



## 第 3 章

# 网络空间的邻近搜索

网络空间邻近搜索的目的是提供网络拓扑感知，即通过测量物理网络中节点之间的网络距离，提供精确、可扩展的物理网络视图。网络距离主要通过网络延迟、网络带宽、丢包率等双向网络参数体现，还可通过节点 CPU 负载、内存容量、网络接入带宽、流量上限体现，其中网络延迟最为常用。

假设一个属于延迟敏感型应用的分布式系统中存在一组具有动态性的服务器面向 Internet 的用户提供服务。每个服务器可能由于更新、维护、失效等而动态地退出系统，同时新的服务器可能动态地加入系统中。满足这种假设的系统如典型的 BitTorrent 应用、P2P 流媒体应用、分布式代理应用等。假定每个在线的服务器向用户提供相同的服务，即将用户的请求定位到不同的服务器，而对应的服务器则根据自己的服务能力向用户提供服务。同时，假定每个服务器的服务能力根据其硬件配置、系统管理需求以及网络状况等因素，定义为一个变量。例如，服务器根据其 CPU 处理能力、出口带宽容量或者流量成本等因素设定一个最大服务数量。一个服务器在其服务数量不超过最大服务数量时，响应来自 Internet 不同位置的服务请求。

### 3.1 网络空间邻近搜索技术

为了支撑网络延迟的获取和应用优化，一个基本的方式是获取网络延迟矩阵。它描述了互联网节点间的双向网络延迟，是网络测量的重要基础。网络延迟矩阵规模巨大，动态波动，难以实时获取。目前主流的方式是通过网络坐标系统为互联网节点提供坐标位置，并通过网络坐标的距离预测网络延迟矩阵。

为了构建和维护网络延迟矩阵，采取分布式的邻近搜索，将服务节点组织为轻量级的无结构化覆盖网，每个服务节点均响应任意用户节点提出的 K 近邻搜索请求，仅利用少量的网络延迟，预测互联网节点间的邻近关系。本章提出分布式 K 近邻搜索框架 DKNNS (distributed K nearest neighbor search)。DKNNS 算法适用于网络延迟空间分簇现象，每个服务节点通过低开销的流言 (gossip) 通信和邻近节点搜索过程，快速地发现位于网络延迟

空间各个距离尺度的邻居节点集合。因此在  $K$  近邻搜索时，每个服务节点均能够利用邻居集合分布式地发现距目标节点最近的邻居。为了快速地完成  $K$  近邻搜索，在搜索时，首先，利用最远节点搜索机制选择  $K$  近邻搜索的发起节点，避免距目标过近导致的搜索过早中止问题；其次，DKNNS 利用分布式搜索结合快速回退的方式在目标节点邻近区域迅速确定  $K$  个近邻节点。

理论分析表明 DKNNS 能够以  $O(K\log\Delta)$  跳步发现  $K$  个近似最近邻居，其中  $\Delta$  为服务节点间最大延迟距离。基于模拟测试发现，DKNNS 算法能够确定近似最优的  $K$  个服务节点，在  $K=10$  时，与真实  $K$  近邻的重合率超过 80%。同时 DKNNS 引发的查询延迟和查询开销并不随着系统规模增大显著提高。基于 PlanetLab 部署的实验证实了 DKNNS 得到的近邻集合接近最优结果，超过 50% 的查询时间不到 10s，平均查询带宽开销不到 5KB，且  $K$  近邻信息具有稳定性，能够满足延迟敏感型应用的服务节点选择需求。

### 3.2 网络邻近搜索概念模型

假定  $\mathbf{V}$  为分布式系统中的服务器以及用户节点的集合。网络延迟为一个二元函数  $d: \mathbf{V} \times \mathbf{V} \rightarrow \mathbf{R}^+$ 。给定一组动态的服务节点集合  $S \in \mathbf{V}$ ，假定每个服务器  $i \in S$  能够接纳的最大访问请求数目为一个常数  $\Delta_i$ ，且在时刻  $t$  已经接纳的用户请求数为  $A_i(t)$ 。

给定一个具有动态性的服务节点集合， $K$  近邻搜索从服务节点集合中可扩展地寻找距任意用户节点网络延迟最近的  $K$  个节点集合。其中  $K$  为系统参数， $K=1$  即为最近服务节点搜索。而  $K>1$  时，搜索后返回距用户节点最近的  $K$  个服务节点，有利于上层应用设计灵活的服务节点筛选策略，从而选择综合性能最好的邻近服务节点。

与  $K$  近邻搜索相关的研究工作按服务节点选择过程可分为集中式排序和分布式搜索两类。集中式排序需要预先测量服务节点与用户节点间的邻近度，导致测量开销较大，且存在单点失效问题，故可扩展性不高。而分布式搜索避免了网络通信瓶颈，降低了测量开销。然而已有的分布式搜索算法容易陷入局部最优值<sup>[1]</sup>。

基于上述假设，在分布式系统中集中式地选择距用户网络延迟最小的服务器面临着可扩展性和通信性能瓶颈等问题，因此，希望通过分布式的方式定位距每个用户（称为一个目标节点）网络延迟最小的服务器。定义分布式最近邻搜索问题如下。

**定义 1** 分布式最近邻搜索问题定义为寻找距目标节点  $T$  网络延迟最小且能够接纳新用户访问请求的服务器  $i^*$ ，即  $A_i(t) < \Delta_i$ 。

考虑一个简化的分布式最近邻搜索问题，即服务器响应任意数目服务请求下（即  $\Delta_i = \text{inf}$ ）的最近邻搜索问题，而在服务器请求数目为有限数值情况下，搜索距目标节点的多个最近服务节点，然后从中筛选距目标节点最近且能够接纳新服务请求的服务器。定义无穷



服务能力下最近服务器搜索问题如下。

**定义 2** 假定  $V$  为分布式系统中的服务器以及用户节点的集合。网络延迟为一个二元函数  $d: V \times V \rightarrow \mathbf{R}^+$ 。给定一组动态的服务节点集合  $S \in V$ ，寻找距目标节点  $T$  网络延迟最小的服务器  $i^*$ 。

由于延迟敏感型网络应用中服务节点集合和目标节点集合具有大规模、动态性的特点，分布式实现  $K$  近邻搜索，避免网络性能瓶颈，成为一个亟待研究的问题。分布式  $K$  近邻搜索通过如下方式定义。

**定义 3(分布式  $K$  近邻搜索)** 假设一个分布式系统中存在一个动态的服务节点集合  $S$ ，给定 Internet 中任意的节点  $T$ ，基于服务节点集合分布式协作，从  $S$  中寻找距  $T$  网络延迟最小的  $K$  个节点集合  $S_1$ ， $K$  为系统参数。

由定义 1 可知，目标节点  $T$  的真实  $K$  近邻可能并不唯一。在分布式环境下，由于每个节点只有少量的邻居节点集合，难以完全收集所有服务节点距目标节点的距离信息，导致分布式  $K$  近邻搜索过程得到的结果容易陷入局部最优值。

为了量化搜索到的  $K$  个近邻距目标节点的邻近度，通过近似度来量化分布式  $K$  近邻搜索的精确度。

**定义 4( $\theta$  近似)** 假设基于分布式  $K$  近邻搜索机制得到  $K$  个近邻节点集合  $S_1$ ，目标节点  $T$  的真实  $K$  近邻节点集合为  $S_2$ ，若  $S_1$  距  $T$  的距离和与  $S_2$  距  $T$  的距离和比值不大于  $\theta$ ，即  $\frac{\sum_{i \in S_1} d_{iT}}{\sum_{i \in S_2} d_{iT}} \leq \theta$ ，则称该分布式  $K$  近邻搜索结果为  $\theta$  近似， $\theta \geq 1$ 。

$\theta=1$  时，搜索到的  $K$  个近邻  $S_1$  就是目标节点  $T$  的一组真实  $K$  近邻。利用反证法证明若将  $S_1$  中每个节点按距目标节点  $T$  的距离升序排序，则序列中第  $i$  个节点距目标节点的距离与第  $i$  个真实近邻距目标节点的距离相等。否则，得到一个距目标节点距离更小的  $K$  近邻序列，这与  $S_2$  为真实的  $K$  近邻前提矛盾。

已有的分布式搜索方法大多基于启发式搜索过程：每个节点通过利用局部信息维护的一组逻辑邻居，在一个节点  $P_i$  接收到一个目标节点为  $T$  的最近服务器搜索请求后，选择逻辑邻居中距目标节点距离比当前节点还近的节点  $P_{itl}$  作为新的搜索节点，并将搜索消息发送给节点  $P_{itl}$ ，最后节点  $P_{itl}$  递归地执行最近服务器搜索过程。上述搜索过程如图 3.1 所示。最近服务器搜索模块类似于 DNS 服务查询部署于 Internet 中，每个用户  $T$  在访问网络服务前通过最近服务器搜索模块定位距本节点最近的服务器。

然而已有的最近服务器搜索研究受到精确度、可扩展性等限制，仍然亟待改进。进一步地将最近服务器搜索的关键需求归纳如下：①精确，搜索到的服务器距目标节点延迟需要足够小，以提高用户的服务访问质量；②快速，分布式搜索过程的完成时间要足够小，以满足用户实时查询的需求；③可扩展，随着系统的增大，搜索过程的探测带宽和逻辑邻居维护带宽开销需要适中；④动态适应性，在服务器动态地加入或者退出时，搜索精确度下

降幅度必须保持平滑。

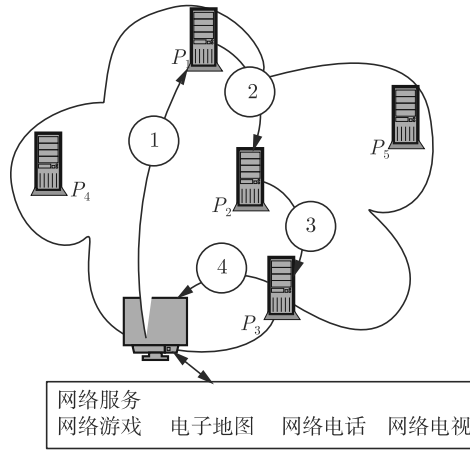


图 3.1 最近服务器搜索过程

选定集合  $V$  中的一个节点  $P$  作为球心，一个以  $r$  为半径的闭球包含的节点集合表示为  $B_P(r)$ ，即  $B_P(r) = \{v | d(P, v) \leq r, P, v \in V\}$ 。一个球的容积 (volume) 定义为球中包含的节点的个数。定义集合的覆盖如下。

**定义 5** 设定  $S$  和  $\Omega$  为两个节点集合，如果  $\Omega \subseteq S$ ，称集合  $S$  覆盖集合  $\Omega$ 。

低度量模型 (inframetric model) 定义如下。

**定义 6** 称一个距离函数  $d: V \times V \rightarrow \mathbf{R}^+$  为一个  $\rho$  低度量模型 ( $\rho > 1$ )，当且仅当针对任意的节点对  $P_1, P_2$ : ①自反性，若  $d(P_1, P_2) = 0$ ，则  $P_1 = P_2$ ; ②对称性， $d(P_1, P_2) = d(P_2, P_1)$ ; ③低度量性，对于任意的节点  $P_3$ ，其中  $P_3 \notin \{P_1, P_2\}$ ，满足  $d(P_1, P_2) \leq \rho \max \{d(P_1, P_3), d(P_3, P_2)\}$ 。

低度量模型具有两个显著特点：

(1) TIV-自适应。由上述条件③可知，随着参数  $\rho$  减小，任意三元组的三个边越来越接近；随着参数  $\rho$  增大，三元组中存在一个边远大于任意的另外两个边，即可能产生三角不等性。

(2) 动态性。为了描述网络延迟的变化，低度量模型通过三元组参数  $\rho$  进行调整。

增量维度<sup>[2]</sup> 定义为相同球心不同半径的两个球包含的节点数目对比。

**定义 7** 在一个  $\rho$  低度量模型中选择任意的半径  $r \in \mathbf{R}^+$ 、 $\gamma_g \in \mathbf{R}^+$  以及球心  $P \in V$ ，如果  $|B_P(\rho r)| \leq \gamma_g |B_P(r)|$ ，那么称  $\rho$  低度量模型的增量维度  $\gamma_g \geq 1$ 。

增量维度越低，说明球  $B_P(\rho r)$  覆盖的节点数目与  $B_P(r)$  覆盖的节点数目越类似。假定节点间的距离满足三角不等性，D.R.Karger 和 M.Ruhl<sup>[3]</sup> 指出，一方面，若沿着节点  $P$  增大球的半径，那么球覆盖的节点数目按照匀速的方式增大；另外，任意的节点  $P$  通过均匀随机采样的方式发现距任意目标节点更近的节点。然而上述结论并不直接适用于三角不等



性违例较为显著的网络延迟空间。

### 3.3 分布式邻近搜索机制

针对网络延迟空间的三角不等性违例，以及分簇现象，为了支持灵活的  $K$  近邻搜索服务，提出了分布式  $K$  近邻搜索算法 DKNNs。DKNNs 算法不依赖三角不等性假设，而是基于低度量模型<sup>[2]</sup>，构建反映节点间分簇特征的邻居列表-多级环，然后，递归地搜索多级环上距目标节点最近的服务节点，查询 1 个近邻，然后通过快速的搜索回退过程，查询剩余的  $(K - 1)$  个近邻，从而实现面向任意目标节点的  $K$  个最近服务节点搜索服务。

在 DKNNs 中，每个服务节点利用一个多级环缓存少量的逻辑邻居集合。多级环由按照距当前服务节点的延迟指数级增大的距离范围划分为多个环。每个环包含少量的逻辑邻居。为了维护多级环，每个服务节点通过定期发出 gossip 消息以及  $K$  近邻搜索消息发现新的服务节点集合，并按照距当前服务节点的延迟，插入多级环的对应环。gossip 消息在逻辑邻居之间进行转发，用于在全局范围均匀随机地选择服务节点，而  $K$  近邻搜索消息用于发现邻近服务节点，避免了均匀随机选择过程容易忽略占有所有服务节点比例较低的邻近服务节点的不足。

针对最近邻搜索需求，假定一个节点  $P$  接收到一个最近邻居搜索请求。节点  $P$  从同心环中选择距目标节点邻近的候选节点集合，然后节点  $P$  从候选邻居中集合延迟预测以及直接探测方式选择距目标节点最近的逻辑邻居。由于目标节点可能是 Internet 的任意一个节点，导致其并没有一个坐标向量。为了预测逻辑邻居到目标节点的延迟，为目标节点分布式地计算一个坐标位置。在确定当前距目标节点最近的节点（记为  $P_*$ ）后，节点  $P$  判断是否终止最近邻居搜索过程：若  $P_* \neq P$ ，那么节点  $P$  将搜索请求转发给节点  $P_*$ ；否则，节点  $P$  终止搜索过程，节点  $P$  向目标节点  $T$  提供服务。

图3.2所示为  $K$  近邻搜索流程示意图。针对  $K$  近邻搜索需求的整体流程如下。首先，响应搜索消息的服务节点确定目标节点的坐标位置，利用逻辑邻居与目标节点的坐标距离并结合直接探测，发现多级环中距目标节点坐标距离最近或最远的少量候选节点集合，然后利用直接探测的方式从候选节点中选择真正最近或者最远的逻辑邻居；其次，当前服务节点利用最近节点搜索过程，通过递归地选择多级环中距目标节点最近的逻辑邻居，选择全局范围距目标节点延迟最大的服务节点执行  $K$  近邻搜索操作，避免因发起  $K$  近邻搜索的服务节点距目标节点过近导致的搜索过程过早终止问题；再次，在  $K$  近邻搜索时，发起搜索节点利用贪心搜索的方式，递归地选择多级环中距目标节点最近的逻辑邻居，分布式地发现距目标节点最近的服务节点；最后，通过回退以及随机重启的方式，在目标节点邻近区域快速地确定剩余的  $(K - 1)$  个最近服务节点。图3.3所示为  $K$  近邻搜索示意图。

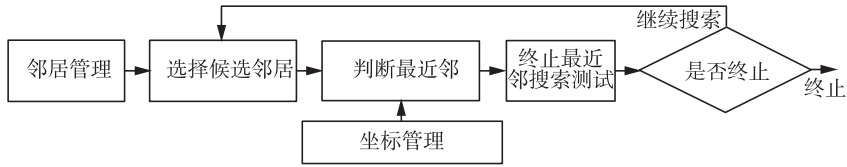


图 3.2 K 近邻搜索流程示意图

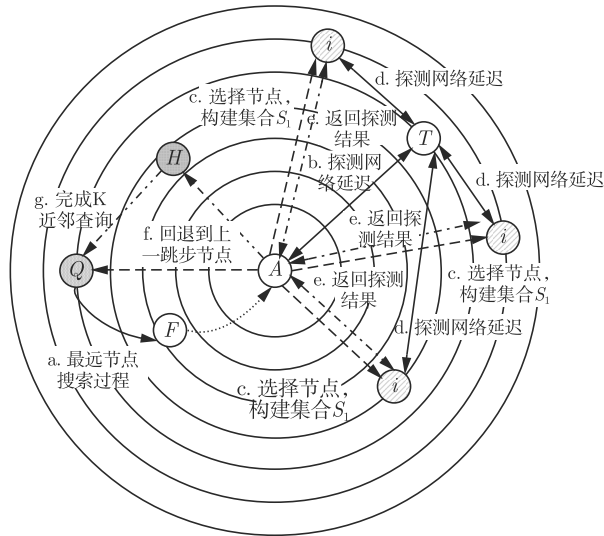


图 3.3 K 近邻搜索示意图

尽管坐标距离具有近似精确度<sup>[4]</sup>, 单纯利用坐标距离并不能发现真实的最近邻, 但是 DKNNS 中采取坐标预测选择少量的候选节点, 然后利用候选节点直接探测目标节点的延迟测量结果, 选择最近的候选节点, 能够以较低的探测开销精确地发现多级环中距目标节点延迟最小或者最大的逻辑邻居。

### 3.4 网络空间邻近搜索关键算法

#### 3.4.1 网络探测优化

为了快速发现距目标节点的邻近节点集合, 降低搜索过程延迟, 基于坐标距离预测的方式进行选择。为了得到准确的目标节点坐标会 K 近邻搜索的最初请求节点  $P$  初始化目标节点坐标位置, 然后每个 K 近邻搜索的中间节点更新目标节点坐标位置。K 近邻搜索的发起节点  $P$  首先向目标节点请求坐标位置, 如果目标节点维护了坐标位置, 则通过请求得



到目标节点的初始坐标位置, 否则, 发起节点  $P$  随机选择  $L$  个邻居节点探测距目标节点的延迟, 然后利用该邻居节点集合的坐标位置信息, 基于改进的 Vivaldi 算法<sup>[5]</sup> 迭代地计算目标节点的初始化坐标, 迭代轮数设定为 15。每个  $K$  近邻搜索的中间查询节点  $I$  在接收到  $K$  近邻搜索消息  $M$  后, 从  $M$  中提取目标节点的坐标位置以及误差值。然后中间查询节点  $I$  探测距目标节点的延迟, 并基于本节点的坐标位置, 用改进的 Vivaldi 算法<sup>[5]</sup> 更新目标节点的坐标位置。

由 3.3 节的搜索流程可知, 每个节点至少需要选择  $3\left(\frac{\rho^2}{\beta}\right)^\alpha$  个候选邻居。参数  $\alpha$  和  $\rho$  取决于网络延迟空间, 无法对其进行修改。而  $\beta$  是一个自由变量。增大  $\beta$  可降低候选邻居数目, 并且能够降低发现的最近服务器距目标节点的延迟。因此, 将  $\beta$  设置为其上限值 1。

由于网络坐标用于延迟预测, 因此避免直接延迟探测导致的带宽开销。利用延迟预测进一步降低测量开销。基于改进的 Vivaldi 算法<sup>[5]</sup> 预测延迟 (记为 TIV-Vivaldi)。该算法迭代地更新坐标位置, 具有较高的可扩展性和快速的收敛速度。然而网络坐标并不与真实的延迟值完全匹配, 错误的预测延迟可能影响搜索过程的精确度。因此单独利用网络坐标进行最近服务器选择并不可靠。所以, 检测网络坐标的精确度, 在错误预测结果时同时进行直接延迟探测以选择距目标最近的候选节点。

为了降低延迟测量开销, 为目标节点计算一个坐标向量, 从而通过延迟预测的方式计算候选邻居与目标节点的网络延迟。首先, 若一个节点  $P$  接收到一个最近邻居搜索请求, 节点  $P$  判断目标节点的坐标位置是否被初始化: 如果搜索请求消息没有目标节点的坐标位置, 那么节点  $P$  初始化目标节点  $T$  的坐标位置; 否则, 节点  $P$  利用本节点到目标节点  $T$  的延迟更新目标节点  $T$  的坐标位置, 以提高目标节点  $T$  坐标的精确度。

在初始化坐标向量时, 节点  $P$  随机选择同心环上一组逻辑邻居, 让这些逻辑邻居探测距目标节点的延迟, 并反馈给节点  $P$ 。节点  $P$  根据逻辑邻居的坐标位置, 为目标节点初始化一个坐标位置。节点  $P$  将目标节点  $T$  的坐标位置和坐标误差存储到搜索请求消息, 然后发送给下一跳步搜索节点。

### 3.4.2 最远节点初始化

若随机选择一个发起节点执行  $K$  近邻搜索操作, 在  $K$  值较大时, 可能因为发起节点距目标节点过近导致无法继续搜索, 而仅能搜索到少数目的近邻。因此设计一个  $K$  近邻初始化节点选择过程, 选择距目标节点最远的节点发起  $K$  近邻搜索过程。

为了避免因为搜索发起节点距目标节点较近导致无法搜索足够数目的近邻, 首先发起一个  $K$  近邻初始化节点选择过程, 选择距目标节点最远的节点发起  $K$  近邻搜索过程。由于一个包含很少节点信息的逻辑邻居难以提供距目标节点的邻近信息, 因此需要被排除在  $K$  近邻搜索过程之外。 $K$  近邻搜索的最初请求节点  $P$  首先选择非空环数目不小于  $R$  的候选节

点子集, 然后从中选择距目标节点坐标距离最远的  $L$  个节点直接探测距目标节点的 RTT 延迟, 判断返回的 RTT 值是否存在大于本节点距目标节点距离阈值  $\beta_f$  ( $\beta_f$  默认为 1.2) 倍的节点 (设为  $P_2$ ), 若存在, 则将最远节点搜索请求发给  $P_2$ ,  $P_2$  继续递归的搜索; 否则, 当前节点  $P$  停止最远发起节点搜索过程, 并选择当前距目标节点最远的节点 (设为  $P_F$ ) 作为  $K$  近邻搜索的发起节点。

### 3.4.3 最近邻搜索

由于搜索的每一跳步都需要选择位于特定区域内足够数目的随机采样节点作为搜索候选节点, 每个节点都需要主动地维护足够数目的逻辑邻居以方便候选节点选择。由于目标节点可能位于网络延迟空间的任意位置, 因此, 每个节点的逻辑邻居集合都需要覆盖网络延迟空间的典型区域。

已有的最近服务器选择方法 (如 Meridian 和 OASIS 算法) 主要通过 gossip 的方式发现逻辑邻居, 并将逻辑邻居组织为同心环结构以方便从球形空间随机采样搜索候选邻居。同心环结构包括多个相同球心不同半径的环。每个环上包含常数数目的逻辑邻居。各个环对应的半径大小呈指数级增长, 以便于采样网络延迟空间距一个节点邻近的区域。然而, 通过实验发现在同心环上内部和外部的环包含的逻辑邻居数目远小于中间部分环包含的逻辑邻居数目, 并且中间部分的环由于逻辑邻居数目过多, 经常触发逻辑邻居替换事件, 增大了同心环维护开销。

首先基于低度量模型构建最近服务器搜索理论。假设任意的节点  $P$  在接收到目标节点为  $T$  的最近搜索消息后, 节点  $P$  在球  $B_P(\rho r)$  包含的节点集合内有替换地均匀随机选择  $3\left(\frac{\rho^2}{\beta}\right)^\alpha$  个节点时, 那么在  $\log_{\frac{1}{\beta}} \Delta_d$  搜索跳步后, 搜索过程终止, 并且发现最近邻距目标节点延迟不大于真实最近服务器距目标节点延迟的  $1/\beta$  倍。其中  $r$  为从节点  $P$  到节点  $T$  的延迟,  $\beta$  为搜索的延迟降低幅度 ( $0 < \beta \leq 1$ ),  $\Delta_d$  为网络延迟空间中节点间的最大延迟与最小延迟的比值, 而  $\alpha$  定义为  $\alpha \in [\log_{\rho} \gamma_g, \leq 2\log_{\rho} \gamma_g]$ 。

接着提出了一个实用的最近服务器搜索方法, 称为 HybridNN。上述最近邻搜索理论结果由于需要严格的随机采样保证而影响了其实用性。HybridNN 在保持理论成立的随机采样前提下, 提高其动态适应性以及测量过程的可扩展性。

基于贪心搜索的思想进行分布式的迭代搜索: 当前节点根据多级环上逻辑邻居距目标节点的延迟信息, 每次选择的下一跳步节点距目标节点需要比当前节点近  $\beta$  倍, 从而以指数级速度逼近目标节点; 在搜索到 1 个近邻后, 由于两个相同发起节点的贪心搜索过程中间经历的节点集合序列相同, 仅区别于最后跳步的节点, 故在目标节点的邻近区域继续发现剩余的近邻节点。

为了动态选择搜索过程的候选节点, HybridNN 利用改进的同心环维护一组在线的逻



辑邻居集合，从而能够及时地筛选出合适的最近邻搜索候选节点。Meridian、OASIS 等工作已经提出同心环的概念。HybridNN 的改进包括扩大同心环的覆盖范围以提高搜索过程的可持续搜索能力以及降低测量开销以提高搜索过程的可扩展性：①通过理论分析中的随机采样数目下限确定同心环中每个环上的逻辑邻居数目，使得 HybridNN 能够以最低数目的逻辑邻居发现距目标节点更近的服务器；②通过带偏好的邻居采样策略，为每个环尽可能多地选择逻辑邻居，以方便定位距不同目标节点更近的服务器；③HybridNN 通过网络坐标进行延迟预测以避免邻居维护过程的探测开销，由于网络坐标的近似精确度，邻居维护结果仍然保持了最大化环上节点覆盖范围的物理意义。

为了提高分布式搜索过程的精确度和可扩展性，采取多种措施在降低测量开销的同时保证搜索过程的健壮性：

(1) HybridNN 通过改进的 Vivaldi 网络坐标方法<sup>[5]</sup> 预测服务器之间以及服务器与目标节点间的延迟，从而避免大量的直接延迟探测开销，并且 HybridNN 基于被动的方式更新坐标位置，避免了主动方式坐标维护过程中的测量带宽开销。

(2) 由于网络坐标的预测延迟并不与真实延迟完全匹配，HybridNN 实时地检测网络坐标的误差率，在延迟预测误差率较高时，HybridNN 利用少量的直接网络延迟探测避免错误率较高的延迟预测。

(3) HybridNN 将  $\beta$  设定为 1，以降低直接探测开销并提高搜索过程的健壮性，已知每个节点需要采样  $3\left(\frac{\rho^2}{\beta}\right)^\alpha$  数目的逻辑邻居作为候选邻居；可通过增大  $\beta$  降低候选邻居数目。由于网络延迟分簇可能造成候选邻居距目标节点延迟类似，进而导致搜索过程过早终止，因此，增大  $\beta$  降低了新的搜索节点的筛选要求，有利于避免因网络延迟分簇造成的搜索过早终止问题，提高搜索过程的健壮性。设定  $\beta$  为 1 保证了 HybridNN 选择最小数目的逻辑邻居作为最近邻搜索过程的候选节点，同时，允许以没有任何延迟降低的步骤进行搜索，有利于适应当前节点与一些候选节点距目标节点延迟相似的现象。

假设当前节点  $P_F$  距目标节点的坐标距离为  $d_T$ ，当前节点从距本节点坐标距离范围在  $[0, \rho \times d_T]$  ( $\rho$  为低度量模型参数，默认为 2.5) 的逻辑邻居集合中。选择不在已经搜索近邻节点集合且非空环数目不小于  $R$  的候选节点子集，然后选择  $L$  个最靠近目标节点的候选节点。实验发现  $R$  为 4 且  $L$  为 4 时精确度较高。接着，每个候选节点主动地探测距目标节点的距离，节点  $P_F$  根据直接探测距离从候选节点中选择距目标节点最近的节点 (设为  $P_C$ )。

首先，如果  $P_C$  与目标节点的距离小于当前节点与目标节点距离乘以参数  $\beta$  ( $\beta$  小于 1，称为下一跳步选择阈值)，则当前节点将  $K$  近邻搜索消息发送给  $P_C$ ， $P_C$  继续进行  $K$  近邻搜索，当前节点成为  $P_C$  的  $K$  近邻搜索父节点；否则，当前节点选择距目标节点最近的邻居 ( $P_C$  或者  $P_F$ ) 为距目标节点最近的一个近邻，并将该节点缓存到  $K$  近邻搜索消息。 $K$

近邻搜索消息中缓存的每个近邻节点标记为已经搜索近邻节点，已搜索过的节点不再参与当前的  $K$  近邻搜索过程。

从候选邻居中筛选距目标节点最邻近的节点子集，从而降低直接探测开销。首先从候选邻居中选择距目标节点坐标距离最近的  $m$  个邻居，记为  $S_c$ 。其次，选择候选邻居中坐标误差超过误差阈值的邻居，坐标误差阈值默认为 0.7。此外，还选择坐标距离与真实的网络延迟值偏差超过 50ms 的节点，以适应由于三角不等性违例造成的坐标误差。图3.4综合了上述选择规则，记上述选择过程得到的逻辑邻居集合为  $S_* = S_c \cup S_e \cup S_t$ 。

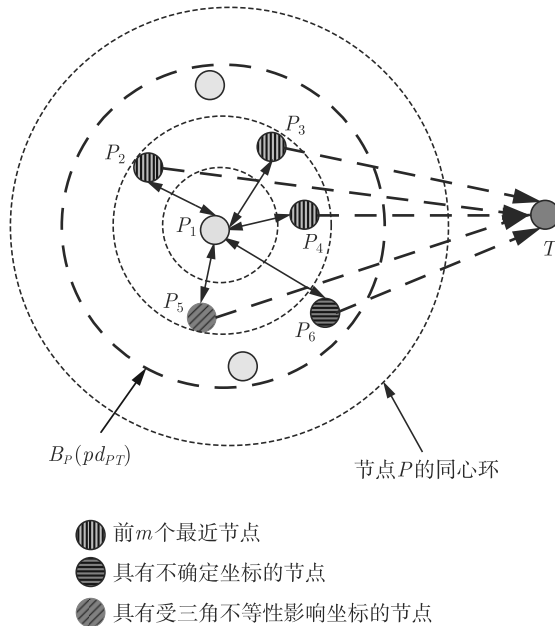


图 3.4 筛选候选逻辑邻居

节点  $P$  接着请求集合  $S_*$  内的节点探测距目标节点  $T$  的网络延迟，然后节点  $P$  从中选择距目标节点最近的逻辑邻居。若多个逻辑邻居距目标节点的延迟相同，那么随机地从中选择一个节点。

### 3.4.4 搜索终止判定机制

如前文所述，HybridNN 设定搜索阈值  $\beta$  为 1。因此，若 3.4.3 节筛选的逻辑邻居比当前节点  $P$  距目标节点更远，那么终止搜索过程，并将节点  $P$  作为向目标节点  $T$  提供服务的节点。



### 3.4.5 逻辑邻居维护机制

#### 1. 同心环

由于 3.4.3 节搜索流程要求在球形区域内随机采样候选节点，因此采取同心环存储逻辑邻居。同心环上第  $i$  个环包含距当前节点延迟值位于  $(2^{i-1}, 2^i]$  区间的节点。假定需要选择半径为  $d_2$  的球形区域内的节点作为候选节点，为不遗漏候选节点，需要选择同心环上第 1 个到第  $\lceil \log_2 d_2 \rceil$  个环包含的逻辑邻居。

同心环的一个重要参数是环上逻辑邻居数目  $\Delta$ 。为了以至少 95% 的概率定位距目标节点更近的一个候选邻居，需要选择  $3\left(\frac{\rho^2}{\beta}\right)^\alpha$  个逻辑邻居。在  $\beta$  为 1 时，逻辑邻居数目位于区间  $[3\gamma_g^2, 3\gamma_g^4]$ 。由于增量维度  $\gamma_g$  较为适中<sup>[3]</sup>，一般低于 5，因此，设置环大小  $\Delta$  为较小的整数值，默认值为 8。

#### 2. 邻居发现机制

由于需要为同心环的各个环选择逻辑邻居，因此提出基于 gossip 方式和随机行走方式的邻居发现机制。gossip 方式中，每个节点定期地与其逻辑邻居交换邻居集合，实现均匀随机的逻辑邻居采样过程。而随机行走方式中，每个节点定期地发起随机行走消息，该随机行走消息按照逐个跳步的方式进行随机转发。每个中间转发节点筛选距发起节点最近或者最远的  $K$  个节点。随机行走消息缓存的节点作为发起节点更新其同心环的内部环和外部环的候选邻居。综上，gossip 方式和随机行走方式具有互补性，并且二者的通信开销以及计算开销均较低，适合作为频繁执行的通信过程。

此外，每个节点基于邻居发现过程中的延迟探测结果维护其坐标位置。因此坐标维护过程是被动的，不引发额外的延迟探测开销。

##### (1) 基于 gossip 的邻居发现。

为了实现 gossip 方式的邻居发现，每个节点定期地与其逻辑邻居交换邻居信息。节点  $P$  从同心环上随机选择一个邻居  $Q$  作为交换 gossip 消息的节点，然后节点  $P$  从同心环的每个环上的逻辑邻居中随机地选择一个邻居，并将选择的邻居的通信地址存储到一个 gossip 请求消息中。节点  $P$  将该 gossip 请求消息发送至节点  $Q$ 。在节点  $Q$  接收到该请求消息后，立即发送一个 gossip 响应消息至节点  $P$ 。节点  $Q$  针对 gossip 请求消息中存储的每个节点进行直接延迟探测，然后根据成功的延迟探测结果将对应的节点插入其同心环上。

##### (2) 基于随机行走的邻居发现。

为了向同心环的内部和外部的环上插入逻辑邻居，每个节点（称为 1 个发起节点）通过随机行走的消息转发机制采样距节点较近或者较远的节点集合。每个节点发送具有一定生存周期的随机行走消息至随机选择的逻辑邻居。在消息转发过程中，每个中间节点将距发起节点最近或者最远的  $K$  个节点存入该随机行走消息。最终，该随机行走消息反馈给

随机行走的发起节点。随机行走消息缓存的  $2K$  个节点作为发起节点，更新其同心环的内部环和外部环的候选邻居。随机行走过程的参数  $K$  取决于系统需求。实验发现适中的  $K$  值即可满足精确的最近邻搜索需求（默认值为 10）。

利用一个实例介绍随机行走的基本过程。假设节点  $P$  需要利用随机行走过程选择邻近或者最远的逻辑邻居。节点  $P$  首先从同心环上随机选择一个邻居  $Q$  作为通信节点，然后节点  $P$  将本节点的坐标位置以及随机行走跳步数（记为 hop）存储到一个随机行走消息并发送给  $Q$ 。节点  $Q$  接收到该随机行走消息后，从本节点的同心环上选择距节点  $P$  坐标距离最近和最远的  $K$  个邻居，将这些邻居的地址以及坐标位置存储到该随机行走消息，并将消息的 hop 值减 1。然后，节点  $Q$  将该消息发送至一个随机选择的逻辑邻居  $Q_2$ 。类似节点  $Q$ ，节点  $Q_2$  判断 hop 是否为零，若为零，则将该随机行走消息立即转发给节点  $P$ ；否则，节点  $Q_2$  从其同心环上选择距  $P$  最近或最远的  $K$  个邻居，然后将这些节点与随机行走消息中包含的  $2K$  个节点进行对比，从中选择距节点  $P$  最近和最远的  $K$  个节点，并利用计算后的  $2K$  个节点替换随机行走消息中缓存的节点。节点  $Q_2$  将消息的 hop 值减 1，并随机地选择一个新的节点递归地进行转发。最终，节点  $P$  收到随机行走消息，选择消息缓存的  $2K$  个节点作为候选逻辑邻居，针对每个候选邻居进行延迟探测，并根据成功的延迟探测结果将对应的候选邻居插入逻辑环中。

### (3) 邻居替换机制。

为了控制邻居集合的存储开销，在一个环上的邻居数目超过最大值  $\Delta$  后，将多余的节点删除。节点删除过程需要最大化邻居集合的多样性，以发现与目标节点邻近的逻辑邻居。利用最大化超体积算法<sup>[6]</sup>可保留环上节点间距离最大的  $\Delta$  个逻辑邻居。

然而最大化超体积算法需要环上节点间延迟探测作为算法输入，会引发平方量级的探测开销。为了避免探测开销，可利用环上邻居的坐标位置预测邻居间的延迟距离，从而提高邻居替换机制的可扩展性。

## 3.4.6 回退策略

若已经得到  $K$  个近邻，则  $K$  近邻搜索过程结束；若当前节点  $P_F$  没有父节点，则重启  $K$  近邻搜索过程，否则，进行回退搜索。当前节点将  $K$  近邻搜索消息发送至上一跳步父节点，父节点从候选节点集合中排除当前已搜索节点，进行继续搜索。

在重启  $K$  近邻搜索过程时，当前节点从多级环上选择距目标节点最远且不在已搜索过节点集合的逻辑邻居  $P_N$ ，然后将  $K$  近邻搜索请求发送给  $P_N$ ， $P_N$  从候选节点集合中排除当前已搜索节点，进行最近邻搜索。



### 3.5 分布式邻近搜索理论分析

利用  $d_{uv}$  代表  $u$  与  $v$  之间的延迟。  $B_u(r)$  为以节点  $u$  为圆心,  $r$  为半径包含节点构成的闭球, 定义  $B_{ui} = B_u(2^i)$ 。每个节点  $u$  维护  $\log(\Delta)$  个环, 其中  $\Delta$  为最大距离, 每个环  $S_{ui} \subset B_{ui} \setminus B_{(u,i-1)}$ , 且环的大小为  $m$ 。

**定义 8(良好)** 如果  $S_{ui}$  为  $B_{ui} \setminus B_{(u,i-1)}$  的  $m$  个随机节点构成的节点子集, 则环  $S_{ui}$  为良好 (well-formed)。

**定义 9 ( $\epsilon$ -nice)** 如果对于任意的节点对  $u$  和  $v$ , 节点  $u$  的多级环具有一个邻居节点  $w \in S_{ui}$ , 使得  $d_{vw} \leq \epsilon d_{uv}$ ,  $d_{uv}(1 + \epsilon) \leq 2^i$ , 则多级环为  $\epsilon$ -nice。

可利用网格维度表达不同半径球之间的大小关系。给定一个  $n$  维网格, 任意的球覆盖的节点数目至多小于同样球心及  $x$  倍半径的球覆盖节点数目的  $x^\alpha$  倍。一个网格维度 (grid dimension) 为上述属性成立的最小的  $\alpha$ 。同时列出两个网格维度的关键性质。

**引理 1** 如果一个  $\rho$ -低度量模型具有增量 (growth)  $\gamma_g \geq 1$ , 对于任意的  $x \geq \rho$ , 任意的球覆盖的节点数目至多小于同样球心及  $x$  倍半径的球覆盖节点数目的  $x^\alpha$  倍, 其中  $\rho$ -低度量模型的网格维度为  $\log_\rho \gamma_g \leq \alpha \leq 2 \log_\rho \gamma_g$ 。

**引理 2** 设定一个  $\rho$ -低度量模型为  $\gamma_d$ -倍增的, 对于任意的  $x \geq \rho$ , 任意的球覆盖的节点数目至多小于同样球心及  $x$  倍半径的球覆盖节点数目的  $x^\alpha$  倍, 其中  $\rho$ -低度量模型的网格维度的可行区间为  $\frac{1}{4} \log_\rho \gamma_d \leq \alpha \leq \log_\rho N$ ,  $N$  为节点总数。

**证明:** (1) 假设  $\rho$ -低度量模型网格维度为未知数  $\alpha$ , 对于任意的  $u$  和  $v$ ,  $v \in B_u(\rho r)$ , 其中  $r > 0$ 。对于任意满足  $x \in B_u(\rho^2 r)$  的节点  $x$ , 根据低度量模型定义,  $x$  距  $v$  的距离  $d_{vx}$  满足

$$d_{vx} \leq \rho \max \{d_{uv}, \rho^2 r\} = \rho^3 r$$

因此  $x \in B_v(\rho^3 r)$ , 故

$$B_u(\rho^2 r) \subseteq B_v(\rho^3 r)$$

根据网格维度的定义可知:

$$|B_u(\rho^2 r)| \leq |B_v(\rho^3 r)| \leq (\rho^3)^\alpha |B_v(r)| \leq (\rho^4)^\alpha |B_v(r/\rho)|$$

(2) 采取贪心的方式利用半径为  $r$  的球构建球  $B_u(\rho r)$  的覆盖  $F$ 。设  $m \leftarrow 0$ , 若球  $B_u(\rho r)$  存在没有被包含的节点子集, 即

$$B_u(\rho r) \setminus \bigcup_{1 \leq i \leq m} B_{v_i}(r) \neq \text{NULL}$$

则任意选择一个新的球心节点  $v_{m+1}$ , 满足

$$v_{m+1} \in B_u(\rho r) \setminus \bigcup_{1 \leq i \leq m} B_{v_i}(r)$$

然后更新  $m \leftarrow m + 1$ , 递归执行上述过程。

对于上述基于贪心方式构建的一个覆盖  $F: B_u(\rho r) = \bigcup_{1 \leq i \leq \gamma_d} B_{v_i}(r)$ , 标记  $C$  为  $F$  中所有球  $B_{v_i}(r)$  的球心组成的集合。根据球心的选择方式可知, 球心集合  $C$  中任意节点  $v_i$  位于球  $B_u(\rho r)$  内部, 即  $d_{v_i u} \leq \rho r$ , 并且任意两个球心  $v_i$ 、 $v_j$  之间的距离  $d_{v_i, v_j}$  大于  $r$ 。

基于反证法证明以上述集合  $C$  为球心, 以  $r/\rho$  为半径的球两两不相交。任意选择两个球心  $v_i \in C$ ,  $v_j \in C$ , 假设存在节点  $w$  满足

$$w \in B_{v_i}(r/\rho) \cap B_{v_j}(r/\rho)$$

则根据低度量模型定义可知

$$d_{v_i, v_j} \leq \rho \max \{d(v_i, w), d(v_j, w)\} \leq r$$

而这与已知的  $d_{v_i, v_j} > r$  矛盾, 故以  $C$  中节点为球心,  $r/\rho$  为半径的任意两个球  $B_{v_i}(r/\rho)$  与  $B_{v_j}(r/\rho)$  是互不相交的。

(3) 综合 (1) 与 (2) 可知, 以球心集合  $C$  中节点为球心,  $r/\rho$  为半径的球的并集满足

$$|\bigcup_{1 \leq i \leq \gamma_d} B_{v_i}(r/\rho)| \geq \gamma_d |B_u(\rho^2 r)| / (\rho^4)^\alpha$$

由 (2) 贪心覆盖构造过程可知, 每个球心  $v_i$  都位于  $B_u(\rho r)$  内部, 即  $v_i \in B_u(\rho r)$ , 根据低度量模型定义可知

$$B_{v_i}(r/\rho) \subseteq B_u(\rho^2 r)$$

因此

$$(\rho^4)^\alpha \geq \gamma_d$$

即

$$\alpha \geq \frac{\log \gamma_d}{4 \log \rho} = \frac{1}{4} \log_\rho \gamma_d$$

故  $\rho$ -低度量模型为  $\gamma_d$ -倍增时, 网络维度  $\alpha \geq \frac{1}{4} \log_\rho \gamma_d$ 。

(4) 由于  $|x^\alpha| \leq N$ , 因此在  $x \geq \rho$  时,  $\alpha \leq \log_\rho(N)$ ,  $N$  为系统中节点总数。综上所述, 引理 2 得证。

本节逻辑邻居维护质量利用了引理 3。引理 3 基于网络维度进行定义, 通过综合 B.Wong 等<sup>[6]</sup> 推导的定理可得。



**引理 3** 给定一个节点集合，由节点间距离组成的距离空间具有网格维度  $\alpha$ ，设  $\delta \in (0, 1)$ ， $\varepsilon < 1$ ， $\gamma$  为常数， $N$  为节点集合规模，每个环的大小为  $O(1/\varepsilon)\log(N/\delta)$ ，任意选择两个节点  $u, v$ ，设  $r = \varepsilon d_{uv}$ ，选择最小的  $i$  使得  $d_{uv} + r \leq 2^i$ ，如果存在  $|B_{ui}| \leq \gamma^\alpha |B_v(r)|$ ，那么按照切尔诺夫界 (Chernoff bound)，从环  $S_{ul}(l \leq i)$  选择一个点，落在  $B_v(r)$  外面的失效概率小于  $\delta/N^2$ 。

引理3给出了网格维度下多级环满足  $\varepsilon$ -nice 的一个充分条件：在网格维度下对于系统中任意两个节点  $u, v$ ，若以  $u, v$  为球心，半径分别为  $2^i$  和  $r$  的两个球包含节点的个数比例不大于一个常数  $\gamma^\alpha$ ，那么，从节点  $u$  的多级环上选择一个邻居满足与  $v$  的距离降低至不大于  $\varepsilon d_{uv}(\varepsilon < 1)$  的成功概率接近 1，即多级环为  $\varepsilon$ -nice 的。

### 3.5.1 最近节点搜索

为了便于分析，根据 3.4 节分布式最近服务器选择定义，将分布式搜索过程的每一步定义如下。

**定义 10** 在最近节点搜索的每一步，一个节点  $P_1$  需要选择另外一个节点  $P_2$ ，使得  $P_2$  与目标节点  $T$  的距离不大于本节点与该目标节点距离的  $\beta$  倍，即  $d_{P_2 T} \leq \beta \times d_{P_1 T}$ 。同时在搜索过程的每一步，只从当前节点集合均匀随机地选择一组采样节点。

首先分析搜索过程的随机采样数目需求，然后分析整个搜索过程的精确度、搜索时间以及开销等。不失一般性，假定要保证至少以 95% 的成功率定位到满足定义10需求的节点  $P_2$ ，分析所需的随机采样数目。而随着概率的增大，所需的随机采样数目将会增大。

首先计算低度量模型下相同球心不同半径的球所包含的节点数目的比例。

**引理 4** 假定一个  $\rho$ -低度量模型的增量维度为  $\gamma_g \geq 1$ ，给定任意的正实数  $x \geq \rho$ ， $r > 0$  以及节点  $P$ ，球  $B_P(r)$  的容积不小于球  $B_P(xr)$  容积的  $x^\alpha$  倍，其中  $\log_\rho \gamma_g \leq \alpha \leq 2\log_\rho \gamma_g$ 。

证明：首先，根据增量维度定义可得

$$|B_P(xr)| \leq \gamma_g \left| B_P\left(\frac{x}{\rho}r\right) \right|$$

然后递归调用  $\lceil \log_\rho x \rceil$  次增量维度定义，直至  $\frac{x}{\rho^{\lceil \log_\rho x \rceil}} < 1$ ，可得

$$\begin{aligned} |B_P(xr)| &\leq \gamma_g^{\lceil \log_\rho x \rceil} |B_P(r)| = x^{\log_x \gamma_g^{\lceil \log_\rho x \rceil}} |B_P(r)| = x^\alpha |B_P(r)| \\ \alpha &= \log_x \gamma_g \times \lceil \log_\rho x \rceil \end{aligned}$$

因此，根据上限函数  $\lceil \cdot \rceil$  定义，计算  $\alpha$  的下界：

$$\alpha \geq \log_x \gamma_g \times \log_\rho x = \log_\rho \gamma_g$$

由于  $x \geq \rho$ ,  $\gamma_g > 1$ , 因此可得

$$\log_{\rho} \gamma_g = \frac{\log \gamma_g}{\log \rho} \geq \frac{\log \gamma_g}{\log x} = \log_x \gamma_g$$

求解  $\alpha$  的上界:

$$\begin{aligned} \alpha &\leq \log_x \gamma_g \times (\log_{\rho} x + 1) \\ &= \log_{\rho} \gamma_g + \log_x \gamma_g \\ &\leq \log_{\rho} \gamma_g + \log_{\rho} \gamma_g \\ &= 2 \log_{\rho} \gamma_g \end{aligned}$$

结论得证。

给定任意的参数  $x$ , 引理4说明相同球心不同半径的球的节点个数比至多为  $x^{\alpha}$ 。参数  $\alpha$  是一个变量, 取决于不同的球包含的节点数目。统计不同半径  $r$  下  $\alpha$  的变化情况。图3.5显示大多数的  $\alpha$  不大于 3, 并且随着半径的增大或者参数  $x$  增大迅速降低。因此, 相同球心不同半径的球所包含的节点数目差异幅度适中。

接着计算满足定义10所需的随机采样节点数目, 也就是满足至少存在一个节点被球  $B_T(\beta r)$  包含所需的采样数目。

**引理 5** 假定一个  $\rho$ -低度量模型的增量维度为  $\gamma_g \geq 1$ ,  $r$  为一个正实数, 且满足当前节点  $P$  与目标节点  $T$  的距离为  $d_{PT} \leq r$ 。在从球  $B_P(\rho r)$  均匀随机地选择  $3\left(\frac{\rho^2}{\beta}\right)^{\alpha}$  个节点时, 以至少 95% 的概率保证至少一个节点被球  $B_T(\beta r)$  包含, 其中  $\log_{\rho} \gamma_g \leq \alpha \leq 2 \log_{\rho} \gamma_g$ ,  $\beta \leq 1$ 。

证明: 首先分析低度量模型中不同球心、不同半径的球之间的包含关系, 作为引理5的分析基础。

**引理 6** 给定任意的节点对  $p$  和  $q$  以及  $d_{pq} \leq r$ , 可知

$$B_q(r) \subseteq B_p(\rho r) \subseteq B_q(\rho^2 r) \quad (3.1)$$

证明: 假设任意的节点  $i$  满足  $d_{qi} \leq r$ , 即  $i \in B_q(r)$ , 根据低度量模型定义可知  $d_{pi} \leq \rho \max\{d_{pq}, d_{qi}\} \leq \rho r$ , 因此  $i \in B_p(\rho r)$  成立。那么

$$B_q(r) \subseteq B_p(\rho r) \quad (3.2)$$

成立。对于任意的节点  $j$  满足  $j \in B_p(\rho r)$ , 由低度量模型定义可知

$$d_{qj} \leq \rho \{d_{pq}, d_{pj}\} \leq \rho^2 r \quad (3.3)$$

综合式 (3.2) 和式 (3.3), 引理得证。

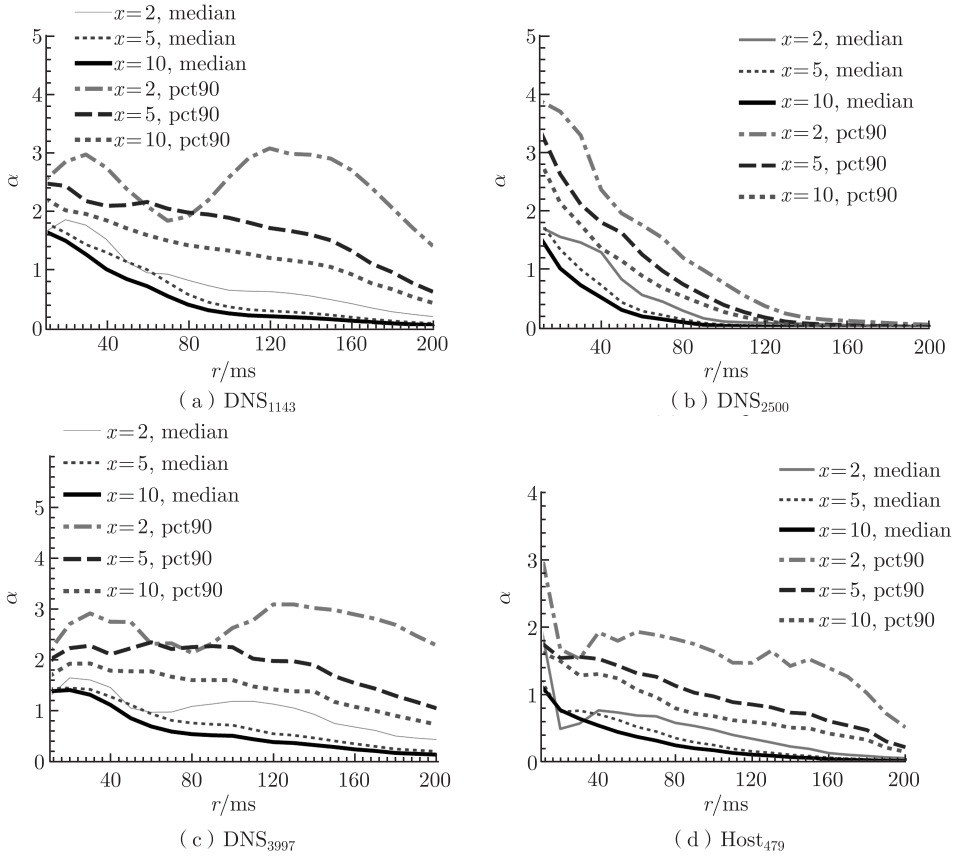


图 3.5 参数  $\alpha$  随球半径  $r$  以及参数  $x$  增长的变化

根据引理6可知,  $B_T(\beta r) \subset B_T(r) \subseteq B_P(\rho r)$ , 所有被球  $B_T(\beta r)$  包含的节点同样被球  $B_P(\rho r)$  包含。因此, 只需要在球  $B_P(\rho r)$  采样足够数目的节点以寻找位于球  $B_T(\beta r)$  内的一个节点。

对于满足  $d_{PT} \leq r$  的两个节点  $P$  和  $Q$  可知

$$|B_P(\rho r)| \leq |B_T(\rho^2 r)| = \left| B_T\left(\frac{\rho^2}{\beta} \beta r\right) \right| \quad (3.4)$$

由于  $\rho > 1$ ,  $\beta \leq 1$ , 可知  $\frac{\rho^2}{\beta} \geq \rho^2 \geq \rho$ 。因此引理4的前提条件成立, 可知球  $B_P(\rho r)$  与  $B_T(\beta r)$  的大小关系为

$$|B_P(\rho r)| \leq \left| B_T\left(\frac{\rho^2}{\beta} \beta r\right) \right| \leq \left(\frac{\rho^2}{\beta}\right)^\alpha |B_T(\beta r)| \quad (3.5)$$

其中,  $\log_\rho \gamma_g \leq \alpha \leq 2 \log_\rho \gamma_g$ 。因此, 从球  $B_P(\rho r)$  包含的节点集合均匀随机地选择一个节

点被球  $B_T(\beta r)$  包含的概率为

$$\frac{|B_T(\beta r)|}{|B_P(\rho r)|} \geq \frac{|B_T(\beta r)|}{\left(\frac{\rho^2}{\beta}\right)^\alpha |B_T(\beta r)|} = \frac{1}{\left(\frac{\rho^2}{\beta}\right)^\alpha} \quad (3.6)$$

那么随机选择  $3\left(\frac{\rho^2}{\beta}\right)^\alpha$  个采样节点均不被球  $B_T(\beta r)$  包含的概率至多为

$$\left(1 - \frac{1}{\left(\frac{\rho^2}{\beta}\right)^\alpha}\right)^{3\left(\frac{\rho^2}{\beta}\right)^\alpha} \leq \left(\frac{1}{e}\right)^3 \leq 0.05 \quad (3.7)$$

式 (3.7) 说明随机选择  $3\left(\frac{\rho^2}{\beta}\right)^\alpha$  个采样节点时选择一个满足定义10的节点的成功率至少为 95%。

从上述引理可知, 随着延迟降低幅度  $\beta$  的增大, 随机采样节点的数目不断降低; 并且随着  $\beta$  接近其上限值 1, 随机采样节点的数目近似为  $3\left(\frac{\rho^2}{\beta}\right)^\alpha \approx 3\rho^{2\alpha} \in [3\gamma_g^2, 3\gamma_g^4]$ 。

基于引理5的随机采样需求, 将整个最近节点搜索过程描述为一个递归的过程: 在搜索过程的每个中间节点  $P$  从球  $B_P(\rho d_{PT})$  包含的节点集合内随机地采样  $3\left(\frac{\rho^2}{\beta}\right)^\alpha$  个节点作为候选节点, 若这些候选节点中距目标节点最近的节点  $P_2$  与目标节点  $T$  的延迟 (不失一般性, 在度量分析中又将延迟称为距离) 不大于节点  $P$  与目标节点延迟的  $\beta$  倍, 那么将最近节点搜索请求发送给节点  $P_2$ , 否则终止搜索过程并将当前已知的最近节点作为向目标节点  $T$  提供服务的节点。

为了量化搜索到的节点与目标节点的邻近度, 定义  $\omega$ -近似度如下:

**定义 11** 给定一个目标节点  $T$ , 若一个节点  $A$  与节点  $T$  的距离  $d_{AT}$  不大于最近节点与目标节点  $T$  的距离 (记为  $d_*$ ) 的  $\omega$  倍, 即  $d_{AT} \leq \omega d_*$ , 那么称节点  $A$  与节点  $T$  为  $\omega$ -近似度。

根据定义11可知, 近似度越低说明发现的节点距目标节点越接近。接着分析搜索到的最近节点的近似度以及搜索速度。

**定理 1** 给定任意目标节点  $T$ , 发现的最近服务器具有  $\frac{1}{\beta}$ -近似度, 并且搜索跳步数至多为  $\log_{\frac{1}{\beta}} \Delta_d$ , 其中  $\Delta_d$  为网络距离空间中最大距离与最小距离的比值。

证明: (1) 假设  $P_*$  为距目标节点  $T$  最近的服务器。首先采取反证法证明搜索节点的近似度。假定搜索过程终止时发现的最近节点为  $P_x$ , 且  $P_x$  与目标节点的距离的近似度大于  $\frac{1}{\beta}$ , 即  $d_{P_x T} > \frac{1}{\beta} d_{P_* T}$ 。由于  $\beta \leq 1$ , 那么  $d_{P_x T} > d_{P_* T}$ 。因此在节点  $P_x$  处继续采取随机采样的



方式发现距目标节点比  $P_x$  更近的节点。这与搜索过程终止的假设矛盾，因此搜索到的最近节点的近似度至多为  $\frac{1}{\beta}$ 。

(2) 假设节点  $P$  将最近搜索请求发送给另外一个节点  $Q$ ，记该次搜索转发过程的搜索改进度为二者与目标节点  $T$  的距离比值  $\frac{d_{PT}}{d_{QT}}$ 。根据引理5可知，一次搜索转发过程的搜索改进度至少为  $\frac{1}{\beta}$ 。设  $d_{\min}$  为网络距离空间中最小距离值且  $d_{\min} > 0$ ， $l$  为搜索转发次数。假定经过  $l$  次搜索转发后，条件

$$\beta^l d_{PT} = d_{\min}$$

成立，那么搜索过程由于没有距目标节点更近的节点而终止。由于  $d_{PT} \leq \Delta_d \times d_{\min}$ ，可知

$$\frac{1}{\beta^l} d_{\min} = d_{PT} \leq \Delta_d \times d_{\min}$$

即

$$l \leq \log_{\beta} \left( \frac{1}{\Delta_d} \right) = \log_{\frac{1}{\beta}} (\Delta_d)$$

由于转发次数  $l$  为整数，因此实际值可能比上述连续值更小。所以在至多  $\log_{\frac{1}{\beta}} \Delta_d$  次搜索转发后，搜索过程终止。综合 (1) 与 (2)，定理得证。

搜索过程中每次发现的节点以至少 95% 的概率与目标节点的距离降低  $(1 - \beta)$  倍。这就意味着若当前搜索过程无法发现满足上述条件的新的节点进行最近邻搜索，那么搜索过程可能在没有发现真实的最近邻前终止。 $l$  搜索跳步时的搜索过程失效概率为  $1 - (95\%)^l$ 。例如假定总搜索跳步数为 4，那么搜索过程提前终止的概率至多为 18.55%。为了避免搜索过程提前终止，通过增大随机采样节点数目来增大每次的成功率。然而实验结果发现，采取适中的随机采样节点数目即可将近似度降低至接近 1。

上述理论分析实际上假定每次搜索过程只比当前节点与目标节点的距离降低  $(1 - \beta)$  倍。然而由于每次搜索转发过程均选择候选节点中与目标节点最近的节点作为新的搜索节点，因此每次搜索过程的距离降低幅度可能远大于  $(1 - \beta)$ 。即使设定  $\beta$  为 1，搜索过程的距离降低幅度也没有变化。本章邻近搜索的算法设计即利用这种不变性优化  $\beta$  值。

### 3.5.2 精确度及搜索时间

**定理 2** (增量维度下的多级环质量) 在  $\rho$ -低度量模型具有增量  $\gamma_g \geq 1$  时，固定  $\delta \in (0, 1)$ ,  $\varepsilon < 1$ ，设定多级环的每个环的大小为  $O\left(\frac{1}{\varepsilon}\right)^{\log \gamma_g} \log(N/\delta)$ ，如果每个环为良好的，则多级环以  $1 - \delta$  的概率为  $\varepsilon$ -nice。

证明：任意选择两个不同的节点  $u$  和  $v$ ,  $d_{uv} > 0$ , 设  $r = \varepsilon d_{uv}$ , 选择最小的整数  $i$  满足  $d_{uv} + r \leq 2^i$ , 因此  $r + d_{uv} > 2^{i-1}$ , 否则与  $i$  是最小的整数矛盾。

根据低度量模型定义, 对于球  $B_{ui}$  中的任意节点  $x$ , 节点  $x$  与节点  $v$  的距离  $d_{xv}$  满足

$$d_{vx} \leq \rho \max \{d_{uv}, d_{ux}\} \leq \rho \times 2^i \leq \rho \times 2(r + d_{uv}) = \rho \times 2 \left(1 + \frac{1}{\varepsilon}\right) r = \rho\gamma r$$

其中,  $\gamma = 2(1 + 1/\varepsilon)$ 。因此, 球  $B_{ui}$  满足如下条件:

$$B_{ui} \subseteq B_v(\rho\gamma r)$$

接着利用引理1结论, 可得

$$|B_{ui}| \leq |B_v(\rho\gamma r)| \leq (\rho\gamma)^\alpha |B_v(r)|$$

其中,  $\alpha$  为倍增维度的网格维度。基于引理3, 位于  $S_{ul}(l \leq i)$  环上的某个节点, 位于  $B_v(r)$  的失效概率小于  $\delta/N^2$ 。定理 2 得证。

定理2说明, 如果多级环的每个环均匀随机地从对应距离范围选择逻辑邻居, 那么对于任意的目标节点, 多级环中均存在一个逻辑邻居, 与目标节点的距离小于当前节点与目标节点距离的  $\varepsilon$  倍,  $\varepsilon < 1$ 。因此, 多级环结构能够支持基于贪心搜索策略的 DKNNS 算法中的 K 近邻搜索过程。

**定理 3**(倍增维度下的多级环质量) 若一个  $\rho$ -低度量模型为  $\gamma_d$ -倍增, 固定  $\delta \in (0, 1)$ ,  $\varepsilon < 1$ , 设定多级环的每个环的大小为  $O\left(\frac{1}{\varepsilon}\right)^{\log \gamma_d} \log(N/\delta)$ , 如果环为良好的, 则多级环以  $(1 - \delta)$  的概率为  $\varepsilon$ -nice。

证明：首先选择两个不同的节点  $u$  和  $v$ ,  $d_{uv} > 0$ , 设  $r = \varepsilon d_{uv}$ , 选择最小的整数  $i$  满足  $d_{uv} + r \leq 2^i$ 。根据低度量模型定义, 对于球  $B_{ui}$  中的任意节点  $x$ , 节点  $x$  与节点  $v$  的距离  $d_{xv}$  满足

$$d_{vx} \leq \rho\gamma r$$

其中,  $\gamma = 2(1 + 1/\varepsilon)$ 。因此, 球  $B_{ui}$  满足如下条件:

$$B_{ui} \subseteq B_v(\rho\gamma r)$$

接着, 直接利用引理2结论, 可得

$$|B_{ui}| \leq |B_v(\rho\gamma r)| \leq (\rho\gamma)^\alpha |B_v(r)|$$

其中,  $\alpha$  为倍增维度的网格维度。最后基于引理3结论, 可知定理 3 成立。