微信朋友圈中人类行为动力学实证分析①

张生太

（北京邮电大学 经济管理学院，北京 100876）

摘 要：为了深入理解基于人类行为动力学的微信用户行为模式，研究实证分析了微信朋友圈中用户的发博行为。结果显示在一段连续的时间内，无论是用户发博数还是间隔时间都服从幂律分布。通过对个体活跃度的深入研究，我们发现，在个体活跃度和间隔时间幂律分布的幂律值之间存在着强烈的正相关关系。同时在阵发性和记忆性方面，研究发现，微信上的用户有着较弱的阵发性和记忆性，这同其他的在线系统的用户行为模式间有着明显的不同。在对阵发性与活跃度间关系进行分析后，得到了合理的解释，这也对传统的人类行为模式提出了一定的挑战。

关键词：微信；人类动力学；幂律分布；博客行为

中图分类号：C93 **文献标识码**：A **文章编号**：（2019）01-0011-08

0 引言

对人类行为的研究一直是社会学、心理学和经济管理学等众多学科关注的焦点问题[1]。但是由于人类行为的复杂性性和多样性，对于所有学科来说认识人类行为特征都是巨大的挑战。传统上，人们往往用齐次泊松过程来描述人的行为。而 2005 年 Barabási 发表在 Nature 上的一篇文章[2]却显示人类行为的时间规律具有高度的非均匀性：在非常长的时间内可能没有做事，而这些长长的空白与空白之间则被阵发的密集活动填充。

Barabási 等人的开创性工作，由于其理论和应用上的双重价值，很快就吸引了数学、系统科学和经济管理学等领域科学家的大量关注。研究者对现实生活、工作中人类活动的大量定量研究成果不断面世，如对商业交易[3]、网页浏览[4,5]、电影点播[6]、手机充值[7]、建筑工人事故发生[8]、邮件通信[9-11]、图书借阅[12]、物流运输[12]、网络购物[13]、网络评论[14]、新浪微博[15]、问答社区[16,17]、在线社交网络[18]、群体或组织行为进行了实证和理论研究，本领域的研究呈现蓬勃发展之势[19]。

微信发展迅速，已成为一种应用最广泛的以强关系社交为基础的移动社交网络应用软件。微信的活跃用户已经超过 10 亿[20]。微信众多使用者是潜在的巨大流量入口和也是强有力的消费群体。因此理解微信用户的人类动力学行为特征，对于理解微信用户的知识共享行为与信息传播的模式有重大的理论意义和应用价值。

但是，由于微信自身的特点使得获得微信数据非常困难，因而很少有人从微信获得数据研究人的行为规律。首先，微信私密性强[21]，微信朋友圈中的信息只有相互加为微信好友的人才能看到，微信中的社交关系以强人际关系为主。添加的好友主要是亲戚、朋友、同事以及同学，多为现实中认识的群体。其次，微信朋友圈中的信息不能像微博式转发形成长链条的多级传播，所以说微信的大众传播能力较弱[22]。与微博相较，微信尚不具备大众媒体的形态，可以说是私人化的交流通信工具且传播的信息内容以分享日常生活、抒发个人情感和评论热点事件为主。最后，微信中的主要用户是 20～30 岁的青年人，其占比达到 74%。而年轻用户往往把微信朋友圈的信息设定很短的时间可以看到。由此可知，微信主要是以青年人为主要用户群体，以强人际关系为主要社交关系的具有私密性特点的在线社交工具[23]。由于以上原因，

① 基金项目：国家自然科学基金项目（71571022）。

作者简介：张生太（1962—），男，山西应县人，博士，北京邮电大学经济管理学院教授，博士生导师，研究方向：知识管理、人力资源管理、企业战略管理，E-mail：sttzz@163.com。

微信用户的数据很难获取，因此通过微信获得大量数据研究人的行为规律的研究还很少。

本文以微信朋友圈为研究对象，从发微信数和时间间隔两个角度来对微信用户的知识分享行为描述，并结合活跃度，阵发性和记忆性等特征进行深入研究。文章的结构如下：在第 1 部分，我们将对数据集和使用方法进行描述；在第 2 部分，我们进行了实证分析，包括数量和间隔时间分布的研究、活跃度、阵发性与记忆性的描述和其中存在着的相关关系；在第 3 部分，我们进行了总结。

1　数据收集及研究方法

我们使用的数据集是 Python 编写爬虫从电脑后台爬取的微信朋友圈数据。这些数据是通过用户许可收集的个人在微信朋友圈较长时间发布信息的数据，总共获取了 12 966 条用户数据。经过数据清洗并剔除掉发博数低于平均值 20% 的数据。最终一共使用来自 628 位用户的 10 196 条数据。每一条数据都包括时间戳、提交者 ID、发表内容三个要素组成，其中时间的单位是分钟。

假设变量 x 服从参数为 a 的幂律分布，则其概率密度函数可以表示为

$$f(x) = Cx^{-\alpha} \tag{1}$$

常用的估计幂律的指数值 α 的方法有最小二乘法和极大似然估计等方法。其中最小二乘法已被证明与真实值之间存在较大差异，极大似然估计的算法则被证实是最为准确的方法。因此本文采用 Ethan 提出的极大似然估计估算 CDF 的方法来拟合幂律值。在原分布满足式（1）的情况下，其累计分布满足：

$$P(X \gg x) = C\int_x^{\infty} x^{-\alpha}\mathrm{d}x = \frac{C}{\alpha - 1}x^{-(\alpha-1)} \tag{2}$$

则可以利用极大似然估计估算，估算的概率累计密度的幂律值会比原分布幂律值 α 小 1，从而通过加上 1 而得到原分布值 α。同时为保证拟合值的准确性，研究将结合 Clauset 提出的 Kolmogorov-Smirnov 距离来做检验。

2　实证分析

图 1 展示的是同一用户连续提交的间隔时间的分布。其幂指数是 3.308，这与其他在线社交系统的值相差比较大，如 GitHub（值为 1.24），新浪微博（值为 2）。造成微信同其他在线社交系统的拟合值相差较大的可能原因是，微信更偏于私密化与生活化，这就使得微信用户在朋友圈中对于日常生活的分享会比较随意而不是和微博一样需要在意别人的看法而比较谨慎。

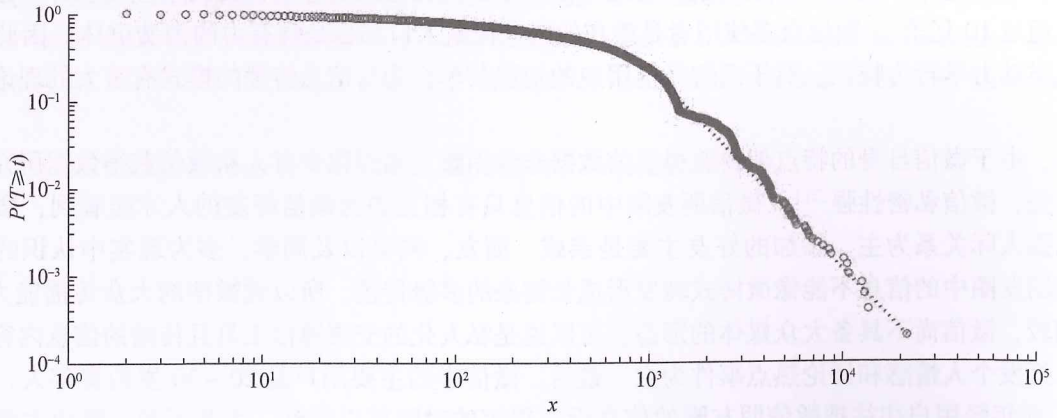

图 1　时间间隔分布

图中横坐标表示间隔分布，纵坐标表示概率累积密度。斜率为 2.308（原分布的幂律值为 3.308）。在［1000，21728］的 K-S 距离为 0.09。

Fig. 1　Inter-event time distribution

The abscissa represents the inter-event time distribution while the ordinate signifies the cumulative probability.

The slope is 2.308 (the power-law exponentof the original distribution is 3.308) . Kolmogorov Smirnovdistanceat [1000, 21728] is 0.09.

2.1 时间间隔分布

2.2 数量分布

如图2所示，发票圈的数量明显表现出胖尾分布的特性。证明大多数用户发票圈数目比较少，而少部分用户则发票圈数目比较多。同时，发票圈的数量明显服从幂律分布。幂律分布的拟合值是2.8，这比传统的在线人类活动值大很多，如电影评级（值为1.19），书签收集（值为1.19），在线拍卖（值为1.21），在线信息发送（值为1.24）和发推特（值为1.35）。造成这一现象的可能原因是微信的用户大多是青年群体，他们之间的差异性更小。

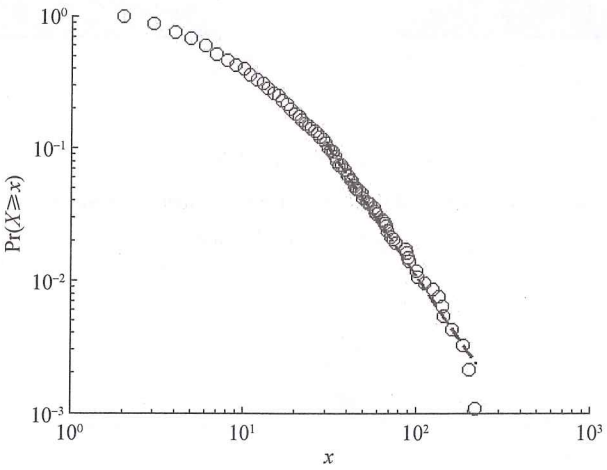

图2 数量分布

图中横坐标表示间隔分布，纵坐标表示概率累积密度。斜率为1.806（原分布的幂律值为2.806）。在［29，221］的K-S距离为0.04。

Fig. 2 Commit number distribution

The abscissa denotesthecommit numberand the ordinate denotesthe cumulative probability density. The slope is 1.806 (the power-lawexponent of the original distribution is 2.806). The Kolmogorov - Smirnov distance at [29, 221] is 0.04.

2.3 阵发性和记忆性影响

为了了解在没有假设分布必须符合幂律或者其他任何形式的分布情况下间隔时间的分布，Goh和Barabási定义了两个参数，阵发性和记忆性。阵发性集中于大量动作在长时间的沉寂之后在短时间内爆发。阵发性（B）被定义如下：

$$B=\frac{\sigma_\tau-m_\tau}{\sigma_\tau+m_\tau} \tag{3}$$

其中σ_τ和m_τ分别表示标准方差和动作时间的均值。我们从定义中可以得知：①B是一个从-1到1的实数。②当$B>0$时，标准差大于均值，这一类有阵发性并且B越大重尾现象越明显。③当$B=0$时，标准差等于均值，这是一中立的类。④当$B<0$时，这一标准差小于均值，这一系类有着规律的特性。

记忆性常被用于了解一个人类活动短（长）时间间隔之后出现短（长）时间间隔的趋势。

为了测度记忆性，这一系类的时间间隔含有$n\tau$个元素，我们从数列中抽取两个数列为S_1和S_2，其中S_1含有前$n_\tau-1$个元素而S_2含有后$n_\tau-1$个元素。记忆性（M）被定义为，S_1和S_2之间的皮尔逊相关系数：

$$M=\frac{1}{n_\tau-1}\sum_{i=1}^{n_\tau-1}\frac{(\tau_i-m_1)(\tau_{i+1}-2)}{\sigma_1\sigma_2} \tag{4}$$

式中，n_τ表示的是原始时间间隔数列中个数。τ_i和τ_{i+1}都是原始数列中的元素。而$\sigma_{1(2)}$和$m_{1(2)}$分别表示标准方差和均值。从记忆性的定义中我们可以知道：①M的范围从-1到1；②$M>0$表示S_1和S_2之间有正相关关系，即短（长）时间间隔后通常会伴随着短（长）时间间隔；③当$M<0$时表示S_1和S_2之间有负相关关系，即短（长）时间间隔后通常会伴随着长（短）时间间隔。

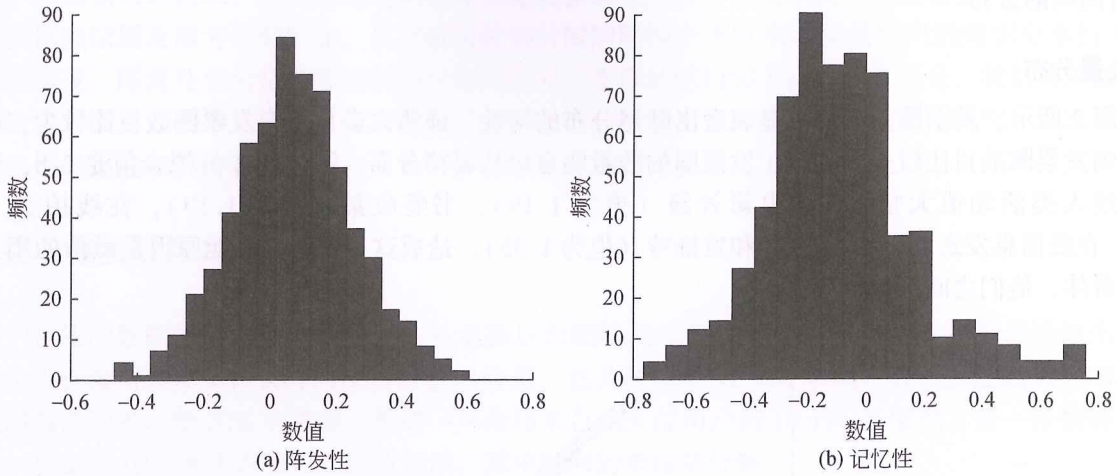

图 3　阵发性和记忆性影响

Fig. 3　The abscissa of (a) and (b) represent value of burstiness and memory respectively while the ordinate represents frequency of each group

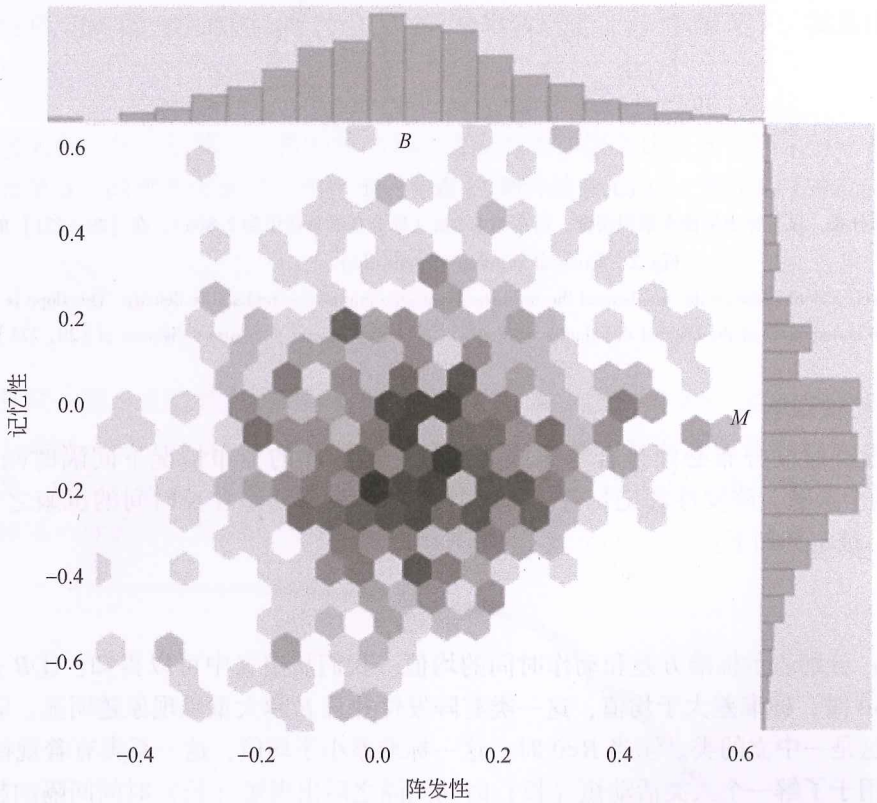

图 4　用户行为模式图

横轴表示阵发性，纵轴表示记忆性。图中六边形颜色的深浅，表示某一特定行为模式的用户的多少，颜色越深，用户越多。

Fig. 4　Diagram of user behavior patterns

The abscissa represents the burstiness while the ordinate represents the memory. The shadeof color of a hexagon signifies the relative group density of a particular user behavioral pattern. The darkerthe color, the moreusers belong tothat pattern.

从图 3 中我们能看出，阵发性和记忆性都是单峰的，这表示大多数用户的阵发性和记忆性集中在某一特定值。因此，我们可以使用平均值去描述这个系统。所有用户的阵发性和记忆性的均值是 0.06 和 -0.08，表示微信用户有着较弱阵发性和记忆性。图 4 展示了大多数微信用户的行为模式处于 low-B，low-M区域但。这与前人的研究相违背，Goh 和 Barabási 提出的人类的行为模式应该处于 high-B，low-M 区域。而微信用户显然不是，关于这部分的解释我们将在下一部分中进行讨论。

2.4 活跃度

为了更好地理解微信用户的行为特征图。我们分析了微信用户和间隔时间分布之间的关系。Ghoshal等提出了一个活动参数A_i，去衡量一个使用者在一个系统中的活跃度。对于任意一个用户i的活跃度定义为：

$$A_i=\frac{n_i}{T_i} \tag{5}$$

式中，n_i为用户i总共的活动数；T_i为用户第一个动作和最后一个动作之间的时间差。

2.4.1 活跃度对时间间隔分布的影响

我们将用户按照活跃度分为4组，并对每个组的间隔时间分布和活跃度进行研究。图5展示了每组用户的幂律拟合值。图6则展示了用户活跃度和间隔时间分布的幂律值成正相关的关系。且随着活跃度的升高，幂指数的上升速度变慢。这与之前的研究是一致的。这个结果意味着活跃的微信用户更倾向于有规律的方式分享自己的生活。

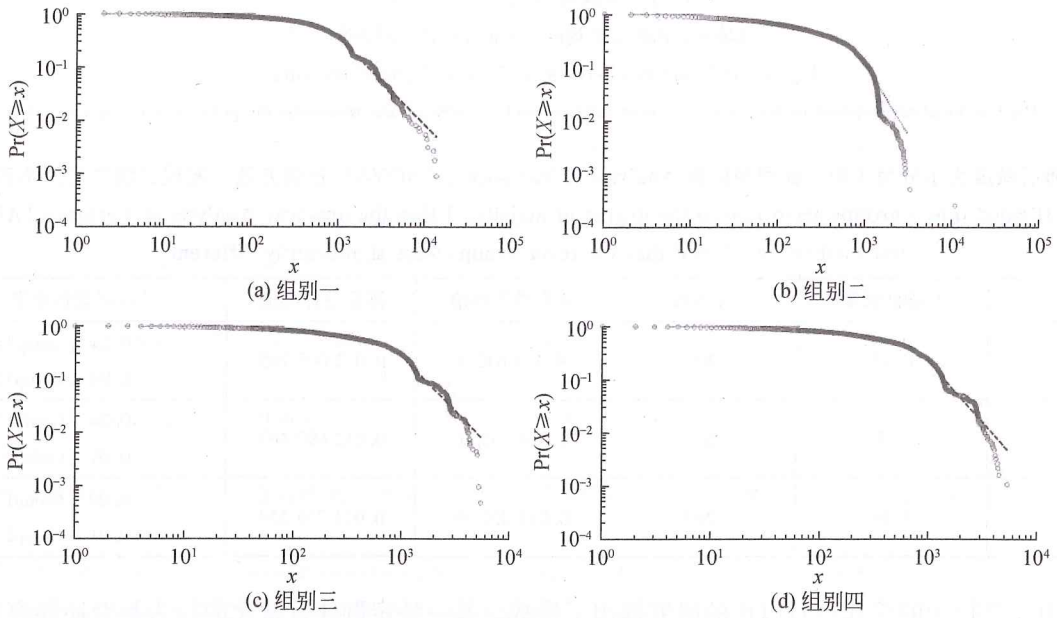

图5 每组用户的幂律拟合值

（a）（b）（c）（d）横轴表示间隔时间分布，纵轴表示概率累计密度。（a）幂律拟合值为2.645，K-S距离为0.08。（b）幂律拟合值为3.161，K-S距离为0.09。（c）幂律拟合值为3.389，K-S距离为0.1。（d）幂律拟合值为3.668，K-S距离为0.14。

Fig. 5 Power-lawvalue' svariation alongwithuser activitychanges

The abscissa represents average activity while the ordinate denotes the power-law fitting exponent. Group (a): Power-law fitting exponentis 2.645. Kolmogorov-Smirnovdistance is 0.08. Group (b): Power-law fitting exponentis 3.161. Kolmogorov-Smirnovdistance is 0.09. Group (c): Power-law fitting exponentis 3.389. Kolmogorov-Smirnovdistance is 0.10. Group (d): Power-law fitting exponentis 3.668Kolmogorov-Smirnovdistance is 0.14.

2.4.2 活跃度对阵发性的影响

表1展示了每组阵发性的数值。经过单因素ANOVA检验，我们可以知道，每组之间差异都很显著。证明活跃度越高的组，其阵发性值将越低。关于这一点很好解释。活跃度高的用户会将朋友圈中的知识分享当作日常活动，因此就不会存在有长时间寂静过后的阵发性特点。基于这一点并结合微信的特性，我们可以解释为什么微信用户的行为图处于low-B，low-M的区域。首先微信的青年人占比非常大，青年人日常使用微信的频繁非常高，几乎每天都会发朋友圈，所以他们的活跃度都处于比较高的水平。此外，微信中有一部分是微商型用户，他们发表朋友圈比较有规律，同时他们的活跃度更高，因此这也将拉低整体微信用户的阵发性。

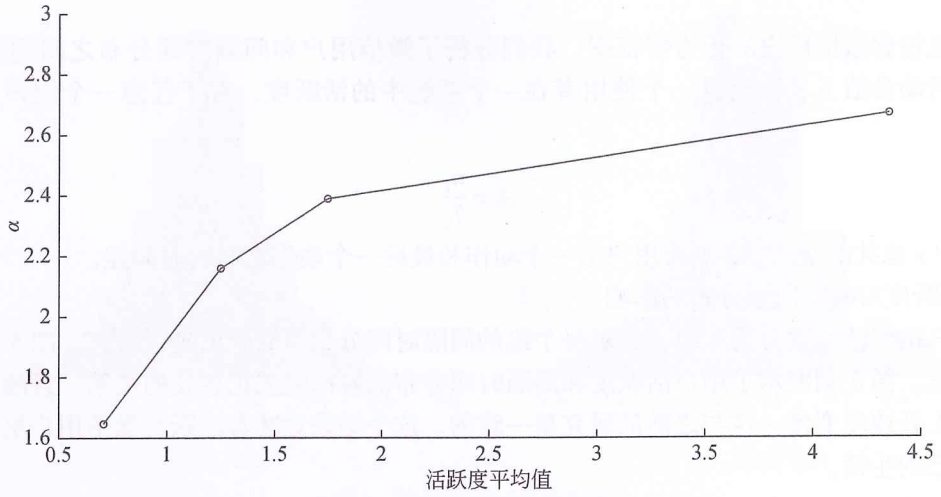

图 6　活跃度与幂律值关系图

横轴表示活跃度平均值，纵轴表示幂律拟合值。

Fig. 6　Relationship between activity and power-law value

The horizontal axis represents the average value of activity, and the vertical axis represents the power-law fitting value.

表 1　按照活跃度大小分为 3 组。使用单因素 Analysis of Variance（ANOVA）检验方法，发现三组之间互相有显著性

Table 1　Divided into 3 groups according to the degree of activity. Using the one-way Analysis of Variance (ANOVA) test method, we found that the three groups were significantly different

分组	平均活跃度	样本数	阵发性平均值	阵发性标准值	显著性水平
1	0. 57	209	0. 113 642 3	0. 012 007 795	0. 24 （Group2） 0. 00 （Group3）
2	1. 3	209	0. 074 735 16	0. 012 690 467	0. 24 （Group1） 0. 01 （Group3）
3	3. 16	210	0. 017 051 39	0. 011 734 554	0. 00 （Group1） 0. 01 （Group2）

微信中人类行为的模式也对以往的研究提出了挑战。某一特定群体日常化的行为是否还能遵从传统的人类行为模式将提出质疑。在合理的实证分析之下，可以说以往的研究确实可能存在缺陷。

3　结论与讨论

为理解微信用户的行为模式，我们已经实证分析了微信用户的发朋友圈行为。结果显示微信用户的发票圈数符合重尾分布，这表示大多数用户发票圈数比较少，而小部分的用户发票圈数比较多。同时发票圈数和间隔时间的分布都符合幂律分布。间隔时间的幂律值是 3. 308。

微信用户的时间间隔分布与其他的在线系统较为不同，显著的大于其他的在线系统。这可以归结于微信中主要群体之间的行为差异性不大。同时我们的结果显示微信用户的行为和其他的在线系统是不同的，其幂律值大于其他的在线系统，如在线信息的发送和推特信息的发送。我们可以把这一原因归结为微信的私密化与生活化。在这一情境之下，微信用户的知识共享的行为也会偏向规律而不是和微博一样因需要在意别人的看法而比较谨慎。

在记忆性和阵发性方面，微信用户的阵发性和记忆性都不显著。相比于其他的在线系统，微信用户的阵发性显著偏弱。同时与前人有关人类行为模式的研究不完全相符。在将时间间隔分布和阵发性与用户的活跃度结合分析后，我们知道个体活跃度和间隔时间的幂律分布的指数值之间存在着强烈的正相关关。同时对活跃度与阵发性结合分析后发现活跃度越高的组，其阵发性值将越低。这样便能解释微信用户行为模式出现反常的原因。微信主要用户群体所具有的高活跃度与日常化的行为使得用户行为的阵发性变低。这

也对传统的人类行为模式的分布提出了挑战。

本文中我们很难完全排除掉微信用户中某些微商型用户的存在，由于这些数据的存在会对实证分析的结果产生一定的影响。同时我们对微信用户行为模式反常的原因没有提出对模型进行深一步的验证与研究。在未来的研究中，我们将收集更多的数据并对反常的微信用户的行为模式进行深入研究。

参考文献：

[1] 韩晓燕，朱晨海．人类行为与社会环境［M］．上海：格致出版社，2009：28-31.
Han X Y, Zhu C H. Human Behavior and Social Environment［M］. Shanghai: Gezhi Press, 2009: 28-31. (in Chinese)

[2] Barabási A L. The origin of bursts and heavy tails in human dynamics［J］. Nature, 2005, 435: 207-211.

[3] Vazquez A, Oliveira J G, Dezso Z, et al. Modeling bursts and heavy t ails in human dynamics［J］. Physical Review E, 2006, 73 (3): 036127.

[4] Dezso Z, Almaas E, Lukacs A, et al. Dynamics of information access on the web［J］. Physical Review E, 2006, 73 (6): 066132.

[5] 郭进利，周涛，张宁，等．人类行为的动力学模型［M］．香港：上海系统科学出版社，2008：47-53.
Guo J L, Zhou T, Zhang N, et al. Model of Human Dynamics［M］. Hong Kong: Shanghai Systems Science Press, 2008: 47-53. (in Chinese)

[6] Zhou T, Kiet H A T, Kim B J, et al. Role of activity in human dynamics［J］. Europhysics Letters, 2008, 82 (2): 28002.

[7] Wang P, Ma Q. From heavy-tailed to exponential distribution of interevent time in cellphone top-up behavior［J］. Physica A: Statistical Mechanics and its Applications, 2019, 473: 10-17.

[8] Zhou C, Chen R. Jiang S, et al. Human dynamics in near- miss accidents resulting from unsafe behavior of construction workers［J］. Physica A, 2019, 8 (13): 121495.

[9] 李凯，张锡哲，申毓佩，等．基于手机通话网络的人类交互行为分析［J］．计算机科学与探索，2017，5：777-784.
LIK, Zhang X Z, Shen Y P, et al. Analysis of human interactive behavior based on phone communication networks［J］. Journal of Frontiers of Computer Science and Technology, 2017, 5: 777-784. (in Chinese)

[10] Oliveira J G, Barabasi A L. Human dynamics: Darwin and Einstein correspondence patterns［J］. Nature, 2005, 437: 1251-1253.

[11] 李楠楠，张宁，周涛．人类通信模式中基于时间统计的实证研究［J］．复杂系统与复杂性科学，2008，5（3）：43-47.
Li N N, Zhang N, Zhou T. Empirical analysis on temporal statistics of human correspondence patterns［J］. Complex Systems and Complexity Science, 2008, 5 (3): 43-47. (in Chinese)

[12] 樊超，郭进利，纪雅莉，等．基于图书借阅的人类行为标度律分析［J］．图书情报工作，2010，54（15）：35-39.
Fan C, Guo J L, Ji Y L, et al. Analysis of human behavior scaling law based on library loans［J］. Libary and Information Service, 2010, 54 (15): 35-39. (in Chinese)

[13] Wang Q, Guo J L. Human dynamics scaling characteristics for aerial inbound logistics operation［J］. Physica A Statistical Mechanics and its Applications, 2010, 389: 2127-2133.

[14] 徐雪娟，郭进利，何静．网络购物行为的人类动力学模式［J］．复杂系统与复杂性科学，2013，10（4）：69-75.
Xu X, Guo J L, He J. Human activity pattern on on-line shopping［J］. Complex System and Complexity Science, 2013, 10 (4): 69-75. (in Chinese)

[15] Tian Y, Xin F, Ye W. Human dynamics in repurchase behavior based oncomments mining［J］. Physica A: Statistical Mechanics and its Applications, 2018, 502: 518-533.

[16] Yao W Y, Jiao P F, Wang W J. Understanding human retweeting patterns on Sina Weibo from a global perspective［J］. Physica A: Statistical Mechanics and its Applications, 2018, 12 (10): 374-383.

[17] 沈洪洲，史俊鹏．基于人类动力学的社会化问答社区优秀贡献者行为研究——以“知乎”为例［J］．情报科学，2019，5：85-91.
Shen H, Shi J. Research on the behavior of outstanding contributors in social Q&A community based on human dynamics—taking “Zhi hu” as an example［J］. Information Science, 2019, 5: 85-91. (in Chinese)

[18] 李瑾颉，吴联仁，齐佳音，等．基于人类动力学的在线社交网络信息传播研究［J］．电子与信息学报，2017，39（4）：785-793.
Li J, Wu L R, Qi J Y, et al. Research on information dissemination in online social network based on human dynamics［J］.

Journal of Electronics & Information Technology, 2017, 39 (4): 785-793. (in Chinese)

[19] 樊超，郭进利，韩筱璞，等．人类行为动力学研究综述 [J]．复杂系统与复杂性科学，2011，6 (2)：1-17.
Fan C, Guo J L, Han X P, et al. A review of research on human dynamics [J]. Complex Systems and Complexity Science, 2011, 6 (2): 1-17. (in Chinese)

[20] https://baike.baidu.com/item/%E5%BE%AE%E4%BF%A1/3905974?fr=aladdin.

[21] 方兴东，石现升，张笑容，等．微信传播机制与治理问题研究 [J]．现代传播，2013，(6)：122-127.
Fang X D, Shi X S, Zhang X R, et al. Research on the communication mechanism and governance of We Chat [J]. Modern Media, 2013, 35 (6): 122-127. (in Chinese)

[22] 王欢，祝阳．人际沟通视阈下的微信传播解读 [J]．现代情报，2013，33 (7)：24-27.
Wang H, Zhu Y. Micro-letters dissemination analysis from interpersonal communication perspective [J]. Journal of Modern Information, 2013, 33 (7): 24-27. (in Chinese)

[23] 张生太，金丹．移动社交网络微信的发展及潜在价值分析 [J]．北京邮电大学学报（社会科学版），2015，17 (2)：1-7.
Zhang S T, Jin D. Analysis of development and potential value of we chat based on mobile social networking [J]. Journal of Beijing University of Posts and Telecommunications (Social Sciences Edition), 2015, 17 (2): 1-7. (in Chinese)

An Empirical Analysis of Human Behavior Dynamics in WeChat Friends

Zhang Shengtai

School of Economic Management, Beijing University of Posts and Telecommunications, Beijing 100876, China

Abstract: In order to deeply understand WeChat users' behavior pattern based on human behavior dynamics, the research empirically analyzes the blog sending behavior of WeChat friends. The results show that in a continuous period of time, both the number of users' blogs and the interval time obey the power law distribution. Through in-depth study of individual activity, we find that there is a strong positive correlation between individual activity and the power-law value of power-law distribution at intervals. At the same time, in terms of paroxysmal and memory, the study found that users on WeChat have weak paroxysmal and memory, which is obviously different from the user behavior patterns of other online systems. After analyzing the relationship between paroxysmal and activity, a reasonable explanation has been obtained, which also poses a certain challenge to the traditional human behavior pattern.

Key words: WeChat; Human Dynamics; Power Law Distribution; Blog Behavior