**partial order based subgraph containment query**

A Dissertation Submitted to

Southeast University

For the Academic Degree of Master of Engineering

BY

Li Fu-hao

Supervised by

Associate Professor Zhang Bai-li

School of Computer Science and Engineering

Southeast University

June 2016

**东 南 大 学 学 位 论 文 独 创 性 声 明**

本人声明所呈交的学位论文是我个人在导师指导下进行的研究工作及取得的研究成果。尽我所知，除了文中特别加以标注和致谢的地方外，论文中不包含其他人已经发表或撰写过的研究成果，也不包含为获得东南大学或其它教育机构的学位或证书而使用过的材料。与我一同工作的同志对本研究所做的任何贡献均已在论文中作了明确的说明并表示了谢意。

研究生签名： 日 期：

**东 南 大 学 学 位 论 文 使 用 授 权 声 明**

东南大学、中国科学技术信息研究所、国家图书馆有权保留本人所送交学位论文的复印件和电子文档，可以采用影印、缩印或其他复制手段保存论文。本人电子文档的内容和纸质论文的内容相一致。除在保密期内的保密论文外，允许论文被查阅和借阅，可以公布（包括刊登）论文的全部或部分内容。论文的公布（包括刊登）授权东南大学研究生院办理。

研究生签名： 导师签名： 日 期：

摘 要

在实际的应用中，不确定性是许多系统的固有特性，如电力传输网络中的元件、数据通信网络中的节点都存在着发生故障的概率，交通运输网络中道路有发生拥塞的概率等。可以将此类的问题建模成不确定图，从而将问题转化为不确定图的问题来解决。根据最大流最小割定理，网络的最大流受限于网络中存在的最小割集，也就是说最小割集中的任意一条边断掉都会使整个网络可以传输的最大流量下降，这样的边就成了整个网络相对于流量的薄弱环节。同样的，不确定图中的边可以看做是从源点到汇点的信息传输通道，而关键边对于网络的信息传输可靠性起到支撑的作用，一旦发生故障，将对整个网络的信息传输产生巨大影响，甚至导致整个网络瘫痪，这样的边就构成了网络相对于可靠性的薄弱环节。为此，针对不确定图中边对于可靠性的重要度进行评估，寻找关键边就成为一项有意义的工作，这对于后期网络维护具有重大的决策和指导意义，对于薄弱环节的改进也可以显著的提高系统的可靠性。

基于此，本文将从一个新的角度，即从不确定图流分布的可靠性和流量的角度来构建边关键度识别模型，根据边对于不确定图可靠性的影响，对关键边进行识别。同时为更加准确，多角度的体现出不确定图中边的关键程度，本文在流分布可靠性指标的基础上添加不确定图能够达到最大流量大小以及对应的随机流网络的可靠性这两个指标，多角度、综合对关键边进行识别。

为研究该问题，本文首先基于状态划分算法，提出一种以增量的方式计算移除边之后不确定图最可靠最大流可靠性的变化，该方法利用边移除之前计算状态的中间结果，根据边在子图区间的存在多少的状况对不确定图的边进行定性分类，然后结合求得的子图区间，定量的求移除边对于不确定图状态的影响。相对于重新计算而言，使用增量计算能有效减少复杂度。同时考虑到传统状态划分算法需要划分的状态空间较大，本节在原有计算方式的基础上提出一种基于确定边过滤的状态划分算法，该算法的主要思路是在使用状态划分算法之前，首先给出能够确定类别的边（定性的边），从而减少状态划分的空间。这些确定边包括割集中的边和悬挂边。

最后本文基于边移除后对流量及可靠性产生的损失这一系统论的角度提出（1）字典序模型、（2）化多目标为单目标模型、（3）多重排序模型，三种基于流量和流可靠性指标的边关键度识别模型，然后研究他们的算法性能，并根据归一化误差方法衡量三种模型在识别关键边上的优劣，最后结合智能电网这一具体的应用背景，综合研究这三个模型，以实现对关键边更灵活、更准确、更快速的识别。

关键词: 不确定图, 关键边, 流可靠性, 流量

Abstract

The importance of the network edge evaluation is an important content of research network. This paper constructed key edge evaluation model based on the flow and reliability （distribution reliability and capacity reliability） in uncertain graph, the key edge model for uncertain graph is on the flow and the reliability of relative loss after the edge is removed（or failure）, to comprehensive evaluation the degree of an edge. The model will consider flow as the most important factors for the key edge, when the flow is consistent, compared to the maximum flow distribution of reliability （after remove edge）, in order to increase the degree of differentiation, at the same time considering capacity reliability. As for the calculation of relative changes of the flow and reliability of uncertain graph after the edge being removed, at first, this paper designed the BASE algorithm, repeated computation algorithm after removing uncertain graph of maximum flow and distribution reliability and capacity reliability, this article consider BASE algorithm as a basic algorithm to measure the key edge; Then aiming at the shortcomings of high repeat BASE algorithm computational complexity, we propose a kind of incremental algorithm of ICA, the main ideas of the algorithm is to divided the edge into ABC three kinds based on the number that the edge in the subgraph which satisfy the maximum flow. When A edge to be removed, use STPA\_CUT algorithm, tree pruning cut edge is used to obtain the interval after removing still meet the maximum flow graph set; When B side to remove, use B - CESA algorithm to obtain still meet the range of maximum flow; Because C edge has not affect, so don't need to calculate. At the end of the paper, through the experiment compares the performance of the two algorithms. The experimental results show that the ICA algorithm relative to the BASE of its space complexity increases to a certain extent, but the time complexity has great advantage.

**Key words**: Uncertain Graph, Key Edge, maximum flow and distribution of the reliability, capacity reliability

目录

[摘 要 I](#_Toc449306438)

[Abstract II](#_Toc449306439)

[目录 III](#_Toc449306440)

[第1章 绪论 1](#_Toc449306441)

[1.1 问题研究背景与意义 1](#_Toc449306442)

[1.2 研究现状 2](#_Toc449306443)

[1.3 研究内容 4](#_Toc449306444)

[1.4 本文内容组织 5](#_Toc449306445)

[第2章 不确定图及关键边相关理论 7](#_Toc449306446)

[2.1 最大流问题 7](#_Toc449306447)

[2.1.1 基本概念和相关定义 7](#_Toc449306448)

[2.1.2 常见的最大流算法 8](#_Toc449306449)

[2.2 随机流网络可靠性相关问题 10](#_Toc449306450)

[2.2.1 基本概念和相关定义 10](#_Toc449306451)

[2.2.2常见的随机流网络容量可靠性算法 11](#_Toc449306452)

[2.3 不确定图数据 12](#_Toc449306453)

[2.3.1 不确定图及可能世界模型 12](#_Toc449306454)

[2.3.2 不确定图分布可靠性 13](#_Toc449306455)

[2.3.3 不确定图容量可靠性 15](#_Toc449306456)

[2.4 本章小结 16](#_Toc449306457)

[第3章 基于流量和容量可靠性的关键边衡量方法 17](#_Toc449306458)

[3.1 基于流量和容量可靠性的不确定图边关键度衡量模型 17](#_Toc449306459)

[3.2 基于重复计算的BASE\_FCP算法 20](#_Toc449306460)

[3.3 基于流量和容量可靠性的不确定图关键边增量算法ICA\_FCP 21](#_Toc449306461)

[3.3.1 算法思想 21](#_Toc449306462)

[3.3.2 基于割集的状态树剪枝A边算法STPA\_FCP 22](#_Toc449306463)

[3.3.3 B类边搜索算法BSA\_FCP 24](#_Toc449306464)

[3.3.4 算法复杂度分析 25](#_Toc449306465)

[3.4 Top-K关键边搜索策略 26](#_Toc449306466)

[3.5 实验及分析 26](#_Toc449306467)

[3.5.1 实验数据 26](#_Toc449306468)

[3.5.2 实验结果与分析 26](#_Toc449306469)

[3.6 本章小结 28](#_Toc449306470)

[第4章 基于流量和分布可靠性的关键边衡量方法 29](#_Toc449306471)

[4.1 基于流量和分布可靠性的不确定图边关键度衡量模型 29](#_Toc449306472)

[4.2 基于重复计算的BASE\_FDP算法 31](#_Toc449306473)

[4.3 基于流量和分布可靠性的不确定图的关键边增量算法ICA\_FDP 32](#_Toc449306474)

[4.3.1 算法思想 32](#_Toc449306475)

[4.3.2 A类边计算 33](#_Toc449306476)

[4.3.3 B类边计算 33](#_Toc449306477)

[4.3.4 算法复杂度分析 35](#_Toc449306478)

[4.4 基于增广路径添加的近似算法 36](#_Toc449306479)

[4.4.1 算法基本思想 36](#_Toc449306480)

[4.4.2 算法实现与分析 37](#_Toc449306481)

[4.4.3 近似算法相似度 38](#_Toc449306482)

[4.4.4 近似算法下界 38](#_Toc449306483)

[4.5 Top-K关键边搜索策略 40](#_Toc449306484)

[4.6 实验及分析 40](#_Toc449306485)

[4.6.1 实验数据 41](#_Toc449306486)

[4.6.2 实验结果与分析 41](#_Toc449306487)

[4.6.3 正确性（相关性）比较 42](#_Toc449306488)

[4.7 本章小结 42](#_Toc449306489)

[第5章 不确定图关键边增量算法优化策略 43](#_Toc449306490)

[5.1 基于确定边过滤的增量算法ICA\_FEA 43](#_Toc449306491)

[5.1.1 算法基本思想 43](#_Toc449306492)

[5.1.2 割集中的边计算 43](#_Toc449306493)

[5.1.3 悬挂边的计算 45](#_Toc449306494)

[5.1.4 基于确定边过滤的关键边优化算法ICA\_FEA 46](#_Toc449306495)

[5.1.5 实验及分析 47](#_Toc449306496)

[5.2 基于下界子图树的增量优化算法 47](#_Toc449306497)

[5.2.1 算法的基本思想 47](#_Toc449306498)

[5.2.2 构造下界子图树与子树搜索过程 47](#_Toc449306499)

[5.2.3 实验结果及分析 47](#_Toc449306500)

[5.3 本章小结 47](#_Toc449306501)

[第6章 总结和展望 48](#_Toc449306502)

[致谢 49](#_Toc449306503)

[参考文献 50](#_Toc449306504)

# 第1章 绪论

## 1.1 问题研究背景与意义

图是离散数学和计算机科学中一种重要的数据结构，它可以很好的表示数据的内部结构及数据之间的关联，因其能表示丰富的信息和复杂的结构而在众多的领域有着广泛的应用。对于实际网络系统如数据通讯网络、输变电网络、交通网络等，不确定性是其固有特性，这种不确定性或是由于测量数据存在误差，或是应用本身包含概率特征而引入。如电能传输网络中，输变电设备都有发生故障的可能性，而交通网络中每条道路都有一个发生拥塞的概率，当这种不确定性表达于图数据中，则形成不确定图[1-3]。

最大流问题是传统图论中一类经典的组合优化问题，在众多的领域有着广泛而深入的研究，一般情况下最大流通常对应多个不同的分布方案，对于确定图而言，这些不同的分布方案对于流的传递是等价的，但是对于边、节点都可能存在不确定性的不确定图来说，不同的流分布方案对应着不同的可靠性，从而存在着最可靠最大流问题。例如：在输变电网络中, 电能从发电站A点传输到目的地B点可能存在着多种传输方案，而这些方案由于涉及到具体不同可靠性的输电设备和输电线路而具有不同的可靠性，为了保证电网电能传递具有更好的安全可靠性，就需要找到最可靠的一种传输电能的方案。

在不确定图中, 对于流分布的可靠性研究已经形成两种较为普遍的方法，一种是流量组合的算法[5]，另外一种是状态划分算法[6]。假设一个不确定图处于最可靠最大流的状态，能够体现不确定图的状态的有不确定图能够达到的最大流（连通性）及可靠性（包括随机流网络的可靠性和不确定图流分布的可靠性），然而由于不确定图中边的不确定性，任何一条边发生故障都有可能对于整个不确定图的状态造成影响，即任何一条边发生故障都可能使得不确定图不能达到原有的最大流最可靠性的状态（即原有的连通性和可靠性）。

根据最大流最小割定理，网络的最大流受限于网络中存在的最小割集，也就是说最小割集中的任意一条边断掉都会使整个网络可以传输的最大流量下降，这样的边就成了整个网络的薄弱环节，或者可以说是相对于流量的薄弱环节。同样的，不确定图中的边可以看做是从源点到汇点的信息传输通道，而关键边对于网络的信息传输可靠性起到支撑的作用，一旦发生故障，将对整个网络的信息传输产生巨大影响，甚至导致整个网络瘫痪[4]。为此，针对不确定图中边对于可靠性的重要度进行评估，寻找关键边就成为一项有意义的工作。同时，根据关键边对于网络的连通性和可靠性的影响程度，对关键边的关键程度进行衡量，对于后期网络维护具有重大的决策和指导意义，对于薄弱环节的改进也可以显著的提高系统的可靠性。

通过阅读文献[1-22]，对于随机流网络的可靠性的研究已经十分成熟，近些年来对于不确定图的可靠性研究也越来越多，然而很少有基于流可靠性指标的对于不确定图边关键度研究的案例，而这样一个问题又确实有着十分重要的研究价值，如：在电力传输网络中，发电站A正以最可靠最大流的分布向目的地B传输电能。这样一种传输分布方式有可能受到不确定图中任何一条边的影响，然而每条边发生故障对不确定图整个运行状态造成的影响大小是不一定的，有可能使不确定图的最大流减少（连通性改变），也有可能使不确定图的可靠性下降（可靠性改变），这样便体现出了边的关键性，同样的这些可靠性衡量指标又对于以后的不确定图模型的维护提供依据。

## 1.2 研究现状

针对最大流问题，它是网络流理论的重要组成部分，是一类经典的组合优化问题，同时也可以看作是特殊的线性规划问题，针对这个问题的研究有40多年的历史[2]。典型的最大流算法包括：网络单纯形法[1,5]﹑Ford-Fulkerson方法[3]﹑Edmonds-Karp方法[4]和Dinic方法[1,5]等，当前的最大流算法主要可以分为两大类：增载算法[5-7]和预流推进算法[8-10]。

针对随机流网络及其可靠性的研究，也一直是各研究者研究的热点问题[1-14]，网络的复杂性、动态性、多态性等特点使得传统成熟的可靠性评估方法，如可靠性框图（RBD）法、故障树分析（FTA）法等很难用于网络。目前，对网络可靠性的研究取得了很多成果，网络可靠性的内涵也由传统的基于网络拓扑结构的连通可靠性逐渐拓展到考虑网络流的容量可靠性，并向基于业务等考虑用户需求的性能可靠性延伸。

网络可靠性评估主要分为3类：（1）网络连通可靠性评估，指的是仅考虑网络拓扑结构，将“网络实现连通功能的概率”作为可靠性度量；(2)网络容量可靠性评估，它在考虑网络是否连通的基础上，还考虑了网络中链路和节点的容量，将“存在满足一定流量需求的连通路径的概率”作为可靠性度量；(3)网络性能可靠性评估，关注的是网络性能的动态变化对可靠性的影响，多以“某些性能参数不超过其规定阈值的概率”作为可靠性的度量[网络可靠性评估方法综述]。本文主要研究的是随机流网络的容量可靠性。

对于不确定图，近年来的大部分研究都是针对网络的可靠性问题研究展开的[6-20], 即给定一个流量约束阈值d，如何得到满足d的网络可靠度。目前，网络可靠性的计算已经证明是NP-hard问题。多状态（图中每条边上容量均服从一定的离散概率分布）网络可靠度的计算可以分为3类：两终端可靠性﹑全终端可靠性和k-终端可靠性。对于网络可靠性的算法通常可以分为两大类：精确的可靠性算法[11,13,18-20]和近似的可靠性算法[15]。精确的可靠性算法中最基本的就是穷举算法, 通过枚举所有可能的状态来得到最终网络可靠度, 在此基础上, 基于状态空间划分的方法在算法效率上则就有很明显的提高[16-20]。文献[10-12]和文献[13]则分别采用了基于路径集和割集的方式来求取网络可靠性。典型的近似算法包括：神经网路算法﹑遗传算法和Monte-Carlo[15]模拟算法等。此外，文献[12,14]还研究了不确定图网络中节点发生故障情况下可靠性问题。

在随机流网络中，每个网络元件由于完全故障或部分故障等诸多因素而存在多种状态，也就是说每个网络元件有多种容量状态，而这些容量对应的概率服从一定的概率分布。随机流网络可靠性问题是一种概率问题，描述的是从网络给定的源点s到汇点t成功传递流量需求d的概率，该问题已经被证明是NP-hard[7-8]。当前，可靠性的分析已经成为许多系统规划、设计和运行中不可缺少的重要组成部分[9]。

关于随机流网络可靠性的算法研究大致可以分为两大类：精确方法和近似方法。精确方法中完全枚举算法是最基本的算法，它通过直接枚举网络的所有可能状态得到网络的可靠度。Chin-Chia Jane与Yih-Wenn Laih采用了基于d-flows[10]和基于d-cuts[11]的不同状态空间划分方法，将状态空间划分成能够满足流量阈值需求的合格状态集，不能够满足流量阈值需求的不合格状态集和同时包含了合格状态和不合格状态的待划分状态集，通过迭代的方式进一步处理待划分状态集，最终通过计算合格状态集中所有合格状态的概率之和来得到网络可靠性。一些研究人员还提出了基于最小路径集[12-14]或最小割集[15]的算法对状态空间进行过滤，然后运用容斥原理来计算网络可靠性。对于复杂的系统，精确算法的运行时间代价往往过高而让人难以容忍，为了从准确性和算法执行效率中找到一种折中的方案，实际系统中通常会选取近似算法来代替精确算法。为此，许多近似算法被相继提出，Ramirez-Marquez和Coit[16]提出了蒙特卡洛模拟算法， Rocco和Muselli[17]则采用机器学习方法成功的获得网络可靠性。

Lin[14,18-19]还将网络可靠性问题扩展到节点不可靠的情况，采用最小路径集的方法进行了研究。考虑到网络中的每条边及节点都有一定的传输代价参数，Lin and Yeh 研究了在实际应用中基于传输代价限定的网络可靠性问题 [20-24]，这其实是一类具有多条件约束的网络可靠性问题。对于一个拓扑结构给定的网络，如何从一堆给定的多状态元件中选取该网络需求数量的元件来构建一个实际的网络，同时又需要满足该网络所要满足的传输流量、代价、时间、可靠性等诸多因素限制的实际需求，也是一类比较经典的网络元件分配与网络可靠性结合的研究问题。针对这类问题，Yi-Kuei Lin和Cheng-Ta Yeh 提出了相应的遗传算法[25]。

不确定图流分布的可靠性是网络流问题在不确定图上的自然延伸，它是一类重要的最优决策问题，关系到如何最大限度的利用现有网络条件选择最可靠的最大流传输方案，对解决实际中诸如构建可靠性网络、选取可靠传输路线以及分析系统薄弱环节等一系列问题有重要的研究意义[34]。目前不确定图流分布的可靠性的计算方法主要是基于简单路径组合分流的最可靠最大流算法SPCA和基于极小子图的最可靠最大流算法MSBA，这两类算法分别从不同的角度有效地解决不确定图最可靠最大流问题[35]。

寻找网络中关键边的方法有很多，1989年Ball等人提出了通过移除网络中某条边后最短路径的变化情况的方法[23]来判断该条边的重要性，但对于节点之间仅有一条路径连接的情况，该方法就不适用；2002年Girvan和Newman在介数中心度的基础上，提出了边介数[24]（edge-betweenness）概念。通过计算网络中边介数的大小来反应边对网络资源传输能力和控制能力的强弱。边介数越大，表明网络中任意节点对经过该边的次数就越多，对网络资源的传输能力和控制能力就越强，在网络中所起到的作用也就越大。因此，边介数在一定程度上反应了边的重要程度。但仅仅以边介数作为度量关键边的方法[25]忽视了边的2个端点对其本身的影响，具有片面性[4]。目前对于网络中边关键性的衡量，大多都是判断边对网络资源传输能力或控制能力强弱来判断，如随机流网络中的一条边E发生故障，使得随机流网络能够达到的最大流减少大小等指标进行判断，很少有使用边发生故障对于随机流网络的可靠性造成的影响进行判断的；在不确定图中则对应的是流分布的可靠性。

评估网络中节点重要性的方法很多，最简单的方法是以节点的连接度（节点连接的边数）作为节点重要度的衡量标准[6]，认为与节点相连的边越多则该节点越重要。这种评估方法具有片面性，有些重要的“关键节点”并不一定具有较大的连接度，比如只有两条边相连的“桥节点”。2002年Girvan和Newman在介数中心度的基础上，提出了边介数[7]（edge-betweenness）概念，通过计算网络中边介数的大小来反映边对网络资源的传输能力和控制能力的强弱，边介数越大，表明网络中任意节点对经过该条边的次数就越多，对网络资源的传输能力和控制能力就越强，在网络中所起到的作用也就越大，因此，边介数在一定程度上反映了边的重要程度；文献[8]提出了一种基于生成树数目的节点删除法，定义最重要的节点为去掉该节点使得生成树数目最小的节点。节点删除法的问题是如果多个节点的删除都使得网络不连通，那么这些节点的重要度将是一致的，从而使得评估结果不精确。针对已有复杂网络节点重要性评估方法中片面强调节点的度而忽略了边对与之相连节点的支撑作用的缺陷，熊金石等人提出了节点度和边介数共同作用下的评估数学模型[9]，以体现边对其端节点的支撑作用。

目前对于随机流网络的可靠性的研究已经相当成熟，对于不确定定图流分布可靠性的研究也已经达到一定的阶段，但是尚没有或很少有基于流可靠性指标的对于不确定图边关键度研究的案例，而这样一个问题又确实有着十分重要的研究价值，本文将从此着手，基于流量及其可靠性指标，对于不确定图的关键边识别。

## 1.3 研究内容

本文主要是针对不确定图的关键边进行研究，首先对于问题相关概念进行了介绍，然后提出了基于流量和容量可靠性的不确定图关键边评估模型，基于此模型提出了增量算法ICA\_FCP。而由于计算不确定图的容量可靠性复杂度较高，问题接着提出了一种基于流量和不确定图分布可靠性的关键边衡量模型，并基于此模型提出了一种增量算法ICA\_FDP，因为计算不确定图的容量可靠性需要获取所有满足给定容量的子图，而根据极小子图理论，不确定图的分布可靠性计算只需要在状态划分的过程中，比较子图集合的下界即可，这在时间和空间复杂度上都有所减少。但是，因为计算不确定图分布可靠性依然是一个NP-Hard问题，所以本文接着提出另一种基于增广路径添加的近似算法，该算法避免了不确定图边移除之后的重复计算，使得复杂度大大降低。最后，通过实验分析，比较各算法的时间和空间复杂度，同时验证了近似算法对于不同稠密度和图规模下的性能波动。

综上所述，本文的主要工作如下：

（1）基于流量和可靠性构建了不确定图边关键度识别模型。

目前对于网络中边关键度的衡量，一般根据边介数的大小来反应边对网络资源传输能力和控制能力的强弱[24]，也有通过移除网络中某条边后最短路径的变化情况或者流量变化的情况[23]来判断该条边的重要性，然而使用这些方法求得的关键边，比较单一，如通过移除边后流量变化的情况获得的关键边，只体现出边对于流量的影响，这样获得的关键边是单一的、片面的。本文从一个新的角度，构建了基于流量和不确定图容量可靠性为指标的关键边识别模型，根据边移除之后，对于不确定图的流量和容量可靠性的相对影响来评价。为减少计算的复杂度，文本又提出一种基于流量和分布可靠性的关键边识别模型，该模型只需要计算分布可靠性，使得算法的空间和时间复杂度都有所减少。经研究，涉足这方面的研究知之甚少。

（2）快速计算边发生故障后不确定图最大流，容量可靠性和分布可靠性的计算。

本文主要研究的是基于流量和可靠性（容量可靠性和分布可靠性）的不确定图关键边识别，针对不确定图中边移除之后，对于流量和可靠性的相对影响来综合识别。关于移除边前后计算不确定图的状态，主要包括不确定图的最大流，最可靠最大流分布以及网络可靠性。

如何计算最可靠最大流分布，现已有了较深入的研究[1-10]。通过重复计算移除边前后状态，获得状态改变量，显然是可行的，但该方法复杂度较高，难以应对大规模图模型。本文提出一种增量的计算方法，利用边移除之前计算状态的中间结果，根据边在子图区间的存在多少的状况对不确定图的边进行定性的分类，然后结合求得的子图区间，定量的求移除边对于不确定图状态的影响。相对于重新计算而言，使用增量计算能有效减少复杂度。

针对精确算法复杂度较高，难以适应大规模图的情况，本文提出了一种基于增广路径添加的近似算法，该算法避免了不确定图边移除之后的重复计算，使得复杂度大大降低。最后，通过实验分析，比较各算法的时间和空间复杂度，同时验证了近似算法对于不同稠密度和图规模下的性能波动。

（3）针对不确定图关键边增量算法的优化策略

考虑到传统状态划分算法需要划分的状态空间较大，在原有计算方式的基础上提出一种基于确定边过滤的状态划分算法，该算法的主要思路是在使用状态划分算法之前，首先给出能够确定类别的边（定性的边），从而减少状态划分的空间。这些确定边包括割集中的边和悬挂边。同时考虑到在状态划分的过程中，需要多次比较区间下界是否能够满足最大流，本文为简化这个过程，提出了基于下界子图树的优化算法，每次在需要比较的过程中，通过搜索下界子图树，快速判断区间下界是否满足最大流，从而减少很多次的最大流计算，进而减少算法复杂度。

## 1.4 本文内容组织

本文的内容组织如下：

第1章 绪论。本章论述了不确定图等相关问题的研究背景，以及不确定图中关键边识别的意义，并调研了该问题的研究现状，简要介绍了本文研究的主要内容。

第2章 不确定图及关键边相关理论。本章主要介绍了一些基本概念。由于本文是基于网络最大流、随机流网络可靠性和不确定图分布可靠性对不确定图关键边进行的相关研究，因此主要介绍这些概念的相关研究。首先介绍了最大流的相关概念和常用的算法，然后介绍了随机流网络可靠性的基本概念，最后介绍了不确定图的相关概念和不确定图数据的一般建模方法，以及不确定图流分布及其可靠性的相关问题。对相关概念给出了详细的实例，方便理解。

第3章 基于流量和容量可靠性的关键边衡量方法。针对不确定图关键边的衡量，本章节构建了一种基于流量和容量可靠性的关键边衡量模型，使用不确定图中边移除（故障）后对流量和容量（最大流）可靠性产生的相对损失这一角度，对边的关键度进行综合评估。对于构建的基于流量和容量可靠性的模型，本章节首先设计了一种基于重复计算的基本算法BASE\_FCP。因为BASE\_FCP算法复杂度较高，本章提出一种基于流量和容量可靠性的不确定图关键边增量算法ICA\_FCP，并重点介绍了算法的思想与实现，同时对算法进行了正确性证明以及时间复杂度和空间复杂度的分析。紧接着，针对实际情况的需要，在很多时候，并不需要把所有不确定图的边的关键程度都衡量出来，因此，本章同时还提供了一种针对TopK需求的搜索策略。最后通过实验分析，比较BASE\_FCP算法和ICA\_FCP算法在空间复杂度和时间复杂度上的差别。

第4章 基于流量和分布可靠性的关键边衡量方法。本章主要构建了一种基于流量和分布可靠性的关键边衡量模型，针对不确定图边移除之后，相对流量和分布可靠性的减少量来衡量。对于构建的基于流量和容量可靠性的模型，本章节首先设计了一种基于重复计算的基本算法BASE\_FDP。因为BASE\_FDP算法复杂度较高，本章提出一种基于流量和分布可靠性的不确定图关键边增量算法ICA\_FDP，并重点介绍了算法的思想与实现，同时对算法进行了正确性证明以及时间复杂度和空间复杂度的分析。因为不确定图计算分布可靠性依然是一个NP-Hard问题，精确算法复杂度很高，因此本文章针对增量算法的A类边提出了对于A边的基于增广路径添加的近似算法，同时分析了算法复杂度，给出了近似算法和精确算法的相似度度量，并分析了近似算法的下界。最后通过试验分析，比较BASE\_FDP算法和ICA\_FDP算法在空间复杂度和时间复杂度上的差别。

第5章 主要介绍了针对不确定图关键边增量算法的优化策略，首先，考虑到传统状态划分算法需要划分的状态空间较大，在原有计算方式的基础上提出一种基于确定边过滤的状态划分算法，这些确定边包括割集中的边和悬挂边。同时，考虑到在状态划分的过程中，需要多次比较区间下界是否能够满足最大流，本章为简化这个过程，提出了基于下界子图树的优化算法，每次在需要比较的过程中，通过搜索下界子图树，快速判断区间下界是否满足最大流，从而减少很多次的最大流计算，进而减少算法复杂度。

第6章 总结和展望。本章总结了本文所做的主要工作，并对下一步的研究进行展望。

# 第2章 不确定图及关键边相关理论

本文是基于网络最大流、随机流网络可靠性和不确定图分布可靠性对不确定图关键边进行的相关研究。因此，本章在2.1节首先介绍了最大流的相关概念和常用的算法；2.2节介绍了随机流网络可靠性的基本概念；2.3节介绍了不确定图的相关概念和不确定图数据的一般建模方法，以及不确定图流分布及其可靠性的相关问题。

## 2.1 最大流问题

### 2.1.1 基本概念和相关定义

最大流问题（maximum flow problem）一种组合最优化问题。网络流理论研究的一个基本问题是求网络中一个可行流f，使其流量达到最大，这种流f称为最大流，这个问题称为（网络）最大流问题。最大流问题是一个特殊的线性规划问题，是图论的一个重要分支。一般用标号法寻求最大流比用求线性规划问题的一般方法要方便得多。为了系统的描述最大流问题，本节首先给出了最大流的基本概念和相关定义。

**定义2-1.[26**,蔡伟的引用**] （流网络）** 一个流网络*G*（*V*,*E*）是一个有向图，它满足以下两个特征：（1）对于任意的边*e*（*u*, *v*）*∈E*，均有一个非负容量*c*（*e*）≥ 0；（2）存在单一的源点*s*与单一的汇点*t*，且*s∈V*, *t∈V*。其中，*V*是*G*的顶点集，*E*是*G*的边集。

**定义2-2.（*s-t*流）** 在流网络*G*中，给定源点*s*，汇点*t*，*s-t*的流是一个边集*E*到实数集*R*的映射*f* : *E* → *R*，且*f*满足如下两条性质：

（1） 容量约束：对所有的*e∈* *E* ，*f*（*e*） ≤ *c*（*e*）；

（2） 流守恒性：对所有顶点*u*和*v* （*u* ∈ *V*，*v* ∈ *V*-{*s*, *t*}），满足Σ*e∈*{（*u*, *v*）|（*u*, *v*） *∈ E*} *f*（*e*） = Σ*e∈*{（*v*, *u*）|（*v*, *u*） *∈ E*} *f*（*e*），即流入顶点*v*的流量之和等于从顶点*v*流出的流量之和，对于源点*s*与汇点*t*满足Σ*e∈*{（*s*, *u*）|（*s*, *u*） *∈ E* } *f*（*e*） - Σ*e∈*{（*u*, *s*）|（*u*, *s*） *∈ E* } *f*（*e*） = Σ*e∈*{（*u*, *t*）|（*u*, *t*） *∈ E* } *f*（*e*） - Σ*e∈*{（*t*, *u*）|（*t*, *u*） *∈ E* } *f*（*e*），表示从源点流出的净流量（*s*流出的所有流量与流入*s*的所有流量之差）最终都会流入汇点，并且该净流量为*s-t*流值。

**定义2-3.（最大流）** 在流网络*G*中，给定源点*s*，汇点*t*，其中最大的s-t流值为s到t的最大流。

如图2-1所示，（a）为流网络，（b）为*s-t*流，源点*s*到汇点*t*的最大流为2。

（a）流网络 （b）*s-t*流

图2-1 流网络及*s-t*流

在实际应用中，许多最大流问题中的容量及流量通常都是整数，而对于容量及流量为有理数的情况，也可以通过适当的比例变换，将它们转换成整数，因此，为了将问题一般化，本文是在整数容量及流量的基础下进行研究的。

### 2.1.2 常见的最大流算法

关于最大流问题的相关算法有很多，其中比较基础的算法是Ford-Fulkerson[27]算法，该算法依赖三种重要的思想：剩余图，增广路径和割[26]，它们也是其它许多最大流算法的实现基础。本节将主要介绍一些常用的最大流算法及算法对应的时间复杂度。

（1）Ford-Fulkerson算法

Ford-Fulkerson是目前基本的解决最大流的方法，算法是一种迭代算法，它的核心思想是：在初始状态下，对所有的*u*, *v ∈ V*，*f*（*u*, *v*）=0，即每条边上流量的初值等于0，然后，在算法的每次迭代过程中，通过不断在*f*对应的“剩余图”*Gf*上寻找一条“增广路径”*p*来增加流值。其中“增广路径”可以简单的看成是从源点*s*出发到汇点*t*的一条路径，沿着这条路径可以将*s*的流量更多的压入到*t*，从而增加流*f*的值。反复执行这个过程，直到*Gf*中不存在增广路径为止。下面给出的是Ford-Fulkerson算法的伪代码实现。

|  |
| --- |
| **算法2-1.** 最大流算法Ford-Fulkerson算法 |
| **输入**：流网络*G*,源点*s*, 汇点 *t*  1. **for** each edge （*u,v*） ∈ *E*（*G*）  2. *f*（*u, v*） ← 0  3. *f*（*v, u*） ← 0  4. *Gf* ← *G*  5. **while** there exists a path *p* from *s* to *t* in *Gf*  6. *cf（p）* ← min{*cf*（*u,v*） where （*u,v*） ∈ *p*}  7. **for** each edge （*u,v*） ∈ *p*  8. *f*（*u, v*） ← *f*（*u,v*） + *cf（p）*  9. *f*（*v, u*） ← - *f*（*u,v*）  10. recomputed *Gf* for flow *f*  **输出**：*s*-*t*最大流 |

下面对Ford-Fulkerson算法相关的一些概念及理论进行说明。

**定义2-4. （剩余图）** 对于给定的流网络*G*（*V*, *E*），源点为*s*，汇点为*t*，设*f*为*G*中*s-t*流，则*f*在*G*中对应的剩余图*Gf* （*V’*，*E’*）定义如下：

1. *V’* = *V*；
2. 对顶点*u*, *v* *∈ V*，如果存在边*e*（*u*, *v*） *∈ E*，则存在*e’*（*u*, *v*） *∈ E’*， 且*e’* 容量*c*（*e’* ） = *c*（*e*）–*f*（*e*），此时称该边为正向边；
3. 如果*G*中边*e*（*u*, *v*）上的流量*f*（*e*） ≠ 0，则在*Gf*中存在从*v*到*u*的边*e*（*v*, *u*），其容量*c*（*e’* ） = *f*（*e*），此时称该边为反向边。

如图2-2，*Gf*是图2-1（b）中流*f*对应于2-1（a）中流网络*G*的剩余图。



图2-2 剩余图*Gf*

在最大流问题的研究中，最大流最小割定理是最大流算法分析的理论基础，下面简单介绍一下割的概念及该定理。

**定义2-5. （*S*-*T*割）** 流网络*G*（*V*, *E*）的割（*S*, *T*）是顶点集*V*的一个划分，它将*V*划分成*S*和*T*（*T* = *V* - *S*）两个部分，且该划分满足源点*s ∈ S*， *t ∈ T*，同时，将割的容量定义为从*S*到*T*的边容量之和，即*c*（*S*, *T*） = Σ*e*∈{（*u*, *v*） | *u**∈**S* ∧ *v**∈ T* ∧ （*u*, *v*） *∈ E*} *c*（*e*）。

**定理2-1.[26]（最大流最小割定理）**在流网络*G*（*V*, *E*）中，*f*是源点*s*到汇点*t*的一个流，则下面三个条件具有等价关系：

1. *f*是*G*的一个最大流。
2. 剩余图*Gf*中不包含*s*-*t*增广路径。
3. *G*中存在某个割（*S*, *T*）满足| *f* | = *c*（*S*,*T*）。

最大流最小割定理不仅说明了Ford-Fulkerson算法的正确性，同时也说明了网络的最大流等于最小割。

Ford-Fulkerson算法的运行时间与增广路径的选取密切相关，算法运行时间的上界为*O*（|*E||fmax*|）[26]。如果| *fmax* |很大，而迭代过程每次选择的增广路径能够增加的流量比较小，则算法的效率会很低。

为了合理的选择增广路径，Edmonds和Karp[28]采用广度优先的搜索方式，在剩余图中，将每条边的距离看成单位1，寻找*s*到*t*的最短路径作为增广路径来增加流。已经证明Edmonds-Karp算法的时间复杂度为*O*（|*V*||*E*|2）。

（2）Dinic算法

Dinic[29]算法是基于分层网络和阻塞流思想的最大流算法，它将剩余图*Gf*中的每条边长度看成1，在每次迭代过程中，根据*Gf*中的顶点与源点*s*的距离来构建*Gf*对应的分层网络*MSN*，通过在*MSN*中寻找*s*到*t*的阻塞流对原始流*f*进行更新，直到汇点*t*不在当前的分层网络中时停止。下面给出的是Dinic算法的伪代码实现。

|  |
| --- |
| **算法2-2.** 最大流算法Dinic算法 |
| **输入：**流网络*G*,源点*s*, 汇点 *t*  1. **for** each edge （*u,v*） ∈ *E*（*G*）  2. *f*（*u,v*） ← 0  3. *f*（*v, u*） ← 0  4. *Gf* ← *G*  5. compute the *MSN* for *Gf*  starting from source *s*  6. **while** sink *t* is in *MSN*  7. find a blocking flow *f*\* in *MSN*  8. **for** each edge （*u,v*） in *G*  9. *f*（*u,v*） ← *f*（*u,v*） + *f*\*（ *u,v*）  10. compute *Gf*  for flow *f*  11. recomputed *MSN* for *Gf*  **输出：***s*-*t*最大流 |

Dinic算法和push-relabel[26]算法都是当前使用较为广泛的最大流算法，它们的时间复杂度均为*O*（|*V*|2|*E*|），而King, Rao和Tarjan[30]则提出了复杂度为*O*（|*V*||*E*|log|*E|*/（|*V|*/lg|*V|*）|*V*|）的最大流算法。

## 2.2 随机流网络可靠性相关问题

### 2.2.1 基本概念和相关定义

网络的可靠度是网络性能的一个重要指标，可靠性分析一直是各研究者研究的热点问题[1-14]，网络的复杂性、动态性、多态性等特点使得传统成熟的可靠性评估方法，如可靠性框图（RBD）法、故障树分析（FTA）法等很难用于网络。目前，对网络可靠性的研究取得了很多成果，网络可靠性的内涵也由传统的基于网络拓扑结构的连通可靠性逐渐拓展到考虑网络流的容量可靠性，并向基于业务等考虑用户需求的性能可靠性延伸。

网络可靠性评估主要分为3类：（1）网络连通可靠性评估，指的是仅考虑网络拓扑结构，将“网络实现连通功能的概率”作为可靠性度量；(2)网络容量可靠性评估，它在考虑网络是否连通的基础上，还考虑了网络中链路和节点的容量，将“存在满足一定流量需求的连通路径的概率”作为可靠性度量；(3)网络性能可靠性评估，关注的是网络性能的动态变化对可靠性的影响，多以“某些性能参数不超过其规定阈值的概率”作为可靠性的度量[网络可靠性评估方法综述]。本文主要研究的是随机流网络的容量可靠性。

根据网络或其组成部分的状态的多少网络可分为二态（binary state）网络和多态（multi-state）网络。

（1）二态（binary state）网络是指网络或网络的每个元素（节点或弧只有两种状态：好/坏。网络的每条弧的容量的取或为0，或为某一个整数。

（2）有的网络的组成部分具有多种状态或容量。在某些况下，除了考虑网络状态的好与坏，还要研究不同的性能指标。这种网络我们称为多态网络，这样的网络又称为随机流网络。

本文使用的所有图和网络，在不特殊说明的情况下都是二态的。

**定义2-6.（随机流网络的容量可靠性）**设G=（N，A，M）是从源点s到汇点t的一个随机流网络，其中N为节点集，A={ai | 1≤i≤n}为弧集，M={M1，M2，…，Mn}，Mi为弧ai的最大容量，i=1,2，…，n。那么G的d容量可靠性可以表示为，其中为满足d流的子图，即。

特别的，当d为最大流的时候，Pr为随机流网络的最大流容量可靠性。本文在不特殊说明的情况下，容量可靠性都指的是最大流容量可靠性。

### 2.2.2常见的随机流网络容量可靠性算法

传统的网络流理论虽然考虑了网络容量的问题，但都是针对网络容量固定的网络。而现实世界中的网络系统受到多种不确定因素（如网络构件的降级运行、网络阻塞等）的影响，可能会导致网络拓扑结构、链路容量发生变化，从而表现出网络容量的随机性和多态性。所以，使用传统的网络流理论没有考虑到网络容量的这两个特性，用来解决随机环境下的网络实际问题已经不再合适。

目前已有的计算流网络的可靠性的方法，大部分都需要已知最小路径或最小割集，再利用不交积和、容斥原理、复合路径等方法求解。

使用一定的规则生成网络的状态树，使得每一个分支都是全序集合。在生成状态树的同时搜索每一个分支，对状态采用基于割集的方法进行判断。每个分支上的最小的有效状态就是网络的小下界点。求得所有的小下界点，进而求出网络的可靠性。[一种计算随机流网络可靠-陛的新算法,王芳]

常用的状态搜索规则如下：

|  |
| --- |
| **算法2-3.** 空间划分搜索算法SDA |
| **输入：**流网络*G*,源点*s*, 汇点 *t*  1. init P = 0 , y← (0,0,…,0|E|)  2. push all the sub graphs states {(0,0,…,0|E|),(1,1,…,1|E|) , , I =0, j=0} into stack  3. while stack is not empty  4. get current closed space [*α*,*β*] of sub graphs according to j in the top state of stack  5. if j < I , then j++  6. else  7. pop stack  8. get the max flow value f on current sub graph   9. if f ≥ Fmax , then  10. calculate the possibility p(*α*)  11. if p(*α*) > P , then P = p(*α*) and y←*α*  12. else  13. get the max flow value f and its distribution fv on sub graph   14. if f ≥ Fmax , then  15. use fv to get pivot sets IC and push {*α*, *β*, *IC* ,| *IC* |,1} into stack  16. get *C0* = [*α’*,*β*] by {*α, β*, *IC* ,| *IC* |,1}  17. calculate the possibility p(*α’*)  18. if p(*α’*) > P, then P = p(*α’*) and y ←*α’*  **输出：***s*-*t*的最大流网络可靠新 |

那么基于状态划分规则的随机流网络容量可靠性算法如下：

|  |
| --- |
| **算法2-4.** 随机流网络容量可靠性算法 |
| **输入：**流网络*G*,源点*s*, 汇点 *t*  1. Fmax = Dinic(MSG(G), s, t) //运行Dinic算法得到最大流值Fmax  2. (subgraph set) ← SDA(G, Fmax, s, t)  3. Pr← pr of all subgraph that satisfy d flow  **输出：***s*-*t*的最大流网络可靠性Pr |

本文使用的网络可靠性一般指的是最大流对应的容量可靠性。使用的计算方法是基于状态划分的d-下界点计算方式。

## 2.3 不确定图数据

### 2.3.1 不确定图及可能世界模型

针对不确定图的研究大多数是建立在标准的概率图模型基础之上的，即给定一个图*G*（*V*, *E*），其中*V*是顶点集，*E*是边集，对任意的边*e* *∈* *E*，*e*都对应一个存在概率*pe*和不存在概率*qe*，且满足*pe* + *qe* = 1。如图2-3中的*G*是一个不确定图，边（*A*, *B*）上的数值0.9表示该边在*G*中存在的概率，对应的不存在概率为0.1。



图2-3 不确定图*G*

当前，关于不确定图的理论和应用研究包括许多方面。Petteri Hinstsane[31]就如何寻找不确定图中最可靠子图的问题展开了深入研究，国内哈工大的邹兆年、张硕等人分别针对不确定图频繁子图模式挖掘和不确定图数据库的查询处理等问题进行了深入的分析与研究，就如何构建挖掘模型、选择*k*极大频繁子模式、如何降低子图同构计算的复杂度以及图数据库索引设计等问题，提出了许多高效可行的方法[32-34]。而对于不确定图的相邻性与可达性也有很多的研究成果，张海杰 [35]等人考虑不确定图中的top-*k*近邻查询、Potanmias[36]等人关于*k*-NN查询问题，都分别提出了相应的距离函数和高效可行的查询算法；然而，东北大学袁野[37]等人则关注于不确定图中给定顶点的可达性，提出了一种有效的随机算法，此外，对于不确定图中的top-*k*生成树及相似性查询问题也有研究，许多有效的解决思路和算法[38] 被相继提出。

下面给出不确定图的相关概念：

**定义2-7. （不确定图）**一个不确定图G是一个五元组G=（V，E，s，t，（C，P）），其中，V是有向图G中顶点的集合，E是G中边的集合，s和t分别为G的源点和汇点，（C，P）是一个二元组且C：E->N是边上容量函数，P：E->（0，1]是边上的概率函数，表明该边能通过的最大容量为C时的概率为P，当边不存在，即边上能通过的容量为0时对应的概率为1-P。

**实例2-1.** 图2-3所示为一个不确定图G，该不确定图的源点为s，汇点为t，除此之外还包含其他的顶点v1和v2，边集为{E1，E2，E3，E4，E5}，以边E4为例，边上的容量c（E4）=1，概率p（E4）=0.8，也就是说E4边能够达到流量为1的概率为0.8，而容量是0的概率为1-p（E4）=0.2。

**定义2-8. （剩余不确定图**） 对于一个不确定图G，如果有一条边e’ 被移除，剩下的不确定图被称为原不确定图G移除边’之后的不确定图G’，则G’可以表示为G’=（V，E-e’，s ，t，（C，P）-e’（c，p）），其中剩余不确定图的顶点和原不确定一致，且同时删除被移除边上的容量和概率的对应关系。



图2-4 不确定图G在移除边E4后的剩余不确定图G’

图2-4为不确定图G移除边E4之后的剩余不确定图G‘，G’中的顶点与原不确定图一致，但是边以被移除边上的流量概率对应被移除。

可能世界模型[39-40]是不确定图问题研究中的基本模型，它将不确定图扩展成为可能世界实例的集合，而每个可能世界实例都对应着一个概率值，且每个可能世界的实例可以被看成是一个确定图，这样就可以把不确定图中的相关问题转换为确定图问题。

对于一个给定的不确定图*G*，其可能世界空间包含有2|*E*|个可能世界实例。对于图2-3中的不确定图*G*对应的可能世界空间大小为32 = 25，每个可能世界实例*g*的概率*P*（*g*） = Π*e∈E*（*g*） *pe* × Π*e∈E\E*（*g*） （1-*pe*），其中：*E*（*g*）表示*g*的边集，*E*\*E*（*g*） = {*e* | *e∈E∧e∉E*（*g*）}。

### 2.3.2 不确定图分布可靠性

不确定图流分布问题是在传统的流网络问题的基础上加入了边的存在概率，因此对于不确定图流分布问题而言，边上存在着两种属性，一种是容量属性，另一种是概率属性。

**定义2-9. （不确定图的子图）**不确定图G=（V，E，s，t，（C，P））的子图g（V’，E’，s，t ，C’）是一个确定图，其中V’ = V，E’∈E，C’是容量的集合，且C’满足∀e∈E’，C’（e）=C（e）。如果E’ = E，则称g为不确定图G的最大子图，记作MSG（G）。



图2-5 不确定图G的子图g1 图2-6 不确定图G的子图g2

如图2-5和图2-6所示，g1和g2是图2-3中不确定图G的两个子图，根据定义子图g2为不确定图G的最大子图，MSG（G）。

**定义2-10. （不确定图的子图概率）**不确定图G=（V，E，s，t，（C，P））的一个子图为g（V’，E’，s，t ，C’），则该子图g的概率为。即子图中边的存在概率和不确定图中其余边的不存在概率之积。

**实例2-2.** 根据上述的子图概率计算的公式可得，不确定图G的子图g1的概率为P（g1）=0.5\*0.4\*0.5\*0.8\*（1-0.8）=0.016，同样的，可以知道其最大子图g2的子图概率为P（g2）=0.5\*0.5\*0.4\*0.8\*0.8=0.064。

**定义2-11. （不确定图s-t流）**给定源点s和汇点t，不确定图G上的s-t流f是一个映射，它把每条边 对应到一个非负实数，f：E→R+；值f（e）直观上表示由边e携带的流量。一个流f必须满足下面两个性质：

（1）（容量条件）对每条边e∈E，有0≤f（e）≤c（e）。

（2）（守恒条件）除了s和t之外，对每个结点 ，有，即：所有进入结点 的流值之和等于所有从 出来的流值之和。

**定义2-12.（不确定图最大流）**不确定图G中，在每条边都存在时，能够从源点s到汇点t传输的最大流值fmax为不确定图G的最大流。

**实例2-3.** 图2-3所示的不确定图G能够达到的最大流为2。

**定义2-13.（不确定图的d容量可靠性）**不确定图的d容量可靠性P（d）可以表示为不确定图所有能够满足d流的子图概率之和[10]，即，特别的，当d为不确定图的最大流时，表示的为不确定图相对于最大流的容量可靠性。

**实例2-4.** 如图2-7为图2-3所示不确定图G的三个能够满足最大流2的所有子图，那么不确定图相对于最大流的容量可靠性可以表示为P=p（g1）+ p（g2）+ p（g3）=0.176。



图2-7 不确定图G的三个能够满足最大流2的所有子图g1，g2，g3

**定义2-14. （不确定图流分布及其可靠性）**给定一个不确定图G=（V，E，s，t，（C，P））及其上的一个流分布*F*，则*F*的可靠性是不确定图G所有包含该流分布*F*的子图概率之和，可以表示为 同时也等于不确定图G中所有流量不为0的边的概率之的边的概率之积[10]，即。特别的，当流分布达到的流量为不确定图最大流时，为不确定图的最大流分布，对应的为不确定图的最大流分布可靠性。

**实例2-5.** 如图2-8为不确定图G的2个最大流分布F1和F2，分析可知，最大流分布F1被子图g2和g3所蕴含，则P（F1）=P（g2）+p（g3）=0.16，此时也可以根据分布中流量不为0的概率之积来计算，此时的计算过程为P（F1）=0.5\*0.5\*0.8\*0.8=0.16，结果一直。同理，对于分布F2，通过定义可知P（F2）=0.08。

图2-8 不确定图G的两个最大流分布F1和F2

**定义2-15. （不确定图最可靠流分布）**一个不确定图最可靠流分布是一个不确定图流分布，其可靠性不小于该不确定图上的任一同等流分布。特别的，当流量为最大流时，为最可靠最大流分布，其对应的可靠性为最可靠最大流分布可靠性。

**实例2-6.** 如图2-8中的两个分布是不确定图G的两个最大流分布，其中P（F1）>P（F2），且不存在一个其他的最大流分布F，使得P（F）>P（F1），则F1是不确定图G的最可靠最大流分布，其对应的可靠性为不确定图G的最可靠性最大流分布可靠性。

### 2.3.3 不确定图容量可靠性

**定义2-16.（不确定图的容量可靠性）**设G=（N，A，M）是从源点s到汇点t的一个不确定图，其中N为节点集，A={ai | 1≤i≤n}为弧集，M={M1，M2，…，Mn}，Mi为弧ai的最大容量，i=1,2，…，n。那么G的d容量可靠性可以表示为，其中为满足d流的子图，即。

特别的，当d为最大流的时候，Pr为随机流网络的最大流容量可靠性。本文在不特殊说明的情况下，容量可靠性都指的是最大流容量可靠性。

**实例2-7.** 对于图2-3的不确定图G，其所有能够达到最大流的子图如图2-7中g1，g2，g3所示，那么G的最大流容量可靠性为Pd=pd（g1）+ pd（g2）+ pd（g3）。

## 2.4 本章小结

本章主要介绍了一些基本概念。由于本文是基于网络最大流、随机流网络可靠性和不确定图分布可靠性对不确定图关键边进行的相关研究，因此主要介绍这些概念的相关研究。首先介绍了最大流的相关概念和常用的算法，然后介绍了随机流网络可靠性的基本概念，最后介绍了不确定图的相关概念和不确定图数据的一般建模方法，以及不确定图流分布及其可靠性的相关问题。对相关概念给出了详细的实例，方便理解。

# 第3章 基于流量和容量可靠性的关键边衡量方法

本章介绍了基于流量和容量可靠性的不确定图边关键度衡量方法。本章的组织组织结构如下：3.1节构建了一种基于流量和容量可靠性的关键边衡量模型，3.2节设计了一种基于重复计算的基本算法BASE\_FCP，并分析了算法复杂度，3.3节提出一种基于流量和容量可靠性的不确定图关键边增量算法ICA\_FCP，并重点介绍了算法的思想与实现，同时对算法进行了正确性证明以及时间复杂度和空间复杂度的分析，3.4节提供了一种针对TopK 需求的搜索策略，3.5节过实验分析，比较BASE\_FCP算法和ICA\_FCP算法在空间复杂度和时间复杂度上的差别。

## 3.1 基于流量和容量可靠性的不确定图边关键度衡量模型

对于一个网络而言，能够达到的最大流的大小，体现出了网络的联通能力和传输能力，例如，网络的最大流受限于网络的最小割集，也就是说最小割集中的任意一条边断掉都会使得真个网络可以传输的最大流下降，这样的边就成了整个网络相对于流量的关键环节。在不确定图中，相对于可靠性的关键边对网络信息传输的可靠性起到关键作用，这样的环节一但发生故障，会使得整个网络的可靠性下降。

基于此，本文构建了基于流量和容量可靠性作为指标的不确定图关键边评估模型，针对不确定图中边移除（故障）后对流量和容量可靠性产生的相对损失，来评估关键边。该模型将流量作为衡量关键边的最重要因素，当流量一致时，比较最大流容量可靠性。假设f，pc分别表示不确定图边移除之后剩余不确定图能够达到的最大流和最大流容量可靠性，Kf，Kpc分别表示为流量和容量可靠性的关键系数，那么模型中的指标关键程度可以表示为Kf>Kpc。

根据这两个指标，可以将不确定图的边分为三类，即边发生故障之后剩余不确定图和原不确定图进行对比：

（1）流量和容量可靠性都减少，作为A类边；

（2）流量不变，容量可靠性减少，作为B类边；

（3）流量和容量可靠性都不边，作为C类边。

根据这两个指标的关键程度Kf>Kpc，易得出ABC三类边的关键程度依次减少。

在不确定图求解最可靠最大流分布的过程中，通过状态划分规则可以求解不确定图满足最大流的所有子图区间[28]，根据边能够被满足最大流的子图的个数可对边进行分类。根据不确定图子图的向量表示法[22]以及向量的偏序关系，本文在获取的子图区间的基础上定义边在所有满足最大流子图的存在率Si。

首先定义sij为边ei在子图区间j的存在状况，根据偏序规则，其中ei=1为在子图区间j中的所有子图都必须包含边ei；ei=0表示在子图区间j中的所有子图都不包含边ei；如果ei=x则表示子图区间j中的子图有包含和不包含两种情况，定义sij如公式3-1所示：

（3-1）

其中i表示边的序号，j表示子图区间的序号。

对于不确定图中的每一条边ei都有边存在率Si，定义如公式3-2所示：

（3-2）

其中i为边的序号，i∈{1….n}，n为不确定图G边的个数，j为子图区间的序号，j∈{1…*τ*}，*τ* 为使用状态划分后闭合区间的个数。

根据定义可知，Si有性质，其中*τ*为使用状态划分区间算法获得的最大流区间的个数。

根据边存在率Si的定义可知，如果Si=*τ*，那边边ei存在于所有的满足最大流子图中，因此当ei发生故障时，所有满足最大流的子图都遭到破坏而不能满足最大流，从而相应的最大流容量可靠性也发生改变，因此当对于边ei，如果Si=*τ*，那么ei为A类边；同理如果0＜Si＜*τ*，表示边ei存在于部分，满足最大流的子图中，那么当边ei故障之后，依然有剩余的一些满足最大流的子图可以不被破坏，这时依然可以达到最大流，但是最大流容量可靠性降低，此时边ei为B类边。

**实例3-1.** 对于图2-3中的不确定图G，其经过状态划分算法获取所有满足最大流的区间表示如表3-1所示，根据公式3-1，可知对于边e1，有s11=1，s12=1，根据公式3-2可知，对于边e1的边存在率S1=s11+s12=2=*τ*，所以对于不确定图G而言，边e1是一个A类边，同理可以知道边e1，e2和e5为A类边，e3和e4为B类边。

表3-1 不确定图G经过划分后的所有满足最大流的区间

|  |  |
| --- | --- |
| 编号 | 满足最大流的区间 |
| 1 | 11x11 |
| 2 | 11101 |

综上所述，根据边存在率Si，对边进行定性的分类如下：

1、Si=*τ* ，那么ei为A类边；

2、0＜Si＜*τ*，那么ei为B类边；

3、Si=0，那么ei为C类边；

根据以上的定性分析，可以得到如下的一系列的推论：

**推论3-1.** 对于不确定图G中的一条边ei，如果边ei是A类边，那么移除边ei会使得不确定图不能满足原有最大流和容量可靠性。

**证明3-1**. 想要证明移除A类边，会使得不确定图G最大流减少，需要证明两点（1）A类边移除，不确定图G不能达到最大流；（2）非A类边移除，不确定图G任然可以达到最大流。下面将按照这两条分别证明；

（1）因为ei是A类边，有Si=*τ*，根据定义边ei存在于所有的满足最大流的不确定子图+中，假设经过划分算保存下来的所有闭合子图空间集为{*C*1，*C*2，…，*Cτ* }，满足*Ci* ∩ *Cj =* ∅，*C*1 ∪ *C*2 ∪ … ∪ *Cτ* = *A*，其中τ 为满足最大流的子图区间的个数，因为Si=*τ* ，所以满足最大流的子图g一定有g∈A且有ei∈g，如果当ei发生故障之后，剩余不确定图的任意子图都不包含ei，即，则必有，因此对于ei发生故障后的剩余不确定图的任意子图都不能达到最大流，那么也不能达到最大流；

（2）同理，对于一条非A类边，因为Si＜*τ*，则在A中必存在一个子图g，使得，如果边ei发生故障，不确定图可以使用子图g传递流量，同时因为g∈A，即g能传递最大流，所以对于非A类边发生故障，不确定图G任然能够达到最大流。

的证。

**实例3-2.** 对于图2-3中的不确定图G，经过状态划分过程，获取的所有满足最大流区间如表3-1所示，因为经过划分有两个区间，则*τ*=2，根据公式3-1，可知对于边e1，有s11=1，s12=1，根据公式3-2可知，对于边e1的边存在率S1=s11+s12=2=*τ*，所以对于不确定图G而言，边e1是一个A类边，A发生故障使得不确定图的最大流从原来的2下降为1。

**推论3-2.** 对于不确定图G中的一条边ei，如果边ei是B类边，那么移除ei，不确定图G依然可以达到最大流。

**证明3-2.** 对于不确定图G的任意一条边ei，如果边ei是B类边，即如果有0<Si＜*τ* ，则ei移除之后，可以找不包含边ei的子图达到最大流，使得最大流不会下降，但是ei的断掉使得一部分能够达到最大流的子图遭到破坏，使得随机流网络的可靠性可能下降。

经过划分算保存下来的所有闭合子图空间集为{*C*1，*C*2，…，*Cτ* }，满足*Ci* ∩ *Cj =* ∅，*C*1 ∪ *C*2 ∪ … ∪ *Cτ* = *A*，其中τ 为满足最大流的子图区间的个数，因为0<Si＜*τ* ，则对于任意的ei∈g子图，可以将子图的集合A分为g∈A1和，其中满足，A1∪A2=A。

在ei断掉之前，随机流网络的可靠性可以为A中的所有子图概率之和，可以表示为P（A），而当ei断掉之后，集合A1中的子图不一定能够满足最大流，但是集合A2中的流量可以满可以满足最大流，因此保证了不确定图的连通性不变；同时ei断掉使得A1中至少有一个子图不能满足最大流，因此使得随机流网络的可靠性下降，此时即证明ei为不确定图的可靠性关键边。

的证。

**实例3-3.** 对于图2-3所示的不确定图G，经过状态划分过程，获取的满足最大流的区间分别是11x11和11101，其中*τ*=2，根据公式3-1，可知对于边e3，有s31=0，s32=1，根据公式3-2可知，对于边e3的边存在率S1=s11+s12=1<*τ*，所以对于不确定图而言e3是一个B类边，则e3移除，使得子图11111和11101不能满足最大流，但是11011依然满足最大流，此时的最大流可靠性下降，仅为11011子图的可靠性。

**推论3-3.** 对于不确定图中的一条边ei，如果ei是C类边，那么移除边ei，不确定图的最大流，不确定图的容量可靠性都不发生改变。

**证明3-3.** 若要证明对于不确定图G中的任意一条边ei，如果ei是C类边，即如果Si=0，则ei断掉之后，不确定图的流量和容量可靠性都不发生改变，只需要证明（1）C边断掉之后，不确定图G的最大流不变（2）C边断掉之后，不确定图G的容量可靠性不变。下面将对这两点分别证明：

（1）对于不确定图G，经过划分算保存下来的所有闭合子图空间集为{*C*1，*C*2，…，*Cτ* }，满足*Ci* ∩ *Cj =* ∅，对于不确定图G中的C边ei有Si=0，即当ei发生故障之后，至少存在一个子图g∈*C*1 ∪ *C*2 ∪ … ∪ *Cτ*且，所以当ei移除之后，依然有g满足最大流，即ei边断掉之后，不确定图G的最大流不变；

（2）首先说明当Si=0，那么对于所有的子图区间每一个对应的边ei处都为标记“x”，即任意一个子图区间可以表示为C=（c1，c2…x…cn），其中ci表示为子图区间C的各条边对应的表示，i={1…i-1，i，i+1…n}，则C可以分为两个子图区间（或子图）C1=（c1，c2…0…cn）和C2=（c1，c2…1…cn），那么在移除边之前，区间C的可靠性可以表示为P=p（c1）\*….\* p（ei-1）\*p（ei=0）\*p（ei+1）…\*p（cn）+p（c1）\*….\* p（ei-1）\*p（ei=1）\*p（ei+1）…\*p（cn）= p（c1）\*….\* p（ei-1）\*p（ei+1）…\*p（cn）\*（p（ei=0）+ p（ei=1）），又因为p（ei=0）+ p（ei=1）=1，所以C边断掉之后，不确定图G的容量可靠性不变。

的证。

**实例3-4**：假设对于一个不确定图，其经过划分获取的子图区间只有一个，11x11，那么根据公式3-1和公式3-2，边e3为C类边，因此e3移除之后，依然可以满足最大流。移除e3之前，容量可靠性pc=p（e1）p（e2）（1-p（e3））p（e4）p（e5）+ p（e1）p（e2）（e3）p（e4）p（e5）= p（e1）p（e2）p（e4）p（e5），移除e3之后，容量可靠性表示为pc’= p（e1）p（e2）p（e4）p（e5），显然移除e3之后不确定图的容量可靠性不变。

## 3.2 基于重复计算的BASE\_FCP算法

基于重复计算的基础算法（BASELINE algorithm with Flow and Content Probability）的思想是，首先计算不确定图每一条边移除之后能够满足的最大流，然后通过比较不确定图在移除边之后能够达到的最大流，只有当流量相同的情况下，才计算比较容量可靠性，本章将BASE\_FCP算法作为基于流量和容量可靠性模型的关键边基本算法。

根据上述分析，下面给出重复计算的基础代码实现（伪代码）：

|  |
| --- |
| **算法3-1.** 基于重复计算的BASE\_FCP算法 |
| **输入**：不确定图G，源点source，汇点sink  1. Input uncertain Graph G and its source point and sink point；  2. Fmax 🡨 Dinic（G，source，sink）；  3. Get G maxflow when every single edge being cut off ALLEdgeFlowDecrease[|E|]；  4. Sort keyedge by ALLEdgeFlowDecrease[|E|]；  5. When the flow G-ei equles the flow G-ej then；  6. reCompute the maxFlow of G-ei，pd using SDBA[10]；  7. Soet KeyEdge by pc；  8. Return sorted key\_edge\_set；  **输出**：根据关键程度排序好的keyEdgeSet，以及移除各边之后不确定图对应的最大流和容量可靠性 |

基于重复计算的BASE\_FCP算法首先要使用getALLEdgeFlowDecrease函数获取所有边发生故障之后能够满足的最大流，然后使用soetKeyEdge\_by\_flow函数对计算的流量进行排序，当前后两条边断掉之后获得的流量不相等时，根据模型定义，直接就可以比较出两条边的关键程度，而只有当两条边断掉获得流量相等的情况下，无法比较的情形，才会重复计算不确定图的容量可靠性，使用容量可靠性来区分这两条边的关键程度。

基于重复计算的基础算法的运行效率与重复计算的次数紧密相连，在算法的首先需要计算所有边发生故障之后的最大流，本文使用的计算最大流的算法是Dinic算法，运行Dinic算法所需的处理时间为O（|V|2|E|），因此在处理这部分是复杂度是O（|V|2|E|2）；然后当移除边之后流量相同的情况下，需要重复计算，本文使用的重复计算方式为基于状态划分的算法[10]，其复杂度为O（k|V|2|E|），其中K为划分过程中需要运行Dinic算法的次数。因此整个BASE算法的复杂度为O（km|V|2|E|） + O（|V|2|E|2），其中m为需要重复计算的次数，一般m的范围是0≤m≤|E|。

## 3.3 基于流量和容量可靠性的不确定图关键边增量算法ICA\_FCP

### 3.3.1 算法思想

虽然BASE\_FCP算法相对于完全重复计算的方式减少了不少的重复计算，但是如果图规模较大的情况下，任然需要多次的重复计算，这显然对于大规模图或者高稠密图是不适用的。本节提出了一种基于状态划分树的增量算法ICA\_FCP（Increment Cost Algorithm with Flow and Content Probability），该算法的主要思想是根据模型中不同类型的边设置不同的增量算法计算，以减少计算复杂度。当A边移除，使用STPA\_FCP（State Tree Pruning Algorithm with Flow and Content Probability）算法，在区间状态树上利用割集中的边剪枝获取移除之后依然满足最大流的子图集合；当B边移除，使用BSA\_FCP（B Class Edge Search Algorithm with Flow and Content Probability）算法获取依然满足最大流的区间；因为C边移除之后不影响，所以不需要计算。

基于以上步骤，ICA\_FCP算法可以表示为如图3-1步骤：



图3-1 算法ICA\_FCP流程图

### 3.3.2 基于割集的状态树剪枝A边算法STPA\_FCP

对于不确定图G而言，求解最可靠最大流分布的方式有很多，目前普遍使用的是算法SDBA[28]，SDBA算法中使用到了状态划分算法，在状态划分的过程中，可以生成一个区间状态树，这样的一个区间状态树生成如算法3-2所示：

|  |
| --- |
| **算法3-2.** SDBA算法生成区间状态树算法 |
| **输入**：不确定图G，源点source，汇点sink  1. stack S 🡨 C[L, U], C[L, U] is the root node of tree T;  2. while S is not empty then  3. C 🡨 pop S;  4. check the lower bound and upper bound of C with DINIC;  5. if lower(C) < max\_flow and upper(C) = max\_flow then  6. resolve C as the child node to C and push unknow child into S;  7. if S is not empty goto step 2;  8. if S is empty return;  **输出**：构造树T |

如图3-2所示，为图2-3中不确定图G的状态划分过程，其中斜体下划线的状态为未知状态，需要继续划分，加粗字体的子图区间为满足最大流的状态区间，而其他的区间为



图3-2 完全状态划分树State-Tree

对于A类边，由推论3-1可知，其断掉会使流量减少，即不能达到原有的最大流，此时原有最大流的网络可靠性和流分布的可靠性都为下降为0，理论上新的最大流相关的计算需要重新计算。但是可以根据该边对原有的完全状态划分树State-Tree进行处理，将原有的A边位置为1状态的去除，那么留下来的状态划分树的区间是移除A边之后的所有子图，在搜索子图区间树的过程中，使用割集中的边，对State-Tree剪枝，可以有效的减少待搜索区间个数。

**实例3-5.** 图7为不确定图G的完全状态划分树，对于不确定图G的A边e5，移除之后按照以上的处理方式，获取的状态树如下图3-3所示，所有的叶子节点为移除e5之后的子图空间的划分。



图3-3 移除e5之后的状态区间树

A类边根据搜索树的规则，对于每一个区间改边位置都设置为0，目的是获取不想交的的区间。对于新构成的一个子图树，本节提出一种基于割集的状态树剪枝算法STPA\_FCP（State Tree Pruning Algorithm with Flow and Content Probability），该算法的主要是利用了割集中的边必定在最大流子图这一性质，通过割集中的边对于子图状态树进行剪枝，以达到减少搜索的目的。因为割集中的边一般在子图树的第二层就能达到剪枝的效果，所有该方法能够有效的减少搜索效果。

该算法首先获取根节点，对于其所有子节点，通过割集中的边进行剪枝，因为满足最大流的子图必定包含割集中的边，所以不包含割集中边的中间节点和叶子节点都必须舍去，对于依然保存的子图区间，进行进一步划分即可，即（1）当区间的下界子图满足最大流，则整个区间满足最大流，保留区间。（2）当区间上界不满足最大流，则整个区间不满足最大流，舍弃区间。（3）区间下界不满足最大流，上界满足最大流，则需要二次划分区间。通过这样的方式便可以有效的获取移除边e之后所有满足最大流的区间子图，进而计算移除e之后不确定图的容量可靠性。

**实例3-6.** 如图3-4是经过割集剪枝之后的子图树。其中有背景的圆部分及其子树都被剪枝掉。

图3-4 经过剪枝后的子图树

综上所述，STPA\_FCP算法的伪代码可以表示如下：

|  |
| --- |
| **算法3-3.** 基于割集的状态树剪枝算法STPA\_FCP |
| **输入**：割集边CutSet，子图区间树root，不确定图G，源点source，汇点sink，移除边e  1. get CutSet，root，init stack S；  2. push root in S；  3. if S is not empty，then pop S as C；  4. if the cut edge’s position is 0 then move out C；  5. Else if C’s lower bound graph satisfy Fmax then keep C  6. If C’s upper bound graph not satisfy Fmax then move out C；  7. If C’s lower bound graph not satisfy Fmax，but upper bound graph satisfy Fmax then divide C  8. Return all graph that satisfy Fmax after G move edge e  **输出**：不确定图G在移除边e之后所有能够满足最大流的子图集合set |

### 3.3.3 B类边搜索算法BSA\_FCP

对于B类边，同样由推论3-2可以知道，断掉B类边，流量不会减少，此时的重点在于计算剩余的不确定图的容量可靠性。为了计算移除B边之后的容量可靠性，主要是计算移除B边，原有只满足最大流的状态划分树MaxFlow-State-Tree中还剩哪些满足最大流的子图，换言之，就是要移除不能满足最大流子图，且不能存在重复的子图。本章提出算法BSA\_FCP（B Class Edge Search Algorithm with Flow and Content Probability），用于计算获取B边移除之后，依然能够满足最大流的所有子图。该算法步骤如下：

|  |
| --- |
| **算法3-4.** B类边搜索算法BSA\_FCP |
| **输入**：满足最大流的状态划分树MaxFlow-State-Tree，移除边e  1. For all leaf node in MaxFlow-State-Tree as C  2. If the position of e in C is 1，then move out C；  3. If the position of e position in C is 0，then push C in set；  4. If the position of e in C is X，then divide C to Ce=1 and Ce=0 and goto step （2）（3）；  5. Until MaxFlow-State-Tree is all be searched，then return set.  **输出**：移除e之后，不确定图依然满足最大流的所有子图集合set |

**证明3-4**. 对于不确定图G，经过划分后获取的满足最大流的所有区间为{*C*1，*C*2，…，*Cτ* }，满足*Ci* ∩ *Cj =* ∅，*C*1 ∪ *C*2 ∪ … ∪ *Cτ* = *A*，其中*τ* 为满足最大流的子图区间个数，假设边ei为G的B类边，根据B类边定义，存在一个k∈{1，2…*τ* }，使得 与 ，令，因为中的所有子图位置为0，所以有

若要证明 为G-之后，所有满足最大流的子图集合，需要证明以下两点：

（1），都有 ；

（2），若 ，则必存在一个，使得。

首先证明（1）：已知条件都有，且，所以可以推出对于任意的子图g，如果，都有g满足最大流，即，都有。

然后证明（2）：反正法。假设，若有，则对于，都有。

对于，且有，所以，又因为，所以，这与不存在一个使得矛盾。假设不成立。

的证。

**实例3-7.** 对于如图2-3中的不确定图G，经过状态划分获取满足最大流的子图区间有两个，分别是11x11和11101，e3是一个B边，根据以上规则，11x11区间在e3位置上的状态是x，可以分为11111和11011两个区间，因为11111和11101两个区间在e3位置上的状态是1，所以舍去，而只保留区间11011。所以最终移除e3之后，依然满足最大流的子图区间是11011。

对于C类边的计算，根据推论3-3，其断掉不会影响不确定图的最大流和容量可靠性，也就是说C类边断掉不会对不确定图的状态照成任何的影响，因此不需要计算。

### 3.3.4 算法复杂度分析

根据推论3-3，C类边移除之后不需要计算，直接使用原有不确定图的最大流和容量可靠性即可。

基于以上处理方式，ICA\_FCP算法的伪代码如下：

|  |
| --- |
| **算法3-5.** 基于流量和容量可靠性的不确定图关键边增量算法ICA\_FCP |
| **输入**：不确定图G，源点source，汇点sink  1. Input uncertain Graph G and its source point and sink point；  2. Get maxFlow ，StateMtrix and build StateTree using algorithm SDBA  3. Get G maxflow when every single edge being cut off ALLEdgeFlowDecrease[|E|]；  4. Sort keyedge by ALLEdgeFlowDecrease[|E|]；  5. Classify ABC edge by StateMtrix；  6. When the flow （G-ei） equles the flow （G-ej） then  7. If ei is A class edge then use algorithm STPA\_FCP to get pc；  8. If ei is B class edge then use algorithm BSA\_FCP to get pc；  9. If ei is C class edge then pc is same as before ei be moved。  10. Soet KeyEdge by pc；  11. Return sorted key\_edge\_set；  **输出**：根据关键程度排序好的keyEdgeSet |

基于状态划分树的增量算法ICA\_FCP首先需要使用getALLEdgeFlowDecrease函数获取所有边发生故障之后能够满足的最大流，然后使用soetKeyEdge\_by\_flow函数对计算的流量进行排序，当前后两条边断掉之后获得的流量不相等时，根据模型定义，直接就可以比较出两条边的关键程度，只有当两条边移除之后最大流一致的情况下，会使用状态划分树的增量算法计算，该部分的计算最大流的算法是Dinic算法，运行Dinic算法所需的处理时间为O（|V|2|E|），因此在处理这部分是复杂度是O（|V|2|E|2）。然后将不确定图的边分为ABC三类边，然后根据三类边的不同性质对于不同性质的边选择不同的算法，从而简化计算过程。

首先对于A类边，考虑最坏的情况，就是重复完全计算，该算法运算复杂度主要体现在K次运行Dinic算法，因此对于A边的复杂度为O（ak|V|2|E|），其中a为不确定图G中a边的个数，对于B类边，需要遍历满足最大流的所有区间，其复杂度为O（b*τ*），其中b为不确定图B边的个数，*τ* 为满足最大流区间的个数，对于C边来说，因为不需要计算，只需要保留原有的最大流和容量可靠性，所以复杂度为O（c），其中c为不确定图C类边的个数。综上所述，基于状态划分树的增量算法的整体复杂度是O（|V|2|E|2）+O（ak|V|2|E|）+ O（b*τ*）+O（c），其中a+b+c=|E|。由公式可见，ICA\_FCP在最坏的情况下，复杂度和BASE\_FCP算法的复杂度一致。

## 3.4 Top-K关键边搜索策略

该问题待定

TopK搜索出来的K个关键边是否一致。（不同的图规模，不同稠密度）

时间空间复杂度的比较

## 3.5 实验及分析

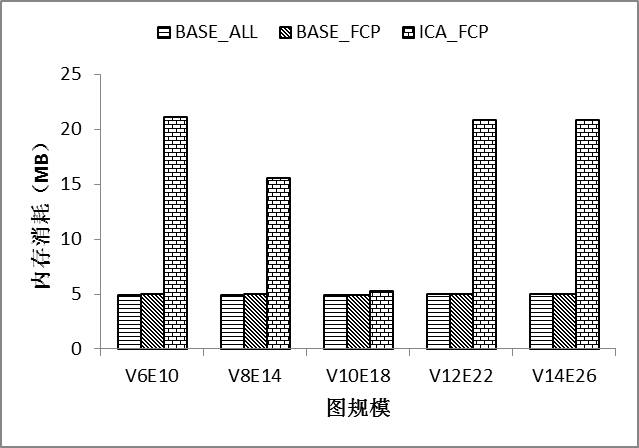
为了验证本文所提出算法的运行效率及分析影响算法性能各种因素，本文进行了一系列的实验，实验平台为一台Intel Core的PC机（CPU i7-3770，3.40GHz，内存8GB，64位windows 7操作系统），算法采用C++在VS2010上实现。

### 3.5.1 实验数据

本文采用了与文献[10]相同的数据集，使用NETGEN生成器[11]产生V6E10、V8E14、V10E18、V12E22、V14E26共5组不同图规模的二态有向图集合（本文实验数据集合大小为5），其中VnEm表示有n个顶点、m条边组成的图，图中边的容量与对应概率满足均匀分布。通过在不同图规模的情况下，比较算法BASE\_FCP和ICA\_FCP算法在运行时间及内存消耗放方面的差异。实验中的BASE\_ALL算法指代的是完全重复计算的方式。

### 3.5.2 实验结果与分析

**实验3-1.** 不同图规模对于算法性能的影响

（a）时间消耗 （b）内存消耗

图3-5 不同图规模情况下BASE\_FCP和ICA\_FCP算法时间和内存消耗

根据图3-5所示，算法BASE\_ALL算法是完全重复计算的方式，BASE\_FCP算法是在当流量一致的情况下才计算分布可靠性和容量可靠性。如图3-5（a）所示，BASE\_FCP算法相对于BASE\_ALL算法在运行时间上有一定的减少，但是依然不适应大规模图，而ICA\_FCP算法在运行时间上有了很大的减少，适应性更好。如图3-5（b）所示，ICA\_FCP算法相对于BASE\_FCP算法在空间复杂度上有了一定程度的增加，但是随着图规模的增加，内存的使用并没有增加太多，依然在可以接受的范围内。

**实验3-2.** 不同图稠密度对于算法性能的影响

为更好的反应算法BASE\_FCP算法和ICA\_FCP算法的性能差异，使用NETGEN生成器生成V15E21，V15E32，V15E42，V15E53四种不同稠密度的图，通过在不同图稠密度的情况下，比较算法BASELINE和ICA算法在运行时间及内存消耗放方面的差异。实验中的BASE\_ALL算法指代的是完全重复计算的方式。试验结果如下：

（a）时间消耗 （b）内存消耗

图3-6 图稠密度对BASE\_FCP和ICA\_FCP时间和内存消耗影响

如图3-6所示，相对于不同稠密度的图，ICA\_FCP算法相对于BASE\_FCP算法在空间复杂度上有一定的增加，但是时间复杂度上有较大的优势。

在实验的过程中，发现对于不确定图使用状态划分来求解所有满足最大流的区间子图，来求解不确定图的最大流容量可靠性，这些子图区间划分的个数与大小与使用DINIC算法获取的“分割线”有直接的联系，对于一块连续的满足最大流的区间，如果“划分线”获取的不是该区间的下界，那么最后获取的子图区间是有可能合并的。这样使得原本是C类边被归结为B类边，但是这不影响算法的正确性，因为C类边不需要计算，这样的结果是的算法的复杂度变高一点，但是由于BSA\_FCP算法复杂度很低，所有依然是可以接受的。

## 3.6 本章小结

针对不确定图关键边的衡量，本章节构建了一种基于流量和容量可靠性的关键边衡量模型，使用不确定图中边移除（故障）后对流量和容量（最大流）可靠性产生的相对损失这一角度，对边的关键度进行综合评估。对于构建的基于流量和容量可靠性的模型，本章节首先设计了一种基于重复计算的基本算法BASE\_FCP。因为BASE\_FCP算法复杂度较高，本章提出一种基于流量和容量可靠性的不确定图关键边增量算法ICA\_FCP，并重点介绍了算法的思想与实现，同时对算法进行了正确性证明以及时间复杂度和空间复杂度的分析。紧接着，针对实际情况的需要，在很多时候，并不需要把所有不确定图的边的关键程度都衡量出来，因此，本章同时还提供了一种针对TopK需求的搜索策略。最后通过实验分析，比较BASE\_FCP算法和ICA\_FCP算法在空间复杂度和时间复杂度上的差别，实验结果表明ICA\_FCP算法相对于BASE\_FCP算法其空间复杂度有一定的增加，但是时间复杂度方面具有较大的优势。

# 第4章 基于流量和分布可靠性的关键边衡量方法

在上一章中，基于流量和容量可靠性的不确定图关键评估方法使用的是移除边之后，相对流量和容量可靠性的减少量，然而根据容量可靠性的定义，计算最大流容量可靠性需要获取不确定图所有满足最大流的子图，也就是说计算容量可靠性本身是一个非常复杂的过程。为了减少衡量不确定图关键边的复杂度，本章提出了基于流量和分布可靠性的不确定图边关键度衡量方法，与上一节不同的是，该方法使用流量和分布可靠性（最可靠最大流分布可靠性）作为衡量关键边的指标，针对边移除（故障）后对于不确定图最大流和分布可靠性造成的相对损失这一角度来评估。

本章的组织组织结构如下：4.1节构建了一种基于流量和分布可靠性的关键边衡量模型；4.2节设计了一种基于重复计算的基本算法BASE\_FDP，并分析了算法复杂度；4.3节提出一种基于流量和分布可靠性的不确定图关键边增量算法ICA\_FDP，并重点介绍了算法的思想与实现，同时对算法进行了正确性证明以及时间复杂度和空间复杂度的分析；4.4节提出了一种对于A边的基于增广路径添加的近似算法，并分析了算法复杂度，给出了近似算法和精确算法的相似度度量，并分析了近似算法的下界；4.5节提供了一种针对TopK需求的搜索策略；4.6节通过试验分析，比较BASE\_FDP算法和ICA\_FDP算法在空间复杂度和时间复杂度上的差别，比较ICA\_FDP相较于ICA\_FCP算法的优势，并分析了近似算法的波动性。

## 4.1 基于流量和分布可靠性的不确定图边关键度衡量模型

本课题第三章研究的是基于流量和容量可靠性的不确定图关键评估方法，然而计算不确定图的容量可靠性复杂度很高，该算法只能适应较小规模的图。因此本章提出一种基于流量和分布可靠性（最可靠最大流分布可靠性）的不确定图关键边衡量模型。该模型针对不确定图中边移除（故障）后对流量和分布可靠性产生的相对损失，来评估关键边。该模型将流量作为衡量关键边的最重要因素，当流量一致时，比较最可靠最大流分布的可靠性。假设f，pd分别表示不确定图边移除之后剩余不确定图能够达到的最大流和最大流容量可靠性，Kf，Kpd分别表示为流量和容量可靠性的关键系数，那么模型中的指标关键程度可以表示为Kf>Kpd。

因为计算不确定图的（最大流）容量可靠性，需要将不确定图所有满足最大流的子图都获取，然后计算所有子图的概率之和，而相对于此，计算不确定图的分布可靠性就简单很多，同样也需要状态划分算法，但是根据极小子图理论[28]，只需要获取极小子图即可。但是不确定图两条边断掉之后，分布可靠性是一样的时候，容量可靠性有可能不同，因此使用不确定使用流量和分布可靠性作为指标的关键边衡量模型的区分度没有第三章的模型好，但是算法复杂度降低了许多。

同样，根据流量和（最可靠最大流）分布可靠性，可以将不确定图的边分为三类，即边发生故障之后剩余不确定图和原不确定图进行对比：

（1）流量和分布可靠性都减少，作为A类边；

（2）流量不变，分布可靠性减少，作为B类边；

（3）流量和分布可靠性都不边，作为C类边。

根据这两个指标的关键程度Kf>Kpd，易得出ABC三类边的关键程度依次减少。

在不确定图求解最可靠最大流分布的过程中，通过状态划分规则可以求解不确定图满足最大流的所有子图区间，根据边能够被满足最大流的子图的个数可对边进行分类。根据不确定图子图的向量表示法以及向量的偏序关系，本文在获取的子图区间的基础上定义边在所有满足最大流子图的存在率Si。

首先定义sij为边ei在子图区间j的存在状况，根据偏序规则，其中ei=1为在子图区间j中的所有子图都必须包含边ei；ei=0表示在子图区间j中的所有子图都不包含边ei；如果ei=x则表示子图区间j中的子图有包含和不包含两种情况，定义sij如下：

（4-1）

其中i表示边的序号，j表示子图区间的序号。

对于不确定图中的每一条边ei都有边存在率Si，定义如下

（4-2）

其中i为边的序号，i∈{1….n}，n为不确定图G边的个数，j为子图区间的序号，j∈{1…*τ*}，*τ* 为使用状态划分后闭合区间的个数。

根据定义可知，Si有性质，其中*τ*为使用状态划分区间算法获得的最大流区间的个数。

根据边存在率Si的定义可知，如果Si=*τ*，那边边ei存在于所有的满足最大流子图中，因此当ei发生故障时，所有满足最大流的子图都遭到破坏而不能满足最大流，从而相应的最大流容量可靠性也发生改变，因此当对于边ei，如果Si=*τ*，那么ei为A类边；如果0＜Si＜*τ*，且边ei存在于G的极小子图g中，那么边ei发生故障之后，依然可以使用其他不包含该边的子图传递最大流，但是因为极小子图被破坏，所以最大流分布的可靠性下降，这样的边为B类边；同理，如果0≤Si＜*τ*，且边ei不存在于G的极小子图g中，那么边ei发生故障之后，依然可以使用其他不包含该边的子图传递最大流，同时极小子图没有被破坏，所以最大流分布的可靠性不变，这样的边为C类边，不需要重新计算。

**实例4-1.** 对于图2-3中的不确定图G，其经过状态划分算法获取所有满足最大流的区间表示如表3-1所示，其中极小子图为11011，根据公式4-1，可知对于边e1，有s11=1，s12=1，根据公式4-2可知，对于边e1的边存在率S1=s11+s12=2=*τ*，所以对于不确定图G而言，边e1是一个A类边，同理可以知道边e1，e2和e5为A类边；但是对于e4而言，S4=1<2=*τ*，且e4在极小子图上，所以e4为B类边，移除e4之后，依然可以使用子图11101满足最大流，但是因为极小子图11011被破坏，所以最可靠最大流分布可靠性下降；对于e3而言，同样有S3=1<2=*τ*，但是e3不在极小子图11011上，因此e3移除之后，极小子图没有被破坏，所以最可靠最大流分布可靠性也不变。

综上所述，根据边存在率Si，对边进行定性的分类如下：

1、Si=*τ* ，那么ei为A类边；

2、0＜Si＜*τ*且ei存在于极小子图，那么ei为B类边；

3、0≤Si＜*τ*且ei不存在于极小子图，那么ei为C类边；

根据以上的定性分析，可以得到如下的一系列的性质。

**推论4-1.** 对于不确定图中的一条边ei，如果边ei是B类边，那么移除ei，不确定图依然可以达到最大流，但分布可靠性下降。

**证明4-1.** 对于不确定图G的任意一条边ei，如果边ei是B类边，即如果有0<Si＜*τ*且ei存在于极小子图，下面将分两步证明（1）移除B边，不确定图最大流不变（2）移除B边。最可靠最大流分布可靠性下降。

（1）假设经过划分算保存下来的所有闭合子图空间集为{*C*1，*C*2，…，*Cτ* }，满足*Ci* ∩ *Cj =* ∅，*C*1 ∪ *C*2 ∪ … ∪ *Cτ* = SET，其中τ 为满足最大流的子图区间的个数，因为0<Si＜*τ* ，因此必存在一个子图，且g的ei位置为0，即移除ei不会对子图g造成影响，又因为g能满足最大流，所以移除ei不会对不确定图的最大流造成影响。

（2）因为ei存在于极小子图，所以移除ei会使得不确定图的极小子图被破坏，那么不能满足原有的最可靠最大流分布可靠性。

的证。

**推论4-2.** 对于不确定图中的一条边ei，如果ei是C类边，那么移除边ei，不确定图的最大流，分布可靠性不发生改变。

**证明4-2.** 对于不确定图G中的任意一条边ei，如果ei是C类边，即如果有0<Si＜*τ*且ei不存在于极小子图，下面分两步证明（1）移除C边，不确定图最大流不变（2）移除C边，最可靠最大流分布可靠性不变。

（1）同证明4-1（1）。

（2）因为ei不存在于极小子图，所以移除ei不会使得不确定图的极小子图被破坏，那么依然能够满足原有的最可靠最大流分布可靠性。

的证。

## 4.2 基于重复计算的BASE\_FDP算法

基于重复计算的基础算法（BASELINE algorithm with Flow and Distribution Probability）的思想是，首先计算不确定图每一条边移除之后能够满足的最大流，然后通过比较不确定图在移除边之后能够达到的最大流，只有当流量相同的情况下，才计算比较最可靠最大流分布及其可靠性，本文将BASE\_FDP算法作为改模型下衡量关键边的基本算法。

根据上述分析，下面给出重复计算的基础代码实现（伪代码）：

|  |
| --- |
| **算法4-1.** 基于重复计算的BASE\_FDP算法 |
| **输入**：不确定图G，源点source，汇点sink  1. Input uncertain Graph G and its source point and sink point；  2. Fmax 🡨 Dinic（G，source，sink）；  3. Get G maxflow when every single edge being cut off ALLEdgeFlowDecrease[|E|]；  4. Sort keyedge by ALLEdgeFlowDecrease[|E|]；  5. When the flow G-ei equles the flow G-ej then；  6. reCompute the maxFlow of G-ei，pd using SDBA[10]；  7. Soet KeyEdge by pd；  8. Return sorted key\_edge\_set；  **输出**：根据关键程度排序好的keyEdgeSet，以及移除各边之后不确定图对应的最大流和最可靠最大流分布及其可靠性 |

算法3-1首先要使用getALLEdgeFlowDecrease函数获取所有边发生故障之后能够满足的最大流，然后使用soetKeyEdge\_by\_flow函数对计算的流量进行排序，当前后两条边断掉之后获得的流量不相等时，根据模型定义，直接就可以比较出两条边的关键程度，而只有当两条边断掉获得流量相等的情况下，无法比较的情形，才会重复计算不确定图的最可靠最大流分布及其，使用分布可靠性来区分这两条边的关键程度。

基于重复计算的基础算法的运行效率与重复计算的次数紧密相连，在算法的首先需要计算所有边发生故障之后的最大流，本文使用的计算最大流的算法是Dinic算法，运行Dinic算法所需的处理时间为O（|V|2|E|），因此在处理这部分是复杂度是O（|V|2|E|2）；然后当移除边之后流量相同的情况下，需要重复计算，本文使用的重复计算方式为基于状态划分的算法[10]，其复杂度为O（k|V|2|E|），其中K为划分过程中需要运行Dinic算法的次数。因此整个BASE算法的复杂度为O（km|V|2|E|） + O（|V|2|E|2），其中m为需要重复计算的次数，一般m的范围是0≤m≤|E|。

## 4.3 基于流量和分布可靠性的不确定图的关键边增量算法ICA\_FDP

### 4.3.1 算法思想

类似的，由于BASE\_FDP算法的复杂度较高，在本章中只是作为一个基础算法作为对比，本节提供了一种基于流量和分布可靠性的不确定图的关键边增量算法ICA\_FDP（Increment Cost Algorithm with Flow and Distribution Probability），该算法的主要思想是根据模型中不同类型的边设置不同的增量算法计算，以减少计算复杂度，对于ABC三类边做不同的处理。

基于以上步骤，ICA\_FDP算法可以表示为如图4-1步骤：



图4-1 算法ICA\_FDP流程图

### 4.3.2 A类边计算

对于不确定图G而言，求解最可靠最大流分布的方式有很多，目前普遍使用的是算法SDBA，SDBA算法中使用到了状态划分算法，在状态划分的过程中，可以生成一个区间状态树，这样的一个区间状态树生成如算法3-2所示。

A类边根据搜索树的规则，对于每一个区间改边位置都设置为0，目的是获取不想交的的区间。对于新构成的一个子图树，本文提出一种基于流量和分布可靠性的割集状态树剪枝算法STPA\_FCP（State Tree Pruning Algorithm with Flow and Distribution Probability），该算法的主要是利用了割集中的边必定在最大流子图这一性质，通过割集中的边对于子图状态树进行剪枝，以达到减少搜索的目的。因为割集中的边一般在子图树的第二层就能达到剪枝的效果，所有该方法能够有效的减少搜索效果。

该算法首先获取根节点，对于其所有子节点，通过割集中的边进行剪枝，因为极小子图满足最大流，所以割集中的边必包含于极小子图，所以不包含割集中边的中间节点和叶子节点都必须舍去，对于依然保存的子图区间，进行进一步划分即可，指导获取极小子图。即（1）当区间的下界子图满足最大流，则整个区间满足最大流，保留下界子图，并计算下界子图概率并与原有极小子图比较。（2）当区间上界不满足最大流，则整个区间不满足最大流，舍弃区间。（3）区间下界不满足最大流，上界满足最大流，则需要二次划分区间。

|  |
| --- |
| **算法4-2.** 基于割集的状态树剪枝算法STPA\_FDP |
| **输入**：割集边CutSet，子图区间树root，不确定图G，源点source，汇点sink，移除边e  1. get CutSet，root，init stack S；  2. push root in S；  3. if S is not empty，then pop S as C；  4. if the cut edge’s position is 0 then move out C；  5. Else if C’s lower bound graph satisfy Fmax then keep C  6. If C’s upper bound graph not satisfy Fmax then move out C；  7. If C’s lower bound graph not satisfy Fmax，but upper bound graph satisfy Fmax then divide C  8. Return all graph that satisfy Fmax after G move edge e  **输出**：不确定图G在移除边e之后所有能够满足最大流的子图集合set |

### 4.3.3 B类边计算

对于B类边，同样由推论4-1可以知道，断掉B类边，流量不会减少，此时的重点在于计算剩余的不确定图的分布可靠性。为了计算移除B边之后的分布可靠性，主要是计算移除B边，原有只满足最大流的状态划分树MaxFlow-State-Tree，找到新的极小子图即可，换言之，就是要移除不能满足最大流子图，且不能存在重复的子图。本章提出算法BSA\_FDP（B Class Edge Search Algorithm with Flow and Distribution Probability），用于计算获取B边移除之后，新的极小子图。该算法步骤如下：

|  |
| --- |
| **算法4-3.** B类边搜索算法BSA\_FCP |
| **输入**：满足最大流的状态划分树MaxFlow-State-Tree，移除边e  1. For all leaf node in MaxFlow-State-Tree as C  2. If the position of e in C is 1，then move out C；  3. If the position of e position in C is 0，then push C in set；  4. If the position of e in C is X，then divide C to Ce=1 and Ce=0 and goto step （2）（3）；  5. Until MaxFlow-State-Tree is all be searched，then return set.  **输出**：移除e之后，不确定图依然满足最大流的所有子图集合set |

**证明4-3**. 对于不确定图G，经过划分后获取的满足最大流的所有区间为{*C*1，*C*2，…，*Cτ* }，满足*Ci* ∩ *Cj =* ∅，*C*1 ∪ *C*2 ∪ … ∪ *Cτ* = *A*，其中*τ* 为满足最大流的子图区间个数，假设边ei为G的B类边，根据B类边定义，存在一个k∈{1，2…*τ* }，使得 与 ，令，因为中的所有子图位置为0，所以有

若要证明 为G-之后，所有满足最大流的子图集合，需要证明以下两点：

（1），都有 ；

（2），若 ，则必存在一个，使得。

首先证明（1）：已知条件都有，且，所以可以推出对于任意的子图g，如果，都有g满足最大流，即，都有。

然后证明（2）：反正法。假设，若有，则对于，都有。

对于，且有，所以，又因为，所以，这与不存在一个使得矛盾。假设不成立。

的证。

**实例4-2.** 对于如图2-3中的不确定图G，经过状态划分获取满足最大流的子图区间有两个，分别是11x11和11101，e3是一个B边，根据以上规则，11x11区间在e3位置上的状态是x，可以分为11111和11011两个区间，因为11111和11101两个区间在e3位置上的状态是1，所以舍去，而只保留区间11011。所以最终移除e3之后，依然满足最大流的子图区间是11011。

对于C类边的计算，根据推论4-2，其断掉不会影响不确定图的最大流和分布可靠性，也就是说C类边断掉不会对不确定图的状态照成任何的影响，因此不需要计算。

### 4.3.4 算法复杂度分析

根据性质4-2，C类边移除之后不需要计算，直接使用原有的最大流，分布可靠性和分布可靠性。

基于以上处理方式，ICA\_FDP算法的伪代码如下：

|  |
| --- |
| **算法4-4.** 基于流量和容量可靠性的不确定图关键边增量算法ICA\_FCP |
| **输入**：不确定图G，源点source，汇点sink  1. Input uncertain Graph G and its source point and sink point；  2. Get maxFlow ，StateMtrix and build StateTree using algorithm SDBA  3. Get G maxflow when every single edge being cut off ALLEdgeFlowDecrease[|E|]；  4. Sort keyedge by ALLEdgeFlowDecrease[|E|]；  5. Classify ABC edge by StateMtrix；  6. When the flow （G-ei） equles the flow （G-ej） then  7. If ei is A class edge then use algorithm STPA\_FCP to get pc；  8. If ei is B class edge then use algorithm BSA\_FCP to get pc；  9. If ei is C class edge then pc is same as before ei be moved。  10. Soet KeyEdge by pc；  11. Return sorted key\_edge\_set；  **输出**：根据关键程度排序好的keyEdgeSet |

基于状态划分树的增量算法ICA\_FDP首先需要使用getALLEdgeFlowDecrease函数获取所有边发生故障之后能够满足的最大流，然后使用soetKeyEdge\_by\_flow函数对计算的流量进行排序，当前后两条边断掉之后获得的流量不相等时，根据模型定义，直接就可以比较出两条边的关键程度，只有当两条边移除之后最大流一致的情况下，会使用状态划分树的增量算法计算，该部分的计算最大流的算法是Dinic算法，运行Dinic算法所需的处理时间为O（|V|2|E|），因此在处理这部分是复杂度是O（|V|2|E|2）。然后将不确定图的边分为ABC三类边，然后根据三类边的不同性质对于不同性质的边选择不同的算法，从而简化计算过程。

首先对于A类边，考虑最坏的情况，就是重复完全计算，该算法运算复杂度主要体现在K次运行Dinic算法，因此对于A边的复杂度为O（ak|V|2|E|），其中a为不确定图G中a边的个数，对于B类边，需要遍历满足最大流的所有区间，其复杂度为O（b*τ*），其中b为不确定图B边的个数，*τ* 为满足最大流区间的个数，对于C边来说，因为不需要计算，只需要保留原有的最大流和容量可靠性，所以复杂度为O（c），其中c为不确定图C类边的个数。综上所述，基于状态划分树的增量算法的整体复杂度是O（|V|2|E|2）+O（ak|V|2|E|）+ O（b*τ*）+O（c），其中a+b+c=|E|。由公式可见，ICA\_FDP在最坏的情况下，复杂度和BASE\_FDP算法的复杂度一致。

## 4.4 基于增广路径添加的近似算法

### 4.4.1 算法基本思想

上节给出了基于流量和分布可靠性的关键边算法，虽然较BASE\_FDP算法有了较大的复杂度改进，但是对于现实中成千上万条边的网络而言，显然还是不适用的。这主要是因为对于模型中的A类边的计算，依然需要对于状态划分树中的部分未知的状态区间进行二次划分，这将耗费很大的复杂度。

本节针对A边故障之后需要二次划分的情况，提出一种对于模型中A类边计算的基于增广路径添加的近似算法ACES\_APA（A Class Edge Search Base on Augmenting Path Add）。该算法的主要思想是利用原不确定图在计算最可靠最大流分布的过程中的剩余图和简单路径，如果当模型中的A边发生故障之后，将所有包含该边的简单路径上的流量在剩余图中返还，同时将“故障边”在剩余图总移除，然后在剩余图总查找所有满足能够联通源点S和汇点T的简单路径，选择一条“可靠性性消耗最小”的一条简单路径，所谓“可靠性消耗最小”指的是这条简单路径需要更少的“未使用边”，而尽量使用更多的“未饱和边”。最后重复以上查找简单路径的步骤，知道没有从源点S到汇点T的简单路径即可，此时求得的便是移除该边之后，剩余不确定图的最大流，求得的最大流分布是相对可靠的最大流分布。本章将使用该方法求得的流分布可靠性近似的作为衡量关键边的指标。为说明问题，特设置如下实例。

**实例4-3.** 如图4-2所示不确定图和它的最可靠最大流分布，其中边E1，E2，E3，E8，E9为它的A边，表4-1为不确定图G在求得最可靠最大流分布过程中获得的简单路径，可以知道原有不确定图的最大流Fmax=4。



图4-2 不确定图G与它的最可靠最大流分布

表4-1 简单路径组合

|  |  |  |
| --- | --- | --- |
| 编号 | 路径 | 容量 |
| sp1 | s-v2-v6-t | 2 |
| sp2 | s-v1-v2-v6-t | 1 |
| sp3 | s-v1-v2-v4-t | 1 |

当边v2-v6发生故障（移除）之后，发下表4-1中的简单路径sp1和sp2都受到影响，那么移除v2-v6之后的图流分布如图4-3所示。



图4-3 移除v2-v6后的图流

此时在剩余不确定图中获取所的简单路径，图表4-2所示。

表4-2 移除边v2-v6之后的所有简单路径

|  |  |  |
| --- | --- | --- |
| 编号 | 路径 | 瓶颈容量 |
| Sp1 | S-v2-v4-t | 1 |
| Sp2 | s-v2-t | 1 |
| Sp3 | s-v1-v2-v4-t | 1 |
| Sp4 | s-v1-v2-t | 1 |

根据算法原理，可知简单路径sp1，sp2，sp4都使用到了“未使用”边，而简单路径sp3使用的都是“未饱和边”，所以在这样的情况下，添加sp3之后的瓶颈流量之后，并不会是不确定图的流分布可靠性下降。在添加这个简单路径之后，继续以上步骤，直到在剩余图中没有简单路径为止，便获得了一个最大流分布，以为该最大流分布尽量使用了原有分布中的“未饱和边”，而尽量少使用了“未使用边”，因此该最大流分布可以认为是一个较可靠性的最大流分布。

### 4.4.2 算法实现与分析

根据上述分析，下面给出基于增广路径添加的近似算法ACES\_APA的伪代码：

|  |
| --- |
| **算法4-5.** 基于增广路径添加的近似算法ACES\_APA |
| **输入**：input  1. 确定输入的边  2. 查看破坏的简单路径  3. 调整剩余图  4. 查找可靠性简单路径  5. 添加流量  6. 直到没有简单路径为止  **输出**： output |

伪代码解释：

算法复杂度分析

|  |
| --- |
| **算法4-6.** 由基于增广路径添加的近似算法ACES\_APA组成的增量算法 |
| **输入**：input  1. 输入不确定图  2. 当边是A百年的时候怎么做  **输出**： output |

伪代码解释：

算法复杂度分析。

### 4.4.3 近似算法相似度

因为不确定图的关键边算法获取的是不确定图边的关键度排序，这样不方便直接比较近似算法获取的结果的精确程度，因此，本小节设置了近似算法的相似度W。相似度W采用了算法获取的边排序精确结果与近似结果的平均欧几里得距离，那么相似度W可以表示如下：

公式4-1

如公式所示，W为精确算法和近似算法的相关性，表示为边ei在精确算法中的位置，为边ei使用近似算法求得的排序中的位置，N表示为不确定图边的个数。

**实例4-5.** 如表4-3所示，为一个有6条边的不确定图，通过精确算法和近似算法获取的不同的关键排序。那么根据相似度W的定义可以知道，W=0.33。根据相似度定义，可以知道如果相似度越小，那么说明近似算法获取的不确定图边关键度排序越接近于精确解。同时根据大量图的近似计算，不同图规模、不同图稠密度的计算可以分析近似算法对于不同类型图的计算效果的波动性。

表4-3. 近似算法与精确算法比对表

|  |  |  |
| --- | --- | --- |
| 关键度排序 | 精确算法求得的排序（数字代表边号） | 近似算法求得的排序（数字代表边号） |
| 1 | 1 | 2 |
| 2 | 2 | 1 |
| 3 | 3 | 4 |
| 4 | 4 | 3 |
| 5 | 5 | 5 |
| 6 | 6 | 6 |

### 4.4.4 近似算法下界

根据相似度的定义，W是算法获取的表排序精确记过与近似结果的平均欧几里得距离。根绝分析可以知道，对于一段排序的相关性，最差情况可能存在两种情况：

（1）首末位置顺序颠倒，然后对剩下的串继续执行颠倒操作，直到结束。如对于一个精确排序为1-2-3-4-5-6，与之相关的最差排序是6-5-4-3-2-1。

（2）等步长排序，即每一个边位置的相差都是等步长的。如对于一个1-2-3-4-5-6，与之相关的最差的排序是4-5-6-1-2-3。

根据分析，对于猜测的两种情况，首末位置顺序颠倒才是最差的情况。下面给出证明过程。

证明4-11。

根据相似性的定义，对于情况（1）首末倒置的情况有：

公式4-2

对于情况（2），根据相似性定义有：

公式4-3

对于情况（2）的相似度的表示，可以跟N为奇数和偶数分别表示如下：

公式4-4

如果要证明，则只需要分别证明当Ｎ为奇数和偶数的情况下成立即可。下面将分别使用数学归纳法证明。

I：当N为奇数时：

若要证明，只需证明，只需证明即可。

1. 当N=１时，有0≥0成立；
2. 假设当N=K时，且K为奇数，有成立。
3. 则当N=K+2时（N仍然为奇数），若要证明，经过化简只需要证明，又因为当N=K时，有成立，经过化简只需要证明即可，即证明，显然对于K=3,5,7….的奇数，是恒成立的。

所以当N为奇数时，有成立。

II：当N为偶数时：

若要证明，只需证明，即只需要证明即可。

1. 当N=2时，有2≥2成立；
2. 假设当N=K时，且K为偶数，有成立。
3. 则当N=K+2时（N仍然为偶数），若要证明，经过化简只需要证明，又因为当N=K时，有成立，所以只需要证明，经过化简只需要证明，显然对于K=4,6,….的偶数，是恒成立的。

所以当Ｎ为偶数时，也有成立。

综上所述，当Ｎ>1时，都有，所以相对于一段排序，其相关性的下界是，其中N为不确定图边的个数。

由于基于增广路径添加的近似算法主要是对于增量算法相对于A类边二次状态划分的改进，在选择增广路径的过程中，会优先使用“未饱和边”来寻找一个较可靠的最可靠最大流分布。因此算法只会使得A边的排序造成影响，所以近似算法的下界可以表示为，其中表示的是基于流量和分布可靠性的不确定图关键边模型中A类边的个数。

## 4.5 Top-K关键边搜索策略

待定

## 4.6 实验及分析

实验部分设置：

相对于 图规模，，，，，图稠密度

完全重复计算，BASELINE，ICA\_FPC的时间复杂度，空间复杂度的比较

TopK搜索出来的K个关键边是否一致。（不同的图规模，不同稠密度）

时间空间复杂度的比较

分布可靠性模型和容量可靠性模型，形同的图（图规模，图稠密度）是否在时间上和空间上有优势。

分布可靠性和容量可靠性 模型计算出来的关键边的相关性。

近似算法的比较

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各种因素，本文进行了一系列的实验，实验平台为一台Intel Core的PC机（CPU i7-3770，3.40GHz，内存8GB，64位windows 7操作系统），算法采用C++在VS2010上实现。

### 4.6.1 实验数据

本文采用了与文献[10]相同的数据集，使用NETGEN生成器[11]产生V6E10、V8E14、V10E18、V12E22、V14E26共5组不同图规模的二态有向图集合（本文实验数据集合大小为5），其中VnEm表示有n个顶点、m条边组成的图，图中边的容量与对应概率满足均匀分布。通过在不同图规模的情况下，比较算法BASE\_FDP和ICA\_FDP算法在运行时间及内存消耗放方面的差异。实验中的BASE\_ALL算法指代的是完全重复计算的方式。

最好添加实际的数据

### 4.6.2 实验结果与分析

**实验4-1.** 不同图规模对于算法性能的影响

（a）时间消耗 （b）内存消耗

图4-4 不同图规模情况下BASE\_FDP和ICA\_FDP算法时间和内存消耗

根据图4-4所示，算法BASE\_ALL算法是完全重复计算的方式，BASE算法是在当流量一致的情况下才计算分布可靠性。如图4-4（a）所示，BASE\_FDP算法相对于BASE\_ALL算法在运行时间上有一定的减少，但是依然不适应大规模图，而ICA算法在运行时间上有了很大的减少，适应性更好。如图4-4（b）所示，ICA\_FDP算法相对于BASE\_FDP算法在空间复杂度上有了一定程度的增加，但是随着图规模的增加，内存的使用并没有增加太多，依然在可以接受的范围内。

**实验4-2.** 不同图稠密度对于算法性能的影响

为更好的反应算法BASE\_FDP算法和ICA\_FDP算法的性能差异，使用NETGEN生成器生成V15E21，V15E32，V15E42，V15E53四种不同稠密度的图，通过在不同图稠密度的情况下，比较算法BASE\_FDP和ICA\_FDP算法在运行时间及内存消耗放方面的差异。实验中的BASE\_ALL算法指代的是完全重复计算的方式。试验结果如下：

（a）时间消耗 （b）内存消耗

图4-5 图稠密度对BASE\_FDP和ICA\_FDP时间和内存消耗影响

如图4-5所示，相对于不同稠密度的图，ICA\_FDP算法相对于BASE\_FDP算法在空间复杂度上有一定的增加，但是时间复杂度上有较大的优势。

### 4.6.3 正确性（相关性）比较

（算法波动性）

不同的数据有好处

## 4.7 本章小结

本章主要构建了一种基于流量和分布可靠性的关键边衡量模型，针对不确定图边移除之后，相对流量和分布可靠性的减少量来衡量。对于构建的基于流量和容量可靠性的模型，本章节首先设计了一种基于重复计算的基本算法BASE\_FDP。因为BASE\_FDP算法复杂度较高，本章提出一种基于流量和分布可靠性的不确定图关键边增量算法ICA\_FDP，并重点介绍了算法的思想与实现，同时对算法进行了正确性证明以及时间复杂度和空间复杂度的分析。因为不确定图计算分布可靠性依然是一个NP-Hard问题，精确算法复杂度很高，因此本文章针对增量算法的A类边提出了对于A边的基于增广路径添加的近似算法，同时分析了算法复杂度，给出了近似算法和精确算法的相似度度量，并分析了近似算法的下界。最后通过试验分析，比较BASE\_FDP算法和ICA\_FDP算法在空间复杂度和时间复杂度上的差别，实验结果表明ICA\_FDP算法相对于BASE\_FDP算法其空间复杂度有一定的增加，但是时间复杂度方面具有较大的优势。ICA\_FDP相较于ICA\_FCP算法，在时间和空间复杂度上都有了优化，近似算法的波动性不强，能够适应不同图规模和稠密度的图。

# 第5章 不确定图关键边增量算法优化策略

本章部分的优化策略是是针对增量算法的，这些优化策略适用于所有增量模型。

## 5.1 基于确定边过滤的增量算法ICA\_FEA

### 5.1.1 算法基本思想

考虑到传统状态划分算法需要划分的状态空间较大，本节在原有计算方式的基础上提出一种基于确定边过滤的状态划分算法ICA\_FEA（Filter Edge Algorithm），该算法的主要思路是在使用状态划分算法之前，首先给出能够确定类别的边（定性的边），从而减少状态划分的空间。这些确定边包括割集中的边和悬挂边。

假设不确定图G的边集为E，那么在传统的状态划分算法中，需要划分的状态区间的个数有2|E|个，而在基于过滤集合的状态划分算法中，只要能够确定一条确定边（包括割集中的边或者悬挂边），都能够是状态划分的区间减少一半，显然尽可能多的确定这些边能够很大程度上简化算法复杂度。该算法对于连接度比较高的图效果更好。

FEA算法的是一个基于ICA的算法，只是减少了ICA算法状态划分的初始空间。因此FEA算法的主要着手点就是如何找到所谓的确定边。确定边包括割集中的边和悬挂边。

对于割集中的边，目前求解割集中的边已经被证明是一个NP-Hard问题[23]，本文提供一种计算割集中的边算法，该方法不追求找到所有割集中的边，而是尽可能多的找到割集中的边，且满足找到的边一定是割集中的边。该方法的主要思想是首先用最大流最小割算法找到一组割，获取的最大流为全局最大流，然后，根据割集将图分割为两部分，并补全虚拟源点或虚拟汇点，连接割集边上的顶点到虚拟源点或者虚拟汇点作为一条边，并设置边的容量为一个大于原容量的值；接着，对构造的新图递归使用最大流最小割定理计算局部最大流，如果局部最大流等于全局最大流，则表示该最大流表示的割集也是所求割集，并将该子图继续分割，如果局部最大流大于全局，则舍弃该子图，直到没有可以计算的子图为止，获得的边即为所求部分割集中的边。

对于悬挂边，指的是只有入度或者只有出度的边（特殊的对于源点的入度边、汇点的出度边归为此类），不管单入度边还是单初度边对于不确定图的最大流都是可有可无的，在计算的时候只需要通过图中点的出入度判断即可。

### 5.1.2 割集中的边计算

目前求解图中所有割集的方法有割集矩阵算法[23], 然而求解图中所有的割集是一个NP问题[24]，当图的规模越来越大，算法复杂度会上升很快，显然是不合适的，因此本文使用一种递归的方式求解割集中的边，该方法的目的是尽可能多的求解割集中的边，而不能保证求得的边包含所有割集中的边。

在一般的图中，割集之间一般有两种形式存在，一种是独立的割集，另一种是关联割集。独立割集就是多个割集之间没有重合的割边，而关联割表示，多个割集中至少存在一条边公用。

图4.5某不确定图的最大子图 图4.6 不确定图最大子图的割集表示

如上图所示，图4.5是某个不确定图的最大子图MSG（G），这个最大子图的其中一个割集如图4.6所示，那么MSG（G）的的割集可以表示为{（ST，SV）（ST，VT）}，可以看出（ST，SV）（ST，VT）两个割集中包含相同的边，则表明两割集是关联割集。同理如果两个割集中不含有相同边，表示两割集为独立割集。

如果求解如图2中的割集，经过划分获得如下的两个子图，如图3.



图4.7 求得一个割集之后获得的两个子图

使用递归的方式求解所有子图中的局部最小割，直到没有子图生成为止，返回所得割集中的边，就是割集中的边。

在一般的图中也会存在一些悬挂边，这些边对于计算不确定图关键程度之前可以确定下来，已减少状态划分算法的划分区间，从而减少算法的复杂度。

算法流程图如下：



**图xxx** 求解部分割集中的边的算法流程图

算法伪代码：

证明：使用该算法获取的边都是割集中的边。（正确性证明）

**待证明。**

### 5.1.3 悬挂边的计算

悬挂边，指的是只有入度或者只有出度的边（特殊的对于源点的入度边、汇点的出度边归为此类），不管单入度边还是单初度边对于不确定图的最大流都是可有可无的，在计算的时候只需要通过图中点的出入度判断即可。

如下图所示为一个比较特殊的不确定图的最大子图，课件点v1只有初度而没有入度，V2只有入度而没有初度，那么v1便是单出度点，V2是单入度点，又因为这些点都是中间节点，因此与这些点相连的边都不会出现在流分布当中，自然也不会出现在最大流分布当中，所以都可以去除，从而减少状态划分算法的划分空间，从而减少算法的复杂度。



图4.8 某不确定图的最大子图

计算悬挂边的算法流程如下图：



**图xxx** 计算悬挂边流程图

算法伪代码

### 5.1.4 基于确定边过滤的关键边优化算法ICA\_FEA

FEA算法的主要框架还是基于ICA算法，只是在使用ICA的状态划分算法之前，确定了部分确定边，使得状态划分算法的划分空间减少，从而减少复杂度。对于出后后的不确定图部分边是确定的，部分是不确定的。对于确定的边使用向量表示方法做如下设置：

对于割集中的边，始终满足空间中存在，即为1；

对于悬挂边，始终满足划分空间中不存在，即为0；

然后对于预处理的初始划分空间，使用状态划分空间求得所有满足最大流的不确定的子图区间。

使用确定边过滤之后的状态划分算法可以有效的减少算法的复杂度。

使用该方法的流程图如下：



算法xxx FEA算法整体流程图

### 5.1.5 实验及分析

## 5.2 基于下界子图树的增量优化算法

### 5.2.1 算法的基本思想

解释算法的思想

### 5.2.2 构造下界子图树与子树搜索过程

### 5.2.3 实验结果及分析

验证算法是否有效

## 5.3 本章小结

说明情况

# 第6章 总结和展望

致谢

致谢

参考文献

1. Kang H, Butler C, Yang Q, et al. A new survivability measure for military communication networks[C]//Military Communications Conference, 1998. MILCOM 98. Proceedings. , IEEE. IEEE, 1998, 1: 71-75.
2. TAN Y, WU J, DENG H. Evaluation method for node importance based on node contraction in complex networks[J]. Systems Engineering-Theory & Practice, 2006, 11（11）: 79-83.
3. Newport K, Schroeder M, Whittaker G M. Techniques for evaluating the nodal survivability of large networks[C]//Military Communications Conference, 1990. MILCOM'90, Conference Record, A New Era. 1990 IEEE. IEEE, 1990: 1108-1113.
4. Nardelli E, Proietti G, Widmayer P. Finding the detour-critical edge of a shortest path between two nodes[J]. Information Processing Letters, 1998, 67（1）: 51-54.
5. Nardelli E, Proietti G, Widmayer P. A faster computation of the most vital edge of a shortest path[J]. Information Processing Letters, 2001, 79（2）: 81-85.
6. Callaway D S, Newman M E J, Strogatz S H, et al. Network robustness and fragility: Percolation on random graphs[J]. Physical review letters, 2000, 85（25）: 5468.
7. Freeman L C. A set of measures of centrality based on betweenness[J]. Sociometry, 1977: 35-41.
8. Chen Y, Hu A Q, Hu J, et al. A method for finding the most vital node in communication networks[J]. High Technology Letters, 2004, 1（2）: 573-575.
9. Xiong J, Li J, Shen D, et al. Evaluation Method for Node Importance of Information System Networks Based on Edge-betweenness[J]. Keji Daobao/ Science & Technology Review, 2013, 31（14）: 53-55.
10. Cai W, Zhang B L, Lv J H. Algorithms of the most reliable maximum flow on uncertain graph[J]. Jisuanji Xuebao（Chinese Journal of Computers）, 2012, 35（11）: 2371-2380.
11. Klingman D, Napier A, Stutz J. NETGEN: A program for generating large scale capacitated assignment, transportation, and minimum cost flow network problems[J]. Management Science, 1974, 20（5）: 814-821.
12. Aggarwal C C, Yu P S. A survey of uncertain data algorithms and applications[J]. Knowledge and Data Engineering, IEEE Transactions on, 2009, 21（5）: 609-623.
13. Yi-Kuei Lin, Cheng-Ta Yeh. Evaluation of optimal network reliability under components-assignments subject to a transmission budget. IEEE Trans. Reliability, 2010, 59: 539-550.