

# 不确定图最可靠最大流算法研究

蔡 伟 张柏礼 吕建华

(东南大学计算机科学与工程学院 南京 210096)

(东南大学计算机网络和信息集成教育部重点实验室 南京 210096)

**摘 要** 文中首先基于可能世界模型提出了不确定图的最可靠最大流问题和可靠性计算模型,这对于构建可靠性网络、可靠传输路径选择以及系统薄弱环节分析等一系列实际问题具有重要意义;然后基于简单路径组合思想提出了一种求解最可靠最大流的算法 SPCA,通过简单路径流量的组合,在无需求得所有最大流分布的情况下获得最可靠最大流,并在组合过程中引入概率剪枝与约束剪枝策略,对无效组合进行过滤,从而显著地提高了算法效率;接着文中针对 SPCA 算法易受路径数量及瓶颈容量影响的问题,又提出一种基于状态空间划分的最可靠最大流算法 SDBA,该算法的主要思想是将不确定图所蕴含的子图空间划分为互不相交且满足最大流值的闭合区间集合,进而寻找所有闭合区间中概率最大的下界状态,经证明这个下界状态对应子图中的最大流分布为最可靠最大流;最后通过实验,比较了两种算法的性能.实验结果表明 SDBA 算法相对于 SPCA 算法其空间复杂度有一定的增加,但时间复杂度方面具有较大的优势,能够很好地解决 SPCA 算法性能受制于容量的问题,具有更好的性能与适用性.

**关键词** 不确定图;可能世界模型;最大流;流可靠性

中图法分类号 TP311

DOI号: 10.3724/SP.J.1016.2012.02371

## Algorithms of the Most Reliable Maximum Flow on Uncertain Graph

CAI Wei ZHANG Bai-Li LV Jian-Hua

(School of Computer Science and Engineering, Southeast University, Nanjing 210096)

(Key Laboratory of Computer Network and Information Integration of Ministry of Education, Southeast University, Nanjing 210096)

**Abstract** Reliability is one of the most important issues on system design and maintenance, such as reliable network construction, reliable transportation path selection, etc. This paper defines the most reliable maximum flow problem (MRMF) on uncertain graphs based on the possible world model, and introduces the reliability calculation model of MRMF. A simple path combination (SPCA) based algorithm is put forward, which can get the most reliable maximum flow without calculating all the maximum flow distributions. Furthermore, a space decomposition based algorithm (SDBA) is proposed to avoid the influence of the numbers and the bottleneck capacity of simple paths. SDBA divides the sub-graph space of an uncertain graph into a collection of closed intervals, which are disjoint and satisfy the maximum flow constraints. Among the lower bounds of all the intervals, the one with greatest probability is proved to be the most reliable maximum flow. Finally, experimental results show the effectiveness and efficiency of the proposed algorithms.

**Keywords** uncertain graph; possible worlds; maximum flow; flow reliability

收稿日期:2012-06-05;最终修改稿收到日期:2012-08-14. 本课题得到国家自然科学基金项目(61073059,61232007)、江苏省自然科学基金项目(Bk2010409,Bk2011335)资助. 蔡 伟,男,1987 年生,硕士研究生,主要研究方向为不确定图数据管理. E-mail: bihaichaosheng4008@163.com. 张柏礼,男,1970 年生,博士,副教授,主要研究方向为 Web 数据管理、传感器网络、数据仓库. 吕建华,男,1977 年生,博士,副教授,主要研究方向为传感器网络、数据流、数据挖掘.

## 1 引 言

随着研究和应用的深入,人们发现不确定性是实际系统的固有特性,如电力传输网中元器件、数据通信网中节点都存在着发生故障的概率,交通运输网也有着发生堵塞的概率等,当这些不确定性表现在上述网络系统中时,可以借助不确定图来描述这类系统中事物之间复杂的数据关系,其数据刻画能力往往优于其它数据模型(其中包括关系数据模型),因而有关不确定图的理论和应用研究,已逐渐引起了人们的关注,很多有价值的问题及其处理方法<sup>[1-11]</sup>被提出,其中对于不确定子图,文献<sup>[1-2]</sup>就如何在不确定图中寻找最可靠子图问题进行了深入的研究,而国内哈尔滨工业大学的邹兆年、张硕和韩蒙等人<sup>[3-5]</sup>分别对不确定图中频繁子图模式挖掘、不确定图数据库的高效查询处理等方面的问题进行了深入的研究,针对挖掘模型构建、降低子图同构计算复杂度、子图索引设计以及  $k$  极大频繁子图模式选择等问题,提出了很多有效算法. 其次,对于不确定图的可达性、相邻性也有不少的研究成果,文献<sup>[6-7]</sup>关注于不确定图中  $k$ -NN 查询、top- $k$  近邻查询,分别提出了各自的距离函数和高效的查询处理算法;东北大学的袁野等人<sup>[8]</sup>则采用一种有效的随机算法来快速估算可达概率,并结合条件概率对算法进行了改进;另外,对于最短路径、top- $k$  最短路径等重要问题在不确定图中的拓展,研究人员定义了一些很有价值的问题,并给出了一系列非常有参考意义的解决思路和算法<sup>[9-11]</sup>.

不确定图上的最大流问题是传统最大流问题在不确定图上的自然延伸,同时又孕育出很多有价值的问题,而下面所列出的则是最直接的两个:

(1) “对于存在多种不确定因素的网络,它在传

递最大流(量)或额定工作流(量)时,其可靠性如何评价?”这是网络可靠承载(最大)流能力的反映,这一论题是随机流网络中的经典问题,已被证明是 NP-hard 问题<sup>[12-13]</sup>. 这方面的算法可以分为两大类:精确算法和近似算法. 精确算法中最基本的是完全枚举的算法,通过直接枚举所有可能的状态组合得到网络的可靠度. 文献<sup>[12-15]</sup>则分别采用不同的状态划分方式,将状态空间划分成能够满足流量需求的合格状态集,不满足流量需求的不合格状态集和同时包含合格状态和不合格状态的待划分状态集,通过迭代的方式处理待划分状态集,来得到所有合格状态集的概率. 一些研究人员<sup>[16-18]</sup>提出了基于路径集或割集的方法对状态空间进行过滤,然后通过 inclusive-exclusive 原理计算网络可靠性. 对于一个复杂系统,精确算法执行效率往往不能令人满意,为此,一些近似算法被相继提出,其中 Ramirez-Marquez 和 Coit<sup>[19]</sup>提出了蒙特卡洛近似算法,而 Rocco 和 Muselli<sup>[20]</sup>采用机器学习方法得到网络可靠性.

(2) “一个(最大)流量很可能对应着多个最大流分布,在多个(最大)流分布中,哪一个最值得信赖,可作为流传递的首选方案”,这同样是一个重要的决策问题,关系到如何最大限度利用现有的网络条件选择最可靠的流实施计划. 如在图 1 中,通过 Ford-Fulkerson 最大流算法<sup>[21]</sup>可知,原始图中源点  $s$  到汇点  $t$  之间的最大流值为 3,它却对应 3 种不同的最大流分布(分别为分布 1、2、3). 而这些流分布存在的概率是不一样的,也就是流分布的可靠性不同. 而可见的文献<sup>[1-24]</sup>表明:目前没有不确定图中流分布的可靠性这方面的研究. 然而这一问题有着十分重要的研究价值,可用来有效解决诸如构建可靠性网络、可靠传输路径选择以及系统薄弱环节的分析等一系列实际问题.

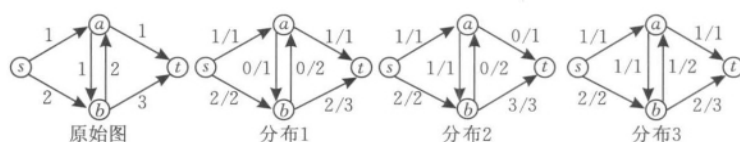


图 1 原始图及对应的最大流分布

为此,本文基于可能世界模型,首先对不确定图最可靠最大流问题进行定义,并给出最大流可靠性的概率计算模型. 然后,借鉴随机流网络中相关的算法,针对问题提出一种基于简单路径组合思想的基本算法 SPCA(Simple Path Combination Algorithm),

该算法基本思想是在最大流值与流守恒性<sup>[18]</sup>的约束下,通过对简单路径分配的流量进行组合,可以在不需要得到所有的最大流分布的情况下获得最可靠最大流,并在组合过程中引入概率剪枝与约束剪枝策略,过滤许多不必要的组合,从而显著地提高了算

法效率. SPCA 算法对于稀疏网络具有良好的适用性,但随着简单路径数量与路径的瓶颈容量增加,该算法性能下降较快,为此,本文又提出一种基于状态空间划分的最可靠最大流算法 SDBA (Space Decomposition Based Algorithm),该算法的主要思想是将不确定图所蕴含的子图空间进行划分,借助文献[14]中提出的状态空间划分方法得到不相交且满足最大流的闭合区间集合,依据概率计算模型,寻找所有闭合区间中概率最大的下界状态,在这个下界状态对应的子图中运行最大流算法得到的最大流经证明即为原图的最可靠最大流,该算法能够很好地解决基本算法对于容量的依赖,效率与适用性相对更好.最后,通过对比实验,比较了两种算法的性能,并分析了影响算法性能的因素.

## 2 问题定义

可能世界模型被广泛应用于不确定数据的建模.在该模型中,不确定数据被扩展成可能世界实例的集合,每个可能世界实例具有一定的概率,所有可能世界实例共同组成一个可能世界空间.

下面将基于不确定图的可能世界模型(假设图的边概率统计独立,边上的流量与容量取值都是非负整数,且流量要满足流守恒性<sup>[18]</sup>)给出不确定图最可靠最大流问题的定义,另外,本文研究的是一种情况比较简单的二态网络,即网络中每条边的容量

为 0 或  $C$ ,两种容量存在的概率之和为 1.

**定义 1.** 不确定图. 一个不确定图  $G$  是一个四元组  $G(V, E, \Sigma, \Gamma)$ , 其中,  $V$  是有向图顶点集合;  $E$  是边集合;  $\Sigma$  是二元组  $(c, p)$  的集合,  $c$  表示容量,  $p$  表示容量对应的概率;  $\Gamma$  是一个映射:  $E \rightarrow \Sigma$ , 对于  $\forall e \in E$ ,  $\Gamma(e) = (c, p)$  表示边  $e$  的容量取  $c$  时,对应的概率为  $p$ , 每条边对应的二元组可取值范围为  $\{(c, p), (0, q)\}$ , 且  $p + q = 1$ .

**例 1.** 图 2 给出一个不确定图  $G$ , 任取其中一条边,不妨取  $e_1$ , 边上二元组  $(2, 0.8)$  表示:  $e_1$  的容量取 2 时对应的概率为 0.8, 如果  $e_1$  的容量取 0 ( $e_1$  不存在), 对应的概率为 0.2.

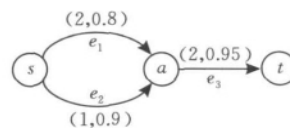


图 2 不确定图  $G$

对于一个有  $|E|$  条边的不确定图,其蕴含的确定子图个数为  $2^{|E|}$ . 图 3 所示为图 2 中不确定图  $G$  的  $2^3 = 8$  个蕴含子图,且每个子图存在的概率  $P_G(g) = \left\{ \prod_e p(e, c) \mid e \in E \right\}$  (其中  $p(e, c)$  表示边  $e$  取容量  $c$  时的概率). 当且仅当  $G$  中每条边都取最大容量时,对应的确定子图为  $G$  的最大子图,记作  $MSG(G)$  (Max Sub Graph). 图 3 中  $G \rightarrow g_8$  为  $G$  的最大子图.

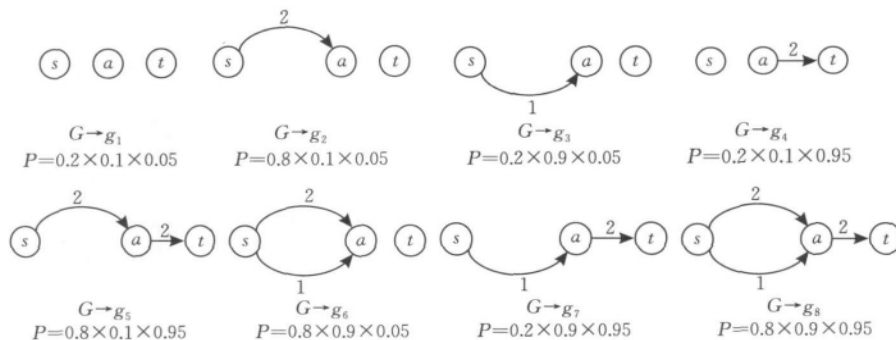


图 3  $G$  蕴含的确定子图及存在概率

不确定图  $G$  所蕴含的  $2^{|E|}$  个确定子图所构成的空间与  $|E|$  维 0-1 向量  $V(y_1, y_2, \dots, y_m)$  ( $m = |E|$ ) 之间可以建立一一映射.  $g_1$  到  $g_8$  对应的 0-1 向量分别为  $(000), (100), (010), (001), (101), (110), (011), (111)$ . 对  $V$ , 给出如下的 0-1 向量概率计算公式:

$$P(V) = \prod_k p_k \quad (1)$$

其中  $p_k = \begin{cases} p(e_k, c_k), & y_k = 1 \\ 1, & y_k = 0 \end{cases}$ ,  $p(e_k, c_k)$  表示  $G$  的第  $k$  条边  $e_k$  容量取  $c_k$  ( $c_k > 0$ ) 的概率. 如图 3 中  $g_2$  对应的 0-1 向量  $(100)$  的概率为  $P(100) = 0.8 \times 1 \times 1$ .

**定义 2.** 不确定图  $s-t$  的最大流值. 设不确定图  $G$  蕴含的最大确定子图为  $MSG(G)$ ,  $MSG(G)$  中源点  $s$  到汇点  $t$  的最大流值为  $F_{\max}$ , 则称  $F_{\max}$  为不确

定图  $G$  的  $s-t$  的最大流值.

例 2. 图 2 中不确定图  $G$  的  $s-t$  最大流值为图 3 中  $g_8$  的  $s-t$  的最大流值  $F_{\max}=2$ .

定义 3. 最大流分布及其可靠性. 对于不确定图  $G$  所蕴含的最大子图  $\text{MSG}(G)$ , 若  $\text{MSG}(G)$  中的源点  $s$  到汇点  $t$  取得最大流值  $F_{\max}$  时, 各边上通过的流量所构成的向量为  $\mathbf{FV}=(v_1, v_2, \dots, v_m)$  (其中  $m$  表示  $G$  的边数,  $v_i$  表示边  $e_i$  上通过的流量值, 这个值不超过该边的容量), 这样的向量  $\mathbf{FV}$  就是  $G$  的一个从  $s$  到  $t$  的最大流分布, 简称最大流, 且这个最大流  $\mathbf{FV}$  的可靠性  $P_G(\mathbf{FV})=\sum_g P_G(g)$  (其中:  $g$  为  $G$  所蕴含的子图, 且  $g$  中存在  $\mathbf{FV}$  这个最大流).

例 3. 图 2 中  $G$  的  $s-t$  的最大流值  $F_{\max}=2$ . 图 3 中  $g_8$  为不确定图  $G$  的最大子图,  $g_8$  在取  $F_{\max}$  时, 各条边上通过流量的取值情况如图 4 所示,  $\mathbf{FV}_1$  与  $\mathbf{FV}_2$  为  $G$  的两种可能最大流, 这两种最大流的概率分别为  $P_G(\mathbf{FV}_1)=P_G(g_5)+P_G(g_8)$  和  $P_G(\mathbf{FV}_2)=P_G(g_8)$ .

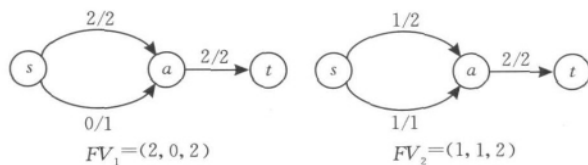


图 4 不确定图  $G$  的最大流

最大流可靠性计算模型. 根据定义 3, 最大流  $\mathbf{FV}$  的可靠性有如下的计算公式:

$$P_G(\mathbf{FV}) = \sum_g P_G(g) \quad (2)$$

其中  $g$  是  $G$  所蕴含的子图, 且  $g$  中存在  $\mathbf{FV}$  这个最大流, 通过分析可以发现这样的子图概率之和  $\sum_g P_G(g)$  其实就等于  $\mathbf{FV}$  中  $v_k$  大于 0 的边概率之积, 因此:

$$P_G(\mathbf{FV}) = \sum_g P_G(g) = \prod_k x_k \quad (3)$$

其中  $x_k = \begin{cases} p(e_k, c_k), & 0 < v_k \leq c_k \\ 1, & v_k = 0 \end{cases}$ ,  $p(e_k, c_k)$  表示  $G$  的第  $k$  条边  $e_k$  容量取  $c_k$  的概率.

按照式(3), 图 4 中最大流  $\mathbf{FV}_1$  与  $\mathbf{FV}_2$  的可靠性  $P_G(\mathbf{FV}_1)=0.8 \times 1 \times 0.95$ ,  $P_G(\mathbf{FV}_2)=0.8 \times 0.9 \times 0.95$ .

显然, 如果将  $\mathbf{FV}=(v_1, v_2, \dots, v_m)$  映射成 0-1 向量  $\mathbf{V}=(y_1, y_2, \dots, y_m)$ , 其中  $y_k = \begin{cases} 0, & v_k = 0 \\ 1, & v_k > 0 \end{cases}$ , 由式(1)、(3)可知  $P_G(\mathbf{FV})=P(\mathbf{V})$ .

基于上述相关定义, 下面给出不确定图最可靠最大流的定义.

定义 4. 不确定图最可靠最大流. 给定不确定图  $G$ , 源点  $s$  与汇点  $t$ , 设  $\mathbf{FV}_1, \mathbf{FV}_2, \dots, \mathbf{FV}_k$  是  $s$  到  $t$  的所有可能的最大流,  $P_G(\mathbf{FV}_i)$  为  $G$  中  $\mathbf{FV}_i$  这个最大流的可靠性, 使得  $P_G(\mathbf{FV}_i)$  值最大的最大流  $\mathbf{FV}_i$ , 称为  $G$  的  $s-t$  最可靠最大流.

例 4. 对图 2 中的不确定图  $G$ ,  $s-t$  的所有可能最大流为  $\mathbf{FV}_1, \mathbf{FV}_2$ ,  $P_G(\mathbf{FV}_1)=0.76$ ,  $P_G(\mathbf{FV}_2)=0.684$ ,  $P_G(\mathbf{FV}_1) > P_G(\mathbf{FV}_2)$ , 因此,  $G$  的  $s-t$  的最可靠最大流为  $\mathbf{FV}_1=(2, 0, 2)$ .

### 3 基于简单路径组合算法

本节首先对不确定图  $s-t$  简单路径进行定义, 然后构建所有  $s-t$  简单路径上通过的流量需要满足的约束模型, 接着给出简单路径流量组合可靠性计算方法, 最后给出基于简单路径组合算法 SPCA.

定义 5. 不确定图  $s-t$  简单路径. 不确定图  $G$  所蕴含的最大子图  $\text{MSG}(G)$  中, 从  $s$  到  $t$  的顶点和边都不重复的路径, 称作  $G$  的  $s-t$  简单路径, 记作  $sp$  (simple path),  $sp$  的瓶颈容量为  $sp$  所包含全部边的容量最小值, 记作  $BC$  (Bottleneck Capacity).

例 5. 如图 2 所示:  $e_1 e_3$  与  $e_2 e_3$  是  $G$  的两条  $s-t$  的简单路径, 对应的瓶颈容量分别是 2 和 1.

#### 3.1 流约束模型

设  $s-t$  的所有简单路径记作:  $sp_1, sp_2, \dots, sp_r$ , 对  $1 \leq i \leq r$ ,  $sp_i$  的瓶颈容量为  $BC_i$ , 当  $s-t$  取最大流值  $F_{\max}$  时, 每条简单路径上通过的流量为  $pf_i$ , 则  $pf_i$  满足如下约束模型:

$$\sum_{i=1}^r pf_i = F_{\max} \quad (4)$$

$$pf_i \leq BC_i \quad (5)$$

$$\sum_{i=1}^r \{pf_i | e_j \in sp_i\} \leq c_j, 1 \leq j \leq m \quad (6)$$

其中, 约束(4)保证  $s-t$  取到  $G$  的最大流值; 约束(5)保证通过简单路径的流量小于瓶颈容量; 约束(6)保证通过每条边的总流量不超过该边的容量.

#### 3.2 简单路径流量组合的可靠性

设  $\mathbf{PF}=(pf_1, pf_2, \dots, pf_r)$  是简单路径上的一种流量组合, 将所有简单路径上的流量映射到  $G$  中的边所构成向量为  $(v_1, v_2, \dots, v_m)$ ,  $v_j = \sum_{i=1}^r \{pf_i | e_j \in sp_i\}$  表示映射后每条边上通过的流量,  $(v_1, v_2, \dots,$

$v_m$ ) 对应的 0-1 向量为  $V(y_1, y_2, \dots, y_m)$ , 设式(1)计算  $V$  概率为  $P(V)$ , 于是可将简单路径流量组合  $PF$  的可靠性  $P(PF)$  定义为  $P(V)$ .

如果流量组合  $PF$  满足约束模型, 由约束(4)可知  $(v_1, v_2, \dots, v_m)$  为  $s-t$  的一个最大流  $FV$ . 此时, 简单路径流量组合  $PF$  的可靠性等于最大流  $FV$  的可靠性, 即  $P(PF) = P_G(FV)$ .

### 3.3 SPCA 算法

SPCA 算法主要由 5 步骤构成: 第 1 步在不确定图  $G$  的最大子图  $MSG(G)$  上运行最大流算法得到  $s-t$  的最大流值  $F_{\max}$ ; 第 2 步搜索  $G$  中  $s-t$  所有简单路径; 第 3 步构建所有简单路径上通过的流量需要满足的约束模型; 第 4 步搜索满足约束模型的所有  $s-t$  简单路径通过流量的组合, 找到可靠性最大的流量组合; 第 5 步将该流量组合中所有简单路径上通过的流量分配到每条边上, 得到边上通过的流量所构成的向量  $FV$ , 输出最可靠最大流  $FV$ .

**例 6.** 如图 2 所示: (1)  $s-t$  最大流值  $F_{\max} = 2$ ; (2)  $s-t$  所有简单路径为  $sp_1 = e_1 e_3, sp_2 = e_2 e_3$ ; (3) 构建  $sp_1, sp_2$  上通过的流量  $pf_1, pf_2$  需要满足的约束条件,  $pf_1 + pf_2 = 2, pf_1 \leq 2, pf_2 \leq 1, pf_1 + pf_2 \leq 2$ ; (4) 搜索所有简单路径流量组合为  $PF_1 = (2, 0), PF_2 = (1, 1), P(PF_1) = P_G(2, 0, 2) = 0.8 \times 1 \times 0.95, P(PF_2) = P_G(1, 1, 2) = 0.8 \times 0.9 \times 0.95, P(PF_1) > P(PF_2)$ , 可靠性最大的流量组合为  $PF_1$ ; (5) 将  $PF_1$  映射到每条边为  $FV = (2, 0, 2)$ , 因此,  $G$  的  $s-t$  最可靠最大流为  $(2, 0, 2)$ .

SPCA 算法步骤 4 以约束(5)作为搜索空间, 让  $PF = (pf_1, pf_2, \dots, pf_r)$  从  $(0_1, 0_2, \dots, 0_r)$  到  $(BC_1, BC_2, \dots, BC_r)$  进行枚举, 为了提高搜索过程的效率, 可以采用如下两种剪枝策略.

(1) 概率剪枝策略. 设  $PF = (pf_1, pf_2, \dots, pf_r)$  是已经找到的满足约束模型的一个流量组合, 这个流量组合的可靠性为  $P$ , 假设当前搜索约束模型下一个流量组合时,  $sp_1, \dots, sp_{i-1}$  上已经取定流量  $pf'_1, \dots, pf'_{i-1}$ , 此时对  $sp_i$  分配流量前, 如果  $(pf'_1, \dots, pf'_{i-1}, 0_i, \dots, 0_r)$  的可靠性为  $P'$ , 且  $P' \leq P$ , 则  $(pf'_1, \dots, pf'_{i-1}, 0_i, \dots, 0_r)$  到  $(pf'_1, \dots, pf'_{i-1}, BC_i, \dots, BC_r)$  的所有组合可直接过滤掉.

(2) 约束剪枝策略. 设  $sp_1, \dots, sp_{i-1}$  上已经取定流量  $pf'_1, \dots, pf'_{i-1}$ , 则  $sp_i$  上的流量取值  $pf_i$  依次从  $\min\{F_{\max} - \sum_{j=1}^{i-1} pf_j, BC_i\}$  取到 0, 如果  $pf_i$  的某个

取值  $f$  满足  $F_{\max} - \sum_{j=1}^i pf_j > \sum_{j=i+1}^r BC_j$ , 则所有简单路径上通过的流量组合从  $(pf'_1, \dots, pf'_{i-1}, 0_i, \dots, 0_r)$  到  $(pf'_1, \dots, pf'_{i-1}, f, BC_{i+1}, \dots, BC_r)$  可以直接过滤掉.

#### 算法 1. SPCA 算法.

输入: 不确定图  $G$ , 源点  $s$ , 汇点  $t$

输出:  $s-t$  最大流值  $F_{\max}$ ,  $s-t$  最可靠最大流  $FV$  及其概率  $P$

1.  $F_{\max} = \text{getMaxFlow}(MSG(G), s, t)$ ;
2. Find all simple path from  $s$  to  $t$   $\{sp_1, sp_2, \dots, sp_r\}$ ;
3. Construct the constraint model which  $\{pf_1, pf_2, \dots, pf_r\}$  should be satisfied by  $F_{\max}, \{sp_1, sp_2, \dots, sp_r\}, G$ ;
4. Using the constraint pruning strategy and probabilistic pruning strategy to search for the flow combination which can satisfy the constraint model;
5. Get the most reliable flow combination  $PF = \{pf_1, \dots, pf_r\}$  and its probability  $P$  by step 4, and transfer  $PF$  to  $FV$ ;
6. Return  $(F_{\max}, FV, P)$ .

假设所有  $s-t$  简单路径已知, 则算法时间复杂度

的上界为  $O(rm \prod_{i=1}^r BC_i)$ , 其中  $r$  为简单路径数量,  $m$  为边数,  $BC_i$  为每条简单路径的瓶颈容量.

## 4 基于空间划分算法

由第 2 节可知: 不确定图所蕴含的子图与 0-1 向量之间存在一一映射的关系, 即一个不确定图所蕴含的子图可以用 0-1 向量  $V$  描述成子图  $V, G$  的整个子图空间可以描述为  $\Omega = [(0, 0, \dots, 0), (1, 1, \dots, 1)]$ .

**定义 6.** 偏序关系. 设定偏序关系  $\geq$  和  $>$ , 对于两个 0-1 向量  $V_1(y_{11}, y_{12}, \dots, y_{1m})$  和  $V_2(y_{21}, y_{22}, \dots, y_{2m})$ , 若对任意的  $i, i \in \{1, \dots, m\}, y_{1i} \geq y_{2i}$ , 则  $V_1 \geq V_2$ ; 若  $V_1 \geq V_2$ , 且存在至少一个  $i, i \in \{1, \dots, m\}$ , 使得  $y_{1i} > y_{2i}$ , 则  $V_1 > V_2$ .

基于上述偏序关系, 本节提出了 SDBA 算法, 该算法主要由 4 步构成: 第 1 步在不确定图  $G$  的最大子图  $MSG(G)$  上运行最大流算法得到  $s-t$  的最大流值  $F_{\max}$ ; 第 2 步对  $G$  子图空间  $\Omega$  进行划分, 得到所有能够取得最大流值  $F_{\max}$  的不相交闭合空间; 第 3 步使用式(1)对每个闭合空间下界对应 0-1 向量计算概率, 获得概率最大的 0-1 向量  $V$ ; 第 4 步在子图  $V$  上运行最大流算法, 得到一个最大流  $FV$ , 输出最可靠最大流  $FV$ .

为了能够取得不确定图  $G$  的所有满足  $s-t$  最大流值  $F_{\max}$  的不相交闭合子区间所构成的集合  $R$ , 这里借鉴文献[14]中的空间划分方法, 具体过程如下:

对于一个给定的 0-1 闭合子图空间  $C[L, U]$ ,  $L=(l_1, \dots, l_m)$ ,  $U=(u_1, \dots, u_m)$ ,  $F(L^C)$ ,  $F(U^C)$  分别表示子图空间  $C$  的下界子图  $L$  和上界子图  $U$  的  $s-t$  最大流值, 如果  $F(L^C) \geq F_{\max}$ , 则  $C \in R$  (表示  $C$  中所有子图都满足  $s-t$  最大流值  $F_{\max}$ ), 如果  $F(U^C) < F_{\max}$ , 则  $C \notin R$  (表示  $C$  中所有子图都不满足  $F_{\max}$ ), 如果  $F(L^C) < F_{\max} \leq F(U^C)$ , 则  $C$  中同时存在满足  $F_{\max}$  与不满足  $F_{\max}$  的子图,  $C$  与  $R$  关系不能确定, 需要进一步划分.

**规则 1.** 设  $C[L, U]$  与  $R$  的关系不确定,  $FV=(v_1, v_2, \dots, v_m)$  是  $C$  的上界子图  $U$  上的一种最大流,  $V(y_1, y_2, \dots, y_m)$  是  $FV$  映射成的 0-1 向量, 令  $I_c = \{i | i \in \{1, \dots, |E|\} \wedge y_i > l_i\} = \{i_1, i_2, \dots, i_q\}$ ,  $q = |I_c|$ , 通过  $I_c$  可以将  $C$  划分成  $q+1$  个非空且不相交的闭合子区间  $C_0, C_1, \dots, C_q$ , 且满足  $C_0 \in R$ , 其中  $C_k = [(\alpha_1, \dots, \alpha_m), (\beta_1, \dots, \beta_m)]$ , 当  $k=0$  时, 对  $i \in I_c$ ,  $\alpha_i = \beta_i = u_i$ ,  $i \notin I_c$ ,  $\alpha_i = l_i$ ,  $\beta_i = u_i$ ; 当  $0 < k \leq q$  时, 对  $1 \leq j < k$ ,  $\alpha_{i_j} = \beta_{i_j} = u_{i_j}$ , 对  $j=k$ ,  $\alpha_{i_j} = \beta_{i_j} = l_{i_j}$ , 对  $i \in \{1, 2, \dots, m\} \setminus \{i_1, i_2, \dots, i_k\}$ ,  $\alpha_i = l_i$ ,  $\beta_i = u_i$  [14].

**例 7.** 图 2 不确定图  $G$  的 0-1 闭合子图空间  $C[L, U] = [(0, 0, 0), (1, 1, 1)]$ ,  $F_{\max} = 2$ ,  $F(L^C) = F(0, 0, 0) = 0$ ,  $F(U^C) = F(1, 1, 1) = 2$ ,  $F(L^C) < F_{\max} \leq F(U^C)$ ,  $FV = (2, 0, 2)$ , 映射成 0-1 向量  $V = (1, 0, 1)$ , 按照规则 1 得到  $I_c = \{1, 3\}$ ,  $q=2$ , 通过  $I_c$  将  $C[L, U]$  划分成  $C_0 = [(1, 0, 1), (1, 1, 1)]$ ,  $C_1 = [(0, 0, 0), (0, 1, 1)]$ ,  $C_2 = [(1, 0, 0), (1, 1, 0)]$ , 且  $C_0 \in R$ .

基于上述空间划分方法的最可靠最大流算法 SDBA 伪代码描述如下.

**算法 2.** SDBA 算法.

输入: 不确定图  $G$ , 源点  $s$ , 汇点  $t$

输出:  $s-t$  最大流值  $F_{\max}$ ,  $s-t$  最可靠最大流  $FV$  及其概率  $P$

1. Init  $P=0$ ,  $left \leftarrow (0, 0, \dots, 0_{|E|})$ ;
2.  $F_{\max} = \text{getMaxFlow}(\text{MSG}(G), s, t)$ ;
3. Push all the subgraphs states  $\{(0, 0, \dots, 0_{|E|}), (1, 1, \dots, 1_{|E|}), \emptyset, I=0, j=0\}$  into stack;
4. While stack is not empty
5. Get current closed space  $[\alpha, \beta]$  of subgraphs according to  $j$  in the top state of stack;
6. If  $j < I$ , then  $j++$ ; else pop stack;
7. Get the max flow value  $f$  on current subgraph  $\alpha$ ;

8. If  $f \geq F_{\max}$ , then
9. Calculate the possibility  $p(\alpha)$ ;
10. If  $p(\alpha) > P$ , then  $P = p(\alpha)$  and  $left \leftarrow \alpha$ ;
11. Else
12. Get the max flow value  $f$  and its distribution  $F_W$  on subgraph  $\beta$ ;
13. If  $f \geq F_{\max}$ , then
14. Use  $F_W$  to get pivot sets  $I_C$  by rule 1 and push  $\{\alpha, \beta, I_C, |I_C|, 1\}$  into stack;
15. Get  $C_0 = [\alpha', \beta]$  by  $\{\alpha, \beta, I_C, |I_C|, 1\}$  according to rule 1;
16. Calculate the possibility  $p(\alpha')$ ;
17. If  $p(\alpha') > P$
18. Then  $P = p(\alpha')$  and  $left \leftarrow \alpha'$ ;
19. Get current flow distribution  $FV$  on subgraph  $left$ ;
20. Return  $(F_{\max}, FV, P)$ .

假设 SDBA 算法进行状态划分所需要的迭代次数为  $k$ , 最大流算法选用时间复杂度为  $O(mn^2)$  的 Dinic 算法, 则 SDBA 时间复杂度为  $O(kmn^2)$  (其中  $m = |E|$ ,  $n = |V|$ ).

对于算法的正确性, 给出如下的定理.

**定理 1.** 设不确定图  $G$  的  $s-t$  最大流值为  $F_{\max}$ , 假设所有可以达到  $F_{\max}$  的子图空间被划分成  $A_1, \dots, A_r$  (其中  $A_i$  是闭合区间  $[\alpha_i, \beta_i]$ ),  $\alpha_i = (\alpha_{i1}, \dots, \alpha_{im})$ ,  $\beta_i = (\beta_{i1}, \dots, \beta_{im})$ , 且  $A_i \cap A_j = \emptyset$ , 设  $\alpha_1, \dots, \alpha_r$  中  $P(\alpha_i)$  最大, 则下面结论成立:

(1) 子图  $\alpha_i$  中去掉任意一条边后的子图中,  $s-t$  最大流值必定小于  $F_{\max}$ ;

(2) 子图  $\alpha_i$  中必定存在不确定图  $G$  的一种  $s-t$  最可靠最大流.

证明.

(1) 反证法. 假设  $\alpha_i$  去掉任意一条边后的子图  $\alpha'_i$  的最大流值为  $F_{\max}$ , 则  $\alpha'_i$  必定属于某个  $A_j$  ( $j \neq i$ ,  $j \in \{1, 2, \dots, r\}$ ), 即  $\alpha'_i \geq \alpha_j$ , 所以  $P(\alpha'_i) \leq P(\alpha_j)$ , 而  $P(\alpha'_i) > P(\alpha_i)$ , 因此  $P(\alpha_j) > P(\alpha_i)$ , 这与  $P(\alpha_i)$  最大矛盾, 即结论 (1) 成立.

(2) 结论 (1) 说明在  $\alpha_i$  上得到的最大流可靠性为  $P(\alpha_i)$ , 因此, 证明结论 2 等价于证明  $G$  的  $s-t$  最可靠最大流  $FV$  的可靠性为  $P(\alpha_i)$ , 反证法: 假设  $s-t$  的最可靠最大流  $FV$  可靠性  $P_G(FV)$  大于  $P(\alpha_i)$ , 不妨设  $FV = (v_1, v_2, \dots, v_m)$ , 将  $FV$  映射成 0-1 向量  $V = (y_1, y_2, \dots, y_m)$ , 其中  $y_k = \begin{cases} 1, & 0 < v_k \leq c_k \\ 0, & v_k = 0 \end{cases}$ ,  $V$  对应不确定图  $G$  的一个子图, 由于  $G$  中所有能够取得最大流值  $F_{\max}$  的子图空间为  $A_1, \dots, A_r$ , 因此,  $V$  必定

属于某个  $A_j (j \in \{1, 2, \dots, r\})$ , 则  $P(V) \leq P(\alpha_j)$ , 而  $P(\alpha_j) \leq P(\alpha_i)$ , 所以  $P(V) \leq P(\alpha_i)$ , 而  $P_G(FV) = P(V)$ , 因此  $P_G(FV) \leq P(\alpha_i)$ , 这与假设矛盾, 因此结论 2 成立. 证毕.

定理 1 说明将不确定图  $G$  中能够取到最大流值  $F_{\max}$  的子图空间划分成若干不相交闭合区间, 通过式(1)中给出的 0-1 向量概率计算方法, 可以在这些不相交闭合区间下界对应的 0-1 向量中找到一个概率最大的 0-1 向量  $\alpha_i$ , 在  $\alpha_i$  子图上运行最大流算法得到的最大流必定是最可靠最大流, 这说明 SDBA 算法在理论上是正确的.

## 5 实 验

为了验证本文所提出算法的运行效率及分析影响算法性能的各种因素, 我们进行了一系列的实验, 其中所有算法都在 STL 库支持下用 Visual C++ 实现, 本文实验以普通 PC 为平台, 基本配置为: 处理器 Intel(R) Core(TM) i3, CPU 2.93 GHz, 内存 4 GB, 64 位 Windows 7 操作系统(下一步工作将基于图分割策略, 研究算法由单机向多机并行扩展).

由于最可靠最大流问题是本文提出的新问题, 目前还没有这方面公认的数据测试集, 本文实验是在图相关研究中普遍采用的有向图生成器 NETGEN<sup>[25]</sup> 基础上, 按照一定的规则合成测试数据集, 主要分为 3 类:

(1) 通过有向图生成器生成  $V_6 E_{10}$ 、 $V_8 E_{14}$ 、 $V_{10} E_{18}$ 、 $V_{12} E_{22}$ 、 $V_{14} E_{26}$  5 组有向图, 其中  $V_n E_m$  表示由  $n$  个顶点、 $m$  条边所构成的图集合(本实验中集合大小为 5), 图中边的容量与对应概率按照随机、均匀分布、正态分布( $\sigma=0.2, 1.0, 1.8$ )进行组合分配, 它们的组合情况如表 1 所示.

表 1 边容量与对应概率组合

容量 $c$	概率 $p$				
	随机分布	均匀分布	正态分布(0.5, 1), $\mu=0.75$		
随机分布	$c \in [1, 40]$ $p \in (0.5, 1)$	$c \in [1, 40]$ $p \in (0.5, 1)$	$c \in [1, 40]$ $\sigma=0.2$	$c \in [1, 40]$ $\sigma=1.0$	$c \in [1, 40]$ $\sigma=1.8$
均匀分布	$c \in [1, 40]$ $p \in (0.5, 1)$	$c \in [1, 40]$ $p \in (0.5, 1)$	$c \in [1, 40]$ $\sigma=0.2$	$c \in [1, 40]$ $\sigma=1.0$	$c \in [1, 40]$ $\sigma=1.8$

为了避免随机生成的图中源点与汇点的任意选择, 不能保证源点与汇点之间有足够多的简单路径, 无法反映出算法效率与路径数量的关系, 实验中对这类数据  $s, t$  的选取策略是使用两点之间简单路数量最多的一对顶点作为源点与汇点.

(2) 在图的顶点个数固定(实验中顶点为 15), 边的数量按照文献[26]所定义的图稠密度  $d = 2|E|/|V| \times (|V| - 1)$ (实验中稠密度  $d$  分别取为 0.3, 0.35, 0.4, 0.45, 0.5)确定的前提下, 图集合(本实验集合大小为 5)通过有向图生成器产生, 然后给图的每条边都分配大小为 10 的容量, 容量对应的概率在(0.5, 1)之间随机取值, 实验中这类数据集的源点与汇点选取策略同第(1)类数据.

(3) 在文献[14]中 15 个顶点、30 条边的有向图基础上, 通过给边依次分配容量(本实验分配的容量依次为 5, 10, 15, 20, 25, 30)及对应的概率(概率在(0.5, 1)之间随机取值), 产生大小为 6 的图集合, 继续使用文献[14]中的  $s, t (s=1, t=15)$  作为本文实验的源点与汇点.

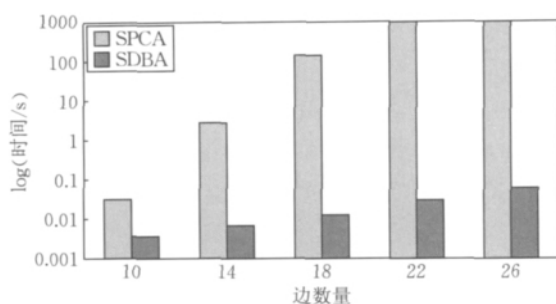
实验 1. 首先考察图的规模对 SPCA 算法与 SDBA 算法运行时间与消耗内存方面的影响, 实验选择第(1)类数据中容量及概率均为随机的组合作为数据集.

从图 5(a)可以看出, 随着边数的增加, 两种算法运行时间的增长速率有较大差别, SPCA 算法因边数的增加导致源点与汇点间简单路径数量增加, 简单路径组合数量将指数增长, 而 SDBA 算法也因边数的增加会导致状态空间变大, 算法需要的空间划分次数增多, 运行时间变长, 但相对于 SPCA 算法, 增涨的幅度明显较为缓慢, 在边数达到 22 时, SPCA 算法已经不能在有效的时间内运行出来; 从图 5(b)可以看出, 随着边数的增加, 算法消耗的内存会随之增加, 可以看出 SPCA 算法空间性能优于 SDBA 算法, 这是由于 SPCA 算法消耗的内存主要用于存储约束条件, 而 SDBA 算法则需要存储迭代划分过程产生的子状态空间, 导致内存消耗稍高, 但随图规模增长其增幅并不明显. 从总体来看, SDBA 算法的整体性能优于 SPCA 算法, 考虑到 SPCA 算法适应性较弱, 后续的实验 2、实验 3 主要针对 SDBA 算法展开.

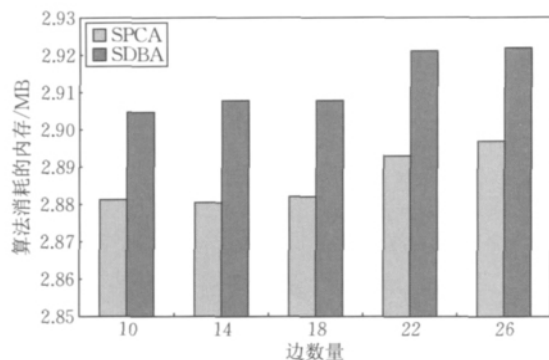
另外实验中对于每一组数据, 两种不同算法得到最大流的可靠性都是相等的, 这也从实验角度再次验证了算法的正确性.

实验 2. 考察不确定图边上容量分布与概率分布对 SDBA 算法运行时间的影响, 实验使用第(1)类数据集, 结果如图 6 所示.

图 6(a)可以看出, 边数大于等于 22 且容量服从均匀分布时, SDBA 算法的运行时间会略低于容量随机取值的情况, 同时可以看出, 在容量都服从均

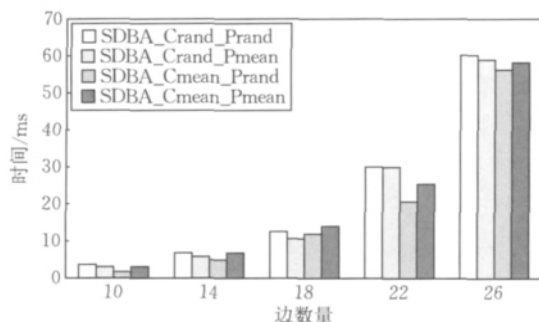


(a) 运行时间与边数量关系



(b) 内存消耗与边数量关系

图 5 算法性能与边数量关系



(a) SDBA在4种分布组合下运行时间

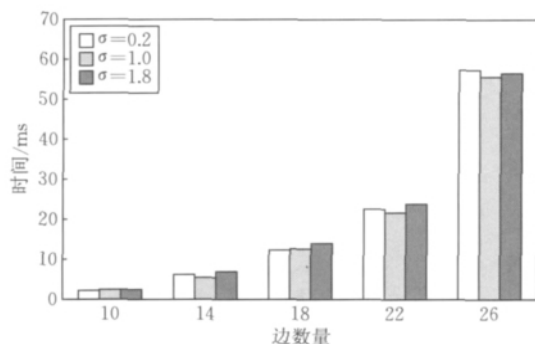
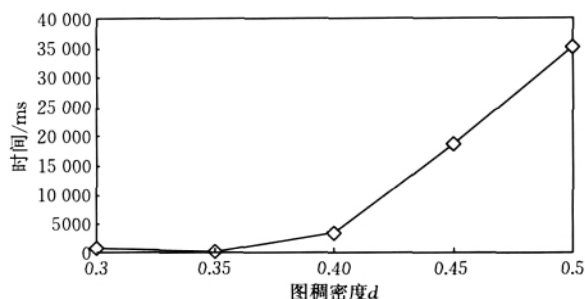
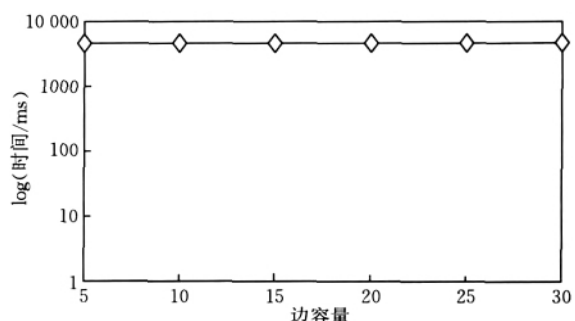
(b) 概率正态分布参数 $\sigma$ 不同取值下SDBA运行时间

图 6 容量概率分布对 SDBA 算法影响

均匀分布的情况下, 概率随机取值时算法运行时间相对均匀分布情况较短, 而在容量都随机取值情况下, 概率服从均匀分布时算法运行时间相对随机情况较短; 图 6(b) 说明在其它条件相同情况下, 概率服从相同均值  $\mu$ , 不同均方差  $\sigma$  的正态分布, 对 SDBA 算

法的运行时间影响可以忽略不计。

实验 3. 考察图的稠密性与边的容量取值对 SDBA 算法运行时间的影响, 实验使用第(2)、(3)类数据集, 结果如图 7 所示。

(a) 不同图稠密度  $d$  下 SDBA 运行时间

(b) 相同条件不同容量下 SDBA 运行时间

图 7 稠密性与容量对 SDBA 算法影响

从图 7(a) 可以看出, SDBA 算法运行时间会随着图稠密度  $d$  的提高而增加, 主要原因是图稠密性会直接影响状态划分算法的状态空间大小, 稠密度越高, 状态空间越大, 算法需要的迭代次数也会随之增加, 因而算法运行时间会增大; 图 7(b) 说明在其它条件相同的情况下, 边的容量对 SDBA 算法运行时间的影响很小, 可以忽略不计。

## 6 结 论

本文首先基于可能世界模型提出了不确定图的最可靠最大流问题和可靠性计算模型. 然后提出了一种基于简单路径组合的基本算法 SPCA. 针对基本算法适用性与效率方面存在的问题, 又提出了一种更加有效且适用性更强的基于空间划分的算法 SDBA. 实验结果表明本文提出的算法能够有效地解决不确定图最可靠最大流问题. 相对于 SPCA 算法而言, SDBA 算法不会受到简单路径数量与瓶颈容量的影响, 且性能相对较好。

本文的下一步工作一方面是研究算法的可扩展性, 希望通过图的分割等策略实现算法的多机并行,



另一方面研究如何将算法拓展至多态模型及流可靠性 top- $k$  查询等方面。

## 参 考 文 献

- [1] Hintsanen P. The most reliable subgraph problem//Proceedings of the 11th European Conference on Principles and Practice of Knowledge Discovery in Databases. Warsaw, 2007: 471-478
- [2] Hintsanen P, Toivonen H. Finding reliable subgraphs from large probabilistic graphs. *Data Mining and Knowledge Discovery*, 2008, 17(1): 3-23
- [3] Zou Zhao-Nian, Li Jian-Zhong, Gao Hong, Zhang Shuo. Mining frequent subgraph patterns from uncertain graph data. *IEEE Transactions on Knowledge and Data Engineering*, 2010, 22(9): 1203-1218
- [4] Zhang Shuo, Gao Hong, Li Jian-Zhong, Zou Zhao-Nian. Efficient query processing on uncertain graph databases. *Chinese Journal of Computers*, 2009, 32(10): 2066-2079(in Chinese)  
(张硕, 高宏, 李建中, 邹兆年. 不确定图数据库中高效查询处理. *计算机学报*, 2009, 32(10): 2066-2079)
- [5] Han Meng, Zhang Wei, Li Jian-Zhong. RAKING: An efficient  $K$ -maximal frequent pattern mining algorithms on uncertain graph database. *Chinese Journal of Computers*, 2010, 33(8): 1387-1395(in Chinese)  
(韩蒙, 张伟, 李建中. RAKING: 一种高效的不确定图  $K$ -极大频繁模式挖掘算法. *计算机学报*, 2010, 33(8): 1387-1395)
- [6] Potamias M, Bonchi F, Gionis A, Kollios G.  $K$ -nearest neighbors in uncertain graphs//Proceedings of the VLDB. Singapore, 2010: 997-1008
- [7] Zhang Hai-Jie, Jiang Shou-Xu, Zou Zhao-Nian. An efficient algorithm for top- $k$  proximity query on uncertain graphs. *Chinese Journal of Computers*, 2011, 34(10): 1885-1896(in Chinese)  
(张海杰, 姜守旭, 邹兆年. 不确定图上的高效 top- $k$  近邻查询处理算法. *计算机学报*, 2011, 34(10): 1885-1896)
- [8] Yuan Ye, Wang Guo-Ren. Answering probabilistic reachability query over uncertain graphs. *Chinese Journal of Computers*, 2010, 33(8): 1378-1386(in Chinese)  
(袁野, 王国仁. 面向不确定图的概率可达查询. *计算机学报*, 2010, 33(8): 1378-1386)
- [9] Hua M, Pei J. Probabilistic path queries in road networks: Traffic uncertainty aware path selection//Proceedings of the 13th International Conference on Extending Database Technology (EDBT 2010) Lausanne. Switzerland, 2010: 347-358
- [10] Kurzhanski A B, Varaiya P. On reachability under uncertainty. *SIAM Journal on Control and Optimization*, 2002, 41(1): 181-216
- [11] Rasteiro D D M L, Anjo A J B. Optimal paths in probabilistic networks. *Journal of Mathematical Sciences*, 2004, 120(1): 974-987
- [12] Doulliez P, Jamoulle E. Transportation networks with random arc capacities. *RAIRO*, 1972, 6(3): 45-59
- [13] Alexopoulos C. Note on state-space decomposition methods for analyzing stochastic flow network. *IEEE Transactions on Reliability*, 1995, 44(2): 354-357
- [14] Jane Chin-Chia, Lai Yih-Wenn. A practical algorithm for computing multi-state two-terminal reliability. *IEEE Transactions on Reliability*, 2008, 57(2): 295-302
- [15] Jane Chin-Chia, Lai Yih-Wenn. Computing multi-state two-terminal reliability through critical arc states that interrupt demand. *IEEE Transactions on Reliability*, 2010, 59(2): 338-345
- [16] Zhao Pei-Xin, Zhang Xin. A survey on reliability evaluation of stochastic flow networks in terms of minimal path//Proceedings of the International Conference on Information Engineering and Computer Science. Wuhan, China, 2009
- [17] Lin J S, Jane C C, Yuan J. On reliability evaluation of a capacitated flow network in terms of minimal path sets. *Networks*, 1995, 25(3): 131-138
- [18] Lin Yi-Kuei. System reliability evaluation for a multi-state supply chain network with failure nodes using minimal paths. *IEEE Transactions on Reliability*, 2009, 58(1): 34-40
- [19] Ramirez-Marquez J E, Coit D W. A monte-carlo simulation approach for approximating multi-state two terminals reliability. *Reliability Engineering & System Safety*, 2005, 87(2): 253-264
- [20] Rocco C M, Muselli S M. Approximate multi-state reliability expressions using a new machine learning technique. *Reliability Engineering & System Safety*, 2005, 89(3): 261-270
- [21] Ahuja R K, Magnanti T L, Orlin J B. *Network Flows: Theory, Algorithm and Applications*. Beijing: China Machine Press, 2005
- [22] Jane C C, Lin J S, Yuan J. On reliability of a limited-flow network in terms of minimal cut sets. *IEEE Transactions on Reliability*, 1993, 42(3): 354-361
- [23] Lin Yi-Kuei. A simple algorithm for reliability evaluation of a stochastic flow network with node failure. *Computer & Operations Research*, 2001, 28(13): 1277-1285
- [24] Lee S H. Reliability evaluation of a flow network. *IEEE Transactions on Reliability*, 1980, 29(1): 24-26
- [25] Klingman D, Napier A, Stutz J. NETGEN: A program for generating large scale capacitated assignment, transportation and minimal cost flow network problems. *Management Science*, 1974, 20(5): 814-821
- [26] Han W S, Lee J, Pham M D. iGraph: A framework for comparisons of disk-based graph indexing techniques//Proceedings of the 36th International Conference on Very Large Data Bases (VLDB). Singapore, 2010: 449-559



**CAI Wei**, born in 1987, M. S. candidate. His research interests focus on uncertain graph data management.

**ZHANG Bai-Li**, born in 1970, Ph. D. , associate professor. His current interests include WEB, WSN and data warehousing.

**LV Jian-Hua**, born in 1977, Ph. D. , associate professor. His current interests include WSN, data stream and data mining.

### Background

This work is supported by the National Natural Science Foundation of China under Grant Nos 61073059 and 61232007.

Currently, researchers focus on the management of uncertain graph data. Some meaningful research problems and useful solutions are put forward. This paper proposes a new

research problem, the most reliable maximum flow problem (MRMF) on uncertain graphs. It is meaningful to solve many practical problems, such as the reliable network construction and the reliable transportation path selection. To solve the new problem, SPCA and SDBA algorithms are proposed.