

不确定图最可靠最大流近似算法研究

学生姓名：李富豪 学号：0906010410 指导教师：曹 敬

摘 要

由于测量手段和问题特性的限制，数据的不确定性普遍存在，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。不确定图中最大流理论有着重要的研究价值，可有效的解决诸如构建可靠性网络、选择可靠性方案等相关问题。本文针对现有的精确求解最可靠最大流算法普遍存在的算法复杂度偏高，难以应用于大规模网络的情况，本文提出了两种不确定图最可靠最大流的近似算法。

算法一是简单路径流量调整算法，算法二是基于剩余稳定性的批自适应最大流算法，该算法首先构造可靠性残留图，其顶点与原来网络完全相同，而是把原来网络中的弧分解成两条方向相反的弧并按一定的关系赋值，从流量为 0，可靠性为 1 初始状态开始，每次在可靠性残留图中寻找一条可靠性最大的增广链，据此增广链在残留容量网络中调整容量，在可靠性残留图中调整残留可靠性，直至没有增广链为止，便获得了最大的流量，由于每次查找的都是可靠性最大的弧，因此方案的可靠性为较大可靠性。

为了综合衡量在算法的迭代过程中流量的增加与可靠性下降的变化情况，本文在算法二的基础上引入了增广效益（指标），该指标主要为了控制迭代次数，不用找到所有的增广路径。最后通过实验仿真，比较现存的精确算法与本文提出的两种算法性能差别，实验结果表明算法一在损失精度的情况下，性能并没有得到很大改善，算法二虽然损失了精度，但是性能得到了显著改善，如使用 75 顶点 3000 条边的不确定图时，该算法运行时间保持在 3 秒中以内，其他算法现有算法则不可计算，增广效益（指标）也得到了可靠验证，能够表示流饱和程度与迭代次数的关系。

ABSTRACT

The measuring method and characteristic of the problem constraints, the data uncertainty is prevalent, such as the electricity transmission network, power transmission equipment has the possibility of failure, and each traffic network in the road there is a congestion probability, when this uncertainty is expressed in the form of map data, uncertain graph. Do not determine the maximum flow graph theory has important research value, can effectively resolve the reliability of the network, such as the construction of relevant questions about reliability program. In this paper, the exact solution of widespread existing most reliable maximum flow algorithm complexity is high, it is difficult to apply in the large-scale networks, this paper presents two kinds of uncertain graph is the most reliable approximation algorithm for maximum flow.

Algorithm is simple path flow adjustment algorithm, algorithm two is the number of remaining stability based on adaptive maximum flow algorithm, the algorithm first constructs a graph whose vertices and reliability of residue, the original network is identical to the original, but of arcs in the network is decomposed into two opposite arc and the assignment relationship, from the flow of 0 the reliability of the initial state, 1, every time the reliability of residue for a maximum reliability augmented chain graph, thus augmenting chain adjustment capacity in the residual capacity in the network, the reliability of residual reliability adjustment in the picture, until no augmented chain, can obtain the maximum flow, because each search are the maximum reliability the arc, so the scheme for greater reliability reliability.

In order to increase the reliability of comprehensive measure changes with flow in the iterative process of decline, this paper introduces extended benefits in algorithm two is based on (index), the main indicators in order to control the number of iterations, not all find augmenting path. Finally, through simulation, two kinds of algorithm performance difference is more accurate algorithm and the existing proposed, experimental results show that the algorithm in the loss of precision, performance is not improved, algorithm two although the loss of precision, but the performance is significantly improved, such as the use of 75 vertex uncertain graph 3000 edges time, the running time of this algorithm maintains in 3 seconds within other existing algorithms, algorithm is immeasurable, extended benefits (index) has also been verified, can express the relationship flow saturation degree and the number of iterations.

正文摘录

一、研究背景与选题依据

图最为一种通用的数据结构,以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堵塞的概率,当这种不确定性表达与图数据中,则形成不确定图。

最大流最为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,对于最大流通常对应多个不同的分布,对于确定图而言,这些不同的分布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 A 点传输到 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能,就需要找到最可靠的一种传输电能的方案。

二、研究内容与研究目标

本文主要对不确定图最可靠最大流问题进行了相关的研究。

首先阐述了该问题的研究背景及重要意义,然后针对现有的精确求解最可靠最大流算法普遍存在的算法复杂度偏高,难以应用于大规模网络的情况,本文提出了两种不确定图最可靠最大流的近似算法。算法一是简单路径流量调整算法,算法二是基于剩余稳定性的批自适应最大流算法。

算法一是在现有最大流最可靠性的精确算法-流量组合算法过度而来,其损失了算法的精度,但是算法的性能并没有降低很多,原因是该算法仍然要在简单路径的空间进行搜索,假设由 n 个菱形组成的不确定图,那么该不确定图的边数为 $4n$,但是其简单路径的条数达到了 2^n (指数级别)。在此基础上提出的基于剩余稳定性的批自适应最大流算法,解决了此问题,并很大程度上提高了算法的性能。

本文主要研究目标:考虑到现实情况,在实际应用过程中,如在电力的传输过程中,并不需要时刻以最大电压方式传输,可能会对电力设备有损坏,在实际的传输中只需要以额定的电压传输即可。

现存对不确定图的最大流最可靠性的精确算法,但相关算法的复杂度普遍存在算法复杂度偏高,难以应用与大规模复杂网络的情况,本文基于此提出该问题的近似算法,既能很大程度上降低算法的复杂度,又能很好的应用现实问题。

三、不确定图最大流可靠性算法

1. 基于简单路径流量调整算法

该算法首先寻找所有简单路径，并在所有简单路径中寻找可靠性最大的那条，然后根据此条简单路径的信息调整其他相关简单路径的流量和可靠性，当简单路径的流量为零时将其删除，直至没有简单路径为止，从而获得最大流的可靠性方案。

2. 基于剩余稳定性的批自适应最大流算法

该算法首先构造可靠性残留图，其顶点与原来网络完全相同，而是把原来网络中的弧分解成两条方向相反的弧并按一定的关系赋值，从流量为 0，可靠性为 1 初始状态开始，每次在可靠性残留图中寻找一条可靠性最大的增广链，据此增广链在残留容量网络中调整容量，在可靠性残留图中调整可靠性，直至没有增广链为止，便获得了最大的流量，由于每次查找的都是可靠性最大的弧，因此方案的可靠性为较大可靠性。算法二有效的解决了算法一容易受到简单路径数量影响的问题。

3. 增广效益（指标）

对于现实情况可能在增加最后的残留容量的时候，可靠性瞬间降低，甚至导致方案的可行性大大降低，同时考虑到已有算法的复杂度过大，在实际应用到大型复杂的不确定图时，算法基本不可用，因此本文在基于剩余稳定性的批自适应最大流算法的基础上提出参考增广效益（指标），该指标与残留容量，每条增广链增加的容量以及在添加此条增广链可靠性减低的数值有关，指标越大，则表示添加此条增广链的需求越大。

在基于剩余稳定性的批自适应最大流算法中，每添加一条增广链都会相应的以减少部分可靠性为代价，增加部分流量，直到没有可增加的流量为止便得到最大流量。再每选择添加一条增广链的时候，会对总体的可靠性方案产生响应的影响，称作增广效益（指标），记作 Beta，其表示方法为：

$$\text{Beta} = k * \frac{\text{RemainFlow}}{\text{MaxFlow}} * \frac{\text{ChangementFlow}}{\text{MaxFlow}} * \frac{1}{\text{ChangementP}} \dots \dots (4)$$

其中式中的 MaxFlow 表示的是不确定图 G 最大子图 MSG(G) 的最大流，RemainFlow 表示的是残留容量网络中离最优值的距离，ChangementFlow 表示的是此条增广路径可以增加的流量，ChangementP 为此条添加此条增广路径，最大流方案的可靠性降低的量，K 表示的是系数，一般设置为 100，可以理解为此条增广路径的可选择性，当 Beta 越大，那么可选择性就越大，当 Beta 越小那么选择此条增广路径的可行性很小，即很可能是添加很少的流量的时候，可靠性降低很大。

五、实验与结论

为了验证本文所提到的算法的运行效率及其分析影响算法性能的各种因素，我们进行了一系列的实验，其中算法都在 STL 库支持下用 Visual C++6.0 实现，本文实验以普通 PC 为平台，基本配置为：处理器 inter(R) Core(TM) 2 duo，CPU2.20GHz，内存 2.00GB，32 位 Windows7 操作系统。

本文实验使用的实验用例，都是在图相关研究中普遍采用的有向图生成器 NETGEN 基

基础上，按照一定的规则合成测试数据集。

实验 1. 首先考虑图的规模对与基于简单路径流量组合算法和基于可靠性网络算法运行时间与消耗内存方面的影响。

(a) 算法运行时间与图规模的关系

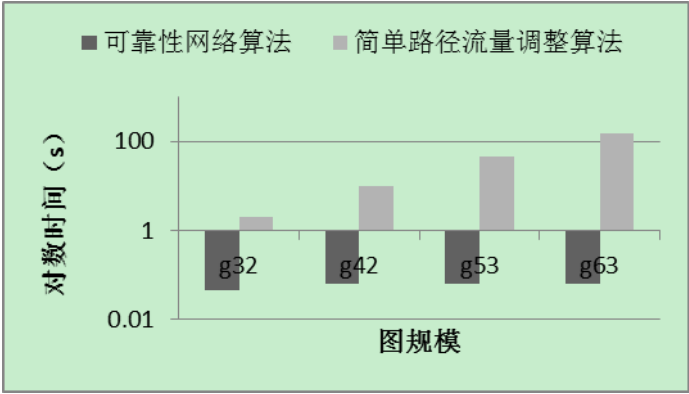


图 4-1 运行时间与边数量的关系

图中可以看到当图规模越来越大的时候，基于可靠性网路的算法的时间一直都很小，保持在 1 秒以内，而基于简单路径流量组合原理的算法在图规模还比较小的时候，算法的时间尚可以接受，但是在图规模越来越大的时候，时间已经不可计算了，图中当边的数量为 84 的时候，已经没有计算的价值了。

然而本文提到的两种算法都是近似算法，在舍弃精度的同时如果不能在算法性能上得到很大的改善，算法的改进就没有了意义，上述实验的实验用例图的图规模都太小，不足以证明算法的性能。因此在次基础上，又找到更为复杂的实验用例，即 75 个定点，3000 条边的实验用例，实验证明，可靠性残留图算法运行时间均保持在 3 秒以内，算法的时间复杂度得到了很大的改善。

(b) 内存消耗图规模的关系

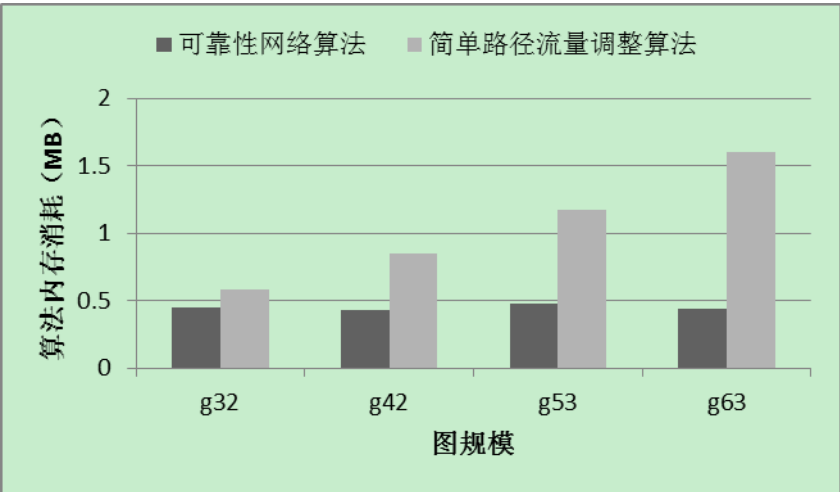


图 4-2 内存消耗与边数量之间的关系

由图 4-2 可以看出，随着边数量的增加，算法消耗的内存也会随之增加，可以看出，基于可靠性网络的算法的性能由于基于简单路径流量调整算法。而对于基于可靠性网络算法，其内存消耗存在不确定的情况，这是因为其内存的消耗与需找增广链的次数有关。

实验 2. 考察基于简单路径流量组合算法和基于可靠性网络算法两会总算法，对于不同规模的图是否能够对可靠性存在影响。

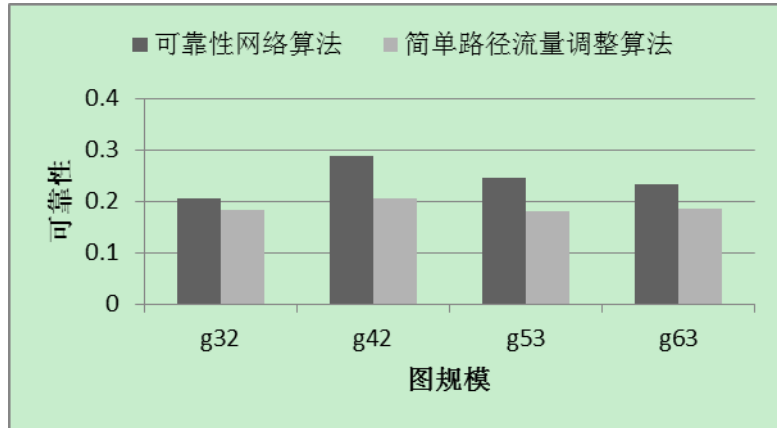


图 4-3 两种算法对于最大流方案可靠性的影响

由图 4-3 可以发现，对于不同的不确定图，基于可靠性网络算法求得的可靠性总体要高于基于简单路径流量组合算法求得的可靠性，两种算法求得的最大流都是一样的。

实验 3. 验证阈值在基于可靠性网络算法中的影响与作用。

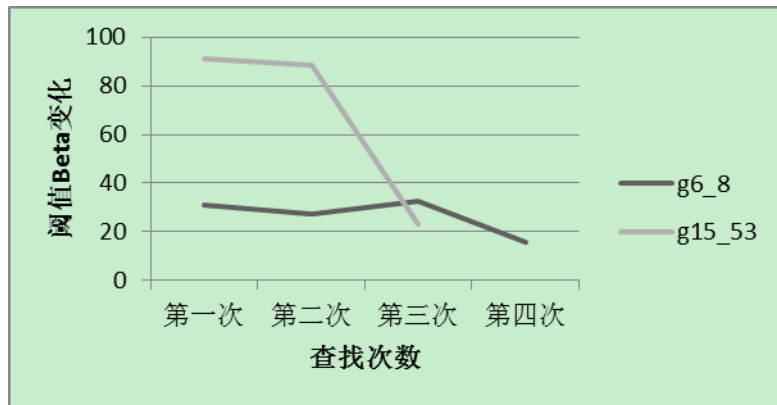


图 4-4 基于可靠性网络算法中阈值的作用图

图 4-4 是基于可靠性网络算法对于 g6_8 和 g15_53 两个不确定图运行是得到的阈值，其变化趋势总体大致呈下降趋势，阈值表示的是增广链添加的必要性，当最后一次添加的时候，可能就没有这个添加的必要性，因为可能在提高很少的流量的时候可靠性降低很多。

图 4-4 中表示的图只是边数较为简单的图，不能很好的表现可靠性残留图算法的优势，因此增加一个复杂用例，即 75 个顶点和 3000 边的不确定图，这样能更好的表现在流量增广

的过程中，增广效益（指标）的变化过程，也就是在实际应用中调整的价值。图 4-5，图 4-6 为两个复杂图用例增广效益与迭代次数之间的关系。

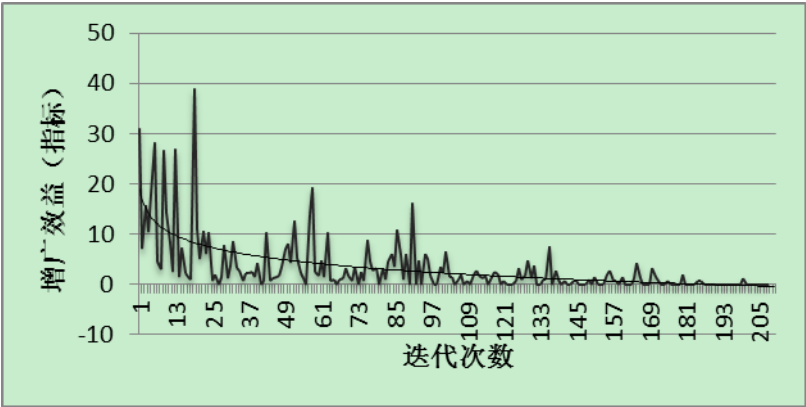


图 4-5 基于可靠性网络算法中阈值的的作用图

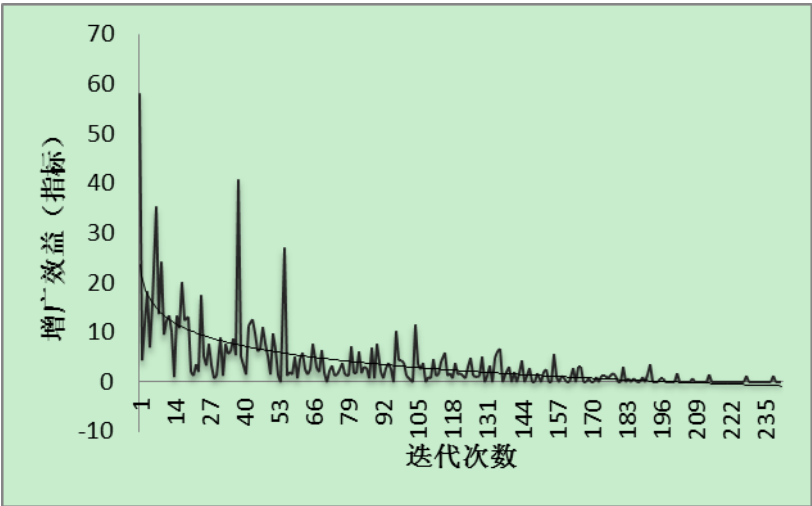


图 4-6 基于可靠性网络算法中阈值的的作用图

参考文献

- [1] 蔡伟. 不确定图最可靠最大流算法研究[N]. 计算机学报. 2012.
- [2] 高宏, 张炜. 不确定图数据管理研究现状.中国计算机学会通讯, 2009, 5(4): 31-36.
- [3] 周傲英, 金澈清, 王国仁, 李建中. 不确定性数据管理技术研究综述[J]. 计算机学报. 2009, 31(1): 1-16.
- [4] 李建中, 于戈, 周傲英. 不确定性数据管理的要求和挑战.中国计算机学会通讯[J]. 2009, 5(4): 6-14.
- [5] Sarma A D, Halevy A., Widom J. Working models for uncertain data. Proceedings of the 22th International Conference on Data Engineering(ICDE 2006).Atlanta, Georgia, USA, 2006.
- [6] 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.
- [7] Hintsanen P, Toivonen H. Finding reliable subgraphs from large probabilistic graphs. Data Mining and Knowledge Discovery, 2008, 17(1): 3-23.
- [8] Potamias M, Bonchi F, Gionis A, Kollios G. k-Nearest Neighbors in Uncertain Graphs. Proc of VLDB, September 2010, Singapore.
- [9] 邹兆年, 李建中, 高宏, 张硕. 从不确定图中挖掘频繁子图模式. 软件学报, 2009, 20(11): 2965-2976.
- [10] Zou Z, Li J, Gao H, Zhang S. Mining frequent subgraph patterns from uncertain graph data. IEEE Transactions on Knowledge and Data Engineering, 2010, 22(9): 1203-1218.
- [11] 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.
- [12] 韩蒙, 张炜, 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.
- [13] 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.
- [14] 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, March 22-26, 2010.
- [15] Kurzhanski A B, Varaiya P. On reachability under uncertainty. SIAM J. Control Optim., 2002, 41(1): 181-216.
- [16] Rasteiro D DM L and Anjo A J B . Optimal paths in probabilistic networks. Journal of Mathematical Sciences, 120(1):974{987, 2004}.
- [17] Aggarwal C C, Yu P S. A survey of uncertain data algorithms and applications, IBM Research Report[M], October 31, 2007.