

基于EIN覆盖网的复杂网络最短路径近似算法

冯冠钦, 林 穗

广东工业大学, 计算机学院, 广东 广州

收稿日期: 2022年3月2日; 录用日期: 2022年3月31日; 发布日期: 2022年4月7日

摘 要

在网络规模远超出最短路径经典算法适用范围的情况下, 最短路径近似算法成为有效的替代解决方案。针对现有近似算法存在的预处理阶段计算效率低、算法性能受网络规模影响较大等问题, 提出一种基于EIN覆盖网的大规模复杂网络最短路径近似算法。算法基于边递归网络(The network created by edge iterations, EIN)的生成和标号方式在实际复杂网络上抽象出具有确定性拓扑结构的标号覆盖网络, 结合覆盖网络标号节点间确定的位置信息快速推导出实际复杂网络中最短路径的近似解, 在确定性网络层面高效解决非确定性复杂网络的最短路径问题。真实网络数据集上的实验结果表明, 所提方法在大规模复杂网络上能保证较高精确度的同时, 大幅度降低计算成本。

关键词

复杂网络, 最短路径, 近似算法, 覆盖网

The Approximate Shortest Path Algorithm for Complex Networks Based on EIN Overlay Network

Guanqin Feng, Sui Lin

School of Computer Science, Guangdong University of Technology, Guangzhou Guangdong

Received: Mar. 2nd, 2022; accepted: Mar. 31st, 2022; published: Apr. 7th, 2022

Abstract

The shortest path approximation algorithm becomes an effective alternative solution when the network scale is far beyond the scope of the classical shortest path algorithm. Aiming at the problems existing in the existing approximation algorithms, such as low computational efficiency in the

preprocessing stage, and the algorithm performance that is greatly affected by the network scale, a shortest path approximation algorithm for large-scale complex networks based on EIN overlay network is proposed. The algorithm abstracts the label overlay network with deterministic topology structure on the actual complex network based on the generation and labeling method of EIN (The network created by edge iterations), and quickly deduces it by combining the location information determined between the label nodes of the overlay network. The approximate solution of the shortest path in the actual complex network can efficiently solve the shortest path problem of the non-deterministic complex network at the deterministic network level. The experimental results on real network datasets show that the proposed method can greatly reduce the computational cost while ensuring high accuracy on large-scale complex networks.

Keywords

Complex Network, Shortest Path, Approximation Algorithm, Overlay

Copyright © 2022 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

信息时代已成为人类社会发展的必然趋势, 科学信息化、社会信息化带动着世界性的巨大变革。在大数据时代的背景下, 复杂网络自然而然地成为了一门深刻的研究课题。复杂网络是现实世界中各种实体之间关系的高度抽象, 包括社交网络、交通系统、电网和病毒传播。因此, 关于复杂网络的理论研究涵盖了实际生活中的众多领域, 具有巨大的应用价值。

在复杂网络中寻找任意两个节点之间的最短路径[1]一直是众多专家研究和讨论的一个重要问题。现有的最短路径精确算法, 例如 Dijkstra [2]和 Floyd [3], 是通过大量的计算来求得节点间的精确距离, 这很适用于早期的小型网络。但随着数据规模的迅猛增长, 精确算法效率低的缺点被不断放大。对于大规模网络和实时网络, 学者们更关注于算法的计算效率。在误差可接受的范围内, 高效计算出良好的近似路径来替代精确路径已被证实为可行的替代解决方案。最短路径近似算法的常用设计方法包括启发策略 [4] [5]、分层策略[6] [7]、地标点策略[8] [9]等, 都是通过预处理保留网络部分重要信息, 达到提高计算效率的目的。文献[10]提出了基于区域中心点距离(centers distance of zone, CDZ)的最短路径近似算法, 该算法结合了中心性度量方法筛选出中心节点集合, 基于区域中心点间的精确路径去估算全局节点间的最短路径。CDZ 算法的准确性在大规模网络上表现突出, 但其预处理阶段花费的时间较长, 算法的时间效率受网络规模的影响较大。文献[11]中提出的基于 k-shell 的最短路径(K-shell shortest path approximation algorithm, KS)近似算法将网络划分为不同层次的 k-shell 子图, 利用高层网络间的精确路径引导路径搜索。虽然 KS 算法相比于 CDZ 算法降低了中心节点的规模, 但仍旧依赖于区域节点间的精确计算, 导致计算效率偏低。Dennis [12]提出超点聚合的分层方法来降低网络的规模, 通过压缩搜索空间来提高计算效率。但算法在聚合中心节点的过程中需要构建多层网络, 消耗了过多的存储空间。

尽管近年来围绕着大规模复杂网络的最短路径近似算法方面的研究已经取得了一些突破和成果, 但是, 总体上该研究仍然处于发展中阶段。现有的最短路径近似算法大都或多或少的存在一些缺陷与不足, 在搜索效率、计算精度以及存储消耗方面通常各有利弊, 很多算法效率的提高都是以牺牲大量的预处理或搜索精度为代价的, 因而如何在保障精确度的情况下不消耗过多的预处理时间及存储空间来找到近似路径, 尚有待于进一步研究。复杂网络拓扑的随机性是加重最短路径计算成本的主要原因之一, 目前针

对复杂网络随机性展开的研究寥寥无几。而确定性复杂网络[13]相较于随机网络的优势在于能依据确定的拓扑关系推断出复杂网络结构性质的解析解。鉴于此, 本文结合覆盖网络和确定性网络的概念, 提出基于 EIN [14]覆盖网络(EIN overlay network, 简称 EON)的最短路径近似算法, EON 算法基于边递归网络模型 EIN 的生成方式与拓扑特性构建确定性覆盖网络并对节点进行标号, 将随机网络任意两点间的最短路径问题抽象到确定性网络的层面处理, 利用节点标号性质更加准确高效地搜索大规模复杂网络节点之间的最短路径。

2. 基本概念

2.1. 基于标号的确定性路由算法

确定性网络作为研究人员设计出来分析计算复杂网络的工具, 结构和特征都接近于实际中的网络。相比于随机网络, 确定性网络的优势在于能通过严格的推导出特定解。文献[15]提出了一种用于 Farey 族顶点的标号方法, 并基于标号推导出 Farey 族网络任意节点间的最短路径。

Farey 网由著名的 Farey 序列[16] [17] [18]推导而来。假设 Farey 图为 $F(t)$, 其中是 t 迭代的步数, $F(t)$ 通过以下两个步骤迭代生成:

- 1) 当 $t = 0$ 时, $F(0)$ 由两个初始节点和一条活动边组成。
- 2) 当 $t \geq 0$ 时, $F(t)$ 由 $F(t-1)$ 得到, 方法是在所有 $t-1$ 时刻加入的活动边上增加一个新节点, 并将新节点与边两端的节点相连。

边递归网络 EIN 是 Farey 网的衍生, 它由三个 Farey 网络而组成, 记为 $N(t)$ 。其生成方式为以下两步:

- 1) 当 $t = 0$ 时, $N(0)$ 由三个初始节点和三条活动边组成。
- 2) 当 $t \geq 0$ 时, 在 $N(t-1)$ 中所有 $t-1$ 时刻加入的活动边上增加一个新节点并连接至边的两端, 得到 $N(t)$ 。

结合确定性网络的生成方式对节点进行标号, 是推导最短路径的关键, 标号能反映出节点间的拓扑关系。Farey 网和 EIN 网的生成和标号过程分别如图 1 和图 2 所示。

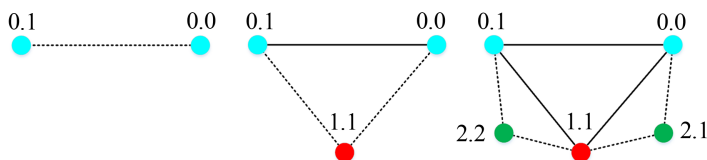


Figure 1. The labeling of $F(t)$ for $t = 0, 1$ and 2

图 1. $t = 0, 1$ 和 2 时刻递归生成的标号 $F(t)$

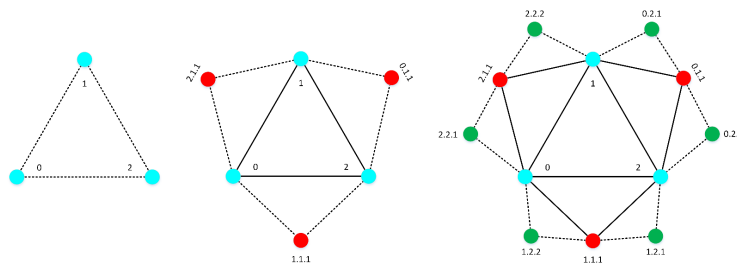


Figure 2. The labeling of $N(t)$ for $t = 0, 1$ and 2

图 2. $t = 0, 1$ 和 2 时刻递归生成的标号 $N(t)$

Farey 网和 EIN 网络具有以下标号性质:

性质 1: Farey 网和 EIN 网中的任意节点能根据标号推导出所有相邻节点标号。

性质 2: 对于 $t \geq 1$ 的 $F(t)$, 可视为由两个子网 $F_1(t-1)$ 和 $F_2(t-1)$ 组成。不同子网节点对间最短路径经过的中间节点, 可由节点对所处区域(根据标号来判断)间的关系推导出。

性质 3: 对于 Farey 网中的每个节点对, 存在一个最小公共子网(MCSG)包含其所有最短路径。

基于以上标号性质, 可以线性时间复杂度找出 Farey 网和 EIN 网中任意标号节点对间的所有最短路径。

算法 1: (SPAF) Farey 网的最短路径算法主要分为三步。首先, 通过性质 3 确定目标节点对的最小公共子网 MCSG。其次, 根据性质 2 找出最短路径经过的中间节点。最后, 目标节点与中间节点组成新的节点对继续重复整个流程, 直到根据性质 1 判断出所有新的节点对都互为邻居为止, 得出目标节点对之间的所有最短路径。

算法 2: (SPAN)将 $N(t)$ 看作由三个 Farey 子网 $F_0(t)$ 、 $F_0(t)$ 和 $F_2(t)$ 组成。当目标节点对位于相同子网时, 求最短路径的过程与上述 SPAF 算法相同。当目标节点对位于不同子网时, 两个子网可构成一个完整的 $F(t+1)$, 输入标号改变后可由 SPAF 算法求出最短路径。

SPAF 算法和 SPAN 算法都能在线性时间 $O(n)$ 内确定任意节点对间最短路径的所有可能, 求出所有节点对的最短路径也仅用次平方的时间复杂度。

2.2. 覆盖网络

覆盖网络是面向应用层的顶层网络, 常为特定应用提供专用的虚拟网络拓扑。覆盖网络中的边是底层网络中物理路径的抽象, 节点针对需求以多种方式逻辑互连。这种灵活性让覆盖网络成为了互联网中特定问题的有效解决方案, 文献[19]提出了一种覆盖网络对 IP 层路径异常进行快速检测和恢复, 以此提高用户体验。本文 EON 算法结合覆盖网络思路, 构建虚拟 EIN 网络拓扑, 实现大规模复杂网络近似路径的高效搜索。

3. 基于 EIN 的最短路径近似算法 EON

EON 算法的思路是将实际非确定性复杂网络任意两点间的最短路径问题带到确定性网络的层面上来解决。需要结合覆盖网络技术, 基于边递归网络 EIN 的生成和标号方式在实际复杂网络上抽象出服务于高效最短路径近似搜索的 EIN 覆盖网络。

3.1. EIN 覆盖网络构建

节点集 V 和边集 E 组成的无向无权网络记为 $G=(V,E)$, EON 算法构建出来的覆盖网络则表示为 $G_o(t)=(O,E_o)$ 。其中覆盖网的节点集合 $O=(O_1,O_2)$ 由两类节点组成, O_1 表示 EIN 覆盖网中的标号节点集合, O_2 则表示 EIN 覆盖网中标号节点的邻居节点集合。覆盖网标号节点依据 EIN 网的生成方式进行覆盖, 将 t 时刻的活动边记为 $E_a(t)$ 。EON 算法构建 EIN 覆盖网络基于以下步骤:

1) 选择 EIN 覆盖网络的三个初始节点。

从 EIN 网的生成过程来看, 图的所有顶点在每一步迭代中将它们的度数增加两个单位, 越早加入的节点度数越高。因此本文对底层网络节点 $v \in V$ 进行度值排序, 依照度值逆序在前 1% 的节点中找到能形成三角结构且度值之和最大的三个节点作为初始标号节点 0, 1, 2。

2) 基于 EIN 网络的生成方式进行覆盖网络扩展。

标号节点间的路径是 EON 算法进行最短路径近似搜索的依据。大部分实际复杂网络具有无标度特

性, 其度分布均符合或接近幂律分布[10]。当标号节点度数之和超过全图的 50% 时, 标号节点直接涉及全图大部分的边。因此, EON 算法基于以下步骤进行覆盖网络的构建:

Step 1: 将初始节点 0, 1, 2 加入集合 O_1 , 此时 $t=0$, $E_a(0)=E_o(0)=\{(0,1),(0,2),(1,2)\}$ 。

Step 2: 依次找到 $E_a(t)$ 中每条活动边两端节点 s 和 t 的度值最高的共同邻居 v_n 加入 $O_1(t+1)$, 其组成的边 (s, v_n) 和 (v_n, t) 则加入 $E_a(t+1)$ 中。

Step 3: 判断 $E_a(t+1)$ 是否为空, 是则表示无符合生成方式的节点进行扩展, 依次找到 G_o 最外层中每条边的两端节点 s 和 t 的度值最高的邻居 v_n 加入 $O_1(t+1)$, 其组成的边 (s, v_n) 和 (v_n, t) 则加入 $E_a(t+1)$ 中, 此时边 (s, v_n) 和 (v_n, t) 与底层网络对应边的映射关系如图 3 所示, (s, v_n) 在覆盖网络的权重为 2。

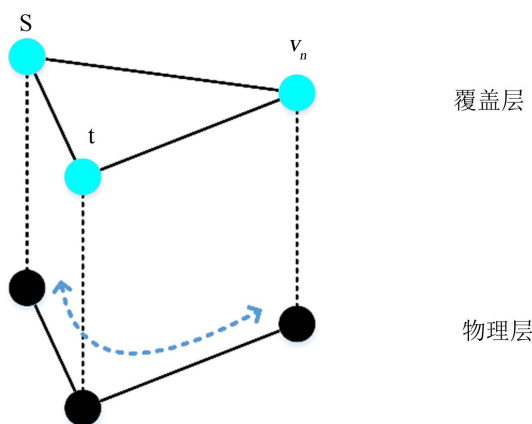


Figure 3. Mapping of conflicting nodes
图 3. 冲突节点的映射方式

Step 4: 如果集合 O_1 中节点的度数之和大于全图度数之和的 50%, 则转到第五步; 反之, 使 $t=t+1$, $E_o(t)=E_a(t)$, 继续执行第二步。

Step 5: O_1 节点间在实际网络中的路径映射到覆盖网络中, 覆盖网络 G_o 初步构造完成, 将 $V-O_1$ 的节点加入集合 O_2 。

3) 基于 EIN 网络的标号方式对覆盖网络的节点进行标号。

节点标号格式为 $a.b.c$, 其中 $a \in \{0,1,2\}$ 表示节点所属的 Farey 子网 $F_a(t)$, $b \in \{1,2,\dots,t\}$ 和 $c \in \{1,2,\dots,2^{b-1}\}$ 则表示节点在子网 $F_a(t)$ 中的位置。

4) 建立标号节点与其余节点的联系。

通过广度优先遍历依次找到最接近标号节点 $s \in O_1$ 的邻居 $t \in O_2$, 将其在实际网络中的路径抽象到覆盖网络上的边 (s_o, t_o) 。

图 4 为构建 EIN 覆盖网的一个简单示例。图 4(a)~(d)展示了基于边递归网络 EIN 的生成方式进行覆盖网络初步覆盖的过程, t 时刻的活动边用红色标注。图 4(e)和图 4(f)表示 GEIN 覆盖网的标号过程。完整的 EIN 覆盖网如图 4(g)所示。

3.2. 最短路径近似搜索

在实际复杂网络上抽象出 EIN 覆盖网络后, 标号节点间的位置信息能通过标号快速推导出来。EON 算法利用标号节点的位置信息, 在目标节点对进行最短路径近似搜索的过程中, 利用 EIN 覆盖网络标号节点对间的距离判断节点出现在最短路径的可能性, 排除低概率的节点方向来限制搜索范围, 从而提高路径搜索的效率。EON 算法的最短路径近似搜索过程如下:

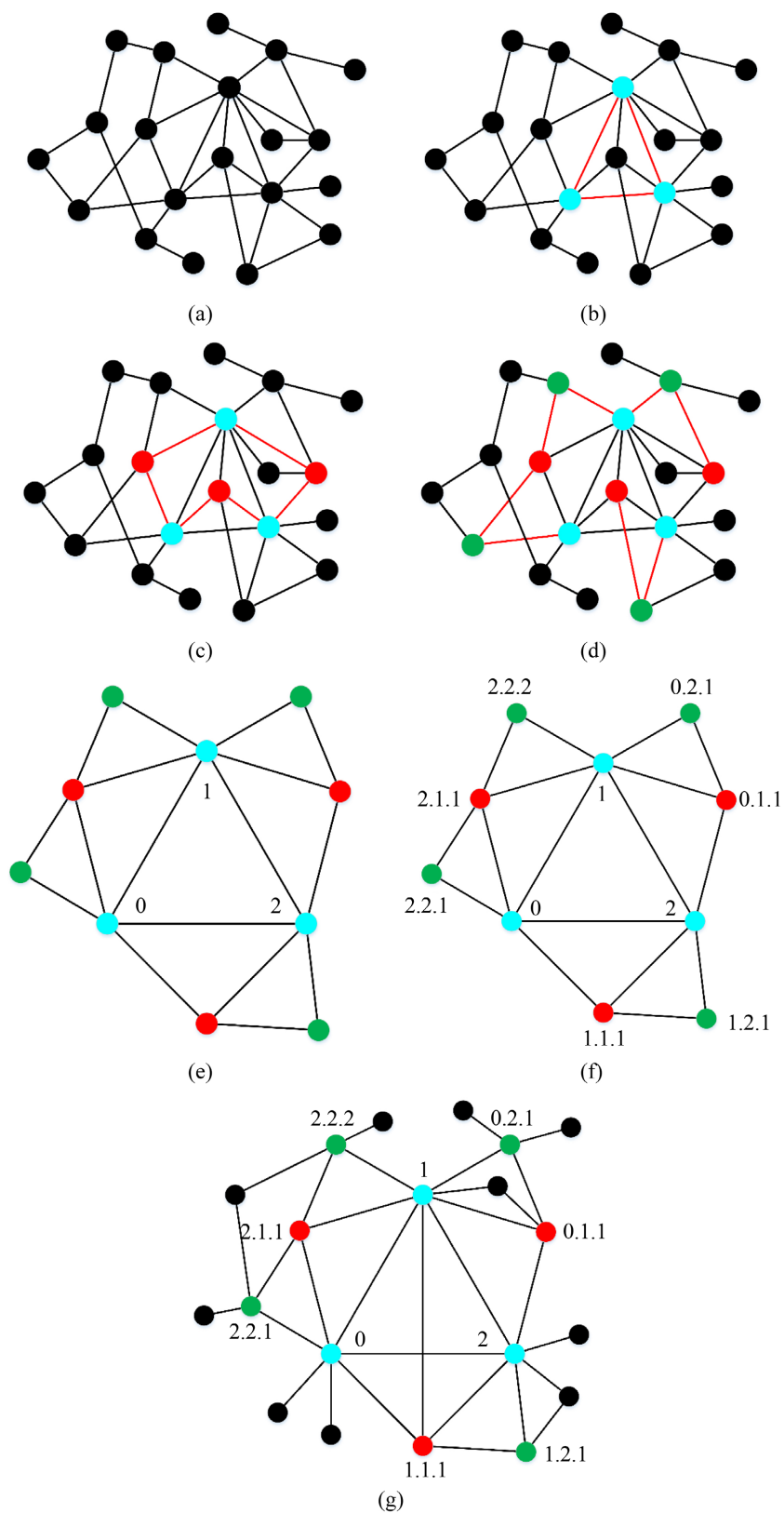


Figure 4. Construction process of EIN overlay network
图 4. EIN 覆盖网的构建过程

1) 计算 EIN 覆盖网络标号节点间的最短路径距离。

记 $d_E(l_s, l_t)$ 为标号节点 s, t 在 EIN 网络拓扑上的最短路径距离, 其中 l_s, l_t 为节点标号。 d_E 作为 EON 算法对路径近似估算的判断, 可用 SPAN 算法高效得出。在算法 SPAN 中, 求得全节点对间最短路径(The all-pairs shortest paths problem, APSP)需要次平方的运算时间 $O(n^{2-\delta})$, 其中 $\delta > 0$ 。由于边递归网络 EIN 是递归构成的, 外层节点对间的最短路径包含内层节点对的最短路径。因此由内向外去计算标号节点对间的最短路径距离, 能提高 SPAN 算法的计算效率。

2) 基于 d_E 值迭代搜索最短近似路径

节点在搜索去往目标节点最短路径的过程中, 选取 d_E 值与边权重之和最低的邻居加入最短近似路径序列, 并将其作为新的起始节点, 直到与目标节点直接相连。

实际网络中任意节点间的距离 $d(s, t)$ 可以表示为:

$$d(s, t) = \begin{cases} d_o(l_s, l_t) & s, t \in O_1 \\ d(s, n_{s_o}) + d_o(l_{n_{s_o}}, l_t) & s \in O_2, t \in O_1 \\ d(s, n_{s_o}) + d_o(l_{n_{s_o}}, l_{n_{t_o}}) + d(t, n_{t_o}) & s, t \in O_2, l_s \neq l_t \\ d(s, LCA_{st}) + d(t, LCA_{st}) & s, t \in O_2, l_s = l_t \end{cases} \quad (1)$$

如果目标节点都属于标号节点集合 O_1 , 节点间的距离为其在 EIN 覆盖网络上的最短近似路径长度 d_o 。

如果目标节点对中在 O_2 集中的节点, 节点间的距离除了包括标号节点对间的最短近似路径长度, 还需要考虑节点 $s \in O_2$ 和其覆盖网络邻居 n_{s_o} 在实际网络上的映射路径长度。

如果目标节点都属于 O_2 集合, 且它们在覆盖网络上具有相同的标号邻居。通过节点对到标号节点的最短路径序列可以快速找到它们的最近公共祖先。目标节点间的距离为其通往最近公共祖先的路径长度之和。

图 5 为 EON 算法的一个简单示例。节点 9 到节点 7 的最短近似路径搜索过程中, 核心在于选取节点 5 之后最有可能存在最短路径上的节点。此时节点 5 的邻居标号分别为 1、2 和 2.1.1, 通过 SPAN 算法可推导出节点 1, 3, 4 在 EIN 网络拓扑上到节点 7 的最短路径集合分别为 $\{\{1, 2.1.1, 2.2.1\}, \{1, 0, 2.2.1\}\}$ 、 $\{\{2, 0, 2.2.1\}\}$ 和 $\{\{2.1.1, 2.2.1\}\}$, 由此可推断节点 4 为最可能出现在最短路径上的节点。最终得到节点 9 到节点 7 的最短近似路径 $\{9, 8, 5, 4, 7\}$ 。

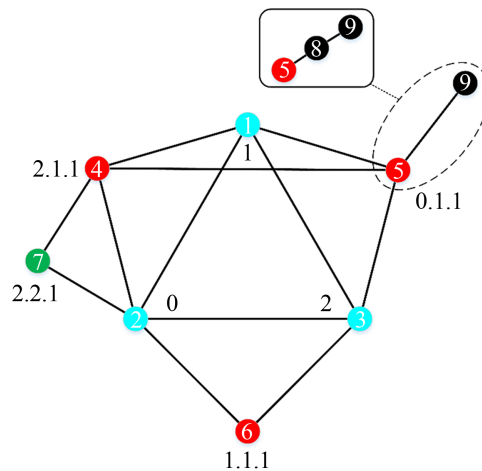


Figure 5. Example diagram of EON

图 5. EON 算法示例图

4. 实验与分析

4.1. 数据集及评估指标

EON 算法在 PyCharm 集成开发环境上使用 Python 语言实现, 运行环境处理器为 Intel Corei5-8265U @1.80 GHz, 内存为 8 GB。选用四个不同规模和类型的无向无权真实网络评估本文算法, 分别是 ego-Facebook [20]、email-Enron [21]、com-DBLP [22]和 soc-BC。ego-Facebook 数据的采集来自 Facebook 应用程序调查的参与者, 节点代表参与者的账号, 边则代表两个账号之间的好友关系。email-Enron 涵盖了上万封 Enron 公司管理的电子邮件通信记录, 数据集最初由联邦能源监管委员会在调查期间公开并发布到网络上。soc-BC 为博客目录网络, 包含了用户之间的所有链接。com-DBLP 来源于计算机科学杂志, 网络代表同期刊或会议等出版地点的作者形成的真实社区。如表 1 所示, 通过 4 个指标来展示实验网络的基本信息, 分别是节点数量 V 、边的数量 E 、平均聚类系数 ACC 和网络直径 D 。

Table 1. Basic information on the five networks

表 1. 实验网络的基本信息

| 指标 | ego-Facebook | email-Enron | soc-BC | com-DBLP |
|-------|--------------|-------------|---------|-----------|
| V | 4039 | 36,692 | 41,600 | 317,080 |
| E | 88,234 | 183,831 | 221,698 | 1,049,866 |
| ACC | 0.6055 | 0.4970 | 0.5121 | 0.6324 |
| D | 8 | 11 | 13 | 21 |

本文选择平均路径比 p [23]作为最短路径近似算法准确性的度量指标。 p 的定义如下:

$$p = \frac{\sum_{i=1}^{p_r} p_{A_i}}{\sum_{i=1}^{p_r} p_{D_i}} \quad (2)$$

其中 p_r 是节点对的数量, p_{A_i} 代表由近似算法求出的第 i 对节点间的近似距离, p_{D_i} 代表由 Dijkstra 算法求出的第 i 对节点间的精确距离。近似算法的精确度取决于 p 接近 1 的程度。

最短路径近似算法的计算效率由预处理时间 T_{init} 和平均查找时间 T_s 来反映。实验通过在测试网络上随机得到 10,000 个节点对, 取其平均运行时间作为算法的 T_s 。

4.2. 实验结果与分析

表 2 展示了 EON 算法、KS 算法和 CDZ 算法在 4 个目标网络的准确性实验结果。三种算法在 4 个实验网络中的平均路径比 p 都低于 1.1, 说明 EON 算法、KS 算法和 CDZ 算法计算出的近似路径仅与精确路径存在不到 10%的误差, 都具有较高的准确率。KS 算法和 CDZ 算法利用不同的中心性指标去评估节点的重要性, 选取一定规模的重要节点作为中心节点, 通过中心节点间的精确路径来保证算法的准确性。EON 算法则是利用实际网络中的确定性拓扑结构的位置信息来保证算法的准确性。

Table 2. Accuracy of approximation algorithms in different networks

表 2. 近似算法在不同网络上的准确性

| 算法 | ego-Facebook | email-Enron | soc-BC | com-DBLP |
|-----|--------------|-------------|--------|----------|
| EON | 1.079 | 1.022 | 1.031 | 1.020 |

Continued

| | | | | |
|-----|-------|-------|-------|-------|
| KS | 1.079 | 1.024 | 1.032 | 1.021 |
| CDZ | 1.081 | 1.026 | 1.034 | 1.020 |

从表 3 中可以发现, EON 算法在计算效率上相较于其他两种近似算法有明显的优势。KS 算法和 CDZ 算法在预处理时间 T_{init} 和平均查找时间 T_s 这两个数据方面远劣于本文算法, 特别是在大规模网络 com-DBLP 的实验中, 时间开销是 EON 算法的上百倍。

Table 3. Computational efficiency of approximation algorithms in different networks

表 3. 近似算法在不同网络上的计算效率

| 网络 | EON | | KS | | CDZ | |
|--------------|-----------------|------------|-----------------|------------|-----------------|------------|
| | T_{init} (ms) | T_s (ms) | T_{init} (ms) | T_s (ms) | T_{init} (ms) | T_s (ms) |
| ego-Facebook | 2312 | 0.53 | 2626 | 1.47 | 16,106 | 1.62 |
| email-Enron | 35,544 | 3.91 | 149,208 | 20.68 | 721,170 | 72.13 |
| soc-BC | 40,476 | 4.60 | 240,616 | 28.73 | 7,298,342 | 729.852 |
| com-DBLP | 78,320 | 8.69 | 2,046,107 | 204.81 | 23,190,519 | 2320.12 |

CDZ 算法和 KS 算法在最短路径近似搜索的过程中依赖于中心/高层节点间的精确路径, 需要在预处理阶段花费 $O(d^3)$ 的时间, 其中 d 为中心/高层节点的数量。虽然中心/高层节点在全图的占比(约占节点总数 5%~10%)并不高, 但放大到大规模实际网络中, 这任是一笔不可忽视的巨大开销。EON 算法的近似搜索则依赖于标号节点间在 EIN 拓扑结构上的精确路径, 结合包含节点间拓扑关系的标号信息, 仅需 $O(t^{2-\delta})$ 的时间复杂度可推导出, 其中 t 为覆盖网络标号节点的数量。

EON 算法基于边递归网络 EIN 的生成方式构建覆盖网络, 标号节点间的拓扑结构是固定的, 因此提前计算并存储一定规模 EIN 网络标号节点间的最短路径, 不同实际网络的近似路径搜索情况下, 能节省大量的重复计算时间。复杂的网络系统是呈现动态的[24], 这种动态由多种变化过程产生, 例如进化、学习和衰老。巨大的预处理开销在需要频繁的对网络进行重复计算或者只计算部分重要节点对的最短路径的情况下显然是不能被接受的。相比于现有的最短路径近似算法, EON 算法能快速适应网络变化, 更具灵活性。

EON 算法和 CDZ 算法在 email-Enron 网络和 soc-BC 网络不同规模子网络上的性能变化如表 4 和表 5 所示。相比于 CDZ 算法, EON 算法具有明显的性能优势, 特别是在平均计算时间数据方面的表现上。5000 到 30,000 节点规模的跨度中, CDZ 算法的平均计算时间横跨了三个计数单位, 本文算法在保证高准确度的同时, 能维持平均计算时间的增长在一个计数单位内。证明 EON 算法适应网络规模增长代价低, 更加适用于节点规模大的实际复杂网络。

Table 4. Performance of approximation algorithms in different scale email-Enron networks

表 4. 近似算法在不同规模 email-Enron 网络上的表现

| Algorithm | Parameter | N | | | | | |
|-----------|------------|-------|--------|--------|--------|--------|--------|
| | | 5000 | 10,000 | 15,000 | 20,000 | 25,000 | 30,000 |
| EON | T_s (ms) | 0.25 | 0.29 | 0.52 | 0.79 | 1.10 | 1.29 |
| | P | 1.021 | 1.034 | 1.029 | 1.033 | 1.030 | 1.025 |

Continued

| | | | | | | | |
|-----|------------|-------|-------|-------|-------|-------|-------|
| CDZ | T_s (ms) | 0.60 | 4.03 | 12.03 | 19.26 | 29.37 | 46.63 |
| | P | 1.050 | 1.069 | 1.055 | 1.039 | 1.041 | 1.032 |

Table 5. Performance of approximation algorithms in different scale soc-BC networks
表 5. 近似算法在不同规模 soc-BC 网络上的表现

| Algorithm | Parameter | N | | | | | |
|-----------|------------|-------|--------|--------|--------|--------|--------|
| | | 5000 | 10,000 | 15,000 | 20,000 | 25,000 | 30,000 |
| EON | T_s (ms) | 0.55 | 2.62 | 4.07 | 5.85 | 6.99 | 9.10 |
| | P | 1.060 | 1.040 | 1.039 | 1.029 | 1.031 | 1.030 |
| CDZ | T_s (ms) | 2.57 | 18.66 | 80.38 | 167.89 | 315.49 | 514.18 |
| | P | 1.062 | 1.040 | 1.043 | 1.033 | 1.034 | 1.039 |

5. 总结

本文针对复杂网络最短路径近似算法预处理阶段时间或空间成本过高、计算效率受网络规模影响较大的问题, 利用确定性网络能够基于节点间确定的位置关系快速推导出最短路径解析解这一特点, 提出基于 EIN 覆盖网络最短路径近似算法 EON。经过在社交网络、引文网络、博客目录网络等实际复杂网络的一系列实验结果表明, EON 算法在不同规模的无向无权网络中能保证高准确度和计算效率的同时, 以较低的代价适应网络规模的增长, 更适用于大规模实际复杂网络。

从确定性模型的层面去解决随机场景下的网络节点间最短路径问题是近似算法的一种新的解决思路。深入研究不同类型的实际复杂网络与确定性网络拓扑之间的联系, 评估与改进覆盖网络的构建方法, 是我们未来研究工作的重点。

参考文献

- [1] 陆锋. 最短路径算法: 分类体系与研究进展[J]. 测绘学报, 2001, 30(3): 269-275.
- [2] Dijkstra, E.W. (1959) A Note on Two Problems in Connexion with Graphs. *Numerische Mathematik*, **1**, 269-271. <https://doi.org/10.1007/BF01386390>
- [3] Myoupo, J.F. and Fabret, A.C. (1996) A Modular Systolic Linearization of the Warshall-Floyd Algorithm. *IEEE Transactions on Parallel & Distributed Systems*, **7**, 449-455. <https://doi.org/10.1109/71.503769>
- [4] Li, J., Deng, K., Huang, X., et al. (2019) Analysis and Applications of Location-Aware Big Complex Network Data. *Complexity*, **2019**, Article ID: 3410262. <https://doi.org/10.1155/2019/3410262>
- [5] Kuipers, F., et al. (2002) An Overview of Constraint-Based Path Selection Algorithms for QoS Routing. *IEEE Communications Magazine*, **40**, 50-55. <https://doi.org/10.1109/MCOM.2002.1106159>
- [6] Song, Q. and Wang, X.F. (2011) Efficient Routing on Large Road Networks Using Hierarchical Communities. *IEEE Transactions on Intelligent Transportation Systems*, **12**, 132-140. <https://doi.org/10.1109/TITS.2010.2072503>
- [7] Wang, Y., Wu, J.J., et al. (2017) An Improved Hierarchical A* Algorithm in the Optimization of Parking Lots. *AIP Conference Proceedings*, **1864**, 20031-20031. <https://doi.org/10.1063/1.4992848>
- [8] Qiao, M. (2013) Query Processing in Large-Scale Networks. The Chinese University of Hong Kong, Hong Kong.
- [9] Shlomi, M., Rami, P. and Guy, S. (2017) Shortest Path Tree Sampling for Landmark Selection in Large Networks. *Journal of Complex Networks*, **5**, 795-815.
- [10] 唐晋韬, 王挺, 王戟. 适合复杂网络分析的最短路径近似算法[J]. 软件学报, 2011, 22(10): 2279-2290.
- [11] 张昕, 严沛, 郭阳, 王慧慧. 基于 k-shell 的复杂网络最短路径近似算法[J]. 计算机工程与应用, 2019, 55(14):

54-60.

- [12] Mensah, D., Gao, H. and Yang, L.W. (2020) Approximation Algorithm for Shortest Path in Large Social Networks. *Algorithms*, **13**, 36. <https://doi.org/10.3390/a13020036>
- [13] 黄韬, 汪硕, 黄玉栋, 郑尧, 刘江, 刘韵洁. 确定性网络研究综述[J]. 通信学报, 2019, 40(6): 160-176.
- [14] Xiao, Y. and Zhao, H. (2013) Counting the Number of Spanning Trees of Generalization Farey Graph. 2013 *IEEE 9th International Conference on Natural Computation (ICNC)*, Shenyang, 23-25 July 2013, 1778-1782. <https://doi.org/10.1109/ICNC.2013.6818271>
- [15] 翟因虎. 复杂网络确定性模型研究[D]: [博士学位论文]. 广州: 广东工业大学, 2016.
- [16] Wright, E.M. and Hardy, G.H. (1981) *Wright: An Introduction to the Theory of Numbers*. Clarendon Press, Oxford.
- [17] Cobeli, C. and Zaharescu, A. (2003) The Haros-Farey Sequence at Two Hundred Years. A Survey. *Acta Universitatis Apulensis. Mathematics. Informatics*, **5**, 1-38.
- [18] Zhang, Z., Rong, L. and Guo, C. (2006) A Deterministic Small-World Network Created by Edge Iterations. *Physica A: Statistical Mechanics and Its Applications*, **363**, 567-572. <https://doi.org/10.1016/j.physa.2005.08.020>
- [19] Chung, W.-C., et al. (2014) Maintenance of Cooperative Overlays in Multi-Overlay Networks. *IET Communications*, **8**, 2676-2683. <https://doi.org/10.1049/iet-com.2013.0709>
- [20] Mcauley, J.J. and Leskovec, J. (2012) *Learning to Discover Social Circles in Ego Networks*. Curran Associates Inc., Red Hook.
- [21] Leskovec, J., Lang, K.J., Dasgupta, A., et al. (2008) Community Structure in Large Networks: Natural Cluster Sizes and the Absence of Large Well-Defined Clusters. *Internet Mathematics*, **6**, 29-123. <https://doi.org/10.1080/15427951.2009.10129177>
- [22] Yang, J. and Leskovec, J. (2012) Defining and Evaluating Network Communities Based on Ground-Truth. *Knowledge & Information Systems*, **42**, 181-213. <https://doi.org/10.1007/s10115-013-0693-z>
- [23] Rattigan, M.J. (2006) Using Structure Indices for Efficient Approximation of Network Properties. In: *Proceedings of the Twelfth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, Knowledge Discovery Laboratory, Department of Computer Science University of Massachusetts, Amherst, 357-366. <https://doi.org/10.1145/1150402.1150443>
- [24] Carley, K.M., Diesner, J., Reminga, J., et al. (2007) Toward an Interoperable Dynamic Network Analysis Toolkit. *Decision Support Systems*, **43**, 1324-1347. <https://doi.org/10.1016/j.dss.2006.04.003>