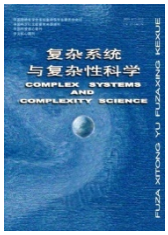

复杂系统与复杂性科学
*Complex Systems and Complexity Science*
ISSN 1672-3813,CN 37-1402/N

# 《复杂系统与复杂性科学》网络首发论文

题目：网络直播大数据：统计特征与时序规律挖掘
作者：郭淑慧，吕欣
收稿日期：2021-09-06
网络首发日期：2022-10-21
引用格式：郭淑慧，吕欣．网络直播大数据：统计特征与时序规律挖掘[J/OL]．复杂系统与复杂性科学.
https://kns.cnki.net/kcms/detail/37.1402.N.20221021.1413.002.html

网络首发时间：2022-10-21 15:44:10
网络首发地址：https://kns.cnki.net/kcms/detail/37.1402.N.20221021.1413.002.html

# 网络直播大数据：统计特征与时序规律挖掘

郭淑慧[1] 吕欣[1]
（1. 国防科技大学系统工程学院 长沙 410073）

**摘要：** 为挖掘网络直播领域数百万主播与数亿计观众的活跃互动下大规模人群行为学特征，以斗鱼和虎牙直播平台为例，统计分析了连续 123 天、涉及 240 多万名主播、超过 7.26 亿条的直播数据，总结了直播平台的负载时序规律和用户行为特征。发现直播负载存在明显的日内效应和周内效应，不同直播模式的主播在观众数、粉丝数等统计特征上存在显著的组间差异，主播生存期和直播间观众数量符合幂律分布，随着平台发展，主播和观众数量呈现较强的线性相关性，但其波动性也逐步增大，体现出系统越来越强的异质性和非均匀性。对理解网络直播复杂系统中的用户行为模式、挖掘用户分布规律及变化趋势、设计商业模式如个性化推荐等方面具有重要意义。

**关键词：** 网络直播；直播平台；大数据；流量分析；行为动力学
**中图分类号：** TP391，G358　　**文献标识码：** A

# Data Mining of Live Streaming Platforms: Statistical Characteristics and Temporal Pattern

Guo Shuhui[1], Lü Xin[1]
[1](College of Systems Engineering, National University of Defense Technology, Changsha 410073, China)

**Abstract:** To explore the behavioral characteristics of massive crowds under the active interaction of millions of anchors and viewers in the field of live streaming, this paper summarized the temporal patterns of live streaming workload and user behavior characteristics of the live streaming platform, taking Douyu and Huya live streaming platforms as examples, a statistical analysis of 123 consecutive days, involving more than 2.4 million anchors, and more than 726 million live streaming data. The live streaming workload has obvious intra-day and intra-week effect. Different live streaming modes have significant differences in live streaming characteristics such as average audience and fan number. The lifetime of anchors and the number of viewers conform to a power law distribution. With the development of the platform, there is a strong linear correlation between the number of anchor and viewer, but its volatility is gradually increasing, reflecting the increasingly strong heterogeneity and non-uniformity of the system. It is of great significance for understanding user behavior patterns in complex systems of live streaming, mining user distribution laws and changing trends, and designing business models such as personalized recommendations.

**Key words:** Live streaming; live streaming platform; big data; workload analysis; behavioral dynamics

## 1 引言

近年来，随着移动通信和互联网技术的发展进步，网络直播逐渐成为了新媒体环境下人们青睐的在线娱乐和信息传播方式。目前除了应用于娱乐性的真人秀、电竞赛事之外，还广泛应用于课堂教学[1]、品牌营销[2]、传统文化与工艺技术传承[3]、政务会议与庭审过程公开[4]等方面。中国互联网网络信息中心第 49 次互联网发展报告显示，截止 2021 年 12 月，我国网络直播用户规模已达 7.03 亿，占网民总体的 68.2%[5]；艾媒咨询发布的《2021Q3 中国在线直播行业研究报告》显示，中国在线直播行业的发展态势稳定，泛娱乐直播、电商直播、以及企业直播等领域都吸引了更广泛的用户群体[6]。数百万主播与数亿计观众的活跃加入和互动，产生了海量的在线人群行为活动数据，为开展大规模人群行为动力学研究、优化直播平台性能和用户体验等提供了丰富的实验场景。与此同时，大数据储存和处理水平的提高为网络直播平台流量量化研究提供了技术支持，为研究网络直播平台中大规模人类行为动力学提供了重要推力。

目前，网络直播领域的研究主要集中在通过分析真实直播流量数据挖掘直播平台负载水平[7-9]、观众行为[10,11]、主播行为[12,13]以及社群网络[14,15]的特征和变化规律[16]，对大规模人群行为动力学特征[17-20]、直播平台优化方法[17,21]、直播行业发展状况[22]等方面进行分析研究。基于直播平台大规模用户参与及交互数据的统计规律挖掘和行为动力学研究，对信息传播、网络营销、舆情监测引导等领域具有重要的参考和指导意义，但在目前已有的社交媒体复杂系统分析中，被广泛应用的社会媒体平台主要是微博、Twitter、百度指数、谷歌趋势等以文本为主的社会媒体，对网络直播复杂系统的分析和研究较为不足，对直播平台统计特征及时序规律的定量研究较少，对网络直播情境下负载规律以及用户特征等方面有待进一步探索和挖

**收稿日期：** 2021-09-06；2022-03-17
**基金项目：** 国家杰出青年科学基金（72025405）；国家自然科学基金重大研究计划（91846301）；国家社科基金重大项目（22ZDA102）。
**第一作者：** 郭淑慧（1996-），女，博士研究生，主要研究方向为社交媒体大数据分析挖掘。
**通信作者：** 吕欣（1984-），男，教授，博士，主要研究方向为大数据挖掘、复杂网络、应急管理、人类行为动力学。

掘。

为了量化研究网络直播情景下的大规模人群动力学特性，挖掘网络直播复杂系统的统计特征和独特规律，本文以斗鱼和虎牙平台为期 123 天，涉及 240 多万名主播、超过 7.26 亿条的直播数据为例，从大规模人群动力学的挖掘角度出发，基于直播平台大规模用户参与及交互数据统计分析了直播平台的负载时序规律、主播直播规律、观众分布规律，以大规模真实时序直播数据的多方面统计特征，多方位展现了主播与用户共生、主播异质性尤其明显的直播生态系统，为以直播为背景的大规模人群行为、用户社群网络分布和演化规律以及平台优化等研究提供了坚实的数据基础和理论支撑，以网络直播复杂系统分析为例为其他社会媒体复杂系统的分析挖掘提供了泛化性较强的研究框架。

## 2 直播数据集概览

斗鱼 TV 和虎牙 TV 是国内直播市场占有率较大的两个直播平台，大量用户活跃其中。直播平台的用户分为两种角色，即主播和观众。直播平台为注册并审核通过的主播提供模拟房间即直播间，主播可以在平台定义的直播类型列表中自定义直播间的直播类型，通过网络游戏或其他内容在直播间内向观众展示自己，观众可以向主播赠送虚拟礼物表达对主播的支持。除了主播对观众的视频内容传输，每个直播间都设有内置的弹幕交流区，用于用户主播之间用文字和表情符号等进行互动（如图 1 所示）。

(a) (b)


**图 1 (a) 斗鱼 TV 直播间示意图 (b) 虎牙 TV 直播间示意图**

**Fig.1 (a) Douyu TV live streaming room. (b) Huya TV live streaming room.**

与斗鱼 TV 和虎牙 TV 类似，Twitch.TV 是一家国际性直播平台，直播内容聚焦在电子竞技类直播，包括多人在线战术竞技（MOBA）、射击、策略、格斗、军事类电子游戏的视频直播，而斗鱼 TV 和虎牙 TV 的直播类型除了电子竞技类直播外，还包含语音直播、颜值互动、科技文化等类型的直播，涵盖的直播类型相对较广。据公开资料显示，2021 年斗鱼、虎牙和 Twitch 的月活跃用户数（MAU）分别为 6190 万、8510 万和 1900 万。本文运用 python 爬虫技术对斗鱼和虎牙平台内全部直播间的真实运行情况进行连续爬取，得到了关于两大直播平台的大量直播间数据，经过数据筛选和处理，作为直播平台流量分析的数据集，数据字段、获取方式及数据集统计信息见表 1。

**表 1 直播数据集统计信息**

**Tab.1 Live streaming dataset statistics**

| | 斗鱼平台 | 虎牙平台 |
|---|---|---|
| 数据起止日期 | 2019.03.06—2019.07.06 | |
| 获取时长 | 123 天 | |
| 获取时间间隔 | 10 分钟 | |
| 平台快照数（个） | 16,938 | 16,626 |
| 直播数据数（条） | 329,108,181 | 396,945,241 |
| 不重复主播（名） | 992,758 | 1,485,322 |
| 直播数据属性字段 | 直播间及主播的 ID 和名称、直播类别、观众数、粉丝数、当次直播开播时间、数据采集时间 | 直播间 ID、直播间名称、主播 ID、主播名称、直播类别、观众数、数据采集时间 |
| 斗鱼直播数据样例 | cate_id fans nick_name now_time online owerner_id room_id room_name show_time<br>0 563 14320 GOD白神 2019-03-06 21:01:18 23038 197963358 4710750 白神：努力呀 2019-03-06 19:34:38<br>1 181 3664911 王者荣耀官方赛事 2019-03-06 21:01:18 2907243 118232622 1863767 首战！三冠王Hero vs 宇宙战舰QG 2018-12-28 02:27:20<br>2 350 1723737 Dy花老湿 2019-03-06 21:01:18 1136854 64593276 952595 空投猎手！22:30抽1000红包 2019-03-06 20:06:07 | |
| 虎牙直播数据样例 | fans game_id game_name introduction nick_name now_host now_time recommend room_id room_name user_id<br>0 35204 4913 探灵笔记 【探灵笔记】天级人皇！千级小红！ 秋玥梦 429961707 2019-03-06 21:01:53 0 15812161 玥梦【主机游戏】千元榜冠名 506824865<br>1 13133 4913 探灵笔记 来了老弟！！！ 葫芦娃-MG 2085222204 2019-03-06 21:01:53 0 308639 虎牙直播 1832575997<br>2 13098 4913 探灵笔记 【狗叔】嗯 不知道写啥 狂鸟、狗叔-90327 sougou 2019-03-06 21:01:53 0 10085379 狗叔：各类单机游戏直播！ 16823928 | |

直播数据集的时间跨度为 2019 年 3 月 6 日到 2019 年 7 月 6 日共计 123 天，以 10 分钟为时间间隔，

通过斗鱼平台开放数据接口（API）对直播平台所有开播的直播间信息进行抓取，除去网站结构变化等原因造成的少部分数据漏抓，斗鱼平台直播数据集包括 329,108,181 条直播数据，涉及 992,758 个唯一主播。在相同的时间段，以相同的时间间隔，获取虎牙平台内所有开播直播间的实时数据，除去网站结构变化等原因造成的少部分数据漏抓，虎牙平台直播数据集包含了 1,485,322 个唯一主播和 396,945,241 条直播数据。

## 3 直播平台负载时序规律

受时间节律的影响，金融市场的流动性[23]、人类的情绪积极程度[24]、反应灵敏度[25]、器官工作机能[26]等都会在一天内不同时段表现出显著差异，股市收益率和波动还存在明显的周内效应[27]。鉴于日内效应和周内效应在人类行为各领域上的广泛存在，本节对直播负载的日内效应和周内效应进行发掘和讨论。

### 3.1 日内效应

直播平台主播和观众数量在 24 小时内的变化曲线展现了直播平台负载的变化趋势。从图 2（a）中可以观察到直播平台负载存在明显的日内效应，呈现降低-升高-降低的循环模式，经单因素方差分析发现，不同时刻下主播数量及观众数量具有显著性差异（$P<0.001$）。直播平台的主播和观众数量都是在早上 6 至 7 时跌至谷底，然后白天持续增长，21 至 22 时左右达到峰值之后回落，在一天之内呈“倒 N 型”的变化趋势，符合年轻用户偏好晚间娱乐的生活作息规律，而且与已有的直播平台负载变化趋势基本一致[7,9,28]。

从负载规模和波动来看，相同时刻的主播规模基本相同，波动性较小，主播数量在 0.59 万到 4.10 万之间变动；斗鱼的观众规模在各个时刻都领先于虎牙，两个平台的观众数量峰值分别为 4.63 亿和 2.84 亿，但虎牙平台的观众数量波动更强。观众规模及波动性存在差异，一方面由于斗鱼和虎牙平台分别由数据接口和网站页面获取，另一方面由于斗鱼平台的官方游戏直播间粉丝基数更大，此类官方直播的权威性和垄断性致使观众群的观看粘性强，所以斗鱼平台的观众更多、更稳定。

从主播数量增加时间为 6 时到 21 时（除 16 时至 17 时外）而观众数量增加时间为 7 时到 22 时、主播数量在 21 时到 6 时持续减少而观众数量则在 22 时到 7 时持续减少、主播数量在 19 时到 20 时增速最快而观众数量在 19 时到 22 时增速最快的现象中可以发现，主播数量和观众数量的增减变化趋势基本一致，而且主播数量增加能够起到带动观众数量增加的作用。

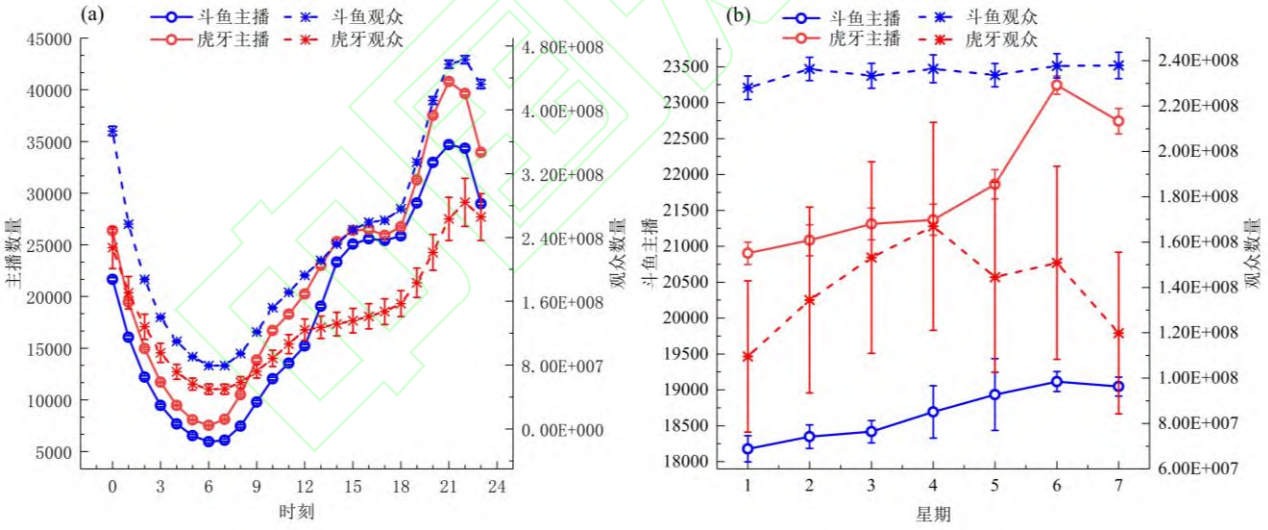


**图 2 (a) 主播及观众数量日内变化图 (b) 主播及观众数量周内变化图**

**Fig.2 (a) Intra-day pattern in the number of anchors and viewers. (b) Intra-week pattern in the number of anchors and viewers.**

### 3.2 周内效应

从直播平台周内负载变化曲线来看（图 2（b））, 包括主播和观众数量在内的直播平台负载从周一到周日的变动趋势存在周内效应，经单因素方差分析发现，周内各天的主播数量日均值及观众数量日均值均具有显著性差异（$P<0.001$）。

斗鱼平台的主播数量从周一至周六呈持续增加趋势，由周一的最小值 1.82 万增长到周六的最大值 1.91 万，周六至周日存在略微降低（0.01 万/天）；但观众数量呈波动增加的变动趋势，除周二到周三、周三到周四呈现下降趋势外均呈上升趋势，观众数量在周一取得最小值 2.28 亿，在周日取得最大值 2.38 亿。虎牙平台的主播数量增长趋势与斗鱼平台类似，从周一的 2.10 万持续增加至周六的 2.32 万，在周五至周六

增速最快为 0.14 万/天，但在周六到周日出现下降趋势（0.05 万/天）；观众数量从周一到周日呈先增长后下降趋势，在周四达到峰值 1.67 亿，最小值在周一取得为 1.10 亿。

斗鱼和虎牙平台的主播数量水平从周一到周六均呈现由低到高的变化，双休日的主播数量水平明显高于其他时段，这与已有国外直播平台 Twitch 的负载研究中双休日在线主播数量略高于工作日[7]的结论基本一致，说明了直播平台存在大量主播仅选择在双休日进行直播，体现了这部分主播选择进行网络直播来填充大量空闲时间的特征。但直播平台观众数量在一周中的变化趋势并不是很统一，双休日的观众数量也并没有显著高于工作日。综合上节提到的观众和主播数量的日内变化趋势中观众数量在 19 时到 21 时增速最快、峰值出现在晚间 21 时至 22 时，可以发现，现阶段观众观看网络直播具有明显的娱乐性特征，在双休日以及工作日的晚间时段均有大量观众进入直播间，造成直播平台负载迅速增加。

### 3.3 长期变化趋势

本节以天为测量窗口统计分析了长达 123 天的主播和观众日内均值的变化情况，体现了直播平台、直播行业的发展前景和未来走向（见图 3）。从中可以看出，直播平台的主播和观众数量均有上升趋势。主播数量基本处于波动上升的状态，斗鱼平台从 3 月 6 日的 16859 位主播上升到 7 月 6 日的 19253 位主播，虎牙平台主播数量从 20175 上升到 23938。斗鱼平台的观众数量基本处于稳定的小幅度增长状态，从 1.80 亿增长到了 2.58 亿；虎牙平台观众数量在 6 月 12 日之前呈现缓慢的增长，6 月 12 日出现了一个大幅度跃升之后维持高位并继续增加，从 0.53 亿增长到了 5.55 亿。从图 3 中可以看出虎牙平台主播数量的波动变化呈现出较为明显的周期性特征，经分析发现其主播数量时间序列与滞后 7 天的时序序列的自相关系数最大（$r$=0.75，$p$<0.01），进一步说明直播平台主播的直播模式具有明显的周内效应。

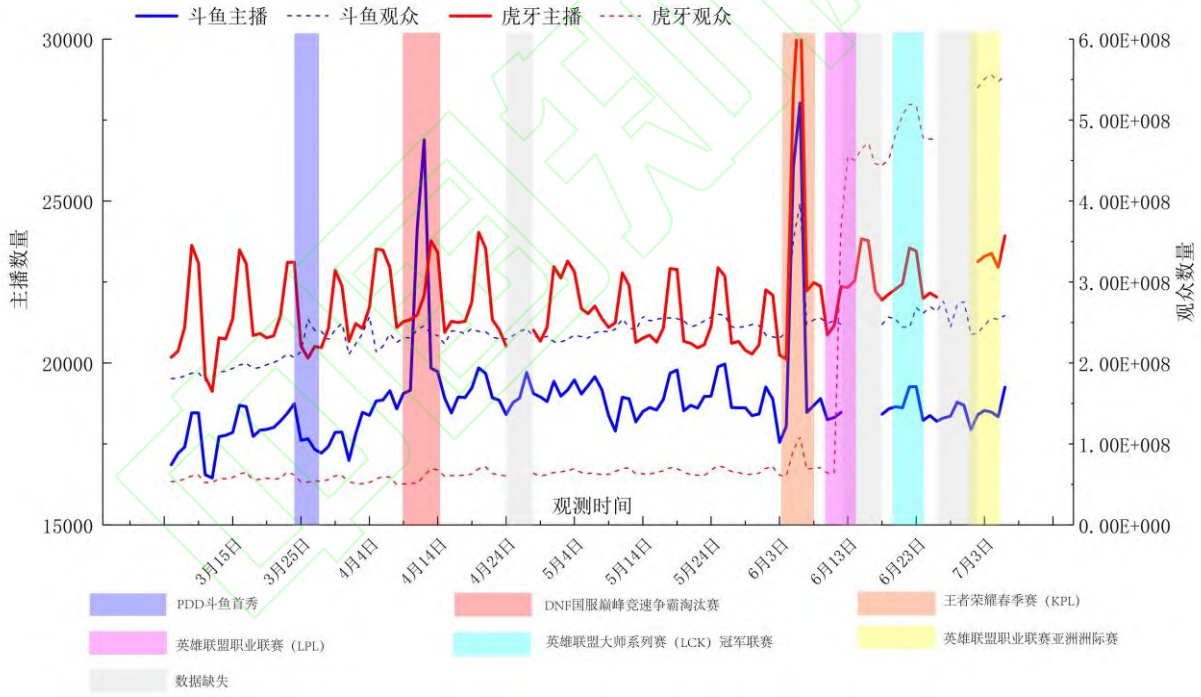


**图 3 直播平台主播和观众数量时序变化与直播大事件**

**Fig.3 Temporal pattern of the number of anchors and viewers; major live streaming events.**

虎牙平台的观众量级在统计后期达到 5-6 亿左右，虽然不能排除直播平台可能为吸引更多观众而展示远超实际数量的观众数，但观众数量变化曲线的多次跃升仍能体现观众规模增长的趋势。体现出直播行业发展前景向好，直播用户规模未来一段时间内仍将继续扩大。

经分析发现，直播平台负载突增基本都是由直播平台官方直播间或者网红直播间举办的、经过提前预告的直播大事件引起的（如 2019 年 3 月 25 日 PDD 斗鱼首秀造成的观众突增，4 月 14 日 DNF 国服巅峰竞速争霸淘汰赛吸引的众多主播开播）。一方面由于网红主播或游戏官方直播间的品牌效应能够吸引大量观众参与观看直播内容，另一方面由于观众提前通过预告了解了自己关注的直播内容的开播时间，所以在相应时间大量观众同时访问直播间，造成直播平台负载突增。

## 4 主播直播规律

本节通过分析主播生存期[29]的统计分布，总结了两种典型的直播模式——短暂出现和重复出现[30]，并讨论了不同直播模式的特征。

### 4.1 主播生存期分布

生存期为主播第一次直播到最后一次直播之间的时间差。它表现了直播平台客户的粘性，即直播平台吸引以及留住主播的能力。图 4（a）展示了直播平台主播生存期的概率分布，横轴代表生存期，纵轴代表相应生存期的主播占平台全部主播的比例，统计期全长 123 天。拟合结果显示直播平台主播生存周期的概率分布均符合幂律分布，用公式表示为

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

其中，斗鱼和虎牙平台参数 $\alpha$ 的取值分别为 1.23 和 1.36。直播平台主播生存期分布体现出明显的重尾效应[31]，即大量生存期非常短的主播和少量的生存期较长的主播并存。虽然大量主播生存期都非常短，但是小部分长生存期的主播对主播生存期的均值和方差起决定性作用。对比斗鱼和虎牙平台主播生存期分布形式，虎牙平台的幂律参数 $\alpha$ 值略大，说明主播生存期异质性更强。分析原因是虎牙平台中有更大比例的短生存期主播，长生存期主播比例更低，观众更集中于少数直播间，整个平台的主播生存期分布的异质性更强。

### 4.2 直播模式及特征

从主播生存期的分布特点出发，本节定义“短暂出现”为直播天数在 15%统计期以下，“长期直播”则是直播天数在 85%统计期以上。对直播平台两类直播模式的粉丝数量、观众数量、直播时长、直播间隔等主播特征统计量的均值进行双样本双边 T 独立性检验，结果表明 T 检验显著性概率（$p$ 值）均小于 0.01，即两类直播模式的主播在观众青睐、直播内容和规律性等方面均存在显著差异（见表 2）。

1. 观众青睐指标。短暂出现的主播粉丝数量、在线观众数量等观众青睐的表现都明显低于长期直播的主播，且长期直播的观众数量最大值远远超过观众数量均值，说明长期直播的主播能吸引平时几倍的流量，是有直播亮点的主播。

2. 直播类型分析。鉴于出现的天数较短，短期出现的直播类型相对固定，两个平台的主播直播类型数量分别是 1.2 和 1.3 种，而长期直播的主播直播类别分别在 1.8 和 3.7。由此来看斗鱼平台的主播直播类型比虎牙平台更固定。

3. 直播时长规律。短暂出现主播的直播时长均值仅有 1 小时左右，而长期直播的主播则在 2 小时以上，短暂出现的生存期和直播间隔更短。短暂直播主播出于新鲜感等原因尝试直播，频繁且短暂地直播了几次之后就退出了直播平台。

**表 2 直播平台不同直播模式统计特征**

**Tab.2 Statistical characteristics of different live streaming modes.**

| | 斗鱼平台 | | | 虎牙平台 | | |
|---|---|---|---|---|---|---|
| | 短期 | 长期 | *P* 值 | 短期 | 长期 | *P* 值 |
| 粉丝最大值 | 218 | 32370 | - | 0 | 20 | - |
| 粉丝均值 | 211 | 31041 | <0.01 | 0 | 14 | <0.01 |
| 观众最大值 | 835 | 89344 | - | 360 | 15658 | - |
| 观众均值 | 545 | 24958 | <0.01 | 236 | 4974 | <0.01 |
| 直播类别数 | 1.2 | 1.8 | - | 1.3 | 3.7 | - |
| 直播时长 | 74 | 299 | <0.01 | 40 | 161 | <0.01 |
| 直播间隔 | 364 | 620 | <0.01 | 285 | 477 | <0.01 |

注：虎牙平台的粉丝数量以主播的观众推荐数替代。

由于观众更青睐经常直播的主播[14]、互动交流的主播[32]以及曾经观看过的主播[10]，而短暂出现主播的直播天数短、直播种类少、直播时长短而且直播间隔长，导致观众不能准确把握直播信息，不能及时关注短暂出现的主播，所以粉丝和观众数量少于长期直播的主播。这种粉丝量少、观众少造成打赏和收入也相应少的局面也使该类主播无法长期坚持直播，造成生存期短的结果。

## 5 观众分布规律

本节通过统计分析观众总量变化规律、在众多直播间中的数量分布形式及其时序变化情况，挖掘观众在直播平台内的分散状态和变化规律，分析观众对直播内容或主播习惯的偏好特征，进而可以为观众进行个性化直播推荐，同时对主播的直播内容进行引导。

### 5.1 观众与主播数量变化关系

本节从观众与主播的数量关系入手对直播平台观众总量变化规律进行探索。从图 4（b）统计期各个时刻在线主播与直播平台观众总量变化的关系中可以看出，直播平台的在线主播数量和直播平台观众数量之间存在较强的正相关性，即在线主播数量增加，观众数量也相应增加，反之同理，体现了主播与观众的“共生”关系。另外，随着平台发展，直播平台内的主播和观众数量均逐渐增加，主播和观众数量呈现较强的线性相关性，但其波动性也逐步增大，体现出系统越来越强的异质性和非均匀性。对主播数量和观众数量的线性拟合关系参数值如表 3 所示。

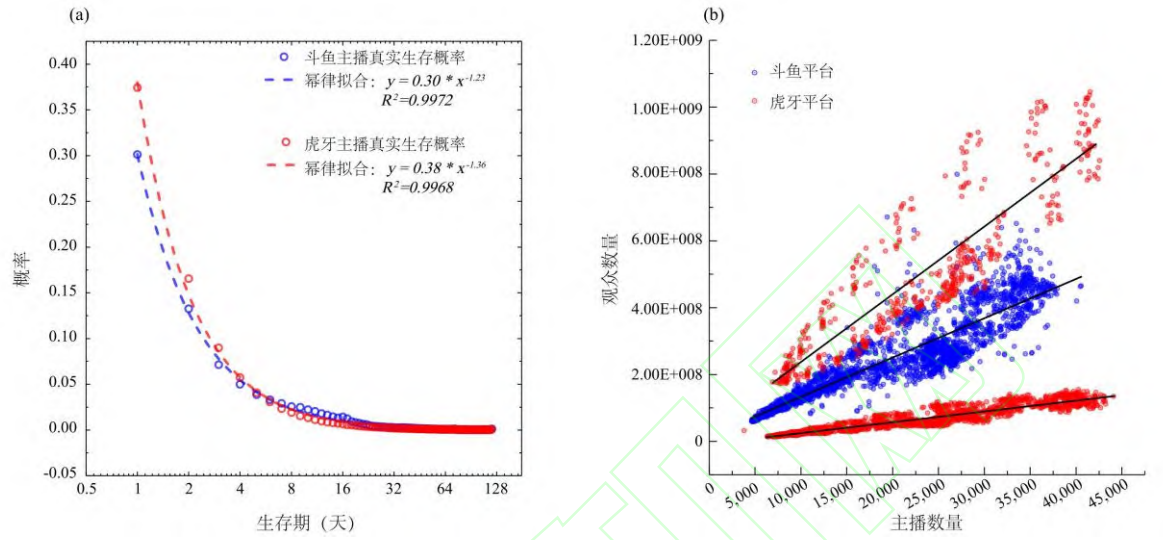


**图 4 (a) 直播平台主播生存期概率分布图 (b) 主播数量与观众数量关系图**

**Fig.4 (a) Probability distribution of anchor lifetime. (b) The relationship between the number of anchors and the number of viewers.**

虎牙平台主播、观众数量关系的分段形式显示出主播数量与观众数量的正相关关系，但由于日内效应等因素的存在，在观众数量超过 1.7 亿之后，主播-观众数量关系的斜率出现了明显抬升。说明直播平台内观众总量达到一定水平之后，主播数量变动对观众数量变动的影响效果会比观众总量较少时更强。

**表 3 观众与主播数量线性关系参数表**

**Tab.3 Parameter table of linear relationship between the number of viewers and anchors.**

| | 斗鱼平台 | 虎牙平台 | |
|---|---|---|---|
| 斜率($a$) | 11769 | 20390 | 3215 |
| 截距($b$) | 15,155,500 | 27,637,100 | -7,572,070 |
| 皮尔逊相关系数 | 0.90 | 0.89 | 0.93 |
| $R^2$ | 0.82 | 0.79 | 0.87 |

注：虎牙平台的线性关系式在观众数量大于 1.7 亿和小于 1.7 亿时分段，参数拟合值分别在第 2 和第 3 列。

## 5.2 直播平台内观众总体分布

由于直播平台存在成千上万直播间，观众在开播直播间的数量分布展现了直播间的吸引力差异。大量研究[8,10-13,18-20,28,30]发现，直播平台观众分布均基本符合幂律形式。其中，Karine 等[8,12]对国外直播平台 Twitch 的观众数量分布研究发现，分布形式近似为齐普夫分布：

$$x \sim r^{-\beta} \tag{2}$$

其中，参数 $\beta$ 的取值在 1.3 至 1.6 之间。

在分析拟合了直播平台内主播数量与观众数量的正向相关变化关系之后，本节对直播平台所有开播直播间吸引观众数量在某一时刻的分布形式进行探究。以 2019 年 5 月 1 日 22 时的观众数量分布为例（如图 5（a）所示），直播平台的观众数量分布均符合指数截断的幂律分布[30]。即观众分布在主播排名靠前的范围呈幂律分布，以齐普夫分布形式展示为：

$$y = cx^{\beta} \tag{3}$$

但是在分布的尾部出现了明显的下降，与指数形式高度相符，公式形如：

$$y = ae^{(-\frac{x}{t})} + y_0 \tag{4}$$

其中分布形式及参数取值见表 4。

**表 4 直播平台观众分布形式拟合结果**

**Tab.4 Fitting results of distributions of the number of viewers.**

| | | 斗鱼平台 | 虎牙平台 |
|---|---|---|---|
| 齐普夫分布 | $x$ 范围 | [100,10000] | [10,10000] |
| | 参数取值 | $c$=3.44×10$^7$ | $c$=1.21×10$^7$ |
| | | $\beta$=-0.93 | $\beta$=-0.99 |
| | $R^2$ | 0.99 | 0.99 |
| 指数分布 | $x$ 范围 | [10000,34005] | [18000,37731] |
| | 参数取值 | $a$=18343.77 | $a$=-165.98 |
| | | $t$=11872.69 | $t$=-31369.56 |
| | | $y_0$=-1477.63 | $y_0$=585.78 |
| | $R^2$ | 0.99 | 0.97 |

从指数截断幂律的分布形式可以看出，观众分布表现出很强的非均匀性。某小部分直播间吸引了绝大部分的观众，而尾部大量直播间则仅有极小部分观众观看。超强异质性的分布规律导致了直播平台内少数主播成为网红主播，对观众的吸引力和号召力比普通主播更强，印证了网红主播开播或官方直播间举办活动导致直播负载大幅度增加的合理性。

观众分布形式中齐普夫分布的参数$\beta$（公式 3）在 0.9-1 之间，即幂律分布的参数$\alpha$（公式 1）在 2-2.1 左右，与诸多已发现人类社会的幂律分布如性伙伴数量分布[33]、演员合作度分布[34]、文献引用度分布[35]、财富分布[36]等相比，观众分布的非均匀性稍弱，即观众在排名靠前的主播房间中的集中性没有上述财富分布等在头部的集中性强。排名较后的主播（在 10000 名之后），观众分布近似指数分布，与已有的对直播平台观众分布[30]、对观看请求次数分布[37]的研究结果类似，排名较后的主播对观众的吸引能力太弱，造成了观众分布形式在排名靠后的部分出现了突然下降的指数形式。

### 5.3 观众分布均匀度时序变化

不同时刻观众数量的幂律分布$\alpha$值的变化情况可以表现不同时刻观众分布的时序特征，$\alpha$值越大代表观众分布越不均匀。本节对直播平台观众在一天之中的分布均匀程度进行比较（如图 5（b）所示）。拟合结果显示斗鱼和虎牙平台观众数量分布的幂律参数在 1.9-2.4 之间变动，与国外直播平台 Twitch 的观众幂律分布参数在 1.3-1.6 之间[12]相比，观众分布的均匀性更弱。结合直播平台负载日内效应发现，直播负载水平高的时段（如 18 时至 1 时）$\alpha$值较低，观众分布的均匀性较强；直播负载水平低的时段（如 2 时至 6 时）$\alpha$值较高，观众分布的均匀性较弱。直播平台负载水平高的时段，大部分观众选择观看直播作为娱乐方式，以随机的方式而不是专门为某几个主播而观看直播；但在直播平台负载水平低的时段，即 2 时至 6 时的深夜时段，仍留在直播平台的观众更可能是为了喜爱的主播而逗留，导致观众整体分布的不均。

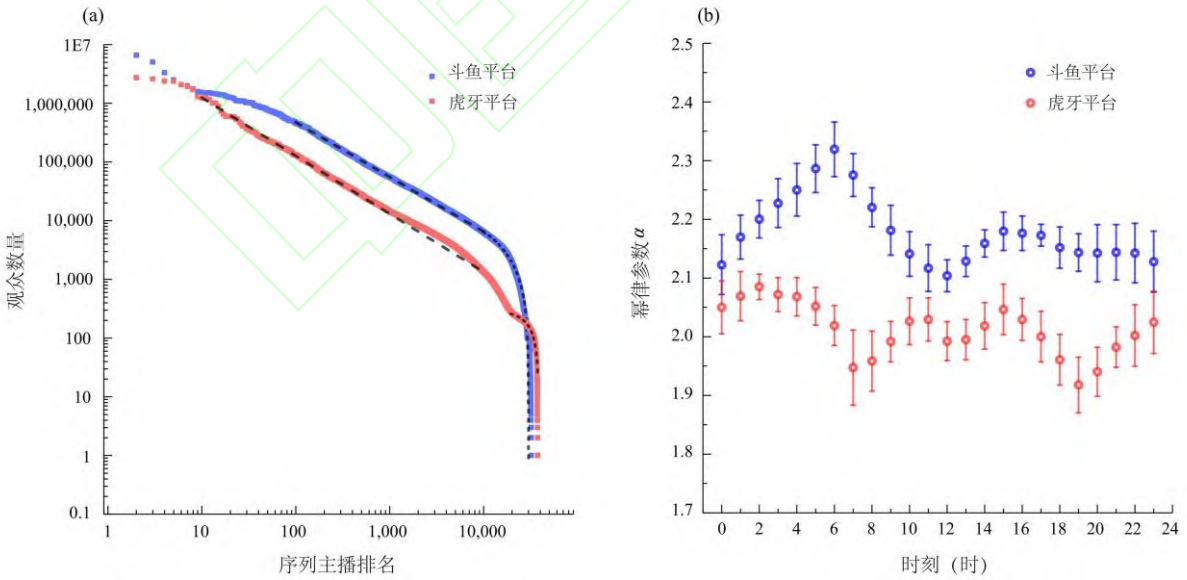


**图 5 (a) 直播平台内观众分布 (b) 直播平台内观众分布均匀程度日内变化**

**Fig.5 (a) Viewer distribution in the live streaming platform. (b) Daily variation of the evenness of viewer distribution in the live streaming platform.**

由于不同平台的直播类型、主播和观众规模不完全一致，斗鱼和虎牙平台的观众分布均匀程度略有差异。斗鱼平台的观众规模比虎牙平台的更高，观众分布$\alpha$值在每个时刻都比虎牙平台的更大，即斗鱼平台的观众分布均匀度在全天都比虎牙平台更低。斗鱼平台的观众分布$\alpha$值在 6 时呈现一个明显的波峰，12 时呈现波谷，其余时刻差异较小；虎牙平台的$\alpha$值的变化相对平缓，全天呈现波浪状的变化趋势，在 2 时

取得最大值，19 时取得最小值。

总体来看，斗鱼平台的观众分布异质性更强且观众分布均匀性受时序变化的影响更大。这种规律与诸多社会系统如城市人口数量[38]、个人收入[39]等分布形式的演化规律一致，随着直播系统的观众规模增加，观众对大型直播间的偏好性增强，观众更加集中在少数几个直播间中，少量超大型直播间（类似于超级大城市、超级富豪）逐渐形成且核心地位越来越明显。

## 6 总结与展望

网络直播作为新兴社交媒体逐渐成为人们学习、生活、娱乐的重要方式，数百万观众同时在线观看直播的情形时有发生。本文首次对国内直播平台的大规模流量数据进行特征挖掘，发现直播平台时序负载存在显著的日内效应和周内效应，主播生存期和观众数量分布呈现幂律乃至指数等极端不均匀的分布形式，主播数量与观众数量正向变化，但主播及观众数量越大时，观众分布越陡峭。研究结果对理解网络直播复杂系统中的用户行为模式、挖掘用户分布规律及变化趋势、设计以直播用户行为动力学特征为基础的商业模式如个性化推荐等方面具有重要的理论和实践意义。

由于网络直播以及相关研究发展的时间尚且较短，对网络直播流量特征的分析和应用仍有待进一步探索，直播平台中各种社群网络的形成和演化机制、进一步优化直播平台等方面的研究是网络直播领域研究的未来发展趋势。分析社群网络组成和演化，探究直播平台的观众流动、规模演化等内在机制，并根据已有的平台负载、观众分布、社群演化等规律和模型进行网络直播平台的特征分析和建模，优化平台性能，加强 5G 技术、虚拟现实等在网络直播领域的研究与应用等。考虑到直播平台目前存在的各种乱象和法律问题，加强制定针对网络直播行业的法律法规，确保网络直播内容在合法的基础上更健康、更有益。

## 参考文献：


[1] CHEN X, CHEN S, WANG X, et al. " I was afraid, but now I enjoy being a streamer!" Understanding the challenges and prospects of using live streaming for online education[J]. Proceedings of the ACM on Human-Computer Interaction, 2021, 4(CSCW3): 1-32.

[2] LIU L, AREMU E O, YOO D. Brand marketing strategy of live streaming in mobile era: a case study of Tmall platform[J]. Journal of East Asia Management, 2020, 1(1): 65-87.

[3] LU Z, ANNETT M, FAN M, et al. " I feel it is my responsibility to stream" Streaming and engaging with intangible cultural heritage through livestreaming[C]// BREWSTER S, FITZPATRICK G, COX A, et al. Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems. Scotland, UK: ACM, 2019: 1-14.

[4] FAN H, LEE F L F. Judicial visibility under responsive authoritarianism: a study of the live broadcasting of court trials in China[J]. Media, Culture & Society, 2019, 41(8): 1088-1106.

[5] 中国互联网网络信息中心. 第 49 次中国互联网络发展状况统计报告[EB/OL]. (2021.07.01) [2022.07.04]. http://www.cnnic.net.cn/hlwfzyj/hlwxzbg/hlwtjbg/202202/P020220311493378715650.pdf.

CENTER C I N I. The 49th Statistical Report on Internet Development in China[EB/OL]. (2021.07.01) [2022.07.04]. http://www.cnnic.net.cn/hlwfzyj/hlwxzbg/hlwtjbg/202202/P020220311493378715650.pdf.

[6] 艾媒咨询. 2021Q3 中国在线直播行业研究报告[EB/OL]. (2021.11.09) [2022.07.04]. https://www.iimedia.cn/c400/81868.html.

CONSULTING I R. 2021Q3 China online live streaming industry research report[EB/OL]. (2021.11.09) [2022.07.04]. https://www.iimedia.cn/c400/81868.html.

[7] CLAYPOOL M, FARRINGTON D, MUESCH N. Measurement-based analysis of the video characteristics of Twitch. tv[C]// BERRY J, BERTOZZI E, FIELLIN L, et al. 2015 IEEE Games Entertainment Media Conference (GEM). Toronto, Canada: IEEE, 2015: 1-4.

[8] PIRES K, SIMON G. YouTube live and Twitch: a tour of user-generated live streaming systems[C]// OOI W T, FENG W-C, LIU F. Proceedings of the 6th ACM multimedia systems conference. Oregon, USA: ACM, 2015: 225-230.

[9] ZHU Z H, YANG Z, DAI Y F. Understanding the gift-sending interaction on live-streaming video websites[C]// MEISELWITZ G. International Conference on Social Computing and Social Media. Vancouver, Canada: Springer, 2017: 274-285.

[10] NASCIMENTO G, RIBEIRO M, CERF L, et al. Modeling and analyzing the video game live-streaming community[C]// BAEZA-YATES R. 2014 9th Latin American Web Congress. Minas Gerais, Brazil: IEEE, 2014: 1-9.

[11] ZHAO J, MA M, GONG W, et al. Social media stickiness in mobile personal livestreaming service[C]// LAB C. 2017 IEEE/ACM 25th International Symposium on Quality of Service (IWQoS). Vilanova i la Geltrú, Spain: IEEE, 2017: 1-2.

[12] PIRES K, SIMON G. Dash in twitch: Adaptive bitrate streaming in live game streaming platforms[C]// HASSAN M, BEGEN A C, TIMMERER C. Proceedings of the 2014 Workshop on Design, Quality and Deployment of Adaptive Video Streaming. Sydney, Australia: ACM, 2014: 13-18.

[13] ZHANG C, LIU J. On crowdsourced interactive live streaming: a twitch. tv-based measurement study[C]// FENG W-C, ZINK M. Proceedings of the 25th ACM Workshop on Network and Operating Systems Support for Digital Audio and Video. Oregon, USA: ACM, 2015: 55-60.

[14] HAMILTON W A, GARRETSON O, KERNE A. Streaming on twitch: fostering participatory communities of play within live mixed media[C]// JONES M, PALANQUE P, SCHMIDT A, et al. Proceedings of the 32nd annual ACM conference on Human factors in computing systems. Toronto, Canada: ACM, 2014: 1315-1324.

[15] LYKOUSAS N, GóMEZ V, PATSAKIS C. Adult content in social live streaming services: characterizing deviant users and

relationships[C]// BRANDES U, REDDY C, TAGARELLI A. 2018 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM). Barcelona, Spain: IEEE, 2018: 375-382.
[16] 郭淑慧, 吕欣. 网络直播平台数据挖掘与行为分析综述[J].物理学报, 2020, v.69(08): 117-126.
GUO S, LU X. Live streaming: Data mining and behavior analysis[J].Acta Physica Sinica, 2020, v.69(08): 117-126.
[17] BORGES A, GOMES P, NACIF J, et al. Characterizing sopcast client behavior[J]. Computer Communications, 2012, 35(8): 1004-1016.
[18] VELOSO E, ALMEIDA V, MEIRA W, et al. A hierarchical characterization of a live streaming media workload[C]// KüHLEWIND M, KUTSCHER D. Proceedings of the 2nd ACM SIGCOMM Workshop on Internet measurment. Marseille France: ACM, 2002: 117-130.
[19] DENG J, CUADRADO F, TYSON G, et al. Behind the game: exploring the twitch streaming platform[C]// NETGAMES. 2015 International Workshop on Network and Systems Support for Games (NetGames). Zagreb, Croatia: IEEE, 2015: 1-6.
[20] JIA A L, SHEN S, EPEMA D H, et al. When game becomes life: the creators and spectators of online game replays and live streaming[J]. ACM Transactions on Multimedia Computing, Communications, and Applications (TOMM), 2016, 12(4): 47.
[21] FALLICA B, LU Y, KUIPERS F, et al. On the quality of experience of SopCast[C]// AT-BEGAIN K, CUEVAS A. 2008 The Second International Conference on Next Generation Mobile Applications, Services, and Technologies. Cardiff, Hnited Kingdom: IEEE, 2008: 501-506.
[22] 中国信息通信研究院. 2018 下半年中国网络直播行业景气指数及短视频报告[EB/OL]. (2019.07.11) [2022.07.04]. http://www.caict.ac.cn/kxyj/qwfb/ztbg/201907/P020190711347399467992.pdf.
TECHNOLOGY C A O I A C. China's online live streaming industry prosperity index and short video report in the second half of 2018[EB/OL]. (2019.07.11) [2022.07.04]. http://www.caict.ac.cn/kxyj/qwfb/ztbg/201907/P020190711347399467992.pdf.
[23] KHADEMALOMOOM S, NARAYAN P K. Intraday effects of the currency market[J]. Journal of International Financial Markets, Institutions and Money, 2019, 58(1): 65-77.
[24] PINK D H. When: The scientific secrets of perfect timing[M]. New York: Penguin Press, 2019: 15-20.
[25] HINES C B. Time-of-Day Effects on Human Performance[J]. Journal of Catholic Education, 2004, 7(3): 390-413.
[26] BERNARD T, GIACOMONI M, GAVARRY O, et al. Time-of-day effects in maximal anaerobic leg exercise[J]. European Journal of Applied Physiology and Occupational Physiology, 1997, 77(1-2): 133-138.
[27] MüLLER U A, DACOROGNA M M, OLSEN R B, et al. Statistical study of foreign exchange rates, empirical evidence of a price change scaling law, and intraday analysis[J]. Journal of Banking & Finance, 1990, 14(6): 1189-1208.
[28] STOHR D, LI T, WILK S, et al. An analysis of the YouNow live streaming platform[C]// KANHERE S, TöLLE J, CHERKAOUI S. 2015 IEEE 40th Local Computer Networks Conference Workshops (LCN Workshops). Florida, USA: IEEE, 2015: 673-679.
[29] GUPTA S, HANSSENS D, HARDIE B, et al. Modeling customer lifetime value[J]. Journal of service research, 2006, 9(2): 139-155.
[30] SRIPANIDKULCHAI K, MAGGS B, ZHANG H. An analysis of live streaming workloads on the internet[C]// LOMBARDO A, KUROSE J. Proceedings of the 4th ACM SIGCOMM conference on Internet measurement. Sicily, Italy: ACM, 2004: 41-54.
[31] 樊超, 郭进利, 韩筱璞, et al. 人类行为动力学研究综述[J].复杂系统与复杂性科学, 2011, 8(02): 1-17.
FAN C, GUO J, HAN X, et al. A review of research on human dynamics[J].Complex Systems and Complexity Science, 2011, 8(02): 1-17.
[32] 李爽, 陈亚荣. 网络直播环境下人际互动对用户行为意愿的影响研究[J].中国市场, 2018, 1(7): 18-20.
LI S, CHEN Y. Research on the influence of interpersonal interaction on user behavior intention in the environment of online live streaming[J].China Market, 2018, 1(7): 18-20.
[33] LILJEROS F, EDLING C R, AMARAL L A, et al. The web of human sexual contacts[J]. Nature, 2001, 411(6840): 907-8.
[34] BARABASI, ALBERT. Emergence of scaling in random networks[J]. Science (New York, N.Y.), 1999, 286(5439): 509-12.
[35] REDNER, S. How popular is your paper? An empirical study of the citation distribution[J]4(2): 131-134.
[36] REPETOWICZ P, HUTZLER S, RICHMOND P. Dynamics of money and income distributions[J]. Physica A: Statistical Mechanics and its Applications, 2005, 356(2-4): 641-654.
[37] ALMEIDA J M, KRUEGER J, EAGER D L, et al. Analysis of educational media server workloads[C]// NIEH J. Proceedings of the 11th international workshop on Network and operating systems support for digital audio and video. New York, USA: ACM, 2001: 21-30.
[38] DA SILVA D F C, NETO R D M S. Population Dynamics and Spatial Dependence: Evidence from Brazilian Cities[J]. Review of Regional Studies, 2019, 49(3): 454-473.
[39] GUO Q, GAO L. Distribution of individual incomes in China between 1992 and 2009[J]. Physica A: Statistical Mechanics and its Applications, 2012, 391(21): 5139-5145.