

# 微博社区中用户行为特征及其机理研究

闫 强, 吴联仁, 郑 兰

(北京邮电大学经济管理学院 北京 海淀区 100876)

**【摘要】**基于微博数据的实证分析表明, 微博用户发布微博的数量及时间间隔均服从幂律分布。通过分析微博用户之间的社会网络关系与用户行为之间的联系, 发现微博消息数量的分布幂指数与用户互动指数呈反向关系, 且随着互动指数的升高, 幂指数的下降趋缓。在排除用户之间的评论、转发行为的影响后, 个体用户发微博的时间间隔分布幂指数与用户互动指数呈正相关关系。基于上述实证分析, 引入用户间评论转发行为、用户间的社会网络关系作为微博用户行为的驱动因素, 提出了一种基于社会关系的动力学模型, 模型仿真结果与实际数据一致。

**关键词** 人类动力学; 互动指数; 微博; 幂律分布; 社会网络

中图分类号 N949

文献标志码 A

doi:10.3969/j.issn.1001-0548.2013.03.002

## Research on User Behavior Characters and Mechanism in Microblog Communities

YAN Qiang, WU Lian-ren, and ZHENG Lan

(School of Economics and Management, Beijing University of Posts and Telecommunications Haidian Beijing 100876)

**Abstract** According to the empirical study about microblog, the message volume of and releasing intervals of microblog users observe power-law distribution. Through analyzing the relationship between the social network of microblog users and their behaviors, the distribution exponent of message volume is proved to be inversely proportional to the interaction exponent of each user. With the increase of interaction exponent, the power-law exponent decreases much slower. By excluding the influence from other users' forwarding and commenting behavior, it is confirmed that the power-law exponent of the interval distribution is positively correlated with users' interaction exponent. Accordingly, a social-relation-based dynamic model is proposed to reflect the behavior of posting microblogs, such as forwarding, commenting, and social network relation among users. The simulation results of the model match well with empirical data.

**Key words** human dynamics; interaction exponent; microblog; power-law distribution; social networks

近年来, 微博应用发展迅速, 已成为一种重要的信息传播途径, 其在个人信息分享、企业口碑传播以及政务信息发布等方面扮演着越来越重要的角色。对于这种新兴的信息传播媒介, 用户使用行为的动力学特征是什么? 用户使用行为的驱动机制是什么? 回答这些问题对微博的运营与使用具有重要意义。

本文以新浪微博为研究对象, 从发布微博的数量和时间两个角度分析了用户行为的动力学特征, 并基于微博用户间的社会网络关系, 提出了用户行为的动力学模型。

### 1 相关研究进展

2005年, 文献[1]研究了用户电子邮件发送、回

复等待等人类行为的时间间隔分布。研究表明, 电子邮件发送和等待的时间分布均满足幂指数为-1的幂律分布, 显示出长时间的静默与短期内的高频率爆发共存的特性, 显然不符合泊松分布。文献[2]发现同一用户连续访问同一网址的时间间隔分布符合幂律分布, 幂指数约为-1。文献[3]对美国一个移动服务提供商提供的一周内的移动搜索日志进行了统计分析, 结果表明, 人类使用移动设备搜索信息的检索间隔时间在一天内服从幂律分布, 超出一天后以指数形式衰减。文献[4]指出用户在门户网站的子页上花费的时间服从幂律分布。文献[5]分析了网络论坛上人们浏览和回复帖子的行为特征, 指出浏览和回复帖子的数量均服从幂律分布, 且幂律系数在

收稿日期: 2012-04-05; 修回日期: 2012-11-13

基金项目: 教育部新世纪优秀人才支持计划(NCET-11-0597); 中央高校基本科研业务费(2012RC1002)

作者简介: 闫强(1972-), 男, 博士, 教授, 主要从事网络动力学、商务智能方面的研究。

-3.5和-1.6之间。通过分析回复帖子时间的突发性和记忆性,发现所有帖子的时间序列有相似的量化特征。文献[6]分析了用户点播电影的行为特征,研究表明,在群体水平上,所有用户间隔时间分布在超过两个数量级的范围服从幂律分布,幂指数和对应人群观看电影的活跃程度之间存在单调的关系;在个体层次上,观察到了具有胖尾特征分布,以及分布宽度和活跃程度之间的负相关性。文献[7]对在线音乐服务系统的日志数据进行研究后发现,个体行为和音乐流行度均服从介于指数分布和幂律分布之间的一种扩展的指数分布,连续收听两首音乐的间隔时间分布具有胖尾特征,随着用户活动的减少,胖尾特征会变得越来越不规则。文献[8]将网络的节点分成稳定节点和随时间不断变化的新闻文档类节点,研究表明,不同用户浏览网页的特点决定了一个文档访问量随着时间幂律减小。

对于人类在线行为的理论解释,一种观点认为人类活动遵循任务优先选择机制<sup>[1]</sup>,高优先级的任务优先执行。老龄化机制认为任务的优先权与时间有关,截止时间最早的任务最先被执行<sup>[9]</sup>。文献[10]分析了截止时间对任务等待时间分布的影响。文献[11]从经济最优化的角度解读人类任务的完成过程。另一种观点从兴趣的角度解释人类的行为特征。人类自身兴趣的周期性涨落导致行为时间间隔上的胖尾分布<sup>[12-13]</sup>。文献[14]则认为记忆作为人类的一个重要属性,记忆时间长度对人类动力学特征有着重要的影响。文献[15-16]认为,当前网络用户多为年轻人,他们在追求个性化的同时,也希望通过在线评论等行为引起他人的关注和认同,因而提出了社会认同驱动的动力学模型,指出了用户活跃度与行为动力学特征之间的相关关系。还有研究指出,用户之间的互动沟通会影响到用户对互联网业务的使用,用户的行为会受到网络中相邻节点的影响<sup>[17]</sup>。关于人类行为动力学的研究,文献[18-20]作出了较好的综述。

上述研究成果揭示了人类在线行为时间间隔的幂律分布特征,这些特征在微博环境中是否依然存在?对微博用户行为的动力学特征是否可以从其他角度进行理论解释?本文将对这些问题进行研究。

## 2 实证数据分析

### 2.1 数据采集

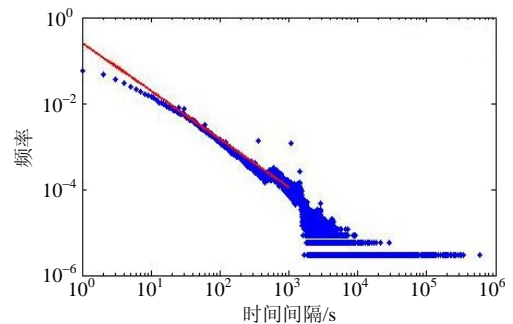
本文数据采集于新浪微博,新浪微博是中国用户规模最大、活跃程度最高的微博网站之一,据

Hitwise数据显示,2011年4月,新浪微博的访问量占到了中国互联网总访问量的0.63%,相当于每158次访问量就有一次是访问新浪微博。因此,对新浪微博用户行为的分析,在一定程度上可以反映微博用户的一般行为特征。

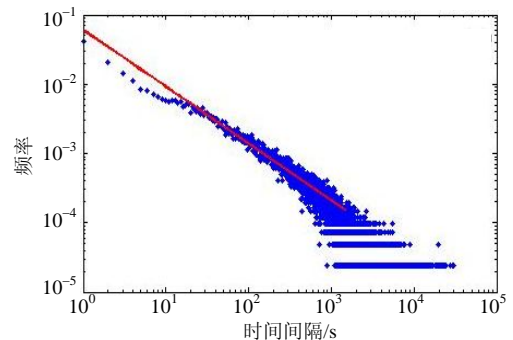
本文采用滚雪球抽样获取了41 667个用户的基础信息,并随机收集了其中291名用户从开始使用微博之日起至获取数据日为止的每条微博信息,共计345 096条,以此为基础对用户的微博使用行为进行分析。

### 2.2 用户行为分析

本文对微博用户行为的分析如图1所示,主要包括两方面,一是使用微博的间隔时间分析,从时间维度上观察用户行为特征;二是发出的微博数量,从总量维度上观察用户行为特征。由图1a可以看出,用户发微博间隔时间呈现幂律分布的特征。这表明对于整体用户而言,频繁的发表微博通常是在大段的停顿之后,存在阵发特征。用户发表微博的总数量同样呈现幂律分布,如图1b所示。只有较小比例的用户会很活跃地发表原创微博,大部分用户发表微博数目较少。



a. 双对数坐标下用户发表微博的时间间隔分布,幂指数为1.123



b. 双对数坐标下用户发表微博的数量分布,幂指数为0.818 7

图1 微博用户行为的总体情况

已有研究表明,用户之间的相互关系会影响用户对互联网业务的使用行为。因此,下面将分析微博用户之间的社会网络关系与用户行为之间的联系。在微博中,用户之间的社会网络关系表现为关

注与粉丝两种情况。如果把用户看作社会网络分析中的节点，用户之间的联系作为边，则关注与粉丝分别对应为节点的出度与入度。本文采用互动指数衡量用户间的社会网络关系与互动水平。具体来说，对某一个给定的用户其互动指数定义为用户的关注数与粉丝数之和，即节点入度与出度之和。

本文将用户按互动指数从小到大的顺序分为6组，每组的人数大致接近，保证互动指数相同的用户分在同一个组内，如表1所示。根据图2所示的6组用户发表微博的数量分布，可以发现随着用户互动指数的增强，用户发表微博数量分布的幂指数逐渐减小。图3为互动水平与微博数分布幂指数间的对应关系。

表1 互动指数分组情况

组别	用户数	互动指数	幂指数
1	6 257	35	1.468
2	6 298	97	0.827 8
3	6 328	195	0.580 6
4	6 335	393	0.451 3
5	6 330	1 333	0.448 9
6	6 433	128 252	—

可以看出，微博数量分布的幂指数与用户互动指数呈反向变动趋势。在互动指数相对较低的区间，幂指数下降的速度较快；而互动指数较高时，幂指数下降的速度变慢，表明群体用户微博使用行为与用户之间的社会关系水平具有较明显的相关关系。但当互动指数水平上升到一定程度时，微博数分布的幂指数变化不再明显，此时群体用户微博使用行为不会再随社会关系水平的变化发生变动了，二者之间的相关关系也较弱。当前，微博运营中普遍使用个性化推荐鼓励用户间的相互关注行为，在一定互动关系水平范围内，该措施固然可以增强群体用户的微博使用行为。但超过一定限度，该方法反而收效甚微。因此，对于微博运营商来说，一方面可以改良现有的个性化推荐算法，提供更优质的关注推荐给用户；另一方面，对那些互动关系水平较高的用户群体，则不能再仅仅依靠推荐其关注更多的人来鼓励他们使用微博了，而是需要从其他方面来刺激其微博使用行为。

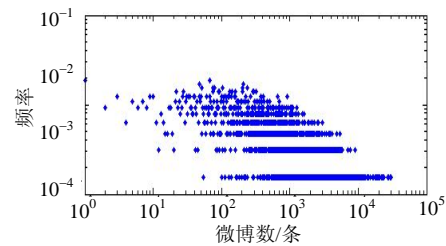
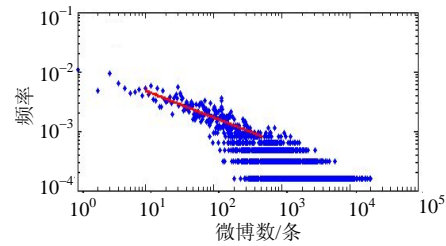
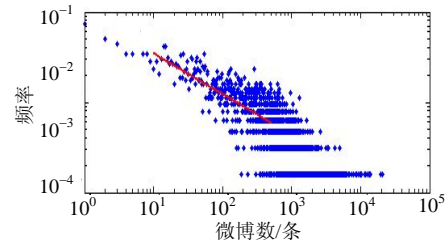
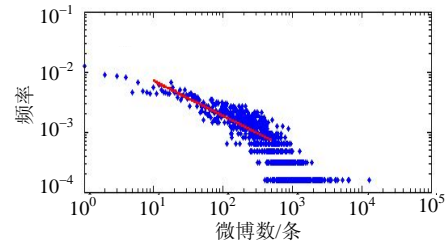
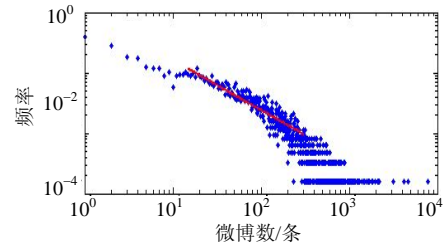
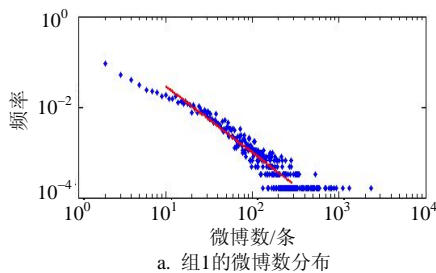


图2 具有不同互动水平的用户发表的微博数量分布

—●— 按互动关系数分组 —■— 按粉丝数分组 —▲— 按关注数分组

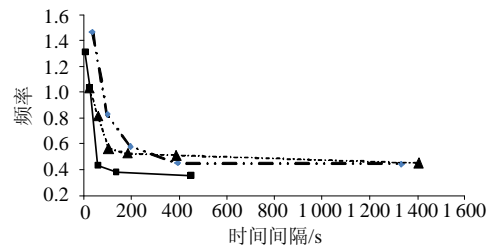


图3 不同互动水平下用户微博数分布的幂指数变化趋势

用户发表微博的时间间隔,除了和互动水平相关外,还与其微博内容受到其他用户的评论、转发程度有关。为消除用户接受到的转发评论对其行为的影响,选取接受到转发评论水平一致的6名用户进行分析,考察个体层面上用户间互动水平与其发微博行为间的关系。如表2所示,在个体层面上,随着互动指数的下降,用户发微博间隔时间分布幂指数减小,这与文献[6]的结论是一致的。幂指数变小,说明在该间隔时间分布中,较小的间隔时间占比较低,而较大的间隔时间占比较高,最终导致用户发微博的数量变少。特别地,当互动指数下降到一定水平时(如表2中用户6的情况),间隔时间特别长,导致分布的幂律现象不再明显。

表2 具有不同互动指数的用户发微博间隔时间分布幂指数

用户	互动指数	幂指数
1	1 396	1.285
2	711	1.034
3	503	1.000 2
4	275	0.928 1
5	161	0.675
6	46	—

### 3 基于社会关系的用户行为模型

上述分析表明,用户间的社会互动关系水平与用户的微博使用行为间存在相关关系。本节将基于社会关系对用户微博使用行为进行建模。

调查研究表明,用户使用微博的行为多发生在日常工作、生活中的碎片时间,更多的是受兴趣的驱动,而非任务的驱动。对用户兴趣的度量,本文采用了微博中的转发与评论行为、用户间的社会网络关系水平来衡量,后两者比兴趣更易于客观评价。

#### 1) 假设条件。

① 行为诱因1:用户发微博行为受其微博中好友(包括用户的关注和其粉丝)转发/评论行为的影响。用户接收到来自其关注或粉丝的转发/评论越多,用户发表微博的概率越大。

② 行为诱因2:如果用户微博好友数越多,用户发表微博的概率越大。

#### 2) 模型描述。

①  $s_i^t$ 表示节点*i*在时刻*t*的状态, $s_i^t=0$ 表示节点*i*在时刻*t*未发微博, $s_i^t=1$ 表示节点*i*在时刻*t*发表了微博。在仿真时,初始时刻随机选取*r*个节点使其状态变为1。

② 时刻*t*,用户*i*的转发评论行为会对其邻居节点时刻*t+1*的发微博行为产生影响。用*E(i,t)*表示时刻*t*节点*i*对邻居节点产生的影响, $E(i,t)=c_i s_i^t$ ,

$c_i$ 表示节点*i*的权威性,也即用户*i*在该微博社会网络所具有的影响力大小,用网络中节点的相对点度中心度量表示为 $c_i=k_i/(N-1)$ , $k_i$ 表示节点*i*的度数, $N$ 为网络规模。

③ 如果节点*j*有 $k_j$ 个邻居节点,那么行为诱因1来自其 $k_j$ 个邻居节点行为,行为诱因2则源自自身的社会网络结构特性,因此节点*j*受到的影响因素为 $k_j$ 个邻居节点再加上其自身,共计 $k_j+1$ 个,每个影响因素对*j*产生的影响只占总影响的 $1/(k_j+1)$ 。

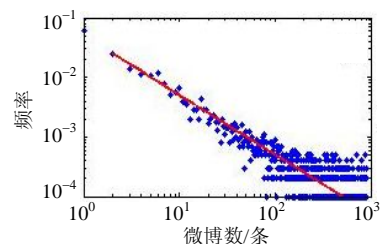
④ 在*t+1*时刻,节点*j*受到其邻居节点*i*的影响为 $d_{ij}E(i,t)$ 。 $d_{ij}$ 为节点*i*对节点*j*的影响因子,与节点*j*所拥有的邻居节点个数有关。如果节点*j*有 $k_j$ 个邻居节点,那么节点*i*所施加的影响力只能发挥 $d_{ij}=1/(k_j+1)$ 。节点*j*在*t+1*时刻接收到 $k_j$ 个邻居节点的总影响为 $I_{j1}=\sum_{i=1}^{k_j} \frac{1}{k_j+1} c_i s_i^t$ 。

⑤ 以 $k_j$ 反映节点*j*的社会网络关系对其微博使用行为的影响,该影响只能占节点*j*受到的总影响的 $1/(k_j+1)$ ,令 $I_{j2}=k_j/(k_j+1)$ 。

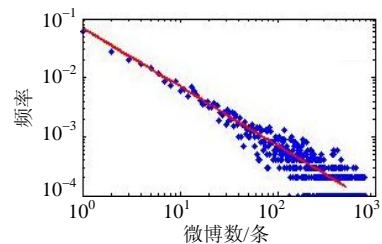
⑥ 最后,节点*j*在*t+1*时刻受到的总影响力为 $I_j=I_{j1}+I_{j2}$ ,从节点集中选取*M*个受到总影响力较大的节点将其状态置为1,表示该用户节点在*t+1*时刻会发表微博。剩下的节点不实施任何行为。

#### 3) 数值模拟。

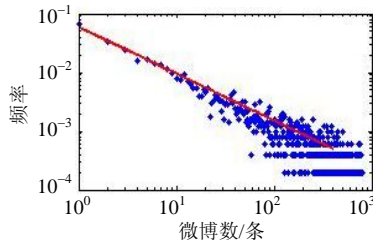
对上述提出的微博社会网络模型进行数值模拟。首先设定微博用户总数即节点数 $N=1\ 000$ ,模拟时间步是1 000。为了与表1中的实证数据进行比较,分别设定互动指数为35, 97, 195, 393和1 333。模拟结果及其与实证数据的比较如图4和图5所示。



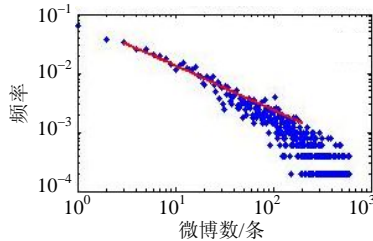
a. 互动指数为35, 相应的幂指数分别为1.45



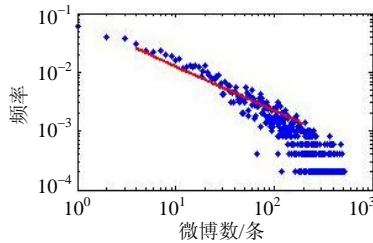
b. 互动指数为97, 相应的幂指数分别为0.9



c. 互动指数为195, 相应的幂指数分别为0.7



d. 互动指数为393, 相应的幂指数分别为0.55



e. 互动指数为1333, 相应的幂指数分别为0.5

图4 模型仿真结果

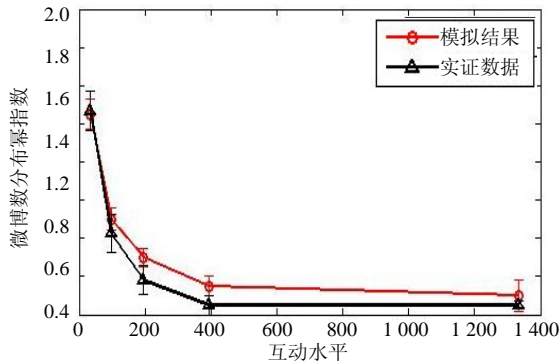


图5 模拟结果与实证数据比较

### 4 结论

微博近年来成为一种发展迅猛的网络信息发布媒介。本文分析了微博用户在信息发布时间与数量方面的行为特征,探讨了用户之间关注、粉丝等社会网络关系与用户行为特征之间的联系。基于新浪微博数据的分析表明,微博用户发布微博的数量及时间间隔均服从幂律分布。通过区分微博用户互动指数的强度,分析结果表明,微博数量的分布指数与用户度呈反向关系,且随着度的升高,指数的下降趋势趋缓。在排除用户之间的评论、转发行为的影响后,个体用户发微博的时间间隔分布指数与用

户度呈正相关关系。基于上述实证分析,论文提出了一种基于社会关系的动力学模型,指出微博用户的行为更多的是受兴趣的驱动,而非是任务的驱动,并通过微博中的转发与评论行为、微博用户间的社会网络关系水平作为对用户行为驱动机制的衡量指标,模型的仿真结果与实际情况较为吻合。

但是,本文提出的模型在用户行为诱因方面作了相应的简化,也未将本模型与其他模型的效果进行对比分析,今后将在这些方面进一步深入研究。

### 参 考 文 献

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

[2] GONÇALVES B, RAMASCO J J. Human dynamics revealed through Web analytics[J]. Phys Rev E, 2008, 78(2): 026123.1-026123.7.

[3] VOJNOVIĆ, M. On mobile user behaviour patterns[C]//Proceedings of International Zurich Seminar on Communications. Switzerland, Zurich: IEEE, 2008: 26-29.

[4] CHMIEL A, KOWALSKA K, HOŁYST J A. Scaling of human behavior during portal browsing[J]. Phys Rev E, 2009, 80(2): 066122.1-066122.7.

[5] YU J F, HU Y Q, YU M, et al. Analyzing netizens' view and reply behaviors on the forum[J]. Physica A-Statistical Mechanics and Its Applications, 2010(389): 3267-3273.

[6] ZHOU T, KIET H A T, KIM B J, et al. Role of activity in human dynamics[J]. Euro Phys Lett, 2008(82): 28002-28006.

[7] HU H B, HAN D Y. Empirical analysis of individual popularity and activity on an online music service system[J]. Physica A, 2008(387): 5916-5921.

[8] DEZSÖ Z, ALMAAS E, LUKÁCS A, et al. Dynamics of information access on the web[J]. Phys Rev E, 2006, 73(6): 066132.1-066132.7.

[9] BLANCHARD P H, HONGLER M O. Modeling human activity in the spirit of barabasis queueing systems[J]. Phys Rev E, 2007, 75(2): 026102.

[10] 邓竹君, 张宁, 李季明. 截止时间对人类动力学模型的影响[C]//人类行为的动力学模型. 上海: 上海系统科学研究院, 2007: 123-127.

DENG Zhu-jun, ZHANG Ning, LI Ji-ming. Influence of deadline on human dynamic mode[C]//Dynamic Model of Human Behavior. Shanghai: Shanghai System Science Publishing House, 2007: 123-127.

[11] DALL'ASTA L, MARSILI M, PIN P. Optimization in task-completion networks[J]. Journal of Statistical Mechanics: Theory and Experiment, 2008(2): P02003.2-P02003.19.

[12] 郭进利. 博客评论的人类行为动力学实证研究和建模[J]. 计算机应用研究, 2011, 28(4): 1422-1433.

GUO Jin-li. Empirical study and modeling of human behaviour dynamics of comments on Blog posts[J]. Application Research of Computers, 2011, 28(4):

- 1422-1433.
- [13] SHANG M S, CHEN G X, DAI S X et al. Interest-driven model for human dynamics[J]. Chinese Phys Lett, 2010, 27(4): 48701-48703.
- [14] VÁZQUEZ A. Impact of memory on human dynamics[J]. Phys A, 2007(373): 747- 752.
- [15] YAN Q, WU L R, YI L L. Research on the human dynamics in mobile communities based on social identity[J]. Discrete Dynamics in Nature and Society, 2012: 672756.1-672756.13.
- [16] YAN Q, YI L L, WU L R. Human dynamic model co-driven by interest and social identity in the microblog community[J]. Physica A: Statistical Mechanics and Its Applications, 2012(391): 1540-1545.
- [17] CHRISTINE K, MARTIN B. Leveraging network effects for predictive modeling in customer relationship management[EB/OL]. (2005-10-10). <http://ssrn.com/abstract=883588>.
- [18] 韩筱璞, 汪秉宏, 周涛. 人类行为动力学研究[J]. 复杂系统与复杂性科学, 2010, 7(2-3): 132-144.  
HAN Xiao-pu, WANG Bing-hong, ZHOU Tao. Researches of human dynamics[J]. Complex System and Complexity Science, 2010, 7(2-3): 132-144.
- [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, 2010, 8(2): 1-17.  
FAN Chao, GUO Jin-li, HAN Xiao-pu, et al. Review of human dynamics[J]. Complex System and Complexity Science, 2010, 8(2): 1-17.
- [20] 李楠楠, 周涛, 张宁. 人类动力学基本概念与实证分析[J]. 复杂系统与复杂性科学, 2008, 5(2):15-24.  
LI Nan-nan, ZHOU Tao, ZHANG Ning. Concept and empirical results of human dynamics[J]. Complex System and Complexity Science, 2008, 5(2): 15-24.

编辑 蒋 晓