

人在社交网络中的投资行为研究

赵倩文, 吴 晔, 王胜烽

(北京邮电大学理学院 北京 海淀区 100876)

【摘要】以二元关系为基础的社交网络对于理解人的社会性非常重要,人们在社交关系上的投资模式很大程度上决定了社交网络的稳定性,而通讯是反映人与人之间情感交流的典型行为之一。本文通过中国某地区的通讯数据研究人们在构建和维系社会关系中的资源投入和分配行为,提出了一个基于熵的衡量人们资源分配均匀性的指标。并利用该指标对不同年龄、性别的手机用户进行通话行为分析,发现男性、女性处于不同的年龄阶段时,在社交网络中资源的投入和分配有着明显的差异。

关键词 熵; 资源分配; 性别差异; 社交网络

中图分类号 TN92; O41

文献标志码 A

doi:10.3969/j.issn.1001-0548.2018.02.022

Study on the Pattern of People's Investment Behavior Based on Mobile Phone Data

ZHAO Qian-wen, WU Ye, and WANG Sheng-feng

(College of Science, Beijing University of Posts and Telecommunications Haidian Beijing 100876)

Abstract Social networks based on binary relationships are very important for understanding people's sociality. People's social relationships largely determine the stability of social networks among which mobile communication is one of the typical way reflecting people's emotional connection. In this paper, we used a large mobile phone dataset from an area of China to study that how people build and maintain social relationships. To measure the uniformity of human emotional behavior, an entropy-based index is proposed and used to analyze the mobile phone data. The results reveals that there are obvious sex and age differences that people invest and assign their resources in social networks.

Key words entropy; resource allocation; sex differences; social network

一直以来,人们都对日常生活中隐含的自身行为特性抱着强烈的好奇心,并不停地展开探索和研究^[1-7]。在大数据时代,人们的电子足迹越来越容易被各种电子产品所捕捉。电信、金融、交通及媒体等领域每时每刻都产生着大量数据。人们通过各种数据分析技术搜集、分析这些数据,试图挖掘出有意义的信息,以加深对自身的了解,促进行业发展。

运营商所拥有的海量通讯数据一直是科研人员研究人类行为模式的重要来源。通过分析这些通话数据,人们不仅发现了不为人知的人类行为特性,如人的行为不是一个泊松过程,而是间隔时间是幂率分布的复杂过程。这个发现得到了广泛的认可和应用^[8-9]。同时有些学者还通过数据分析的方式佐证了部分社会学和心理学领域的研究论点,如文献[10]用facebook数据验证了邓巴观点,文献[11-12]分别提出和验证了弱链接、强作用的理论。

关于人类如何构建和维系自己的社交关系网络同样也备受关,国内外已有很多文章就此展开讨论。如文献[13]认为手机用户的联系人通话频次排名反映了用户特有的个人属性,而且这种属性不轻易受到网络动荡的影响;文献[14]通过分析用户通话的时间间隔模式对社交网络中的用户行为进行了分类;文献[15]观察了某在线游戏中男女玩家的不同游戏表现,证实男女玩家无论在个人层面还是网络层面,沟通模式都有很大的差别。通讯行为是人与人之间的典型社交行为之一,在一定程度上可以体现人与人之间的情感交互特点。本文通过中国某城市的运营商提供的用户之间的通讯数据,从通话行为的角度分析用户在维护社交关系中的资源投入和分配行为,并研究了性别、年龄等因素的影响,从而加深人们对复杂社会行为模式的理解。

收稿日期: 2017-03-24; 修回日期: 2017-09-12

基金项目: 国家自然科学基金(71231002, 11375033)

作者简介: 赵倩文(1991-),女,主要从事数据处理、行为预测等方面的研究。

1 数据

本文的数据来源于中国某地区电信运营商提供的33万手机用户的基本属性信息, 主要关注其年龄和性别, 以及他们在2012年2月期间的所有通话数据(用户作为拨打者), 包括主叫号码、被叫号码、通话开始时间、通话时长。在不失一般性的前提下, 对这些手机用户数据进行了以下处理:

1) 由于20岁以下和60岁以上的人群中手机普及率比较低(在对33万手机用户的年龄统计后证实了此观点), 因此本文只关注年龄在20~60岁之间的用户。

2) 通过观察用户的联系人数量, 发现有少量用户在1个月内与数千名联系人有电话联系, 根据“邓巴数字”^[16](即“150定律”: 人类智力允许人类拥有稳定社交网络的人数是148人, 四舍五入大约是150人)、社会经验等, 这些用户被认为是机器人用户、诈骗用户等异常用户^[14]。本文研究人们在日常生活中维持社交关系的资源投入和分配行为特性, 所以只选取了在这个月期间, 联系人数量在5~200人之间的手机用户的数据。

经过以上数据处理, 最后被采用的数据集包括253 037名用户的信息及其1个月内的通话数据, 其中包括男性用户170 906名, 女性用户82 131名, 年龄跨度为20~60岁。其中各年龄段男、女性人数所占比例如图1所示。

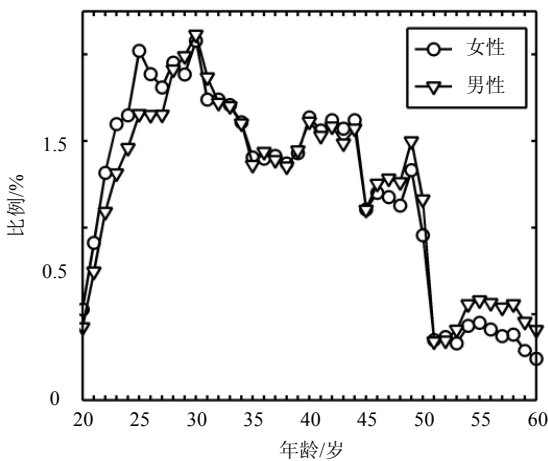


图1 各年龄段的男女人数比例

2 结果分析

2.1 资源投入

人们通过不同的社会关系构成了一个复杂的社交网络^[17-18]。本文在邓巴数的基础上进一步分析了

由于个人精力的局限性所引起的每个人在社交网络中资源投入的差异性, 研究不同年龄、性别个体在资源投入模式方面的差异。

为了研究人在社交网络中的资源投入模式, 本文认为人们的平均通话次数 $\langle M \rangle$ 、平均联系人数量 $\langle k \rangle$ 从一定程度上体现了这个人在社交网络中的资源投入程度。 $\langle k \rangle$ 、 $\langle M \rangle$ 随年龄变化趋势如图2所示, 可以看出: 1) 无论男女, $\langle M \rangle$ 、 $\langle k \rangle$ 随年龄的变化趋势都有一个先增加后减小的过程, 说明一个人随着年龄的增加社交圈在不断增大, 需要投入更多的精力, 而在其社交网络相对稳定后, 需要投入的精力就会逐渐减少。2) 男性在34岁, 女性在35岁的时候, $\langle M \rangle$ 、 $\langle k \rangle$ 达到峰值, 因为这个年龄一般是人们事业的高峰期和家庭的稳定期, 社交关系相对稳固和强健。为了维护这种状态下的社交网络, 人们投入更多的精力是容易理解的。3) 此外, 在25岁左右, $\langle M \rangle$ 、 $\langle k \rangle$ 达到一个“小高峰”, 因为这期间人们一般处于求学及建立恋爱关系的阶段, 是社交网络扩大的时期。

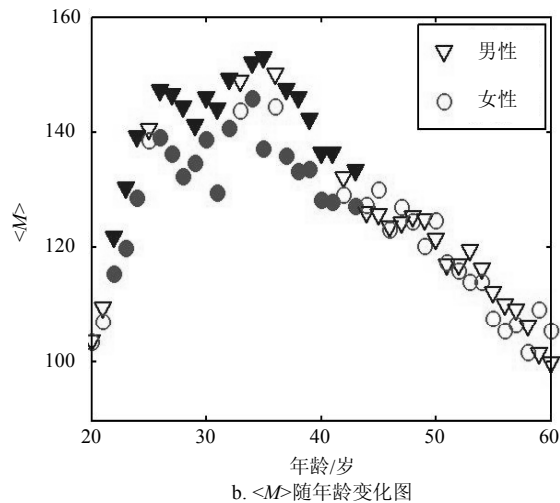
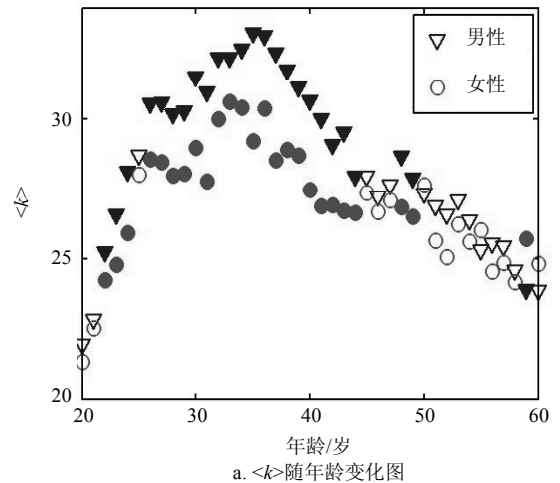
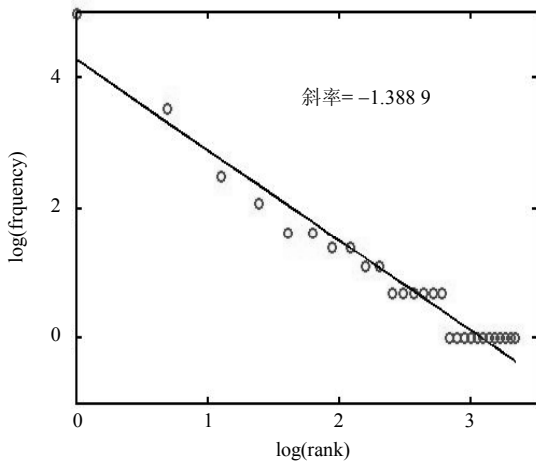


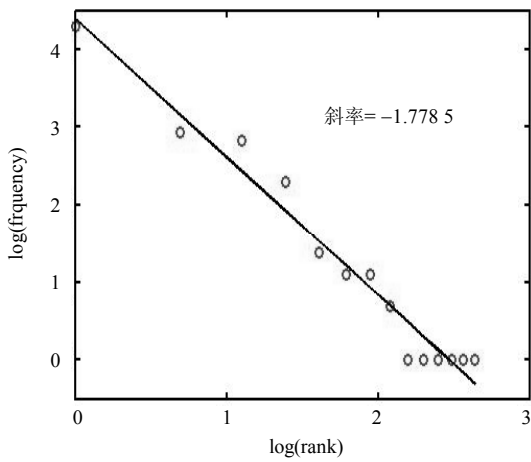
图2 $\langle k \rangle$ 、 $\langle M \rangle$ 随年龄变化图

此外,图2还显示了 $\langle k \rangle$ 分布、 $\langle M \rangle$ 分布的统计显著性t检验,其中拒绝概率 $P=0.05$, $H=1$ 。实心点表示具有统计显著性,空心点反之。可以看出,图中大部分点都具有统计显著性,表示用平均值 $\langle k \rangle$ 、 $\langle M \rangle$ 来刻画社交关系投入程度是合理的。从图2还可以看出,在20~45岁时期,男性的平均通话次数、平均联系人数量都比女性多,且大部分具有统计显著性。这说明男性在此时期投入的用于维护社交网路的资源更多。而随着年龄的增加,特别是45岁以后,男性和女性的差异变小。

2.2 资源分配



a. 用户A的通话频次图



b. 用户B的通话频次图

图3 随机两位用户通话频次图

前文主要讨论了不同性别、年龄的人在维系社交关系时的资源投入差别;本节将主要讨论个体在精力有限的前提下,对资源的分配模式差异。如有的人选择将精力均匀分配给所有朋友,有的人选择将资源集中分配给少数好友。如用户A与10个好友通话100次,通话次数序列为:80,8,4,2,1,1,1,1,1,1,而用户B与6个好友通话100次,通话次数

序列为:25,20,18,17,10,10,这两人对资源分配就存在着明显差异。为了定量地研究性别和年龄对这种资源分配的影响,本文首先把该用户与所有好友的通话次数按从大到小排序,得到通话次数的序列 $X_i, i=1,2,3,\dots$ 。本文随机挑选了两名用户的通话频次情况,发现通话次数与排名的关系为幂率分布,满足Zipf定律^[19],如图3所示。这一定律在文献[20-22]的相关研究中也类似发现。不同的人的Zipf定律的幂指数不一样,幂指数的大小在一定程度上体现了该人对资源的分配模式。幂指数越大,说明该人倾向于把大部分精力分配给少数的人;而幂指数越小,说明该人倾向与把自己的资源平均分配给朋友。

Zipf定律是在分析自然语言的语料库里单词出现频次时发现的,由于单词量的庞大,可以比较清楚地观察到幂律分布现象,然而在本研究1个月的通话数据中,人们的联系人数量比较有限,且部分联系人的联系频次一样。为了更准确细致地理解人们资源分配的模式,本文提出了一个基于熵的刻画分配均匀性的指标 Q 。熵是一种简洁但是十分有效的计算指标^[23],在一些人类行为动力学研究的文章中被广泛应用。如文献[24]利用相对熵刻画舆论传播中的信息变化;文献[25]通过熵指标来度量会话对象的可预测性。本文引用了文献[25]中的两种熵:随机熵和独立熵。随机熵为:

$$H_i^0 \equiv \log_2 k_i \tag{1}$$

式中, i 表示用户的标号; k_i 表示该用户的联系人数量。独立熵为:

$$H_i^1 \equiv -\sum_{j \in N_i} P_i(j) \log_2 P_i(j) \tag{2}$$

式中, N_i 表示用户 i 的联系人所构成的集合,集合中联系人的总数量就是式(1)中的 k_i ; $P_i(j)$ 表示用户 i 与所有联系人打电话的总次数中,与用户 j 打电话的次数所占的比例。因此, $\sum_{j \in N_i} P_i(j) = 1$,且 H_i^0 、 H_i^1 均为非负数。

本文主要考虑人们在维持朋友关系中的成本投入和资源分配模式,前者通过用户的朋友数 k_i 以及其通话的总次数 M_i 来刻画,后者在 H_i^0 、 H_i^1 的基础上提出一个新的衡量指标 Q_i 。可以证明,当用户与每名联系人通话次数一样,即资源精力分配均匀时, H_i^1 达到最大值,此时与 H_i^0 相等,而分配越不均匀的时候, H_i^1 越小。但是对比不同的用户的资源分配行为时,仅仅比较 H_i^1 是不够的,其原因就在于:对

于不同的用户, 联系人数量(即 k_i) 在大多数情况下是不相等的。因此, 本文提出了一个定义资源分配均匀性的指标:

$$Q_i = (H_i^0 - H_i^1) / H_i^0 \quad (3)$$

不难看出, Q 的大小可以体现出用户的资源分配倾向性模式。 Q_i 越小, 表示该名用户在维系朋友关系时, 资源分配越平均; 反之, Q_i 越大, 表示用户在联系朋友时倾向性越明显。

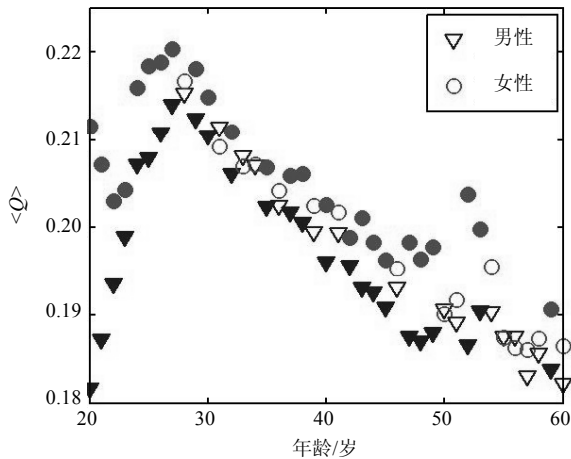


图4 $\langle Q \rangle$ 随年龄的变化

图4反映了 $\langle Q \rangle$ 随年龄的变化情况, 并对当前年龄下的男女对应的 $\langle Q \rangle$ 做了两个分布是否具有统计显著性的 t 检验, 拒绝概率 $P=0.05$ 。可以发现: 无论男性、女性, 其平均的资源分配倾向性在 20~28 岁左右期间逐渐增强, 之后减弱, 但是在 50 岁左右时有一个小强度的增强, 之后又缓慢减弱。20~28 岁期间随着年龄的增加、个性发展的增强、以及寻找异性伴侣的需要, 人的资源分配逐渐聚集到少数人身上, 以求达到个人利益的最大化, 随着这种伴侣关系的稳定, 工作关系的拓宽需要, 分配的倾向性逐渐减弱。

在人们 50 岁左右的时候, 子女一般处于 20~25 岁左右, 这正是他们初遇人生各种挑战机会的黄金时期。于是在这段时期, 父母沟通倾向性又明显地偏向于孩子, 和孩子之间的交流会格外频繁, 显得分配不均匀^[26]。从图4还可以看出, 在 50 岁左右, 女性对资源的分配比男性更不均匀, 这个是由于男女在祖父母关系中具有资源分配的差异性。

3 结束语

本文通过分析 20 余万手机用户的通信数据, 从通话行为的角度分析人类在社交关系中的投资行为, 发现人们在社交网络的资源投入方面具有这样

的特点: 平均层面上, 男性比女性的资源投入更多, 而且都是在 35 岁左右时投入达到最高峰。进一步的研究发现, 男性的成本投入存在明显的异质性(45 岁之前), 相当比例的男性或者极少投入, 或者投入极大, 而女性则稳定得多; 而在资源分配方面, 女性的资源分配倾向性更加明显, 而且, 随着年龄的增长, 无论男性、女性, 在 25~28 岁期间, 倾向性会达到巅峰, 之后逐渐减弱, 但在 50 岁左右时, 会有小阶段的增强。需要指出的是, 本文所采用的通话数据只是人们社交生活中的一部分, 只能在一定层面上刻画人们的情感交互行为, 如能结合微博, 微信等其他社交数据, 会更好体现人与人之间的情感交互规律。本文的贡献在于提出了从资源投入和资源分配两个角度分析人们在社交网络中的资源整合模式。此外, 还提出了一个基于熵的资源分配性指标 Q , 以便于定量分析。本文的研究有助于理解人作为一种群居性动物的内在复杂性。

参考文献

- [1] BARABASI A L. The origin of bursts and heavy tails in human dynamics[J]. Nature, 2005, 435(7039): 207-211.
- [2] OLIVEIRA J G, BARABÁSI A L. Human dynamics: Darwin and Einstein correspondence patterns[J]. Nature, 2005, 437(7063): 1251.
- [3] WEI H, XIAO P H, TAO Z, et al. Heavy-tailed statistics in short-message communication[J]. Chinese Physics Letters, 2009, 26(2): 028902.
- [4] ZHAO Z D, XIA H, SHANG M S, et al. Empirical analysis on the human dynamics of a large-scale short message communication system[J]. Chinese Physics Letter, 2011, 28(6): 68901.
- [5] JO H H, KARSAI M, KERTÉSZ J, et al. Circadian pattern and burstiness in mobile phone communication[J]. New Journal of Physics, 2012, 14(1): 013055.
- [6] ZHOU T, KIET H A T, KIM B J, et al. Role of activity in human dynamics[J]. Euro Physics Letters, 2008, 82(2): 28002.
- [7] KARSAI M, KASKI K, KERTÉSZ J. Correlated dynamics in egocentric communication networks[J]. Plos One, 2012, 7(7): e40612.
- [8] BARABÁSI A L, ALBERT R. Emergence of scaling in random networks[J]. Science, 1999, 286(5439): 509-512.
- [9] BARABASI A, FRANGOS J. Linked: the new science of networks science of networks[J]. Physics Today, 2002, 6(5): 243-270.
- [10] CAMERON M. Maintained relationships on facebook [EB/OL]. [2009-03-09]. <http://overstated.net/2009/03/09/aintained-relationships-on-facebook>.
- [11] GRANOVETTER M S. The strength of weak ties[J]. Social Networks, 1973, 78(6): 1360-1380.
- [12] BURT R S. Structural holes: the social structure of

- competition[M]. Cambridge, MA: Harvard University Press, 2010.
- [13] SARAMÄKI J, LEICHT E A, LÓPEZ E, et al. Persistence of social signatures in human communication[J]. Proceedings of the National Academy of Sciences, 2014, 111(3): 942-947.
- [14] JIANG Z Q, XIE W J, LI M X, et al. Calling patterns in human communication dynamics[J]. Proceedings of the National Academy of Sciences, 2013, 110(5): 1600-1605.
- [15] SZELL M, THURNER S. How women organize social networks different from men[J]. Scientific Reports, 2013, 3(7435): 1214.
- [16] Wikipedia. Dunbar's number[EB/OL]. [2017-12-06]. http://en.wikipedia.org/wiki/Dunbar%27s_number.
- [17] SCOTT J. Social network analysis[M]. LA: Sage Publications, 2012.
- [18] LAZER D, PENTLAND A S, ADAMIC L, et al. Life in the network: the coming age of computational social science[J]. Science, 2009, 323(5915): 721.
- [19] RIEMER S. Book review: Human behavior and the principle of least effort. An introduction to human ecology George K Zipf[J]. American Journal of Sociology, 1949, 110(110): 306.
- [20] ZHANG Z K, LÜ L, LIU J G, et al. Empirical analysis on a keyword-based semantic system[J]. The European Physical Journal B, 2008, 66(4): 557-561.
- [21] LÜ L, ZHANG Z K, ZHOU T. Zipf's law leads to Heaps' law: Analyzing their relation in finite-size systems[J]. Plos One, 2010, 5(12): e14139.
- [22] LÜ L, ZHANG Z K, ZHOU T. Deviation of Zipf's and Heaps' laws in human languages with limited dictionary sizes[J]. Scientific Reports, 2013, 3: 1082.
- [23] Wikipedia. Entropy(information theory)[EB/OL]. [2016-12-10]. [http://en.wikipedia.org/wiki/Entropy_\(information_theory\)](http://en.wikipedia.org/wiki/Entropy_(information_theory)).
- [24] BERGE-HOLTHOEFER J, PERRA N, GONÇALVES B, et al. The dynamics of information-driven coordination phenomena: a transfer entropy analysis[J]. Science advances, 2016, 2(4): e1501158.
- [25] TAKAGUCHI T, NAKAMURA M, SATO N, et al. Predictability of conversation partners[J]. Physical Review X, 2011, 1(1): 011008.
- [26] PALCHYKOV V, KASKI K, KERTÉSZ J, et al. Sex differences in intimate relationships[J]. Scientific Reports, 2012, 2(7394): 370.

编辑 蒋晓