

# 第一章 维基百科中的用户分类研究

在虚拟社区中，用户是内容的创建者和分享者。社区的活力取决于用户的参与和活跃程度，并最终影响到社区能否持续存在。虚拟实践社区中个人的特质、参与动机、参与行为和资历的各不相同，每人的积极程度、个人性格、表达能力和知识水平也不同，个体在社区互动中扮演的角色不同，因而在同一虚拟社区内的用户会在社区中具有不同的行为模式，并自然分化为多个不同的群体。

大量的研究表明，虚拟社区中总是存在着一小部分热忱的参与者和人数众多的潜水者。积极参与的社区成员实际上介入了社区的绝大部分活动，包括内容的创建，各类社区活动的发起和组织，以及各类日常的维护工作。Bryant 曾经对英文维基百科的活跃用户进行过调查，发现这些活跃用户所参与的工作要远远多于那些非活跃用户，除了条目内容本身的编写外，还参与社区政策的指定，讨论社区的发展方向，以及其他为保证社区正常运转的活动<sup>[2]</sup>。而大多数人则只是作为社区中的看客，很少介入到社区的活动中，尤其是那些需要投入时间和精力活动，例如创建新内容。这种特性已经在包括社会网络、论坛、邮件组、维基和邮件列表等在内的各类虚拟社区中得到了验证。

通过定量分析可以发现，大部分虚拟社区的内容创建是服从幂率分布的：即一小部分用户贡献了社区中的绝大部分内容。维基百科作为典型的以内容创建为目标的虚拟时间社区也不例外，已经有很多研究验证了不同语言版本的维基百科都具有这个特点，图 1.1 展示了英文维基百科中内容的贡献于用户数量之间的关系：

成员是虚拟社区形成和存在的基础。分清成员在虚拟社区中扮演何种角色，以及不同类型的成员所占的比例是关系到社区能否进一步发展的重要问题。Amsrtnog 和 Hgael(1997) 提出，虚拟社区用户是被感兴趣的话题及内容所吸引，在错综复杂的网络环境中找到社区。尽管他们具有加入社区的潜在倾向，但还是要根据兴趣及满意程度决定是否从单纯的浏览者转变为社区用户，反复进入其中并参与互动，乃至成为生产内容的社区建设者。他们根据参与程度和价值两个维度将虚拟社区成员分为如下四种：浏览者 (Browser)、潜水者 (Lurker)、贡献者 (Contributor)、购买者 (Shopper)。Adler 和

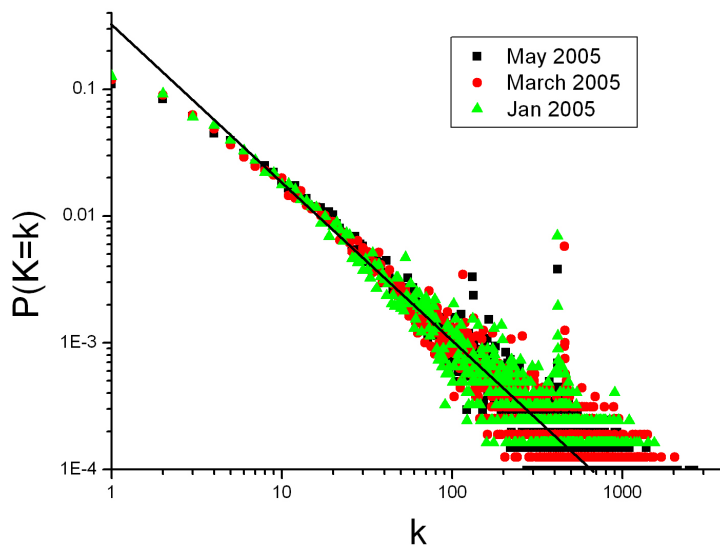


图 1.1: 幂率分布

Christopher(1998) 则根据参与形式将虚拟社区成员分为以下四种: 被动者、主动者、诱导者、管理者 [20]。巴特勒 (Bartle, 1990) 在对基于游戏编程的 MUD 型虚拟社区进行多年群体动力学观察基础上, 从使用动机及相关行为的角度将其用户大致分为恶作剧者 (killers)、成就者 (achievers)、社交者 (socializes)、探索者 (explorers) 四类 [21]。Kozinest(1999) 在研究虚拟消费社区时根据成员与消费活动的关系和成员与虚拟社区的关系两个因素来将虚拟社区的成员划分为不同的类型: 浏览者、社交者、贡献者以及内部者。Kozinets 认为一个虚拟社区成员的最后确认要基于两个非独立的因素: 个人与特定的消费行为之间的关系以及个体与社区中其他成员社会关系的强度, 这两个因素是相关的。依据这两个因素可以将社区中的成员划分为如图 1 所示的四种类型 [22]:

- (1) 浏览者: 与群体缺乏强的社会联系, 对消费行为只有比较小的或短暂的兴趣。
- (2) 社交者: 与群体的社会联系强, 但对核心的消费行为兴趣不大。
- (3) 贡献者: 对于消费行为有着很强的兴趣和热情, 但与群体的社会联系比较弱。
- (4) 内部者: 与消费行为的个体联系以及与群体的社会联系都比较强。

杨堤雅进一步将虚拟社区成员分为了八种类别, 并总结了每种类型成员的特征, 如表 1 所示:

用户的行为受其动机水平影响, 导致不同类型的用户有其不同的行为方式。同时

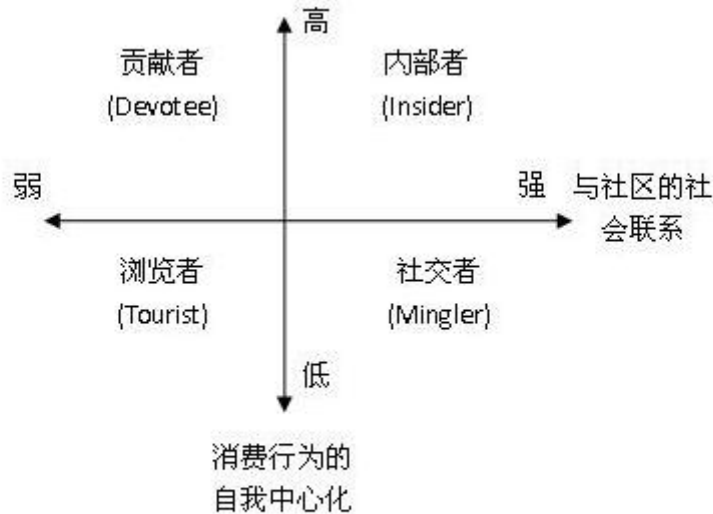


图 1.2: 虚拟社区的四种成员类型

表 1: 虚拟社区成员角色及特征

特性 \ 角色	参与程度	专业知识	成员互动	文章主要内容
成员领袖	高 \ 中	高	高	提供意见、分享经验
意见呼应者	高 \ 中	中 \ 低	高	提供意见、分享经验
自我揭露者	中 \ 低	低	低	分享经验
经验意见分享者	高 \ 中 \ 低 \ 路过	高 \ 中	中	提供意见、分享经验
查询者	低 \ 路过	中 \ 低	中 \ 低	提问
信息推广者	高 \ 中 \ 低 \ 路过	高 \ 中	高 \ 中 \ 低	信息推广、建立关系
浏览者	—	低	低	无
干扰者	低 \ 路过	低	低	其他

行为的结果也会影响到用户的动机水平，引起用户动机的增强或者减弱。即使是同一种动机因素，对于不同类型的用户所起到的作用也可能是不同的。因此，有必要对用户的类型加以区分，在此基础上分别研究动机因素对每类用户的影响。通过对现有的文献进行分析可以发现，大部分关于维基百科用户动机因素的研究均将社区中的所有用户视为一个整体，而并未针对不同群体的用户特征进行专门的分析。这就使得研究的结果模糊了不同用户群体间的区别，同时也无法提出有针对性的管理建议。本文的研究力求弥补这方面的研究空白，从用户的知识协同动机角度出发，分析和研究用户

的分类。

## 1.1 数据处理

维基基金会大约每隔 3 周就会对下属所有语言版本的维基百科的数据下载备份, 形成一个时间点的归档。备份的目的除了用于灾难回复外, 更重要的是为所有有志于参与维基百科学术研究的个人和团体提供数据支持。备份的内容除了所有条目的内容外, 还包括页面链接的列表以及图片元数据等内容。备份按照包涵内容的不同分为 3 种类型:

- **pages-articles** 备份。该类型的备份包含所有条目、模板和其它页面的当前版本, 去除了条目的讨论页面和用户主页, 适用于在其他地方再发布维基百科的内容。
- **pages-meta-current** 备份。提供所有页面的当前版本。
- **pages-meta-history** 备份。提供所有页面的所有历史版本, 适合于学术研究。

本研究为了分析用户的协同行为并据此为其分类, 因此需要研究用户的协同贡献, 比较所有历史版本同最后版本间的内容的区别。显然, **pages-meta-history** 备份包含了研究所需的全部数据。维基基金会提供了不同时段 **pages-meta-history** 备份, 随着时间的推移, 数据量显著增大。其中, 最近的备份数据量在未压缩的情况下达到了 140Gb, 数据的处理量非常大。为此, 本研究选择了一个较小的数据集, 囊括了中文维基百科从创立值 2008 年 8 月的所有历史数据。数据集在未压缩情况下大小为 70Gb, 存储格式为 XML, 其基本的数据结构如下:

```
<page>
  <title>Wikipedia:Upload log</title>
  <id>1</id>
  <restrictions>sysop</restrictions>
  <revision>
    <id>288633</id>
    <timestamp>2002-12-11T09:10:02Z</timestamp>
    <contributor>
      <username>Formulax</username>
```

```

    <id>3</id>

    </contributor>

    <comment>uploaded &quot;Shanghai1.jpg&quot;;: 上海夜景</comment>

    <text xml:space="preserve">Below is a list of the most recent file uploads.
        .....
    </revision>
</page>

```

数据由一个个 `<page>` 组成，每个 `page` 是维基百科中的一个页面（如条目页面、讨论页面、用户页面等）。每个 `page` 包括标题、`id`、以及所有的历史版本等信息。其中 `page` 的标题由两部分组成，用“:”分割开：第一部分是该 `page` 所属的命名空间；第二部分是该 `page` 实际的标题。这样，就可以通过标题判断出该 `page` 的实际所属类型。通常一个 `page` 中会包括多个版本。各个版本的内容记录在 `<revision>` 标签中，包括版本的 `id`、该版本修订的时间、作者的相关信息、每个版本的实际文字内容和修订备注。利用这些信息，可以得到所有的用户协同行为的量化的属性值，为下一步分析用户行为并分类做好数据准备。

### 1.1.1 协同用户的选择

本文研究的对象是维基百科社区中的用户的协同行为。社区用户应该是具有正常行为，努力为社区发展贡献力量的用户。但是，由于维基百科社区的开放性，各类非正常的用户不断地涌现，不仅危害了社区的正常发展，也为相应的学术研究带来了困难。维基百科社区中不正常的用户主要包括以下几种：

1. 被封禁的用户。这里用户通常是在社区中有不恰当的行为，例如攻击他人、广告宣传以及恶意地破坏等。封禁的目的只有一个，就是防止维基百科遭到持续或严重破坏。对于各种不正常行为一旦确认，用户的账户即被封禁。本文认为被封禁的用户及其行为不应该被纳入到研究中来，因为他们的行为并不属于本文所定义的协同行为。
2. 傀儡用户。维基百科社区规定每位用户只能拥有一个帐号，防止用户使用傀儡来伪造对某件事的支持度，误导他人，挑起争端，协助破坏，或回避封禁。除非有及特殊的原因，傀儡帐号不用来参与同社区的知识协同活动，而是进行伪冒、滋扰等任何违反维基百科方针的行为。傀儡也不会纳入到本文的研究范围。

3. 用户名不规范的用户。用户名是每一个社区用户注册后的唯一标识。维基百科不允许使用带有误导性、宣传性、侮辱性或破坏性的用户名；也不允许用户名与现实世界中各组织或团体相关，或令人混淆于具体的人或组织。一旦发现具有以上特征的用户名，用户可以报告至“维基百科: 需要管理员注意的用户名”页面并附上相关的解释，管理员可于审核以后把之封禁。用户被举报名不意味着马上就会封禁，也就是说不会立刻出现在封禁列表中，但是历史记录表明绝大部分上报的用户名都受到了封禁。因此，本文排除了所有用户名不规范的用户。
4. 机器人用户。机器人是实际用户创建的为了完成特定目的，按照一定规则以自动化方式运行的账户。机器人对内容的操作极其有限，大部分机器人只能用于维护跨语言链接和修复重定向。除非有社群的批准，否则机器人不能用于拼写检查和纠正错字，以及处理繁简转换问题，尤其是对条目内容的检查。机器人本身并不创建任何内容，也不反应所属用户的协同行为和协同动机。本文的研究将机器人用户排除在外。
5. 匿名用户。关于匿名用户在维基百科社区中的作用一直是有争议的研究问题。就社区本身来说，尽管并不限制匿名用户的编辑，但是社区仍然鼓励用户注册一个账户，而非匿名编辑。维基百科认为匿名编辑的质量在本质上是存疑的，因此应该尽量避免。另一方面，匿名用户为识别用户带来了困难。匿名用户使用其 IP 地址作为其标识，这就意味着来自同一个 IP 地址的编辑可能实际上是由多个用户完成的，而单个用户也可能使用多个 IP 地址进行内容编辑。匿名用户本身并不是一个正常的用户状态，用户的行为也难以确认。同时，匿名用户本身的数量在所有用户中所占比例非常小，其贡献完全可以忽略。

在所有的内容条目中，有一些条目也不适于纳入到研究范畴，需要从数据集中剔除，主要包括：

1. 重定向页面。在维基百科中，经常会出现不同的用户针对相同的内容创建了不同的条目，例如“中华人民共和国”和“中国”两个条目本身的内容是一致的。为了避免这种情况，需要对两个条目进行合并，避免用户做重复工作，也使读者不会因此而产生困惑。重定向是一种特殊的页面，它提供一种运作机制，使得人们在输入该名称进入条目或者点击指向该名称的内部链接时，系统能够自动导航到重定向页面内部指定的另一相关页面中，从而实现相关页面可以以多个名称进行

访问<sup>1</sup>。被重定向的页面将不再包含条目的实质内容，因此也就不存在用户的协同。

2. 消歧页和列表页。消歧义在维基百科中指消除由于不同条目拥有相同名称（一词多义）所引起的歧义。用户建立消歧页面，通过一些简单的注释内容消除混淆条目间的歧义，并引导读者链接到准确的具体的条目。列表页是为同一类别下的条目建立的索引式入口，如“亚洲国家列表”条目。列表页提供了该类下的所有条目的链接，方便导航读者。消歧页和列表页严格来说不算是合格的维基百科条目，而更像是方便读者的辅助性页面。另外，这两类页面的内容也常常由机器人完成。

### 1.1.2 数据分析

通过对数据集的处理，提取内容条目，在去除了重定向页面、消歧页和列表页之后，共得到 205378 个条目页面。所有的用户都围绕这些条目开展知识协同。社区中共有注册用户 55218 名，占总用户数的 93%；另有数百名匿名用户和近百个机器人用户。这近六万名用户协力完成了 4592042 次编辑，平均每个用户参与了 77 次编辑，每个条目得到了 22 次编辑。

利用上一章提出的计算用户协同贡献的方法，对数据进行处理。首先计算所有条目的条目质量，计算的结果进行归一化。随后分别计算每个条目中参与用户各自的贡献，最终可以得到每个用户在每次编辑后所做的贡献。将同一条目的编辑贡献累加可以得到该用户为该条目所做的协同贡献，将用户所有编辑的贡献累加则可以得到用户为整个社区所做的协同贡献。协同贡献计算完成后，本文进一步检验了匿名用户在社区中的行为活动。匿名用户共参与了 96389 个条目的知识协同活动，贡献了 907763 次编辑，分别占总条目数的 46.9% 以及总编辑次数的 19%。这个结果表明匿名用户似乎是维基社区中的一股重要力量，不应该轻易舍弃。但是协同贡献的计算结果表明，匿名用户的贡献度要远远低于注册用户的贡献度。注册用户平均每次编辑的贡献达到了 0.049，而匿名用户的平均贡献只有 0.008，不及注册用户的六分之一。这也验证了前文提到的匿名用户从总体上来说协同的质量并不高，对于社区的影响也非常弱。就总的协同贡献来说，匿名用户的贡献只占总量的 3%。因此从匿名用户的数量和对社区的贡献两方面因素考虑，本研究不包括匿名用户的协同行为是合理的选择。

<sup>1</sup><http://zh.wikipedia.org/zh-cn/Help>: 重定向

在将所有不适合参与研究的用户从数据集中剔除，剩下的正常用户即为本文的研究对象，即维基百科中至少参与过一次知识协同的用户。这些用户在参与社区的协同编辑过程中逐渐分化为不同的类别，每种类别均有各自不同的行为模式和特征。但是，同其他类型的虚拟社区不同，已有的关于用户分类的研究并不完全适合本文的研究对象。实际上，本文所研究的用户实际上是社区中所有用户的一个子集，即那些曾经参与过知识协同的用户。尽管这些协同用户参与维基百科社区具有相似的目的，但是仍可以根据其行为特点划分出不同类别。因此传统的针对虚拟社区所有用户的分类并不适合本文的研究，需要寻找新的分类方法和维度。

首先，参与知识协同的用户已经将传统意义上的潜水者（Lurker）和浏览者（Browser）排除在外。潜水者和浏览者均是社区的外围成员，他们加入社区的主要目的是为了消费社区的知识产品和信息产品，几乎没有贡献内容的行为。Yong Du 分析了这两类用户产生的原因，包括：1) 认为即使自己不贡献内容也不影响社区的运行；2) 希望通过加入社区进一步了解社区的相关情况；3) 不愿意与他人分享自己的知识和经验<sup>[2]</sup>。由于维基百科的开放性，如果一个用户试图从维基百科获取有用信息，根本就不需要进行注册，只是单纯浏览就可以了。注册用户都是在维基百科精神的感召下，对社区已经有了一定的了解，愿意为之贡献自己的一份力量而进行注册的。即使一部分用户注册完之后一次编辑行为也没有，这通常是由于没有找到合适的、可以参与的条目，或者是不熟悉界面操作等因素造成的。从广义上讲，本文所研究的协同用户都是知识协同过程的参与者，即都是内容的生产者而非消费者。因此，本文研究的用户群并不具备这两类用户的特征。

其次，不以协同创建内容为目的的用户也已经被排除在外。大部分虚拟社区都不可避免地存在一些不受欢迎的用户，如破坏者和广告发布者等。这些人的存在对于社区的繁荣和发展毫无益处，本文在数据处理过程中已经将这些用户排除。

### 1.1.3 分类维度

用户分类可以有多种维度，每一种分类维度都同最终研究的目标紧密相关。通过分析现有文献可以发现，很少有文献从分析用户参与的动机角度研究用户的分类维度。本文将用户的协同行为分为从两个角度来考察，这些角度反映了用户协同的特点，从而最终形成分类的依据。

用户对每个条目的平均贡献是区分不同用户类型的重要维度。用户对条目的平均



贡献反映了一个用户的投入程度及其价值。条目的贡献取决于用户自身的知识水平和其将知识转化为符合维基百科标准的内容的能力。用户所拥有的知识越多，越是积极参与该条目的编写，那么最终在条目中所作的贡献就越大。不同的用户的平均贡献相差非常大。图 1.1.3 显示了用户贡献不均衡的分布情况。

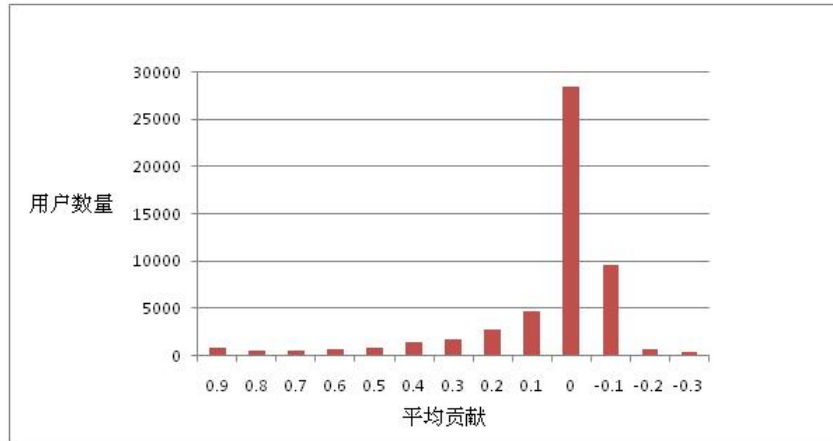


图 1.3: 用户平均贡献分布

图中显示了每个贡献度区间内的用户数量。可以看到贡献度的分布明显可以分为几个部分。从 0.9 至 0.5 这个区间可以看出用户分布非常均匀，几乎每个区间段的用户数量差别都不大。从 0.5-0.1 区间，用户数量开始稳步提升。在区间  $[0.1, -0.1]$  之间用户数量暴涨，在这个区间的用户数达到了总用户数的 69%。随后用户数量迅速减少。

不同区段的用户，其行为特点也非常鲜明。对于平均贡献大于 0.5 的用户来说，这意味着一个条目中有超过一半的内容是由该用户贡献的。称这名用户为该条目的主导者毫不为过。这类用户除了亲自编写内容外，往往还会负责引领条目的编写方向，规划内容的结构等，是组织条目编写的积极的带头人。对于平均贡献在 0.1 至 0.5 这个区间的用户来说，他们尽管在知识协同中不占据主导力量，却是整个协同过程中不可或缺的中坚力量。毕竟条目带头人的群体只占总用户数量的 6.6%，面对数量如此众多的条目，仅靠这一小部分人是无法完成的。这类用户是前一类用户积极的追随者和稳定的协同实践者，按照条目编辑的预定目标，最大程度地贡献自己的力量。最后一类用户是整个用户群体的低端用户，他们的平均贡献不超过 0.1。这意味着他们在条目编辑过程中所起的作用微不足道。甚至，平均贡献为负值的用户占到了总用户数的 20.5%。从数值上看，很难说这个用户群体究竟对维基百科社区起到了关键作用。但是，这个数量庞大的群体却是社区存在的坚实基础。首先，所有的高级用户均产生于这个用户群体。任何的高级用户必然要经历一个从初级用户到高级用户的过程，在这个过程中

不断增加对社区的了解，不断提升自己的技能和经验，最终才能带领其他用户共同建设一个繁荣的社区。只有拥有数量庞大的初级用户群体，才会不断地产生更多的高级用户。其次，这些初级用户所参与的工作大部分是维护性的工作，如修正条目中的错别字，改进内容的表达方式，补充内容来源等工作。这些工作虽小，但是对于改进读者的阅读体验却有很大帮助。因此，尽管这个用户群短期内对社区发展的影响很小，但是社区仍应将其视为社区发展的重要资源。

另一个分类维度是用户参与编辑的条目数量。这个角度反映了用户的参与和活跃程度。积极的用户往往不满足于参与有限的几个条目编辑中去，而是希望能尽可能地为更多的条目贡献自己的力量。用户越是活跃，越是积极参与，那么该用户所涉及到的条目也就越多。

同用户的平均贡献类似，用户参与的条目数量的分布也是极为不均衡。绝大部分在其加入社区的整个生命周期内只参与了一、两个条目的协同。图 1.4显示了参与条目数量在 10 以内的用户分布。

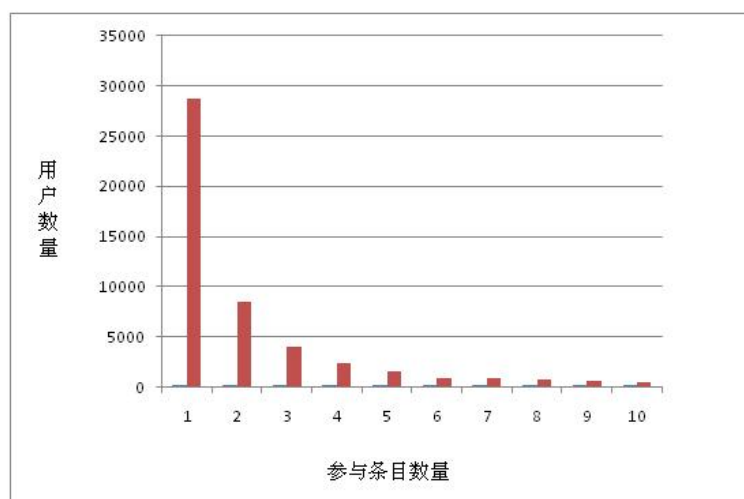


图 1.4: 用户参与的条目数量

从图中可以看出，只参与了一个条目编写的用户数量达到了 28749 人，占总用户数量的 52%。即有超过一半的用户处于这种极度不活跃的状态。而编辑条目数在 5 个以下的用户数共计 45187 人，占总用户数量的 81.83%。这一部分用户可以视为社区中的不活跃人群，他们因为各种原因没能找到发挥自己才能的条目，成为了沉默者。社区流失的成员大部分来自于这个群体。从用户的分布数量看，编辑条目数量在 1-5 之间的用户数量从最大值 28749 人急剧减少，随后，人数呈缓慢的下降趋势，由于数值太小，难以看出分布随后的变化趋势。图 1.5 进一步显示了较活跃的用户数量的分布。

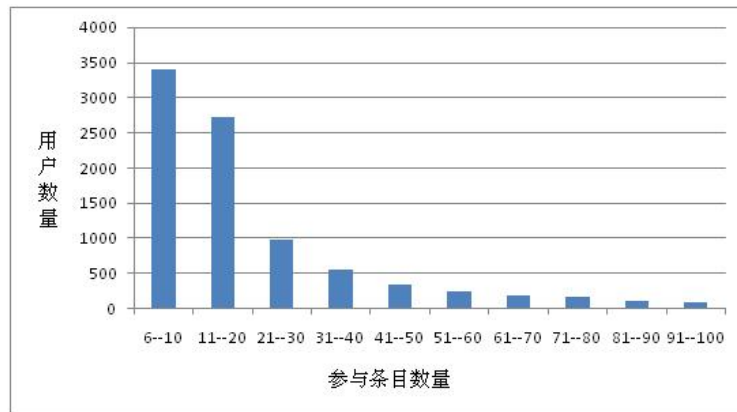


图 1.5: 用户参与的条目数量 2

图 1.5 也表现出类似的用户分布。用户数量先是显著下降，当到达 51–60 这个区间段后开始平缓下降。可以认为从这个区间段开始，用户表现出了非常显著和活跃的协同行为。当一个用户参与条目的数量超过了 50，表明该用户已经完全熟悉并掌握了社区的基本规则，确立了自身的定位，并以一种非常积极的态度参与到社区的协同活动中来。这一群体的用户是维基精神的积极践行者。尽管他们可能不具有很好的专业知识，不能引领每一个条目的发展方向，但是他们尽可能地发挥自身的优势，以巨大的热情为社区做出自己的贡献。在这类用户中还存在着是一些“超人”用户：有 10 个用户参与的条目数量超过了 10000，其中参与条目最多的用户达到了 23991 个条目。正是这类活跃用户的努力繁荣了整个社区。第三类用户是处于上述两类用户之间的“中间用户”。这类用户逐渐从不活跃的状态向活跃状态转变，开始有意识地寻找一些自己关心或是感兴趣的条目，试图从中发现可以贡献自身知识的机会。另一方面，这一类用户由于时间、精力等方面的因素，没能象活跃用户那样全情投入。尽管如此，由于人数上的优势（约为活跃用户的 4 倍），这类用户仍然是社区繁荣的可靠力量。

通过以上分析的两个分类维度以及每个分类维度下的分类原则，最终可以得到适用于分析用户协同行为及其特征的分类法。所有的用户按照该分类法都会分到一个恰当类别中去，分类的结果如表 2 所示：

初始的分类结果形成了 9 个分类，并且不同分类间用户数量相差非常大。由于本研究的目的是考察不同类型的用户参与知识协同的动机因素，因此形成的用户分类应该符合两个特点：1) 分类之间的界限明显，分类应该突出突出本类别用户的明显特点；2) 分类应该与时间关联较小。用户由于加入社区在时间上有早晚，因此用户特征会随时间变化，从一种类型的用户转变为另一种类型的用户。如果分类本身与时间的

表 2: 用户分类结果

分类	条目平均贡献	参与条目数量	人数	百分比
分类 1	高	高	61	0.11%
分类 2	高	中	224	0.41%
分类 3	高	低	3182	5.76%
分类 4	中	高	1027	1.86%
分类 5	中	中	2835	5.13%
分类 6	中	低	6866	12.43%
分类 7	低	高	1044	1.89%
分类 8	低	中	4839	8.76%
分类 9	低	低	35139	63.64%

关联度很高，则意味着该类别的用户转换速度非常快，该分类很可能只是用户的过渡状态，而用户的真实行为特征并不一定与分类特征项符合。基于以上原因，故考虑合并一些分类。合并的依据是，对于人数较少的分类合并到相似的分类中；对于区分度不够明显的分类合并为一类。经过进一步分析数据，分类 2 和分类 3 的界限并不明显，分类 2 的用户平均贡献度同分类 3 非常接近，而参与编辑条目的数量大部分集中 6–7 之间，仅略高于分类 3 的边界 5。同时分类 2 的用户数量非常少，故将分类 2 和分类 3 合并。对于分类 4、分类 5 及分类 6，这三类的用户有着相似的用户贡献度，而参与的条目数量却同时的关联很紧密，经过一段时间后会有相当一部分用户会从“低级”的分类向“高级”的分类转化，因此将分类 4、分类 5 及分类 6 合并。基于同样的原因将分类 7 和分类 8 合并。最终将维基百科的协同用户划分为 5 个类别。

1. 领导者。这类用户的特点是广泛深入地参与到社区的协同活动中去。不但参与了大量条目的编辑工作，而且对于每个条目都积极投入，是协同编辑主要的领导者和最重要的贡献者。这类用户所占的比例极小，但是却起到了引领社区前进和发展的作用。领导者是社区最重要的用户群体。
2. 领域专家。这类用户的特点是对某些领域的知识精通。对于该领域下的条目具有独立撰写、或者领导其他用户共同完成编写条目的能力。对于每个参与的条目，他们都能深入地参与并贡献高质量的内容。同领导者群体不同，领域专家参与社区活动的热情要小得多，只求做好自己擅长的工作即可。因此他们实际参与的条目数量都比较小。

3. 内容贡献者。这类用户是协同活动的积极参与者。尽管他们限于自身的知识水平和个人能力还不足以起到领导知识协同的作用，但是他们是领域专家和领导者的追随者。他们的工作对于补充、丰富条目的内容起到了重要的作用。对于前两类用户所忽视或者涉及不到的内容，都是由这类用户完成的。尽管由于时间和精力的原因他们参与的广泛程度不同，但是参与的动机基本是一致的。
4. 内容维护者。内容维护者的主要作用在于修补条目内容的疏漏，及时更新过时、无效的信息。对于每个条目，内容维护者参与的程度都不高，但是其参与的范围却很广泛。内容维护者自身并不具有太多的专业知识，但是他们通过维护内容的方式弥补了自身的不足，为改进条目质量和促进社区发展贡献力量。
5. 边缘用户。边缘用户是所有用户类别中数量最为庞大的群体。他们很少参与到协同活动中，即使偶尔参与协同也往往会由于经验不足而贡献低质量的内容，很快被回退或被其他用户修改。从主观上讲，边缘用户愿意参与到社区活动中，这是他们区别于破坏者和潜水者的显著特征。但是他们的意愿收到了某种客观条件的制约，一旦条件成熟，将促进边缘用户向更高级的用户转变。

不同类型的用户间的关系如图 1.6 所示。

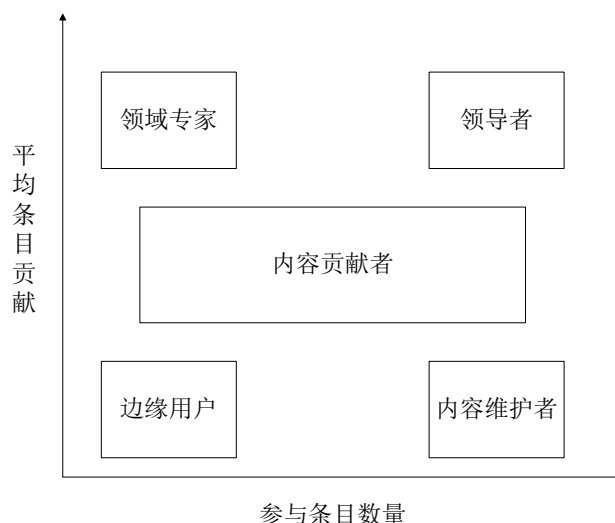


图 1.6: 不同类型的用户间的关系

## 1.2 用户协同的社会网络分析

至此，维基百科社区中的用户依据其贡献度和参与程度两个维度共分为 5 个类别。同一类型用户之间的协同以及不同类型之间的用户协同构成了社区中的所有协同活动。用户在协同中形成了用户间的复杂网络。运用社会网络分析手段，可以更为清晰地理解维基百科社区中用户的协同模式。

对于复杂网络的度量，往往使用图论来研究网络属性。由用户形成的复杂网络中，用户是网络中的节点，这些节点通过某种关系而连接起来，形成了网络中的边。如果两个用户共同参与编辑了一个条目，则认为二者进行了知识协同，形成了网络中的关联关系。

度 (Degree) 是复杂网络中最重要的度量之一。对于节点  $i$ ，定义度  $k_i$  为与该节点连接的其他节点的数目。在有向网络中还要进一步定义出度和入度。由于协同关系是双相关联，即如果用户  $A$  与用户  $B$  协同一定意味着用户  $B$  与用户  $A$  协同，因此用户网络实际上相当于无向图。在分析节点的度是主要考虑的指标有三个：节点的平均度  $\bar{k} = \frac{1}{N} \sum_i k_i$ ；节点的最大度  $k_{max} = \max k_i$  以及节点的度分布  $P(k_i)$ ，它是节点度的概率分布函数，表示节点  $i$  有  $k$  条边的概率。

集聚系数是衡量一个网络的集团化程度的重要的特征参数。在社会化网络中，如果两个人同为一个人的朋友，往往这两个人也是朋友。集聚系数就是为了表示网络中朋友圈的凝聚成度，描述这种集团化现象的属性。一个节点  $i$  的集聚系数可以定义为该节点邻接的节点之间实际的连接数量同可能的最大连接数量的比，即： $C_i = \frac{2E_i}{k_i(k_i-1)}$ 。其中  $\frac{k_i(k_i-1)}{2}$  表示于节点  $i$  邻接的节点间最大可能邻接数，而  $E_i$  是邻接节点间的实际连接数。这样，通过集聚系数，可以反映不同群体间用户关联的紧密程度，从而分析其协同模式。

领导者是所有类别中人数最少的，但是却是最为投入的群体。领导者共参与了 37939 个条目的编写，平均每个人参与了 622 个条目。该类用户另外一个特点是：用户参与的条目贡献均值很高，但是贡献的方差却很大。在参与的所有条目中有 41.7% 的条目贡献度超过了 0.8，几乎达到了单个领导者独立编写的程度；与之相对的有 39.2% 的条目领导者用户的贡献度不足 0.1。这个结果意味着领导者用户实际上存在着两种特征，除了自身用具有的特征外还同时具有维护者的特征。通过进一步的分析发现，领导者用户所做的维护工作同维护者所做的维护工作略有不同。领导者用户不是以消除

文字错误、更新信息等为目的，而是为纠正其他用户错误的或不适当的行为为目的。由于维基百科对用户参与的要求较高，除了必须有一定的独立撰写能力、遵从维基百科的编写规范，还必须熟悉编写系统和标记语言的用法。对于没有经验的用户来说很容易犯各种类型的错误。领导者于是承担起教育用户的责任，希望用户能在参与编辑的过程中不断提升自身的水平。尽管这类用户同时存在这两种特征，但是在接下来对用户参与知识协同活动的动机的研究中，将主要考虑领导者的特征，而不考虑其维护者的特征。

领导者用户的最大度为 42，但是平均度却只有 9.2，超过一半的用户度在 6 一下，说明领导者用户间很少有合作编写条目的经历，这和之前分析的领导者用户的特征是一致的。同时，领导者的聚类系数只有 0.17，远远小于整个社区的聚类系数 0.53。这说明用户之间并没有形成紧密联系的群体，一个用户也没有太多的意愿去参与由另一个领导者主导的条目。用户间彼此的协同更多地是在偶然的情况下发生的。

领域专家拥有和领导者类似的平均条目贡献，但是其参与条目的数量要小的多。领域专家共计 3406 人，只参与了 7424 个条目的编写，平均每个人不到 2.2 个条目。同领导者不同，领域专家的条目贡献离散程度要小的多，说明该类用户的协同模式非常稳定：每参与到一个条目中，就尽全力把工作做好；而对于其他条目则完全不理睬。

领域专家用户的最大度为 4，平均度只有 0.04，用户间的关联非常稀疏。网络的聚集系数更是趋紧于 0。这类用户比领导者用户还要独立，用户间的协同几乎不发生。另外，领导者和领域专家两类用户之间也鲜有协同发生。领域专家和领导者两类用户都属于能够主导条目编写的用户，同时有领导者和领域专家参与编写的条目共计 403 个，只占领域专家参与条目总数的 5.4%，对于领导者来说更是微不足道。如果把两类用户合成一类共同考察的话，在新形成的用户网络中，节点的最大度也只有 103，而平均度也只达到了 0.43，仍然是非常稀疏的连接。网络的聚集系数值为 0.03，显示用户之间几乎没有任何凝聚效应。

但是，这两类用户彼此之间很少发生协同行为并不意味着这两类用户真的是特立独行的。事实上，在有领域专家和领导者参与的条目中，参与协同的用户数量平均为 127 人/条目，远远高于社区的均值 46 人/条目。尽管这两类用户掌控了条目的编辑工作，但是似乎“独裁”并未影响用户的参与程度，反倒是由于这两类用户的积极投入，给其他用户带来了更多的参与机会去丰富内容和提升质量。

内容贡献者是所有用户中参与范围最广的群体，共参与了 184746 个条目的编写，

几乎占整个维基百科条目数量的 90%。平均每个用户参与编写 17.2 个条目。巨大的参与数量意味着该类用户同其他几类用户都具有较强的联系。其中，内容贡献者于领导者协作参与了 24954 个条目，占领导者参与总量的 65.77%；与领域专家协作参与了 5552 个条目，占领域专家参与总量的 74.79%。说明内容贡献者积极地参与了这两类用户领导的条目的编写工作。另外也应该注意到，尽管这两类用户都是维基百科社区的优质用户、但是其参与的条目总数只占社区中条目总数的 21.9%。社区的精英并不能完成所有的工作，还必须要依靠那些热心的普通用户配合。每个人的精力都是有限的。对于维基百科这样完全依靠参与者在业余时间自愿投入的社区来说，明星用户永远是少数，群体的力量才是真正可以依靠的力量。内容贡献者本身就承担了大量的内容编写工作，是整个社区中力量最大的群体。

同领导者和领域专家不同，内容贡献者群体内部的协同非常频繁和活跃。网络节点的最大度为 7109，平均度为 195，均远远高于其他两个类别。网络的聚集系数为 0.65，体现出了非常明显的聚集特征，意味着这一类用户的协同效应非常显著。尽管他们每个人的能力都是有限的，但是他们依靠集体的力量，相互合作，最终为社区做出自己的贡献。

维护者也是一个广泛参与知识协同的群体。该类用户共参与了 143863 个条目的编写，约占条目总量的 70%。平均每个用户参与编辑 24.5 个条目。维护者与内容贡献者的具有类似的网络特征。同上述几类用户的联系也非常紧密。其中，维护者于领导者共同参与了 21192 个条目，占领导者参与总量的 55.86%；与领域专家共同参与了 5410 个条目，占领域专家参与总量的 72.87%；与内容贡献者共同参与了 132539 个条目，占维护者参与总量的 92.13%。可以看到，维护者于社区中的主要成员关系都非常密切，尤其是和内容贡献者。

维护者群体内部的知识协同是所有类型用户中最为频繁和活跃的。其网络节点的最大度数为 5104，虽然不及内容贡献者的最大度，但是该类用户的数量只有内容贡献者的一半，所以其平均度达到了 320，为所有类别中最高。网络的聚集系数为 0.63，其聚集特征更为明显，成员之间的关系也更为密切。

边缘用户是人数最多的一类用户，但是却只参与了 27732 个条目的编写，平均没人参与的条目为 0.97 个，是所有类型用户中最少的。边缘用户与其他类别的用户协同则呈现明显的两极分化的特征。其于领导者共同参与了 4299 个条目，占边缘用户参与总量的 15.49%；与领域专家共同参与了 1270 个条目，占领域专家参与总量的



17.11%。而与之相对的，与内容贡献者共同参与了 27254 个条目，占边缘用户参与总量的 98.28%，与维护用户共同参与了 26134 个条目，占边缘用户参与总量的 94.24%。边缘用户几乎所有的协同都是和内容贡献者和维护者之间发生的，而与领导者和领域专家间的协同则非常少。这一方面是由于领导者和领域专家所参与条目的数量比较少，同时也是由于这两类用户领导的条目质量普遍较高，并没有给边缘用户太多的参与机会。而内容贡献者和维护者受自身条件的限制，给边缘用户留下了参与协同的空间。

边缘用户之间基本上没有什么直接联系，只有少数人具有较高的度。最大度为 151，但是有近 1/3 的用户度为 0，也就是没有和任何其他边缘用户间发生过协同，而网络平均度为 9，表明连接是非常稀疏的。边缘用户的聚类系数为 0.09，证明用户之间非常独立，没有聚集特征。

通过社会网络分析，本文分析了维基百科社区中 5 类用户的协同行为和其特征。可以明显地看出社区中主要存在两种形式的协同：一种是以领导者或者领域专家为主导，内容贡献者和维护者参与辅助性工作最终完成条目的编写；另一种是参与编写的条目中没有真正的主导者，而是由多个内容贡献者和维护者通力合作，利用集体的力量共同完成的。这个结果同时也回答了人们一直以来对维基百科如何取得成功而产生的疑问：究竟是“少数人的力量”还是“多数人的智慧”。一部分学者认为如此数量巨大的内容不可能仅仅由少数人来完成，必然是很多人齐心协力的结果；而维基百科的创始人 Jimmy Wales 则以英文维基百科为例，列举了用户参与的数据：“有 50% 的编辑是由不到 0.7% 的用户完成的，最活跃的 2% 的用户，约 1400 人，则贡献了总编辑数量的 73.4%。本文的研究结果表明，两种协同模式的存在使得维基百科既体现了“少数人的力量”，也体现了“多数人的智慧”。少数明星用户是社区最积极的推动力量，并取得了非常耀眼的成就，但是这并不意味着其他用户就处于陪衬的角色。不同类型的用户通过不同的协同模式为社区贡献力量，最终才造就了维基百科的成功。

### 1.3 本章小结

本章利用上一章研究的用户贡献算法，对维基百科中的数据进行整理和分析，并从用户贡献和用户参与程度两个维度对维基百科社区中的用户进行了分类。分类经过调整和合并后共分为五类：领导者、领域专家、内容贡献者、内容维护者以及边缘用户。随后分析了这五类用户的特征，并使用社会网络分析研究了不同用户的协同模式。分析结果表明，这五类用户各自都有独特的协同模式，因此其参与协同的动机可能各

不相同。在随后的章节中将分别讨论各个动机因素对每种用户的影响，构建每类用户的动机模型。