ISSN 1673-9418 CODEN JKYTA8
Journal of Frontiers of Computer Science and Technology
1673-9418/2018/12(12)-1903-11
doi: 10.3778/j.issn.1673-9418.1710041

E-mail: fcst@vip.163.com
http://www.ceaj.org
Tel: +86-10-89056056

差分隐私下满足一致性的轨迹流量发布方法*

张双越[1],蔡剑平[2],田 丰[1],吴振强[1+]
1. 陕西师范大学 计算机科学学院,西安 710119
2. 福州大学 数学与计算机科学学院,福州 350116

Trajectory Flow Releasing Method with Consistency Constraint under Differential Privacy*

ZHANG Shuangyue[1], CAI Jianping[2], TIAN Feng[1], WU Zhenqiang[1+]
1. College of Computer Science, Shaanxi Normal University, Xi'an 710119, China
2. College of Mathematics and Computer Science, Fuzhou University, Fuzhou 350116, China
+ Corresponding author: E-mail: zqiangwu@snnu.edu.cn

ZHANG Shuangyue, CAI Jianping, TIAN Feng, et al. Trajectory flow releasing method with consistency constraint under differential privacy. Journal of Frontiers of Computer Science and Technology, 2018, 12(12): 1903-1913.

Abstract: Vehicles carrying GPS equipment create large trajectory information. Analyzing and publishing trajectory data flow statistics based on road network is beneficial to the improvement of network structure and the realization of intelligent transportation. However, the direct release of trajectory traffic can lead to the disclosure of user privacy, and there is lack of a rigorous and provable privacy method to release traffic flow in road networks. Therefore, this paper presents a differential privacy trajectory flow releasing method. The method is divided into two steps: firstly, the flow value of each section is statistically calculated and the difference privacy noise is added. Secondly the post adjustment algorithm is put forward for the consistency characteristic of the flow graph, so that the adjusted flow graph not only satisfies the consistency characteristic, but also greatly reduces the publishing error. The experiment

* The National Natural Science Foundation of China under Grant No. 61602290 (国家自然科学基金); the Fundamental Research Funds for the Central Universities of China under Grant Nos. GK201603093, GK201501008 (中央高校基本科研业务费专项资金); the Natural Science Basic Research Program of Shaanxi Province under Grant No. 2017JQ6038 (陕西省自然科学基础研究计划).
Received 2017-10, Accepted 2018-01.
CNKI网络出版: 2018-01-19, http://kns.cnki.net/kcms/detail/11.5602.TP.20180118.1740.008.html

on the real road network shows that the method has the ability to deal with large-scale network traffic, and after the optimization of the post adjustment algorithm, the release error is reduced by about 13%.

Key words: trajectory flow; differential privacy; consistency adjustment; road network constraints

摘　要:搭载GPS设备的车辆在运行过程中产生大量轨迹信息,对轨迹流量信息的统计与发布有利于改善路网结构,实现智能交通。但是直接发布轨迹流量可能导致用户隐私的泄露,而目前缺乏严格的可证明的轨迹流量隐私保护发布方法。为此,提出了一种基于路网的差分隐私轨迹流量发布方法。该方法分两步:首先根据轨迹数据统计各个路段的流量值并添加差分隐私噪声;随后针对流量图的一致性特性提出后置调节算法,使得调节后的流量图不仅重新满足一致性特性,而且还极大地减少了发布误差。在真实路网上的实验表明,该方法具有处理大规模路网流量的能力,且经过后置调节算法的优化,发布误差减小了约13%。

关键词:轨迹流量;差分隐私;一致性调节;路网约束

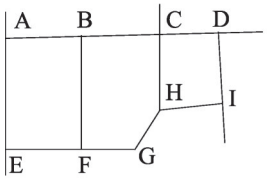
文献标志码:A　**中图分类号:**TP309.2

1 引言

随着移动互联网和GPS定位技术的发展,移动对象不断上传的位置信息形成了轨迹大数据。根据轨迹数据所含的丰富时空信息可以挖掘分析出许多人类行为模式和交通演化规律[1]。在路网交通背景下,对车辆轨迹数据中各个统计指标的发布可为道路管理及应急规划提供有力支持。例如,发布轨迹流量指标的统计信息可反映出每条路段的拥堵程度以及流量特征,这对于用户出行,交通指挥,乃至路网结构调整及整个路网的承载能力和可靠性的提高都有重要的现实意义和应用价值[2-3]。

然而,轨迹数据中隐含了用户的出行习惯、行为模式等隐私信息[4-5]。因此,直接发布轨迹流量值可能造成用户隐私泄露。例如,假设攻击者掌握了某个特定时段内除目标用户之外其他用户的轨迹信息,则他再结合所发布的轨迹流量数据即可追踪出目标用户在该时段内的轨迹。为了保护用户隐私,常用的轨迹发布隐私保护方法主要有假数据法[6-7]、泛化法[8-12]、抑制法[13-15]。这些方法通常通过变换或剔除轨迹中的敏感信息来保护轨迹数据的发布。但是,这些方法对攻击者拥有的背景知识敏感,容易受到背景知识攻击。当攻击者掌握了较多的背景信息时,用户的隐私就难以保障且缺乏严格的数学证明。2008年Dwork针对统计数据库的隐私泄露问题提出了差分隐私模型[16]。该模型保证了攻击者在任何知识背景下都无法准确地从所发布的数据中获得隐私信息且具有严格的概率分布不可区分性证明,引起了广大研究者的兴趣。2011年起,差分隐私模型被应用到轨迹发布隐私保护中,并取得了一系列成果[17-24]。但是,已有的工作专注于发布经过差分隐私保护的整个轨迹数据集,对路网环境下的轨迹流量发布缺乏考虑。本文在路网约束下结合差分隐私在保护统计信息领域的优势提出轨迹流量发布算法,并对该算法进行一致性优化,提高了发布结果的可用性和发布效率。

本文将路网结构抽象为有向图,有向图中的边连接两个相邻的路口,同时将轨迹数据相应地处理为形如图1(b)所示的有序序列,对这些轨迹数据进行统计即可得到路网中各个路段的轨迹流量。以北京某街区为例,图1(a)为该街区道路示意图,图1(b)为该示意图上的车辆轨迹数据样例。经统计得到的轨迹流量图如图2(a)所示。图2(a)中节点和边分别对应图1(a)中的各个路口和街道,边上的权值代表了相应路段的轨迹流量。接下来,在差分隐私机制下,向图2(a)中的边权值添加符合拉普拉斯分布的噪声以保护用户隐私。添加噪声后的流量图如图2(b)所示。不难发现,由于拉普拉斯噪声的引入导致了某些路口出入流量不一致的问题。如B路口所示(非起止路口),添加噪音扰动前车辆驶入该路口的次数等于驶出该路口的次数,如图2(a)都为3;添加



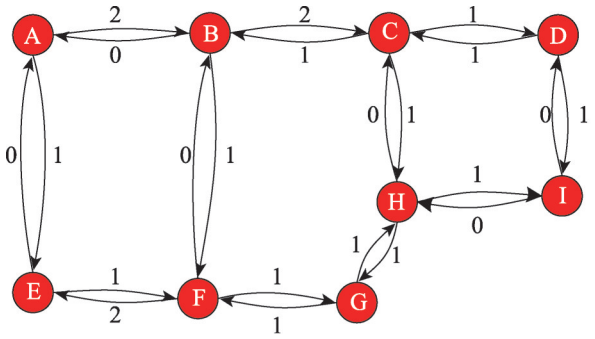
(a) City block
(a) 城市街区

ID	Trajectory
u1	A → B → C
u2	D → C → H → G → F → E
u3	A → B → C → D → I
u4	A → E → F → G → H → I
u5	C → B → F → E

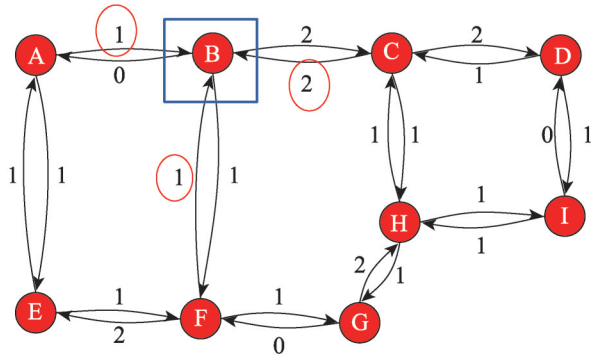
(b) Vehicle trajectory
(b) 车辆轨迹

Fig.1 Vehicle trajectory from a city block

图1 某城市街区及该街区的车辆轨迹



(a) Real traffic flow
(a) 真实轨迹流量图



(b) Protected traffic flow
(b) 隐私保护轨迹流量图

Fig.2 Traffic flow before and after protection

图2 保护前后的轨迹流量图

噪音扰动后，该路口的驶入车次为4，驶出车次为3。这显然与实际情况不符。深入研究发现，不一致问题不仅导致发布结果不符合实际，同时也增大了发布误差。为此，本文提出了一种一致性调节方案来解决上述问题。

对于差分隐私领域的不一致性问题已有学者进行了研究[25]。例如，Hay等人[26]对直方图发布中的不一致性问题提出了后置处理方案。然而由于研究目标不同，该方案无法直接应用于解决本文中的不一致问题。因此，本文在参考Hay等人[26]思路的基础上提出了一种新的后置调节算法。经过大量的理论分析以及实验表明，本文所提后置处理算法不仅能有效地解决上述不一致问题，还在合理的时间内对发布结果的精度有了明显提升。

综上所述，本文完成了以下工作：

(1)将实际城市路网抽象成有向图模型，并向图中引入一个辅助的虚拟节点。利用该虚拟节点形式化表现了流量图中的不一致性问题。

(2)结合拉普拉斯机制提出差分隐私流量图生成算法。

(3)在(2)的基础上提出一致性调节算法。该算法在严格的理论分析和公式推导下有效地解决了一致性问题，并提升了发布精度。

(4)在真实路网上对本文所提算法进行实验验证，结果表明一致性调节算法减小了约13%的发布误差且具有处理大规模数据的能力。

2 相关工作

为了解决轨迹数据发布过程中的隐私泄露问题，文献[17]首次引入了差分隐私机制。该文献利用原始轨迹数据的特性构建噪音前缀树，并向前缀树节点的计数值中添加拉普拉斯噪声保证用户的轨迹隐私。然而随着前缀树规模的增大，树中的节点将呈指数增长，导致落入每个分支的轨迹数量急剧减少，严重降低了数据可用性。随后文献[18]通过n-gram模型实现了可变长度的轨迹数据集发布，在一定程度上改善了文献[17]的不足。但是，文献[17-18]都基于一个共同的假设，即原始数据中存在大量的共同前缀，这在现实中很难满足。文献[20]利用指数机制对轨迹数据进行聚类操作，去除了轨迹数据中存在共同前缀的假设条件。文献[23]提出一种有界噪音泛化算法，减小了信息损失和平均轨迹合并时间。然而，以上文献都是发布经过差分隐私保护的整个轨迹数据集。文献[24]提出l-轨迹差分隐私保护模型，实时发布无限轨迹流上不同位置的用户统计值。然而却鲜有文献考虑路网约束下的轨迹统计值发布。文献[27]指出当发布大量用户位置的聚合信息时，差分隐私能提供较好的隐私保障。因此本文

就以保护路网中的轨迹流量统计值作为切入点展开研究工作。

差分隐私中常见的优化方法主要分为基于数据压缩转换的前置优化算法以及基于规划策略的后置优化算法[26]。Hay等人[26]利用区间树的一致性问题对直方图统计发布进行优化就是一种典型的后置优化算法,该算法利用最优估计理论不仅解决了区间树的不一致性问题,同时还有效地提升了发布数据的可用性。在该理论的启发下,本文提出了一种新的优化模型,有效解决了轨迹流量发布过程中的不一致问题并提升了发布数据的有效性。

3 预备知识

3.1 基本概念

本文在论述过程中涉及到较多的数学符号,因此在介绍相关概念之前,先对这些符号进行说明,以方便读者理解。如表1所示。

Table 1 List of notations

表1 符号表

符号	描述
V, E	V 代表交叉路口集合;E代表各路段集合
p_v	虚拟节点,作为轨迹的起止点,且与路网中其他节点均相连
ε	隐私预算
$\boldsymbol{M}$	路网拓扑图的邻接矩阵
$\boldsymbol{W}$	未加噪的轨迹流量图邻接矩阵
$\tilde{\boldsymbol{W}}$	满足 ε-差分隐私的轨迹流量图邻接矩阵
$\bar{\boldsymbol{W}}$	添加噪声并经过一致性调节后的流量图矩阵
$\boldsymbol{u}$	拉格朗日系数所组成的列向量
$\boldsymbol{I}_n, \mathbb{E}_i$	$\boldsymbol{I}_n$ 表示长度为 n 的全1列向量;$\mathbb{E}_i$ 表示 $\boldsymbol{M}$ 的对角线为1其余为0的方阵
$\boldsymbol{o}_n, \boldsymbol{O}$	$\boldsymbol{o}_n$ 表示长度为 n 的全0列向量;$\boldsymbol{O}$ 表示零矩阵
α	惩罚系数,($\alpha>>1$)
$\boldsymbol{C}$	惩罚系数矩阵,其元素为 α 或1
$\boldsymbol{C}^{\neg}$	$\boldsymbol{C}$ 中的所有元素取倒数得到的矩阵

定义1(路网拓扑图) 将城市道路网抽象成一个有向图 $G=(V,E)$,其中 V 表示道路网中所有交叉路口的集合,E 表示道路网中所有相邻交叉路口之间路段的集合。其相应的邻接矩阵记为 $\boldsymbol{M}\in\{0,1\}^{|V|\times|V|}$。其中,$|V|$ 表示集合 V 的大小,$m_{u,v}=1$ 表示路口 u 和 v 之间存在边,否则不存在边。无特殊说明,以下将路网拓扑简称为路网。

定义2(轨迹空间) 轨迹是移动用户产生的一系列位置 v_i 按照时间排序构成的集合,记为 L。本文对轨迹 L 进行预处理,使得 L 由所有它所经过的交叉路口表示,即 $\forall v_i,v_j\in V$,且每两个相邻节点所构成的边都属于边集 E。由此可对轨迹空间 $\boldsymbol{\Omega}$ 做如下定义:

$$\boldsymbol{\Omega}=\{L:\{v_i|1\leqslant i\leqslant N\}|\forall v_i,v_{i+1}\in V,M_{v_i,v_{i+1}}=1\} \tag{1}$$

其中,N 表示轨迹 L 的长度。

现实中的任何轨迹经过预处理后均是 $\boldsymbol{\Omega}$ 中的一个元素。本文所用的轨迹数据集 $\boldsymbol{D}$ 是 $\boldsymbol{\Omega}$ 的子集。根据轨迹数据集 $\boldsymbol{D}$ 和路网即可构建轨迹流量图。轨迹流量图定义如下。

定义3(轨迹流量图) 根据轨迹数据集 $\boldsymbol{D}$ 计算出路网中每条边的权值,由此构成的有向加权图即为轨迹流量图。边 $u\rightarrow v$ 的权值表示轨迹数据集 $\boldsymbol{D}$ 中经过边 $u\rightarrow v$ 的轨迹数量。轨迹流量图中各个边的权值构成的流量矩阵记为 $\boldsymbol{W}\in\mathbf{R}^{|V|\times|V|}$。$\boldsymbol{W}$ 即为本文发布的轨迹流量图。以下简称流量图。

3.2 差分隐私

定义4(差分隐私) 假设存在一个随机算法 $\boldsymbol{A}$,$\boldsymbol{S}_A$ 表示该算法所有可能的输出构成的集合。则对于两个相邻的数据集 $\boldsymbol{D},\boldsymbol{D}'(|\boldsymbol{D}\Delta\boldsymbol{D}'|\leqslant 1)$ 以及 $\boldsymbol{S}_A$,若算法 $\boldsymbol{A}$ 满足:

$$\Pr[A(\boldsymbol{D})\in\boldsymbol{S}_A]\leqslant\exp(\varepsilon)\times\Pr[A(\boldsymbol{D}')\in\boldsymbol{S}_A] \tag{2}$$

则称算法 $\boldsymbol{A}$ 满足 ε-差分隐私保护。其中参数 ε 称为隐私保护预算[16]。ε 越小表明算法 $\boldsymbol{A}$ 对数据集的保护程度越高。

定义5(全局敏感度) 对于任意函数 $f:\boldsymbol{D}\rightarrow\mathbf{R}^d$,全局敏感度表示向数据集中添加或删除一条记录,对函数 f 产生的最大影响。全局敏感度 Δf 可由以下公式计算:

$$\Delta f=\max_{\boldsymbol{D},\boldsymbol{D}'}\|f(\boldsymbol{D})-f(\boldsymbol{D}')\|_1 \tag{3}$$

其中,$\boldsymbol{D}$ 和 $\boldsymbol{D}'$ 满足 $|\boldsymbol{D}\Delta\boldsymbol{D}'|\leqslant 1$。

拉普拉斯机制[28]是常用的实现差分隐私的技术。它通过向查询结果添加符合拉普拉斯分布的噪音值来实现差分隐私。对于函数 $f:\boldsymbol{D}\rightarrow\mathbf{R}^d$,拉普拉斯机制实现差分隐私保护的过程可表示如下:

$$\boldsymbol{A}(\boldsymbol{D})=f(\boldsymbol{D})+\left(Lap_1\left(\frac{\Delta f}{\varepsilon}\right),Lap_2\left(\frac{\Delta f}{\varepsilon}\right),\cdots,Lap_d\left(\frac{\Delta f}{\varepsilon}\right)\right)^{\mathrm{T}} \quad (4)$$

其中，$Lap_i\left(\frac{\Delta f}{\varepsilon}\right)$ 表示独立同分布的拉普拉斯噪声。Δf 越大，所需的拉普拉斯噪声也越大。由式(4)可知，$\boldsymbol{A}(\boldsymbol{D})$中第 $i(1\leqslant i\leqslant d)$ 个元素由拉普拉斯噪音引起的方差为：

$$E(\boldsymbol{A}_i(\boldsymbol{D})-f_i(\boldsymbol{D}))^2=\frac{2(\Delta f)^2}{\varepsilon^2} \quad (5)$$

4 轨迹流量图发布

流量图的发布过程包括两步：(1)生成差分隐私流量图，即统计路网中各路段的流量值并添加独立同分布的拉普拉斯噪音；(2)一致性调节，即对(1)中的流量图进行出入流量一致性调节。下面分别描述这两个过程。

4.1 差分隐私流量图

轨迹流量图中节点的流入流量表示从各个方向驶向该节点的车次，而流出流量表示所有驶离该节点的车次。通过观察分析不难发现，当节点不是任何轨迹的起止点时，该节点的出入流量相等。为此，本文引入一个虚拟节点 p_v，并让该虚拟节点作为所有轨迹的起止点，使轨迹变为首尾相连的环，且虚拟节点与路网中每个节点都相连。虚拟节点的引入使得流量图中所有节点(包括虚拟节点)的出入流量都相等且不会影响真实流量图中各边的统计值。没有特殊说明，下文中的路网邻接矩阵 $\boldsymbol{M}$ 和流量矩阵 $\boldsymbol{W}$ 中均包含了虚拟节点。

4.1.1 差分隐私流量图算法

使用差分隐私进行隐私保护前首先需要明确相邻数据集和全局敏感度。本文规定如果两个轨迹数据集 $\boldsymbol{D}$ 和 $\boldsymbol{D}'$ 仅相差一个位置点，则称两者为相邻数据集。因此，轨迹数据集 $\boldsymbol{D}$ 的相邻数据集 $\boldsymbol{D}'$(仍为 $\boldsymbol{\Omega}$ 的子集)可通过删除或替换 $\boldsymbol{D}$ 中某条轨迹的一个位置点得到。删除方式引起的流量改变值为3，替换方式引起的流量改变值为4。由定义3可知，发布一个流量图的全局敏感度为4，即 $\Delta f=4$。

差分隐私流量图算法描述见算法1。该算法的输入为轨迹数据集 $\boldsymbol{D}$，路网邻接矩阵 $\boldsymbol{M}$ 以及差分隐私预算 ε；输出为真实的轨迹流量邻接矩阵 $\boldsymbol{W}$ 以及满足差分隐私的流量矩阵 $\tilde{\boldsymbol{W}}$。算法1中首先初始化输出变量，2~5行对轨迹数据集进行处理，使每条轨迹的首尾与虚拟节点 p_v 相连，然后统计路网中每个路段的流量值；6、7行向流量图中存在边的流量值添加噪音扰动，这是由于流量图中不存在的边不可能有轨迹经过，对其添加噪声保护没有意义；最后将流量矩阵返回。接下来对算法1的隐私性和误差进行分析。

算法1 轨迹流量图发布

输入：轨迹数据集 $\boldsymbol{D}$，邻接矩阵 $\boldsymbol{M}$，隐私预算 ε。

输出：加噪前后的流量矩阵 $\boldsymbol{W}$、$\tilde{\boldsymbol{W}}$。

1. 初始化 $\boldsymbol{W}=\boldsymbol{O}$，$\tilde{\boldsymbol{W}}=\boldsymbol{O}$；
2. for each $L\in\boldsymbol{D}$
3. $\quad w_{p_v,p_1}\leftarrow w_{p_v,p_1}+1$，$w_{p_{|L|},p_v}\leftarrow w_{p_{|L|},p_v}+1$；
4. $\quad$ for each $\langle p_i,p_{i+1}\rangle\in L,i\in[1,|L|-1]$
5. $\quad\quad w_{p_i,p_{i+1}}\leftarrow w_{p_i,p_{i+1}}+1$；
6. for each $u\rightarrow v$，满足$m_{u,v}=1\quad L\in\boldsymbol{D}$
7. $\quad\tilde{w}_{u,v}=w_{u,v}+Lap(4/\varepsilon)$；
8. return $\boldsymbol{W},\tilde{\boldsymbol{W}}$；

4.1.2 差分隐私流量图算法分析

本节首先证明差分隐私流量图算法符合差分隐私的定义，即算法1可以保证用户的轨迹隐私；然后分析该算法中由于差分隐私噪声的添加所引起的发布误差。用 $P:\boldsymbol{D}\rightarrow\mathbf{R}^k$ 代表差分隐私流量图生成算法，输入为轨迹数据集 $\boldsymbol{D}$，输出为实数域上的 k 维向量，k 代表路网中所有的边数，其所有可能的输出结果为 $\boldsymbol{S}_P$。$\boldsymbol{D}'$ 与 $\boldsymbol{D}$ 是任意两个相邻的轨迹数据集，则对于任意的输出 $o(o_1,o_2,\cdots,o_k)\in\boldsymbol{S}_P$ 有：

$$\frac{\Pr[P(\boldsymbol{D})=o]}{\Pr[P(\boldsymbol{D}')=o]}=\prod_{i=1}^{k}\left(\frac{\exp\left(-\frac{\varepsilon|P_i(\boldsymbol{D})-o_i|}{\Delta f}\right)}{\exp\left(-\frac{\varepsilon|P_i(\boldsymbol{D}')-o_i|}{\Delta f}\right)}\right)$$

对其进行推导可得：

$$\prod_{i=1}^{k}\exp\left(\frac{\varepsilon\left(|P_i(\boldsymbol{D}')-o_i|-|P_i(\boldsymbol{D})-o_i|\right)}{\Delta f}\right)\leqslant\prod_{i=1}^{k}\exp\left(\frac{\varepsilon|P_i(\boldsymbol{D}')-P_i(\boldsymbol{D})|}{\Delta f}\right)=\exp\left(\frac{\varepsilon\|P(\boldsymbol{D}')-P(\boldsymbol{D})\|_1}{\Delta f}\right)\leqslant\exp(\varepsilon)$$

由前文关于差分隐私的定义可知，轨迹流量图发布算法符合差分隐私的定义。

接下来分析算法1中由于差分隐私噪声的添加引起的发布误差。根据式(5)可知，每添加一次拉普拉斯噪声将引起 $2(\Delta f)^2/\varepsilon^2=2(4/\varepsilon)^2$ 的均方误差。算法1对路网矩阵 $\boldsymbol{M}$ 的每条边均添加拉普拉斯噪声。因此，$\tilde{\boldsymbol{W}}$ 的总体均方误差为 $2(4/\varepsilon)^2\times|E|$，其中 $|E|$ 表示路网中所有的边数。由以上公式可知，算法1所发布的流量图 $\tilde{\boldsymbol{W}}$ 误差大小取决于路网中边的数量以及用户设定的隐私预算 ε，与轨迹数量没有关系。因此，在同一个路网中算法1所引起的误差不会随着轨迹数据集的改变而改变。

由于算法1对轨迹流量图中的每条边独立加噪，因此破坏了路网中节点出入流量相等的特性。针对这个问题，下面将通过二次规划求解，对算法1进行一致性调节。

4.2 一致性调节算法

如图2(b)所示，添加噪音扰动后的流量图大部分节点的出入流量不相等，虽然保护了用户的隐私，但是影响了发布数据的可用性。本节针对该问题提出算法2。对加噪后的流量图进行一致性调节，使得发布的流量图在满足差分隐私的同时具有较高的可用性。推导过程涉及了大量符号，符号说明见表1。

分别记一致性调节前后的流量矩阵为 $\tilde{\boldsymbol{W}}$ 和 $\bar{\boldsymbol{W}}$，其元素分别表示为 $\tilde{w}_{ij}$ 和 $\bar{w}_{ij}$。经分析发现，由 $\tilde{\boldsymbol{W}}$ 得到 $\bar{\boldsymbol{W}}$ 的过程可转化为求解凸优化表达式(6)：

$$\begin{cases}\min\limits_{\bar{\boldsymbol{W}}} f(\bar{\boldsymbol{W}})=\\ \qquad\dfrac{1}{2}\left(\sum\limits_{i\to j\text{存在}}(\bar{w}_{ij}-\tilde{w}_{ij})^2+\alpha^2\sum\limits_{i\to j\text{不存在}}(\bar{w}_{ij}-\tilde{w}_{ij})^2\right)\\ \text{s.t. }\bar{\boldsymbol{W}}^{\mathrm{T}}\boldsymbol{I}_{|V|+1}=\bar{\boldsymbol{W}}\boldsymbol{I}_{|V|+1}\end{cases}\tag{6}$$

其中，$\boldsymbol{I}_{|V|+1}$ 表示长度为 $|V|+1$(包含虚拟节点)且所有元素全为1的列向量。惩罚系数 α^2 ($\alpha>>1$) 的引入是使得路网中不存在的边上的流量值在一致性调节后趋于0。因为当边 $i\to j$ 不存在时 $\tilde{w}_{ij}=0$，若 $\bar{w}_{ij}\neq 0$，则子式 $\alpha^2\sum\limits_{i\to j\text{不存在}}(\bar{w}_{ij}-\tilde{w}_{ij})^2$ 的值会非常大，而式(6)为了取得最小值就会迫使 $\bar{w}_{ij}=\tilde{w}_{ij}$。极限情况下，当 $\alpha\to+\infty$ 时，有 $\bar{w}_{ij}\to 0$。为了式(6)表达更加紧凑，记惩罚系数矩阵为 $\boldsymbol{C}$，其元素满足 $c_{ij}=\begin{cases}\alpha, & \text{if } m_{i,j}=0\\ 1, & \text{if } m_{i,j}=1\end{cases}$。此时式(6)表示为：

$$\begin{cases}\min\limits_{\bar{\boldsymbol{W}}}\dfrac{1}{2}\text{trace}(((\bar{\boldsymbol{W}}-\tilde{\boldsymbol{W}})\odot\boldsymbol{C})^{\mathrm{T}}((\bar{\boldsymbol{W}}-\tilde{\boldsymbol{W}})\odot\boldsymbol{C}))\\ \text{s.t. }\bar{\boldsymbol{W}}^{\mathrm{T}}\boldsymbol{I}_{|V|+1}=\bar{\boldsymbol{W}}\boldsymbol{I}_{|V|+1}\end{cases}\tag{7}$$

式中，$\odot$ 为矩阵Hadamard积运算符。令 $\boldsymbol{X}=\bar{\boldsymbol{W}}\odot\boldsymbol{C}$，$\boldsymbol{X}'=\tilde{\boldsymbol{W}}\odot\boldsymbol{C}$，即 $\bar{\boldsymbol{W}}=\boldsymbol{X}\odot\boldsymbol{C}^{-1}$（$\boldsymbol{C}^{-1}$ 为 $\boldsymbol{C}$ 中的所有元素取倒数得到的矩阵）。式(7)改为如下的等价形式：

$$\begin{cases}\min\limits_{\boldsymbol{X}}\dfrac{1}{2}\text{trace}((\boldsymbol{X}-\boldsymbol{X}')^{\mathrm{T}}(\boldsymbol{X}-\boldsymbol{X}'))\\ \text{s.t. }(\boldsymbol{X}\odot\boldsymbol{C}^{-1})^{\mathrm{T}}\boldsymbol{I}_{|V|+1}=(\boldsymbol{X}\odot\boldsymbol{C}^{-1})\boldsymbol{I}_{|V|+1}\end{cases}\tag{8}$$

当 $\alpha\to+\infty$ 时，$\boldsymbol{C}^{-1}\to\boldsymbol{M}$。因此，可使用邻接矩阵 $\boldsymbol{M}$ 来代替 $\boldsymbol{C}^{-1}$。同时将式(8)中的 $\odot$ 运算部分利用普通乘法公式进行代换，得到与之等价的表达式(9)：

$$\begin{cases}\min\limits_{\boldsymbol{X}}\dfrac{1}{2}\text{trace}((\boldsymbol{X}-\boldsymbol{X}')^{\mathrm{T}}(\boldsymbol{X}-\boldsymbol{X}'))\\ \text{s.t. }\left(\sum\limits_i(\boldsymbol{D}_i\boldsymbol{X}^{\mathrm{T}}\mathbb{E}_i)\right)\boldsymbol{I}_{|V|+1}=\left(\sum\limits_i(\mathbb{E}_i\boldsymbol{X}\boldsymbol{D}_i)\right)\boldsymbol{I}_{|V|+1}\end{cases}\tag{9}$$

其中，$\boldsymbol{D}_i$ 表示取 $\boldsymbol{M}$ 的第 i 行向量并将其对角化后的常数矩阵；$\mathbb{E}_i$ 为第 i 行以及第 i 列为1其他全为0的对角阵。此时 $\boldsymbol{X}\to\bar{\boldsymbol{W}}$，$\boldsymbol{X}'\to\tilde{\boldsymbol{W}}$。因此，式(9)解出的 $\boldsymbol{X}$ 即为式(6)的解 $\bar{\boldsymbol{W}}$。

接下来，为了求解式(9)，使用拉格朗日乘子法将其转化为无约束的对偶问题。$\boldsymbol{L}(\boldsymbol{u},\bar{\boldsymbol{W}})$ 表示如下：

$$\begin{aligned}\boldsymbol{L}(\boldsymbol{u},\boldsymbol{X})=&\frac{1}{2}\text{trace}((\boldsymbol{X}-\boldsymbol{X}')^{\mathrm{T}}(\boldsymbol{X}-\boldsymbol{X}'))+\\ &\boldsymbol{u}^{\mathrm{T}}\left(\sum_i(\boldsymbol{D}_i\boldsymbol{X}^{\mathrm{T}}\mathbb{E}_i)-\sum_i(\mathbb{E}_i\boldsymbol{X}\boldsymbol{D}_i)\right)\boldsymbol{I}_{|V|+1}\end{aligned}\tag{10}$$

式中，$\boldsymbol{u}$ 为拉格朗日系数组成的长度为 $|V|+1$ 的列向量。对其化简可得：

$$\begin{aligned}\boldsymbol{L}(\boldsymbol{u},\boldsymbol{X})=&\frac{1}{2}\text{trace}((\boldsymbol{X}-\boldsymbol{X}')^{\mathrm{T}}(\boldsymbol{X}-\boldsymbol{X}'))+\\ &\text{trace}(\boldsymbol{M}\,\text{diag}(\boldsymbol{u})\boldsymbol{X}^{\mathrm{T}}-\boldsymbol{M}^{\mathrm{T}}\,\text{diag}(\boldsymbol{u})\boldsymbol{X})\end{aligned}\tag{11}$$

将式(11)对 $\boldsymbol{X}$ 求导，得到：

$$\frac{\partial\boldsymbol{L}(\boldsymbol{u},\boldsymbol{X})}{\partial\boldsymbol{X}}=\boldsymbol{X}-\boldsymbol{X}'+\boldsymbol{M}\,\text{diag}(\boldsymbol{u})-\text{diag}(\boldsymbol{u})\boldsymbol{M}$$

令 $\dfrac{\partial\boldsymbol{L}(\boldsymbol{u},\boldsymbol{X})}{\partial\boldsymbol{X}}=\boldsymbol{O}$，得到式(11)取最小值时的解：$\boldsymbol{X}^*=\boldsymbol{X}'-\boldsymbol{M}\,\text{diag}(\boldsymbol{u})+\text{diag}(\boldsymbol{u})\boldsymbol{M}$，并将其代回式(11)并进行一系列化简有：

$$\boldsymbol{L}(\boldsymbol{u},\boldsymbol{X}^*)=-\frac{1}{2}(\boldsymbol{u}^{\mathrm{T}}(\mathrm{diag}((\boldsymbol{M}+\boldsymbol{M}^{\mathrm{T}})\boldsymbol{I}_{|V|+1}-(\boldsymbol{M}+\boldsymbol{M}^{\mathrm{T}})))\boldsymbol{u})+\boldsymbol{u}^{\mathrm{T}}(\boldsymbol{X}'^{\mathrm{T}}-\boldsymbol{X}')\boldsymbol{I}_{|V|+1} \tag{12}$$

根据拉格朗日系数方法，$\boldsymbol{L}(\boldsymbol{u},\boldsymbol{X}^*)$ 取得最大值的解即为式(6)的解。

接下来，求式(12)的最大值。不难发现该式为无约束的二次规划问题，令其二次项系数矩阵 $\mathrm{diag}((\boldsymbol{M}+\boldsymbol{M}^{\mathrm{T}})\boldsymbol{I}_{|V|+1}-(\boldsymbol{M}+\boldsymbol{M}^{\mathrm{T}}))$ 为 $\boldsymbol{Q}$，$\boldsymbol{Q}$ 为拉普拉斯矩阵，即不可求逆的半正定矩阵，由于 $\boldsymbol{M}$ 为连通图的邻接矩阵，因而其秩等于邻接矩阵的点数减去1。对式(12)求导可得：$\dfrac{\partial \boldsymbol{L}(\boldsymbol{u},\boldsymbol{X}^*)}{\partial \boldsymbol{u}}=\boldsymbol{Q}\boldsymbol{u}+(\boldsymbol{X}'^{\mathrm{T}}-\boldsymbol{X}')\boldsymbol{I}_{|V|+1}$。根据 $\boldsymbol{Q}$ 的秩的性质，可采用将 $\boldsymbol{u}$ 的最后一个元素设为0求解 $\dfrac{\partial \boldsymbol{L}(\boldsymbol{u},\boldsymbol{X}^*)}{\partial \boldsymbol{u}}=\boldsymbol{O}$。记 $\boldsymbol{u}=(\boldsymbol{u}'^{\mathrm{T}}\ \ 0)^{\mathrm{T}}=\begin{bmatrix}\boldsymbol{E}_{|V|}\\ \boldsymbol{o}_{|V|}^{\mathrm{T}}\end{bmatrix}\boldsymbol{u}'$。同时由于路网邻接矩阵 $\boldsymbol{M}$ 中的虚拟节点 p_v 与其他节点相连，将 $\boldsymbol{Q}$ 做如下分块 $\boldsymbol{Q}=\begin{bmatrix}\boldsymbol{P} & \boldsymbol{b}\\ \boldsymbol{b}^{\mathrm{T}} & c\end{bmatrix}=\begin{bmatrix}\boldsymbol{P} & -2\boldsymbol{I}_{|V|}\\ -2\boldsymbol{I}_{|V|}^{\mathrm{T}} & 2|V|\end{bmatrix}$，代入式(12)可得：

$$\boldsymbol{L}(\boldsymbol{u}',\bar{\boldsymbol{W}})=-\boldsymbol{u}'^{\mathrm{T}}\boldsymbol{P}\boldsymbol{u}'+\boldsymbol{u}'^{\mathrm{T}}[\boldsymbol{E}_{|V|}\ \ \boldsymbol{o}_{|V|}](\boldsymbol{X}'^{\mathrm{T}}-\boldsymbol{X}')\boldsymbol{I}_{|V|+1} \tag{13}$$

$\boldsymbol{P}$ 为式(13)的二次项系数矩阵，对其分析可知该系数矩阵为正定矩阵。具体分析如下：

$\boldsymbol{Q}$ 为半正定矩阵且 $\boldsymbol{Q}\boldsymbol{I}_{|V|+1}=\begin{bmatrix}\boldsymbol{P}\boldsymbol{I}_{|V|}-2\boldsymbol{I}_{|V|}\\ 0\end{bmatrix}=\boldsymbol{o}_{|V|+1}$，得 $\boldsymbol{P}\boldsymbol{I}_{|V|}-2\boldsymbol{I}_{|V|}=(\boldsymbol{P}-2\boldsymbol{E}_{|V|})\boldsymbol{I}_{|V|}=\boldsymbol{o}_{|V|}$。令 $\boldsymbol{P}'=\boldsymbol{P}-2\boldsymbol{E}_{|V|}$，可知 $\boldsymbol{P}'$ 也是拉普拉斯矩阵，具有半正定性。因此，将任意非零列向量 $\boldsymbol{x}$ 代入下式有：

$$\boldsymbol{x}^{\mathrm{T}}\boldsymbol{P}\boldsymbol{x}=\boldsymbol{x}^{\mathrm{T}}(\boldsymbol{P}'+2\boldsymbol{E}_{|V|})\boldsymbol{x}=\boldsymbol{x}^{\mathrm{T}}\boldsymbol{P}'\boldsymbol{x}+2\boldsymbol{x}^{\mathrm{T}}\boldsymbol{x}\geqslant 2\boldsymbol{x}^{\mathrm{T}}\boldsymbol{x}>0 \tag{14}$$

因此，$\boldsymbol{P}$ 为正定矩阵。式(13)可直接求解得：

$$\boldsymbol{u}'=\boldsymbol{P}^{-1}[\boldsymbol{E}_{|V|}\ \ \boldsymbol{o}_{|V|}](\boldsymbol{X}'^{\mathrm{T}}-\boldsymbol{X}')\boldsymbol{I}_{|V|+1} \tag{15}$$

式(15)中求解 $\boldsymbol{u}'$ 的过程实际上就等价于解方程组 $\boldsymbol{P}\boldsymbol{u}'=[\boldsymbol{E}_{|V|}\ \ \boldsymbol{o}_{|V|}](\boldsymbol{X}'^{\mathrm{T}}-\boldsymbol{X}')\boldsymbol{I}_{|V|+1}$。由于邻接矩阵节点数众多，而且极度稀疏。为使得式(15)具有更高效的求解效率，本文采用了PCG(preconditioned conjugate gradients method)。该算法是现有解决大规模稀疏方程组最有效的方法之一。由式(15)求得的结果 $\boldsymbol{u}'$ 即可得到 $\boldsymbol{u}$，由此求得一致性调节后的流量矩阵 $\bar{\boldsymbol{W}}$ 及最小误差分别为：

$$\bar{\boldsymbol{W}}=\tilde{\boldsymbol{W}}-\boldsymbol{M}\,\mathrm{diag}(\boldsymbol{u})+\mathrm{diag}(\boldsymbol{u})\boldsymbol{M} \tag{16}$$

$$err=\min f(\bar{\boldsymbol{W}})=\frac{1}{2}sum((\bar{\boldsymbol{W}}-\tilde{\boldsymbol{W}})\odot(\bar{\boldsymbol{W}}-\tilde{\boldsymbol{W}})) \tag{17}$$

根据上述推导，最终得到一致性调节算法。

算法2 一致性调节算法

输入：添加噪声保护的流量图 $\tilde{\boldsymbol{W}}$，路网矩阵 $\boldsymbol{M}$。

输出：一致性调节后的流量图 $\bar{\boldsymbol{W}}$。

1. 令 $n=size(\boldsymbol{M})$ 表示路网矩阵的节点规模；
2. 令 $\boldsymbol{Q}=\mathrm{diag}((\boldsymbol{M}+\boldsymbol{M}^{\mathrm{T}})\boldsymbol{I}_n-(\boldsymbol{M}+\boldsymbol{M}^{\mathrm{T}}))$；
3. $\boldsymbol{P}\leftarrow\boldsymbol{Q}(1{:}n-1,1{:}n-1)$;//取 $\boldsymbol{Q}$ 的前 $n-1$ 行和 $n-1$ 列
4. $\boldsymbol{b}\leftarrow(\tilde{\boldsymbol{W}}^{\mathrm{T}}-\tilde{\boldsymbol{W}})\boldsymbol{I}_n$；
5. $\boldsymbol{b}\leftarrow\boldsymbol{b}(1{:}n-1)$；
6. 采用PCG方法求解方程 $\boldsymbol{P}\boldsymbol{u}'=\boldsymbol{b}$，并求得最优解 $\boldsymbol{u}^*=[\boldsymbol{u}'^*\ \ 0]$；
7. 求得 $\bar{\boldsymbol{W}}=\tilde{\boldsymbol{W}}-\boldsymbol{M}\,\mathrm{diag}(\boldsymbol{u}^*)+\mathrm{diag}(\boldsymbol{u}^*)\boldsymbol{M}$；
8. $\bar{\boldsymbol{W}}=\bar{\boldsymbol{W}}(1{:}n-1,1{:}n-1)$;//去除虚拟节点
9. return $\bar{\boldsymbol{W}}$；

为保证算法效率，上述算法中的矩阵均采用系数矩阵存储并采用相关算法进行解释，具体过程文本不做进一步讨论。算法2的实现过程中采用了稀疏矩阵的数据结构及相关算法。关于稀疏矩阵的算法已经有诸多研究以及API的支持。本文在第5章提供了该算法的实现源码下载网址，供读者参考。

5 实验分析

本章主要从效用和执行效率两方面对本文提出的算法进行验证。实验的硬件环境为：Intel®Core™ i7-6700 CPU@3.40 GHz，16 GB内存；用Matlab实现。实验中所用的路网分别来自德国奥尔登堡(Oldenburg)市，美国的圣华金(San Joaquin)以及旧金山湾区(San Francisco Bay Area)三个城市，相应的三个城市中的轨迹数据由Brinkhoff基于路网的轨迹生成器生成。Brinkhoff轨迹生成器和三个城市的路网信息数据可以从网站 http://iapg.jade-hs.de/personen/brinkhoff/generator/下载。本实验中所用到的三个城市的路网信息以及相应路网上的轨迹详细信息如表2所示。

Table 2 Details of road network and trajectory

表2 路网及轨迹信息

路网	边数	路口数	轨迹数	平均轨迹长度
Oldenburg	7 035	14 058	54 792	45
San Joaquin	24 123	48 061	39 554	68
San Francisco	223 606	444 503	98 048	312

其中,边的方向不同视为不同的边,路口数表示路网中交叉路口的数量。本章所涉及到的主要实验源码及部分供验证的数据已经上传到 http://matweb.applinzi.com/paperCode/,供读者下载参考。本章实验包括三部分:(1)分析一致性调节前后算法 1 的相对误差;(2)采用标准误差具体分析一致性调节对算法 1 的优化情况,标准误差采用矩阵 Frobenius 范数计算,即 $err=\|\tilde{\boldsymbol{W}}-\boldsymbol{W}\|_F$;(3)对算法 1 和算法 2 的耗时情况进行分析。下面分别介绍这三部分。

首先,采用奥尔登堡和圣华金的数据分别分析当隐私预算 ε 取 1 时不同轨迹规模下一致性调节前后的相对误差情况。相对误差通过公式 $\delta=\frac{err}{\sum_{i,j}w_{ij}}$ 求得。实验结果如图 3 所示。横坐标为轨迹规模,纵坐标为相对误差。从图中可以看出,同一路网中,随着轨迹规模的增大,相对误差逐渐减小,这是因为轨迹越多路网上流量值越大,相比之下噪音的影响就微乎其微。另外可发现,经过一致性调节后的相对误差明显比调节前的小。接下来具体分析一致性调节算法的优化程度。

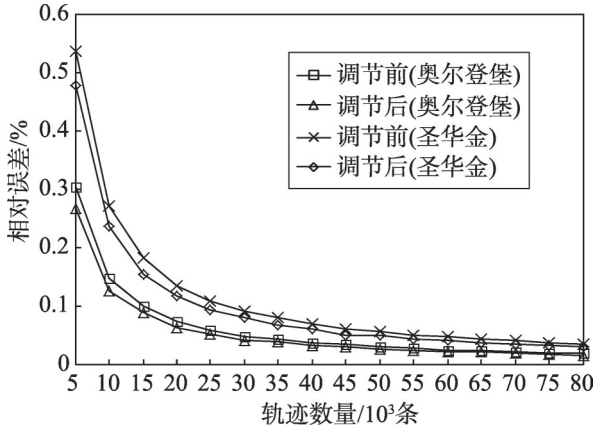


Fig.3 Relative error

图 3 相对误差

为验证算法 2 对流量图发布精度的提升情况,分别计算当 ε 取 0.5、1.0、2.0、5.0 时,三个城市一致性调节前后流量的标准误差,其结果如图 4~图 6 所示。随着 ε 的增大,添加的噪音减小,相应的整体误差减小;随着路网规模的增加,添加噪音的次数增加,相应的整体误差增加,与 4.1.2 节的误差分析结果一

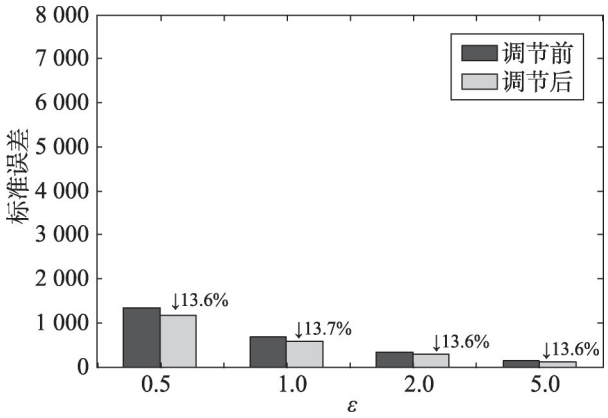


Fig.4 Error of trajectory flow before and after consistency adjustment in Oldenburg

图 4 奥尔登堡市一致性调节前后轨迹流量误差

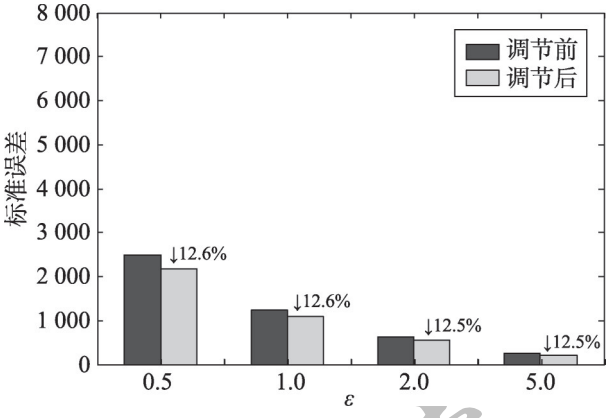


Fig.5 Error of trajectory flow before and after consistency adjustment in San Joaquin

图 5 圣华金市一致性调节前后轨迹流量误差

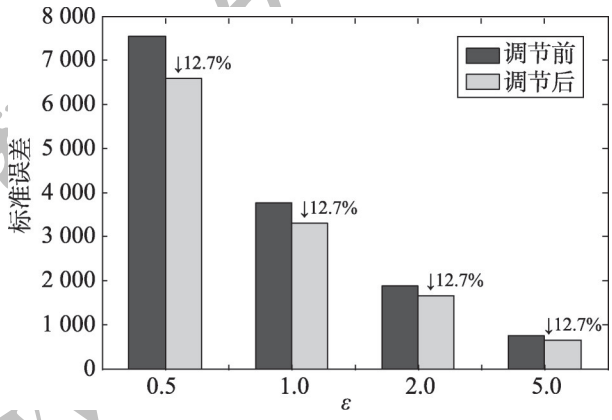


Fig.6 Error of trajectory flow before and after consistency adjustment in San Francisco

图 6 旧金山一致性调节前后轨迹流量误差

致。另外可以看出,通过算法2的优化,整体误差减小了12%~14%,且减小的程度随 ε 的改变无明显变化,这说明算法2在不同的 ε 下是稳定的。

最后对算法1和算法2的耗时情况进行分析,仍以奥尔登堡市、圣华金以及旧金山湾区三个城市路网数据进行实验,实验结果如图7。结合表2可看出根据不同城市路网节点数、边数以及相应的轨迹规模的不同,算法1的耗时相差较大。尤其是路网规模最大的旧金山湾区,由于处理的轨迹数量为98 048条且平均轨迹长度为312,因此耗时达到了340 s左右。而算法2在优化上述三个城市的轨迹流量时用时均不到1 s。因此,本文所提出的一致性调节方法在提升轨迹流量发布效果的同时具有较高的发布效率且具有处理较大规模路网的能力。

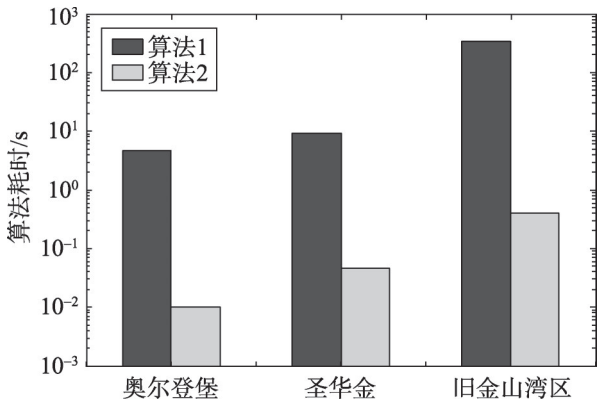


Fig.7 Run time

图7 算法耗时

6 结束语

对蕴含路网信息的轨迹大数据进行分析,并发布一段时间内路网上车流量的统计值,可为现实中的很多应用提供参考信息。本文介绍了在差分隐私机制下发布基于路网的轨迹流量信息过程,并通过求解二次规划问题实现了流量图的后置处理算法。实验证明该一致性调节算法不仅提高了发布数据的精度,而且可处理较大规模的路网数据。但是,由于本文算法是针对于静态的轨迹流量图发布所设计的,无法满足现实中需要实时发布的应用场景。因此,如何将本文所涉及的算法与轨迹流量图发布的实时性相结合是下一步的研究方向。

References:

[1] Giannotti F, Nanni M, Pinelli F, et al. Trajectory pattern mining[C]//Proceedings of the 13th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, California, Aug 12-15, 2007. New York: ACM, 2007: 330-339.

[2] Xu Ming. Research on key issues of road network through large-scale trajectory data mining[D]. Beijing: Beijing University of Posts and Telecommunications, 2015.

[3] Jiang Na. Research on the influence of road networks structural characteristics to the distribution of vehicle flow[D]. Chengdu: Southwest Jiaotong University, 2015.

[4] Gao Qiang, Zhang Fengli, Wang Ruijin, et al. Trajectory big data: a review of key technologies in data processing [J]. Journal of Software, 2017, 28(4): 959-992.

[5] Porta S, Crucitti P, Latora V. The network analysis of urban streets: a primal approach[J]. Environment and Planning B: Planning and Design, 2006, 33(5): 705-725.

[6] You T H, Peng W H, Lee W C. Protecting moving trajectories with dummies[C]//Proceedings of the 8th International Conference on Mobile Data Management, Mannheim, May 7-11, 2007. Piscataway: IEEE, 2008: 278-282.

[7] Lei Kaiyue, Li Xinghua, Liu Hai, et al. Dummy trajectory privacy protection scheme for trajectory publishing based on the spatiotemporal correlation[J]. Journal on Communications, 2016, 37(12): 156-164.

[8] Sweeney L. *k*-anonymity: a model for protecting privacy[J]. International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems, 2002, 10(5): 557-570.

[9] Machanavajjhala A, Kifer D, Gehrke J, et al. L-diversity: privacy beyond k-anonymity[J]. ACM Transactions on Knowledge Discovery from Data, 2007, 1(1): 3.

[10] Li Ninghui, Li Tiancheng, Venkatasubramanian S. t-closeness: privacy beyond *k*-anonymity and *l*-diversity[C]//Proceedings of the 23rd International Conference on Data Engineering, Istanbul, Apr 15-20, 2007. Washington: IEEE Computer Society, 2007: 106-115.

[11] Poulis G, Skiadopoulos S, Loukides G, et al. Distance-based k^m-anonymization of trajectory data[C]//Proceedings of the 14th International Conference on Mobile Data Management, Milan, Jun 3-6, 2013. Washington: IEEE Computer

Society, 2013: 57-62.

[12] Abul O, Bonchi F, Nanni M. Never walk alone: uncertainty for anonymity in moving objects databases[C]//Proceedings of the 24th International Conference on Data Engineering, Cancún, Apr 7-12, 2008. Washington: IEEE Computer Society, 2008: 376-385.

[13] Hoh B, Gruteser M, Xiong Hui, et al. Achieving guaranteed anonymity in GPS traces via uncertainty-aware path cloaking [J]. IEEE Transactions on Mobile Computing, 2010, 9(8): 1089-1107.

[14] Terrovitis M, Mamoulis N. Privacy preservation in the publication of trajectories[C]//Proceedings of the 9th International Conference on Mobile Data Management, Beijing, Apr 27-30, 2008. Piscataway: IEEE, 2008: 65-72.

[15] Zhao Jing, Zhang Yuan, Li Xinghua, et al . A trajectory privacy protection approach via trajectory frequency suppression[J]. Chinese Journal of Computers, 2014, 37(10): 2096-2106.

[16] Dwork C. Differential privacy: a survey of results[C]// LNCS 4978: Proceedings of the 5th International Conference on Theory and Applications of Models of Computation, Xi'an, Apr 25-29, 2008. Berlin, Heidelberg: Springer, 2008: 1-19.

[17] Chen Rui, Fung B, Desai B. Differentially private trajectory data publication[J]. arXiv:1112.2020, 2011.

[18] Chen Rui, Ács G, Castelluccia C. Differentially private sequential data publication via variable-length *n*-grams[C]// Proceedings of the ACM Conference on Computer and Communications Security, Raleigh, Oct 16-18, 2012. New York: ACM, 2012: 638-649.

[19] Jiang Kaifeng, Shao Dongxu, Bressan S, et al. Publishing trajectories with differential privacy guarantees[C]//Proceedings of the 25th International Conference on Scientific and Statistical Database Management, Baltimore, Jul 29-31, 2013. New York: ACM, 2013: 12.

[20] Hua Jingyu, Gao Yue, Zhong Sheng. Differentially private publication of general time-serial trajectory data[C]//Proceedings of the 2015 IEEE Conference on Computer Communications, Kowloon, Hong Kong, China, Apr 26-May 1, 2015. Piscataway: IEEE, 2015: 549-557.

[21] Wang S, Sinnott R. Protecting personal trajectories of social media users through differential privacy[J]. Computers & Security, 2017, 67: 142-163.

[22] Wang Hao, Xu Zhengquan. CTS-DP: publishing correlated time-series data via differential privacy[J]. Knowledge Based Systems, 2017, 122: 167-179.

[23] Li Meng, Zhu Liehuang, Zhang Zijian, et al. Achieving differential privacy of trajectory data publishing in participatory sensing[J]. Information Sciences, 2017, 400: 1-13.

[24] Cao Yang, Yoshikawa M. Differentially private real-time data release over infinite trajectory streams [C]//Proceedings of the 16th IEEE International Conference on Mobile Data Management, Pittsburgh, Jun 15-18, 2015. Washington: IEEE Computer Society, 2015: 68-73.

[25] Zhang Xiaojian, Meng Xiaofeng. Differential privacy in data publication and analysis[J]. Chinese Journal of Computers, 2014, 37(4): 927-949.

[26] Hay M, Rastogi V, Miklau G, et al. Boosting the accuracy of differentially private histograms through consistency[J]. Proceedings of the VLDB Endowment, 2010, 3(1/2): 1021-1032.

[27] Gramaglia M, Fiore M, Tarable A, et al. Preserving mobile subscriber privacy in open datasets of spatiotemporal trajectories[C]//IEEE INFOCOM 2017- IEEE Conference on Computer Communications, Atlanta, May 1-4, 2017. Piscataway: IEEE, 2017: 1-9.

[28] Dwork C. A firm foundation for private data analysis[J]. Communications of the ACM, 2011, 54(1): 86-95.

附中文参考文献:

[2] 许明. 基于车辆轨迹挖掘的城市路网分析关键问题研究[D]. 北京: 北京邮电大学, 2015.

[3] 江娜. 道路网衔接结构对机动车交通流量分布影响研究[D]. 成都: 西南交通大学, 2015.

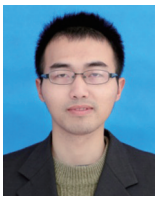
[4] 高强, 张凤荔, 王瑞锦, 等. 轨迹大数据: 数据处理关键技术研究综述[J]. 软件学报, 2017, 28(4): 959-992.

[7] 雷凯跃, 李兴华, 刘海, 等. 轨迹发布中基于时空关联性的假轨迹隐私保护方案[J]. 通信学报, 2016, 37(12): 156-164.

[15] 赵婧, 张渊, 李兴华, 等. 基于轨迹频率抑制的轨迹隐私保护方法[J]. 计算机学报, 2014, 37(10): 2096-2106.

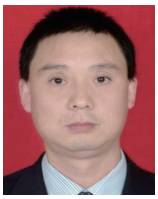
[25] 张啸剑, 孟小峰. 面向数据发布和分析的差分隐私保护[J]. 计算机学报, 2014, 37(4): 927-949.

ZHANG Shuangyue was born in 1990. She is an M.S. candidate at College of Computer Science, Shaanxi Normal University. Her research interest is differential privacy.
张双越(1990—),女,河北人,陕西师范大学计算机科学学院硕士研究生,主要研究领域为差分隐私。



CAI Jianping was born in 1990. He received the M.S. degree form Fuzhou University in 2016. His research interest is differential privacy.
蔡剑平(1990—),男,福建漳州人,2016年于福州大学数学与计算机科学学院获得硕士学位,主要研究领域为差分隐私。

TIAN Feng was born in 1987. He received the Ph.D. degree from Xi'an Jiaotong University in 2015. Now he is a lecturer at Shaanxi Normal University. His research interests include location privacy protection, privacy protection in data mining.
田丰(1987—),男,陕西人,2015年于西安交通大学获得博士学位,现为陕西师范大学讲师,主要研究领域为位置隐私保护,隐私保护的数据挖掘。



WU Zhenqiang was born in 1968. He received the Ph.D. degree from Xidian University in 2007. Now he is a professor at Shaanxi Normal University. His research interests include network science, network security and privacy protection.
吴振强(1968—),男,陕西人,2007年于西安电子科技大学获得博士学位,现为陕西师范大学教授,主要研究领域为网络科学,网络安全,隐私保护。