融合用户兴趣和传播欲望的两阶段位置推广算法

**摘要：**位置推广即在位置社交网络上帮助商家寻找种子用户发布广告，从而用最少的成本换来最大的广告收益。现有研究将位置推广定义为位置社交网络中的影响力最大化问题，并使用启发式算法或贪心算法选择种子用户。相比启发式算法，贪心算法选择的种子用户的影响范围更大，然而其时间复杂度也更高，不适用于大规模网络。此外，这些算法在评估种子用户的传播范围时，大多数仅仅利用用户流动性评估用户对位置的访问概率并将其作为用户间的信息传播概率，而没有考虑多种因素对用户间信息传播概率的综合影响，降低了传播概率度量的准确性，从而影响了选择种子用户的准确率。针对上述问题，本文提出了一种融合用户兴趣和传播欲望的两阶段位置推广算法（Two-Phase Location Promotion algorithm combining with user Interests and spread Desires，InDes-TPLP）。第一阶段，结合用户的社区结构在每个社区中选择具有高访问兴趣和传播欲望的用户作为候选种子集合，在保证种子用户影响范围和选种准确性的前提下降低贪心选种算法的时间复杂度。同时，本文提出了一种综合传播者影响力、用户间社交关系以及被传播者的访问概率三方面特征的多特征独立级联模型（Multi-Feature Independent Cascade Model，MFICM），并在第二阶段使用该模型度量候选种子用户的影响范围并基于贪心思想选择最终的种子用户，从而提高选种过程中传播概率的预测准确性。在真实的LBSN数据集yelp上的实验结果表明，本文提出的InDes-TPLP大大降低了选种过程的时间复杂度，并有效提高了所选种子的准确性。同时，其传播范围和覆盖率仅稍低于性能较好的NewGreedy贪心算法。

**关键词：**基于位置的社交网络；位置推广；影响力最大化；传播模型

# 引言

随着社交网络的不断发展，出现了越来越多的个性化社交网络，其中基于位置的社交网络(Location Based Social Networks，LBSNs)由于包含了用户签到和情感分享等多元的信息，受到越来越多用户的喜爱。典型的LBSNs有Yelp、Gowalla、Foursquare、Twitter、Flickr以及中国的MicroBlog、WeChat等。以Yelp (<https://www.yelp.com/sf>)为例，用户在对商家的点评中，不仅包含了点评内容还包含其他用户对点评的偏好。通过商家页面，可以获取商家的地理位置、类别、签到用户等信息。通过用户个人页面，可以获取用户的年龄等个人信息、朋友列表以及其他用户对当前用户的评价等。因此通过对LBSNs中多方面信息的挖掘，可以发现人们的出行规律以及兴趣爱好，为商家进行位置推广、为用户进行位置推荐、好友推荐，以及预测用户位置、好友关系等提供良好的支持。其中，位置推广作为一个新颖的研究方向受到越来越研究者的关注。

位置推广，即如何寻找最有效的目标用户群体，通过他们的影响力将一个新出现的位置推广出去，这类似于传统社交网络中的影响力最大化问题。传统社交网络中影响力最大化算法的目标是寻找能够将信息最大范围传播出去的目标用户群体。但是在LBSNs中，位置信息的传播不仅与用户间的线上社交关系有关，由于地理位置约束，也与用户的线下活动规律有关。[Zhu et al. 2014][1]首次将位置推广定义为LBSNs中的影响力最大化问题，其后很多研究者也基于此思路进行研究[Li et al 2014][2][Zhou et al. 2015][3][Wang et al. 2016][4][Hai et al.2015][5]。

位置推广涉及两个方面的关键问题：如何高效、准确地选择种子用户；如何从大量的网络数据中提取有价值的信息来准确度量用户间的信息传播概率。

针对第一个问题，现有研究主要采用启发式算法选种[Zhou et al. 2015] [3] [Wang et al. 2016] [4] [Zhu et al. 2015][6]，或贪心算法选种[Li et al. 2014][2] [Hai et al.2015] [5] [Li et al. 2017][7] [Li et al.2017][8] [Wang et al. 2017][9]。传统的贪心算法[Li et al. 2014][2] [Hai et al.2015] [5] [Li et al.2017][8]需要结合传播模型选择种子，因此算法时间复杂度很高，同时传播模型的准确性直接影响了选择种子的准确性。启发式算法虽然时间复杂度低，但其选择的种子影响范围较贪心算法低[Kempe et al. 2015][10]。因此，如何降低选种算法的时间复杂度，同时保证较大的种子影响范围仍然是一个挑战。

传播模型可以用来模拟信息在用户间的传播过程，是影响力最大化算法的关键部分。传统贪心算法使用传播模型度量用户的影响范围，从而选择影响范围最大的目标用户群体，而启发式算法利用传播模型评估种子选择的效果。现有影响力最大化算法主要基于独立级联模型（Independent Cascade Model，ICM）和线性阈值模型（Linear Threshold Model，LTM）模拟信息的传播过程。但在LBSNs中，每个用户访问位置的差异性较大，用户很可能仅受到一个朋友的影响就被激活。因此，位置推广算法大都不使用LTM模型，而使用ICM模型及其改进模型[Zhou et al. 2015][3][Zhu et al. 2015][6]模拟位置信息的传播过程。

传播模型的关键在于如何准确度量用户间的信息传播概率。有些研究者[Li et al.2014][2][Wang et al.2016][4]将传统社交网络中度量传播概率的方法引入到位置社交网络中。Zhu等[Zhu et al. 2016][11]认为，LBSNs中的传播概率随目标位置的变化而变化，但传统社交网络中的静态传播概率不能合理度量这种变化，因此他们提出了两种移动模型，即基于高斯分布的移动模型（Gaussian-based mobility models, GMMs）和基于距离的移动模型（Distance-based mobility models, DMMs）。但是，他们只考虑了用户的地理位置的移动性，没有考虑到不同位置的类别对传播概率的影响。Yu[Yu et al. 2017][12]认为用户对不同类别的位置的偏好不同，Zhu[Zhu et al. 2016][11]认为结合位置的类别能够更准确地度量传播概率。但是，上述算法大都将传播者或者被传播者对位置的访问概率作为传播概率，没有考虑用户兴趣偏好和位置自身吸引力对传播概率的影响，也忽视了用户间潜在的社交关系以及传播者对位置信息传播的影响，导致传播概率的度量不够准确。

本文认为，用户访问位置会考虑多种因素，不仅包括用户与签到位置的距离，还包括诸如位置的类型、商圈、用户对位置的评价等等。用户签到位置的类型和用户评价等反映了用户的兴趣偏好以及用户的个体差异。同时，用户发布签到往往体现了用户对该位置具有较高的传播欲望，而位置信息的传播效果与用户间的社交关系以及兴趣相似度有关。因此，在度量传播概率时，应该综合考虑上述多种因素，从而提高选种算法的准确性。另一方面，社交网络存在社区结构，社区内部节点的紧密度较高，社区之间的节点紧密度较低，结合社区结构选种[He et al. 2015][13] [Zhao et al. 2016][14] [Gong et al. 2016][15]能够有效避免影响力的重叠，从而提高种子的传播范围。

基于上述考虑，本文提出了一种融合用户兴趣和传播欲望的两阶段位置推广算法来更加高效、准确地选择种子用户，以便进行位置推广。第一阶段结合用户的社区结构选择具有高访问兴趣和传播欲望的用户群体候选种子集合，以缩小选种范围，提高选种准确性并降低算法时间复杂度。同时，本文提出了一种综合传播者影响力、用户间社交关系以及被传播者的访问概率三方面特征的多特征独立级联模型，在第二阶段使用该传播模型度量候选种子用户的影响范围并基于贪心思想选择最终的种子用户，从而提高选种算法的准确率。

本文的主要贡献有：

1）针对贪心算法的高时间复杂度，提出了一种融合用户兴趣和传播欲望的两阶段位置推广算法（Two-phase Location Promotion algorithm combining with user Interests and spread Desires，InDes-TPLP），在保证种子用户影响范围和选种准确性的前提下，降低了选种算法的时间复杂度。

2）针对现有传播模型对位置社交网络信息挖掘不充分，导致传播概率度量不够准确的问题，利用位置社交网络的多种特征，提出了一种多特征独立级联模型（Multi-Feature Independent Cascade Model，MFICM），可以更准确地度量位置信息的传播概率。

本文后续章节的内容如下：第二章介绍了位置推广相关的研究工作；第三章给出了位置推广问题的相关定义以及InDes-TPLP的算法框架；第四章和第五章分别对本文提出的MFICM传播模型以及InDes-TPLP选种算法展开详细介绍；第六章对本文方案的效果进行了实验验证和分析；第七章给出了本文的结论。

# 相关工作

在传统社交网络中，影响力最大化问题是指在特定传播模型下，获取一个指定大小的种子节点集合，使得该集合在网络中的聚合影响力最大。在现有位置推广相关研究中，位置推广大多被定位为位置社交网络中的影响力最大化问题。由于缺乏用户的地理位置信息，传统社交网络中的影响力最大化算法不能直接应用于基于位置的社交网络。因此相比传统社交网络中的影响力最大化问题，位置推广涉及更多方面的技术，包括影响力最大化、信息传播模型、用户行为预测和位置推荐等。近些年，有些研究者将影响力最大化算法和社区发现算法相结合，可以有效避免种子节点影响力的重叠问题。本小节将从影响力最大化、信息传播模型和社区划分三个方面分析位置推广相关的研究工作。

## 影响力最大化

传统社交网络的影响力最大化算法主要基于两种思想：启发式算法和贪心算法。启发式算法根据一定的规则选择种子用户，然后通过传播模型模拟种子用户的传播范围，在确定选种规则的情况下，最终的种子用户是一次性确定的。不同于启发式算法，贪心算法需要结合指定的传播模型动态选择种子用户，且每次只选择具有最大传播范围的节点加入到最终的种子集合中，因此算法时间复杂度较高。

Domingos等[Domingos et al.2001][16]首次提出影响力最大化问题。其后，Kempe等[Kempe et al.2003][17]将该问题定义为最优化问题，并提出了基于贪心爬山思想的通用算法框架来解决此类问题。然而，他们的算法效率非常低。为了提高算法的效率，Leskovec等[Leskovec et al.2007][18]提出了CELF算法，并利用子模结构优化了选种时评估节点影响范围的方法，降低了种子选择的时间复杂度。Chen等[Chen et al.2009][19]提出了NewGreedy算法，首先将图中概率较小的边提前剪枝，然后在剪枝后的图中计算每个节点的影响范围，通过多次迭代模拟信息的传播范围。该方法有效降低了选种算法的时间和空间复杂度，算法效率优于CELF。Goyal等[Goyal et al. 2011][20]在CELF算法的基础上，提出了CELF++算法。该方法通过维护当前轮次的节点最大增益影响范围，避免了CELF在迭代过程中的重复计算带来的时间复杂度，有效提高了算法的效率。但是CELF和CELF++由于需要保存每个节点的增益影响范围，因此，需要很大的内存开销。

一些研究者利用基于网络拓扑的启发式算法来研究影响力最大化问题，如基于度中心性、中介中心性、接近中心性等。但是这类启发式算法依赖于网络的拓扑结构，不能和信息的传播过程有效结合，并且存在种子节点的影响力重叠问题。为了降低基于度中心性的选种方法存在的影响力重叠问题，Chen等[Chen et al.2010][21]提出了度折扣算法，即以度中心性选择初始种子，并在后续种子选择的过程中，对和种子集中的节点存在连边的种子的度进行相应折扣。另有一些研究者提出了结合社区结构的选种算法，通过在不同的社区中选择种子用户来降低影响力的重叠，从而提高种子用户的影响范围。He等[He et al.2015][13]基于社区划分的结果，在每个社区中选择中心性较高的节点作为种子节点，同时为了避免后续节点和种子节点的影响力重叠问题，他们对后续节点进行相应的折扣。Zhao等[Zhao et al.2016][14] 首先在网络中寻找度大的节点，然后以他们为初始节点进行标签传播，通过标签传播范围度量用户影响范围。Gong等[Gong et al. 2016][15]基于社区划分结果，选择度较大的节点形成种子向量，并对向量进行交叉变换后，通过衡量向量的影响范围得到种子集合。

上述基于传统社交网络的影响力最大化算法只考虑了用户间的线上社交关系，没有考虑用户线下签到位置的影响。因此，一些结合了用户位置信息的影响力最大化算法被提出。

Zhu等[Zhu et al.2014][1]提出了两种启发式算法，分别基于签到活跃用户和专家用户选种。但该算法不能区分用户在不同位置上的影响力进行选种。为了解决针对不同位置选择种子用户的问题，Zhou等[Zhou et al. 2015][3]提出了一种基于两阶段传播模型的启发式算法（Two-Phase Heuristic Algorithm，TPH），将用户的影响过程分为线上和线下两个阶段，依次计算用户的线上和线下传播概率，并选择具有高传播概率的用户作为种子用户。Wang等[Wang et.al 2016] [4]基于最大树形影响力模型（Maximum Influence Arborescence，MIA）求解种子用户，但是这种启发式算法会偏离具有真正影响力的节点。

部分研究者基于贪心思想求解位置社交网络中的种子用户。Li等[Li et.al 2014] [2]提出了区域影响力最大化算法，将网络按照区域划分，根据每一对节点间的传播概率找到每个区域中具有最大影响力的路径和节点，从而确定种子用户。但这种方法最终选择的种子用户与区域划分的依赖性较大。Hai等[Hai et.al 2015][5]借鉴贪心思想，逐步选择具有最大影响增益的种子用户，但仅仅度量了用户对一跳邻居用户的影响力，忽略了信息从邻居继续向前传播的过程，对用户影响力的传播范围评估不够准确。Li等[Li et al. 2017][7]首次提出了最大地理跨度区域（Maximum Geographic Spanning Regions，MGSR）问题，即通过用户的最大地理传播范围来度量用户的影响力。他们认为，用户的影响力不仅取决于用户所影响的人数，同时也与用户所影响的地理区域有关，覆盖区域和激活人数更大的用户的影响力更强。但是，他们的方法和Li等[Li et.al 2014] [2]的方法一样，都只能求解某个区域内具有高传播影响力的种子用户，不能针对指定位置点发现种子用户，忽略了不同位置点的差异性。Wang等[Wang et al.2017][9]提出一种基于逆影响力抽样（Reverse Influence Sampling，RIS）的方法度量用户影响力，通过逆向随机抽样找到对抽样节点具有最大可达性的用户群体，同时通过对采样集合的大小进行限制，降低了算法的时间复杂度。但是，上述研究均假设每个用户拥有一个固定不变的位置，并通过用户的居住位置选择具有最大影响力的用户。然而，位置社交网络中用户的签到位置往往变化多样，具有高影响力的用户也并非集中于待推广位置附近，这种假设不能体现用户在LBSN中的移动性，影响选择种子节点的准确性。针对贪心算法的高时间复杂度问题，Li等[Li et al. 2017][8]基于改进的贪心算法CELF[Leskovec et al 2007][18]选择种子用户，但是他们的目标在于同时推广连锁店商家的多个位置。

考虑到基于贪心算法选择的种子的影响范围较大，本文结合用户的线下活动以及位置信息，基于贪心思想选择种子用户。同时考虑到贪心算法存在时间复杂性高的问题，与上述文献不同，本文利用用户的兴趣偏好和传播欲望并结合社区结构在各个社区中提前筛选出候选种子用户，缩小利用贪心思想选种的范围并避免影响力重叠问题，从而降低算法时间复杂度并提高选择种子的准确性。

## 信息传播模型

在传统社交网络中，主要基于两种传播模型及其改进模型度量用户的影响范围，即独立级联模型（Independent Cascade Model，ICM）和线性阈值模型（Linear Threshold Model，LTM）[Kempe et al. 2003][17]。ICM和LTM的区别在于，LTM需要考虑已激活邻居的叠加影响，而ICM只考虑用户间的独立激活行为，可以更容易地模拟信息的传播过程。因此，现有位置推广相关研究普遍通过改进ICM来模拟位置信息的传播过程，从而度量用户的影响范围。

Zhu等[Zhu et al. 2016][11]认为，进行位置推广的关键在于对传播概率的度量。针对传统社交网络，一些研究将传播概率设置为统一的值[Goyal et al. 2011][20]、被传播者的入度均分[Wang et al. 2016] [4], Kempe et al.2015] [10], Chen et al.2009[19], Goyal et al.2011] [20]、两个用户间的共同好友数[Goyal et al.2011] [20]等。但LBSN中传播的是位置信息，信息传播会受到位置信息的特殊影响，因此仅仅通过用户间的线上社交关系不能准确度量用户间的信息传播概率。

Zhu等[Zhu et al. 2015][6]基于ICM提出了两种针对位置社交网络的动态移动性模型来度量传播概率，即基于高斯分布的GMMs和基于距离分布的DMMs，并将DMMs进一步细化为基于用户自身规律的自身移动性概率DMM-Basic和考虑用户好友关系影响的DMM-Social。他们将这两种动态传播概率和包括基于签到的用户Jaccard相似度、Cosine相似度及Bernoulli分布在内的8种静态传播概率进行对比，发现动态传播概率的效果要优于静态传播概率，基于距离分布的DMMs要优于GMMs，且DMM-Social的预测相比DMM-Basic更准确。这是因为GMMs主要考虑了用户的签到位置分布，没有考虑用户签到顺序，而DMMs不仅考虑了用户历史签到位置和目标位置的距离分布，还考虑了用户签到位置的顺序。但是，他们的方法没有考虑位置分类对用户偏好的影响。

Barbieri等[Barbieri et al. 2013][22]认为用户在不同主题上的偏好不同，结合传统社交网络中媒体信息的主题对ICM和LTM进行改进，提出Topic-aware Independent Casca（TIC）模型和 Topic-aware Linear Thresho（TLT）模型。Zhu等[Zhu et al. 2016] [11]认为LBSN中的位置分类类似于信息的主题，用户在不同的位置类别上的兴趣偏好也不同，因此进一步考虑了位置的分类信息度量传播概率，结果表明，考虑了位置分类的传播概率的准确率更高。同时考虑到用户访问不同位置存在差异，传播概率应随目标位置的变化而变化，他们提出了一种位置感知的独立级联模型（Location-aware Independent Cascade Model，LICM）。LICM改进了ICM，当目标位置固定后，LICM即具有和ICM相同的属性。LICM的传播过程如下：所有在*t=0*时刻在目标位置*l*签到的用户将会传播位置信息。如果用户*u*在*t*时刻已激活，那么他的朋友*v*将会在*t+1*时刻以概率*pp(u,v|l)*被激活，且激活过程是独立进行的。但是，他们认为传播者只起到对信息的分享作用，用户间的传播概率与传播者的影响力无关，只与被传播者对位置的访问概率有关，因此将传播概率定义为被传播者对位置的访问概率。

实际上，在LBSN网络中，用户的签到不仅起到分享信息的作用，也会对其朋友产生一定的潜在影响力，且不同的传播者对于既定位置的影响力是存在差异的，因此仅仅考虑被传播者的访问概率不够全面和准确。Zhou等[Zhou et.al 2015][3]提出了一种结合线上和线下访问概率的传播模型TP。他们将节点的激活过程分为两个阶段，用户只有在线上浏览到其朋友的签到记录，并且受到影响才会被线上激活；当且仅当用户线上激活时，其才有机会线下激活邻居节点。但是对线上和线下传播概率的关系不好区分，研究者也没有对线上传播概率进行深入研究，只是简单设置几个统一的值进行实验，因此准确性不高。

可见，上述文献在度量信息传播概率时，大都只通过简单的用户签到的距离分布度量用户对位置的访问概率，并以此作为位置信息的传播概率。虽然有些研究者在此基础上考虑到了位置分类的影响，但是考虑仍然不够全面。本文认为，在LBSNs中，用户间的位置信息传播概率受到用户间社交关系、传播者影响力、用户自身偏好等多方面因素的影响，结合多种特征度量传播概率会更加准确。

## 社区发现

基于全局社交网络选择的种子用户可能会聚集在某几个大的社区中，存在影响力重叠问题。如2.1节所述，一些基于社区结构选择种子用户的算法[He et al. 2015][13] [Gong et al. 2016][15] [Zhao et al. 2016] [14]，能够有效避免种子节点聚集在同一个社区中，通过社区划分降低了影响力的重叠，相比启发式算法能够得到更大的影响范围，同时可以降低算法的时间复杂度。社区结构体现了网络中用户的聚集性，社区内用户的联系紧密，社区间用户连接稀疏，因此信息在社区内的传播较容易，在社区间的传播较困难。结合社区结构在不同社区中选择种子节点，能够提高信息的传播范围。但是，上述方法在社区结构的基础上仅结合网络的拓扑结构和节点中心性选择种子用户，没有考虑节点的语义信息。本文希望结合用户的访问兴趣和传播欲望等语义信息在社区中选择种子，增加在既定位置选种的准确性。

现有的社区发现算法可以分为重叠社区发现算法和非重叠社区发现算法[Fortunato et al.][23]。在非重叠社区发现算法中，Lovain算法[Blondel et.al 2008][24]能够快速、准确地发现紧密的社区结构。Lovain算法是一种基于模块度最大化的算法，模块度Q用来衡量网络中各社区内部联系的强度。Lovain算法最初将每个节点分为一个簇，每次将社区外的节点纳入使得整个网络模块度变化最大的社区中，直到社区不再发生变化。为了避免选择的种子用户产生重叠，本文考虑选择非重叠社区发现算法Lovain进行社区划分。

综上所述，针对位置推广问题，现有研究对LBSN中信息的挖掘不够全面，降低了传播模型对用户传播范围评估的准确性。同时，现有算法大都基于用户的居住位置以及用户签到行为的移动性来选择种子用户，而没有结合用户对既定位置的偏好，影响了选种的准确性。此外，基于贪心思想的选种算法虽然传播范围更大，但是时间复杂度也很高，影响选种的效率。

本文认为，在LBSNs中，结合用户间的社交关系、兴趣偏好以及位置的自身属性能够提高传播概率度量的准确性；在划分社区的基础上，结合用户的访问兴趣和传播欲望筛选种子能够降低选种过程的时间复杂度，同时提高选种的准确性。

# 基础准备

在本节中，我们首先给出LBSN网络、位置推广以及传播模型的定义，然后给出了我们提出的位置推广方法框架。

## 3.1 位置推广定义

现有研究主要基于用户和位置两种实体构建位置社交网络，但是，这样建模忽视了在签到或评论中隐含的用户影响力。实际上，用户的朋友对用户签到的评论和点赞在一定程度上反映了用户的影响力，忽略这些因素会降低度量用户影响力的准确性。因此，为了更好地度量用户间的关系，参考刘等[Liu.et.al 2015][25]对位置社交网络的定义，同时考虑签到记录和位置本身的多种信息，本文定义LBSN网络如下：

*定义3.1(LBSN)。*LBSN定义为无权图，如图1所示，包含用户、签到和位置三种实体。为用户集合，包含用户的所在城市、签到数量、朋友数量等属性；为签到记录集合，包含用户签到的时间、签到所获得的点赞数量、评论数量等属性；为位置集合，包含位置的经纬度坐标、分类等属性；为用户间的关注关系，代表用户和用户存在关注关系；为用户签到关系，代表用户*u*访问了位置，同时留下签到记录，。



图1 基于位置的社交网络

Fig. 1. Location based social network

定义了网络后，本文对位置推广的定义如下：

*定义3.2 (位置推广)*。给定位置社交网络以及待推广位置，通过位置推广算法得到在网络中对于位置具有最大传播影响力的个种子用户集合，使得他们的影响范围最大，即受到这些种子用户的影响而访问位置的用户数量最大。

在度量用户影响范围时，需要结合传播模型来模拟用户影响力传播的过程。本文认为，用户往往容易被与其关系亲密、兴趣相似并且自身具有一定权威的用户所影响。因此，为了更准确地度量传播概率，本文对LICM[Zhu et al. 2016][11]中的传播概率进行改进，定义了一种多特征独立级联模型（Multi-Feature Independent Cascade Model，MFICM）来模拟用户的影响力传播过程。

*定义3.3 (MFICM)。*给定位置社交网络、种子集合（最初访问位置*l*的用户，即初始激活用户）以及待推广的位置，对于任意，*s*在*0*时刻开始传播签到信息，如果用户*u*在*t*时刻被激活，即产生签到行为，则用户*u*在*t+1*时刻将以概率尝试激活其邻居节点*v*，其中。尝试激活的过程是独立的，并且只有一次尝试激活的机会。上述传播过程结束后，种子集合激活的用户即的传播范围，也就是产生新的签到行为的用户集合。

MFICM和LICM的区别在于，LICM认为传播概率即被传播者对位置的访问概率，而MFICM考虑传播者的影响力、用户间关注关系以及用户对位置的访问概率三方面因素，并基于此提出了3种度量传播概率的方法，具体度量方法将在4.2节展开介绍。

## 3.2 InDes-TPLP算法框架

位置推广的目的是选择能够将目标位置推广出去的种子用户，这就要求种子用户具备三个条件：1）自身对目标位置感兴趣；2）具有传播欲望，所见所闻乐于分享给其他用户；3）具有高传播影响力。

如果用户对目标位置不感兴趣，那么用户访问目标位置的概率会很低，更不会向他的朋友推广该位置。对于访问兴趣高但签到欲望低的用户，即使用户访问了目标位置，但由于没有将信息分享出去，目标位置也无法得到推广。所以，位置推广需要在满足条件1、2的基础上，通过影响力最大化算法来选择满足条件3的用户群体。

现有位置推广相关研究主要采用贪心算法或者启发式算法选择种子用户。由于贪心算法需要在选种过程中评估用户的影响范围，从而逐步选择具有最大增益影响力的用户加入种子集合，因此时间复杂度较启发式算法高，但是影响范围比启发式算法大。同时，传统贪心算法需要评估所有用户的影响范围，从而找到具有最大影响范围的用户作为种子用户，这样就会对很多不具备条件1、2的用户产生大量冗余计算。

为了能够在保证贪心算法的较大影响范围的基础上，降低算法时间复杂度，本文提出一种融合用户兴趣和传播欲望的两阶段位置推广算法（Two-Phase Location Promotion algorithm combining with user Interests and spread Desire，InDes-TPLP）。在第一阶段，通过度量用户的访问概率和传播欲望对不具备条件1、2的用户进行提前剪枝，筛选出候选种子用户，同时为了避免候选种子用户的影响力重叠，结合社区结构在不同社区中选择候选种子，以增大候选种子的影响范围。在第二阶段，利用MFICM度量候选种子用户的影响范围，并利用基于贪心思想的影响力最大化算法从候选种子用户中选择具有最大影响范围的种子用户，从而在保证种子具有较大传播范围的前提下降低选种算法的时间复杂度。

另一方面，现有位置推广算法大都基于用户居住位置以及用户移动性评估位置信息的传播概率。本文认为，位置信息在用户间的传播概率不仅取决于用户的签到移动性，更取决于传播者的影响力、用户间的兴趣相似度以及被传播者对目标位置的访问概率。因此，在第二阶段选择种子时，本文使用第4章提出的MFICM模型来度量位置信息在用户间的传播概率，从而提高评估候选种子用户传播范围的准确性。

InDes-TPLP框架如图2所示，包括4部分：1）构建3.1节中定义的位置社交网络，包含用户间的社交关系信息以及用户对位置的签到记录；2）对构建好的用户社交关系网络进行社区划分，以便能够在不同的社区中选择候选种子用户，降低种子用户影响力的重叠，并增大其影响范围；3）基于用户对目标位置的访问概率评估用户对目标位置的访问兴趣，并基于用户访问兴趣、传播欲望以及社区划分结果，在不同社区内选择候选种子用户；4）利用MFICM传播模型评估每个候选种子的影响范围，并利用基于贪心思想的影响力最大化算法找到具有最大影响范围的k个用户作为最终的种子用户集合。



图2 InDes-TPLP总体框架

Fig.2. Framework of InDes-TPLP

下面将展开介绍MFICM模型和InDes-TPLP算法。

# MFICM传播模型

在传统社交网络中，研究者大多使用独立级联传播模型（ICM）模拟信息的传播过程[Kempe et al. 2015][17][Chen et al. 2009][19] [Goyal et al. 2011][20]。在度量ICM模型的信息传播概率时，有些研究者考虑到用户所受影响来源于不同朋友，将传播概率定义为被传播者的入度均分[Goyal et al. 2011][20][Li et al. 2017][7]。这种方式潜在地认为用户均等地接受不同用户的影响力，同时对于不同的位置，传播概率不具有差异性。Zhu等[Zhu et al. 2016][11]认为不同位置的传播概率应该是不同的，因此在ICM的基础上定义了位置感知的独立级联模型（LICM）。但他们将传播概率定义为被传播者对位置的访问概率，并认为传播者只起到分享信息的作用，因此不同传播者对于被传播者的影响力也是相同的。但是真实生活中，不同的传播者对于既定位置的影响力是存在差异的，上述方法对传播概率的度量不够准确。

本文认为，用户往往容易被与其关系亲密、兴趣相似并且自身具有一定权威的用户所影响。因此，本文在LICM的基础上，综合考虑传播者的影响力、用户间的社交关系以及用户对位置的访问概率三方面因素度量传播概率，提出了一种多特征独立级联模型（MFICM），旨在提高传播概率的准确性，从而提高利用传播模型评估用户传播范围的准确性。

MFICM包括三部分，即激活规则、传播概率和终止条件。下面将对MFICM传播模型展开介绍。

## 4.1 激活规则

当用户*u*看到朋友发布的签到位置时，可能受到朋友的影响访问该位置，即该用户被激活，这个影响过程是独立进行的。用户*u*在被激活的同时有且仅有一次机会以概率激活其未激活的朋友*v*，其中在4.2节中介绍。不论用户*v*是否被激活，用户*u*都不再有机会对用户*v*进行激活。

## 4.2 传播概率

传播概率是传播模型的关键要素，对传播概率的准确度量不仅能够增加传播模型的准确性，同时也能够增加用户传播范围评估的准确性。考虑到用户是否受到朋友影响而访问目标位置与传播者的影响力、用户间的关系以及被传播者的访问概率都相关，本文结合上述三方面影响因素提出了三种度量传播概率的方法：

1），与现有研究相同，将用户对位置的访问概率作为传播概率，即忽略不同传播者传播影响力的差异性。

2），在考虑用户对位置的访问概率的基础上，进一步考虑了传播者与被传播者的兴趣相似度，即认为不同的社交关系对于传播概率存在不同的影响。

3），在前两者的基础上考虑了传播者的影响力，即不同传播源对于传播概率存在不同影响。

### 4.2.1 将用户对位置的访问概率作为传播概率

用户是否受朋友分享的位置信息的影响而产生签到行为，与用户对目标位置的访问概率密切相关。用户对目标位置的访问概率越大，则其产生签到行为（被激活）的概率越大。Ye等[Ye et al.2011][26]认为，在基于位置的社交网络中，传播者的影响力不重要，可以忽略。和他们的理论相同，Zhu等[Zhu et al.2015][6]使用GMMs和DMMs来度量用户对位置的访问概率，并将其作为传播概率。上述文献在计算用户对目标位置的访问概率时，主要考虑了用户签到位置的移动性以及用户社交圈的影响。Zhu等[Zhu et al.2016][11]进一步考虑了位置的分类因素，认为用户在不同类型位置上的访问概率存在差异，但是他们没有兼顾位置自身的吸引力。同时这些方法基于概率分布来计算用户对位置的访问概率，对不同网络的普适性不强。

为了更准确地度量传播概率，本文结合用户的历史访问兴趣、签到移动性以及位置自身的吸引力三方面的特征来度量用户的访问概率，并将用户的访问概率作为传播概率，即：

 (1)

利用机器学习中的监督回归模型对上述三方面的特征进行学习。用户是否访问位置用0/1标签表示，通过回归模型对目标函数进行拟合。以逻辑回归为例，上述三种特征构成特征向量，则目标函数为：

 (2)

其中，为特征参数向量，通过最大似然估计学习其取值，即可得到用户*u*对位置*l*的访问概率。下面将对（2）中的三方面特征进行度量。

（1）用户的历史访问兴趣

为了观察用户访问兴趣的差异性，我们从yelp数据集中提取了包括美食、住宿、娱乐等15类商家位置类型信息，图3为其中top-9个用户在不同商家位置类型的签到分布。

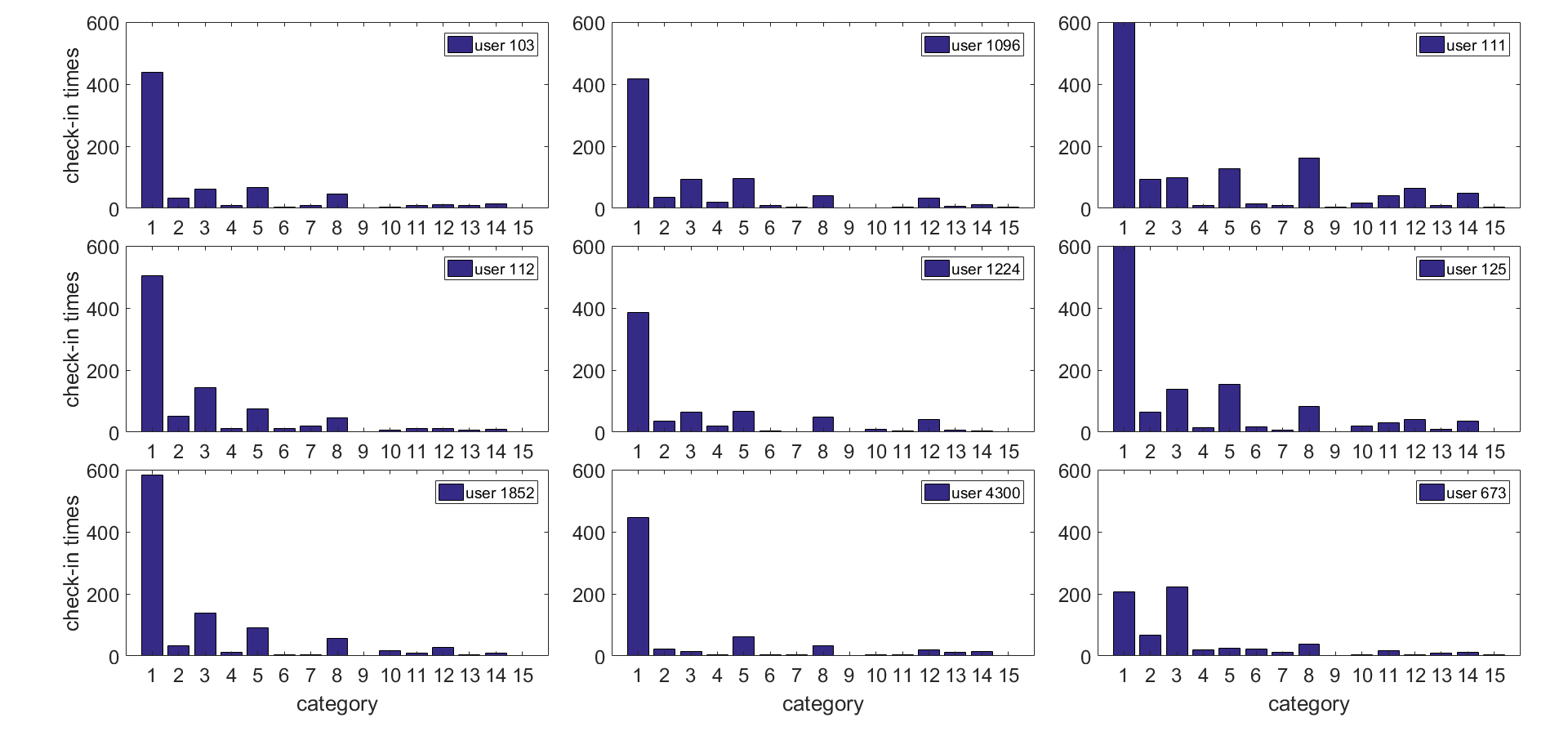


图3 用户的历史访问兴趣分布图

Fig**.** 3. Distribution of users’ historical interests

从图3中可以看出，不同用户对不同商家位置类型的偏好不同，同一用户在不同商家位置类型上的访问偏好也不同。因此，本文定义用户的访问兴趣为用户和位置的兴趣重合度，即该商家位置是否为用户的偏好类型。用户的访问兴趣定义为用户的历史签到中与待推广位置的类型相同的位置数量的比例：

 (3)

其中，表示用户*u*的签到集合，表示位置的类型。

（2）签到移动性

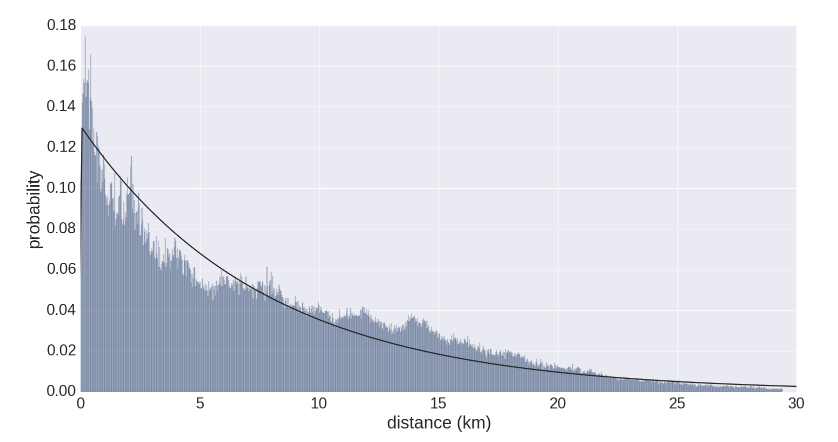
签到移动性描述了用户位置和待推广位置的距离。研究表明，用户历史访问距离的分布呈现幂律分布[Zhou et al. 2015][3] [Li et al. 2014][2]、指数型分布[Wang et al. 2016][4] [Wang et al. 2017][9]，或pareto分布[Zhu et al. 2016][11]。为了研究yelp数据集中用户签到移动性的规律，我们统计了用户按时间升序的相邻签到记录的位置移动距离，如图4所示。

图4 用户连续签到距离分布

Fig. 4. Distance distribution of continuous check-in

图4表明，yelp数据集中的用户签到移动性呈现指数型分布。因此，本文定义签到移动性为：

 (4)

其中，表示用户的上一次签到位置和待推广位置间的距离，为参数。

（3）位置吸引力

代表了位置自身的因素对用户的吸引力，吸引力越大，用户对位置的访问概率越高。比如，对于一些热门的商圈，用户的访问数量较大，对于比较小众的商圈，用户的访问数量相对较低。Hsieh等人认为[Hsieh et al.2015][27]，位置所在区域的密度和地点熵对用户的访问概率有一定的影响。因此，我们将yelp数据集中用户的签到位置分布使用BaseMap可视化工具进行展示，如图5所示。图5中每个点代表一个商家，颜色代表商家位置的类型，点的大小代表商家签到人数的多少。从图5可以发现，商家所处的区域内商户越多，越容易吸引用户；同时，商户所处的区域内商户类型分布越多样，也越容易吸引用户到访。因此，本文结合区域密度和地点熵来衡量位置的吸引力。

 (5)

其中，区域密度代表该位置*l*所在区域内的用户签到数量，区域密度越大，表示访问该位置的用户越多，该位置越吸引用户来访。

 (6)

其中，为网络中的位置集合，代表与位置距离不超过的区域中的位置集合，此处。

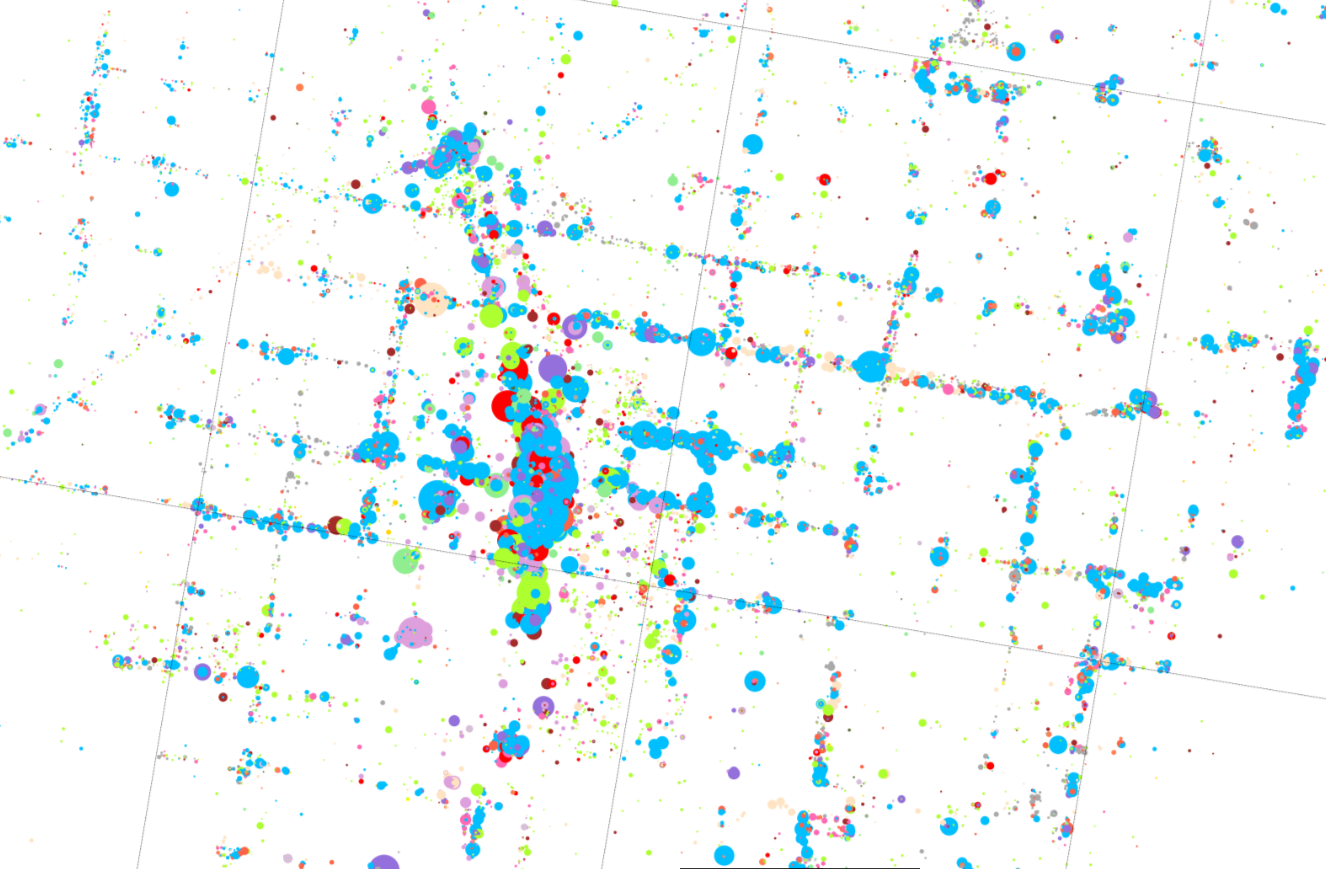


图5 用户签到兴趣以及地理分布

Fig. 5. User’s check-in interests and geographical distribution

地点熵反映了位置所处区域内商家位置类型的多样性，区域内位置类型越多样，并且各个类别的位置数量越多，则该区域的熵值越大。熵值大的位置类似于综合商场，能够吸引更多用户的到访。

 (7)

### 4.2.2 结合用户访问概率以及签到相似度度量传播概率

在基于位置的社交网络中，可以通过用户的历史签到相似度判断两个用户是否会在相同地点产生签到行为，相似度越高，相似用户被激活的可能性越大。因此，在考虑用户对位置的访问概率的基础上，进一步结合用户间的签到相似度，能够更准确地度量位置传播概率。定义结合用户访问概率和签到相似度的传播概率为：

 (8)

用户签到相似度由用户签到位置的地理分布的相似度和位置类别的相似度组成，即：

 (9)

其中，[Zou et al.2015][28]为用户*u*和用户*v*共同签到的位置数，该值越大，说明两个用户访问相同位置的可能性越大，也间接体现了两个用户的生活规律和地理位置的相似性。

代表用户签到兴趣的相似性，即用户在不同类型位置上签到分布的相似程度。兴趣相似性越高，用户访问相同位置的可能性越大。这里使用余弦相似度来度量用户间签到兴趣的相似程度：

 (10)

其中，表示用户在类型位置上的签到集合，为网络中位置类别的总数量。

### 4.2.3 结合传播者影响力、访问概率及签到相似度度量传播概率

用户对从不同渠道得到的信息的接受程度是有所区分的。比如，在传统社交网络中，明星用户发布的博文更具有影响力。位置社交网络也同样，有经验的用户分享的信息更易被其好友接受。因此，本文认为，结合传播者的影响力可以提高传播概率的准确率，因此在考虑用户访问概率和用户间签到相似度的基础上，进一步考虑传播者的影响力，定义传播概率为：

 (11)

考虑到用户历史签到中互动的人数越多，说明用户的知名度和影响力越高，这类用户对于信息的传播更具有优势。同时考虑到不同用户对位置的访问兴趣不同，知名度高但是访问兴趣低的用户，对于推广该位置的贡献将相应降低。因此，本文结合用户在特定位置上的传播欲望、访问概率以及在位置社交网络中的互动情况计算用户的影响力。

 (12)

其中，为用户在位置*l*的综合影响力，为用户的社交影响力，即其在社交网络中的互动情况。

由(2)中定义的用户访问概率和传播欲望组成，即：

 (13)

 (14)

传播欲望的定义源于传统社交网络中的用户活跃度定义。式（14）代表用户*u*的签到数与网络中用户的平均签到数的比例，该值越高，代表用户在位置网络中的活跃度越高，越愿意将自己的到访位置分享给朋友。同时，通过调研yelp数据集，我们发现用户的传播欲望呈现长尾分布，即较多用户签到数量较低，而较少用户拥有较多的签到数量。因此，式（14）结合阶跃函数，将少部分传播欲望过低的用户过滤掉。

 (15)

用户*u*的社交影响力通过用户u历史签到中其所有朋友对其的点赞数量所占比例来评估，点赞表明了朋友对用户的认可和支持，收到的点赞越多，说明该用户自身的影响力越大，越容易受到朋友的信任。式（15）中，*FLE(u)*为用户*u*的朋友集合。

## 4.3 终止条件

当用户*u*的邻居中没有可以尝试激活的节点（所有节点都遍历过，不管激活成功与否），并且用户u的已激活邻居的邻居中没有可以尝试激活的节点，依次类推，所有与u存在可激活连接的节点都被遍历过时，本次传播终止，得到用户u的影响范围，即通过u直接或间接激活的邻居数量。

# InDes-TPLP算法

InDes-TPLP算法利用位置社交网络中用户对位置的访问兴趣和传播欲望筛选候选种子用户，并使用MFICM传播模型度量位置信息的传播过程，能够更准确地评估被激活的用户数量，通过两阶段选种过程能够更高效地选择能激活最大范围用户的种子用户。其中，用户对位置的访问兴趣通过4.2.1节（2）定义的用户访问概率*A(u, l)*来衡量，用户对特定位置的访问概率越，表示用户对该位置的访问兴趣越大。传播欲望*S(u)*则由4.2.3节（14）定义。本小节将基于3.2节给出的InDes-TPLP框架展开介绍该算法的各部分。

## 5.1 第一阶段：选择候选种子用户

为了选择具有高访问概率和高传播欲望的用户作为候选种子用户，同时为了避免选择的种子用户间的影响范围重叠问题，本文首先对用户关注关系网络进行社区划分，然后在不同的社区中筛选出访问概率和传播欲望都较高的用户作为候选种子用户。这样，由于第二阶段只需从候选种子用户中选择最终的种子用户，能够大大减少评估影响范围时涉及的用户数量，从而降低阶段二利用基于贪心思想的影响力最大化算法选择最终种子用户的时间复杂度。

### 社区划分

现有很多研究将用户按照活动区域进行聚类，然后在不同区域中选择种子用户。但是，在线社交网络中距离较远的用户之间也可能存在较大的影响力，按照区域划分用户会限制用户线上信息的传播。另一方面，考虑到用户个人居住位置属于用户的隐私，很多社交网络不会记录用户的具体居住位置，获取用户的居住位置信息存在一定难度。同时，用户虽然对某些类别的位置存在一定偏好，但是这种偏好不局限于固定类别的位置。因此，基于位置区域对用户进行聚类存在一定的局限性。

在社交网络中，用户主要通过朋友圈内好友分享信息。关注关系网络是用户传播信息的重要载体，用户的关注关系很大程度上体现了用户间的社交关系。因此，为了有利于信息的更广泛传播，不同于现有基于位置区域进行用户聚类的方法，本文首先对用户的关注关系网络进行社区发现，使得关系紧密而不是位置相近的用户能够聚集在一个社区中，然后从不同社区中选择种子用户。

Louvain算法[Blondel et.al 2008][24]是一种基于多层次优化模块度的社区划分算法，其优点是快速、准确，是目前社区划分算法中表现最好的算法之一。Louvain算法分为两个阶段，第一阶段将每个节点加入使得社区模块度增益最大的社区中，直到所有节点的社区归属不再变化；第二阶段将社区内的节点合并成为一个超节点，并更新网络的边权。迭代这两个步骤直至算法稳定。因此，本文选取Louvain算法对用户关注关系网络进行社区发现。

### 基于访问概率和传播欲望选择候选种子

本文的最终目标是选择k个种子用户来推广位置，因此，候选种子的数量将直接影响最终种子用户的选择。候选种子越多，阶段二的选种范围就越大。更大的选种范围使得选种算法可以通过评估更多用户的传播范围来综合选择种子用户，从而提高种子用户的传播范围。但是，需要评估的用户节点越多，算法的计算时间复杂度也越高，同时选择过多低访问概率和传播欲望的用户也将降低算法选种的准确性。因此，本文引入参数来控制选择候选种子的比例：

 (16)

其中，为选取的候选种子的数量。即根据指定的种子数量和参数确定候选种子的数量。

特别的，当时（*N* 为网络中用户数），选择的候选种子数量*K*为网络中的用户数量*N*，即在全局网络范围内选择种子用户；当时，候选种子数量*K*为最终选取的种子数量*k*，相当于采用启发式算法选择种子用户，即根据访问概率和传播欲望选择种子用户。

考虑到在真实社交网络中存在很多孤立用户，他们和其他用户的连接性较低，通过线上的社交关系很难对他们产生影响。同时，他们对于信息的传播贡献较小，即使他们有很高的传播欲望和访问概率，但是通过他们分享出去的信息的覆盖面较小。因此，本文采取基于社区规模选择候选用户的策略，社区规模越大，选择的候选种子的数量也越大，这样能够避免将这些孤立点选作候选种子，同时也可以保证在大规模社区中获得更大的影响范围。

根据社区规模在各个社区内选择的候选种子数量如下：

 (17)

其中，为社区i中的用户数量。

当确定了候选种子的数量后，下一个关键的问题就是按照什么样的标准选择候选种子。基于4.2节对种子用户应满足的条件的论证，InDes-TPLP算法第一阶段的目标即要寻找对目标位置有高访问兴趣和传播欲望的用户，即根据(13)定义的包含了用户对位置的访问概率*A(u, l)*和传播欲望*S(u)*的综合影响力，和(17)指定的候选种子数量在各个社区中选择影响力最大的个候选种子。

候选种子选择算法如Algorithm 1。

|  |
| --- |
| **Algorithm 1:** Candidate seeds selection |
| **Input:** , target location, seed\_num, parameter  **Output:** ; Candidate seeds  1: ; Calculate the number of candidate seeds  2: ; Calculate the number of candidate seeds in every community  3: **for** every community  4: **for** every node *u*:  //Calculate the comprehensive influence of node *u* to locationaccording to (13)  5:  6: **end for**  7: Selectusers with max  8: ←{}  9: **end for** |

## 5.2 第二阶段：选择最终种子用户

本阶段的目标是在第一阶段得到的个候选种子中，基于影响力最大化算法选取个具有最大传播影响力的最终种子用户集合。

由于启发式算法依赖于网络拓扑，传播范围受限，因此本文考虑基于贪心算法进行第二阶段选种。研究表明，贪心算法选择种子能够达到最优解的63%[Kempe et al. 2015][17]。但是，贪心算法由于要借助传播模型评估节点的影响范围，时间复杂度非常高。在本文中，由于通过第一阶段选种已经筛选出了候选种子节点，因此大大减少了需要评估的节点数量，从而能够有效降低贪心算法选种过程的时间复杂度，同时也能够保证良好的选种效果。为了有效提高贪心算法的效率，本文使用Chen等[Chen et al.2009][19]提出的NewGreedy算法进行第二阶段选种，该算法对不能被成功激活的边提前剪枝，能够有效降低算法的时间复杂度。为了更准确地评估用户的影响范围，本文使用第3节提出的MFICM传播模型模拟位置信息的传播过程。

选择最终种子用户的算法如Algorithm 2。

|  |
| --- |
| **Algorithm 2: Final seeds selection** |
| **Input：**, target location *l*, seed\_num *k*, candidate seeds, iterations  **Output：**, Final seeds  1: Initializeand  2: **for**to**do**  3: setfor all, initialize the influence score for all nodes in the candidate seed set while not in the final seed set  4: **for**to**do**  5: Computeby removing each edge fromwith probability  6: Compute the influence rangeof seed setusing MFICM  7: Computefor eachusing MFICM  8: **for** each vertex**do**  9: **if****then**  10:  11: **end if**  12: **end for**  13: **end for**  14: set  for all  15:  16: **end** **for**  17: Output |

## 5.3 选种过程示例

假设最终选择的种子数量为k=2，InDes-TPLP算法的选种过程如图6所示。图6（a）为对用户关注关系网络进行社区划分的结果，不同颜色的用户节点代表其归属不同的社区。根据候选种子选择算法，得到4个候选种子，如图6（b）中用红色标注的节点1、5、6和18。这些候选节点的传播范围如图6（c）、（d）、（e）、（f）中紫色节点所示，即传播范围分别为4、10、11和7。注意，此处示例的传播范围为一次迭代计算所得的结果，最终结果应通过多次迭代取平均值。基于上述四个候选节点的传播范围分值，首先选择传播范围最大的6号作为第一个种子节点。在选择第2个种子节点时，由于5号节点已经包含在6号的传播范围中，如图6（e）所示，因此不再考虑。1号和18号候选种子的传播范围见图6（g）和（h），分别为4和7。因此，第18号节点被选为第二个种子节点。由于最终只需要选择2个种子节点，因此选种过程结束，由此得到最终的种子集合为6号和18号用户节点，最终的传播范围为16。

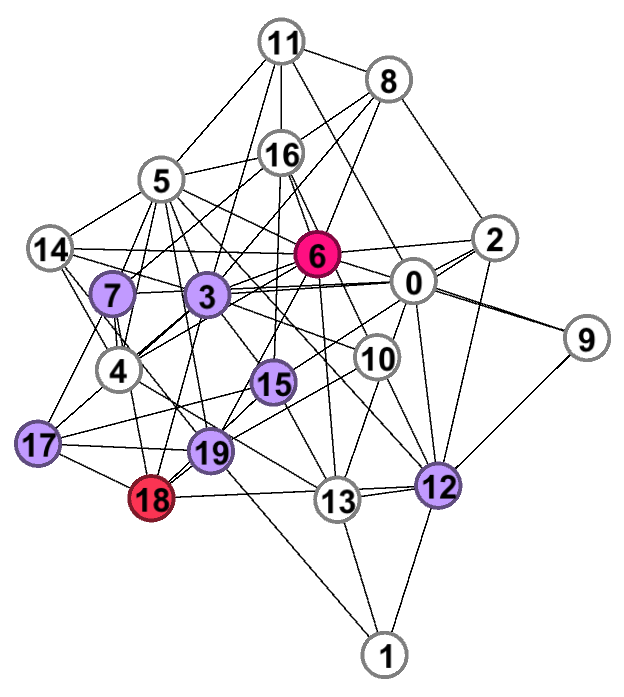
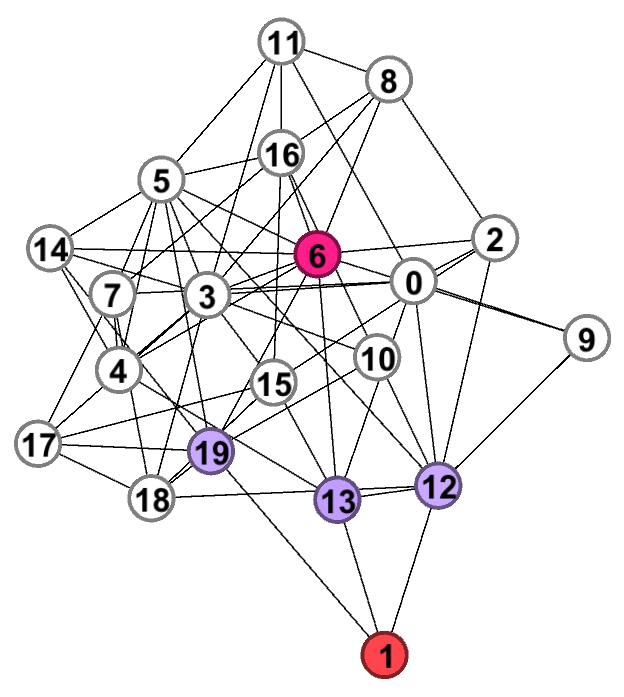
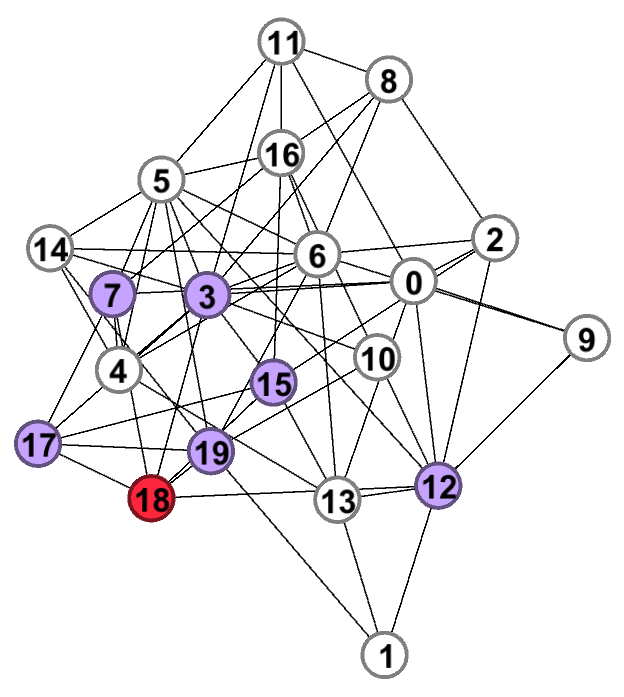
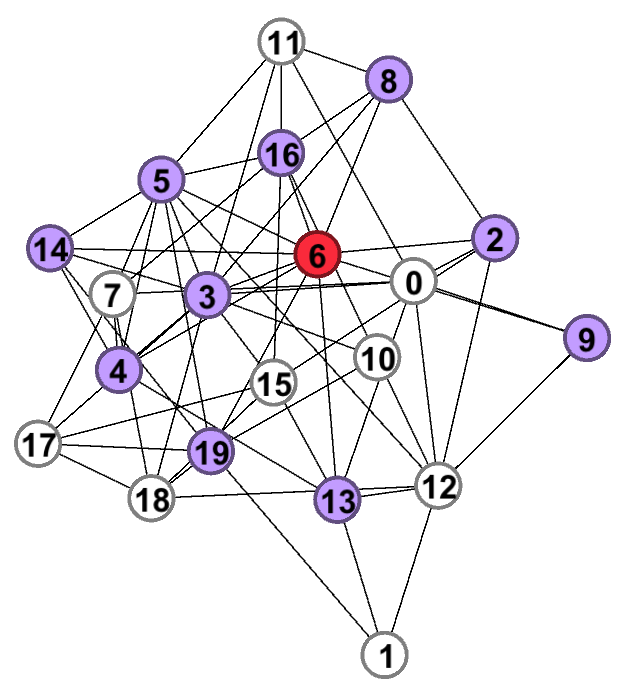
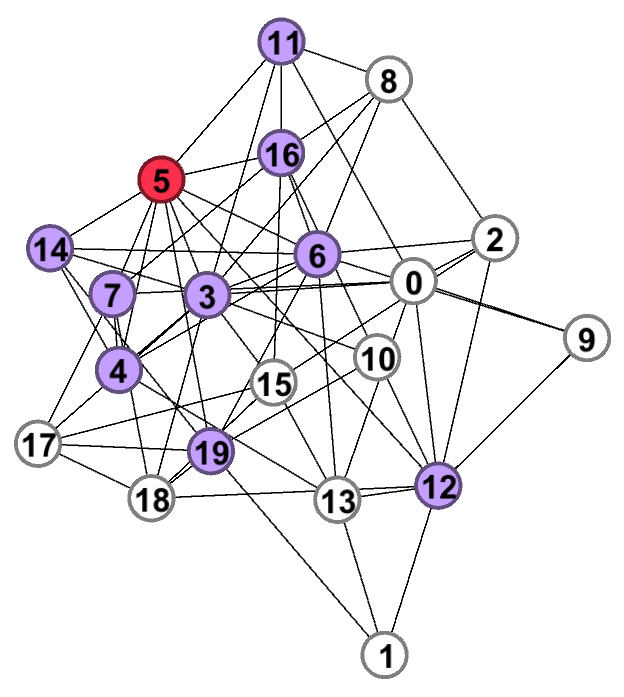
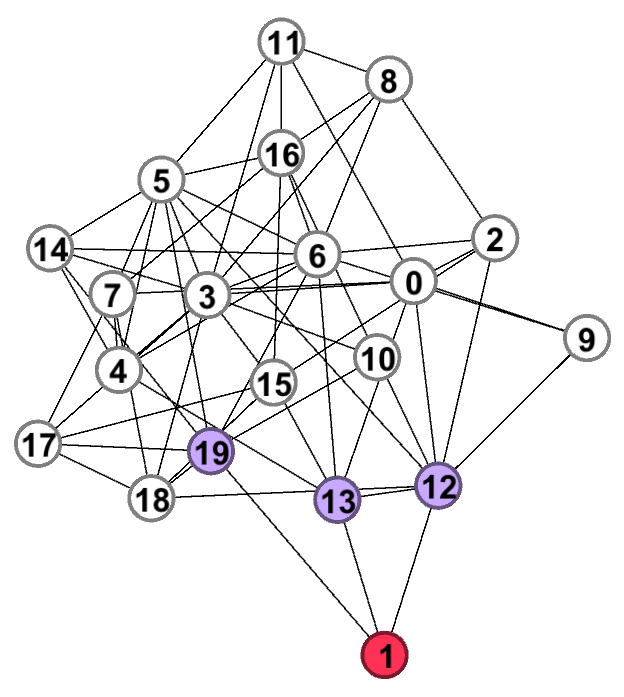
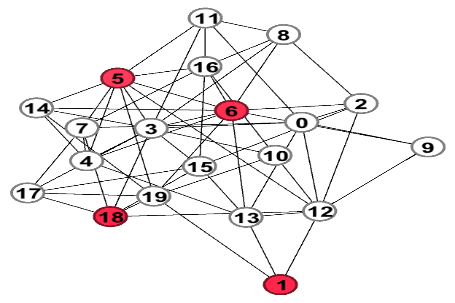
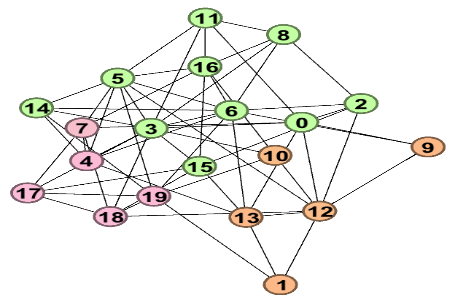


图6 贪心算法选种示意图

Fig.6. Seed selection based on Greedy

(h)

(g)

(f)

(e)

(d)

(c)

(b)

(a)

## 5.4 时间复杂度分析

传统的贪心算法需要评估每一个节点的影响范围，且每个节点的影响范围需要通过多次迭代计算得到。当指定种子数目和迭代次数后，每选择一个种子，都需要通过迭代次来评估不在当前种子集合中的网络节点的影响范围，然后取平均值作为当前节点的影响范围。最后，选取其中影响力增益最大的节点作为第个种子。所以，对于节点个数为，边数为的网络，评估每个节点的影响范围的时间复杂度为，每选择一个种子的时间复杂度为，选取个种子的时间复杂度为。

为了降低时间复杂度，本文选取了NewGreedy算法[Chen t al. 2009][19]进行选种。在该算法中，每选择一个种子，首先从关注关系网络中移除不可达的边，然后在新的图中迭代次计算当前种子集合中每一个节点的影响范围。最后选取平均影响力得分最大且不在最终种子集合的节点为新的种子节点。所以，每选择一个种子节点，通过在随机图上进行一次BFS或DFS算法即可得到节点的传播范围，算法的时间复杂度为，选择个种子需要。

在本文算法中，由于第一阶段选种已经选择了候选种子，所以在评估每一个节点的影响范围时，不需要对全局图进行BFS，而只需要对候选种子集合中的节点进行BFS。因此，每选择一个种子的时间复杂度为，其中，*d*为网络中节点的平均度。由于，，因此，选择个种子用户的时间复杂度，仍为线性时间复杂度。

# 算法评估及结果分析

## 6.1 实验方案

### 数据集

现有针对位置推广的研究大多仅仅使用了用户的社交关系以及用户签到记录的经纬度信息，因此往往使用Gowalla、Brightkite、Foursquare等数据集进行实验验证。随着LBSNs的发展，位置社交网络中包含了更丰富的信息，如位置的分类、评论等信息。为了结合更丰富的语义信息和用户信息对本文方案进行评估，本文使用信息相对全面的Yelp公开的比赛数据集(<https://www.yelp.com/dataset>)进行实验验证。

Yelp是美国最大的点评网站，囊括各地餐馆、购物中心、酒店、旅游等领域的商户，用户可以在Yelp网站中给商户打分，提交评论，交流购物体验等。本文提取了Las Vegas地区的409,314个用户（去除孤立点后剩余191,057个用户）和24,118个商家以及1,324,960条签到记录。

### 评估指标

在现有算法中，研究者主要从影响范围和时间复杂度两个方面评估位置推广算法的性能。但是由于影响范围是通过传播模型进行度量的，因此受传播模型的准确性影响，影响范围与真实的由于受到影响而访问目标位置的用户数量并不完全一致。

Hai等人[Hai et al. 2015][5]提出了一种新的评估方法，即通过度量影响范围与真实访问位置的用户数量的比例来评估位置推广算法的效果。但是，他们只考虑了两者在数量上的差异，没有考虑用户是否真实访问了目标位置，评估方法仍然不够合理。比如，当影响范围大小和真实访问位置的用户数量接近，然而影响范围中真实访问目标位置的用户数量很小时，直观上算法的效果就很差，但是他们的指标却可能接近1。

受到Hai等人的启发，本文认为位置推广的目的在于最大化访问目标位置的用户数，所以对位置推广算法的性能评估不能仅仅局限于影响范围，而应结合真实访问目标位置的用户数来评估算法的效果。因此，我们结合分类预测算法中的评估指标，准确率和召回率，来评估本文提出的位置推广算法和传播模型的效果。由于我们需要评估种子用户的覆盖范围，所以本文将召回率记做覆盖率。准确率和覆盖率高，说明传播模型能够更准确地评估用户的影响范围，相应地，采用该传播模型的推广算法选择种子的准确性更高、覆盖范围更大。

另外，为了更好地和Zhu等人[Zhu et al.2015][6]提出的传播模型进行对比，本文也选取了他们采用的Log-likelihood指标来评估传播模型预测用户访问概率的准确性。

本文涉及的评估指标的定义如下：

* Log-likelihood：该指标用于对比不同的用户访问概率预测方法的效果，Log-likelihood越大，传播模型更好地表示了用户的真实签到。
* influence spread(影响范围)[Li et al. 2014][2] [Wang et al.2016][4] [Li et al.2017][7] [Wang et al.2017][9]：通过传播模型模拟的种子用户所激活的用户数。值越大，算法的效果越好。
* coverage：种子用户所激活的用户中真实访问目标位置的用户数量与网络中所有访问目标位置的用户数量的比值。值越大，算法的效果越好。
* accuracy：种子用户所激活的用户中真实访问目标位置的用户数量与影响范围的比值。值越大，算法的准确率越高。

### 实验方案

本文从用户访问概率预测、传播模型、选种算法三个方面进行实验验证。具体实验方案如下：

实验1：评估本文提出的用户访问概率预测方法的效果。本文按照对比方案DMMs[Zhu et al. 2015][6]的实验设置，提取具有不同签到记录数的用户进行实验，评估本文提出的用户访问概率预测方法对不同活跃度用户群体的预测效果。对于每一个用户，随机选取与其真实访问的位置相同比例的未访问位置作为反例，学习其对位置的访问概率，并使用逻辑回归方法预测用户访问目标位置的概率。

实验2：评估MFICM传播模型的影响范围、准确率和覆盖率。根据用户的历史签到信息和用户关注关系，从每个用户的邻居用户中，找到与用户存在相同签到的用户，然后随机抽取等量的没有共同签到的用户作为反例，根据4.2节的传播概率度量方法计算三种不同的传播概率，并利用决策树模型学习用户间的信息传播概率。最后，随机选取相同数量的种子用户，通过实验对比MFICM与ICM、LICM传播模型的效果。为了避免单次随机选种的不确定性，本文重复了10次选种试验，并对结果取平均值。

实验3：评估本文提出的InDes-TPLP选种算法的效果。由于本文方案在阶段一使用了Louvain算法进行社区发现，因此根据社区划分的实验结果，本文确定了社区模块度最优时的参数，即将Louvain算法的resolution设置为1.0。然后，通过实验确定InDes-TPLP算法的参数，以选择一个最佳的选种比例。最后，基于实验2的结果，选择效果最好的传播模型（即MFICM）进行选种实验，从影响范围、准确率和覆盖率三个方面评估选种算法的性能。

## 6.2 实验结果和分析

### 6.2.1 用户访问概率的预测效果

DMM-Basic[Zhu et al.2015][6]算法基于Pareto分布模拟用户移动性，并基于用户移动性度量LBSN中用户间的信息传播概率。DMM-Social[Zhu et al. 2015][6]在DMM-Basic的基础上考虑了用户社交圈因素，认为用户的访问概率受到用户移动性和社交圈因素的共同影响，因此采用了一种混合Pareto分布模拟用户移动性。DMMs模型能够较好地模拟位置信息的传播过程，但是其认为用户间的信息传播概率等同于被影响者对位置的访问概率。本文的位置推广方案同样考虑了用户对位置的访问概率，但结合了用户的访问兴趣、签到移动性、位置的吸引力来预测用户的访问概率Pv。

按照实验1的设计，本文提取了签到数量分别为表1所示数值的用户，共计3899个，分别采用本文提出的用户访问概率预测方法Pv和Zhu等[Zhu et al.2015][6]提出的DMM-Basic和DMM-Social方法预测不同签到活跃度用户的访问概率，并使用Log-likelihood评估算法的预测效果。实验结果如图7所示。

表格 1 Yelp数据集中具有不同签到数量的用户数量



图 7 用户访问概率预测算法对比

Fig.7. Results of different user access prediction methods

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **签到数量** | **10** | **20** | **30** | **40** | **50** | **60** | **70** | **80** | **90** | **100** | **总用户数** |
| 用户数 | 2842 | 577 | 196 | 111 | 67 | 37 | 24 | 14 | 15 | 16 | 3899 |

从图7可以看出，本文方案Pv明显好于对比方案DMM-Basic和DMM-Social。原因是本文方案结合了用户的访问兴趣以及位置自身吸引力等多种因素来度量用户对位置的访问概率，更符合实际情况，而DMMs仅基于距离分布度量访问概率，因此准确性比Pv低。

### 6.2.2 MFICM传播模型的影响范围、准确率和覆盖率

不同ICM模型的区别主要在于度量信息传播概率的方法。根据Zhu等[Zhu et al. 2015][6]的实验结果：在静态传播概率度量方法中，Unify、In-degree、Jaccard-L、Bernoulli的效果较好；动态传播概率的效果要好于静态传播概率，且在动态传播概率度量方法中，基于距离分布的DMM-Basic和DMM-Socail的效果好于基于地理位置分布的GMM概率度量方法。

因此，为了评估MFICM传播模型的效果，我们从本文基于yelp数据集构建的位置社交网络中随机选择种子，并以相同的种子集合作为初始传播者，然后分别在具有三种传播概率Pv、Psv、Pisv的MFICM，具有Unify、In-degree、Jaccard-L、Bernoulli传播概率的ICM以及具有DMM-Basic和DMM-Social传播概率的LICM上进行传播，并利用影响范围、准确率和覆盖率三个指标对这些传播模型的效果进行对比评估。

其中的传播概率设置如下：

* Unify[Goyal et al.2011][20]：使用统一的传播概率，P=0.1。
* In-degree[Wang et al. 2016[4], Chen et al. 2009[19], Goyal et al.2011[20]]：使用被影响者的朋友关系的入度均分作为传播概率，P=1/deg\_v。
* Jaccard-L[Goyal et al.2011][20]：，即两个用户签到的Jaccard相似度。
* Bernoulli[Sacharidis et al.2014][29]：，其中为用户*v*对用户*u*产生跟随签到行为的次数，即在用户*u*的签到中，对用户*v*产生影响的签到的比例。
* DMM-Basic[Zhu et al. 2015][6]：基于用户签到记录的移动性计算用户的访问概率，并将被影响者的访问概率作为传播概率。
* DMM-Social[Zhu et al. 2015] [6]：在DMM-Basic的基础上，结合用户社交因素，利用EM算法学习用户的访问概率，并将被影响者的访问概率作为传播概率。
* Pv：使用被影响者对目标位置的访问概率作为传播概率，P=Pv。
* Psv：在Pv的基础上，同时考虑用户间的签到相似度的传播概率，P=Psv。
* Pisv：在Psv的基础上，同时考虑传播者的影响力的传播概率，P=Pisv。

为了避免随机选种产生的偶然性，我们对基于每一种传播概率的传播模型重复了10次试验，并对结果取均值。图8展示了ICM、LICM和MFICM三种传播模型在不同传播概率下的影响范围、准确率和覆盖率。其中，不同的传播概率表示采用相应传播概率的传播模型。

|  |
| --- |
| 1. influence spread with different number of seeds |
| 1. accuracy with different number of seeds |
| 1. coverage with different number of seeds |

图 8 不同传播模型的传播效果对比

Fig.8. Results of different propagation probabilities with different number of seeds

从图8可以得到如下结论：

（1）随着种子数量的增加，所有传播概率下传播模型的影响范围和覆盖率都呈现上升趋势，且不同传播概率的差异基本稳定。其中，考虑了多种因素预测传播概率的Pisv的传播效果最好，其次是考虑用户间相似度的Psv。而基于入度均分的In-degree表现最差，甚至低于统一的传播概率Unify。虽然In-degree和Unify的影响范围随着选种数量的增加有一定提升，但其覆盖率的变化并不明显，且这两种算法的覆盖率不及Pisv的10%。说明忽略位置信息和用户签到信息的传播概率度量方法不适于评估LBSNs中位置信息的传播概率。

（2）基于动态传播概率Pisv和Psv的传播模型的效果稍好于基于静态传播概率Bernoulli和Jaccard-L的传播模型的效果。其中，Pisv的覆盖率比Bernoulli和Jaccard-L高出14%，Psv比两者高出10%。这是因为静态传播概率忽略了不同位置对于目标人群的影响。但Jaccard-L和Bernoulli由于考虑到了用户在签到位置上的行为相似性和跟随性，其传播效果明显好于只考虑网络拓扑的静态传播概率In-degree和Unity。

（3）考虑了用户签到距离分布的DMM-Basic和DMM-Social（简称DMMs）在影响范围、准确率和覆盖率上差异不大，且只比Psv和Pisv的效果稍差。这是因为这四种方法都基于用户签到距离的pareto分布度量传播概率。但是基于DMMs的传播模型的效果明显优于基于Pv的传播模型，说明DMMs在度量信息传播概率方面具有优势。另一方面，从6.2.1节中的访问概率预测结果可以看出，Pv对用户的访问概率预测结果明显好于DMMs。因此综合分析基于静态传播概率的传播模型的结果可以得出，不能简单地将传播概率设置为被传播者对位置的访问概率，还应该考虑两个用户间的社交关系、签到相似度等信息对传播概率的影响。DMMs之所以比Psv和Pisv的效果略显逊色，也是因为Psv和Pisv综合了两个用户间的社交关系以及传播者的影响力。

（4）虽然图8(a)和(c)中的影响范围和覆盖率的整体走势相近，但并不是影响范围大，覆盖率就高。例如，在图8(a)中，Pisv、Psv以及Pv的影响范围较接近，但是在图8(c)中的覆盖率却相差很大。Pisv比Psv高出16%左右，而Pv仅有Pisv的50%左右。可见，Pisv对用户间传播概率的度量更加准确，因此其模拟的传播范围较Psv和Pv也更加准确，能够模拟出更多受到影响而真实访问目标位置的用户群体。这说明，仅仅利用影响范围指标并不能够反映真实激活的用户数量，不足以准确评估传播模型的效果。

总之，在基于位置的社交网络中，仅仅考虑网络的拓扑结构度量用户间的传播概率以及仅仅将被影响用户对位置的访问概率作为传播概率都会降低传播模型的效果，同时通过影响范围、准确率和覆盖率三个指标能够更准确地评估传播模型的效果。本文综合考虑了传播者的影响力、用户间的签到相似性以及被传播者对位置的访问概率的传播概率计算方法Pisv能够更准确地度量用户间的信息传播概率，而本文基于Pisv的MFICM在影响范围、准确率和覆盖率三个指标上均取得了最好的效果。

### 6.2.3 InDes-TPLP的选种效果

在本节中，将首先确定InDes-TPLP中的参数k，然后选择合适的β作为选种参数并将InDesTPLP算法和其他选种算法进行对比。

#### 6.2.3.1 InDes-TPLP阶段一选种参数β的确定

参数控制了InDes-TPLP第一阶段选择候选种子的比例，从而影响阶段二选择的种子用户。为了比较不同值对选种结果的影响，本文对在10-400范围内的不同取值进行实验。为了控制算法的时间复杂度，此处假设需要选择10个最终种子用户。实验结果如图9所示。

(a) influence spread with different β (b) accuracy with different β

(c) coverage with different β (d) time complexity with different β

图9 不同β下InDes-TPLP的表现

Fig 9 Results of InDes-TPLP with different β

从图9中可以看出，随着取值逐渐增大，影响范围和覆盖率都呈现先增大后下降的趋势。这是因为，越大，阶段一选取的候选种子用户数量越多，阶段二选择种子用户时需要评估的用户数量越大，选取范围越大，因此可以选择影响范围最大的top-k个用户作为最终种子用户。但是当的取值增大到一定范围后，虽然选取的种子用户的影响范围大，但是由于部分用户在阶段一中的得分不够高，甚至有些用户其实并未真正访问位置，因此通过这些用户评估的影响范围将会降低，从而影响整体的效果。当取300时，覆盖率和准确率都相对较高，影响范围也比较大。因此，考虑到算法的时间复杂度随着的增大而不断上升，在后续实验中，将参数设置为300。

#### 6.2.3.2 InDes-TPLP选种算法效果

在本节中，我们将InDes-TPLP算法和现有算法中表现较好的NewGreedy算法[Chen t al. 2009][19]、基于社区发现的选种算法Novel-IM[He et al. 2015][13]，以及基于两阶段的启发式选种算法TPH[Zhou et al. 2015][3]进行对比。同时，为了评估社区发现以及本文提出的不同传播概率度量方法对InDes-TPLP选种效果的影响，我们在选种时对不同条件进行了控制，设计不同选种算法对比方案如下：

1. Random：随机选种。
2. Novel-IM[He et al. 2015][13]：基于传统社交网络，结合社区结构的影响力最大化算法。同本文方案一样，该算法基于Louvain算法进行社区划分，然后根据节点的中心性在社区中寻找种子用户，最后寻找相互分离的种子节点。
3. NewGreedy[Chen t al. 2009][19]：基于原始贪心算法的改进算法，通过剪枝过滤不可激活的边，然后基于剪枝后的子图计算节点的影响范围。这里将采用传播概率为Pisv的MFICM传播模型度量用户的影响范围，并利用NewGreedy选种方案进行选种。
4. TPH[Zhou et al. 2015][3]：LBSN中基于两阶段传播模型的启发式选种算法。
5. InDes-TPLP-Ncom：不进行第一阶段的社区划分的InDes-TPLP算法。即基于全局网络选择访问兴趣和传播欲望较高的用户作为阶段一的候选种子用户，然后进行阶段二的种子选择。阶段二选种时使用传播概率为Pisv的MFICM传播模型模型度量用户的传播范围。
6. InDes-TPLP-Pv：在InDesTPLP算法的阶段二选种时使用传播概率为Pv的MFICM传播模型度量用户的影响范围。
7. InDes-TPLP-Psv：在InDesTPLP算法的阶段二选种时使用传播概率为Psv的MFICM传播模型度量用户的影响范围。
8. InDes-TPLP-Pisv：在InDesTPLP算法的阶段二选种时使用传播概率为Pisv的MFICM传播模型度量用户的影响范围。

由6.2.2节的实验结果可知，使用Pisv作为传播概率的MFICM模型能够更准确地模拟用户间位置信息的传播过程。所以，在对比选种算法的效果时，本文以基于Pisv的MFICM作为传播模型评估上述8种算法的选种效果。不同算法的选种效果如图10所示：



图 10 不同种子数量下不同选种算法的效果

Fig.10. Results of different seed selection algorithm with different number of seeds

对图10的实验结果分析如下：

(1) 本文提出的InDes-TPLP系列算法是在NewGreedy算法的基础上进一步去除了访问概率和传播欲望较低的用户后，选择访问概率和传播欲望高的用户作为候选种子，因此不仅能够大大降低贪心算法的时间复杂度，同时能够保证选择的种子具有较大的影响范围和覆盖率。特别地，当种子数量较少时，InDes-TPLP-Pisv的覆盖率高于NewGreedy，但是时间复杂度只有NewGreedy算法的50%。当然，由于NewGreedy算法是对全部网络用户进行评估来选种的，因此种子用户的影响范围和覆盖率都高于其他算法。

(2) 从图10（a）、（b）和（c）可以看出，不同选种算法的影响范围、准确率和覆盖率变化并不是完全一致的。相比影响范围和覆盖率，准确率的差异性更大。当种子数量较少时，NewGreedy算法的影响范围较大，但是其覆盖率较低。影响范围越大，说明算法选择的种子的传播范围越广，但是并非实际访问目标位置的用户越多。实验结果中，不同算法的影响范围和覆盖率的变化差异较小，主要是因为所有选种算法都采用了本文提出的准确率较高的MFICM传播模型模拟位置信息的传播过程，因此对于种子用户影响范围的评估也更准确。但是，结合6.2.2节中的实验结果可知，由于现有方案中采用的传播概率度量方法并不合理，对比方案所选种子节点的影响范围和覆盖率的变化将会存在更大的差异。所以，单纯从影响范围的角度评估选种算法的效果是不够的。

(3) 随着种子数量的增加，除Random算法外，其他选种算法的准确率都相对降低，而影响范围和覆盖率的变化幅度也相对较小。这是因为，基于贪心思想的算法每次都会选择具有最大影响范围的种子用户，因此，随着种子数量的增加，所选择的新的种子用户的影响范围将不断降低甚至达到饱和，覆盖率和准确率相比最初的种子也会逐渐降低。总体上看，所有的种子影响范围和覆盖率的变化逐渐平缓，而准确率有所下降。

(4) 对比具有不同传播概率的选种算法发现，本文的InDes-TPLP-Pisv的影响范围比InDes-TPLP-Psv和InDes-TPLP-Pv分别高出10%和30%。可见，考虑更全面的因素有助于提高选种算法的效果。同时，基于社区发现的InDes-TPLP-Pisv算法由于考虑了社区结构，有效降低了种子节点影响范围的重叠，在传播范围和覆盖率上相比不进行社区发现算法的InDes-TPLP-Ncom分别高出40%和30%。因此，考虑社区结构更有利于信息更大范围的传播。尽管InDes-TPLP-Ncom的准确率高于InDes-TPLP-Pisv，但是较低的影响范围使得InDes-TPLP-Ncom相比InDes-TPLP-Pisv的覆盖率也更低。Novel-IM虽然考虑了社区结构，但是由于采用启发式算法选种，且仅仅考虑网络的拓扑结构使得选择的种子范围受到影响。因此当种子数量增加时，考虑局部网络拓扑结构的选种算法只能找到局部最优的种子节点。

# 结论

为了在大规模位置社交网络中更高效、准确地推广位置，本文提出了一种融合用户兴趣和传播欲望的两阶段位置推广算法InDesTPLP。通过结合网络中用户对位置的访问兴趣和传播欲望对网络中的用户提前筛选，有效缩小了种子的选择范围，从而降低了基于贪心思想的选种算法的时间复杂度。同时，本文提出了一种新的传播概率模型MFICM，通过结合用户对目标位置的访问概率、用户间的签到相似性以及传播者的影响力三方面因素度量位置信息的传播概率，有效提高了预测传播概率的准确性，从而提高了传播模型预测用户影响范围的准确性。在yelp数据集上的实验结果表明，本文提出的InDesTPLP算法能够有效降低NewGreedy算法的时间复杂度，同时保证较大的影响范围和覆盖率。

但是，InDesTPLP算法的效果依赖于阶段一选择种子的准确性，即选择的种子对目标位置应有较高的访问概率，同时用户间的位置信息传播概率也会受用户访问概率的影响。因此，提高访问概率预测算法的准确性能够有效提高最终种子影响范围的准确性。除此以外，虽然基于贪心思想选种能够得到较大的影响范围，但是算法时间复杂度仍然比较高。因此，如何结合现有的基于树形结构的启发式算法以及社区发现结果，在社区子网络中并行选种也将是今后的研究方向。

# 参考文献

1. Zhu W Y, Peng W C, Chen L J. Exploiting mobility for location promotion in location-based social networks[C]//Data Science and Advanced Analytics (DSAA), 2014 International Conference on. IEEE, 2014: 76-82.
2. Li G, Chen S, Feng J, et al. Efficient location-aware influence maximization[C]//Proceedings of the 2014 ACM SIGMOD International Conference on Management of Data. ACM, 2014: 87-98.
3. Zhou T, Cao J, Liu B, et al. Location-based influence maximization in social networks[C]//Proceedings of the 24th ACM International Conference on Information and Knowledge Management. ACM, 2015: 1211-1220.
4. Wang A, Zhang A, Zhang A, et al. Distance-aware influence maximization in geo-social network[C]// IEEE, International Conference on Data Engineering. IEEE, 2016: 1-12.
5. Hai N T. A novel approach for location promotion on location-based social networks[C]//Computing & Communication Technologies-Research, Innovation, and Vision for the Future (RIVF), 2015 IEEE RIVF International Conference on. IEEE, 2015: 53-58.
6. Zhu W Y, Peng W C, Chen L J, et al. Modeling user mobility for location promotion in location-based social networks[C]//Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. ACM, 2015: 1573-1582.
7. Li J, Sellis T, Culpepper J S, et al. Geo-social Influence Spanning Maximization[J]. IEEE Transactions on Knowledge & Data Engineering, 2017, PP(99):1653-1666.
8. Li G, Wen Z Y, Zhu W Y. Promoting a bundle of locations via viral marketing in location-based social networks[C]//Technologies and Applications of Artificial Intelligence. IEEE, 2017: 32-39.
9. Wang X, Zhang Y, Zhang W, et al. Efficient Distance-Aware Influence Maximization in Geo-social Networks[J]. IEEE Transactions on Knowledge & Data Engineering, 2017, 29(3):599-612.
10. Kempe D, Kleinberg J M, Tardos É. Maximizing the spread of influence through a social network[J]. Theory of Computing, 2015, 11(4): 105-147.
11. Zhu W Y, Peng W C, Chen L J, et al. Exploiting viral marketing for location promotion in Location-Based Social Networks[J]. ACM Transactions on Knowledge Discovery from Data (TKDD), 2016, 11(2): 25.
12. Yu F, Li Z, Jiang S, et al. Point-of-Interest Recommendation for Location Promotion in Location-Based Social Networks[C]// IEEE International Conference on Mobile Data Management. IEEE, 2017.
13. He J L, Fu Y, Chen D B. A novel top-k strategy for influence maximization in complex networks with community structure[J]. Plos One, 2015, 10(12): e0145283.
14. Zhao Y, Li S, Jin F. Identification of influential nodes in social networks with community structure based on label propagation[J]. Neurocomputing, 2016, 210: 34-44.
15. Gong M, Song C, Duan C, et al. An Efficient Memetic Algorithm for Influence Maximization in Social Networks[J]. IEEE Computational Intelligence Magazine, 2016, 11(3): 22-33.
16. Domingos P, Richardson M. Mining the network value of customers[C]//Proceedings of the seventh ACM SIGKDD international conference on Knowledge discovery and data mining. ACM, 2001: 57-66.
17. Kempe D, Kleinberg J. Maximizing the spread of influence through a social network[C]// ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. ACM, 2003:137-146.
18. Leskovec J, Krause A, Guestrin C, et al. Cost-effective outbreak detection in networks[C]//Proceedings of the 13th ACM SIGKDD international conference on Knowledge discovery and data mining. ACM, 2007: 420-429.
19. Chen W, Wang Y, Yang S. Efficient influence maximization in social networks[C]//Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining. ACM, 2009: 199-208.
20. Goyal A, Lu W, Lakshmanan L V S. Celf++: optimizing the greedy algorithm for influence maximization in social networks[C]//Proceedings of the 20th international conference companion on World wide web. ACM, 2011: 47-48.
21. Chen W, Wang C, Wang Y. Scalable influence maximization for prevalent viral marketing in large-scale social networks[C]// ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. ACM, 2010:1029-1038.
22. Barbieri N, Bonchi F, Manco G. Topic-aware social influence propagation models[J]. Knowledge & Information Systems, 2013, 37(3):555-584.
23. Fortunato S, Hric D. Community detection in networks: A user guide[J]. Physics Reports, 2016, 659:1-44.
24. Blondel V D, Guillaume J L, Lambiotte R, et al. Fast unfolding of communities in large networks[J]. Journal of statistical mechanics: theory and experiment, 2008, 2008(10): P10008.
25. Liu S D, Meng X W. Recommender systems in location-based social networks[J]. Chinese Journal of Computers, 2015, 38.
26. Ye M, Yin P, Lee W C, et al. Exploiting geographical influence for collaborative point-of-interest recommendation[C]// ACM, 2011:325-334.
27. Hsieh H P, Li C T, Lin S D. Estimating Potential Customers Anywhere and Anytime Based on Location-Based Social Networks[C]//Joint European Conference on Machine Learning and Knowledge Discovery in Databases. Springer, Cham, 2015: 576-592.
28. Zou Z, Xie X, Sha C. Mining User Behavior and Similarity in Location-Based Social Networks[C]// Seventh International Symposium on Parallel Architectures, Algorithms and Programming. 2015:167-171.
29. Sacharidis D, Sacharidis D, Bikakis N. Regionally influential users in location-aware social networks[C]// ACM Sigspatial International Conference on Advances in Geographic Information Systems. ACM, 2014:501-504.