常见图论数据的构造算法的研究与实现

葛昌威，王桂平，黄春淦

(重庆交通大学信息科学与工程学院，重庆 404100)

**摘要：**在近些年的算法竞赛与数据结构课程中,与图论相关的题目屡见不鲜。图论中的图是由若干给定的点和连接两点之间的线所构成的图形,这种图形用点代表实体,用线表示两个实体间的关系，常用于描述实体之间的特定关系。然而，算法竞赛与数据结构课程中各种图论模型的传统构造算法在随机性和时间效率方面存在表现不佳的问题。因此，本文针对算法竞赛与数据结构课程中常见的各种图论模型，提出了兼顾随机性与时间效率的构造算法。

**关键词**：图论；算法；调和级数；几何分布

A Study and Realisation of Construction Algorithms for Common Graph Theory data

Ge Changwei, Wang Guiping, Huang Chungan

(Shool of Information Science and Engineering, Chongqing Jiaotong University, Chongqing, China, 404100)

**Abstract：** In recent years, graph theory has been a common topic in algorithm competitions and data structures courses. A graph in graph theory is a figure consisting of a number of given points and a line connecting the two points, which represents an entity by a point and a relationship between two entities by a line, and is often used to describe specific relationships between entities. However, the traditional construction algorithms for various graph theoretic models in algorithmic contests and data structure courses suffer from poor performance in terms of randomness and time efficiency. Therefore, this paper proposes a construction algorithm that balances randomness and time efficiency for different kinds of graph theoretic models commonly used in algorithmic competitions and data structure courses.

**Keywords：** graph theory; algorithms; reconciliation levels; geometric distributions

# 0 引言

图论是数学的一个分支，它以图作为研究对象[1]。图论中的图是由若干给定的点和连接两点的线所构成的图形，这种图形常来用于描述某些事物之间的某种特定关系，其用点代表实体，用连接两点的线表示相应实体间的关系[2,3]。常见的图模型大体上可以分为树，有向无环图，有向带环图，负环图，二分图，仙人掌等。在算法竞赛与涉及图论的课程中，我们往往需要构造图数据以验证算法的正确性[4]。目前，大部分的研究集中在利用图模型来进行后续的研究，如图像分割[5]、数据挖掘[6]、信息安全[7]等领域。然而对图模型本身的构造方法研究较少[8,9]。但是尤其是在算法竞赛与涉及图论的课程[10]中，我们往往需要构造图数据以验证算法的正确性。

在算法竞赛与涉及图论的课程中常见的图大体上可以分为树，有向无环图，有向带环图，负环图，二分图，仙人掌等[11]，本文将按边权的构造，一般图的构造，特殊图的构造依次进行介绍, 并给出兼顾随机性与时间效率的构造算法实现方案。

# 1 边权的构造

构造图的边权等价于构造一个数列，该数列由数列长度，值域下界，值域上界，数列中数值是否可重复，四个限制条件生成。若为，使用随机数生成器构造。若为，考虑下述两种算法。

算法：使用随机数生成器产生数据，通过重复值检查后加入到集合。

算法：将值域中所有整数数值存入数组，随机打乱数组后取其中项。

证明1.1：设，当时，算法1.1的期望时间复杂度近似小于算法。

如果已经得到了个不同的数字，下一次随机取数结果是一个未出现数字的概率为：

在值域中进行若干次随机取值，得到个不同数字的期望次数：

算法期望时间复杂度为(1.2)，算法时间复杂度为。将期望时间复杂度作为判断条件，得表达式。若为真采用算法。反之采用算法。

计算时间复杂度为。考虑近似值优化：当时，左式为调和级数。调和级数有不等式[3]：

将带入，得到左式的一个近似解：

将带入，得证明判别式：

实验1.1：图展示了算法和算法在不同值域下，生成不同数量随机数的平均用时(单位:毫秒)。具体代码开源在<https://github.com/gcwAndHisFriends/cactus_construct_check/blob/main/time_checker.cpp>。

|  |  |
| --- | --- |
|  |  |
|  | |

图 1

可以看出在实际测试中，当，算法1.1的耗时超过了算法。实验结果仅供参考，算法的实际运行耗时受语言，实现细节，CPU等因素影响。表格则展示了算法在不同值域范围下的运行耗时(单位：毫秒)。

表格 1

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| 值域范围 | 100 | 1000 | 10000 | 100000 | 1000000 | 10000000 |
| 耗时 | 0.008 | 0.044 | 0.454 | 4.272 | 59.033 | 1302.48 |

值得注意的是当后，常规内存就难以支持算法的开销。构造较大的数据集，那么分值域依次进行构造是一个比较合适的方法。



# 2 简单有向图的构造

非连通图由多个连通图组合而成，故此仅讨论连通图构造方法。假设要构造一个包含个节点，条边的图。按算法2构造一个基本边集保证图的连通性。



在算法中，除节点之外的节点至少和一个编号小于自身的节点直接连通，可以推出任意节点都与号节点连通，由连通的传递性可知边集中任意两点互相连通。

表示节点在有根树中的深度。选择任意两点，，添加有序顶点对至边集。当且仅当所有有序顶点对，都有，那么构造出的图为有向无环图。只要存在有一个有序点对，满足，那么构造出的图为有向带环图。

简单图要求不能存在重边与自环，在构造边时需根据边密度进行构造。设为图的最大边容量上限，有向无环图的为，有向带环图的为。

当，可以选择在拓扑序中随机选点，判重后连边的方法。反之，枚举出所有边，打乱后随机取其中一部分边的方法更合适。有向无环图具体构造方法可参见算法，有向带环图具体构造方法可参见算法。

# 3特殊图的构造

## 3.1 负环图的构造

如果图中存在一个边权总和为负数的环，这个图就被称为负环图。负环图由节点数量，边数量，值域下界，值域上界，四个限制条件生成。

根据值域和的不同，有以下两种构造方法：

算法：使用算法4随机构造一个有向带环图，再对图中是否存在负环进行判断。

算法：生成一个树，非随机地构造一些负环连接到树上。

由算法4生成的有向带环图中的任意一个环，其为负环的概率为：。

|  |  |
| --- | --- |
|  |  |

如果图中包含个环，则构造出的图中不包含负环的概率为：。

在的情况下，如果，构造出的图中有以上的几率至少包含一个负环。

算法可以对图中是否存在负环进行检验。在算法中每个点在被更新最短距离后会重新入队，由于负环上的边权和为负数，每次遍历负环都会得到更优答案，通过观察算法是否陷入死循环来判断图中是否存在负环。根据引理——入队次数超过节点总数的节点必定在负环上。在一个不包含负环的图中，节点最多只会更新次。当入队次数超过则必然存在一个负环。具体的检验方法可参见算法。



## 3.2 二分图的构造

节点由两个集合组成，且两个集合内部没有边的图被称为二分图。

算法：将所有点随机分为两个集合，在两个集合之间随机选点连边。

算法保证了高随机性，但构造出的二分图无法保证连通。

引理：如果一个图中不存在奇环，那么该图就是二分图，否则该图不是二分图。

对引理的简单证明：假设二分图中存在一个奇环[下文中的环若无特殊说明，均表示简单环]，环上节点编号分别为。任意相邻两点有边连接，且有一条边相连。设属于集合，以此类推。因为有一条边相连，其均属于，与二分图相连结点属于不同点集的定义矛盾，即证二分图不含奇环。

算法：构造一棵节点数为的树，称原树边集。取任意节点为根。设表示节点的深度，为节点的父节点，。对于任意节点，

随机取两节点，，若满足，，，则添加至新增边集。

证明：仅包含新增边与部分原树边的环为偶环。

设两点分别为，，为两点在原树上的距离，表示两点在原树上的最近公共祖先。因为树中两点之间路径唯一，并且

新增边与部分原树边的环为偶环得证。

定义是一个由条新增边和部分原树边组成的环，包含的第条新增边为。为新增边与部分原树边构成环的边集。，表示两个边集的对称差运算。

证明：与最多只有一段路径重叠。

反证法，如果存在两段路径重叠，由于，所以原树上的两条路径，之间有两段重叠路径。设，分别是这两段重叠路径中的点，则可以通过原树上两条不同路径到达，与树的性质相悖，故假设不成立，原命题得证。三段以上重叠区域的证明同理。

证明：表示一个由，与部分原树边构成的，具有唯一性的环(后续均不讨论不存在的情况)。

证明为环：由证明可知，与最多只有一段连续路径重叠，设路径两端节点分别为，。与非重叠的边集所对应的点集除，外两两不同，将并集可以表示为仅有一对重复点对的顶点序列，并且该点集可以对应的边集，故为环。

证明具有唯一性：如果存在两个环，均仅包含与与部分原树边，并且至少存在一条边不同。说明点集中至少存在一个点对，在原树边集中的路径不唯一，这与树的定义相悖，假设不成立，唯一性得证。

证明：表示一个偶环。

由证明，与均为偶环。

引理：任意两个偶数大小的集合的对称差也是偶数大小。

这里给出一个简单的证明：设集合由个独有的元素，个共有的元素，集合由个共有的元素，个独有的元素。。，包含个元素，当为奇数，，均为奇数。当为偶数，，均为偶数。故无论的奇偶性，均为偶数，得证。

证明：。

，，。对称差具有结合律[[1]](#footnote-1)，，原式得证。

证明：表示一个偶环。

由证明：

由证明：是一个由条新增边和部分原树边构成的具有唯一性的环的边集。故。

由证明、证明可知表示偶环，故为偶环得证。

结论：在原树中进行若干次操作，每次操作选择两个深度奇偶性不同的节点连边，最终的到的图不包含奇环，故算法构造出得图为二分图。

具体步骤参见算法6。



## 3.3 仙人掌图的构造

如果一张无向连通图的每条边最多在一个环内，则称它是一棵仙人掌。

设为节点的子节点编号集合，表示集合的大小，表示节点的父亲节点编号。

若节点为根，或，称为开始节点。从开始节点出发，至叶子节点结束的路径称为树链。

算法：从根节点开始进行向下遍历一棵树。对于节点，设为节点所在树链的中深度最浅的节点编号，随机取节点。对于如果，(表示赋值)，否则。若，则，属于同一条树链。设表示值等于的节点集合。对于每个元素数量大于2的集合，随机选择，添加边至边集。

每条树链中最多只有一个环，最终生成的无向连通图中的每条边最多在一个环内。算法可以构造一个高随机性的仙人掌图，但缺点是能构造出的仙人掌图边的最大数量的平均值仅有节点数的倍(见实验)。当我们想要构造一个密度较高，并且仍具有一定随机性的仙人掌图，那么就需要对上述算法进行修改。

算法：从根节点开始进行向下遍历一棵树。对于节点，设为节点所在树链的中深度最浅的节点编号，设等于以节点为根节点的子树大小。选择节点，，。对于如果，，否则。若，则，属于同一条树链。设表示值等于的节点集合。对于每个元素数量大于2的集合，随机选择，添加边至边集。

定义有效树链为长度大于1的树链。设表示以为根的子树中包含的有效树链的结束节点的最大数量。

证明：如果并且，那么。如果并且，那么

因为，所以为一条树链的开始节点。因为，所以为叶子节点。由于树链从开始节点出发到叶子节点结束，故该树链仅包含一个节点，长度为1，不是有效树链，故。

因为，所在的树链至少包含节点与，长度大于1，为有效树链，。证毕。

证明：如果，无论是否等于，不变。

如果，假设存在一个最优剖分方案，有。其中表示以作为结束节点的树链所包含的点集。设，因为子树大小大于1，节点不是叶子节点，故。因为，所以，因此表示一条有效树链。

如果，因为，所以，依然表示一条有效树链，而其他划分方案不变，故不变。证毕。

证明：如果存在一个子节点，，那么令总是不劣。

反证法，如果存在一个最优方案，其中存在两个节点，，满足，，，。设，根据证明可知。如果改变方案令，，根据证明，可知此时，，优于原方案，这与原方案是最优方案的说法相悖。证毕。

证明：算法所剖分出的树链数量不少于算法所剖分出的树链数量。

如果在一次剖分中，仅有大小为1的子树或者仅有大小大于1的子树，根据证明与证明，无论选择哪个子树根节点，令，最终树链数量不变。

如果在一次剖分中，同时包含两类子树，根据证明，选择大小只有1的子树进行剖分总是更优。综合以上两种情况，选择较小的子树根节点，令的方案总是不劣。证毕。

实验3.3.1：下面用实验对算法的优化效果进行对比验证[[2]](#footnote-2)。输入数据为一棵随机构造的树，输出数据为该算法在树上剖分的树链数量。实验数据一共有7组，每组有100个随机生成的树。分别测试在节点数为100，500，1000，5000，10000，50000，100000的情况下，两类算法剖分出的平均树链数量。

表格 2

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| 节点数量 | 100 | 500 | 1000 | 5000 | 10000 | 50000 | 100000 |
| 算法3.3.1 | 20.28 | 107.9 | 217.49 | 1088.97 | 2184.26 | 10912 | 21821.5 |
| 算法3.3.2 | 35.72 | 182.65 | 367.03 | 1835.17 | 3677.19 | 18397.4 | 36788.1 |
| 比值 | 1.514792 | 1.692771 | 1.687572 | 1.685235 | 1.683495 | 1.685979 | 1.685865 |

表格展示了算法与算法对同一组数据剖分出的树链数量对比，对于每个节点数量级均准备了100组测试数据，测试数据使用算法2随机构造。可以看到在不同节点数量级下，算法构造出的平均树链数量均较算法提高了。

图展示了在不同节点数量下，算法3.3.1和算法3.3.2进行100次随机构造与剖分后得到的随机树链数量结果的折线图。不难看出在不同节点数的情况下，算法3.3.1与3.3.2剖分出的树链数量均稳定。算法3.3.1所剖分出的数链数量大致是节点总数的20%，算法3.3.2所剖分出的数链数量大致是节点总数的35%。如果需要构造更稠密的仙人掌图，则建议使用非随机性构造算法，即由每个节点向其父节点连边构造。相关代码与数据已被开源整理到<https://github.com/gcwAndHisFriends/cactus_construct_check>。



图 2

# 4 **结论**

图论中的问题是千变万化的，如何构造图论模型的算法并没有不二法门。简单的问题可以通过常见的图论模型构造方法进行实现，但在更复杂的问题中，有时会感到难以建立适当的模型，这时，需要在不改变问题原型本身的性质的前提下，对原型进行抽象、简化，在此基础上建立合适、有效的模型。本文针对算法竞赛与数据结构课程中常见的图论模型，提出了兼顾随机性与时间效率的构造算法。这对于需要满足在规定的时间内正确的分析问题和快速构造相应的图论模型的实际的竞赛和教学过程中，借此方法解决关键问题提供了极大的帮助。

# 参考文献

1. 王桂平、杨建喜、李韧编著. 图论算法理论、实现及应用(第2版)[M], 北京大学出版社，2022年1月.
2. 杨玉军, 王大勇. 图论在数学建模中的应用[J]. 教育现代化, 2018(4).
3. 王桂平，王衍，任嘉辰. 图论算法理论、实现及应用[M]. 北京大学出版社, 2022.
4. 罗勇军, 郭卫斌. 算法竞赛[M]. 清华大学出版社, 2022.
5. 王兰亭. 基于SLIC超像素的图论图像分割方法[D].吉林大学,2021.
6. 王映龙,宋泽锋,陈卓.基于图的多关系数据挖掘理论研究与方法[J].计算机科学,2008(05):152-153+157.
7. 曾滔.改进的k度匿名图构造算法[J].计算机系统应用,2022,31(05):157-164.
8. 邓慧明,甘新标,白皓等.shiftX:基于异或快速算子图生成并行优化方法[J/OL].计算机工程与科学:1-9[2023-02-22].
9. 宋先铎. 基于层次化循环神经网络的图生成模型研究[D].吉林大学,2022.
10. 刘汝佳. 算法竞赛入门经典[M]. 清华大学出版社, 2009.
11. Cormen T H, Leiserson C E, Rivest R L，et al. 算法导论(原书第3版)[M].算法导论(原书第3版), 2012.

作者简介：

葛昌威（2001—），男，重庆丰都人，学士，研究方向为数据科学与大数据

黄春淦（1998—），男，四川广安人，硕士，研究方向为图神经网络、推荐系统

王桂平（1982—），男，广东肇庆人，硕士，讲师，研究方向为图论算法、大规模图数据分析及处理等

通信作者：王桂平（1979—），男，江西安福人，博士，副教授，硕导，研究方向为图论算法、大规模图数据分析及处理等，E-mail：wgp@cqjtu.edu.cn。

非公开发表部分：

姓名：黄春淦

**联系地址**：重庆交通大学信息科学与工程学院计算机系

**电子邮箱：**[1053538739@qq.com](mailto:1053538739@qq.com)

**联系电话：**15310970050

1. [↑](#footnote-ref-1)
2. [↑](#footnote-ref-2)