

# 社交网络中的“偶像趋同性”现象及其趋同后果

## ——基于新浪微博样本的实证分析

徐 翔

**摘 要:** 本文聚焦社交网络中的“偶像趋同性”现象,并以新浪微博样本实证检视“偶像趋同性”现象与其所产生的用户趋同结构。基于潜在语义分析(LSA),进行用户内容特征的提取与用户相似度计算。在对于“偶像趋同性”基本现象进行分析的基础上,分析在“偶像趋同性”的作用下产生的微博用户单极化、收窄化、大众化的趋同演变,以及这些路径的内在同步性而非独立性或冲突性。微博用户的这些趋同演化趋势,既是内嵌于“偶像趋同性”作用逻辑的有机组成,也是需要深入挖掘的网络信息社会后果。

**关键词:** 偶像趋同性; 微博; 用户内容生产; 内容相似度; 潜在语义分析

**中图分类号:** G206.3 **文献标识码:** A **文章编号:** 1671-0169(2021)04-0133-14

社交网络是用户生产内容(User Generated Content, UGC)的自主平台和自媒体空间,有着各种各样的“圈粉”大咖、大V、“头部用户”和“流量担当”。这些有着高影响力、高粉丝规模、高关注度的用户可称为社交网络中不同程度的“偶像”用户。从最原始的层面,“偶像-追随者/崇拜者”是基本的对应关系,由于其追随者的数量和规模的不同,产生了社交网络偶像在程度上的高低差异。由之而产生的是,任一用户都可以表示出一个或低或高的“偶像”程度,而不仅仅是那些高圈粉量的用户。本文的核心内容,是探讨用户的“偶像”程度与他们之间趋同性的关系。明确提出用户可能存在、易被忽视的“偶像趋同性”现象;阐释和检验由用户的“偶像趋同性”对社交网络产生的趋同后果与连贯效应,及这些不同路径之间的同步性和一致性。

### 一、文献回顾与问题提出

本文以新浪微博为社交网络对象,考察其中的“偶像趋同性”现象。社交网络中的偶像用户千差万别,各个细分领域的意见领袖、内容精英“你方唱罢我登台”,各种大、中、小“咖位”的流量用户百花齐放,容易造成一种错觉,即认为社交网络中的偶像用户是多种多样的。关于网络文化和 Web2.0 不同于传统媒介的去中心化及其多样化话语主体,常成为社交网络偶像多样化的理论依据。例如波斯特对于互联网时代作为“双向的去中心化的交流”的“第二媒介时代”的阐述<sup>[1](P22-28)</sup>,吉摩尔认为自媒体“草根媒体”把“宣讲”变为“对话”并动摇主流媒体的信息垄断<sup>[2]</sup>。较多对意见领袖的调研认为,他们是分布在多种领域、内容主题分化和差异化的<sup>[3][4][5][6]</sup>。

**基金项目:** 国家自然科学基金项目“社交网络互动中用户‘信息窄化’机理分析:基于微博的数据挖掘”(71804126)

**作者简介:** 徐翔,博士,同济大学艺术与传媒学院教授, xuxiang@tongji.edu.cn (上海 201801)

这些理论观点在意见市场中拥有巨大拥趸,使得本文对于用户趋同的强调面临着现实必要性。一些研究涉及和分析了社交网络中用户内容同质化的形成机制,包括网络“回音室”、群体极化、网络“巴尔干化”等的分析,都指出网络中会形成同质化用户的聚集和圈层<sup>[7][8][9][10]</sup>。对网络“巴尔干化”的重要分析表明,信息技术构筑起基于学科专业、社会阶层、个人偏好、社会文化的分化和相互之间的壁垒<sup>[11]</sup>。这些社交网络用户的趋同化为本文的分析提供了有益基础,然而仍有巨大的挖掘空间:(1)对于多种社群、用户阶层和“壁垒”因素的实证分析表明,用户本身的影响力、“圈粉”流量、偶像地位等因素如何影响他们的同质化,尚待明确。(2)对于种种“局部化”用户的趋同研究是当前相关的实证研究的主要倾向之一,但是这些“局部”视野易于忽略整体中的用户趋同化是否发生、如何发生,或者要以整体分化为局部趋同的背景或代价。如果社交网络用户具有趋同化的现象,我们并不仅需要关注局部如何趋同化,还要进一步关注于用户在全局和整体上,是否以及如何体现其“趋同”效应及驱动逻辑?但这方面的研究仍然较为匮乏。

一些研究部分认可社交网络意见领袖和“人气用户”在类型上的偏倚,及其中隐含的集中和趋同。张玉晨等将新浪微博加V用户分为专家型、媒体型、政务型、娱乐型、机构型和其他类型六大类,指出媒体型占据最主要的比重<sup>[12]</sup>。张志安等对于微博的分析结果表明,营销类、娱乐类用户成为微博意见领袖群体的主流<sup>[13]</sup>。佟力强<sup>[14](P27-28)</sup>将新浪微博名人用户主要分为七类,并指出文化传媒类、商业财经类、文体明星类占据着最为主要的比重。但总体上,这些分析还是对于意见领袖的静态描述,而非对其动态化机制的分析:(1)即使在高粉丝规模的意见领袖中,也仍然存在着强弱程度的分化,这些程度差异是否以及如何体现到用户的窄化和偏倚中?(2)如果“头部”用户是具有偏倚性和趋同性的,那么这种趋同化如何形成?形成之后如何演变?如果把用户看成铁板一块的整体,就难以展开对于用户驱同逻辑线条的分析。对于社交网络用户而言,其中的重要动态条件是用户在影响力程度上的差异,以及这种条件差异所带来的趋同程度的结果差异。

社交网络中高影响力用户/偶像用户具有更强的内容扩散力,对其他用户会产生更强的同化作用。Matsumura等提出的影响力扩散模型(Influence Diffusion Model, IDM)<sup>[15]</sup>,反映了意见领袖在内容上向其他用户的扩散力和扩散深度。Borge等<sup>[16]</sup>分析了推特中的政治意见领袖发布的相关信息在推特中的扩散能力。高剂斌等提出在BA无标度网络上模拟个体观念演化驱动群体行为的模型,表明网络中的“超级节点”对达成观念共识和形成全局群体行为具有明显牵引作用<sup>[17]</sup>。李根强等运用计算实验法分析意见领袖对网络集群观点演化的影响作用,指出分散的群体观点最终会朝向意见领袖的观点逐渐偏移<sup>[18]</sup>。李卫东等发现个体的社会网络中心性与意见强化现象有明显的正相关关系<sup>[19]</sup>。Yoo等使用Twitter数据分析了意见领袖所产生的内容级联,具有较大网络的生产者贡献的内容的扩散更有可能被相似内容的扩散所放大,并对其他内容的扩散产生抑制<sup>[20]</sup>。这些显示着大范围的用户节点围绕某种具有强同化能力的用户发生趋同的可能,尤其是围绕高影响力的“偶像”用户发生的趋同。

用户为了获得高影响力而迎合受众、市场的期待与过滤,使得符合这种内容模式的用户更有可能实现高影响力或高偶像程度,从而影响力越高的用户和某种“顶部”用户的内容特征就越切近。用户的内容生产受到以召回率为目标的推荐算法影响,带来内容选题、类型和风格的系统化倾向,影响到内容生产者的旨趣,并促使内容生产变得愈加标准化<sup>[21]</sup>。Masson等建立的动态模型假定用户只会发布其认为可能非常受欢迎的内容,证明用户发帖的收敛性<sup>[22]</sup>。社交网络用户在发帖和分享时会受到“想象的受众”塑造作用,使得自身内容增强朝向某种内容特征的收敛<sup>[23]</sup>。Mikal等认为社交网站具有话语内容的某种潜在特质和标准,网民受到网站中沟通“标准”的修正,从而在发帖时相应地修改自己的内容而趋同于网站内容<sup>[24]</sup>。在这些网站中的内容生产行为不断调适,使得用户内容不是自主化和多样化,而是朝向受众和市场普遍选择的模式发生收敛与趋同;这种趋同程

度越高也意味着受到受众、注意力市场的接受程度也越高，从而产生更高的影响力或偶像程度。

总体而言，高影响力用户具有使得其他用户和自己趋同的更强的同化能力；高影响力用户由于符合某种共通的内容模式、受众期待，而产生多样性的消减、内容同质化的增强；媒介生产中其他用户朝向高影响力的用户发生内容的模仿、学习而提升自己的竞争力与影响力。这些都体现出了高影响力用户在用户趋同中的重要作用。对此，需要将这种同化作用的程度更进一步具体化，并寻求它和用户影响力程度之间的对应关系与共变机制。一方面，用户的偶像程度越高，则使得其他用户和自己趋同的效能越强；另一方面，其他用户对于“头部”用户、已成功获得高影响力的偶像用户的学习和趋近，也反过来使得自身的影响力越可能趋于增强。在这种作用条件和动态描述基础上，本文用偶像用户程度作为指标表示用户的影响力，明确地凝练和提出社交网络用户“偶像趋同性”的基本问题（代称 Q0）：用户作为“偶像”用户的程度高低，密切关系到他和高程度偶像用户似同化的程度；由之，是否存在如下具有规律性的现象——高程度偶像用户和其他任意某用户之间的内容相似程度，和这两个用户的偶像程度高低同时具有正相关？

对于可能存在的“偶像趋同性”现象的基本界定和操作化的形式如下（代称 P0）：对于社交网络中某用户  $x$  而言，其他任意某用户设为  $y$ ，两者的用户内容相似程度设为  $sim(x, y)$ ，这两个用户各自作为偶像用户的程度分别设为  $i_x$ 、 $i_y$ ；当  $y$  为不同于  $x$  的任意用户的情况下，如果用户的“偶像趋同性”程度越强，则  $sim(x, y)$  与  $(i_x + i_y)$  之间的正相关程度就越高。

由 P0 和前文的理论分析所延展的重要假设 H0 是：用户的“偶像趋同性”现象和强度不是随机分布的，而是与用户的偶像程度具有正相关。换言之，高偶像程度的用户，具有比较稳定、强烈、可预期而非若有若无的“偶像趋同性”现象。基于 H0 这一基本假设，对于后文的偶像用户趋同分析提供了重要准备。

针对 P0 的基本现象，以典型社交网络之一的新浪微博为案例对象，展开后续实证考察和机理分析，全文的推进步骤按照任务 1、任务 2、任务 3 展开，依次如下。

任务 1。这是具有基础性的一环，结合用户作为偶像用户的影响力维度（见后文 C0、C2），以及用户之间在内容上的相似程度（C1），提出用户“偶像趋同性”的现象与内涵（P0），测度用户的偶像趋同性现象与程度（N1），并指出越是高偶像程度的用户其“偶像趋同性”就越强烈（H0）。

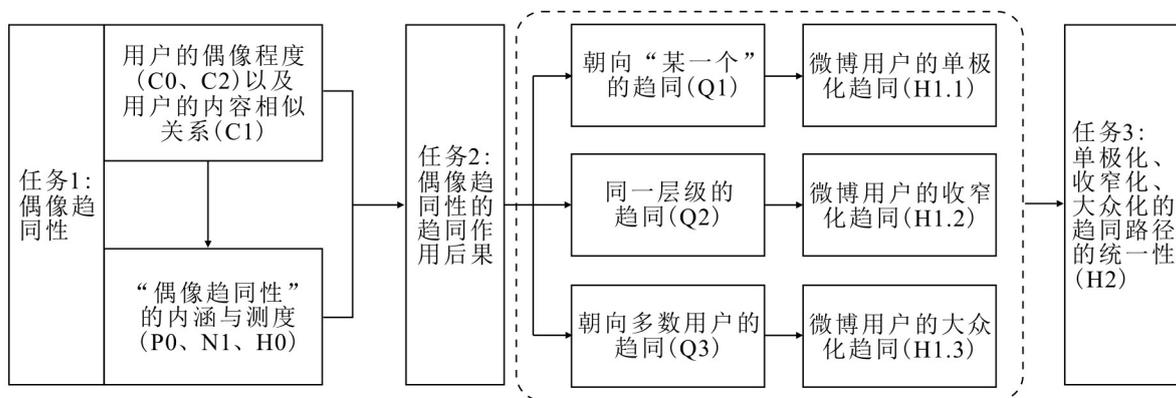


图 1 研究内容与逻辑关系

任务 2。在“偶像趋同性”的内涵、现象与机制的基础上，探讨和检验其蕴藏的社交网络用户单极化、收窄化、大众化的趋同作用后果（分别对应于假设 H1.1、假设 H1.2 和假设 H1.3），它

们是和“偶像趋同性”紧密相关的组成层面。其中,包括一部分不大符合经验,但却包含在偶像趋同性的延伸范围内的可以经受检验的现象。

任务3。对于任务2中涉及的三种后果,探讨这三者的共变是统一的亦或相互对立的、或存在矛盾和冲突?这部分对应于假设H2,同样容易产生一些错误的“观感”。任务3构成对于偶像趋同性所产生的社会后果与微博社会特征的分析。

全文围绕任务1、任务2、任务3,及其紧密关联的H1.1、H1.2、H1.3、H2,从理论分析和实证检验方面系统展开。

## 二、用户“偶像趋同性”的内涵以及其他相关指标的界定

由P0可以给出每个用户的“偶像趋同性”强烈程度的衡量方法(N1):用户 $x$ 分别和 $m$ 个其他用户的内容相似度记为数组 $S_{x1} = [s_1, s_2, s_3, \dots, s_m]$ ;  $x$ 自身作为偶像用户的影响力程度(内涵参见C0,测量方法参见后文C2)记为 $i_x$ ,  $x$ 依次和其他 $m$ 个用户的影响力程度之和记为数组 $I_{x1} = [i_x + i_1, i_x + i_2, i_x + i_3, \dots, i_x + i_m]$ ;则显然,如果 $x$ 所表现出的“偶像趋同性”越是鲜明和强烈,则 $S_{x1}$ 和 $I_{x1}$ 的皮尔逊相关系数也就越高。这两组数之间的皮尔逊相关系数记为 $p_{x1}$ ,表示的含义是:若 $x$ 和其他任意一个用户的偶像程度之和越大;则 $x$ 和该用户的内容相似度也就相应地越高,反之亦然。

由相关系数的定义可知,任意用户的“偶像趋同性”强度 $p_{x1}$ 取值在 $[-1, +1]$ 之间。实际计算过程中,当 $p_{x1}$ 不符合相关系数的显著性要求时,把 $p_{x1}$ 替换为0,表明无相关关系。当 $p_{x1}$ 等于1并且显著时,此时完全可以由 $x$ 和其他任意用户各自的偶像程度,精确地预测两者之间的内容相似度。反之,若 $p_{x1}$ 取值小于或等于0且显著时,则意味着用户 $x$ 不存在P0中所述的偶像趋同性的现象,甚至会发生与P0相反的现象。

对于个体 $x$ 层面的偶像趋同性,如果扩展到宏观和整体也成立,此时对于任意一个个体 $x$ ,其和任意一个其他个体 $y$ ,相互之间存在着“两者内容相似度与两者偶像程度之和具有正相关”的现象。个体层面的“偶像趋同性”,成为关系到全局“偶像趋同性”得以成立的微观基础。用户个体的偶像趋同性程度越高,则该个体和全局任一用户的趋同力就越强烈和鲜明、可预测。如果这些强“偶像趋同性”的个体越多,则整体越有可能表现出任意两个用户之间普遍存在的正向的趋同性。从结果上会导致高影响力的“偶像”用户长得越来越像,而不是由于垂直领域、细分受众等原因长得越来越差异化。

用户趋同度/用户内容相似度(C1)。本研究中,对于用户趋同的现象和程度,基于用户内容相似度进行衡量。实际操作中,把某个用户发布的 $n$ 条帖子作为一个整体,考察其内容特征,进而以此分析用户之间的内容相似、内容趋同的程度。选择向量空间模型(Vector Space Model, VSM) + 潜在语义分析(Latent Semantic Analysis, LSA),把每个用户根据其帖子内容转化为一个若干维的向量。这个向量不仅作为用户的嵌入式表示,也很方便在用户之间进行向量的相似度计算。

社交网络偶像:参见前文C0。

用户的偶像程度/偶像度(C2)的衡量。本文中的“偶像”用户是广义上的,不是某个固定的、铁板一块的人群,而是具有各种不同“偶像程度”,既包括各种“大V”、“中V”,也包括零粉丝的程度为0的“偶像”。采用“追随者”/“粉丝”的规模,反映一个用户的偶像程度/偶像度。微博中,粉丝规模是一个偶像的追随者的最为直接而原始的体现。对于微博意见领袖、社交网络“大V”的研究,突出“粉丝规模”这个指标反映用户影响力和微博话语地位的重要性<sup>[25]</sup>。实际计算中,由于微博用户的粉丝规模悬殊太大,从寥寥几个到数千万不等,因此采取对数转换后的形

式， $x$  为用户粉丝规模的原始值，则偶像程度  $x_{new} = \log_2(x+1)$ 。

### 三、从“偶像趋同性”到微博用户的单极化、收窄化、大众化

“偶像趋同性”现象对于微博会带来什么样的影响与后果？总的说来：越是高影响力的用户，就越是发生多种形态的趋同、相似，而不是容易误认为的细分化、差异化与多样化。

但是这种用户“趋同”有着多种严格的路径及条件，而不能笼统地用“同质化”来概括。就此分别称之为：单极化、收窄化、大众化。其涉及的层面由窄到宽，而指向越来越泛化的趋同。由偶像用户的影响力程度、偶像趋同性和用户趋同现象，进一步的具体问题如下。问题 Q1：用户随着影响力程度的提升，是否和“某一个”用户越来越相似，从而产生单极化的趋同？问题 Q2：用户随着影响力程度的提升，是否在影响力的“同一个层级”内的用户彼此之间越来越相似，从而产生由低到高层级的收窄化的趋同？问题 Q3：用户随着影响力程度的提升，是否和多数的大众、“芸芸众生”用户越来越相似，从而产生大众化的趋同？这几种趋同的态势和路径，和一般的经验直观也有着诸多相悖之处，因此同样需要结合实证释疑。

#### （一）微博用户随着影响力程度增加而表现出的单极化

这里的“单极化”指的是所有的用户不仅随着偶像程度的增加而长得越来越像，更为重要的是：随着偶像程度/影响力程度越高，则用户长得越来越像“同一个模子”里出来的。

为什么如此推测？根据前文的“偶像趋同性”现象和效应、测量方式（P0、N1），对于微博中“偶像趋同性”程度最高（且该程度为正数的情况下）的那个用户（代称 topuser），存在以下情况：其他任意一个用户（代称  $x$ ）和 topuser 的偶像程度之和也越高，此时  $x$  和 topuser 的内容相似度也倾向于越大，两者是正相关的。而且这种正相关程度与 topuser 的“偶像趋同性”程度是等价的。因此可以推测，这种用户“单极化”趋同的假设是有合理性的；而成为这个“单极”的模子，就是最高“偶像趋同性”的那个用户。例如，当 topuser 的“偶像趋同性”程度等于 1 时，则任何一个其他用户  $x$ ，只要  $x$  自身的偶像程度越高，则和 topuser 就必定内容相似度越高，而且是完全确定的一一对应关系。虽然这是理论上的案例而在现实中难以真实发生，但是对于其他高“偶像趋同性”用户来说，这种情况还是存在的，只是其强烈程度稍弱。同时这里需要注意，体现“单极化”方向的是“偶像趋同性”程度最高的用户，而不是“偶像程度”最高的用户。

把关于“单极化”的推测转为可操作化的假设 H1.1：微博中“偶像趋同性”程度最高的用户设为 topuser；其他  $m$  个用户记为  $Y = [y_1, y_2, y_3, \dots, y_m]$ ； $Y$  中每个用户与 topuser 的内容相似度依次记为  $S = [s_1, s_2, s_3, \dots, s_m]$ ； $Y$  中每个用户的作为偶像用户的程度依次为  $I = [i_1, i_2, i_3, \dots, i_m]$ ；则  $I$  和  $S$  之间具有显著的线性正相关。

H1.1 还原为通俗表述是：任一用户，如果他作为偶像用户的程度越高，则他和最高“偶像趋同性”程度的那个“模板”就越相似。

#### （二）微博用户随着影响力程度增加而表现出的收窄化

这里的“收窄化”指的是：把某个用户以及和他同“偶像程度”的多个用户组成一个“用户层级”，则从低到高的层级中，同层级内用户的“彼此相像”程度会越来越高。偶像程度最高的偶像层级拥有最高的彼此相似度；次高层级则拥有次高的彼此相似度；依此类推。

为什么做出这样的推断？依然是基于“偶像趋同性”以及 H0 的简单延伸。由于越是高偶像程度的用户其“偶像趋同性”程度也越强烈，这就意味着，高偶像程度的用户和其他任意一个用户的相似度更容易与两者的偶像程度之和有正相关，而非随机；反之，低偶像程度用户和其他任意一个用户的相似度则更容易随机而缺乏某种约束性。这样就可以推测，高偶像程度用户彼此由于更强的

“偶像趋同性”因而更容易具有高相似度；高偶像程度和低偶像程度用户的相似度同样由于“偶像趋同性”的作用而与两者的偶像程度之和成正相关，因此稍弱；但低偶像程度用户之间的相似度或者由于“偶像趋同性”的作用而偏低，或者由于没有受到“偶像趋同性”的作用而倾向于围绕平均相似度波动于一个随机值。由此，我们提出推测：越是低偶像程度的用户其彼此相似度越低，越是高偶像程度的用户其彼此相似度越高，两者具有正相关性。

将上述推测转换为本文要验证的假设 H1.2：微博  $m$  个用户记为  $Y = [y_1, y_2, y_3, \dots, y_m]$ ； $Y$  中对每个用户而言和自己在偶像程度上最为接近的  $k$  个用户组成一个“用户层级”，这样就得到  $m$  个用户层级。如果每个层级内的用户彼此之间内容相似度的均值依次记为  $S = [s_1, s_2, s_3, \dots, s_m]$ ，每个层级中用户的偶像程度均值依次记为  $I = [i_1, i_2, i_3, \dots, i_m]$ ；则  $I$  和  $S$  之间具有显著的线性正相关。

每层用户内部彼此相似度的计算方法，为后文式 3。

H1.2 通俗表述是：偶像程度越高的用户层级，层级内部各用户越是长的彼此相像、消减掉了彼此的异质性；而偶像程度越低的用户层级，越是长的“五花八门”、分布得更为宽泛。

(三) 微博用户随着影响力程度增加而表现出的大众化

这里的大众化是指，越是高程度的偶像用户，他们的内容就越不是“特立独行”、引领一方或“独树一帜”的意见领袖、内容精英；相反，他们的内容越趋于普通和去个性化、更为“长着大众脸”、和多数人更为相似。

这和一种容易产生的经验是相悖的，也即误认为拥有很多粉丝规模的偶像用户需要独特的、差异化的或者飘然乎众人的精英化内容，才可能成为某种程度上的“意见领袖”或某领域的大咖。为什么高粉丝的用户不是更拥有独特“气质”和“识别度”的用户，反而是与多数人更为相似、“泯然众人矣”？仍然可以从“偶像趋同性”解释与推断。

由前文已得，“偶像趋同性”程度为正的用户  $x$  会和其他任意用户产生趋似性，趋似的程度随着其他任意用户的偶像程度的高低而变化，两者具有和  $x$  的“偶像趋同性”程度相等的正相关系数。那么对于全体的“偶像趋同性”程度为正的用户集  $U = [x_1, x_2, x_3, \dots, x_p]$  而言，其他任意用户的偶像程度和  $U$  中每个个体的相似度都具有正相关性。

在此基础上提出假设 H1.3：微博  $m$  个用户记为  $Y = [y_1, y_2, y_3, \dots, y_m]$ ， $Y$  中用户的偶像程度依次记为  $I = [i_1, i_2, i_3, \dots, i_m]$ ； $Y$  中“偶像趋同性”程度为正的用户记为  $U = [x_1, x_2, x_3, \dots, x_p]$ ； $Y$  中单个用户和  $U$  中各个用户的相似度之和依次记为  $S = [s_1, s_2, s_3, \dots, s_m]$ ；则  $I$  和  $S$  之间具有显著的正相关。

其中，每个用户和一群用户  $U$  的平均相似度的计算方法，为后文式 2。计算得到的结果乘以  $U$  中的人数即为某个用户和  $U$  的“众人”的总相似度。

H1.3 的通俗表述是：偶像程度越高的用户和全体具有正“偶像趋同性”用户的总相似度越高，而这个总相似度越高也等价于平均相似度越高；最后体现为：偶像程度越高的用户和“占大部分比例的一大帮子人”的相似度越高，也即随着自身偶像程度的升高而越是“贴近和隐匿于滚滚人流之中”。

当然，这个“去个性化”的强弱和系统中具有正“偶像趋同性”用户的数量比例有关。根据本研究样本的分析，其中“偶像趋同性”程度为正且显著的用户占到了 63.38%，从而意味着这些用户的“去个性化”的程度较强，和六成以上的用户“大众”保持着“越来越相似、越来越贴近”的关系。

(四) 单极化、收窄化、大众化三种趋同作用之间的统一性

微博用户的单极化、收窄化、大众化之间是什么关系？容易产生的误解之一是，用户越来越变

得和某个顶端“单极”相似的同时，越来越“收窄”的同时，应该是越来越远离“芸芸众生”而变得越来越特别甚至趋于“独一无二的那一个”。也即，三者之间看似存在着某种冲突关系。另一种可能的误解则是，单极化、收窄化、大众化可能是相互独立、互不相干的过程。

但是，既然假设 H1.1、H1.2、H1.3 都是基于“偶像趋同性”机制而产生的，都是伴随着用户偶像程度的增加而不断增强的现象与效应；我们就难以轻易否认它们之间的内在统一性，因为它们是同一种机制、在同一变量变化下的现象演化。因此，这里补充提出假设 H2：用户随着作为偶像用户的程度增加，其单极化、收窄化、大众化的程度是统一的过程，而非相互独立或存有冲突的过程；也即用户作为偶像用户的程度、用户单极化程度、用户收窄化程度、用户大众化程度这四者具有一致性。对 H2 的检验方法，采取常用的克朗巴赫系数（Cronbach's Alpha 系数）。四者的克朗巴赫系数越高，表明相互的统一性越高；反之则一致性越差。

## 四、研究方法 with 实施过程

### （一）样本选取与数据预处理

根据新浪微博数据中心《2018 微博用户发展报告》，新浪微博月活跃用户 4.62 亿，月阅读量过百亿，涉及领域达 32 个<sup>[26]</sup>。采取成本相对较低的多阶段抽样，从新浪微博首页 47 个版块（社会、国际、科技、科普、数码、财经、股市、明星、综艺、电视剧、电影、音乐、汽车、体育、运动健身、健康、瘦身、养生、军事、历史、美女模特、美图、情感、搞笑、辟谣、正能量、政务、游戏、旅游、育儿、校园、美食、房产、家居、星座、读书、三农、设计、艺术、时尚、美妆、动漫、宗教、萌宠、法律、视频、上海）中，获取其发布者和评论者共 3 501 153 个用户的数据库，用八爪鱼抓取软件从中最终抓取到具有有效个人资料的、发帖数不少于 500 条的 88 739 个用户；对这些有效用户抓取每人的前 5 页关注者（新浪微博官方对每个用户最多只公开显示其 100 位关注者）。由于不同的用户所关注的人可能有重复，经过选取发帖数不少于 500 条的用户、去除重复用户、筛选具有个人资料的用户等清洗环节，从这些被关注者中选取到有效用户 130 082 个。

对用户统一横向的比较口径。其一是时间段的统一，帖子一律选取在 2018 年 1 月 1 日到 2018 年 12 月 31 日这一年的。该年度没有经历新冠疫情等特殊的大型公共事件，用户的帖子生产不会被少数共同话题所过度牵引。其二是帖子数量的统一，每个用户一律随机选取 500 条帖子。由于只言片语难以“察其言、观其人”，因此少于该帖数的用户不纳入比较范围。

经过上述取样和数据清洗整理，最后得到的样本用户为 24 779 个。剩下的这些用于计算的用户分布有广泛性与代表性，粉丝规模、发博量、关注数等各主要指标都包含从数十到数万乃至数十万的大范围，各种重度/轻度用户、活跃/不活跃用户都有良好的覆盖。

### （二）用户内容的特征提取与向量化表示

把单个用户的各条帖子无顺序拼接为一个长文本，先通过向量空间模型（Vector Space Model, VSM）得到用户的词频矩阵。词频矩阵的获取采取数据挖掘模块 sklearn 中的 CountVectorizer（）函数，其中最低词频数设为 50，即只计算出现次数达到 50 次的词，最大文档频率设为 0.2，也即在 20% 以上的帖子中都出现的词不予计算，只采用一元词而未采用二元词、三元词等多元词。得到词频矩阵转换为表示每个词语 TF-IDF 值的矩阵 X，转换函数为 sklearn 的 TfidfTransformer（norm='1'）。

对于矩阵 X，采取潜在语义分析（Latent Semantic Analysis, LSA）进行降维和内容特征提取<sup>[27]</sup>。LSA 利用奇异值分解技术，可把数十万以上的高维、稀疏的矩阵降到只有数千、数百的低维表示。本研究中每个用户的词频矩阵原有 113 694 维，一律通过潜在语义分析降到 700 维。降维

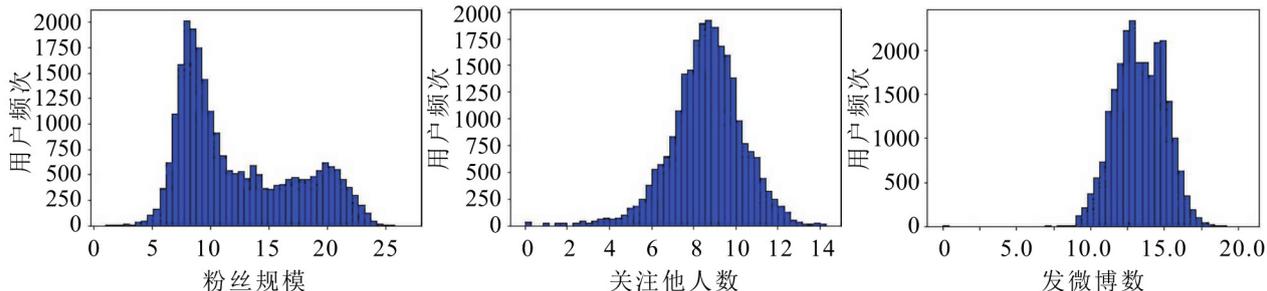


图2 样本用户的特征分布柱状图

注：粉丝数、关注数、微博数均为微博官方原始数值经  $\log_2(x+1)$  转换得到

工具选择目前广泛应用的开源模块 scikit-learn 中的 TruncatedSVD () 函数。

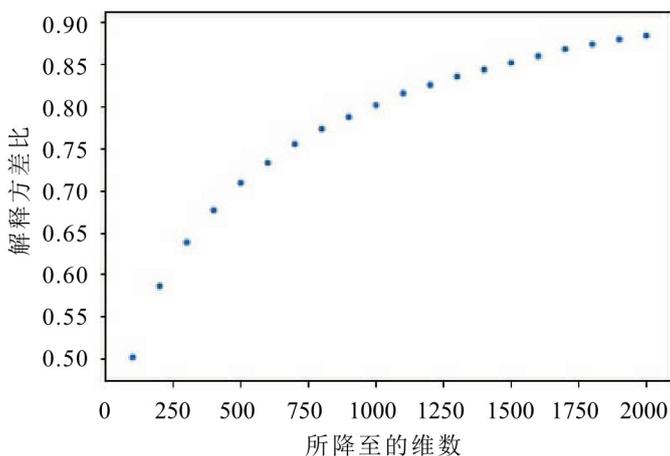


图3 潜在语义分析对用户内容降维的解释方差比

选择 700 维的维度时, 其解释方差比处于图 3 中“肘拐点”的位置, 再增加维度已经对于保留原信息的程度增幅平缓; TruncatedSVD () 计算得到的解释方差比值已达 0.76, 保留了原有的十一余万维矩阵的大部分信息, 用尽可能精简的维度提取和高效表示用户特征, 剔除冗余或噪音信息特征。

### (三) 用户的内容相似度计算

每个用户根据其内容提取为 700 维的向量之后, 就可以直接进行用户相似度计算。选择文本挖掘中稳健的余弦相似度。任意两个用户  $U_m$  和  $U_n$  之间的内容余弦相似度计算方式表示为:

$$R(U_m, U_n) \quad (1)$$

在上述式 (1) 的基础上扩展, 从“1 对 1”的用户相似度扩展到“n 对 n”的两组用户 (每组中用户数量  $n \geq 1$ ) 之间的相似度。任意一组用户  $G_1$  (包含  $n_1$  个用户) 和另一组用户  $G_2$  (包含  $n_2$  个用户) 的内容相似度, 表示为:

$$H(G_1, G_2) = \frac{1}{n_1 n_2} \sum_{x_i \in G_1} \sum_{x_j \in G_2} R(U_{x_i}, U_{x_j}) \quad (2)$$

式 (2) 在式 (1) 的基础上, 采用“类平均法”扩展得到。其中  $G_1$  或  $G_2$  都可以有且仅有一个用户。  $H(G_1, G_2)$  的值越大, 表明  $G_1$ 、 $G_2$  这两组用户之间两两的类同乃至重复程度越高; 若两组用户的异质化内容越大则平均相似度就会越低, 也即  $H(G_1, G_2)$  的值越小。

式(2)中,当 $n_1=n_2=1$ 时,则等同于式1。式(2)中,当 $n_1=1$ 而 $n_2>1$ 时,为计算一个用户和一组用户之间的平均相似度,此时用于H1.3中个体和群组的相似度计算。

式(2)中,当 $G_1=G_2$ 且组内用户数量大于1个时,则为计算组内用户的两两之间彼此相似度的平均值(用于对H1.2中用户层组“收窄化”的计算),也即等价于计算 $G_1$ 内部的自我相似度。此时采取式(2)的微调形式如下,因为不能包含个体和自身的相似度(因为该值=1):

$$H(G_1, G_1) = \frac{1}{n_1 \times (n_1 - 1)} \sum_{x_i \in G_1} \sum_{x_j \in G_1} R(U_{x_i}, U_{x_j}), \text{其中 } i \neq j \quad (3)$$

#### (四) 微博用户在个体与宏观层面的“偶像趋同性”现象预调研

1. 个体层面的用户“偶像趋同性”现象与分布。结合前文P0和N1中的界定,对微博用户在个体表现中的“偶像趋同性”程度和分布进行计算,其分布频次如图4所示,基本统计特征描述和百分位值如表1、表2所示。

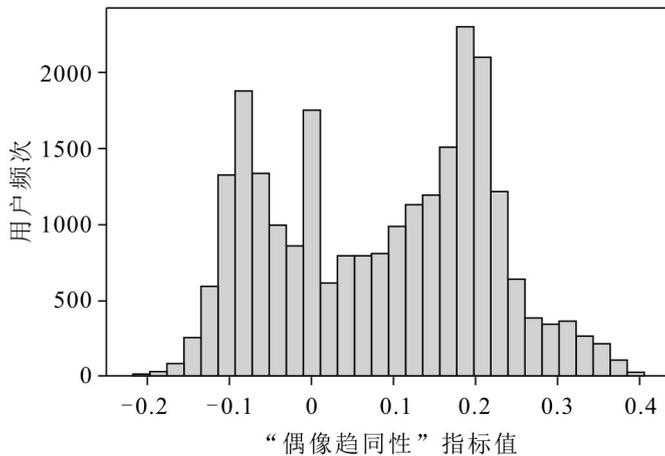


图4 新浪微博用户个体的“偶像趋同性”指标值分布

表1 用户个体“偶像趋同性”程度的统计特征描述

样本数	平均值	中位数	标准偏差	最小值	最大值
24779	0.083759	0.100338	0.1304872	-0.2200	0.4066

表2 用户个体“偶像趋同性”程度百分位数值

百分位数	5%	10%	15%	20%	25%	30%	35%	40%	45%	50%
	-0.11	-0.09	-0.08	-0.06	-0.04	0.00	0.00	0.04	0.07	0.10
百分位数	55%	60%	65%	70%	75%	80%	85%	90%	95%	100%
	0.13	0.15	0.17	0.18	0.19	0.20	0.22	0.24	0.28	0.41

这些结果显示,用户个体表现出“偶像趋同性”不只是一种理论上的想象,而是在现实中表现出的客观现象。各个体的“偶像趋同性”程度值有正有负,但总体上正值居多,平均值为0.084,中位数为0.100。正的最大值为0.406,高于负的最小值-0.22。63.4%的用户存在显著的、正向的“偶像趋同性”,其余的则是无显著效应或指标值为负的用户。多数用户的这种现象的显著使得个体层面的偶像趋同性具有现实性,并会使得宏观总体也表现出一般性的偶像趋同性现象。

2. 用户的偶像程度与“偶像趋同性”程度之间的关系(H0)。用户的“偶像趋同性”的强度不  
万方数据

是任意分布的,而是与用户的偶像程度之间具有正向关系。对于 24 779 个用户样本依次得到其偶像程度值的序列、其“偶像趋同性”程度值的序列,则两个序列之间,皮尔逊相关系数为 0.544 ( $p < 0.001$ ,  $N = 247\ 79$ ),斯皮尔曼相关系数为 0.537 ( $p < 0.001$ ,  $N = 247\ 79$ ),从而验证了  $H_0$ 。

围绕用户偶像程度和“偶像趋同性”程度之间的正向相关性与同步性,将用户按照偶像程度等频划分为 10 个层级(从低到高依次标为第 1、2、3、…、10 层),每层对应的偶像趋同性程度(95%置信区间)如图 5 所示。

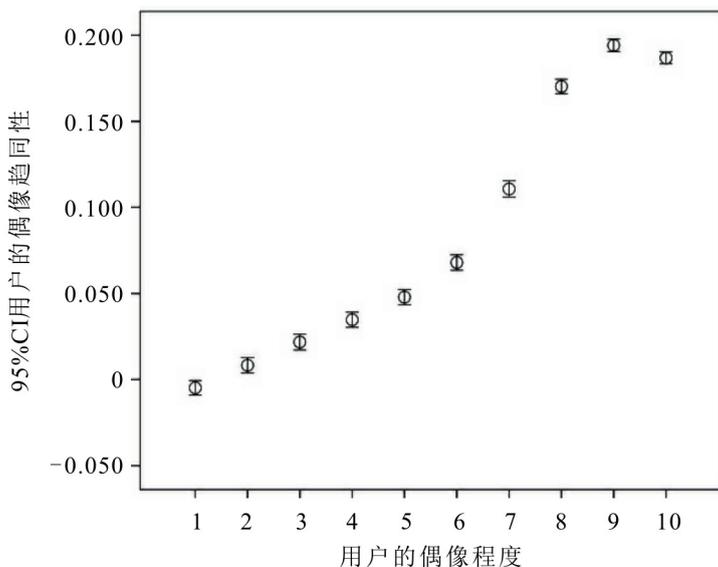


图 5 用户的偶像程度及其所对应的偶像趋同性程度

图 5 显示,偶像程度越高的用户其偶像趋同性的程度也越强烈。偶像程度处于最低层级的用户,其偶像趋同性均值只有-0.005;而偶像程度处于最高层级的用户其偶像趋同性均值已稳步提升到 0.187,大大超过全体均值 0.084。

3. 宏观层面的用户“偶像趋同性”现象与效应。从上述的特定个体的微观层面,自然拓展到总体中任意两个个体的宏观层面。继续对微博宏观条件下任意两个用户之间相互的偶像趋同性是否显著展开检验。由  $P_0$  延伸而得到的扩展性的假设为  $P_1$ :任意两个用户  $x$  和  $y$  的内容相似度,与这两个用户各自作为偶像的程度相加的和,两者之间具有正相关。

对  $P_1$  的检验方法为,抽取新浪微博中的  $n$  个用户,两两配对(不含和自己的配对),形成  $n \times (n-1)$  对用户;然后考察每对用户的内容相似度,和对中两位用户偶像程度之和,是不是具有正相关。对本研究的 24 779 个用户样本两两之间进行分析,得到每对用户的内容相似度(代称序列 a)、每对用户的偶像程度之和(代称序列 b)。序列 a 和序列 b 之间的皮尔逊相关系数为 0.125 ( $p < 0.001$ ,  $N = 613\ 949\ 284$ ),斯皮尔曼相关系数为 0.162 ( $p < 0.001$ ,  $N = 613\ 974\ 062$ )。这显示即使拓展到宏观总体中的任意用户,用户之间的“偶像趋同性”现象也是存在的和显著的;越是强偶像程度的用户相互之间相似度越高,反之则越低。尽管由于受到多种干扰因素,相关系数不算高,但仍然清晰可见这种正向效应的显著和规律性,而不是负向的效应,也不是杂乱无规律。

## 五、核心假设的实证检验

本节在  $P_0$  的基础上,检验“偶像趋同性”所内蕴的微博偶像单极化( $H_{1.1}$ )、收窄化

(H1.2)、大众化 (H1.3) 的态势与趋同路径, 以及这些不同路径之间的内在一致关系 (H2)。

### (一) 微博偶像的单极化、收窄化、大众化

对于 24 779 个样本用户, 首先依次计算每个用户的“偶像趋同性”程度 (记为  $I = [i_1, i_2, i_3, \dots, i_m]$ )。然后依次对于每个用户计算他和具有最高“偶像趋同性”程度用户的内容相似度 (记为数值序列  $S_i = [st_1, st_2, st_3, \dots, st_m]$ )；每个用户所属“偶像趋同性”层级的用户层内平均相似度 (记为数值序列  $S_g = [sg_1, sg_2, sg_3, \dots, sg_m]$ )；每个用户和具有正的偶像趋同性的多个用户的总相似度 (记为数值序列  $S_p = [sp_1, sp_2, sp_3, \dots, sp_m]$ )。

最后通过分析数值序列  $I$  是不是分别和  $S_i$ 、 $S_g$ 、 $S_p$  具有显著的正相关, 从而检验微博用户的单极化、收窄化、大众化规律是否成立, 三者分别依次对应于前文所述的 H1.1、H1.2、H1.3。检验结果如表 3 所示。

表 3 假设 H1.1、H1.2、H1.3 的内容及其检验

假设	H1.1: 微博偶像的单极化规律	H1.2 微博偶像的收窄化规律	H1.3 微博偶像的大众化规律
假设的内涵	微博用户的偶像程度越高, 则其与“偶像趋同性”程度最高的那一个用户越是趋同	微博用户的偶像程度越高, 则其同偶像程度层级的 $k$ 个用户彼此之间越是相似 (本处分别选择 $k=30$ 、 $k=50$ 两种情况验证)	微博用户的偶像程度越高, 则其和“偶像趋同性”程度为正的全体 $k$ 个用户 (本研究 $k=15706$ ) 总相似度/平均相似度越高
检验方法	检验方法: 序列 $I$ 和序列 $S_i$ 具有显著的正相关	检验方法: 序列 $I$ 和序列 $S_g$ 具有显著的正相关	检验方法: 序列 $I$ 和序列 $S_p$ 具有显著的正相关
检验结果	皮尔逊相关系数 0.407 ( $p < 0.001$ , $N=24779$ )； 斯皮尔曼相关系数 0.386 ( $p < 0.001$ , $N=24779$ )	(1) $k=30$ ： 皮尔逊相关系数 0.742 ( $p < 0.001$ , $N=24779$ )； 斯皮尔曼相关系数 0.677 ( $p < 0.001$ , $N=24779$ ) (2) $k=50$ ： 皮尔逊相关系数 0.812 ( $p < 0.001$ , $N=24779$ )； 斯皮尔曼相关系数 0.743 ( $p < 0.001$ , $N=24779$ )	皮尔逊相关系数 0.234 ( $p < 0.001$ , $N=24779$ )； 斯皮尔曼相关系数 0.274 ( $p < 0.001$ , $N=24779$ )

单极化说明, 由于用户个体之间的“偶像趋同性”, 所有用户随着自身偶像程度的增加而朝向“同一个”用户越长越像。偶像用户在消减着自身的多样性, 而非趋于多样化和细分差异化。

收窄化同样是和偶像趋同性高度关联的效应, 并且由于它反映的是层组的趋同逻辑, 会减少个体层面的过大波动, 所以效应更为稳定。对于 24 779 个用户分别所属的 24 779 层用户阶层, 其高达 0.7 乃至 0.8 以上 ( $N=24 779$ ) 的皮尔逊相关系数显现了偶像趋同性效应传导至收窄化的结果, 规律性高度稳定。图 6 鲜明呈现了 24 779 层用户的“收窄化”程度的变化:

对于微博偶像的大众化而言, 同样是显著的效应和规律, 反映着偶像用户越来越去除独特性。马尔库塞曾提出“单向度的人”<sup>[28]</sup>, 反思着现代社会中主体的重复性和单一化。这些批判性的社会理论, 虽缺乏实际证据的验证, 但却留下丰富的启示。本研究对于微博偶像“大众化”的现象检验和证据显示, 越是高粉丝规模的偶像用户就越趋于大众化、相似化和平庸化, 而非“独特可辩”或“各领风骚”; 并且这种大众化的“标的”是具体的、而不是笼统的: 和全体的“偶像趋同性”程度为正的全体用户, 总相似度和平均相似度越来越高, 从而表现出对于同一化和重复化用户的生产。

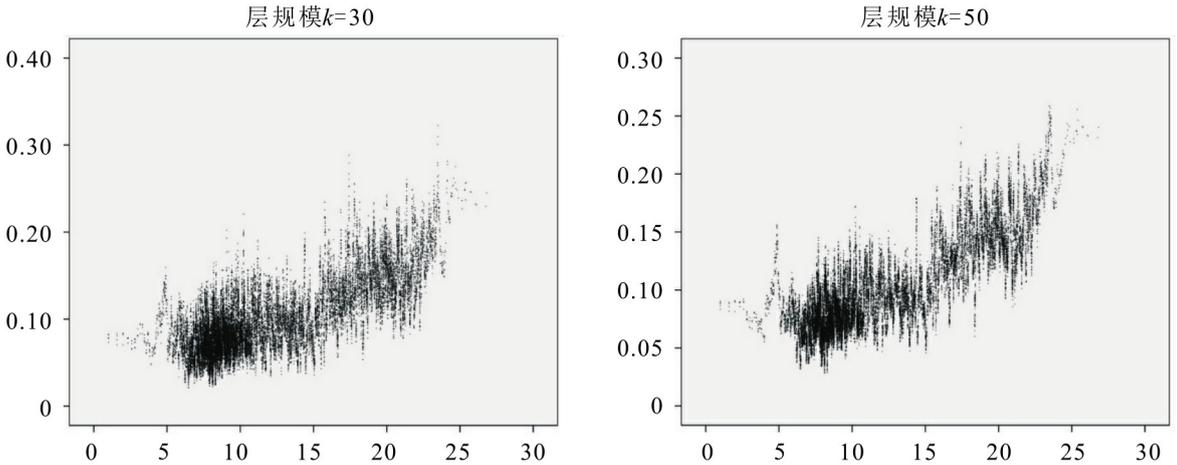


图6 用户所处偶像程度层级提升而“收窄化”的趋势

横轴为用户所属“偶像层级”的平均粉丝规模,纵轴为该层级用户的内容相似度平均值。样本点为24 779个。

## (二) 偶像程度与单极化、收窄化、大众化的趋同作用之间的统一性

本处涉及对假设H2的检验。前文所述的每个用户的偶像程度,以及在H1.1、H1.2、H1.3所分别涉及的单极化、收窄化、大众化程度,这四个变量全部转化为正态得分值,其后计算得到四者的克隆巴赫系数为0.759。Cronbach's Alpha系数达到0.6以上属于可接受,达到0.7以上时表示变量之间具有较高的一致性。这四个变量中,如表4所示,删除任意一个变量不能明显提高Cronbach's Alpha系数的结果(仅在删除“大众化程度”的指标时微微从0.759上升到0.763,但这个变化过于微弱,依然可以视四个变量为统一、同步变化的整体)。总体而言,H2得到实证检验与支持。

表4 用户偶像程度与单极化、收窄化、大众化的统一性

用户指标	作为偶像用户的程度	单极化程度	收窄化程度	大众化程度
项已删除的Cronbach's Alpha值	0.665	0.679	0.699	0.763

随着用户的粉丝规模增加,用户所表现出的单极化程度、收窄化程度、大众化程度的变化,是一致和同步而不是相互独立或冲突的。这与常识观感有相悖之处,值得我们注意。用户在趋于“顶部”最高“偶像趋同性”用户的同时,也在趋于和多数化的众人越来越相似,自身的独特性降低;同时上述这两个过程也伴随着偶像本身彼此的相似程度越来越高。

## 六、结 语

本文的工作主要如下。

1. 针对社交网络中的用户相似关系和趋同问题,明确提出用户“偶像趋同性”的基本现象(P0),并以微博为主体进行实证研究,所得结果验证了本文提出的现象和效果,证明“偶然趋同性”不只是一种理论上的虚构或臆造。“偶像趋同性”的现象体现了微博社会中的用户演变和趋同分布的规律性,背后具有丰富的内涵值得继续从理论上深入挖掘。

2. 在“偶像趋同性”的基础上,延伸出其对于微博社会带来的趋同作用后果:用户的单极化、收窄化、大众化,这三种后果的规律性全部得到实证检验。它们作为偶像趋同性的社交媒介后果,

是与前者紧密相关联的整体，分别是指朝向某一个模子的趋同、同一影响力层级内的趋同、朝向多数用户的趋同。同时，这三者也表现出高程度的同步，显现出趋同路向的统一性，表现出了微博在线社会运行的整体同化逻辑。

用户的偶像趋同性及其所内蕴的单极化、收窄化、大众化等趋同后果，部分程度上是反直观、不符合常识观感的，和一些强调用户多样性或分化性的具有较大拥趸的理论观点具有不符之处，但它们得到了实证结果的支持，且其作用机理也表现出一定的合理性。这警示我们在面对社交网络空间及偶像社会流动的复杂性时，要更谨慎地存疑与求证。同时，偶像趋同性的个体机制、宏观现象及其后果的分析显示出，用户随着其影响力的升高，多种趋同路径都在不断增强，社交网络中的用户变得越来越相似与同质化，并从趋于单极的同质化、同层级的同质化、大众的同质化进行立体化的系统演变。这为我们分析微博社交网络中的趋同化社会特征与相似性传递、同质性社会文化生产的运行机理，提供了可探索路向与延伸空间。

### 参考文献

- [1] [美]马克·波斯特. 第二媒介时代[M]. 范静哗,译. 南京:南京大学出版社,2001.
- [2] [美]丹·吉摩尔. 草根媒体[M]. 陈建勋,方延明,译. 南京大学出版社,2010.
- [3] 王晗啸,于德山. 意见领袖关系及主题参与倾向研究——基于微博热点事件的耦合分析[J]. 新闻与传播研究,2018(1).
- [4] 曹洵,张志安. 基于媒介权力结构的微博意见领袖影响力研究[J]. 新闻界,2016(9).
- [5] 王国华,张勇波,王雅蕾,等. 微博意见领袖的类型特征与内容指向研究[J]. 电子政务,2014(8).
- [6] 生奇志,高森宇. 中国微博意见领袖:特征·类型与发展趋势[J]. 东北大学学报(社会科学版),2013(4).
- [7] Himelboim, I., S. McCreery, M. Smith. Birds of a feather tweet together: Integrating network and content analyses to examine cross-ideology exposure on Twitter[J]. *Journal of Computer-Mediated Communication*, 2013(2).
- [8] 陈福平,许丹红. 观点与链接:在线社交网络中的群体政治极化[J]. 社会,2017(4).
- [9] Lawrence, E., S. H. Farrell. Self-segregation or deliberation? Blog readership, participation, and polarization in American politics[J]. *Perspectives on Politics*, 2010(1).
- [10] Colleoni, E., A. Rozza, A. Arvidsson. Echo chamber or public sphere? Predicting political orientation and measuring political homophily in Twitter using big data[J]. *Journal of Communication*, 2014(2).
- [11] Van Alstyne, M., E. Brynjolfsson. Could the internet balkanize science? [J]. *Science*, 1996(5292).
- [12] 张玉晨,翟姗姗,许鑫,等. 微博“中V”用户的传播特征及其引导力研究——以罗一笑事件为例[J]. 图书情报工作,2018(11).
- [13] 曹洵,张志安. 社交媒体意见群体的特征、变化和影响力研究[J]. 新闻界,2017(7).
- [14] 佟力强. 中国微博发展报告(2012)[M]. 北京:人民出版社,2013.
- [15] Matsumura, N., Y. Ohsawa, M. Ishizuka. Influence diffusion model in text based communication[J]. *Transactions of the Japanese Society for Artificial Intelligence*, 2002(3).
- [16] Borge, B., R., E. D. V. Marc. Opinion leadership in parliamentary Twitter networks: A matter of layers of interaction? [J]. *Journal of Information Technology & Politics*, 2017(3).
- [17] 高剂斌,贾贞,郝佳佳. 基于个体观念演化驱动的社交网络群体行为建模及分析[J]. 桂林理工大学学报, 2018(1).
- [18] 李根强,方从慧. 复杂网络视角下网络集群行为为主体的观点演化研究[J]. 情报科学,2017(5).
- [19] 李卫东,彭静. 社交网络平台信息传播的回声室效应仿真实验分析[J]. 现代传播(中国传媒大学学报), 2019(4).

- [20] Yoo, E. , B. Gu, E. Rabinovich. Diffusion on social media platforms: A point process model for interaction among similar content[J]. *Journal of Management Information Systems*, 2019(4).
- [21] 翟秀凤. 创意劳动抑或算法规训? ——探析智能化传播对网络内容生产者的影响[J]. 新闻记者, 2019(10).
- [22] Masson, A. R. , Y. Hayel, E. Altman. Posting behavior dynamics and active filtering for content diversity in social networks[J]. *IEEE Transactions on Signal and Information Processing over Networks*, 2017(2).
- [23] Litt, E. , E. Hargittai. The imagined audience on social network sites[J]. *Social Media & Society*, 2016(1).
- [24] Mikal, J. P. , R. E. Rice, R. G. Kent, et al. Common voice: Analysis of behavior modification and content convergence in a popular online community[J]. *Computers in Human Behavior*, 2014(6).
- [25] 刘志明, 刘鲁. 微博网络舆情中的意见领袖识别及分析[J]. 系统工程, 2011(6).
- [26] 新浪微博数据中心. 2018 微博用户发展报告[EB/OL]. <https://data.weibo.com/report/reportDetail?id=433>, 2019-03-15.
- [27] Dumais, S. T. Latent semantic analysis[J]. *Annual Review of Information Science and Technology*, 2004(1).
- [28] [美]赫伯特·马尔库塞. 单向度的人[M]. 刘继, 译. 上海: 世纪出版集团, 上海人民出版社, 2008.

## The Phenomenon of “Idol Convergence” in Social Network and Its Convergence Consequences

— Empirical Analysis Based on Sina Weibo Samples

XU Xiang

**Abstract:** This paper clearly and creatively puts forward the theoretical concept and phenomenon of “idol convergence” in social network, and empirically examines the phenomenon of “idol convergence” and its macro effect and user convergence structure with sina Weibo samples. Latent semantic analysis (LSA) is used to extract user content features and calculate similarity. On the basis of the macro level effect test and individual level difference analysis of “idol convergence”, the paper concludes the evolution mechanism of micro blog users’ monopolization, narrowing and popularization, and the internal synchronization rather than independence or conflict of these paths. These evolutionary trends and laws of microblog users are not only the organic components of the logic embedded in the “idol convergence”, but also the consequences of network information society which is in need of in-depth exploration.

**Key words:** idol convergence; microblog; user content production; content similarity; latent semantic analysis

(责任编辑 周振新)