair is the optimal strategy with lowest expected cost. Then, our strategy is the optimal strategy with lowest expected cost. $\square$

Theorem 7 shows that our strategy is the optimal strategy with lowest expected cost. $\square$

3. **(Optional, with additional bonus up to 2% of your final course score.)** For the series-parallel graph consisting of n parallel graph labeled $P_{m_1}^1, \cdots, P_{m_n}^n$, arranged in series. Let the i-th edge in the j-th parallel graph be represented as e_{ij} ($1 \leq j \leq n, 1 \leq i \leq m_j$). Edge e_{ij} exists with probability p_{ij} and has test cost c_{ij}.
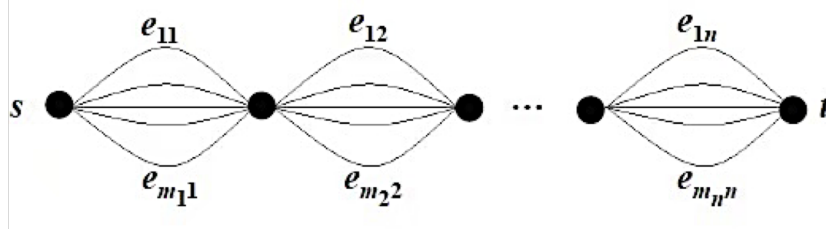

Figure 5: Series-Parallel Graph

Solution. We make the following clarifications.

(a) To prove s is connected with t in the series-parallel graph, every parallel graph should be connected.

(b) To prove s is disconnected with t in the parallel graph, at least one parallel graph should be disconnected.

The series-parallel graph consists of several parallel graph, then according to the proof in Problem 2, we propose the following lemma.

Lemma 8. *The optimal strategy should follows our strategy of Problem 2 in every parallel graph, i.e., checking the edges from highest revenue to lowest revenue.*

Proof. If the optimal strategy does not follows our strategy of Problem 2 in some parallel graphs, we can always swap the consecutive ordered pairs in these parallel graphs to optimize the current strategy and form a better one, which contradicts with the optimality of the strategy. $\square$

We are going to introduce an important lemma in our proof process.

Lemma 9. *Once we check an edge in a parallel graph, the optimal strategy is to check the rest of the edges in this parallel graph, until the connectivity of this parallel graph is determined.*

Proof. If the edge exists in the parallel graph, then the connectivity is already determined, thus the lemma obviously holds. Therefore, we only focus on the situation that the edge does not exist in the parallel graph.

If the graph is finally connected, then the cost remains the same according to the optimal strategy of each parallel graph is fixed, which is proved by Lemma 8.

If the graph is finally disconnected, then given the information that one edge does not exist in current parallel graph, the probability that this parallel graph is disconnected increases, and the cost that this parallel graph is disconnected decreases. To achieve optimal strategy, we need to continue checking edges in this parallel graph.

One may arguing that if the graph is finally disconnected, we can check another parallel graph which can "minimize" the final expected cost. If so, why don't we check this parallel graph first, since the information of the current parallel graph does not affect the connectivity of this parallel graph? This should form a better strategy that also follows this lemma.

Hence, this lemma holds in every situation. $\square$

For j-th parallel graph, if we haven't determined whether this parallel graph is connected or not, we define P_j and C_j as follows.

$$P_j = \prod_{\substack{i \\ e_{ij} \text{ not checked}}} (1 - p_{ij}), \quad C_j = \sum_{\substack{i \\ e_{ij} \text{ not checked}}} c_{ij}$$

Actually, P_j and C_j represent the probability and the cost of the worst case in the j-th parallel graph respectively, *i.e.*, the parallel graph is unconnected. However, for the global series-parallel graph, it is the best case since we do not need to check other edges anymore. Moreover, let D_i denote the expected cost in other situations, that is,

$$D_j = E[\text{cost} \mid \text{the } j\text{-th parallel graph is connected}]$$

Therefore, the expected cost of determining the j-th parallel graph is

$$E_j = P_j C_j + (1 - P_j) D_j$$

Let us define the **global revenue** of the j-th parallel graph as $R_j = D_j - C_j - D_j/P_j$. We present our strategy as follows.

(a) We sort the parallel graphs by their global revenue R_j in a descending order, and check them one by one.

(b) In each parallel graph, we use the strategy presented in Problem 2 to check the connectivity of the parallel graph, *i.e.*, we check edges according to their revenue in the parallel graph.

Now, let us prove the strategy gives optimal solution with minimum expected cost. Similar to Problem 2, we present the following definitions lemmas without proof. The specific proof process can refer to the solution to Problem 2.

Definition 5 (Global static strategy). *A **global static strategy** means that we won't change the checking order of the parallel graph when new information is collected. Let T_i denotes the order that the i-th parallel graph is checked in a strategy. Then, a static strategy can be represented as $\{T_1, T_2, \cdots, T_n\}$.*

Definition 6 (Global dynamic strategy). *A **global dynamic strategy** means that we may change the checking order of the parallel graph when new information is collected.*

Lemma 10. *Our strategy is a global static strategy.*

According to Lemma 9, we can derive the following corollary.

Corollary 11. *The optimal strategy should be a global static strategy.*

Definition 7 (Global ordered pair). *Given a strategy, if the parallel graph G_i is checked before the parallel graph G_j in the strategy and the global revenue of G_i is less than the global revenue of G_j, i.e., $T_i < T_j$ and $R_i < R_j$, then (G_i, G_j) is an **global ordered pair**.*

Lemma 12. *Our strategy has no global ordered pair.*

Definition 8 (Consecutive global ordered pair). *Given a strategy, an ordered pair (G_i, G_j) is a **consecutive ordered pair** if and only if $T_i + 1 = T_j$.*

Lemma 13. *Any strategy that has a global ordered pair must have a consecutive global ordered pair.*

The proof of the previous lemmas and definitions is similar to that in Problem 2. Therefore, we omit the proof. However, we will prove the following important lemma.

Lemma 14. *Given a strategy with consecutive global ordered pairs, we can swap the edge order of a consecutive global ordered pair and get a better strategy, i.e., if a strategy $\{T_1, T_2, \cdots, T_{i-1}, T_i, T_{i+1}, \cdots, T_{j-1}, T_j, T_{j+1}, \cdots, T_n\}$ has a consecutive global ordered pair (G_i, G_j), then strategy $\{T_1, T_2, \cdots, T_{i-1}, T_j, T_{i+1}, \cdots, T_{j-1}, T_i, T_{j+1}, \cdots, T_n\}$ is better than the original strategy.*

Proof. The expected cost of the original strategy and the new strategy are as follows:

$$E_o = \cdots (P_i C_i + (1 - P_i)(D_i + P_j C_j + (1 - P_j)(D_j + \cdots)))$$
$$E_n = \cdots (P_j C_j + (1 - P_j)(D_j + P_i C_i + (1 - P_i)(D_i + \cdots)))$$

Hence,

$$E_n - E_o = P_j C_j + (1 - P_j)(D_j + P_i C_i + (1 - P_i)D_i) - P_i C_i - (1 - P_i)(D_i + P_j C_j + (1 - P_j)D_j)$$

Simplify the previous formula, and we can derive the following result.

$$E_n - E_o = P_i P_j \left(\left(D_i - C_i - \frac{D_i}{C_i} \right) - \left(D_j - C_j - \frac{D_j}{P_j} \right) \right) = P_i P_j (R_i - R_j)$$

Notice that, since (G_i, G_j) is a consecutive ordered pair, then $R_i < R_j$. Therefore, we have $E_n - E_o = P_i P_j (R_i - R_j) < 0$, *i.e.* $E_n < E_o$. Therefore, our new strategy is better than the original strategy in the aspect of expected cost. $\square$

Hence, it is easy to prove the following theorems.

Theorem 15. *The strategy with no global ordered pair is the optimal strategy with lowest expected cost.*

Proof. Suppose another strategy with at least one global ordered pair is optimal, *i.e.*, better than the strategy with no ordered pair. Then according to Lemma 13, the strategy must include at least one consecutive global ordered pair. Then from Lemma 14 we know that, swapping the consecutive global ordered pair can form a better strategy, which contradicts with the optimality of the strategy. Hence, the strategy with no orderd pair is the optimal strategy with lowest expected cost. $\square$

Theorem 16. *Our strategy is the optimal strategy with lowest expected cost*

Proof. From Lemma 12, our strategy has no global ordered pair. According to Theorem 15, the strategy with no global ordered pair is the optimal strategy with lowest expected cost. Then, our strategy is the optimal strategy with lowest expected cost. $\square$

Theorem 16 shows that our strategy is the optimal strategy with lowest expected cost. $\square$