On Distribution Function of the Diameter in Uncertain Graph

Yuan Gao¹, Lixing Yang¹, Shukai Li¹, Samarjit Kar²

¹State Key Laboratory of Rail Traffic Control and Safety, Beijing Jiaotong University
Beijing 100044, China

²Department of Mathematics, National Institute of Technology Durgapur
Durgapur 713209, India

Abstract

In uncertain graphs, the existence of some edges is not predetermined. The diameter of an uncertain graph is essentially an uncertain variable, which indicates the suitability for investigation of its distribution function. The main focus of this paper is to propose an algorithm to determine the distribution function of the diameter of an uncertain graph. We first discuss the characteristics of the uncertain diameter, and the the distribution function is derived. An efficient algorithm is designed based on Floyd’s algorithm. Further, some numerical examples are illustrated to show the efficiency and application of the algorithm.

Keywords: Uncertainty modelling; Graph theory; Uncertain graph; Distribution function; Floyd’s algorithm

1 Introduction

To the best of our knowledge, in the classical graph theory, the edges and vertices are predetermined. Theoretical problems on graph theory are concerned with connectivity, nature of the graph and determination of diameter. To solve these problems, a variety of efficient algorithms have been proposed over the last decades and successfully applied to many real-world problems, such as transportation, communications, supply chain management, etc. More extensive information on classical graph theory can be found in [4, 40].

In practice, indeterminacy is inevitable due to the lack of information. The classical algorithms appear to be very difficult to apply directly the indeterminacy in respect of vertices and edges. In this paper, we have studied the existence of some nondeterministic edges. The existence of such nondeterministic graphs are used to represent relationship network of a group. As an example, Figure 1 depicts “confirmed friends” which represents completely confirmed relationships and “potential friends” represents relationships that are inferred from other information, such as personal data and friends list. In other words, in a nondeterministic graph describing relationship network, “potential friends” are represented by nondeterministic edges. It is shown how an existence chance \( \alpha_{ij} \) is associated to and nondeterministic edge \( e_{ij} \).

To deal with nondeterministic graphs, some researchers introduced probability theory and developed random graphs. Erdős & Réyi [12, 13] first used probability theory for nondeterministic graphs. Moreover, Gilbert [15] in 1959 dealt independently E-R random graphs. Usually, an E-R random graph is obtained by starting with a set of \( n \) isolated vertices and adding successive edges between them with probability

*Corresponding Author: gaoyuan@bjtu.edu.cn
In 1998, Watts and Strogatz [39] proposed the concept of “small world” network between the deterministic graph and E-R random graph, which emphasized the cluster phenomenon of social networks. Similar researches can also be found in literature [28, 29, 34]. The motivation of these researches leads to qualitatively explain phenomena in networks, and the probability is more like a theoretical parameter.

As in relationship networks, the existence of chance of edges is sometimes evaluated by experts, based on which the structure of the nondeterministic graph is analyzed and decisions are then made. It has been emphasized by [23, 37] that the expert data tend to put too much weight on unlikely events. In other words, it is unsuitable to use probability to handle nondeterministic information with expert data. Motivated by this point, some non-probabilistic methods have been proposed and developed since 1960s, such as fuzzy mathematics (Dubois&Prade [11], Zadeh [42, 43]), rough sets theory (Gong et al. [20], Pawlak [30]), evidence theory (Dempster [8]), and methods of Granular Computing (Pedrycz [31], Bargiela & Pedrycz [2]).

These methods are also employed to handle nondeterministic graphs with expert data, for instance fuzzy graphs (Bhattacharya [3], Rosenfeld [33]) and rough graphs (Chen & Li [7], He et al. [22]).

Uncertainty theory, proposed in 2007 and refined in 2010 by Liu [24, 26], is another efficient tool to handle nondeterministic information with expert data. In recent years, uncertainty theory has been widely used in the field of operational research, such as network optimization [17, 19, 21, 44], inventory problem [32, 18, 9, 10] and transportation problem [35, 36], etc. In 2013, Gao & Gao [16] first introduced uncertainty theory into graph theory, and proposed a concept of uncertain graph, in which the existence of chance of edges is described by uncertain measure. In their paper, the connectedness index, which measures the chance that an uncertain graph is connected, was first proposed. To calculate the connectedness index, two algorithms were designed, which were derived from the Kruskal algorithm and Prim algorithm respectively.

To the best of our knowledge, so far there is currently no related study on the diameter of uncertain graphs in literature. In view of this fact, this paper focuses on investigating the diameter of uncertain graphs. Note that the diameter of uncertain graph is an uncertain variable due to the existence of uncertain edges. It is thus an meaningful issue to explore how to formulate and compute the distribution function of the diameter in an uncertain graph. In this paper, we first study the characteristics of diameter in an uncertain graph, and then obtain the corresponding distribution function. Moreover, we explicitly design an algorithm derived from the Floyd’s algorithm [14] to calculate the distribution function. The efficiency of the algorithm is finally shown theoretically and experimentally.

The remainder of this paper is organized as follows. In Section 2, uncertainty theory is introduced briefly for the completeness of this research. Section 3 gives the problem descriptions and defines the concept of the diameter in uncertain graphs. In Section 4, the distribution function of the diameter of uncertain graph is obtained. In Section 5, an efficient algorithm for calculating the distribution function is proposed and illustrated with some numerical examples. Section 6 concludes this paper with a brief summary.
2 Uncertainty Theory

For the completeness of this study, we introduce here some basic properties of uncertainty theory.

Let $\Gamma$ be a nonempty set, and $\mathcal{L}$ a $\sigma$-algebra over $\Gamma$. Each element $\Lambda \in \mathcal{L}$ is assigned a number $M\{\Lambda\}$. In order to ensure that the number $M\{\Lambda\}$ has certain mathematical properties, Liu [24] presented the following three axioms:

Axiom 1. (Normality) $M\{\Gamma\} = 1$.

Axiom 2. (Duality) $M\{\Lambda\} + M\{\Lambda^c\} = 1$ for any event $\Lambda$.

Axiom 3. (Subadditivity) For every countable sequence of events $\{\Lambda_i\}$, we have

$$M\left(\bigcup_{i=1}^{\infty} \Lambda_i\right) \leq \sum_{i=1}^{\infty} M\{\Lambda_i\}.$$

Definition 1. (Liu [24]) The set function $M$ is called an uncertain measure if it satisfies the normality, self-duality, and countable subadditivity axioms. The triplet $(\Gamma, \mathcal{L}, M)$ is called uncertain space.

Uncertain measure $M\{\Lambda\}$ is the chance that event $\Lambda$ occurs. Product uncertain measure was studied by Liu [25] in 2009, which leads to the product measure axiom.

Axiom 4. (Liu [25] Product Axiom) Let $(\Gamma_i, \mathcal{L}_i, M_i)$ be uncertainty spaces for $i = 1, 2, \cdots$. The product uncertain measure $M$ is an uncertain measure satisfying

$$M\left(\prod_{i=1}^{\infty} \Lambda_i\right) = \bigwedge_{i=1}^{\infty} M_i\{\Lambda_i\},$$

where $\Lambda_i$ are arbitrarily chosen events from $\mathcal{L}_k$ for $k = 1, 2, \cdots$, respectively.

Definition 2. (Liu [24]) An uncertain variable is a measurable function $\xi$ from an uncertainty space $(\Gamma, \mathcal{L}, M)$ to the set of real numbers, i.e., for any Borel set $B$ of real numbers, the set

$$\{\xi \in B\} = \{\gamma \in \Gamma \mid \xi(\gamma) \in B\}$$

is an event.

In an uncertain space $(\Gamma, \mathcal{L}, M)$, let $\Gamma = \{\gamma_1, \gamma_2\}$, and $\mathcal{L}$ be the power set of $\Gamma$. Assume that $M\{\gamma_1\} = \alpha$, $M\{\gamma_2\} = 1 - \alpha$. Define

$$\xi(\gamma) = \begin{cases} 1, & \text{if } \gamma = \gamma_1 \\ 0, & \text{if } \gamma = \gamma_2. \end{cases}$$

Then $\xi$ is an uncertain variable. Moreover, $\xi$ is a boolean uncertain variable.

The uncertainty distribution function of an uncertain variable $\xi$ is defined by $\Phi(x) = M\{\xi \leq x\}$.

Definition 3. (Liu [24]) The uncertain variables $\xi_1, \xi_2, \cdots, \xi_n$ are said to be independent if

$$M\left(\bigcap_{i=1}^{n} \{\xi_i \in B_i\}\right) = \min_{1 \leq i \leq n} M\{\xi_i \in B_i\}$$

for any Borel sets $B_1, B_2, \cdots, B_n$ of real numbers.

The following theorem is the operational law of uncertain variables.
Theorem 1. (Liu [25], Operational Law) Let \( \xi_1, \xi_2, \cdots, \xi_m \) be independent uncertain variables, and \( f : \mathbb{R}^m \to \mathbb{R} \) a measurable function. Then \( \xi = f(\xi_1, \xi_2, \cdots, \xi_m) \) is an uncertain variable such that

\[
\mathcal{M}\{\xi \in B\} = \begin{cases} 
\sup_{f(B_1, B_2, \cdots, B_m) \subset B} \min_{1 \leq i \leq m} M_i\{\xi_i \in B_i\}, & \text{if } \min_{1 \leq i \leq m} M_i\{\xi_i \in B_i\} > 0.5 \\
1 - \sup_{f(B_1, B_2, \cdots, B_m) \subset B} \min_{1 \leq i \leq m} M_i\{\xi_i \in B_i\}, & \text{if } \min_{1 \leq i \leq m} M_i\{\xi_i \in B_i\} > 0.5 \\
0.5, & \text{otherwise}
\end{cases}
\]

where \( B, B_1, B_2, \cdots, B_m \) are Borel sets of real numbers.

3 Uncertain Graph and Its Diameter

In this paper, the terminologies associated with deterministic graphs reference from [40]. For illustration, Figure 2 presents two deterministic graphs to show the distance between vertices.

![Figure 2: Deterministic graph](image)

Graph \( G_1 = (V_1, E_1) \) is a \( v_1 - v_5 \) path. The length of \( v_1 - v_5 \) path is 4, which is the number of edges that \( v_1 - v_5 \) path contains. Given vertices \( v_i \) and \( v_j \) in deterministic graph \( G \), the distance \( d(v_i, v_j) \) is the minimal length of a \( v_i - v_j \) path. If there is no such path, then \( d(v_i, v_j) = \infty \).

The diameter is a basic concept in graph theory, which measures the maximal distance between vertices. In other words, the diameter of deterministic graph \( G = (V, E) \) is \( \text{diam}(G) = \max_{v_i, v_j \in V} d(v_i, v_j) \).

If \( G \) is not connected, then \( \text{diam}(G) = +\infty \); if \( G \) is connected and \( |V| = n \), \( \text{diam}(G) \in [1, n - 1] \). In a deterministic graph, since vertices and edges are predetermined, the diameter is a crisp value. In Figure 2, it is easy to verify that \( \text{diam}(G_1) = 4 \) and \( \text{diam}(G_2) = 2 \).

As stated above, computing the diameter of a deterministic graph typically can be regarded as an optimization problem in the corresponding graph. However, it is not the case in uncertain graphs proposed by Gao & Gao [16].

Definition 4. (Gao & Gao, [16]) An uncertain graph \( G \) is a triplet consisting of a vertex set \( V(G) \), an edge set \( E(G) \), and a set of indicator functions \( \xi(G) \) and is denoted by \( G = (V, E, \xi) \).

In other words, in uncertain graph, vertices are predetermined, while edges are uncertain. For convenience, an indicator function \( \xi(e) \), which is a boolean uncertain variable, is used to indicate the existence of corresponding edge. That is, \( \xi = 1 \), if edge \( e \) exists, otherwise it is zero.

Let us assume that \( V(G) = \{v_1, v_2, \cdots, v_n\} \), \( E(G) = \{e_1, e_2, \cdots, e_m\} \) and \( \xi(G) = \{\xi_1, \xi_2, \cdots, \xi_m\} \). Let \( M\{\xi_i = 1\} = \alpha_i \) and \( 1 = \alpha_0 \geq \alpha_1 \geq \alpha_2 \geq \cdots \geq \alpha_m \geq \alpha_{m+1} = 0 \). Figure 3 shows an uncertain graph \( G = (V, E, \xi) \).
For simplicity, uncertain graph $G = (V, E, \xi)$ shown in Figure 3 can also be expressed as shown in Figure 4, where the existence chance $\alpha_i$ for an edge $e_i$ is marked on the corresponding edge.

The uncertain graph $G$ can be considered as a function of $\xi_i$, $i = 1, 2, \cdots, m$, i.e., if all the value of $\xi_i$ are determined, the realization of uncertain graph $G$ is determined; for different values of $\xi_i$, the realization of the uncertain graph $G$ will be different. Note that each realization is a deterministic graph $G_i = (V_i, E_i)$ with $V_i = V(G)$ and $E_i \subset E(G)$. Obviously, if $0 < M\{\xi_i = 1\} = \alpha_i < 1$, $i = 1, 2, \cdots, m$, there are $2^m$ realizations of uncertain graph $G$. Figure 5 shows some realizations of the uncertain graph $G$. It is easy to verify that when $\xi_1 = \xi_2 = \xi_3 = \xi_4 = 0$, the realization is $G_1$; when $\xi_1 = \xi_2 = \xi_3 = 1$ and $\xi_4 = 0$, the realization is $G_2$; and when $\xi_1 = \xi_2 = \xi_3 = \xi_4 = 1$, the realization is $G_3$.

Among all the realizations, there is a special one called the adjoint graph, which is defined below.

**Definition 5.** For uncertain graph $G = (V, E, \xi)$, the adjoint graph $G^*$ is defined as $G^* = (V, E)$.

The adjoint graph is a deterministic graph, which contains all the vertices and edges of the original uncertain graph (See Figure 6).

Since different realizations of uncertain graph may have different diameters, the diameter of uncertain graphs is an uncertain variable rather than a crisp value. Thus, in an uncertain graph, it is practically meaningful to investigate the uncertainty distribution function of the diameter. It is easy to
verify that, in $G = (V, E, \xi)$, the diameter can be considered as a function of indicator functions, i.e., $\text{diam}(G) = \text{diam}(\xi_1, \cdots, \xi_m)$. In this paper, we design an efficient algorithm to compute the value of $\mathcal{M}\{\text{diam}(\xi_1, \cdots, \xi_m) \leq k\}$, where $k$ is a positive integer.

In a deterministic graph, adding an edge will not increase the diameter. So, in an uncertain graph, the diameter $\text{diam}(\xi_1, \cdots, \xi_m)$ is a decreasing function of $\xi_i$, $i = 1, \cdots, m$. This characteristic will play an important role for computing the distribution function of uncertain graphs.

### 4 Distribution Function of Diameter

To further investigate the characteristics of the diameter, this section aims to obtain a brief form of uncertainty distribution function of diameter.

Theorem 1 is a general operational law in uncertainty theory, by which we can directly formulate the uncertainty distribution function of $\text{diam}(\xi_1, \cdots, \xi_m)$, i.e.,

$$
\mathcal{M}\{\text{diam}(\xi_1, \cdots, \xi_m) \leq k\} = \begin{cases} 
\min_{\text{diam}(B_1, B_2, \cdots, B_m) \subset (-\infty, k]} \sup_{1 \leq i \leq m} \mathcal{M}\{\xi_i \in B_i\}, & \text{if } \min_{\text{diam}(B_1, B_2, \cdots, B_m) \subset (-\infty, k]} \sup_{1 \leq i \leq m} \mathcal{M}\{\xi_i \in B_i\} > 0.5 \\
1 - \min_{\text{diam}(B_1, B_2, \cdots, B_m) \subset (k, +\infty]} \sup_{1 \leq i \leq m} \mathcal{M}\{\xi_i \in B_i\}, & \text{if } \min_{\text{diam}(B_1, B_2, \cdots, B_m) \subset (k, +\infty]} \sup_{1 \leq i \leq m} \mathcal{M}\{\xi_i \in B_i\} > 0.5 \\
0.5, & \text{otherwise,}
\end{cases}
$$

where $B_i$ is the subset of $\{0, 1\}$, $i = 1, 2, \cdots, m$.

The above formula is a complex piecewise function. Moreover, each subfunction is difficult to understand and calculate. In essence, the inherent computational difficulties can be overcome, since $\text{diam}(\xi_1, \cdots, \xi_m)$ is a decreasing function and $\xi_i$ is a boolean uncertain variable. We first focus on the simplification of Theorem 1, under the condition that $\xi_i$ is a boolean uncertain variable and $f$ is a monotone function.

**Theorem 2.** Suppose that $m$ is a positive integer and $\xi_1, \xi_2, \cdots, \xi_m$ are independent boolean uncertain variables, i.e.,

$$
\xi_i = \begin{cases} 
1, & \mathcal{M}\{\xi_i = 1\} = \alpha_i \\
0, & \mathcal{M}\{\xi_i = 0\} = 1 - \alpha_i
\end{cases}, \quad i = 1, 2, \cdots, m.
$$

If $f$ is a monotone function, then uncertain variable $\xi = f(\xi_1, \xi_2, \cdots, \xi_m)$ has a distribution function

$$
\mathcal{M}\{\xi \leq k\} = \sup_{f(B_1, B_2, \cdots, B_m) \subset (-\infty, k]} \min_{1 \leq i \leq m} \mathcal{M}\{\xi_i \in B_i\},
$$

Figure 6: Uncertain graph $G = (V, E, \xi)$ and its adjoint graph $G^* = (V, E)$.
where $B_i$ is the subset of \{0, 1\}, $i = 1, 2, \ldots, m$, respectively.

**Proof:** The proof of Theorem 2 is given in the Appendix. □

Theorem 2 is a fundamental theorem in uncertain graphs. According to Theorem 2, we can easily formulate the uncertainty distribution function of \( \text{diam}(\xi_1, \cdots, \xi_m) \) in a much simpler form, i.e.,

**Theorem 3.** In uncertain graph $G = (V, E, \xi)$, for any positive integer $k$, the uncertainty distribution function of $\text{diam}(G)$ satisfies

$$
\mathcal{M}\{ \text{diam}(G) \leq k \} = \sup_{\text{diam}(G) \leq k} \min_{1 \leq i \leq m} \mathcal{M}\{ \xi_i \in B_i \},
$$

where $B_i$ is the nonempty subset of \{0, 1\}, $i = 1, 2, \ldots, m$.

Although we have obtained a simplified form to calculate the distribution function, Theorem 3 is still a fairly abstract formula for practical applications. For understanding convenience, in uncertain graph $G = (V, E, \xi)$, we define a set

$$
\mathcal{D}(k) = \{ H | H \text{ is the spanning subgraph of adjoint graph } G^* \text{ and } \text{diam}(H) \leq k \}.
$$

By integrating the set $\mathcal{D}(k)$, Theorem 3 can be reexpressed in a specific form given in Theorem 4.

**Theorem 4.** In uncertain graph $G = (V, E, \xi)$, the uncertainty distribution function of $\text{diam}(G)$ is

$$
\mathcal{M}\{ \text{diam}(G) \leq k \} = \sup_{H \in \mathcal{D}(k)} \min_{e_i \in E(H)} \mathcal{M}\{ \xi_i = 1 \} = \sup_{H \in \mathcal{D}(k)} \min_{e_i \in E(H)} \alpha_i,
$$

where $e_i \in E(H)$, that is, $e_i$ is the edge of subgraph $H$.

**Proof:** The proof can be divided into two steps.

**Step 1:** We prove that

$$
\sup_{H \in \mathcal{D}(k)} \min_{e_i \in E(H)} \mathcal{M}\{ \xi_i = 1 \} \geq \sup_{\text{diam}(B_1^*, \cdots, B_m^*) \subset [1, k]} \min_{1 \leq i \leq m} \mathcal{M}\{ \xi_i \in B_i \}. \tag{1}
$$

Since $m$ is a finite integer, there exists a series $\{B_1^*, B_2^*, \cdots, B_m^*\}$, where $B_i^* \in \{\{0\}, \{1\}, \{0, 1\}\}, i = 1, 2, \cdots, m$, satisfying

$$
\text{diam}(B_1^*, \cdots, B_m^*) \subset [1, k],
$$

$$
\sup_{\text{diam}(B_1^*, \cdots, B_m^*) \subset [1, k]} \min_{1 \leq i \leq m} \mathcal{M}\{ \xi_i \in B_i \} = \min_{1 \leq i \leq m} \mathcal{M}\{ \xi_i \in B_i^* \}.
$$

Without loss of generality, set $B_1^* = \{0\}$ and

$$
\text{diam}(\{0\}, B_2^*, \cdots, B_m^*) \subset [1, k].
$$

Since $\text{diam}(\xi_1, \cdots, \xi_m)$ is decreasing with respect to $\xi_i$, we have

$$
\text{diam}(\{0, 1\}, B_2^*, \cdots, B_m^*) \subset [1, k].
$$

On the one hand, we have

$$
\sup_{\text{diam}(B_1^*, \cdots, B_m^*) \subset [1, k]} \min_{1 \leq i \leq m} \mathcal{M}\{ \xi_i \in B_i \} \geq \mathcal{M}\{ \xi_1 \in \{0, 1\}\} \wedge \min_{2 \leq i \leq m} \mathcal{M}\{ \xi_i \in B_i^* \}.
$$

On the other hand,

$$
\sup_{\text{diam}(B_1^*, \cdots, B_m^*) \subset [1, k]} \min_{1 \leq i \leq m} \mathcal{M}\{ \xi_i \in B_i \} = \min_{1 \leq i \leq m} \mathcal{M}\{ \xi_i \in B_i^* \} \leq \mathcal{M}\{ \xi_1 \in \{0, 1\}\} \wedge \min_{2 \leq i \leq m} \mathcal{M}\{ \xi_i \in B_i^* \}.
$$
Thus,
\[
\sup_{diam(B_1, \ldots, B_m) \subseteq [1, k]} \min_{1 \leq i \leq m} \mathcal{M}\{\xi_i \in B_i\} = \mathcal{M}\{\xi_i \in \{0, 1\}\} \wedge \min_{2 \leq i \leq m} \mathcal{M}\{\xi_i \in B_i^*\}.
\]

So, we can extend $B_i^*$ from $[0]$ to $\{0, 1\}$. It shows that there exists a series $\{B_1^*, B_2^*, \ldots, B_m^*\}$, where $B_i^* \in \{\{1\}, \{0, 1\}\}$, $i = 1, 2, \ldots, m$, satisfying
\[
diam(B_1^*, \ldots, B_m^*) \subseteq [1, k],
\]
\[
\sup_{diam(B_1, \ldots, B_m) \subseteq [1, k]} \min_{1 \leq i \leq m} \mathcal{M}\{\xi_i \in B_i\} = \min_{1 \leq i \leq m} \mathcal{M}\{\xi_i \in B_i^*\}.
\]

Construct a graph $H^*$ satisfying $V(H^*) = V(G)$ and $E(H^*) = \{e_i | e_i \in E(G), B_i^* = \{1\}\}$. For the choice of $B_i^*$, it is easy to verify that $H^* \in \mathcal{D}(k)$. Then,
\[
\sup_{H \in \mathcal{D}(k)} \min_{e_i \in E(H)} \mathcal{M}\{\xi_i = 1\} \geq \min_{e_i \in E(H^*)} \mathcal{M}\{\xi_i = 1\} = \min_{1 \leq i \leq m} \mathcal{M}\{\xi_i \in B_i^*\},
\]
and we have inequality (1).

**Step 2:** we prove that
\[
\sup_{diam(B_1, \ldots, B_m) \subseteq [1, k]} \min_{1 \leq i \leq m} \mathcal{M}\{\xi_i \in B_i\} \geq \sup_{H \in \mathcal{D}(k)} \min_{e_i \in E(H)} \mathcal{M}\{\xi_i = 1\}.
\]

Since $\mathcal{D}(k)$ is a finite set, there must exist a subgraph $H' \in \mathcal{D}(k)$, satisfying
\[
\sup_{H \in \mathcal{D}(k)} \min_{e_i \in E(H)} \mathcal{M}\{\xi_i = 1\} = \min_{e_i \in E(H')} \mathcal{M}\{\xi_i = 1\}.
\]

Choose a series $\{B'_1, B'_2, \ldots, B'_m\}$ which satisfies $B'_i = \{1\}$, if $e_i \in E(H^*)$; $B'_i = \{0, 1\}$, if $e_i \notin E(H^*)$. Then
\[
diam(B'_1, \ldots, B'_m) \subseteq [1, k],
\]
\[
\min_{e_i \in E(H')} \mathcal{M}\{\xi_i = 1\} = \min_{1 \leq i \leq m} \mathcal{M}\{\xi_i \in B'_i\}.
\]

So we can write,
\[
\sup_{diam(B_1, \ldots, B_m) \subseteq [1, k]} \min_{1 \leq i \leq m} \mathcal{M}\{\xi_i \in B_i\} \geq \min_{1 \leq i \leq m} \mathcal{M}\{\xi_i \in B'_i\} = \min_{e_i \in E(H')} \mathcal{M}\{\xi_i = 1\}.
\]

Thus, we obtain inequality (2).

According to inequalities (1) and (2), the proof of Theorem 4 is proved.

\[ \square \]

## 5 Algorithm and Examples

Theorem 4 is used to calculate the distribution function of the diameter. For finding $\mathcal{M}\{diam(G) \leq k\}$ in an uncertain graph $G = (V, E, \xi)$, we need to traverse the set
\[
\mathcal{D}(k) = \{H | H \text{ is the spanning subgraph of adjoint graph } G^* \text{ and } diam(H) \leq k\}.
\]

Figure 7 is considered as an example to illustrate the proved method.

**Example 1:** Consider a simple uncertain graph $G_0$ shown in Figure 7. We want to calculate the value of $\mathcal{M}\{diam(G_0) \leq 3\}$.

According to Theorem 4, we first find set $\mathcal{D}(3)$, i.e.,
\[
\mathcal{D}(3) = \{H | H \text{ is the spanning subgraph of adjoint graph } G_0 \text{ and } diam(H) \leq 3\}.
\]
Figure 7: Uncertain graph $G_0$

Figure 8 shows all the members of set $\mathcal{D}(3)$, where

$$\mathcal{D}(3) = \{H_1, H_2, H_3, H_4, H_5\}.$$ 

We get the diameter of $H_i$ as $diam(H_1) = 3$, $diam(H_2) = 3$, $diam(H_3) = 3$, $diam(H_4) = 3$ and $diam(H_5) = 2$. According to Theorem 4, we have

\[
\begin{align*}
\min_{e_i \in E(H_1)} M\{\xi_i = 1\} &= \min_{e_i \in E(H_1)} \alpha_i = \min\{0.9, 0.4, 0.8\} = 0.4, \\
\min_{e_i \in E(H_2)} M\{\xi_i = 1\} &= \min_{e_i \in E(H_2)} \alpha_i = \min\{0.9, 0.3, 0.8\} = 0.3, \\
\min_{e_i \in E(H_3)} M\{\xi_i = 1\} &= \min_{e_i \in E(H_3)} \alpha_i = \min\{0.9, 0.3, 0.4\} = 0.3, \\
\min_{e_i \in E(H_4)} M\{\xi_i = 1\} &= \min_{e_i \in E(H_4)} \alpha_i = \min\{0.3, 0.4, 0.8\} = 0.3, \\
\min_{e_i \in E(H_5)} M\{\xi_i = 1\} &= \min_{e_i \in E(H_5)} \alpha_i = \min\{0.9, 0.3, 0.4, 0.8\} = 0.3. 
\end{align*}
\]

Then, we can write

$$M\{diam(G_0) \leq 3\} = \sup_{H \in \mathcal{D}(3)} \min_{e_i \in E(H)} M\{\xi_i = 1\} = \max\{0.4, 0.3, 0.3, 0.3, 0.3\} = 0.4.$$
Here we need to point out that in most cases, it is inefficient to generate the set $\mathcal{D}(k)$ since it needs to traverse all the spanning subgraphs of the adjoint graph. As an example, we consider the graph $H$ shown in Figure 9. Although graph $H$ contains only 10 edges, it has $2^{10} = 1024$ spanning subgraphs, which indicates that traversing the set $\mathcal{D}(k)$ is time consuming and so inefficient. Then in the following discussion, we propose a more efficient algorithm to calculate the value of $M\{\text{diam}(G) \leq k\}$ based the Floyd’s algorithm.

![Figure 9: A more complex uncertain graph](image)

Theorem 4 shows that $M\{\text{diam}(G) \leq k\}$ take values in the set $\{\alpha_1, \alpha_2, \cdots, \alpha_m\}$. Thus, instead of traversing the set $\mathcal{D}(k)$, we can equivalently traverse a smaller set of $\{\alpha_1, \alpha_2, \cdots, \alpha_m\}$. For each $\alpha_i$, we need to find $k_i$ satisfying $M\{\text{diam}(G) \leq k_i\} = \alpha_i$. If $k_{i_0} = \max(k_i | k_i \leq k, i = 1, 2, \cdots, m)$, then the corresponding $\alpha_{i_0}$ is taken as the value of $M\{\text{diam}(G) \leq k\}$.

Following algorithm is described to find the value of $M\{\text{diam}(G) \leq k\}$.

**Algorithm 1.** Algorithm for calculating the value of $M\{\text{diam}(G) \leq k\}$.

**Step 1.** Sort the set $\{\alpha_1, \alpha_2, \cdots, \alpha_m\}$ in descending order. Without loss of generality, it is assumed that $1 = \alpha_0 \geq \alpha_1 > \alpha_2 > \cdots > \alpha_m > \alpha_{m+1} = 0$. Set $j = 1$.

**Step 2.** In uncertain graph $G$, remove the pair $(e_i, \xi_i)$ that satisfies $\alpha_i < \alpha_j$, $i = 1, 2, \cdots, m$. Denote the new uncertain graph by $G_j$.

**Step 3.** Denote the adjoint graph of $G_j$ by $G_j^*$. Calculate $\text{diam}(G_j^*)$, and set $k_j = \text{diam}(G_j^*)$.

**Step 4.** If $j = m + 1$ or $k_j \leq k$, stop; if $k_j > k$, set $j = j + 1$ and go to Step 2.

In Step 3, the diameter of adjoint graph $G_j^*$ can be obtained by Floyd algorithm. The complexity of Floyd algorithm is $O(n^3)$, where $n$ is the number of vertices. Thus, the complexity of this algorithm is $O(m \cdot n^3)$, where $n$ is the number of vertices and $m$ is the number of edges.

In Step 4, if the termination condition is $j = m + 1$, then $M\{\text{diam}(G) \leq k\} = 0$. If $k_j \leq k$, then $M\{\text{diam}(G) \leq k\} = \alpha_j$. Next, we show that Algorithm 1 gives exact results.

**Proposition 1.** The result obtained by Algorithm 1 is exactly equal to the value of $M\{\text{diam}(G) \leq k\}$.

**Proof:** It is obvious that more is the value of index $j$, less is the value of $\alpha_j$. So in Step 2, we remove less edges in $G$. Thus, uncertain graph $G_j$ contains more edges, and so the diameter of adjoint graph $G_j^*$ is smaller. That is, we can write if $j_1 > j_2$, then $\text{diam}(G_{j_1}^*) \leq \text{diam}(G_{j_2}^*)$. In Step 4, there are two cases for termination condition. Following discussion considers both the cases.

Case 1: The termination condition is $j = m + 1$, so $k_m > k$ in the $m$th iteration of Step 2 and Step 3. It implies that $G_m = G$ and $G_m^* = G^*$. Since there exists no sample $H'$ satisfying $\text{diam}(H') < \text{diam}(G^*) = k_m$, it is impossible that $\text{diam}(G) \leq k < k_m$. So, $M\{\text{diam}(G) \leq k\} = 0$. 


Case 2: The termination condition is $k_j \leq k$. So in the $(j-1)$th iteration $k_{j-1} > k$. Step 3 shows $\text{diam}(G_j^*) = k_j$ and $\text{diam}(G_j^*_{k_{j-1}}) = k_{j-1}$, which gives $G_j^* \in \mathcal{D}(k_j)$. According to Theorem 4, we have

$$\mathcal{M}\{\text{diam}(G) \leq k_j\} \geq \min_{e_i \in E(G_j^*)} \mathcal{M}\{\xi_i = 1\} = \alpha_j.$$  

Since $k_j \leq k$, we have

$$\mathcal{M}\{\text{diam}(G) \leq k\} \geq \mathcal{M}\{\text{diam}(G) \leq k_j\} \geq \alpha_j.$$  

Assume that $\mathcal{M}\{\text{diam}(G) \leq k\} > \alpha_j$, i.e., $\mathcal{M}\{\text{diam}(G) \leq k\}$ is located in $\{\alpha_1, \alpha_2, \cdots, \alpha_{j-1}\}$. Without loss of generality, assume

$$\mathcal{M}\{\text{diam}(G) \leq k\} = \alpha_{j-1}. \quad (3)$$

According to Theorem 4, there exists a subgraph $H \in \mathcal{D}(k)$, satisfying

$$\mathcal{M}\{\text{diam}(G) \leq k\} = \min_{e_i \in E(H)} \mathcal{M}\{\xi_i = 1\} = \alpha_{j-1}.$$  

It implies that for any $e_i \in E(H)$, $\mathcal{M}\{\xi_i = 1\} = \alpha_i \geq \alpha_{j-1} > \alpha_j$. On the other hand, according to Step 2, it is easy to see that all the edges $e_i \in E(G)$ satisfying $\alpha_i \geq \alpha_{j-1}$ are contained in the graph $G_{k_{j-1}}$. In other words, $H \subset G_{k_j}$, i.e., $H$ is a spanning graph of graph $G_{k_j}$. So, $\text{diam}(H) \geq \text{diam}(G_{k_{j-1}}) = k_{j-1} > k$, from which we can conclude $H \notin \mathcal{D}(k)$. So, assumption (3) is not true, i.e.,

$$\mathcal{M}\{\text{diam}(G) \leq k\} = \alpha_j.$$  

Hence, Proposition 1 is proved. □

Next, we shall use the proposed algorithm to test the numerical example given below.

**Example 2:** Consider uncertain graph $G$ shown in Figure 10. We use Algorithm 1 to calculate the value of $\mathcal{M}\{\text{diam}(G) \leq 3\}$.

\[ \text{Figure 10: Uncertain Graph } G \text{ of Example 2} \]

According to Theorem 4, $\mathcal{M}\{\text{diam}(G) \leq 3\}$ takes values from the set $\{1, 0.9, 0.8, 0.7, 0.6, 0.5\}$. In this example, $\alpha_1 = 1$, $\alpha_2 = 0.9$, $\alpha_3 = 0.8$, $\alpha_4 = 0.7$, $\alpha_5 = 0.6$, $\alpha_6 = 0.5$. Next, we present the solution procedure in detail.

In the first iteration, i.e., when $j = 1$ and $\alpha_1 = 1$, remove the uncertain edges $(e_i, \xi_i)$ that satisfy $\mathcal{M}\{\xi_i = 1\} = \alpha_i < 1$, and then we obtain a new uncertain graph $G_1$ shown in Figure 11. Denote the adjoint graph of $G_1$ by $G_1^*$. For this iteration, it is easy to see that $G_1^*$ is disconnected and we have $\text{diam}(G_1^*) = +\infty$. Since $\text{diam}(G_1^*) = +\infty > 3$, according to Algorithm 1, we continue the iteration.

In the second iteration, $j = 2$ and $\alpha_2 = 0.9$. In uncertain graph $G$, remove the uncertain edges $e_i$ with $\alpha_i < 0.9$. It gives a new uncertain graph $G_2$ with adjoint graph $G_2^*$ as shown in Figure 12. Now, $G_2^*$ is still disconnected and $\text{diam}(G_2^*) = +\infty > 3$, which indicates to continue the iteration.
In the third iteration, \( j = 3 \) and \( \alpha_3 = 0.8 \). Again from uncertain graph \( G \), remove the uncertain edges \( e_i \) with \( \alpha_i < 0.8 \). Uncertain graph \( G_3 \) is obtained with adjoint graph \( G_3^* \) shown in Figure 13. The adjoint graph \( G_3^* \) is connected for which \( \text{diam}(G_3^*) = 4 > 3 \). According to Algorithm 1, we still need to continue the iteration.

In the fourth iteration, \( j = 4 \) and \( \alpha_4 = 0.7 \). After removing the uncertain edges \( e_i \) with \( \alpha_i < 0.7 \), we obtain an uncertain graph \( G_4 \) with adjoint graph \( G_4^* \) shown in Figure 14. By employing the Floyd Algorithm, we obtain the diameter of \( G_4^* \) and \( \text{diam}(G_4^*) = 3 \). According to the Algorithm 1, we can terminate the iteration. Finally, we obtain \( M\{\text{diam}(G_i) \leq k\} = \alpha_4 = 0.7 \).

**Example 3:** In this example, we consider three large-scale uncertain graphs to show the efficiency and effectiveness of the proposed Algorithm 1, in which experimental uncertain graphs \( G_1, G_2 \) and \( G_3 \) are constructed with 200 vertices and 9000, 12000 and 15000 edges, respectively. The edges of each graph are chosen randomly. The existence chance \( \alpha_i \) of edge \( e_i \) is also a random number in the interval \((0,1)\).

The algorithm is implemented by C++ in Microsoft Visual Studio 2010. Our test platform is a personal computer with Intel(R) Pentium(R) CPU 2.8 GHz and memory size 2 GB. The computation time of \( M\{\text{diam}(G_i) \leq k\} \) is listed in Table 1 for different cases.
Note that for each experiment, the computation time is less than 1.5s, which implies the high efficiency of the algorithm. According to the results, shown in Table 1, we find that for each uncertain graph, the computation time decreases as $k$ increases. This is due to the termination condition $k_j < k$, where the index $j$ is the number of iteration and $k_j$ is taken as the diameter of the adjoint graph in the $j$th iteration. In the algorithm, the value of $k_j$ becomes smaller as the index $j$ becomes larger. In other words, for the smaller value of $k$, it will take more iterations to meet the termination condition. For each $k$, more is the edges in an uncertain graph, less is the computational time of the algorithm, due to the termination condition $k_j < k$. Note that $j$ is the number of iterations and $k_j$ is set as the diameter in the $j$th iteration of the algorithm. It implies that if we fix the number of vertices, more edges will lead to a smaller $k_j$. In other words, for an uncertain graph with more edges, number of iterations and computation time is less.

6 Conclusions and Future Research

In this paper, we have studied the distribution function of the diameter of uncertain graphs. An efficient algorithm was proposed in order to calculate the value of the distribution function. The important feature of the algorithm is that it is a polynomial time algorithm having computational complexity $O(mn^3)$, where $m$ is the number of edges and $n$ is the number of vertices. Moreover, in this paper it is assumed that the vertices are predetermined. It will be interesting to investigate graphs with uncertain vertices. Other topics in classical, such as connectivity, flows, and matching, may be studied in future for uncertain graphs.

Acknowledgments

The research was supported by State Key Laboratory of Rail Traffic Control and Safety, Beijing Jiaotong University (No.RCS2014ZQ001), the Fundamental Research Funds for the Central Universities (No.2014JBM155), the National Natural Science Foundation of China (No. 71271020), Program for New
Century Excellent Talents in University under Grant No. NCET-10-0218, and National Basic Research Program of China (No. 2012CB725400).

References


Appendix

Proof of Theorem 2: Suppose that $f$ is a decreasing function. According to Theorem 1, we have

$$
M \{ \xi \leq k \} = \begin{cases} 
\sup_{f(B_1, B_2, \cdots, B_m) \subset (-\infty, k]} \min_{1 \leq i \leq m} M \{ \xi_i \in B_i \}, \\
\sup_{f(B_1, B_2, \cdots, B_m) \subset (-\infty, k]} \min_{1 \leq i \leq m} M \{ \xi_i \in B_i \} > 0.5 \\
0.5, & \text{otherwise.} 
\end{cases}
$$

(4)

We can write $M \{ \xi \leq k \} = 1$ or $M \{ \xi \leq k \} = 0$, Theorem 2 holds. We need to prove that when $0 < M \{ \xi \leq k \} < 1$, Theorem 2 still holds.

Step 1: We can find a series $\{ \hat{B}_i \}_{i=1}^m$ taking values in $\{ \{0\}, \{1\}, \{0, 1\} \}$, satisfying

$$
\sup_{f(B_1, B_2, \cdots, B_m) \subset (-\infty, k]} \min_{1 \leq i \leq m} M \{ \xi_i \in B_i \} = \min_{1 \leq i \leq m} M \{ \xi_i \in \hat{B}_i \}.
$$

Since $m$ is a finite integer, there exists a series $\{ \hat{B}_1, \hat{B}_2, \cdots, \hat{B}_m \}$, where $\hat{B}_i \in \{ \{0\}, \{1\}, \{0, 1\} \}$, $i = 1, 2, \cdots, m$, satisfying

$$
f(\hat{B}_1, \cdots, \hat{B}_m) \subset (-\infty, k], \\
\sup_{f(B_1, B_2, \cdots, B_m) \subset (-\infty, k]} \min_{1 \leq i \leq m} M \{ \xi_i \in B_i \} = \min_{1 \leq i \leq m} M \{ \xi_i \in \hat{B}_i \}.
$$

Without loss of generality, assume $\hat{B}_1 = \{0\}$, i.e.,

$$
f(\{0\}, \hat{B}_2, \cdots, \hat{B}_m) \subset (-\infty, k].
$$

Since $f(\xi_1, \cdots, \xi_m)$ is decreasing with respect to $\xi_i$, we have

$$
f(\{0, 1\}, \hat{B}_2, \cdots, \hat{B}_m) \subset (-\infty, k].
$$

On one hand,

$$
\sup_{f(B_1, \cdots, B_m) \subset (-\infty, k]} \min_{1 \leq i \leq m} M \{ \xi_i \in B_i \} \geq M \{ \xi \in \{0, 1\} \} \land \min_{2 \leq i \leq m} M \{ \xi_i \in \hat{B}_i \}.
$$

On the other hand,

$$
\sup_{f(B_1, \cdots, B_m) \subset (-\infty, k]} \min_{1 \leq i \leq m} M \{ \xi_i \in B_i \} = \min_{1 \leq i \leq m} M \{ \xi_i \in \hat{B}_i \} \leq M \{ \xi \in \{0, 1\} \} \land \min_{2 \leq i \leq m} M \{ \xi_i \in \hat{B}_i \}.
$$

Thus,

$$
\sup_{f(B_1, \cdots, B_m) \subset (-\infty, k]} \min_{1 \leq i \leq m} M \{ \xi_i \in B_i \} = M \{ \xi \in \{0, 1\} \} \land \min_{2 \leq i \leq m} M \{ \xi_i \in \hat{B}_i \}.
$$

So, we can extend $\hat{B}_1$ from $\{0\}$ to $\{0, 1\}$. Repeat this process, and we finally obtain a series $\{ \hat{B}_1, \hat{B}_2, \cdots, \hat{B}_m \}$, where $\hat{B}_i \in \{ \{1\}, \{0, 1\} \}$, $i = 1, 2, \cdots, m$, satisfying

$$
f(\hat{B}_1, \cdots, \hat{B}_m) \subset (-\infty, k], \\
\sup_{f(B_1, B_2, \cdots, B_m) \subset (-\infty, k]} \min_{1 \leq i \leq m} M \{ \xi_i \in B_i \} = \min_{1 \leq i \leq m} M \{ \xi_i \in \hat{B}_i \}.
$$
Similarly, there always exists a series \( \{ \tilde{B}_i \}_{i=1}^m \) taking values in \( \{ \{0\}, \{0,1\} \} \), satisfying
\[
\sup_{f(B_1, B_2, \ldots, B_m) \subset (k, +\infty)} \min_{1 \leq i \leq m} M\{ \xi_i \in B_i \} = \min_{1 \leq i \leq m} M\{ \xi_i \in \tilde{B}_i \}.
\]

**Step 2:** Next, we prove that
\[
\sup_{f(B_1, B_2, \ldots, B_m) \subset (-\infty, k]} \min_{1 \leq i \leq m} M\{ \xi_i \in B_i \} + \sup_{f(B_1, B_2, \ldots, B_m) \subset (k, +\infty)} \min_{1 \leq i \leq m} M\{ \xi_i \in B_i \} = 1.
\]
Without loss of generality, set \( 1 = \alpha_0 > \alpha_1 > \alpha_2 > \cdots > \alpha_m > \alpha_{m+1} = 0 \). Then, we can find a number \( p \), satisfying
\[
\sup_{f(B_1, B_2, \ldots, B_m) \subset (-\infty, k]} \min_{1 \leq i \leq m} M\{ \xi_i \in B_i \} = \min_{1 \leq i \leq m} M\{ \xi_i \in \hat{B}_i \} = \min M\{ \xi \in \hat{B}_p \} = \alpha_p,
\]
where \( \hat{B}_p = \{1\} \) and \( \hat{B}_i = \{0,1\} \) for all \( i > p \). Similarly, we can always find a number \( q \) satisfying
\[
\sup_{f(B_1, B_2, \ldots, B_m) \subset (k, +\infty)} \min_{1 \leq i \leq m} M\{ \xi_i \in B_i \} = \min_{1 \leq i \leq m} M\{ \xi_i \in \tilde{B}_i \} = \min M\{ \xi \in \tilde{B}_q \} = 1 - \alpha_q,
\]
where \( \tilde{B}_q = \{0\} \) and \( \tilde{B}_i = \{0,1\} \) for all \( i < q \). Now, we need to prove that \( p = q \) by reduction to absurdity. Assume that \( p > q \). Choose a series \( \{ B_i \}_{i=1}^m \), where
\[
B_1 = B_2 = \cdots = B_q = \cdots = B_p = \{1\}, \ B_{p+1} = \cdots = B_m = \{0,1\}.
\]
According to formula (5), we have
\[
f(B_1, B_2, \ldots, B_p, \ldots, B_m) \subset (-\infty, k].
\]
However, if we choose following series
\[
B_1 = B_2 = \cdots = B_q = \cdots = B_{p-1} = \{1\}, \ B_p = \{0\}, \ B_{p+1} = \cdots = B_m = \{0,1\},
\]
we have
\[
f(B_1, B_2, \ldots, B_p, \ldots, B_m) \subset (k, +\infty).
\]
In fact, if when \( B_p = \{0\} \) the following expression holds,
\[
f(B_1, B_2, \ldots, B_p, \ldots, B_m) \subset (-\infty, k],
\]
then we have
\[
f(B_1, B_2, \ldots, \{0,1\}, \ldots, B_m) \subset (-\infty, k].
\]
We can write
\[
\sup_{f(B_1, B_2, \ldots, B_m) \subset (-\infty, k]} \min_{1 \leq i \leq m} M\{ \xi_i \in B_i \} \geq M\{ \xi_{p-1} \in \{1\} \} = \alpha_{p-1} > \alpha_p.
\]
Obviously, the above inequality is inconsistent with formula (5), leading to formula (7). However, formula (7) is inconsistent with formula (6), i.e.,
\[
\sup_{f(B_1, B_2, \ldots, B_m) \subset (k, +\infty)} \min_{1 \leq i \leq m} M\{ \xi_i \in B_i \} \geq M\{ \xi_{p} \in \{0\} \} = 1 - \alpha_p > 1 - \alpha_q,
\]
Thus, the assumption that \( p > q \) is not reasonable. We then prove that \( p \leq q \).
In the similar way, we can prove that \( q \leq p \), that is, \( p = q \).

In a word, the following equation holds
\[
\sup_{f(B_1, B_2, \ldots, B_m) \subset (-\infty, k]} \min_{1 \leq i \leq m} M\{ \xi_i \in B_i \} + \sup_{f(B_1, B_2, \ldots, B_m) \subset (k, +\infty)} \min_{1 \leq i \leq m} M\{ \xi_i \in B_i \} = 1.
\]
Substituting the above equation into formula (4), we obtain that when \( f \) is a decreasing function, Theorem 2 holds. When \( f \) is an increasing function, the proof is similar. \( \square \)