



运营技术广角

基于社交网络分析的流量红包客户挖掘与传播模式

徐海勇, 陶涛, 黄岩, 唐崔巍, 张兆静, 吴晶
(中国移动信息技术有限公司, 北京 100037)

摘要: 电信运营商持续创新流量经营模式与手段, 将流量营销与社交化的红包活动相结合, 创新推出流量红包活动, 激发客户流量使用兴趣。面对社交化流量红包客户特征与传播模式的研究痛点及当前社群建模算法较单一的技术现状, 详细研究、对比 6 种社群建模算法的应用效果, 筛选出适合流量红包的最优算法, 并定位分析核心价值客户群的特征。仿真结果显示, Multi-Level 算法在流量红包场景中表现更好, 基于该算法挖掘种子客户、高价值客户、低价值客户和沉默客户 4 种特征客户群的社交网络结构。社交网络分群结论为运营商精准营销、精准推送等营销活动以及沉默客户促活、流失客户挽回等客户运营管理提供了有效指导。

关键词: 社交网络分析; 流量红包; 客户群划分

中图分类号: TP301

文献标识码: A

doi: 10.11959/j.issn.1000-0801.2020128

Data red envelope clients mining and communication model based on social network analysis

XU Haiyong, TAO Tao, HUANG Yan, TANG Cuiwei, ZHANG Zhaojing, WU Jing
China Mobile Information Technology Co., Ltd., Beijing 100037, China

Abstract: Telecom operators innovate continually in data management modes and combine data marketing with social red envelope activities, consequently launching data red envelope activities, which stimulates customers' interest of data using. Faced with the research pain points of customer characteristics and propagation modes of socialized traffic red envelopes and the current status of the single community modeling algorithm, the detailed research and comparison of the application effects of the six community modeling algorithms were conducted to select the optimal traffic envelope algorithm, and the characteristics of the core value customer groups were located and analyzed. Results show that Multi-Level algorithm performs better in data red-envelope scenario. Based on this algorithm, four characteristic customer groups were mined, namely seed customer, high-value customer, low-value customer and silent customer. The social network clustering conclusion provided effective guidance for operators' precise marketing, precise push and other marketing activities, as well as other customer operation management such as silent customer activation, lost customer recovery and so on.

Key words: social network analysis, data red envelope, customer group division



1 引言

随着 5G 时代的到来,电信运营商持续加强数字化转型,积极拓展传统业务的新玩法,尤其在流量运营方面。电信运营商持续创新流量经营模式与手段,紧抓互联网趋势,不断推出视频、音乐和直播等定向流量包,拓展流量红包等社交化新玩法。流量红包是基于流量账户体系,实现个人客户的流量购买、红包分发和提取,借助互联网社交化的病毒式传播,刺激个人客户的流量增长。在流量红包传播中,发红包客户和领红包客户之间构成了复杂的社交网络关系,研究发现,在社交网络关系中存在社群传播的关键人物或群体,对这部分客户的精准画像与分析有助于以低成本、高效率的方式达到最好的营销推广效果。在海量数据中,如何挖掘客群特征来更好地指导客户精准运营显得尤为重要^[1],本文将基于流量红包分发场景,利用社交网络分析算法对客户进行精准分群,深入挖掘流量红包客群特征与传播模式。

2 技术现状分析

当前社交网络涉及的理论研究主要包括以下两个方面:社交网络个体影响力研究,通过识别社群中具有影响力的核心节点,分析核心节点的特征和影响力传播机制;社交网络群体划分研究,从庞大的社交网络中划分社交子群,分析不同社群的特征差异性,按群体结构可分为重叠社群和非重叠社群,“重叠”指节点同时归属于两个或两个以上的群体。

2.1 基于社交网络个体影响力的研究

吴信东等^[2]从网络拓扑、客户行为和交互信息等方面总结了影响力建模方法,介绍了影响力传播问题的研究现状,并对在线社交网络影响力分析前景进行了展望。毛佳昕等^[3]基于新浪微博数据提出一个通过预测客户传播信息能力大小分

析和度量客户社会影响力的方法。全拥等^[4]针对一种经典的个体影响力计算方法,介绍了该算法的两种并行化实现,并在真实的、大规模在线社交网络数据集上进行了性能测试。王祯骏等^[5]提出一种融合内容信息和社交网络动态时间特性的潜在影响力传播模型,能够准确地建模潜在的影响力传播过程,有效地挖掘出社交网络中的有影响力客户。

2.2 基于社交网络群体划分的研究

Newman^[6]基于非重叠社群提出一种快速 Newman 算法,具有较好的速度优势,但针对重叠社群的社交网络的实用性不强。Raghavan 等^[7]提出一种基于标签传播的重叠社区挖掘算法,该算法仅以网络结构为指导,不需要对预先定义的目标函数进行优化,特点是计算快速和可在短时间内处理大而密集的网络。王庚等^[8]从多方面研究目前基于标签传播的社区挖掘算法的优劣,并详细分析基于标签传播算法在以后研究中的改进思路。

由于当前国内文献大多侧重于社交网络个体的影响力研究,针对社交群体划分建模分析较少,社群建模算法比较单一,缺乏实际生产的案例验证。同时,流量红包在社交网络流转中具有极强的不确定性,这种不确定性将直接影响客户社交关系的强弱,因此需要综合考虑社交网络算法的精度、效率及与模型的配合程度来筛选合适的特征提取方法。本文将针对流量红包客户传播场景,从社交网络社群划分算法入手,结合特征客户的交互行为,从而挖掘出更具体的客户社交网络特征。

3 基于社交网络的流量红包客户建模方案

基于运营商流量红包客户精准分群和精细运营的重要意义以及当前社交网络理论的研究现状,研究、比较了 6 种社群模型算法,选择最优算法进行社交分群,挖掘具有社交关系的客户群体,进而对客户行为特征进行深入分析,观测流量红包在不同客户群体中的传播模式。

3.1 系统体系架构

分析流程包括数据收集、数据处理、模型建立和社交网络可视化,形成的系统体系架构如图1所示。

(1) 数据层

流量红包数据来源于某运营商流量中心系统,通过采集领取用户号码、发送用户号码、年龄、网龄、近3个月平均消费金额、当月套餐内流量、当月DOU(流量使用情况)、红包状态等12个字段,形成社交网络分析宽表。将本网客户和异网客户分别录入,进行区分。

(2) 处理层

由于流量红包客户的异常领取行为可能会影响社交网络的分析结论,因此需要对原始数据进行清洗,清洗剔除重复领取红包形成的冗余数据和领取中未采集到手机号的无效数据,筛选出正常发送和被领取中的数据,并进行用户特征分析和传统聚类分析。

(3) 模型层

步骤1 利用数据层中形成的流量红包数据宽表构建社交网络,将领取人号码(DRAW_

NUMBER)和发红包人号码(SEND_NUMBER)具象为社交网络中的节点,当一位客户领取了另一位客户的流量红包,则构建出节点的连接边。

步骤2 遍历社交网络中的所有节点和边,计算节点度、点度中心度、接近中心度、中间中心度、密度、PageRank值等指标,识别出影响力大的客户。

步骤3 分别基于Fast-Greedy算法、Leading Eigenvector算法、Walktrap算法、Label Propagation算法、InfoMap算法和Multi-Level算法,对社交网络进行社交分群,针对不同群体特征客户的社交结构和流量红包传播模式进行深入分析。

(4) 应用层

将客户节点的分布情况、可视化社群算法的网络社交分群图和特征客户社交关系结构进行可视化展示。

3.2 社交网络理论主要内容

社交网络分析(social network analysis, SNA)是探究社交网络的一门数据挖掘方法,研究社交网络的群体现象,分析节点在社群中和社群之间的作用,在网站信息传播、网络反欺诈、社交媒

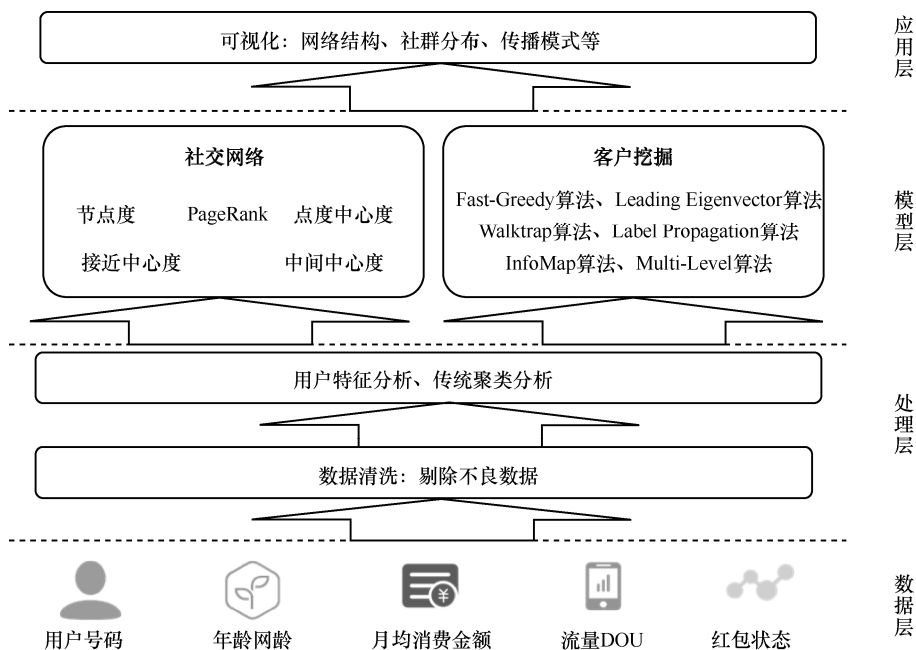


图1 系统体系架构



体影响力等领域中得到了广泛应用,涉及的主要概念见表1。

社交网络分群的目的是将一个完整网络划分为多个节点密集的社群,社群内部节点的连接比较紧密,社群之间的连接相对较稀疏。社群划分的精细程度通常由模块化指标 Q 值来衡量, Q 是一个介于-1和1之间的标量值,主要用于比较不同社交分群方法的质量,同时也是优化分群的目标函数,计算式如下:

$$Q = \frac{1}{2m} \sum_{ij} \left[A_{ij} - \frac{k_i k_j}{2m} \right] \delta(c_i - c_j) \quad (1)$$

其中, A_{ij} 表示节点 i 和 j 之间边的权重, $k_i = \sum_j A_{ij}$ 是连接节点 i 的边的权重之和。 c_i 是节点 i 所在的社群; $m = \frac{1}{2} \sum_{ij} A_{ij}$; 当 $\mu = \sigma$ 时, 函数 $\delta(\mu, \sigma) = 1$, 反之等于0。

3.3 社群划分模型算法原理

当前国内社群建模算法比较单一,针对算法的实用性评估较少,本文研究、比较了社交网络分群的6种算法,其原理如下。

(1) Fast-Greedy 算法

Fast-Greedy 算法^[9]是一种自下而上的分层方法。算法假设原始状态下每个节点属于一个单独

的社群,然后不断迭代、合并社群,使得每次合并达到局部最优。当模块化指标 Q 值不能再增加时,算法自动停止,输出分群和树形图。该算法运行速度快,不需要调整参数,通常作为第一个社群划分尝试的方法。

(2) Leading Eigenvector 算法

Leading Eigenvector 算法^[10]是一种自上而下的分层方法,将原始社群划分为两个群体,使模块化指标 Q 值显著增加。划分通过模块化矩阵的特征向量来评估,并设定一个停止机制,使紧密连接的群体终止划分。

(3) Walktrap 算法

Walktrap 算法^[11]是一种基于随机游走的方法。假设节点在社交网络中随机游走,那么其步行可能保持在同一社群内。与 Fast-Greedy 算法类似,Walktrap 随机游走的步长较短,并以自下而上的方式使随机游走结果合并成单独的社群,并依据模块化指标 Q 值从不同群体划分方式中选择最佳的分群方式。

(4) Label Propagation 算法

Label Propagation 算法^[7]首先给每个节点都分配一个 k 标签,然后迭代重新分配标签,每个节点将获取其邻近点出现次数最多的标签,直到

表1 社交网络分析中涉及的概念解释

分类	主要内容	概念解释
基本概念	节点	研究对象个体,节点集是有限非空集合
	边	对象个体之间的关系,边集是有限集合
	度	从节点出发或指向该节点的边数,度越大,节点连接边数越多
中心性	点度中心度 (degree)	用来计算节点的度中心性,值越大表示节点越在中心,能到达更多其他点
	接近中心度 (closeness)	用来计算节点的紧密中心性,值越大表示节点间距离越短,影响速度越快
	中间中心度 (betweenness)	用来计算节点的中心性之间的最短路径,值越大表示经过该节点时包含信息越多
其他	密度 (density)	用来计算图的密度,空图的密度为0,完整图的密度为1
	广度优先搜索和深度优先搜索	用来搜索图中是否可以从给定节点到达某个节点,也被称为图遍历
	PageRank 算法	用来计算图中节点的排名,值越高说明该节点的影响力越大、越受欢迎

具有相同标签的节点组成社群，则迭代进程停止。该算法运行速度较快，不同的初始值会产生不同的分群结果，可多次运行，直到达到满意效果。

(5) InfoMap 算法

InfoMap 算法^[12]是一种基于信息论的社群算法，认为社交网络节点的连接会产生跨网络移动以及导致系统范围内的相互依赖，与基于随机模型的模块化算法不同，InfoMap 算法使用谱分析和映射方程，在处理网络动态行为、网络之间相互依赖性以及网络结构与系统行为相关性等方面更有优势。

(6) Multi-Level 算法

Multi-Level 算法^[13]是一种基于层次聚类算法，包括重复和迭代两个阶段，首先给节点分配不同社群，计算将节点 i 从社群中移除并将其放在 j 社群中的模块化 Q 值，对所有节点重复执行直到 Q 值达到局部最大值；再以初步划分的社群作为新节点建立新网络，迭代直到达到模块最大化。

4 社交网络构建与社群算法实现

本文采用 2019 年 5 月份某运营商流量红包数据作为研究对象，涉及 814.4 万条数据记录，关键字段见表 2，对数据进行清洗后，首先进行用户特征分析和非社交网络的聚类分析。

4.1 用户特征分析与聚类分析

用户特征分析结果如图 2~图 5 所示。由图 2~图 5 的用户特征分析可知，发送用户中 62% 用户套餐内流量少于 1 GB，领取用户中 64% 用户套餐内流量低于 1 GB。从使用流量情况看，高 DOU 用户占比较大，其中，发送用户中每月 DOU 在 1 GB 以上用户占比 89%，领取用户中每月 DOU 在 1 GB 以上用户占比 86%。反映出用户使用流量较多，而套餐内流量较少，因此对套餐外流量（即流量红包）需求较大。

本文采用 K -means++ 算法分别对发送用户和领取用户进行聚类分析。发送用户雷达图如图 6 所示，

表 2 流量红包数据关键字段

字段名称	字段解释	字段类型
AGE	用户年龄	数值型
NET_AGE	用户网龄	数值型
DATA	参与当月套餐内流量	数值型
LAST_DOU	参与流量红包的发送和领取的上月 DOU	数值型
NOW_DOU	参与流量红包的发送和领取的当月 DOU	数值型
AVER_CHARGE	近 3 个月平均消费金额	数值型
DRAW_NUMBER	领取用户号码	字符串型
SEND_NUMBER	发送红包号码	字符串型
DRAW_TIME	领取时间	数值型
UNDER_SIZE	子红包大小/MB	数值型
TYPE	红包类型	字符串型
CONDITION	红包状态	字符串型

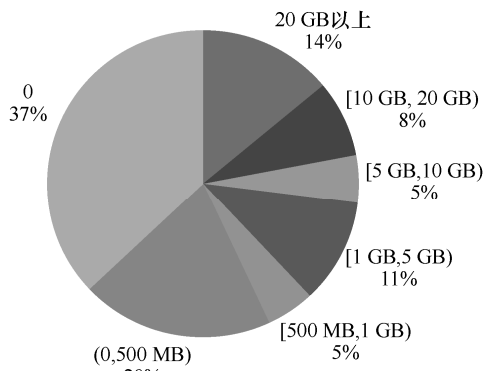


图 2 发送用户参与当月套餐内流量

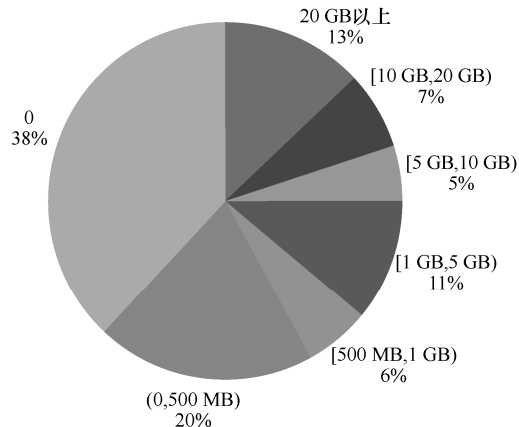


图 3 领取用户参与当月套餐内流量

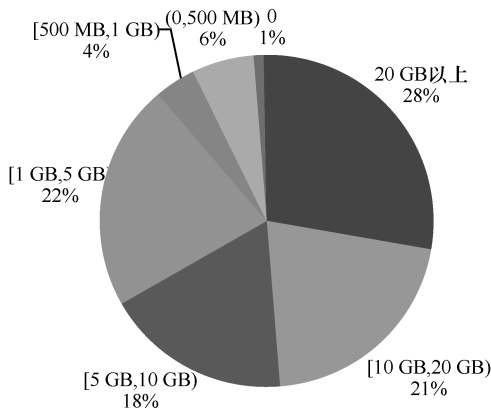


图4 发送用户参与当月DOU

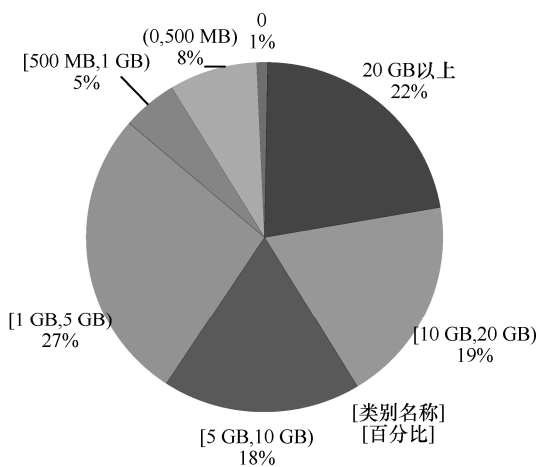


图5 领取用户参与当月DOU

较大。领取用户雷达图如图7所示，与发送用户分群结果类似，图7中领取用户的A类占比82.3%，B类占比17.7%，领取用户A类群体占比比发送用户A类更高，所以领取用户对套餐外流量需求更大。从图6、图7中也可以看出，发送用户和领取用户A、B的用户分群特征较为相似，群数量较少，单个群内用户数较多，划分群体不够细致。因此，通过传统的非社交网络分析方法不能有效地划分用户群体，且无法跟踪流量红包在用户之间的传播路径。

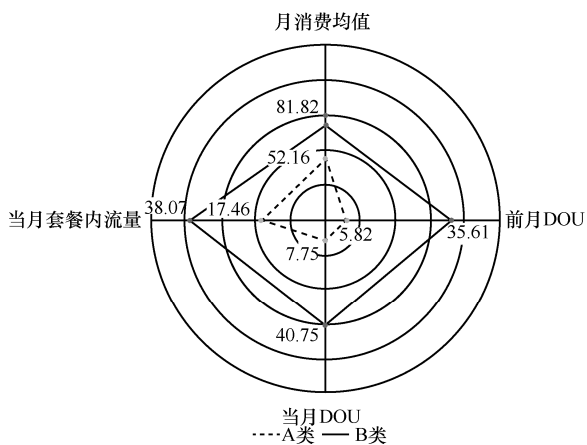


图7 领取用户雷达图

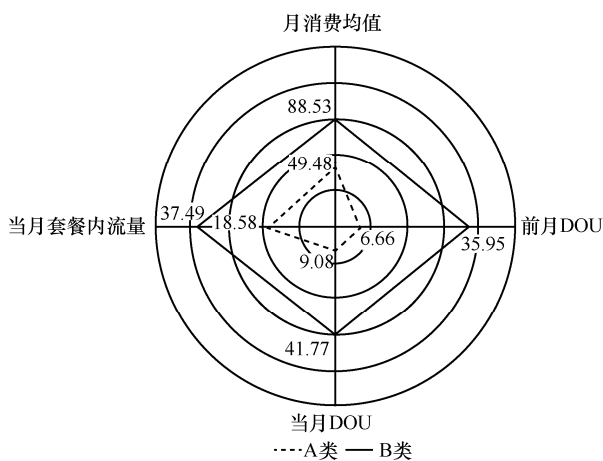


图6 发送用户雷达图

由图6可知，发送用户被分作A、B两类，其中A类占比78.9%，B类占比21.1%，A类用户的月消费均值、套餐内流量、DOU指标均低于B类用户，反映出发送用户群体中低消费、低流量人群占比

4.2 基于社交网络的分群算法

基于社交网络的建模方案，构建出流量红包客户的社交网络，共包含91.46万个节点和81.97万条边，分别计算91.46万个节点的度数，其中最小的度数为1，最大为173，平均度数为1.79，平均每个客户发送或者领取1.79个流量红包，换言之，平均每个客户有1.79个好友。流量红包社交网络节点度数分布散点图如图8所示，可知节点度数具有显著的长尾效应，度数较小的节点占比较大，度数较大的节点占比较小，这类数量较少的客户具有较大的好友圈，可以作为重点营销对象。

通过比较6种算法在不同样本量时的社交化质量，选择出适合流量红包客户的分群算法，从总体研究对象中选取30万和5万个节点作为大样本和小样本进行社交分群，对比分群数、模块化

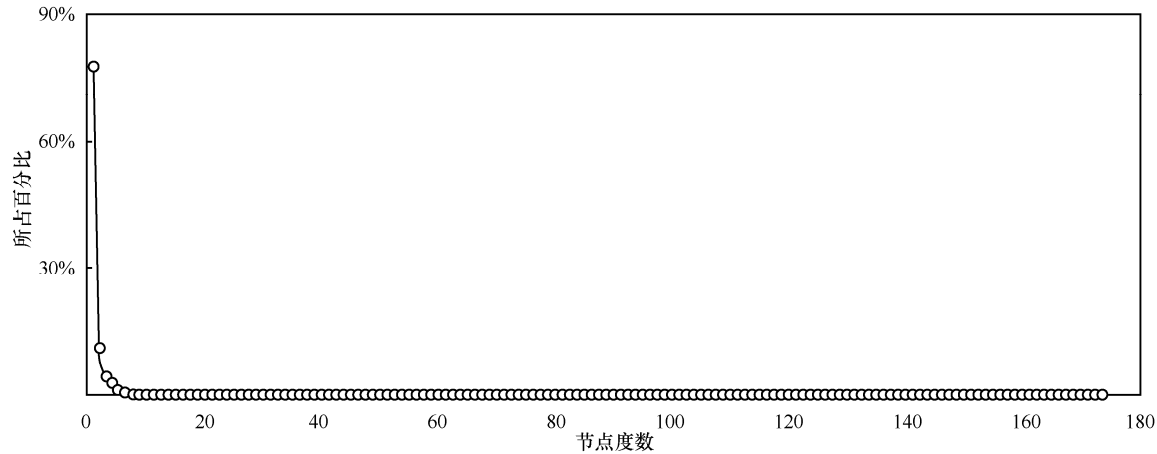


图 8 流量红包社交网络节点度数分布散点图

指标 Q 值和算法运行时间, 结果见表 3 和表 4。

表 3 6 种社群划分算法大样本社交化质量(30 万个节点)

方法	分群数/个	模块化指标 Q	运行时间/s
Fast-Greedy 算法	45 306	0.97	106.39
Leading Eigenvector 算法	45 219	0.90	5.08
Walktrap 算法	45 444	0.97	2 841.08
Label Propagation 算法	47 820	0.92	2.56
InfoMap 算法	46 933	0.93	478.99
Multi-Level 算法	45 281	0.97	0.89

表 4 6 种社群划分算法小样本社交化质量 (5 万个节点)

方法	分群数/个	模块化指标 Q	运行时间/s
Fast-Greedy 算法	1 740	0.992	0.03
Leading Eigenvector 算法	1 740	0.991	0.08
Walktrap 算法	1 741	0.991	0.10
Label Propagation 算法	1 795	0.965	0.03
InfoMap 算法	1 772	0.977	0.74
Multi-Level 算法	1 740	0.991	0.00

通过运行结果来看,6 种算法的分群数比较相近,在大样本中, Fast-Greedy 算法、Walktrap 算法和 Multi-Level 算法模块化指标 Q 值为 0.97, 表现较好, 反映出这 3 种算法划分的社群内部节点的连接比较紧密, 社群之间的连接相对较稀疏; 从运行时间看, Fast-Greedy 算法和 Walktrap 算法花费时间较长, Multi-Level 计算速度较快。在小样本中, Fast-Greedy 算法、Leading Eigenvector 算法、Walktrap 算法和 Multi-Level 算法模块化指标 Q 值均在 0.991 以上, 运行时间在 0.10 s 及以内, Label Propagation 算法和 InfoMap 算法模块化 Q 值相对较低。由此可知, 在小样本中, Fast-Greedy 算法的群体划分准确性相对更好, 而在大样本中, Multi-Level 算法的模块化 Q 值和计算时间表现更好。针对这两种社群算法, 进一步选取 6 月和 7 月的流量红包数据进行验证比较, 两种社群划分算法 6 月和 7 月数据验证比较见表 5, 由表 5 可知, 7 月的 Multi-Level 算法模块化指标 Q 值更高, 计算迭代时间更短, 且更适用于实际

表 5 两种社群划分算法 6 月和 7 月数据验证比较

方法	6 月			7 月		
	分群数/个	模块化指标 Q	运行时间/s	分群数/个	模块化指标 Q	运行时间/s
Fast-Greedy 算法	45 054	0.95	172.48	41 610	0.90	172.83
Multi-Level 算法	44 987	0.95	2.17	41 419	0.91	3.48



生产的大规模社交网络，因此这里选择 Multi-Level 算法对流量红包客户进行社交分群。

4.3 社群划分与客户挖掘

为直观展示客户社交网络的分布情况，选取 5 月份大样本的 30 万个节点，筛选出节点度数大于 10 的节点，并剔除孤立点，得到社交网络如图 9 所示。可以看出流量红包用户具有“小世界特性”，是一个典型的社会网络。图 9 中编号相同的节点为相互有发送和接收红包行为的客户，节点大小反映客户的点度中心度，节点越大，客户影响力越大。

采用 Multi-Level 算法进行社群划分，结果如图 10 所示。网络节点被划分为 36 个社群，其中

超过 10 个节点的大社群有 12 个。社交分群数量分布如图 11 所示。不同社群之间有重叠的社交节点，其中 A 类、B 类、C 类覆盖面较大的社交群为核心节点客户群，在流量红包营销活动中参与度较高，需重视这类核心客户的维系和留存。A 类和 B 类社群之间的 1 号和 2 号节点中间中心度较高（见表 6），反映这些客户在两个社群之间起到重要的连接作用，在流量红包传播的社交网络中作用较大。

4.4 特征客户与红包传播模式

社交网络社群划分后，对子群中的特征客户进行拆分，结合客户在实际中的发红包和领红包

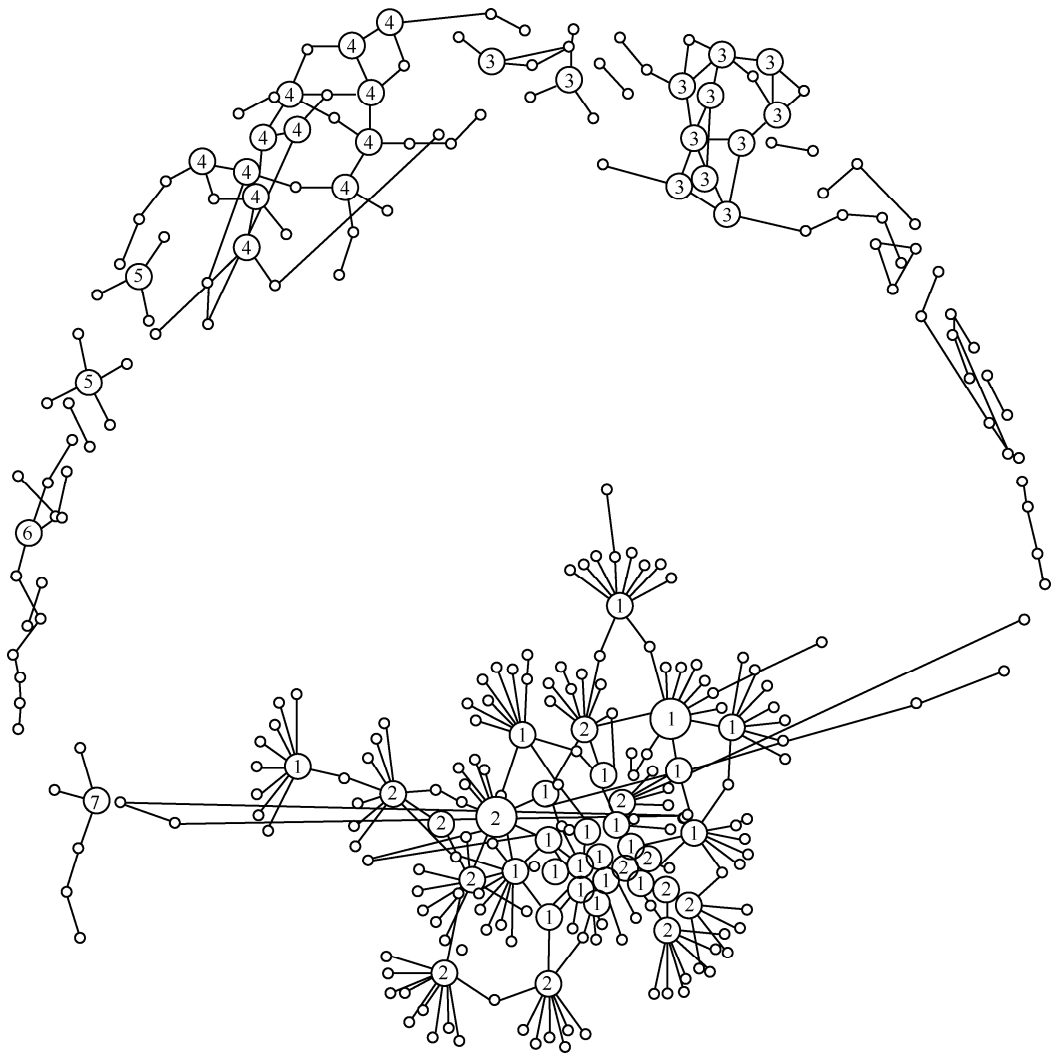


图 9 流量红包样本客户社交网络

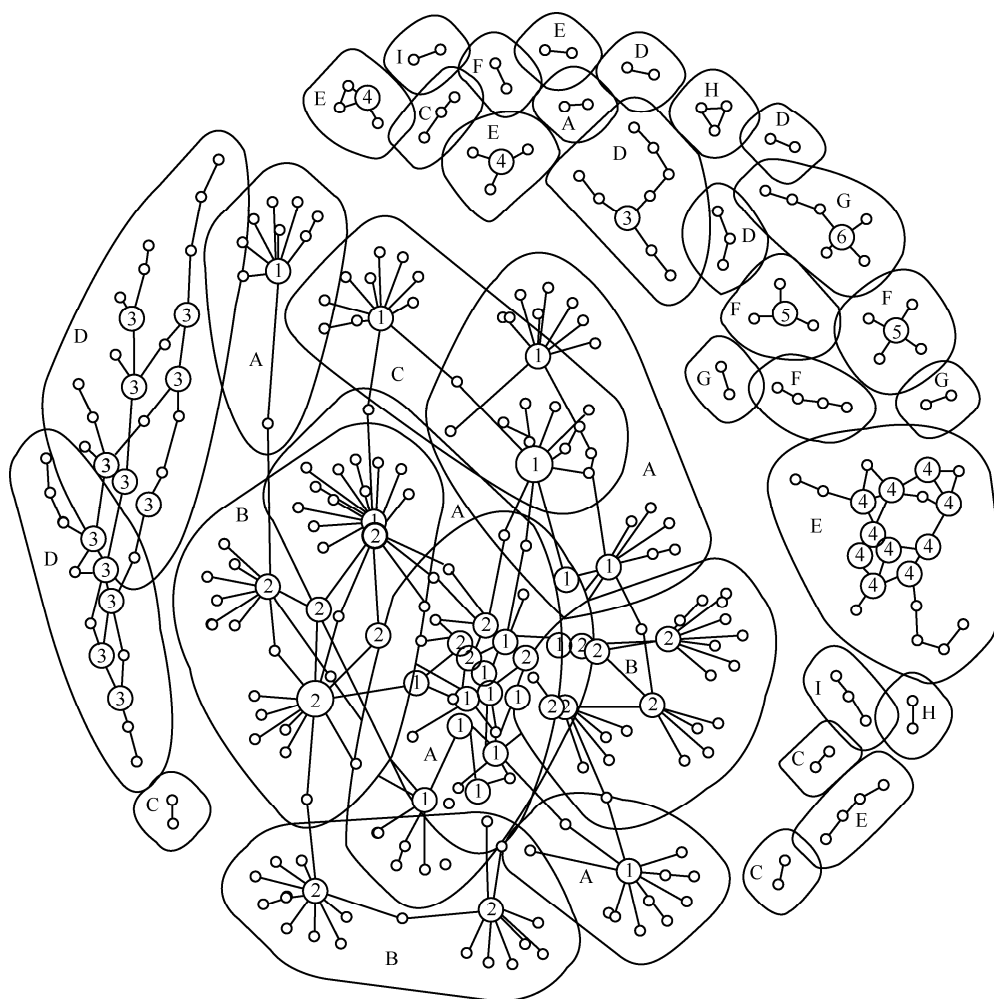


图 10 Multi-Level 算法社交分群结果

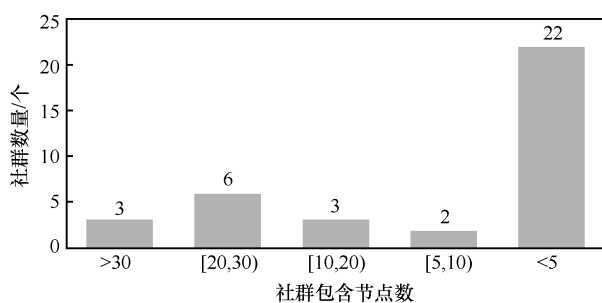


图 11 社交分群数量分布

行为,分析流量红包在不同特征客户中的传播模式,如下几个代表性的社交网络结构中,深色边表示发送母红包号码相同,浅色边表示领取子红包来自同一个母红包。

(1) 种子客户

种子客户如图 12 所示,11 位客户组成一个社

表 6 社交网络分群中的核心节点客户

节点	度	中间中心度	所属社群
138****0889	12	1 953.498	A 类
134****0893	12	4 056.851	A 类
134****2859	11	2 024.253	A 类
137****2554	11	1 654.052	B 类
183****6317	11	1 479.824	A 类
187****6950	11	1 411.204	B 类
158****8081	11	2 023.352	B 类
152****7792	11	1 548.448	B 类
188****8755	11	1 567.933	A 类
135****8460	11	2 929.701	A 类



交网络结构, 编号 467、3770、3773、3767 客户由深色边相互连接, 表明这 4 位客户的号码相同, 即同一客户在不同时段有多次发红包行为。同理, 编号为 489 和 493 的客户为同一用户, 编号为 503 和 1410 的客户也为同一客户, 为主要发红包客户。除了深色边, 这些客户之间也由浅色边连接, 反映发红包客户也参与了领取红包, 既有发送行为, 又有领取行为, 该类特征客户社交网络较庞大, 属于流量红包的忠实客户群, 可在活动中发展为“种子客户”, 进而拉动其他客户参与。

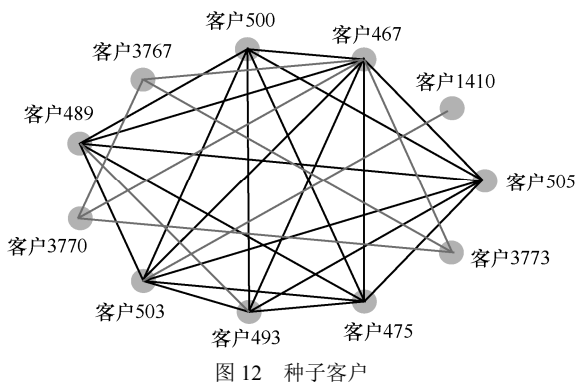


图 12 种子客户

(2) 低价值客户

低价值客户如图 13 所示, 10 位客户全部由浅色边连接, 表明领取了同一个母红包, 通过同一客户的母红包形成一个社交网络, 存在共同的好友社交关系, 每个客户的度为 9。这些客户只有领取红

包行为, 无发送红包行为, 是领流量红包的偏好者, 对流量营销活动参与积极性较高, 且形成一定规模的社交圈子, 属于“薅羊毛”类型客户, 可针对性引导这类客户参与更多相似的流量类营销活动。

(3) 高价值客户

高价值客户如图 14 所示, 4 位客户之间为深色边, 表示同一个发红包客户。根据深度优先搜索 (depth-first-search, DFS) 算法得到边的顺序为 (4071, 4058)、(4058, 4062)、(4062, 4046)、(4071, 4068), 该客户存在多次发送红包的行为, 而仅有一次领取红包行为, 为发红包偏好者, 其对流量红包认可程度高、消费能力强、社交圈子大, 属于高价值客户, 往往具有“引导者”“传播者”的作用, 在营销活动中可以发挥重要的传播价值。

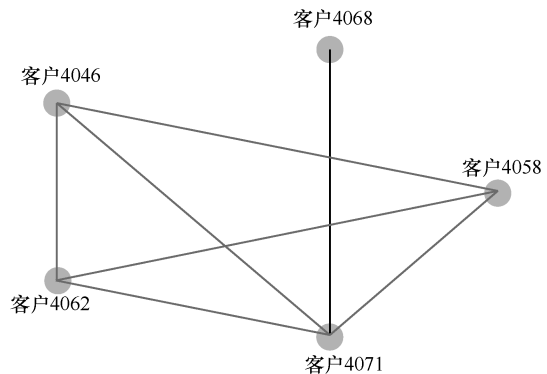


图 14 高价值客户

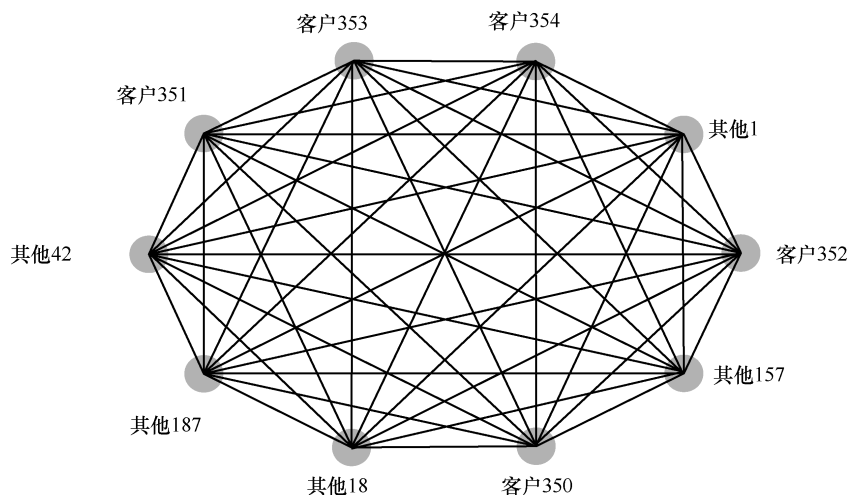


图 13 低价值客户

(4) 沉默客户

沉默客户如图15所示,客户4218和客户4219仅领取了客户184****3602的红包,没有领取其他红包,也没有发送红包,领取红包为偶然事件,属于流量红包不活跃客户。这类特征客户对流量红包流程不熟悉,认可度较低,在营销中应采取主动引导的措施,例如发送邀请体验的种子红包、推送消息促进活跃等。



图15 沉默客户

(5) 异网客户

针对样本数据中279个领取红包的异网客户同步进行了特征分析,发现领取红包的异网客户社交网络包含70个节点和50条边,重合节点为多位异网客户领取了同一位本网客户的流量红包。其中,约70位异网客户通过本网客户发送的红包构建了社交关系,占比25.1%,异网客户社交关系的平均度为1.43。应重视异网客户与本网客户的社交关系,借助异网客户圈较大的本网客户,拉动异网客户的携号转入。

基于上述客户群特征,在制定流量红包营销及推广策略时,应重视核心节点客户在营销活动中的推广作用,充分利用其私域流量优势,通过加大对核心节点客户或客户群体打折、免费赠送、返现等优惠力度或分享计酬等方式,利用其消费示范效应及影响力,拉动非核心节点客户的消费能力进而扩散到整个客户群。通过采取多种组合策略促进核心节点客户所在客户群乃至整个社交网络的活动宣传及客户购买。

5 结束语

本文针对运营商流量红包场景,对比了6种社群算法的分析效果,筛选出最优算法,定位挖掘流量红包核心价值客户群特征。在数据量倍增的5G时代,精确分析客户的访问行为、消费行为、

社交属性和发展周期,是运营商实现数字化转型的重要一步。未来,基于社交网络知识图谱和人工智能算法的大数据分析将更广泛应用于运营商多场景融合的营销服务体系中^[14]。

参考文献:

- [1] 谷红勋, 杨珂. 基于大数据的移动用户行为分析系统与应用案例[J]. 电信科学, 2016, 32(3): 139-146.
GU H X, YANG K. Mobile user behavior analysis system and applications based on big data[J]. Telecommunications Science, 2016, 32(3): 139-146.
- [2] 吴信东, 李毅, 李磊. 在线社交网络影响力分析[J]. 计算机学报, 2014(4): 735-752.
WU X D, LI Y, LI L. Influence analysis of online social networks[J]. Chinese Journal of Computers, 2014(4): 735-752.
- [3] 毛佳昕, 刘奕群, 张敏, 等. 基于客户行为的微博客户社会影响力分析[J]. 计算机学报, 2014(4): 791-800.
MAO J X, LIU Y Q, ZHANG M, et al. Social influence analysis of mirco-blog user based on user behavior [J]. Chinese Journal of Computers, 2014(4): 791-800.
- [4] 全拥, 贾焰, 张良, 等. 在线社交网络个体影响力算法测试与性能评估[J]. 通信学报, 2018, 39(10): 1-10.
QUAN Y, JIA Y, ZHANG L, et al. Performance analysis and testing of personal influence algorithm in online social networks [J]. Journal on Communications, 2018, 39(10): 1-10.
- [5] 王祯骏, 王树徽, 张维刚, 等. 基于社交内容的潜在影响力传播模型[J]. 计算机学报, 2016, 39(8): 1528-1540.
WANG Z J, WANG S H, ZHANG W G, et al. Social content based latent influence propagation model [J]. Chinese Journal of Computers, 2016, 39(8): 1528-1540.
- [6] NEWMAN M E J. Fast algorithm for detecting community structure in networks[J]. Physical Review E, 2004, 69(6): 066133.
- [7] RAGHAVAN U N, RÉKA A, KUMARA S. Near linear time algorithm to detect community structures in large-scale networks[J]. Physical Review E, 2007, 76(3): 036106.
- [8] 王庚, 宋传超, 盛玉晓, 等. 基于标签传播的社区挖掘算法研究综述[J]. 计算机技术与发展, 2013(12): 69-73.
WANG G, SONG C C, SHENG Y X, et al. Research summary on communities mining algorithm based on label propagation[J]. Computer Technology and Development, 2013(12): 69-73.
- [9] CLAUSET A, NEWMAN M E J, MOORE C. Finding community structure in very large networks[J]. Physical Review E, 2004, 70(6): 1539-3755.
- [10] NEWMAN M E J. Finding community structure in networks using the eigenvectors of matrices[J]. Physical Review E, 2006, 74(3): 036104.
- [11] PONS P, LATAPY M. Computing communities in large networks using random walks[J]. Journal of Graph Algorithms and



Applications, 2006, 10(2): 191-218.

- [12] ROSVALL M, BERGSTROM C T. Maps of random walks on complex networks reveal community structure[J]. Proceedings of the National Academy of Sciences, 2008, 105(4): 1118-1123.
- [13] 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(10): P10008.
- [14] 王志宏, 杨震. 人工智能技术研究及未来智能化信息服务体系的思考[J]. 电信科学, 2017, 33(5): 1-11.
- WANG Z H, YANG Z. Research on artificial intelligence technology and the future intelligent information service architecture[J]. Telecommunications Science, 2017, 33(5): 1-11.

[作者简介]



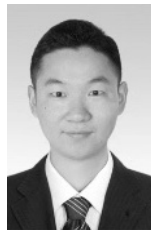
徐海勇（1970—），男，中移动信息技术有限公司总经理、高级工程师，主要研究方向为移动通信、互联网、大数据。



陶涛（1972—），男，博士，中移动信息技术有限公司副总经理、高级工程师，主要研究方向为大数据、通信技术。



黄岩（1976—），男，中移动信息技术有限公司电子渠道运营中心总经理、高级工程师，主要研究方向为大数据、能力开放。



唐崔巍（1992—），男，中移动信息技术有限公司工程师，主要研究方向为社交网络分析、机器学习算法。



张兆静（1990—），女，中移动信息技术有限公司工程师，主要研究方向为流量产品运营、数据分析。



吴晶（1979—），女，博士，中移动信息技术有限公司高级工程师，主要研究方向为计算机应用技术、数据挖掘、用户行为分析。