北京大学

本科生毕业论文

题目：基于社交媒体大数据的游客出行空间模式研究

姓　　名：唐启浩

学　　号：1400012415

院　　系：地球与空间科学学院

专　　业：地理信息科学

指导教师：刘　　瑜

二〇一八 年 六 月

摘要

近年以来，随着生活水平逐步提高，交通方式的逐渐便利，旅游出行不再局限于高收入人群，普通人群中也逐渐流行起来。于是旅游行业也逐渐发展起来。在这样的环境下，不管从城市的旅游产业的发展还是有对游客本身出行体验的角度考虑，对于游客旅游出行的研究就变得十分有意义。而游客的出行空间模式的研究，对于城市和景点对于自身的认知、发展和决策有指导作用，对于游客的出行路线的定制也能提供帮助。

而游客出行的空间模式理论也已经提出了很多年，有不少的学者都提出了自己的理论模型，有的学者也进行实证研究。但这些研究的时间比较久远而且研究所使用的数据基本都是来自于某个城市或者到某个景点游览的游客。数据量十分少，而且在收集数据的过程中，游客的主观因素对于数据的收集也有很大的影响。

随着智能手机的普及以及社交媒体的快速发展，游客的旅游数据在社交媒体上的体现也逐渐频繁。而大数据时代的到来，社交媒体产生的数据量十分客观，只要进行比较深度的挖掘，可以将用户的旅游行为提取出来。因此，使用社交媒体大数据进行游客的出行空间模式研究具有一定的意义。

在本文的研究中，目的在于基于社交媒体大数据对于游客的出行空间模式进行研究，并于学者提出的模型理论进行比较。并使用大数据对游客的出行模式进行利用，从城市和游客的角度探究出行模式对于旅游能带来的作用。从实验的结果来看，之前学者提出的 LCF 模型和城市内微观尺度的游客出行模式得到了一定程度上的验证。两个具体的应用效果也具有比较好的指导效果。

本文综述了学者们已经提出的出行空间模式理论，给出了利用社交媒体大数据对旅游中的出行空间模式进行研究的一个方法，使用微博数据进行实证研究，取得了比较好的结果。后续，可以基于这种处理方法，运用到其他的旅游理论中去；也可以结合更多的数据源，对结果进行更加准确的判定。

关键词：旅游，社交媒体，大数据，出行空间模式，motif

Research on Travel Spatial Patterns of Tourists Based on Social Media Big Data

Qihao Tang (Geographic Information Science)

Directed by Yu Liu

ABSTRACT

In recent years, with the gradual improvement of living standards and the gradual facilitation of transportation, travel is no longer limited to high-income people, and the general population has become increasingly popular. So the tourism industry has gradually developed. Under such circumstances, regardless of the development of the tourism industry in the city or the travel experience of the tourists themselves, the study of tourist travel has become very meaningful. The research on the travel space model of tourists can guide the recognition, development and decision-making of the city and its attractions, as well as the customization of travel routes for tourists.

The spatial mode theory of tourist travel has also been proposed for many years. Many scholars have put forward their own theoretical models, and some scholars have also conducted empirical research. However, these studies have taken a long time and the data used by the institute are basically from tourists visiting a certain city or visiting a certain attraction. The amount of data is very small, and in the process of collecting data, subjective factors of tourists also have a great influence on the collection of data.

With the popularization of smart phones and the rapid development of social media, the travel data of tourists has been increasingly reflected in social media. With the advent of the era of big data, the amount of data generated by social media is very objective. As long as relatively deep excavation is conducted, the user's travel behavior can be extracted. Therefore, the use of social media big data to carry out tourist travel space model has a certain significance.

In the study of this paper, the purpose is to study the travel space model of tourists based on social media big data, and to compare the model theory proposed by scholars. And use big data to make use of the travel mode of tourists, from the perspective of cities and tourists to explore the role of travel mode for tourism. From the results of the experiment, the LCF model proposed by previous scholars and the micro-scale tourist travel mode in cities have been verified to some extent. Two specific application effects also have better guidance effects.

This article reviews the travel space model theory that scholars have put forward, and gives a method of using the social media big data to study travel space modes in tourism. It uses microblogging data to conduct empirical research and has achieved relatively good results. Follow-up, based on this approach, can be applied to other tourism theories; more data sources can also be combined to more accurately determine the results.

Keywords: tourism, social media, big data, travel space model, motif

目录

第一章 概述

1.1 研究背景与意义

游客出行空间模式是描述游客在旅游活动空间活动规律的模式，描述了游客从其常居地到一个或者多个目的地游憩并最终返回常居地的一个出行路线的空间模式。长时间以来人们就认识到了该问题的重要性，但在旅游研究中很少有学者关注，也很少有学者进行实证或者概念上的研究或模拟。而这些已经进行了的模式研究，大多采用的是游客地图问卷的形式来进行研究，这在数据源上来说有其局限性，比如游客的游览范围小，游客随机性弱，数据量少等。随着大数据时代的到来，这些数据上的问题都有了比较好的解决方案，比如，通过社交媒体大数据平台获取的大量数据，从中将游客的数据提取出来，对其旅游行为以及路线进行分析，归纳聚类总结出的旅游空间模式可能更具有说服力。

游客出行空间模式研究的重要性：是研究旅游空间结构的重要构成，游客的出行模式对于旅游区空间的组织和管理具有指导意义，有利于旅游区对自身功能有更深的理解，从而做出调整，使空间结构得到高效利用。

和以往使用问卷等小数据量的研究形式不同，大数据时代的到来给验证游客出行空间模式提供了新的思路。社交媒体平台，如微博、QQ、微信等，能够提供游客的许多信息，通过信息挖掘，可以找出这些信息中包含的诸如用户的位置、时间等信息。而通过这些信息，我们可以将用户的轨迹构建出来，并通过一定的途径验证游客出行的空间模式。使用大数据进行分析，比传统使用调研问卷的方式更加合理，更有说服力，另外，充足的数据量可以支持我们进行更有深度的研究。

因此，在大数据时代中使用新的方式对游客出行空间模式进行研究与验证，具有较强的可行性。而研究的结果又可以应用到一些基于游客的路线和旅游城市的研究中。比如，研究结果对分析游客出行中关键节点和可能路径有很大的帮助，

对于节点自身的发展和规划设计都有重要的指示作用。因此，在大数据时代应该用新的方式对于游客的出行空间模式进行研究，为游客的目的地的市场定位、规划设计、宣传和营销提供参考；而对于游客而言，更加有助于他们对自己旅游路线的规划、增强对旅游时间和地点的把控。简而言之，在大数据时代的游客出行空间模式研究具有极大的应用价值，对于增强空间资源的合理规划与使用和提高游客的出行体验均有重要作用。

1.2 研究现状

游客的出行空间模式很早就已经有人进行探讨，在不同的尺度上，游客的出行模式有其差异。在宏观尺度上，在 1967 年，Campbell 根据目的地类型提出了回路中的游憩与度假旅行的模型。随后，越来越多的学者开始在这方面进行研究。Gunn 认识到不同类型的旅游对于旅游出行模式由较大的影响，提出了单目的地和往返式两种旅游模式。到 1993 年，Lue、Crompton 和 Fesenmaier 总结了五种度假旅行模型，分别包括了单目的地模式、往返模式、营区基地模式、区域旅游和旅行链模式，这也就是 LCF 模式。后来 Oppermann 在其基础上进一步细化，提出了 7 种旅行模式。这些模型被后续的研究者用于国际间游客的出行空间模式的研究。在微观尺度上，旅游出行模式的研究主要体现在城市内部的景点间的旅游路径上。城市内的游客出行模式和城市间以及国际间的模式有着较大的差异。城市间和国际间的关注点主要是区域，而城市内部则主要关注于景点等吸引物。Lew 等人提出了包括了点对点、环游、复杂等模式。

学者们在已有研究的基础上逐渐完善游客出行的空间模式的理论体系，这些体系从总体上来看都是比较接近的。目前在城市尺度上，学者们比较认可的游客出行模式应该就是 LCF 模式，这个模式可以在很大程度上表示游客的出行空间模式。

LCF 模式中的五种模式如下：

模式 1：单一目的地旅游——旅游者的大部分旅游活动集中在一个目的

地;

模式 2：线型旅游——旅游者选择使用一条线路上的多个旅游目的地,但存在主次之分,主要选择使用的目的地只有一个;

模式 3：基营式旅游——旅游者在访问主要目的地的同时也选择访问其他几个目的地,但往往以主体目的地作为大本营;

模式 4：环型旅游——旅游者在既定的目标区域内环旅游好几个目的地,相当于游览线路空间;

模式 5：链式旅游——旅游者以客源地为中心进行的链式游览。

直观的使用图像表示如下图 1-1：

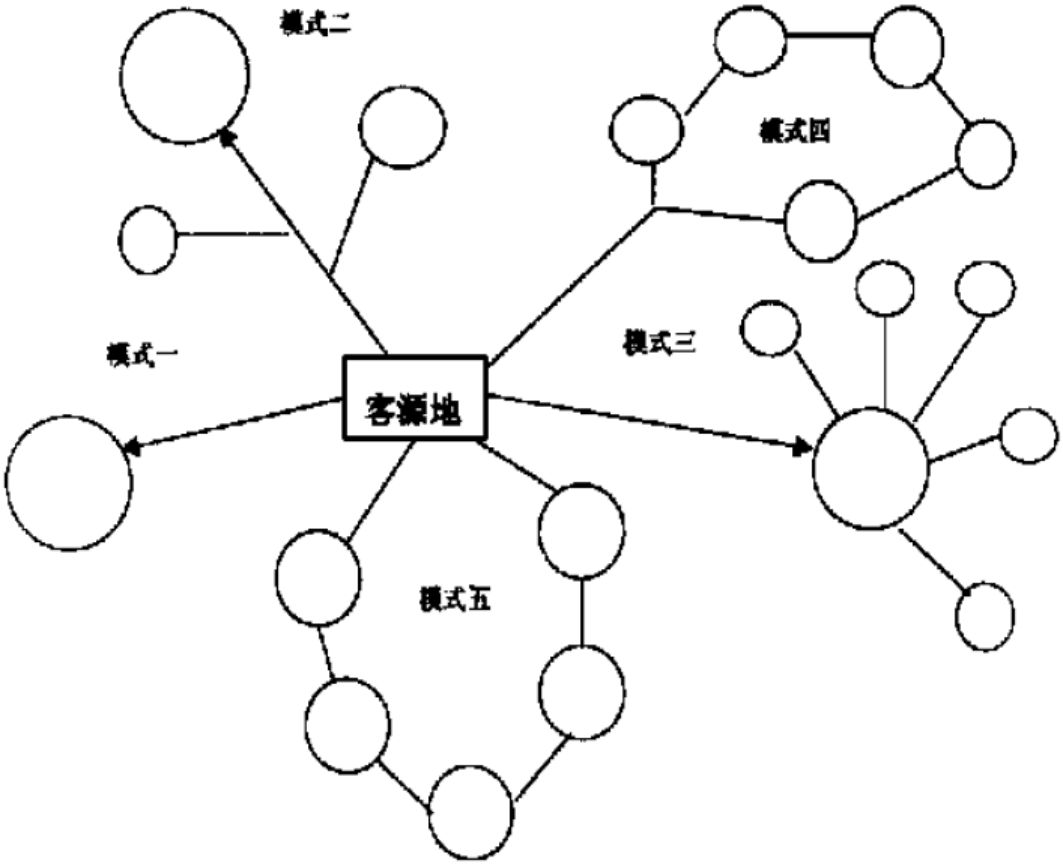


图 1-1 LCF 出行模式

对于城市内部小尺度的旅行模式，可以看作是目的地内部不同景点之间的游客出行模式。这个尺度上的理论比大尺度，如城市、国家尺度的理论研究较少，主要是 Lew 和 McKercher 在 2006 年归纳得到的空间活动模型，包括点对点、环游、复杂三种线型旅游模式。相对于大尺度上的出行模式来说，游客的出行模式

显得比较简单，如图 1-2 所示：

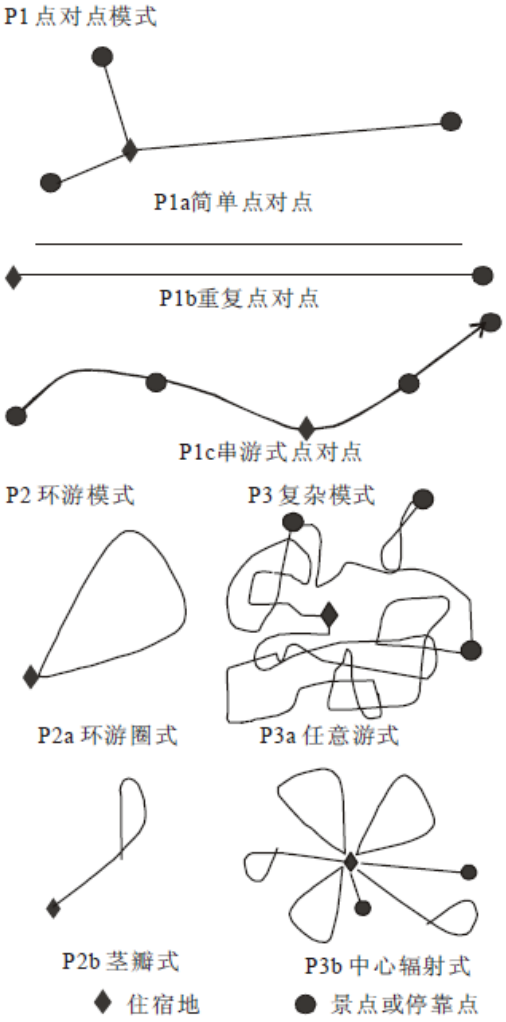


图 1-2 小尺度区域出行模式

除了理论研究，也有学者进行了实证研究，而这些学者的实证研究的方法也大抵相似。Mings 和 McHugh 在 1992 年进行对到皇黄石国家公园的度假者的旅行路线进行了研究，选择了 600 个度假者（实际可用问卷 570 份），分析他们的旅行模式，得出了四种模式：直游模式、部分环游模式、完全环游模式以及飞行+驾车游模式。Oppermann 在 1995 年对到新西兰的国际游客的旅游行程做了分析研究，共有 1000 个国际游客的旅游行程路线作为数据，总结出了基于 LCF 模式的 7 种细化模式。Stewart 和 Vogt 在 1997 年使用到美国密苏里州布兰森的游客进行了研究，得到了 1993 年 LCF 的多目的地的旅游模式。Lew 本人在 2002 年也使用了到香港的国际游客的信息，验证了 LCF 在 1993 年和 Oppermann 在 1995 年得到的多目的地的旅行模式。如下表 1-1 是有个比较完善的实验研究统

计表。

表 1-1 国外旅游路线空间模式相关研究

作者	研究类型	旅行模式
Gunn(1972)	理论研究	目的地模式、环游模式
Mings & Mchugh(1992)	实证研究 到黄石国家公园度假者	直游模式、部分环游模式、完全环游模式、飞行/驱车模式
Lue, Crompton & Fesenmaier(1993)	理论研究	单目的地模式、营区基地模式，往返模式，区域游模式、旅行链模式
Oppermann(1994)	实证研究 到新西兰的国际游客	旅行链模式，完全环游模式
Oppermann(1995)	实证研究 到马来西亚的国际游客	单一和营区基地两种单目的地模式，往返模式、完全环游模式、目的地区域环游、旅行链、复合多目的地区域环游
Stewart & Vogt(1997)	实证研究 到美国密苏里州布兰森的游客	LCF(1993)的多目的地旅行模式
Lew & McKercher(2002)	实证研究 到香港的国际游客	LCF(1993)、Oppermann(1995)的多目的地旅行模式

从上述的实证研究中我们可以看到，这些学者验证的主要方法就是使用游客的调查问卷。问卷的量在几百到几千不等，在当今时期来说，都算不上是很多。

而且，如 Mings 在黄石国家公园中的研究提到的，一些游客在填写问卷时会有诸如仓促、粗略、不认真的情况。这个情况是很难避免的，而且如何判定这些问卷是否在游客认真填写的条件下完成的也是一个问题。另外，这些实验中，因为是问卷类型的数据，其处理的过程基本上是离不开人工的，增加了一些处理上的消耗。还有，不同学者提出的理论的尺度有所差异，各个理论模型没有一个统一完善的验证方法。

为了解决这些问题，我们需要解决数据量的问题，还要给出一个验证游客出行空间模式的方案。而在大数据时代，数据量的问题已经可以很好的解决，使用社交媒体产生的大数据，从中将游客的出行路线勾绘出来，从而进行分析。这就很好的解决了用户问卷游客填写仓促、不认真等问题，因为在社交媒体中，游客总是不自觉的将自己发布的信息和所在的位置展现出来，是一个不依赖于用户主观选择的数据。不仅解决了数据量的问题还解决了问卷数据主观选择影响偏差的问题。而且，只要数据是有效准确的，那么尺度的影响就减小了。

1.3 研究目的和研究内容

本次研究的目的从大数据中提取出城市间和城市内的游客出行路线，基于复杂网络 motif 概念，发掘游客的出行空间模式并于理论模式进行验证，同时将游客出行空间模式进行应用，判断其应用价值。

为了达到上述的目的，需要进行以下的一些实验：

(1) 综述理论：对已有的出行空间模式理论和进行的实证研究进行了综述。阅读旅游相关的论文，综述出已有的理论和实验，分析这些研究方法的优缺点。

(2) 数据准备：从社交媒体大数据中提取出游客行为，确定游客的客源地，构建所有游客的旅游轨迹网络。

(3) 挖掘模式：从游客的旅游轨迹网络中找到出现频率高的模式，借用网络理论中的 motif 的概念，将轨迹网络中出现频率高于随机网络中的子图提取

出来，选择其中符合游客出行规律的子图作为从大数据中发现的旅游出行的空间模式结果。

(4) 模式识别与验证：将游客的出行轨迹与发掘的模式进行统计，研究这些模式是否具有代表游客行为的能力。将能够代表游客行为的 motif 作为发现的游客出行空间模式的最终结果。对这些 motif 与传统的游客出行空间模式理论进行对比分析，加以验证。

(5) 模式应用：在城市的尺度中，将得到的模式与网络中出现的城市进行匹配，模式中每个节点的作用可以对应到城市的目的地类型，如单目的地型、途径地型、门户型、出口型、枢纽型等，由此分析城市功能属性，对于城市的发展策略给予支持。在城市内部，以景区为单位，生成游客的出行网络，可以得到城市内部景点间的出行模式。根据不同的出行模式是提取出对应的行为，借助 K-Means 等聚类方法，可以找出不同出行模式下的优势路线，可用于对游客旅游路线的推荐。

1.4 技术路线

从上述的现状中可以看到，游客出行空间模式已经有了充足的理论基础，而且也有一些学者用问卷的方法进行了检验，只是这些方法存在着一些不足之处。本次提出的方法则可以改正这些不足之处，使用大数据进行验证，同时讨论大数据时代游客出行空间模式的一些应用方式。本次在方法中使用到的 motif 发掘方法以及在应用时使用的 k-Means 空间聚类方法都有比较充足的理论基础，已经被运用与很多的实验之中。

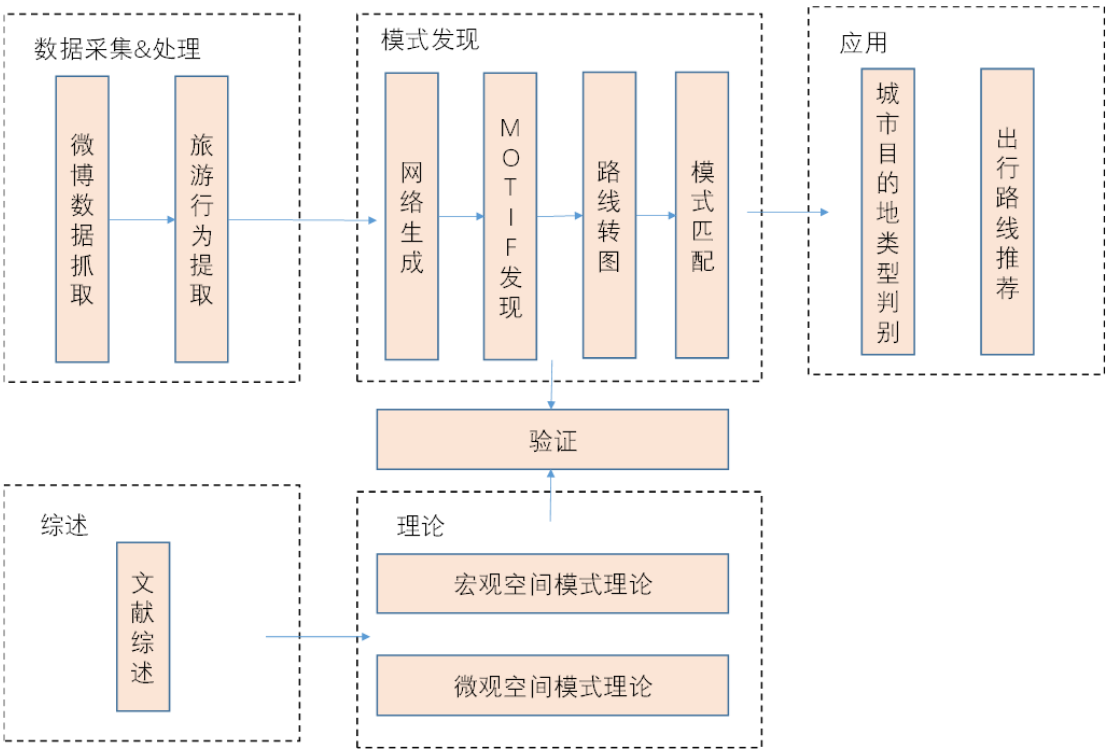


图 1-3 技术框架

技术框架如上图 1-3，可以概括为两个部分：

1. 验证过程：从文献中综述得到现有理论模式；进行数据采集和处理，借用网络理论 motif 概念提取 motif 模式，与实际的行为数据匹配验证，将发现的 motif 模式类型与已有理论进行对比分析，验证概述得到的理论模式。

2. 应用过程：根据城市间的模式 motif 的节点与旅游行为中的城市对应起来，获得城市担当功能的比例；根据城市内部景点间的旅游行为提取出行模式，根据出行模式筛选出一类旅游行为，对这些行为进行 k-mean 聚类，聚类的结果可以用于特定出行模式游客的出行路线的推荐。

第二章 研究方法

2.1 旅游轨迹网络

要进行 motif 的提取，需要先生成旅游轨迹网络，步骤如下图 2-1 所示：

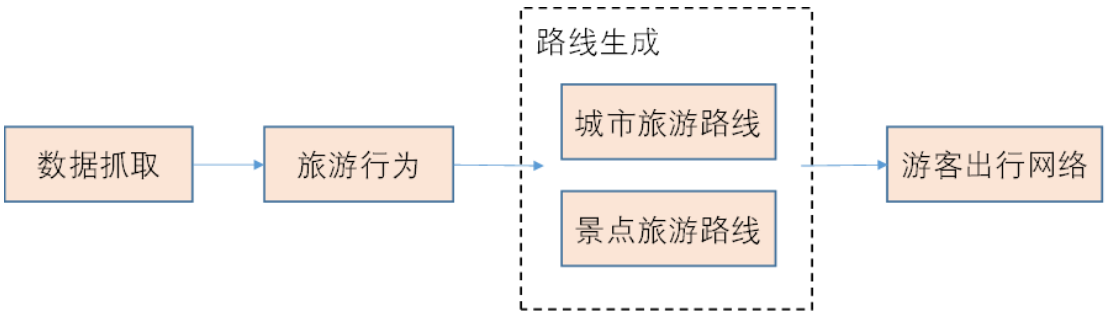


图 2-1 游客轨迹网络生成步骤图

需要先进行数据的抓起，因为使用的是实验室课题组的数据，数据抓取过程是孙奇师兄完成的，详见孙奇 2017 年本科论文《基于社交媒体的旅游时空行为数据库的建立》。随后要进行旅游行为的提取和客源地的发现，从大量的数据中选出游客的路线数据。然后根据规则进行旅游路线的提取，再将所有的路线生成网络，以供 motif 发掘。

2.1.1 旅游行为提取

要从大数据中分析出游客的出行空间模式，首先要做的就是将和游客有关的数据提取出来，而不是将社交媒体中的所有数据都用于研究。这种做法在确保了数据的能够正确的反映游客出行的空间模式而不是人群的移动模式。在目前的一些同样使用大数据进行旅游研究的项目中，有些研究没有进行类似的游客行为的提取，这种做法会在一定程度上对实验的结果有扰动。

判断一个行为是否是旅游行为的方法，可以借助对象所处的地理位置的属性进行判断。比如对于社交媒体中的一个用户，如果它在一个 4A 级包括 4A 级以上的景点发布了消息，那么我们可以认为他是一个游客，同时将他的其他行为（除常居地行为外）认定为也是旅游行为。将游客的所有旅游行为连接起来，则可以得到用户初步的旅游行为链。

2.1.2 客源地

客源地的提取在整个分析过程中也有很重要的意义。客源地，也即常居地，对于划分游客类型如本地客源、外地客源、周边客源有重要作用。其次，客源地的确定对于游客的旅游路线的划分也很重用，以客源地为起点与分割点，才能将游客的整体旅游行为分割为数次有向的旅行。

常居地限定在城市尺度上，因为在更小的尺度上，用户常居地的确认具有更大的不确定性，很难精准的在小尺度上将一个用户的常居地表示出来，准确率不高。虽然微博的用户表中保存有用户的所在城市和省份，但是因为这个信息是用户自身设定的，并不是所有用户都会将正确的常居地填写到上面，因此这个方式也是不可靠的。

客源地的提取方法有很多，这里我们以用户发的消息数最多的城市作为其客源地。

设一个用户U，城市序列为$\mathrm{City}_1, \mathrm{City}_2, \ldots \mathrm{City}_n$，用户在各个城市所发消息数记为$\mathrm{W}_1, \mathrm{W}_2, \ldots, \mathrm{W}_n$，则用户的常居地，记为PCity，应满足

若$\mathrm{City}_k = \mathrm{PCity}$，则有$\mathrm{W}_k \geq \mathrm{W}_i, \forall i, 1 \leq i \leq n$。

若存在一个以上这样的k，则选取用户在待选的常居地城市中，最早的一条消息所在城市视为用户 U 的常居地。

因为使用的是课题组的数据，用户的客源地已经有课题的孙奇师兄完成提取，提取的具体方法见孙奇 2017 年本科论文《基于社交媒体的旅游时空行为数据库的建立》，本次实验使用的是该提取结果。

2.1.3 游客路线的生成

经过旅游行为提取和客源地判定之后，就可以进行游客路线的生成。根据研究的尺度，可以将游客的行为所处的国际、城市、景点等作为一个节点，分别用于进行国际间、城市间和城市内的游客出行模式的研究。

对于社交媒体中的一个用户来说，他的旅游行为可以经过以下几个步骤，将

他所有的旅游行为划分到多次的旅行中：

(1) 去头去尾

因为社交媒体平台获得的数据大多数是一段时间内的，所以需要将第一个常居地之前的数据去掉，将最后一个常居地之后的数据去掉。这样得到了在数据中完全的旅游行为，不存在残缺。如果不去除头尾多余的数据，那首尾的数据将有可能不是依次完整的旅行，对于结果会有偏差。经过这个操作之后，对于城市的尺度来说，旅游行为链中将包含多次的客源地，用于分隔不同的旅行。此时，这些行为链包含的数据是多次旅行数据，后续还需要将其分开。

如一个用户的城市城市列表是

$\mathrm{City}_1, \mathrm{City}_2, \mathrm{HomeCity}, \ldots \mathrm{HomeCity}, \mathrm{City}_{n-1}, \mathrm{City}_n$，经过去改步骤的操作之后就得到了$\mathrm{HomeCity}, \mathrm{City}_3, \ldots \mathrm{City}_{n-3}, \mathrm{HomeCity}$的城市序列。

(2) 去除无用重复常居地

社交媒体大数据而言，用户发布的信息具有很大的随机性，发布的信息在时间上会有一些偏好，除了在旅游过程中发信息，平时在客源地也会发信息，因此，会有多个常居地重复出现。这个操作就是将连续出现的多个常居地中间的数据去掉，所以列表中最多只存在两个连续的常居地。保留两个连续的常居地的意义在于可以用于对前一次旅行结束以及下一次旅行开始的判断。因为两个的时间跨度可能很大，不删除对于以时间分割旅行有帮助。

对于一个用户的城市列表

$\mathrm{HomeCity}, \mathrm{City}_3, \ldots, \mathrm{City}_m, \mathrm{HomeCity}, \mathrm{HomeCity}, \mathrm{HomeCity}, \mathrm{City}_{m+4}, \ldots, \mathrm{City}_{n-3}, \mathrm{HomeCity}$

的处理结果得到

$\mathrm{HomeCity}, \mathrm{City}_3, \ldots, \mathrm{City}_m, \mathrm{HomeCity}, \mathrm{HomeCity}, \mathrm{City}_{m+4}, \ldots, \mathrm{City}_{n-3}, \mathrm{HomeCity}$。

也即将重复出现次数超过三次的常居地去除中间值，保留两端常居地，方便用于按时间进行分割。

(3) 按时间间隔分割

经过了上述的两个步骤之后，旅游行为链还是包含着所有的旅行数据。这个

步骤就是将旅行分割成为多次的旅行数据。游客的两次旅行之间总会有一定的间隔，不同的游客的间隔不同，为了将所有游客的旅行分开来，必须要选择一个适当的时间阈值。将游客的旅行行为中连续两次的时间间隔大于与阈值的分为两次旅行，直到将该游客的旅游行为链全部分割成按时间不能再分割的形态。此时，得到了多条可能的旅行路线。这些路线可能首尾不一定是他的常居地，因为他可能再一次行为之后隔了超过时间阈值之后才发消息。 经过这个操作之后，得到了初步的旅行路线集合，此后如果不需要使用到时间内容，可以不再记录。

对于旅游行为链对应的时间列表$T_1, T_2, \dots, T_n$，对于每一个$T_{i+1} - T_i > \rho$，将路线从T_i处分开，形成两条路线。一直重复，直到所有的路线中不存在两个相邻点的时间差大于ρ。即可得到按时间分割的结果。本次使用的时间间隔$\rho = 3$天。

(4) 按常居地分割

经过上述的步骤之后，得到了游客可能的旅行路线的集合。但这些路线不一定都是正确的，还需要经过常居地的分割，得到以常居地作为首尾是常居地、首尾不是常居地、首或尾是常居地的旅行路线。利用常居地对时间分割结果的旅行路线分割的意义在于一次旅行的出发地和结束地总是常居地，认为回到常居地将是一次旅行的结束。

如果经过时间分割之后的结果中的路线中含有常居地，而且常居地位于路线的中间位置，即形式如$City_1, City_2, HomeCity, \dots HomeCity, City_{n-1}, City_n$，则以常居地为分割点分割路线，得到$City_1, City_2, HomeCity$和$City_4, \dots HomeCity, City_{n-1}, City_n$。对于第二条路径因为还存在可以分割的情况，则进行迭代，继续分割。

(5) 对分割后的行为处理

经过一系列的处理之后，已经得到了初步的旅行行为，还需要经过一些额外的处理，如合并相邻点，路线中可能会再一个地方发布几次信息，这些我们只需要保留一次即可；删除无效路径，因为上述的分割过程分割结果有可能有的是空路径或者只是常居地，这些需要删除；最后，将所有的路径的首尾设定尾常居地，

如果不是，那就需要添加。至此，游客路线生成完毕。

上述为城市间的旅游行为的处理，根据处理的尺度的不同，处理方式会有一些差异。比如城市间的旅游行为一半要求轨迹链能形成一个完整的旅行路线，常居地充当该次旅行的起点和终点。因为就社交媒体平台的特性而言，在大尺度上获取的地理位置信息更加精确。因为用户在社交平台发布的信息普遍是离散的，在很小的尺度上不一定能正确反映用户的实际行为，因为用户不一定在这个尺度上频繁的更新动态，于是就造成了小尺度上离散的状态更加普遍。而在大尺度上，用户发布的状态在空间、时间上有更多的选择，所以在这个尺度构建的旅游行为可以更加严格。

在小尺度，比如城市内部景点间的旅游行为，对于游客而言，大多是将行为依靠在某一个点上，比如一个景点、餐厅等类似的吸引物。而由于去的地点比较多，用户也不可能到一个地点就发布一个消息，因此他的行为在社交媒体平台上就显得零散。这种情况下，对于用户的行为处理就需要稍微放松限制，只需要得到用户的一个有向的路线即可，而不需要严格的要求将在城市内的路线的起点和终点设定为同一个吸引物。因此城市内部的旅行路线跟倾向于一条有向的路线。

对于城市内景点尺度的旅行路线，做的处理和城市尺度旅行路线的处理过程相似，不过在处理的顺序上有些差异，整体流程较为简单，如下：

(1) 根据时间间隔划分旅行路线

对于游客而言，没到一个城市旅游的时间一半不会超过 3 天，因此，在景区尺度的旅行路线划分时可以采用时间阈值为 3 天来对用户在这个城市内的信息进行分割处理，得到包含数次旅游的节点信息。

(2) 合并相邻同个节点

经过时间分割之后，得到了几段分别包含几个节点的组合，对这些组合中依次出现的节点，即相邻出现的同一个节点。因为要记录的是一个具有时间方向的路径信息，合并这些节点才能体现出轨迹的特性，去除冗余数据的干扰。经过这两个处理的流程之后，得到了几次完整的旅行路线。

2.1.4 构建旅游行为网络

构建旅游行为网络是一个重要的步骤，目的在于将所有游客的旅行路线之间的关系合并到同一个网络中。所有的游客旅行路线都是有一个个节点组成的，不同游客有的节点是一样的，甚至旅行路线都是一样的，构建行为网络的意义在于在所有的节点中，将游客的旅行路线信息都加载进去。于是，不同的节点之间的边数，表示这两个节点之间的联系，边数可能为 0，1，2，表示两个几点没有直接联系，有从一个节点之间到另一个节点的行为以及两个节点之间关系密切，可以直接连接，有双向的行为。旅游行为网络是进行游客出行空间模式研究的一个重要步骤，后续需要在这个网络的基础上发掘出 motif，用来对出行模式的分析与比对。

2.2 基于 motif 的行为模式提取

2.2.1 motif 的概念

网络图的 motif 是全图的一种连通的导出子图，具有在原图中出现的次数和在相似的随机子图中出现的次数多得多的性质。这个性质表明这种导出子图在原图中起到了比在任意的随机图中更加重要的作用。

网络 motif 的概念最初由 R. Milo 在 2002 年提出，他做出的定义是：在复杂网络中发现的某种相互连接的模式个数显著高于随机网络。Motif 在不同领域的网络图中表现出来的形式有所不同。而研究人员在诸如工程学网络、生态学网络、生物化学网络和神经网络中都找到了这样的 motif。不同的网络中找到的 motif 都互不相同，有其独特的形式，从这个角度出发，也许 motif 可以解揭示大多数网络构造的基础部分。而近年来，motif 也受到越来越多的人的关注，在一些新领域也发现了 motif，并证实在一些情况下，motif 在原始网络中起到了重要作用。

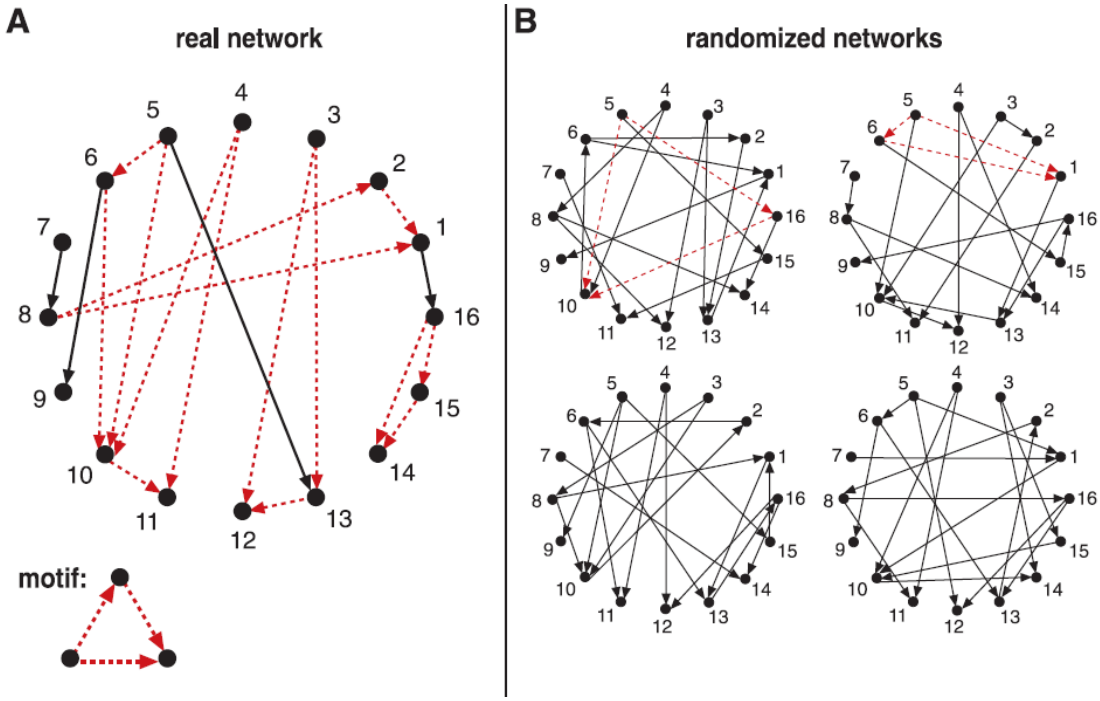


图 2-2 真实网络与随机网络子图出现对比图

如上图 2-2 所示，在 A（real network）中使用三个节点的 motif 出现的频率比 B(randomized networks)中出现的频率要高得多。则这个三节点的子图可以作为原图的一个 motif。

2.2.2 旅游 motif

发杂网络在科学的诸多领域都有研究，而 motif 目前也作为研究网络结构本质的工具被运用其中。在神经网络、互联网等领域，motif 也逐渐被使用起来。近年来，motif 作为研究人类移动行为的工具也逐渐运用于旅游网络中。

Schneider CM 等人将 motif 的概念用于研究人类的日常移动行为，并得到了 17 种 motif，用来表示在城市中一天的移动行为。刘曦运用 motif 分析美国城市间的移民路线问题，杨柳将拓扑 motif 和时间 motif 扩展，提出了场景的语义 motif。但是，这些基于 motif 的研究并没有依靠到游客出行的空间模式上。目前来看，借助复杂网络研究游客的移动行为具有较高的应用价值。

2.2.3 motif 提取方式

根据 motif 的定义，提取 motif 的过程可以分为以下几个过程：

(1) 计算一个图作为原始图的导出子图的次数

(2) 生成一系列和原始图具有相似性质的随机图。所谓的相似性质指的是保证随机图和原始图具有相同的出入度序列。

(3) 计算目标图在每个随机图中出现的次数

(4) 根据一定的规则，比较目标图在原始图和随机图中出现的次数，判定目标图是否是 motif。

对于判定一个图是否是原始图的 motif 并没有一个固定的标准，目前有人使用 Z-Score 或者 p-value 来判定，也有人直接使用目标图出现的频率作为判定标准。总而言之，使用判定的条件需要根据具体使用的场合进行选择。在本次实际数据的处理中，为了确定出一个比较合适的阈值，并筛选出比较多的能过够代表游客旅游路线的 motif，采用了 Z-Score 和 p-value 一起使用的选择方法。先从游客的出行模式中将 3 节点的路径取出来，分析其模式，得到其对应的 motif。再去反查其 Z-Score 和 p-value 作为其他节点 motif 的初始筛选的阈值。

随着 motif 的不断发展，已经有一些比较成熟的软件进行 motif 发掘，不同的软件发掘 motif 的效率不同。目前使用比较多的发掘软件 Minder，FanMode，NetMode 等。本次实验使用的是 NetMODE 软件来进行 motif 发掘，FanMod 在处理节点数目较高的 motif 时效率很低。

根据规则提取 motif 之后，根据需要，可能还需要进一步进行筛选。因为目前提取到的 motif 来源于原始网络图，有的 motif 并不符合实际研究对象的规律。比如研究社交媒体中一个用户的完整的一次旅游行为，那么，他的路径必须是以常居地为起点，同时也以常居地为终点，同时还要保证图是一个欧拉图，将图中所有的边都能经过一次且仅经过一次。

Motif 的表达使用的邻接矩阵的形式，对于如下的一个有向图，使用邻接矩阵表达有下图 2-3 所示：

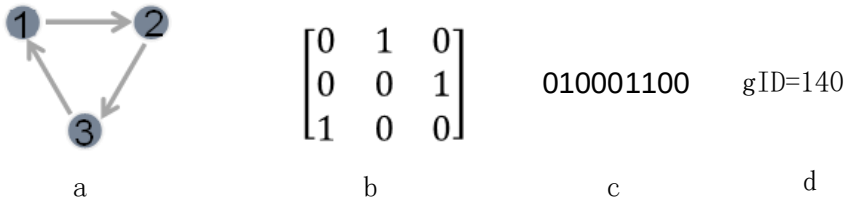


图 2-3 有向图、邻接矩阵、矩阵字符串、唯一 ID 的转换

如图 2-3 中，a 表示原始的节点联系图，b 是用邻接矩阵表示形式，c 是矩阵从左到右，从上到下的一个转成，生成一个唯一的字符串，d 表示将 c 的字符串当成一个二级制数转成的十进制而形成的一个 ID，在使用 motif 提取操作时会记录这个编号。这也是一个子图在提取工具中的一个存储过程。

要根据研究的内容来定义 motif 筛选的规则，因为不是所有的 motif 都能有用来表征用户的旅游行为。本次实验中在提取城市间的旅游行为模式时使用的是欧拉图的定义，不符合这个定义的子图需要删除。

比如 010000010 和 011000010 矩阵表示的三节点的子图，见图 2-4，即使这个图在原始图中出现的频率很高，但是它并不符合研究的要求，没有实际的意义。如果筛选之后的结果中包含有这个图，还需要根据研究的实际意义将其删除。其中，判定一个图是连通采用并查集的方法。

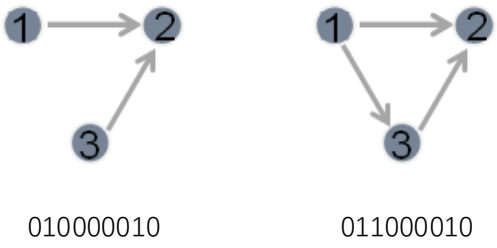


图 2-4 不符合出行路线规律子图示例

2.3 发现模式与理论模式验证

2.3.1 网络 motif 的重新标号

从网络中得到了 motif，与所有游客的出行行为进行比对，观察发现的这些模式是否能够代表整个群体的出行规律，认为如果这些模式能够代表超过 95%的

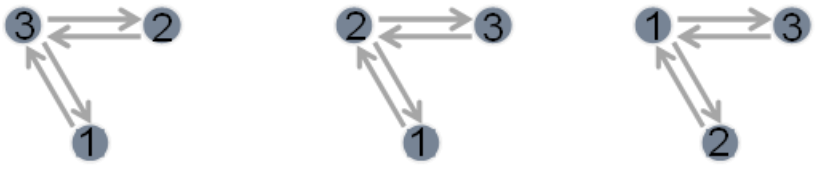
用户的行为就具有代表性。如果不符合，则修正 motif 发现的阈值，使得发现的模式能够代表用户的行为。

将发现的 motif 的模式区分到理论模型中，观察这些模式能够体现理论模型的特征，如果在一定程度上能够符合理论模型，则认为理论模型得到验证，如果存在较大的差异，则分析原因，解释现象。

从提取到 motif 到检验这些 motif 能否代表所有游客的旅行路线，还需要经过寻找有效的同构图的步骤，将具有标号的图转成可以表示旅行路线的图。因为这些提取到的 motif 大部分点并不是按顺序，而游客实际的旅行路线中地点的出现和连接的顺序是有序的。即对与标号而言，必须先经过标号为 0 的点才能到达标号为 1 的点，依次如此，直到发现所有点为止。因此必须要有一个转换的过程，才能将结果用于与游客的旅行模式进行判定。

算法实现需要使用的方法为深度优先搜索，找出所有可能的路线。采用标记的方法，从任意一个顶点出发，从其指向的临近点中选出一个作为新的顶点，删除该边，重复上述步骤，如果遇到没有临近点可选的情况，则判断边是否已经遍历完成，如果没有遍历完成，说明这条路径是行不通的。所以，返回上一个点，同时修复这条边，选择另一个临近点，继续探索，如果没有临界点，则继续返回上一个点并恢复边。如果遇到没有临近点可选的情况，而且所有边都已经遍历结束，并且这个点是最初的顶点，那么将这个结果保存，返回上一个点，同时修复这条边，继续探索，直到所有顶点作为起始点并完成所有探索。

如对初始矩阵为 001001110 的图，可以找的结果有下图 2-5：



001001110　　010101010　　011100100

图 2-5 图的重新标号

在进行匹配时，需要用经过重新标号之后的结果对用户的旅游行为路线进行匹配，而不是使用原始标号的图像的矩阵字符串，因为原始图的标号顺序生

成的矩阵字符串和实际的情况不一定是符合的。

2.3.2 motif 与旅行路线轨迹匹配

经过上述对 motif 的重新标号之后，已经得到了按照标号顺序出现的图，可以用这个图对旅行路线轨迹进行匹配。旅行路线是由一系列可以重复出现的节点组成的，相同节点不能出连续出现。对于城市尺度的旅行路线而言，还要求路线的起始点和终止点是同一个点，而对于城市内部景点尺度的轨迹路线则没有这个要求，只需要根据旅行路线将用户的每一次旅行轨迹所形成的有向图使用邻接矩阵表示即可。

将从复杂网络中提取到的 motif 的邻接矩阵和实际的旅行路线的邻接矩阵相互匹配，即可得到每种 motif 在实际的旅行路线中对应的行为数目。最终，统计所有的 motif 和出现频率很高的旅行路线的行为矩阵，得到所有研究用户的旅行路线的模式，将这些模式和以前的研究者提出的模式进行验证，查看是否符合，是否达到验证旅游空间模式的目的。

2.4 基于大数据提取旅游模式的应用

2.4.1 城市目的地类型判定

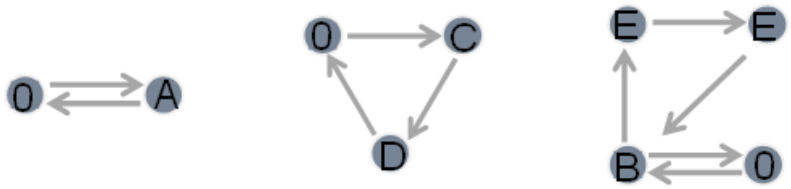
基于城市间的旅行路线提取的游客出行的空间模式，可以用于研究作为节点的城市的目的地类型的判定。

因为已经从游客的旅游网络中获取了可以代表众多游客旅游行为的模式，所以可以很方便的利用这些模式统计出对应的城市的功能。

大多数学者将目的地类型分为以下单目的地型、门户型、出口型、枢纽型、途径型。本次采用的目的地类型的定义借鉴朱明，史春云等人在《基于旅行社线路的国内旅行空间模式研究》提出的定义。

其中，单目的地表示游客在一次旅行中只去了一个目的地城市就返回了常居地，对于去到的那个城市称其为单目的地。门户型目的地指的是除了单目的地类

型的旅行路线之外，游客从常居地出发后所到达的第一个城市。出口型则指的是排除单目的地类型的旅行路线中最后访问的城市，即访问这个城市之后就返回了目的地。枢纽型则指的是访问次数超过两次的城市，有的学者也将枢纽型目的地称为那些同时充当门户型和出口型目的地类型的城市。在本文中也使用这一个说法，将同时具有门户和出口功能的城市设定为枢纽型城市。途径型则是排除上述功能类型之外的节点城市。下图 2-6 是城市目的地类型的一个示例：



O：客源地　A：单目的地　B：枢纽型　C:门户型　D：出口型　E：途径型

图 2-6 城市目的地类型示例

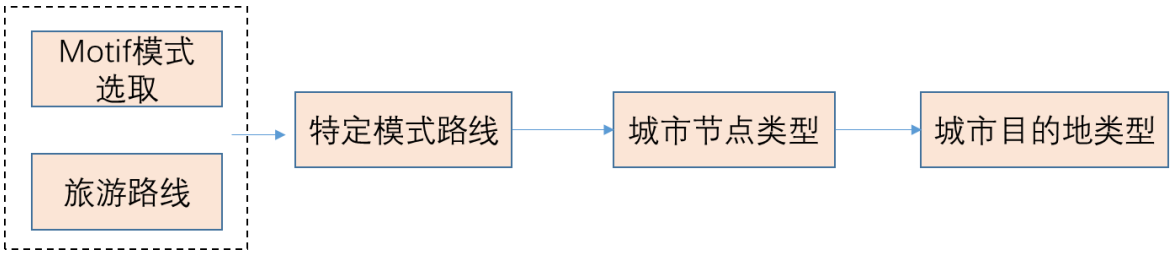


图 2-7 城市目的地类型判定流程

城市目的地的类型判定的流程如上图 2-7 所示，具体流程如下：

(1) 根据研究的目的，在已有的模型中选出用于研究的 motif。因为不同模式的旅行行为的数量的差异时十分大的，比如，大多数游客的旅游行为中一般只访问一个城市，所以在总的旅游行为中，单目的地类型的占比就很高，如果不对这些问题进行一些处理的话，将对判定的结果产生很大的干扰，比如最终所有的城市的最可能的目的地类型都变成单目的地。因此要根据研究的目的将所需要关注的旅游路线类型先行筛选出来用于研究。综上，需要根据 motif 和其对应旅游行为的数量进行一些必要的筛选。

(2) 将旅行路线轨迹矩阵化，对应到具体的 motif 中，并判定节点城市的功能。和验证出行空间模式的类似的，需要将 motif 模型和旅游路线对应起来，

具体的方式和 2.3.2 中匹配的方式一样。

(3) 随后根据上述对于目的地类型的分类的定义，将每个城市节点的目的地类型确定下来。对所有选中的路线进行城市节点目的地类型的判别，即可得到每个城市各个目的地类型的出现的次数。

(4) 根据研究的目的地类型以及城市的目的地类型占比，判定结果。经过上面流程之后，已经得到了每个城市的几种可能的目的地类型。对这些类型占比进行比较，比值最大的目的地类型就时该城市的目的地类型。

2.4.2 旅游路线推荐

城市间的出行空间模式可以用于判定城市的目的地类型，城市内景点间的出行模式则可以用于旅游路线的推荐。

游客出行时的旅游路线非常复杂的，但是从大数据的角度来说，总有一些景点的访问轨迹是相似的，关联性很大。而且对于一些轨迹而言，他们的访问是具有一些普遍性的。因此可以通过一定的方法将这些路线提取出来并推荐给那些对这个旅行模式有兴趣的游客。执行的流程如下图 2-8 所示：

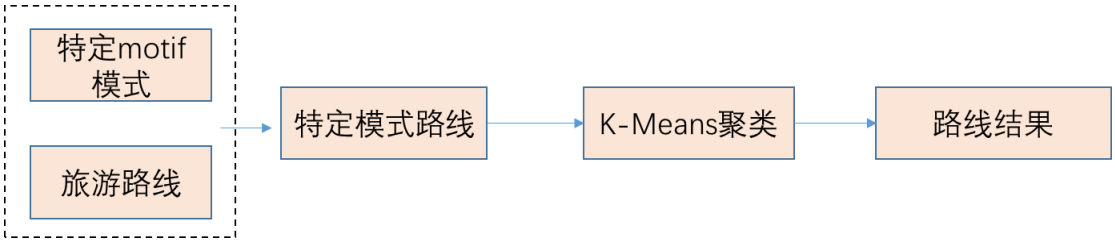


图 2-8 旅游路线提取流程

具体操作方法如下：

(1) 根据不同模式提取出对应模式的旅游行为。如对于要在城市中按照点对点模式，如串游式点对点模式，按照一条简单的路径一直往前，不返回已经访问过的点。将这些旅游行为提取出来，作为分析的数据源，分析出最受欢迎的路线。

(2) 将得到的旅游行为进行聚类，获得具有优势的路线。使用 k-means 聚类的思想，从一些路线中找出，找出由一些特性出现频率高的地点的几条路

线。

(3) 随机选择 k 个不同的路线作为中心。在此次实验中，k 的取值为人工判断的方式。不同的旅游路线中出现的地点是不一样的，为了将其聚类需要确定一个指标，用以表征两条路线间的距离。本次实验采用距离作为相似性指标。

将所有提取到的旅游行为中包含的地点生成一个长度为 n 的列表，每个地点对应到 0~n-1 的数字，用于表示在地点向量中的下标。对于一条路径，如果访问的一个地点，那么在向量中对应的地点处标记为 1，不访问的地点标记为 0。对于路线对象X_i和X_j的距离，计算方式如下：

$$S(X_i, X_j) = \sqrt{\sum_{s=0}^{n-1} (X_{is} - X_{js})^2}$$

其中 n 表示地点数目，X_{is}和X_{js}分别表示路线是否访问了地点 s，访问为 1，不访问为 0。将所有路线聚到和自身最近的中心上。此距离测度函数借用李渊 2016 年《基于 GPS 的景区旅游者空间行为分析》的研究。

(4) 分配路线对象到各个中心后，重新计算各个簇的中心。如果中心和上次的中心不一致，转到步骤 2；如一致则完成一次聚类，返回所有中心以及簇。

(5) 重复进行多次聚类，计算误差平法和测度函数，取测度结果最小的那一次分类作为分类的结果。测定函数如下：

$$E = \sum_{i=1}^{k} \sum_{p \in c_i} dist(p, m_i)$$

其中 k 表示聚的簇的数目，m_i为第 i 簇的中心，p 为分到第 i 簇的路线，dist 函数用于计算两条路线的距离。

(6) 获得聚类结果，将路线推荐给特定类型的用户。给定簇数 k，最终可以聚类出 k 类路线，则每一类的中心路线都将是该类中与其他路线“距离”最近的路线，一般的，其中包含的景点也是受较多人欢迎的景点。于是可以将这 k 条路线推荐给这个旅游类型的用户。

第三章 数据处理结果

3.1 旅游数据

从前面的几个章节中，已经提出了根据社交媒体大数据进行验证游客出行的空间模式的方法，但是还缺少一个实际论证的过程。这里给出一个实际的处理流程，用于检验方法的可行性。研究游客在城市间的旅行路线的空间模式以及城市内部景点间的流动的空间模式，研究城市内景点尺度时采用苏州市作为研究对象，分析其特征。

本次实验使用的社交媒体平台为新浪微博，使用的数据也是新浪微博数据。微博是一个拥有大量用户的社交平台，是提供微型博客服务类型的社交网站。自2009年8月的内测版以来，微博已经成为了全球用户规模最大的社交媒体公司。截至2018年3月，微博的月活跃用户达到4.11亿。用户可以通过智能手机、PC终端等多种客户端进行个人社区的搭建，十分方便快捷。因此，用户发布的微博数也很多。保证了具有充足的数据量，而且用户的随机性强，涵盖的范围广，用于数据分析，具有的代表性更强。

本次使用的微博数据是由实验室的师兄孙奇下载的。通过他的算法下载到的微博数据有以下的一些信息，兴趣点信息、用户信息、地理微博信息。其中，兴趣点信息包括的字段有 POI 编号（POIID）、名称（Title）、 经度（Longitude）、纬度（Latitude）、 类别（Category）、街道地址（Address）、在此 POI 签到的微博数量（Checkin_num）和在此 POI 签到的用户数量（Checkin_user_num）。而用户信息则包括了用户编号（Userid）、微博昵称（Screen_name） 性别（Gender）、用户签名（Description）、省份代码（Province）、 城市代码（City）、 用户注册地位置（Location）、 好友数量(Friends_count)、粉丝数量（Followers_count）、 微博数量（Statuses_count）、微博注册时间（Created_at）等等。为保证社交媒体登的隐私安全，用户信息中并不涉及用户邮箱、手机、真实姓名、 生日等敏感信息。而要进行和地理位置相关的研究，就必须有用户包含地理信息的微博，这些

微博就称作地理微博。地理微博包括微博 id（ID）、 发布用户的 ID（Userid）、微博内容（Text）、 经度（Longitude）、纬度（Latitude）、发布时间（Created_at）、转发数量（Reposts_count）、 评论数量（Comments_count）。如果用户在发布微博时位于某一 POI 附近，并且用户手动选择了这一个 POI，则用户进行了一次签到，这样的微博称之为签到微博。以下表 3-1 是一段地理微博的例子：

表 3-1 地理微博示例

微博id	用户id	正文	经度	纬度	发布时间	转发数	评论数	签到	签到POI名称
3424557241	179884	我在这里: #中国电信西街营业厅#最怕到这个电信交费，太慢了，等 SHI 了! http://t.cn/zOx7s2D	120.5445	31.86326	Sat Mar 17 13:18:53 +0800 2012	0	0	B2094650D064A0F4459A	中国电信西街营业厅
3508640809	158905	rings for friendship..			Sun Nov 04 13:56:59 +0800 2012	0	0	B2094650D064A0F44798	名典咖啡语茶
6718900	6216210	我在#名典咖啡语茶# http://t.cn/zWlQ4iR	120.5461	31.86368					
3416595462	177764	我喜欢鸡心喜欢到不行啊! 看了半天还是决定鸡心、 弄好了到我婶娘这里来happy 啦! 我在这里:张家港 #逸品精品店 #			Fri Feb 24 14:01:39 +0800 2012	0	6	B2094650D064A0F449F	逸品精品店
4258600	5451	http://t.cn/zO4PF8U	120.544	31.86433					
3599662710	251078	饿了[熊猫][熊猫] 我在这			Sat Jul 13 18:05:31 +0800 2013	0	0	B2094650D064A0F449E	机关宾馆
0211200	3693	里 :http://t.cn/zQyFeVZ	120.5424	31.8638					
3501077314	196039	每天下班都有好景色^O^今天是鱼鳞云←_← 我在#商业大厦#			Sun Oct 14 17:02:21 +0800 2012	0	2	B2094650D064A0F54598	商业大厦
0179200	5121	http://t.cn/zOgisar	120.5454	31.8626					
3452074664	144472	我在这里:#江南人家龙虾馆 #			Fri Jun 01 11:43:19 +0800 2012	0	0	B2094650D064A0F5419C	江南人家龙虾馆
9230900	7221	http://t.cn/zOggASQ	120.5445	31.86494					

					Mon Oct					
36383	314				28			B20946		
60696	514				12:57:30			50D064	国	际
62732	142	[威武][威武] 我在这	120.5	31.86	+0800			A0F544	购	物
0	0	里:http://t.cn/z8LTyGS	463	245	2013	0	0	93	中心	

而对于研究用户的旅游行为来说，我们需要使用的字段主要有:用户id，发布的每条微博的时间，发布时的地点，发布的微博所依附的兴趣点。本次实验使用的微博数据为 2012 年，2013 年，2016 年在苏州市发布过微博的用户的所有微博。表 3-2 景点数与路线数目的统计

3.2 数据预处理结果

采用 2.1 节提到的方法对这些微博数据进行处理，得到城市间和城市内景点间的旅行路线数据如下：

城市间：

旅游用户数：202383

旅游行为数：1088673

城市内景点间：

旅游用户数：258745

旅游行为数：705494

表 3-2 景点数与路线数目的统计

经过城市数目	行为数	经过景点数目	行为数
1	865296	1	595997
2	125466	2	55875
3	49124	3	11322
4	20799	4	3383
5	11423	5	1282
6	6253	6	548
7	3592	7	209
8	2232	8	113
9	1381	9	37
10	887	10	14
11	558	11	7

12	402	12	7
13	296	13	6
14	197	14	4
15	175		
16	124		
17	69		
18	64		
19	61		
20	42		
21	33		
22	29		
23	22		
24	22		
25	17		
26	16		
27	11		
28	9		
29	8		
30	6		
31	2		
32	10		
33	4		
34	3		
35	5		
36	5		
37	1		
38	4		
39	3		
40	2		
41	2		
42	3		
44	1		
46	2		
48	1		
49	2		
50	1		
51	1		
52	1		
56	1		
62	1		
65	1		
69	1		
91	1		

115	1		

路线访问的景点数与路线数目的统计如上表 3-2 所示。城市间的旅游行为中有的用户在一词旅行中访问的城市的数目非常多，这种情况可能的原因是对于其常居地的判定有问题，造成了使用常居地进行路线的切割时，没能正确分割。当然也有可能是这个用户确实是进行这种模式的旅行，类似于全球旅行，在一个地方停留的时间很短，却去了很多城市。

但是，不管是城市间旅行路线还是城市内部景点间的旅行路线，这些一次性访问多个节点的行为在整个数据整体中还是占少数的，而且，就实际情况而言，一般地，一个用户选择进行城市间旅行时不会超过 4 个城市，在城市内景点间的旅行路线也不会超过 5 个。考虑到 motif 提取算法的局限性，在这里只分析 5 个节点的城市路线以及 6 个节点的景点路线，至于其他节点数目，可以借助一般方法进行模式发现并和已有的 motif 进行分析比对，判断其发生的可能性。

第四章 结果分析

4.1 motif 提取结果

4.1.1 提取结果

根据上述的处理方法，游客在城市间的旅行路线可以生成一个网络图，这个网络图的节点即是这些游客所游览过的城市，网络图的边则表示有游客经过这条路线。网络具有 359 个节点，有 33133 条边，属于一个稀疏网络。

使用 NetMODE 工具对这个网络图中的 motif 进行分析，分析节点数为 3，4，5，6 的 motif，节点数为 2 的 motif，根据前面提到的规则，只有一种情况，即邻接矩阵为 0110 的有向图，无需进行 motif 发掘。需要注意的是，因为这些 motif 是一个欧拉图来表示一个完整的旅行轨迹，所以常居地也会被包含其中。也就是说，节点数比访问的城市数目会多 1。城市间的景点的旅行路线则不必考虑这一点，因为不用考虑其住宿地点，若住宿地点出现在轨迹中，也将其视为一个普通的景点。

目前的 motif 提取算法基本上是使用正规标号来表示的，是解决图同构问题的一个重要方法。图重构是指两个图之间当且仅当存在一个置换，将其中一个图的节点标号序列映射到另一个图的节点序列之后，得到的图和另一个图在节点之间的关系上完全一致的情况。这是 motif 提取算法的一个难点之一。

因此经过 NetMODE 等软件进行 motif 提取之后得到的网络图是用正规标号的形式给出的。所以得到的这些 motif 只是表示了节点之间的联系关系，还需要将图的所有同构图求出并按照节点出现的顺序表示成邻接矩阵时才能用于表示游客出行路线。对三个节点的旅游行为数据进行统计，得到出现频率比较高的 motif 的邻接矩阵，反查它们对应的 Z-Score 和 P-Value，以其为一个参照阈值去筛选其他节点的 motif。

对 Motif 提取的原始结果的邻接矩阵如下表 4-1：

表 4-1 城市尺度根据阈值筛选 motif 结果

序号	邻接矩阵
1	0110
2	001001110
3	001100010
4	0001000100011110
5	0001000101001010
6	0001001001011010
7	0001001010000100
8	0001001010010110
9	0001001101011110
10	0001001111000110
11	000010000100001000111110
12	000010000100001001011010
13	000010000100010010111010
14	000010000100010010010100
15	000010000100010010110110
16	000010000100010011001010
17	000010000100010110000110
18	000010000100011001011110
19	000010000100011011001010
20	000010000101000100000110
21	000010000101010101000110
22	000010001000011001011100
23	000010001000011011001010
24	000010001000011011011010
25	000010001000011101000110
26	000010001000011101010110
27	000010001001000001011001
28	000010001001000100000100
29	000010001001000100010011
30	000010001001001001011011
31	000010001001001101000011
32	000010001001010001011010
33	000010001001010101010011
34	000010001011000001010011
35	000010001100011011001110
36	000010001100011011011111
37	000010001100011111000111
38	000010001101001001011111
39	000010001101001011001011

40	0000100011010011010001110
41	0000100011010100110010100
42	0000100011010100110110110
43	0000100011010101010001100
44	0000100011010101010101110
45	0000100011010101110000110
46	0000100011100000100101110
47	0000100011100000110001010
48	0000100011100010110001110
49	0000100011100100110001100
50	0000100011100100110101110
51	0000100011110000110000110
52	0000100110010100110110010
53	0000100110010101000101100
54	0000100110010101010101010
55	0000100110010110110110110
56	0000100110010111010001100
57	0000100110010111010101110
58	0000100110010111110000110
59	0000100111010110110111110
60	0000100111010111110001110
61	0100000010000001000000100000001100000
62	0100000010000001000000110001001000000
63	0100001010000001000000100000001010000
64	0100000001000000010000000100000001000000011000000

以提取到的这些 motif 的原始图为基础，找到他们的同构图，对这些符合条件的同构图进行重新标号，共得到 207 年符合条件的同构图。完全列出来太长，此处略过。

对于城市内景点的旅行路线，根据游客的行为生成的网络图中一共有 42 个节点，1168 条边。每条边表示有游客的路线连接了这两个景点。

使用 netmode 进行 motif 提取，根据景点间的旅行路线的筛选规则，motif 必须是可以形成欧拉路径的有向图。因为在城市内景点间的 motif 的要求没有像城市间的那么严格，所以会选择出很多符合条件的 motif。根据这些条件于是根据这些条件先筛选出满足条件的 motif 有 8090 个，再将这些 motif 进行重新标号。得到的所有重新标号的 motif 有 47207 个，数目较多，在此不再列举。随后还需

要将这些符合条件的 motif 和实际的旅游行为结合起来，只有那些匹配到一定程度的 motif 才能选为最终的 motif，作为代表所有游客出行的空间模式。

4.1.2 motif 与旅游路线

上述提到，要将 motif 重新标号之后才能和旅行路线轨迹进行匹配。从网络中发掘 motif 之后根据筛选的规则，将符合条件的 motif 挑选出来，使用重新标号的算法进行标号。而对于旅游路线，则根据节点出现的次序标号，对已经标号的节点直接联系即可。如旅游路线：[北京市，上海市，苏州市，成都市，上海市，北京市]，北京市作为常居地，其他城市按照出现次序标号为下表：

表 4-2 路线标号示例

城市名称	标号
北京市	0
上海市	1
苏州市	2
成都市	3

这个轨迹生成的轨迹图的邻接矩阵表示为 0100101000010100，对应下图 id 为 6 的 motif。于是就可以将这个轨迹和 id 为 6 的 motif 关联起来。

利用这些邻接矩阵去和游客的出行路线进行比对，得到如下图 4-1 的结果，其中，motif 下方的数字标号表示其 id，id 越小表明出现的频率越大：

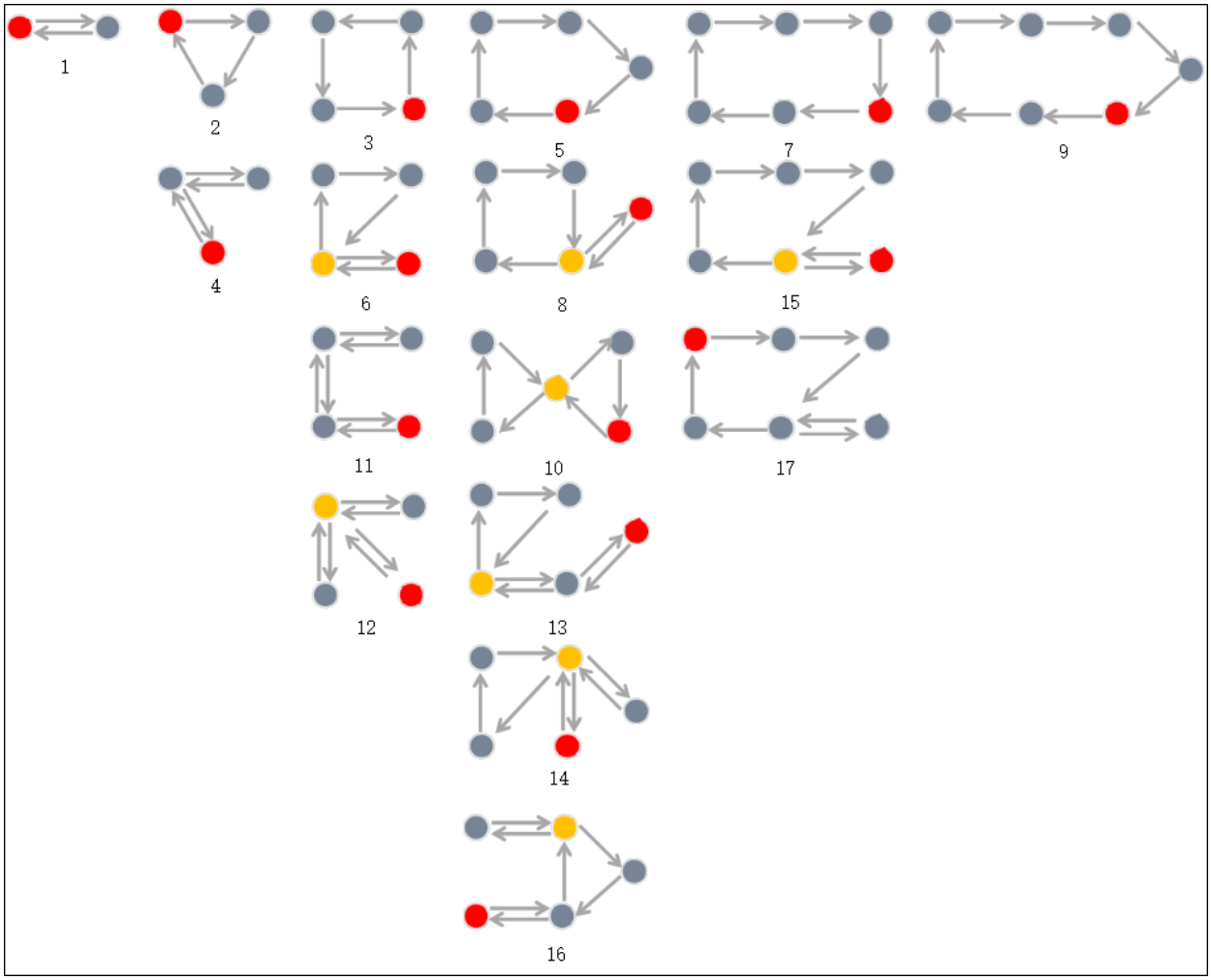


图 4-1 城市尺度 motif 提取结果

上述的 motif 中，红色节点表示为这个 motif 中可以作为常居地的节点，橙色节点表示重要节点，这些节点的连通性比较强。每个模式对应的行为以及占比如下表 4-3：

表 4-3 城市间行为与 motif 模式匹配表

ID	邻接矩阵	行为	占比	累计
1	0110	865296	79.48%	79.48%
2	001100010	125466	11.52%	91.01%
3	0001001010000100	31670	2.91%	93.92%
4	001001110	17784	1.63%	95.55%
5	000010001001000100000100	10964	1.01%	96.56%
6	0001000101001010	9298	0.85%	97.41%
7	010000001000000100000010000001100000	4456	0.41%	97.82%
8	000010000100010010010100	3590	0.33%	98.15%
9	010000000100000001000000010000000100000001100000	2118	0.19%	98.34%
10	000010000101000100000110	1110	0.10%	98.45%

11	0001001001011010	918	0.08%	98.53%
12	0001000100011110	532	0.05%	98.58%
13	000010001001000000010110010	456	0.04%	98.62%
14	000010000100000100100011010	406	0.04%	98.66%
15	010000101000000100000010000001010000	401	0.04%	98.69%
16	000010000100010011001010	375	0.03%	98.73%
17	010000001000000100000011000100100000	303	0.03%	98.76%

将城市内景点间的旅游路线也按照出现的次序标号，生成图，和所有符合筛选条件的城市内部的 motif 进行匹配，最终得到的 motif 如下图 4-2 所示：

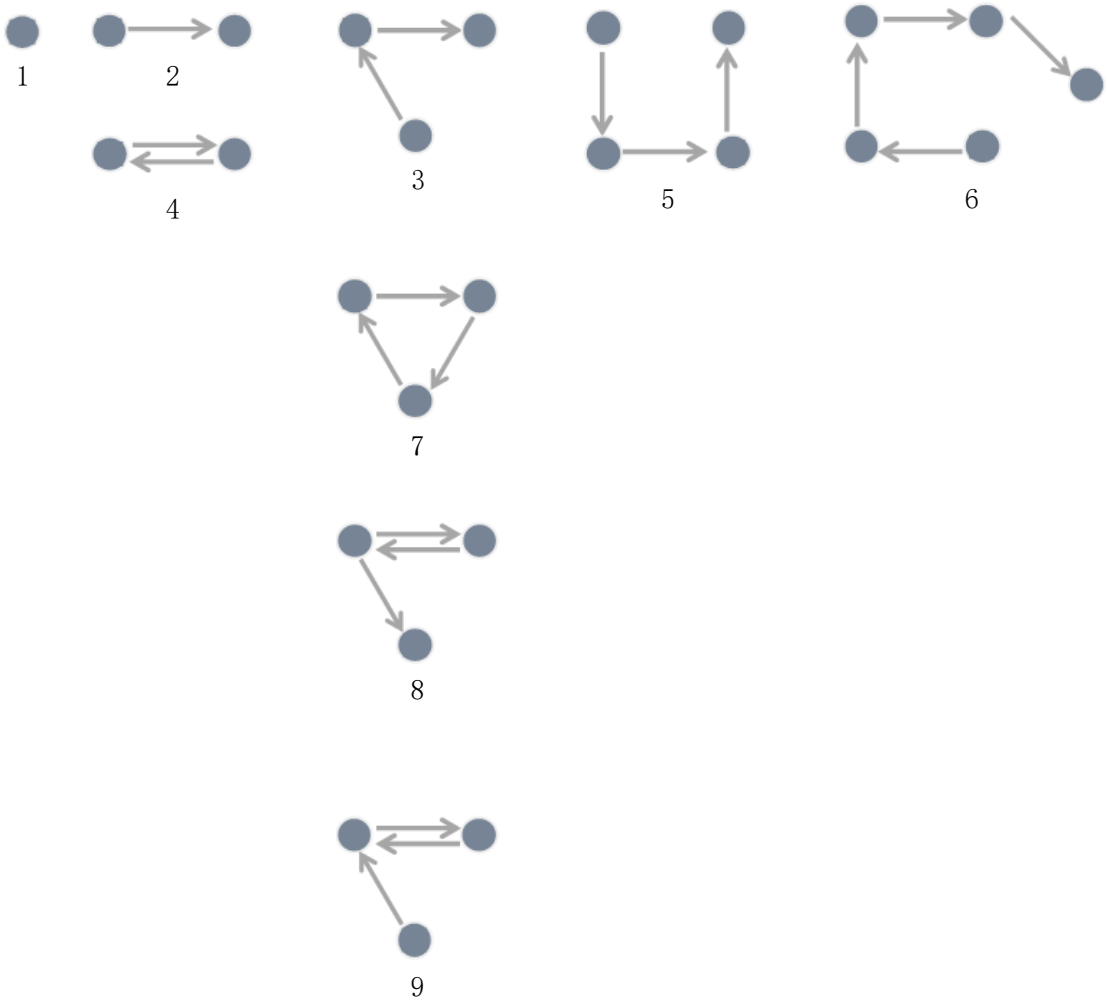


图 4-2 城市内尺度 motif 提取结果

模式与旅游行为匹配表如下表 4-4：

表 4-4 城市内景点行为与 motif 模式匹配表

ID	邻接矩阵	匹配行为数	占比	累计占比
1	0	645833	91.54%	91.54%
2	0100	43459	6.16%	97.70%

3	010001000	7725	1.09%	98.80%
4	0110	2772	0.39%	99.19%
5	0100001000010000	1988	0.28%	99.47%
6	010000010000010000010000	634	0.09%	99.56%
7	010001100	416	0.06%	99.62%
8	011100000	402	0.06%	99.68%
9	010001010	363	0.05%	99.73%

在城市间的旅游路线和预设的 motif 的匹配情况来看，这 17 个 motif 已经可以表示出 95%以上的行为，因此这些 motif 对于这些行为来说具有一定的代表性。对于景点间的旅游行为，预设的 motif 中匹配最频率最高的 9 个已经能够表示表示 99%以上的旅游行为，因此，这 9 个 motif 也可以用来表征这些游客出行的空间模式。所以，用这些被选出来的 motif 表示游客的出行空间模式是可信的。

4.2 motif 与游客出行空间模式验证

4.2.1 城市尺度旅游出行空间模式

以往提出的游客出行的空间模式理论中受到较多学者认可的应该就是由 Lue、Crompton 和 Fesenmaier 总结了五种度假旅行模型，即 LCF 模型。根据 LCF 模型的分类规则，将城市尺度最终提取到的 motif 进行分类，得到如下的分类结果如图 4-3：

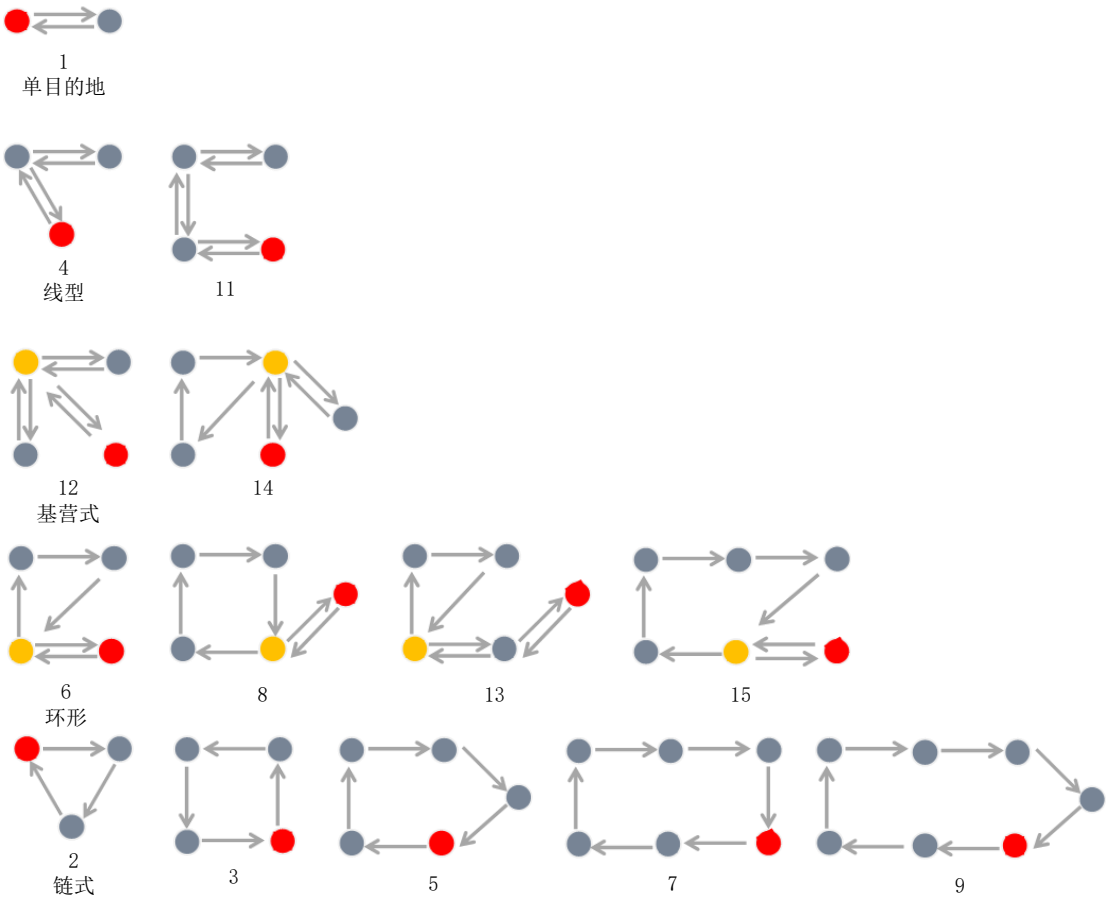


图 4-3 城市尺度 motif 分类结果

其中 LCF 的 5 个模式的占比如下表 4-5：

表 4-5 城市尺度模式占比表

模式	占比
单目的地	79.48%
链式	16.04%
线型	1.72%
环形	1.26%
基营式	0.09%

从上述的处理过程，得到以下结果：

(1) 发现的模式没有超出理论模式。从得到的 motif 识别到的几种模式和 LCF 模式能很好的匹配，基本都能归类到理论模式中，没有超出理论模式的类型。这可能和筛选的阈值有关，本次实验忽略了出现次数很少的模式，认为这些出现很少的情况是某种特例。

(2) 理论模式都包含了发现模式。LCF 理论模式共分了五种类型，而在发现到的模式中，这几种类型都有一些 motif 可以归类到模式上。所以理论模式能够涵盖整个发现模式。

(3) 发现模式的集中形式。从上面的统计中可以看到，发现模式中单目的地和和链式模式占据了极大的比重，其他三种模式的比重远远小于前面两种模式。但是，发现的模式可以完整的包含理论模式的类型，而理论模式也可以完整包含发现模式的类型，因此，发现模式可以用来表征理论模式，理论模式得到了验证。

从上面几点可以看出，利用大数据时代的社交媒体的用户数据所提取的旅游行为分析得到的游客出行的空间模式和 LCF 模式具有很高的重合度，LCF 模式所提到的可能性在大数据时代的游客出行路线中都所有体现。因此，LCF 模式得到了一个验证，这个验证时基于大数据量，基于用户的客观行为而不是主观意愿（个人花费时间填写问卷）得到的，具有一定的合理性。

4.2.2 城市内部景点间的出行空间模式

对于城市间的出行模式，比较接受认可的还是 Lew 和 McKercher 在 2006 年归纳得到的空间活动模型，包括点对点、环游、复杂三种线型旅游模式。根据这几个模式的定义，将从城市内景点间提取到的游客出行的空间模式进行归类，得到如下图 4-4 结果：

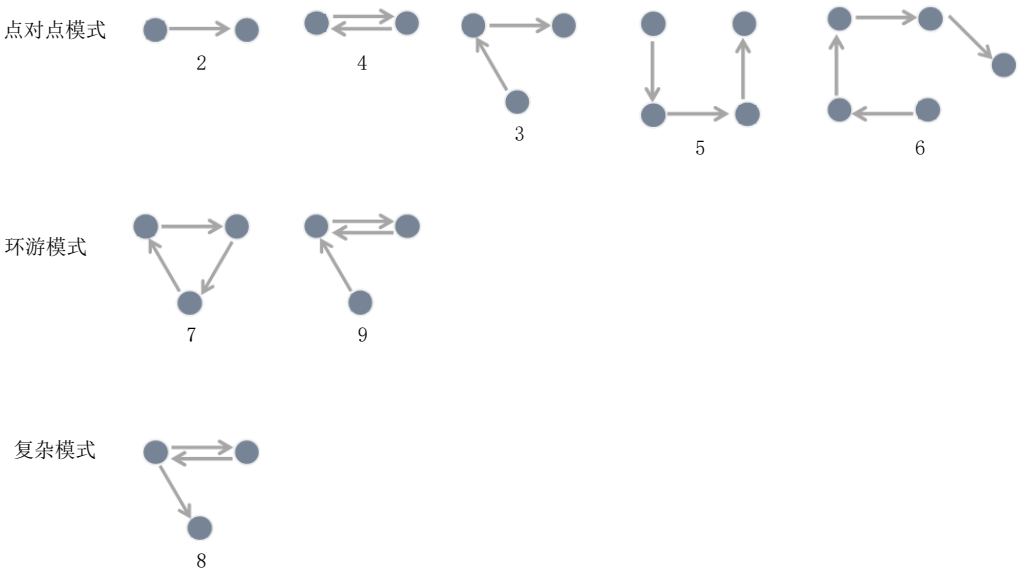


图 4-4 城市内尺度 motif 分类结果

各个模式的占比如下表 4-6：

表 4-6 城市内景点尺度模式占比表

模式	占比
点对点	8.02%
环游	0.11%
复杂	0.06%

从景点间的游客出行的空间模式中可以看到，大多数的游客选择的都是比较简单的出行方式，比如简单的点对点模式以及环游模式。选择以复杂模式进行游览的游客还比较少，这个模式有点像城市尺度下的基营式出行模式，而且也偏少。但是，就所有游客在景点间的出行空间模式而言，这些出行也和学者提出的模式有很大的重合度，在一定程度上也可以验证了游客出行的空间模式。

从上述的流程中得到以下的结果：

（1） 发现的模式没有超出原有理论模式。发现模式可以归类到理论模式中，没有不可用理论模式的类型进行解释的 motif。因此，发现模式很好的表征出理论模型的几种类型的特点。

（2） 理论模式都包含了发现模式。理论模式共分成了三大类型，点对点模式、环游模式、复杂模式，而发现模式的 motif 都可以归类到这三种类型种，所以理论模式都包含了发现模式。理论模式中还细分了一些小模式，这些小模式有的没有发现，但是大模式已经全部被发现。

（3） 发现模式频率的集中程度。从发现模式的统计中可以看到，点对点模式占据了绝大部分，其他两个模式虽然占比较少，但是还是存在的。所以，发现模式和理论模式都存在相同的模式分类，两者可以互相印证。所以发现模式在一定程度上验证了理论模式。

4.3 城市和景点尺度的游客出行模式的应用

4.3.1 城市目的地类型判定

可以根据游客在城市间的出行模式来判断城市的目的地类型，有利于城市对自身的了解，也有利于对城市的发展和规划指定决策。根据 2.4.1 中的方法，对目的地类型城市进行分析，以枢纽城市为例，从游客在城市间的出行空间模式中选出具有枢纽型节点的 motif，用这些 motif 对游客出行的路线进行统计，并对节点城市功能进行分析，即可得到枢纽类型的城市。

所以选中的 motif 的 ID 为：4，6，8.10，11，12，13，14，15，16，17。这些 f 是可能出现枢纽类型节点的 motif。对用户的行为进行判定，得到如下的结果 4-7：

表 4-7 城市目的地类型判定结果表

城市	枢纽	门户	出口	途径	归类
上海市	3417	1118	1091	1975	枢纽
成都市	1070	242	340	498	枢纽
广州市	788	316	285	773	枢纽
昆明市	652	222	284	184	枢纽
三亚市	479	43	133	337	枢纽
咸阳市	445	192	168	208	枢纽
长沙市	335	139	122	229	枢纽
贵阳市	246	59	81	96	枢纽
郑州市	238	128	134	238	枢纽
乌鲁木齐市	176	46	33	65	枢纽
哈尔滨市	165	70	53	165	枢纽
太原市	156	62	74	118	枢纽
海口市	156	55	54	88	枢纽
兰州市	88	51	48	87	枢纽

银川市	56	34	26	52	枢纽
揭阳市	51	30	25	38	枢纽

从上表可以看到，上海市、成都市、广州市、昆明市这几个城市作为枢纽型城市的特征十分明显。从现实的角度来看，排名靠前的几个城市的交通十分便利，都具有飞机、高铁、火车等路线通过，因此游客来旅行时作为门户和出口的选择比较多，计算的结果也十分符合现实情况。而后方的几个城市，虽然他们在城市类型判定上枢纽型的可能性也很高，但是因为城市本身参与游客出行轨迹的构成次数很少，所以这个类型的判定对于可能会有所偏差，可能是会作为单目的地类型的存在。

对于其他城市类型的判定也可以采用类似的方法来进行判断。如对苏州的判定结果分析如下表 4-8：

表 4-8 苏州市目的地类型分析表

城市	枢纽	门户	出口	途径	归类
苏州市	2247	796	697	5047	途径

其中，苏州市作为途径类型的目的地城市的次数最多，所以将其判定为路径类型，这个结果和苏州市旅游局给出的结果一致，说明使用游客出行的空间模式对城市的目的地类型进行判定，是可行的，具有应用价值的。

在本次实验中，因为单目的地类型的旅游路线占据了很大的比重，因此在执行判断规则时需要对其做特殊处理，比如上述枢纽型城市的选择时，可以去除了不可能出现枢纽类型的模式，避免其影响结果的占比。

4.3.2 旅游路线推荐

游客在城市间的出行空间模式，可以用于对城市的目的地类型进行判断。而在更小的尺度上，比如城市内部的景点尺度上，因为游客的出行基本上会和景点的分布有所关联，所以出行模式和城市景点的联系更加紧密。所以可以根据游客出行的空间模式来进行游客旅游出行路线的推荐。

按照 2.4.2 节的方法，我们以进行四个景点的点对点模式的游客的旅游路径作为分析的对象，用来聚类出 k=4 的情况下的优势路线。共筛选出了 1988 条路

线，最终聚成 4 类。共进行了 50 次聚类，取目标函数最小的结果，得到的三个中心的路线分别为：

(1) ['平江路', '观前街', '金鸡湖', '山塘']

(2) ['山塘', '寒山寺', '虎丘山', '拙政园']

(3) ['平江路', '观前街', '拙政园', '苏州博物馆']

(4) ['拙政园', '狮子林', '观前街', '寒山寺']

结合实际的数据，从得到的几个路线来说，路线中所提到的节点比较基本都是苏州市内比较出名的景点，而且从距离上来说，每天线路中的景点也比较接近，是比较常见的游览路线。另外，将这些结果和马蜂窝上的旅游推荐路线比较起来，也有一定的相似性。马蜂窝推荐的路线：

拙政园 →苏州博物馆 →平江路 →金鸡湖

寒山寺 →虎丘 →七里山塘 →山塘昆曲馆

寒山寺 →虎丘 →七里山塘 →山塘昆曲馆

拙政园 →苏州博物馆 →平江路 →金鸡湖

（马蜂窝推荐路线数据来源：路线来源：

http://www.mafengwo.cn/mdd/route/10207.html）

聚类结果和马蜂窝推荐的结果中都存在观前街、寒山寺、金鸡湖、拙政园、山塘、平江路这些景点，所以用这种方法进行旅游路线的推荐具有一定的实用性，模式的应用效果良好。

第五章 结论与展望

5.1 结论

本文主要提供了在大数据时代下运用社交媒体大数据对学者提出的游客出行空间模式理论进行实证研究的一个方法。本文首先总结了已有的游客出行的空间模式理论，并对一些学者进行的实证研究进行了分析，阐述这些实证研究方法的时代局限性。随后提出了在大数据时代使用社交媒体的数据来提取用户的旅游行为一个方法，并给出从用户发布的信息到用户每次旅行的旅游路线的一个提取方法；之后，借用复杂网络的 motif 的概念，将所有研究用户的旅游行为联系起来形成一个网络，利用 netmode 工具提取 motif，经过一些列的筛选之后，将结果和游客的实际出行路线进行匹配，得到占比较大的 motif，将其与已有的出行空间模式进行比对，验证结果。最后，使用苏州为例子，对上述过程进行了一次实证研究，并利用游客的出行空间模式进行了两个实际应用举例，表明再大数据时代游客的出行空间模式的应该采用新的思路，为游客的出行空间的决策以及游客本身出行体验的提高做一个参考。

根据上述的研究过程，本文所取得的研究成果如下：

(1) 从大数据中提取游客旅游行为，基于复杂网络 motif 概念进行游客的出行模式的识别与验证的方法是可行的。根据微博数据对方法的验证，城市间和城市内的游客出行模式得到了验证，验证效果良好，说明方法可行有效。

(2) 使用该方法在进行游客出行模式的验证效果良好，城市间的效果优于城市内的效果。对两个尺度的出行模式进行了验证，大尺度即城市间的出行模式验证结果中发现的模式和理论模式都能对应上，而且每一类的数量都比较多。在小尺度即城市内景点的出行模式验证时，发现的模式和理论模式在大的模式框架中时重合的，但是在细分的情况下，一些小模式没有完全覆盖。

(3) 该方法提取的出行模式可方便的进行具体的应用，且应用效果良好。进行了城市目的地类型的判别以及特定模式出行路线推荐两个具体的应用，得到的效果和已有的资料对比发现效果良好。因此，游客出行空间模式具有具体的应用价值，利于城市或景点加深对自身的了解，给游客的出行提供便利，为游客目的地的发展提供新的认识视角，为游客提供更加合理的出行方案等。

5.2 局限与不足

在本次实验中，虽然验证的结果比较理想，但是从实验过程中遇到的问题来说，这个方法也具有一些局限性与不足之处。

(1) Motif 提取方法耗时长，不同提取工具结果存在差异。目前用于提取 motif 的工具也比较多，比如 FanMod、Mfinder、MAVisto、MODA、Kavosh 和 NetMODE。这些提取工具内部的算法有些许差别，所以最终的提取结果有所差异，而这些差异有时候会比较明显。而且，不同的工具在提取不同节点的 moitf 的效率差别也很大，特别是当网络的节点很大时，发掘 motif 的过程就变成了非常耗时的过程。存在差异的结果和巨大的时间效率差异，使得选择哪个工具进行 motif 提取成为一个问题。

(2) Motif 结果不能反映多重边的特性。上述提到的 motif 提取软件没有处理多重边的问题。虽然在形成图时，它们所表现出来的形式时一样的，但是在现实的网络中还是会存在多重边的问题。比如游客在一次完整的旅游中重复经过同一个条路径几次，虽然这种情况比较少见，但是确实时存在的。而且在验证小尺度即景点尺度的出行空间模式时，点对点模式的子模式重复点对点模式，使用 motif 就不能表示。因此 motif 对于这些模式并不能很好的表示。

(3) 用户的客源地提取方式有待完善。实验的结果非常依赖于用户的旅游路线的提取，而如果用户的客源地提取的不正确，那么分割其地点轨迹将成

为一个问题，将会得到错误的旅游行为。如使用微博数据进行城市间的游客出行模式进行研究时，从 3.2 节的表格中可以看到，有些游客的一次出行中访问的超过 10 个城市以上的行为数也是不少的。出现这种情况的原因有两个，一个是客源地选取错误，造成分割结果不正确，另一个可能的原因就是用户确实是进行类似的旅游路线，如果是这种情况，就需要人工进行处理。不过对于这些路径，可以反馈到客源地的提取过程中，进行某种迭代的方式，寻找更有可能的客源地。

(4) 社交媒体数据的离散型强，在提取行为时存在缺失和稀疏的问题。社交媒体数据因为是离散类型的数据，用户每到一个点并不一定都会发布微博，所以存在缺失的问题，对于路径的生成会有所影响。这个问题在小尺度上表现得更加明显，在大尺度的范围中只要存在一条信息就能确定用户的位置，而在小尺度范围内发布的信息数更少，缺失会更加明显。由于社交媒体数据的这种潜在因素，用户的路径就会显得稀疏，不完全。

(5) 对于尺度比较小的出行模式的使用性较差。因为使用的是社交媒体大数据，得到路径会存在有些误差，数据缺失严重时影响路径的正确性。小尺度上的行为比较随意，用户的随机性强，和大尺度上的准确度相比会差一些。

5.2 展望

在本次实验中遇到了不少困难，比如 motif 的提取效果不是很理想，效率低下；不能反映多重边的性质；客源地提取方法不完善等，将来可以进行一些如下的尝试：

(1) 总结目前已有的 motif 提取的算法，寻找优化的方式。目前几年内使用 motif 这个概念进行研究的工作越来越多，也有不少新的理论正在形成，可以选择这些理论对 motif 提取算法进行优化，让其在节点数目更大的复杂网路中能够表现出更好的效率。

(2) 对于出现多重边的旅行行为单独分析，寻找合理的解决方式。在进行游客出行模式验证时，发现游客行为中存在一些多重边的行为，这些行为不同使用已有的 motif 进行分析。因此需要特殊对待，将其单独提取出来，分析多重边出现的原因以及对一些决策进行支持。

(3) 完善客源地提取的方法。可以采用多种方法和属性对客源地进行提取，使得结果更加真实准确。目前的方法中使用的时用户发布信息最多的城市作为客源地，将来可以添加时间熵等概念来辅助进行判定，这样的结果应该会比单一的用户信息数量这一个指标准确，有说服力。

参考文献

1. Milo, R.; Shen-Orr, S.; Itzkovitz, S.; Kashtan, N.; Chklovskii, D.; Alon, U. Network Motifs: Simple Building Blocks of Complex Networks. Science 2002, 298, 824–827

2. Schneider, C.M.; Belik, V.; Couronné, T.; Smoreda, Z.; González, M.C. Unravelling Daily Human Mobility Motifs. J. R. Soc. Interface 2013, 10, 20130246.

3. Shan Jiang, Joseph Ferreira, Jr., and Marta C. Gonzalez; Activity-Based Human Mobility Patterns Inferred from Mobile Phone Data: A Case Study of Singapore；2017

4. Liu Yang; Lun Wu; Yu Liu and Chaogui Kang; Quantifying Tourist Behavior Patterns by Travel Motifs and Geo-Tagged Photos from Flickr

5. Oppermann M. Length of Stay and Spatial Distribution [J]. Annals of Tourism Research, 1994, 21 (4): 834-836

6. Oppermann M. A Model of Travel Itineraries [J]. Journal of Travel Research, 1995, 33:57-61

7. Lue C, Crompton J L, Fesenmaier D R. Conceptualization of Multi-Destination Pleasure Trips [J]. Annals of Tourism Research, 1993,20:289-301.

8. Stewart S I, Vogt C A. Multi-destination Trip Patterns [J]. Annals of Tourism Research, 1997, 24(2): 458-461.

9. Lew A, McKercher B. Modeling Tourist Movements: A Local Destination Analysis [J]. Annals of Tourism Research, 2006, 33(2):403-423.

10. 杨新军, 牛栋, 吴必虎. 旅游行为空间模式及其评价[J]. 经济地理, 2000, 20(4):105-108.

11. 朱明, 史春云, 袁欣,等. 基于旅行社线路的国内旅行空间模式研究[J]. 旅游学刊, 2010, 25(9):32-37.

12. 史春云, 朱传耿, 赵玉宗,等. 国外旅游线路空间模式研究进展[J]. 人文地理, 2010(4):31-35.

13. 李旭, 马耀峰. 海外旅游者对旅游目的地和旅游路线的选择研究[J]. 陕西师范大学学报(自然科学版), 2003, 31(2):115-119.

14. 钟行明, 喻学才. 国外旅游目的地研究综述 --基于 Tourism Management 近 10 年文章[J]. 旅游科学, 2005, 19(3):1-9.

16. 李鑫. 网络图的 motif 发现算法研究[D]. 南开大学, 2013.

17. 牟廉明. 求有向图的所有 Hamilton 回路的新方法[J]. 计算机工程, 2007, 33(17):208-210.

18. 李锋, 商慧亮. 有向图的同构判定算法:出入度序列法[J]. 应用科学学报, 2002, 20(3):258-

262.

致谢

时间飞快，在这个美丽的燕园里已经度过四年时光。四年时间，足以对一个人的人生产生巨大的影响。从入学时的喜悦，大一时的无知，大二时的逃避，到大三时的觉悟，到大四的平稳，这是一个充满着神奇色彩的旅程。这一路上不断的收获，不断的成长，每次回首，总会发觉自己和曾经有所改变，有所成长。谨以此文，感谢这一路上给我帮助、支持的老师同学朋友和亲人们，感谢一切在这个旅途中遇到的人遇到的事，正是有了这些，证明我曾经于此处的存在，证明我曾经历过的无限美好、多彩的、绚烂的本科生活。感谢你们！

首先，我要感谢我的导师张毅老师。张老师治学严谨，学识渊博，在进行毕业设计的过程中给予我很大的帮助。当我遇到难题时，张老师总能给我提供参考，耐心为我分析存在的问题。不管是在论文方向，研究方法还是平时的生活态度上，张老师都给予我很多的指导和建议。张老师平易近人，幽默风趣，每次和张老师进行沟通时氛围都是十分和谐，感到轻松自然。十分感谢张老师在各个方面给我带来的影响，受益良多。

感谢在实验室大组中的刘瑜老师，邬伦老师，田原老师，高勇老师，黄舟老师等人以及班主任刘岳峰老师，感谢大组中的师兄师姐们。首先，刘岳峰老师作为班主任，对于我在生活上和学习上的状态非常关心，为人友善，为我们着想。在每周大组的分享报告中，师兄师姐们进行分享时都做的非常好，有的研究内容也给我的毕业设计提供了一些思路。感谢刘瑜老师每次在大组会时的分析与总结，刘老师通常都能给出很好的建议，同时，他对我们在学术中的要求也比较严格，对于我们的发展十分的重视与关心。另外，要感谢大组中的杨柳师姐，因为我们的研究方式存在相似之处，平时也会请教师姐问题，十分感谢师姐给我答疑解惑。

感谢实验室课题组的师兄和同学，包括王雯夫师兄，陈子豪师兄，孙奇师兄，吴梦彤同学。几位师兄在进行毕业设计时提供了很多技术上和理论上的支持，对于我遇到的问题也十分关心，整个实验室的氛围我也非常喜欢。感谢吴梦彤同学的思路交流，在遇到问题时也给了不少有创新性的见解。

感谢我的舍友以及 14 级 GIS 班的同学们，在这四年里面，我们互相帮助，互相鼓励，留下了许多美好的回忆。

感谢我的好友蒙维洒，覃欢谈，张幸宁，卢永乐，韦果等人。在四年里面时常相聚，互相鼓励，互相学习，让我度过了学校之外的美好时光。

最后，感谢我的父母，感谢你们多年来的养育之恩，即便远隔千里，仍然对我情绪的波动了若指掌，在我苦闷的时候，及时地开导我，安慰我，给我无微不至的关怀。常常叮嘱我吃早饭多运动多锻炼，教给我做人的道理和为人处世的方式。感谢你们对我的关爱和付出，愿你们永远健康快乐！

到这里，我的论文也结束了，四年本科生涯也落下帷幕。愿所有人都能有一个美好的明天，一个向往的生活。不忘初心，奋勇前行。

唐启浩

2018 年 6 月于北京大学