

从开源科技的数字化洞察看开源教育的未来

作者：王伟、赵生字



导语：本文为 X-lab 开放实验室在《新程序员》第五期（开源深度指南&新金融背后的科技力量）上刊出的一篇文章，和大家一起交流学习。开源以开放、分享、对等、协作、以及全球化运作的方式，正在深刻影响企业的商业行为与治理模式，而开源科技的飞速发展也给开源教育带来了巨大的机遇与挑战。本文通过介绍 X-lab 开放实验室在开源全域数据分析与洞察的研究经历，并结合作者实际开设的开源课程，给出若干开源教育发展的具体建议，希望为国内开源生态的持续健康发展提供启发。

一、开源协作的趋势与整体现状

开源软件已经成为人类数字社会的基石，是全人类共同努力的结晶，开源协作对人类数字文明的发展起到了巨大的推动作用。基于 **Git** 的分布式协作成为全球范围内最主要的开源创新模式，无数个开源社区在其上孕育而生，其背后海量的开发者行为数据蕴含了大量的个体贡献规律、群体协作模式、社区健康状况、生态发展趋势、以及商业战略价值。

2020 年突如其来的新冠疫情，加速了开源协作的全数字化进程，以 **GitHub** 平台为例：

- **GitHub** 的日志数在 2020 年达到了 8.6 亿条，相较 2019 年增长了 42.6%，而 2021 年更突破了 10 亿条，足以凸显疫情期间爆发的巨大生产力；
- 2020 年活跃代码仓库达到了 5,421 万个，相较 2019 年增长了 36.4%，而 2021 年达到了 6,700 万个（增长率为 23.6%）；
- 2020 年活跃开发者数达到了 1,454 万人，相较 2019 年增长了 21.8%，而 2021 年达到了 1,748 万人，继续保持 20% 以上的增长速度。

这类趋势在国内诸如 Gitee 平台也非常一致，甚至更为显著：Gitee 新增用户数超过了 180 万，达到 800 万，新增活跃的仓库数超过了 200 万，开发者贡献活跃度增长了 102%。

全球开源协作的活跃，带动了整个开源科技产业的发展^[1-5]，部分研究要点摘录如下：

- 全球开源事业大发展，社区活跃行为、开发者数量、开源仓库数量均大幅提升；
- 开源软件生产流水线自动化程度大幅提升，多样化的数字协作机器人成为主流；
- 美洲开发者分布最多，欧洲拥有最高的单时区开发者比例，亚洲开发者数量依然较少，中国相较其他亚洲国家具有较高的开源活跃度；
- 谷歌、微软等老牌企业依旧为活跃的开源贡献大户，国内企业中，阿里巴巴活跃度排名第一，PingCAP 的表现非常亮眼；
- 主流技术领域的开源生态已经形成，新的开源社群也不断涌现，然而，极少量项目还处在协作孤岛；
- CNCF、LF、Apache 等基金会在技术领域各有侧重，通过开源象限等分析工具，能够进一步区分同类项目的发展阶段与成熟水平；
- 开发者时区分布图和协作网络成为开源社区多样性与健壮性的有效分析手段，能够更好地指引社区经理的开源治理与运营工作。

具体的详细分析与内容，大家可以参考我们发起和参与的几个主要报告^[1-5]。也正是得益于这些研究工作，使得我们成长为一个极具特色的开源共同体。

我们所在的 X-lab 开放实验室作为一个开源研究与开放创新的群体，长期思考并实践开源战略、开源测量学、开源数字生态系统等主题，并在每年联合多家科研机构与开源社区共同发起、制订反映全球开源现状与趋势的各类指标与洞察报告^[1-5]。这些报告涵盖了当今全球开源的宏观、中观、微观等不同层面的分析，如图 1 所示。

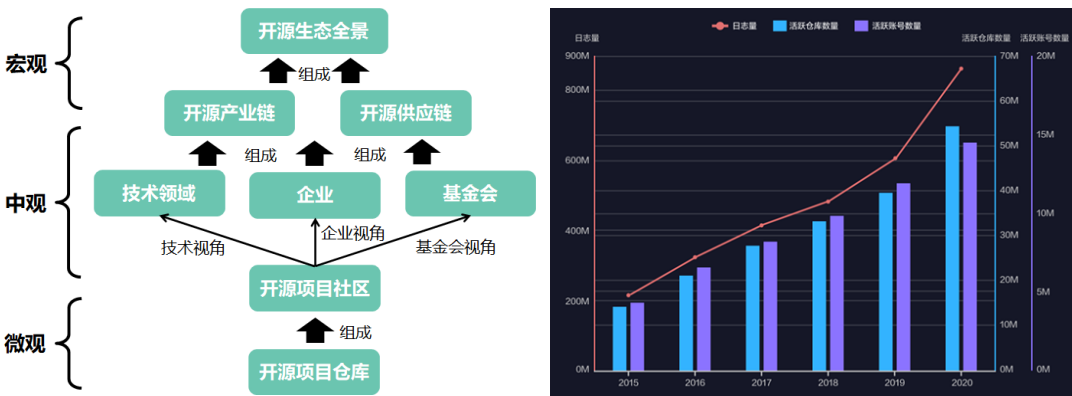


图 1: 开源洞察框架（左）以及 2020 年度全球开源数据总体分析（右边）

我们将正在实践的这套方法称之为“面向开源数字生态的开源测量学”，本质上就是用数据的方式来研究开源，通过采集能够反映全球开源协作的各类行为数据，结合开源生产、开源治理、社群运营等关键活动中的专业知识，利用数据科学的手段方法，对开源数字生态进行深入分析与洞察，最后开发各种数据与信息工具，为政策制定者、企业决策者、社区管理者、开源贡献者等不同角色提供服务。目前已经在包括开源治理与社区运营的标准、国内部分企业与社区的开源项目健康评估、国际开源度量指标体系的建设等方面做出了一定的成果与影响力。而这些研究经历与成果，又可以很好的转化到开源教育的课程当中，结合课程内容与教学工具的开源方法，形成了一个“产-教-学-研”的深度融合。接下来的两节，就分别从开源研究与教学两个方面介绍。

2、数据驱动的开源测量学研究

管理学大师彼得·德鲁克曾经说过：“你如果无法度量它，就无法管理它”（If you can not measure it, you can not manage it），进而也无法提高它，而软件行业至今也还没有找到一个可以有效度量软件开发生产效率（Productivity）的方法。

整个开源生态系统更是如此，个体如何度量、社区如何度量、管理者如何利用这些数字做更好的决策，这些都是问题。但在我们看来，这些既是挑战，也是机遇。要想高效的发展开源生态建设、开展开源社区治理与运营等工作，特别是促进大规模全球化的开源创新协作，科学合理的度量与指标体系非常关键。

度量也是一把双刃剑。度量具有极强的引导性，它会激励你重视并改善能够度量的元素，但也可能使你忽视无法度量的元素并使之恶化。在全球大规模的开源社区与生态构建的过程当中，如何找到合理的度量，并合理的利用这些度量呢？这些都是非常值得探讨的问题。

基于此，X-lab 开放实验室基于我们长期在开源生态数据测量与分析方面的积累，对开源生态进行了系统性的洞察与分析，并系统性的提出了几个关键指标（指数），来帮助大家理解开源社区的情况，整个过程实际上也是我们实验室的一次开源教育之路，并且还在不断前行。

（1）活跃度

活跃度指标大家普遍达成共识的指标：

- 站在开源办公室的角度，必须有一个“北极星”指标。一些开源创业公司或具体的开源项目团队，可以通过监控或观察多个指标来判断项目健康与否。但对于有两千多个开源项目的企业而言，要同时监控项目的健康度，就需要一个聚合的指标，否则人力成本过高。
- 由于 star、fork 等行为属于开发者的单向行为，虽然表示了对项目的一种关注，但并不对项目产生具体贡献，所以没有纳入到活跃度的计算之中，即刷 star 等行为在活跃度算法下无效。

- 即便当时默认的贡献者（contributor）的定义为代码贡献者，但从实际角度出发，参与到社区中的所有开发者，包括提交 bug、参与讨论、参与代码 review 的开发者事实上都对项目是有贡献的，所以在计算中并不是仅计入代码贡献，而是将讨论等也纳入。
- 这些事件对应的权重应该是多少，这个需要在一个组织中战略级开源项目的负责人共识下产生。例如，阿里就对 PR 的 review 给出了非常高的 4 分，也是鼓励大家多进行基于 GitHub 的异步 review。

这个活跃度的计算方式是一种简单的基于贡献者行为数据的加权统计算法，权值属于专家经验赋值，带有一定的价值导向。并且从开发者活跃度到项目活跃度的计算也具有一定的价值导向，即间接的将贡献者数量作为一个重要因素引入到项目活跃度中。

活跃度指标的一个最直接的应用就是构建活跃度指数排行榜，帮助大家直观了解国内外项目、企业、开发者的整体现状。例如，下面两张图即为我们在《中国开源年度报告》[4]以及《中国开发者十年》[3]中提供的分析结果。

排名	项目名	活跃度	参与开发者数量	issue comment	open issue	open pull	pull review comment	merge pull	star	fork
1	pingcap/tidb	210.1	5,831	53,022	2,801	4,969	10,928	3,459.2	4,862	1,052
2	ant-design/ant-design	193.3	23,620	32,026	4,836	3,131	3,320	2,130.7	12,709	8,052
3	PaddlePaddle/Paddle	127.4	4,842	15,329	2,256	5,656	9,625	3,478.2	3,574	786
4	tku/tkv	81.7	2,593	17,817	997	2,019	5,547	1,279.9	2,129	434
5	apache/shardingsphere	75.3	5,267	9,055	1,713	3,235	1,858	2,539.5	3,834	1,443
6	apache/incubator-tvm	70.4	2,148	7,961	437	2,112	8,506	1,540.1	1,454	662
7	pingcap/docs-cn	65.1	532	8,202	96	2,965	6,959	2,315.9	140	320
8	apache/incubator-echarts	64.2	11,638	7,650	1,620	324	346	194.5	6,664	4,463
9	pingcap/pd	60.9	437	13,325	667	1,667	4,972	1,297.7	214	224
10	alibaba/nacos	59.9	9,956	7,042	1,640	706	827	410.0	6,347	3,450
11	NervJS/taro	54.7	7,469	9,339	2,231	917	135	551.5	5,250	1,012
12	youzan/vant	54.2	9,806	4,897	1,661	715	201	554.4	4,672	4,502
13	pingcap/docs	53.9	314	7,014	64	2,736	5,226	2,257.8	90	164
14	ElemeFE/element	52.7	11,749	4,993	1,762	297	10	33.3	6,853	3,411
15	apache/skywalking	51.9	5,556	6,783	1,084	860	3,455	583.4	4,201	1,471
16	PaddlePaddle/PaddleOCR	47.9	9,394	4,039	1,033	573	622	420.0	8,430	1,664
17	apache/incubator-dolphinscheduler	47.1	2,588	9,364	1,269	1,407	730	902.7	1,835	909
18	apache/apisix	45.4	2,923	5,855	1,109	1,029	3,383	715.0	2,496	579
19	seata/seata	45.1	7,339	3,754	785	517	1,805	313.5	5,261	2,296
20	pingcap/tidb-operator	45.1	425	8,627	703	1,498	3,683	1,172.1	240	140

GitHub 2020 中国项目活跃度 TOP 20

2012	2013	2014	2015	2016	2017	2018	2019	2020	2021
Alibaba	Alibaba	Alibaba	Alibaba	Alibaba	Alibaba	Alibaba	Alibaba	Alibaba	Alibaba
Netease	Baidu	Baidu	Baidu	Baidu	Baidu	Baidu	Baidu	PingCAP	PingCAP
Xiaomi	Netease	Linux China	PingCAP	PingCAP	Tencent	Tencent	PingCAP	Baidu	Baidu
Douban	Tencent	Netease	360	Tencent	PingCAP	PingCAP	Tencent	Tencent	Tencent
Tencent	Xiaomi	Xiaomi	Bilibili	Juejin	Juejin	Juejin	JD	JD	Huawei
Meituan	Linux China	Tencent	Xiaomi	Bilibili	Bilibili	Meituan	Huawei	Huawei	QingCloud
Deepin	Douban	Douban	Tencent	360	360	JD	DiDi	QingCloud	JD
Baidu	Meituan	Meituan	Linux China	DCloud	Meituan	Youzan	Juejin	Youzan	Vesoft
	360	360	Douban	Xiaomi	Huawei	360	Youzan	Vesoft	Bytedance
	Deepin	Deepin	Netease	Netease	Youzan	Huawei	Meituan	DiDi	Juejin

图 2：基于活跃度的项目排行榜（左）以及国内企业排行榜（右边）

但活跃度的计算也存在一些显而易见的问题，例如：

- 到底哪些数值需要被纳入，例如 star、fork 是否应该被纳入，是不太确定的。尤其是在活跃度这样一个概念下，那些对项目没有带来实际反馈的行为是否要纳入进来极为微妙。

- 不同行为的权重是人为指定的，虽然包含了一定的专家经验，但这些数值的大小其实是还是具有相当的主观性，尤其是在项目之间比较时，权重的微小差别就会带来一些总体活跃度的波动。
- 只要是简单的统计指标来进行计算，就一定无法避免刷指标的行为。

虽然活跃度依然存在许多问题，尤其是在持续的大规模运算和可扩展性上的问题。但由于其直观、易理解，可解释性强，事实上依然是我们实验室目前广泛使用的一种计算方式。并且也已经在很多项目中有落地，但我们也还是希望可以有更好的指标体系和算法框架，来更好的利用开源生态和网络来对项目做出更加有效的衡量，这就是协作影响力指标。

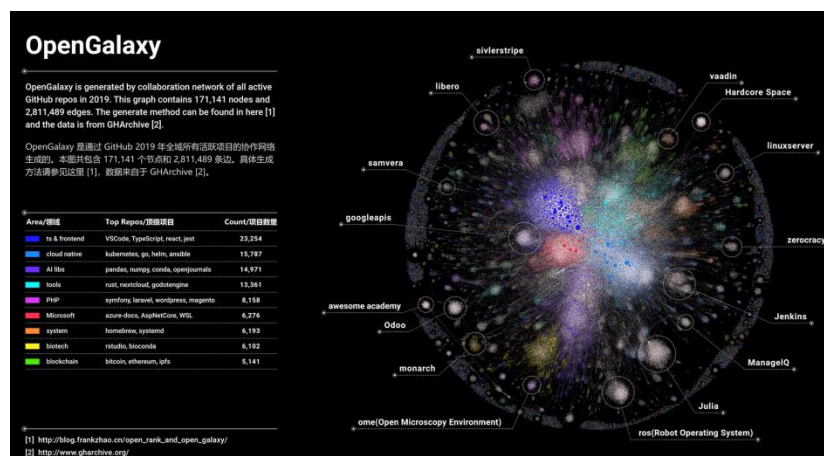
（2）影响力

开源协作网络的构建思想非常朴素，基本逻辑是：如果有开发者同时在两个项目上都非常活跃，那么这两个项目就存在着较高的协作关联度。在这里我们暂不考虑这种关联产生的动机，只是一种对开发者行为的观察。事实上根据后续的采样，发现基本符合最初的想法。即大部分情况下，是由于项目存在着上下游的关系，或存在某种使用上的依赖或合作关系，才会有开发者同时在两个项目上高度活跃。

因此，我们构建了开源全域生态中所有项目的协作网络，并利用一些图分析的算法（如PageRank）来计算每个项目的协作影响力，其背后的思想非常类似：即一个影响力较大的项目，会和更多的项目有协作关系；对于影响力较大的项目，与其协作关联度较高的项目影响力也会较大。

基于此，我们提出了两个分析工具来分析开源科技下的不同技术领域的发展趋势。

开源星系（OpenGalaxy）作为一个开源生态的宏观分析工具。例如，GitHub 2020 最活跃的 22.1 万个开源项目组成的协作网络图，该图中节点的大小表示项目的影响力大小，节点的着色表示节点所属的协作聚类结果，同时可以得到 GitHub 2020 全域影响力最高的项目 Top 20，如图 3 所示。



#	项目名	影响力
1	microsoft/vscode	977.3
2	flutter/flutter	593.3
3	tensorflow/tensorflow	484.8
4	microsoft/TypeScript	448.5
5	DefinitelyTyped/DefinitelyTyped	446.7
6	microsoft/WSL	431.7
7	golang/go	404.6
8	gatsbyjs/gatsby	381.3
9	kubernetes/kubernetes	370.1
10	vercel/next.js	354.5
11	pytorch/pytorch	340.2
12	NixOS/nixpkgs	335.4
13	rust-lang/rust	328.9
14	home-assistant/core	317.9
15	Homebrew/homebrew-core	311.5
16	microsoft/terminal	301.1
17	nodejs/node	292.9
18	MicrosoftDocs/azure-docs	285.5
19	electron/electron	276.5
20	facebook/create-react-app	276.3

图 3: GitHub 2020 开源星系（上）全域影响力排名项目 Top 20（下）

可以看到，VSCode 的影响力第 1（活跃度排名中则为第 5），且高于排名第 2 的 flutter 约 64.7%，以巨大的优势成为全球最具影响力的项目。事实上这是由于 VSCode 在成为全球最流行的 IDE 的同时，也与其他各领域的顶级项目产生了大量的协作关联。

可以看出，相对于活跃度，影响力是一个图数据指标，更能客观全面的体现一个项目在整个开源生态下的重要性。以国内比较热门的一个开源数据库项目为例（TiDB），我们分别绘制出了相应的活跃度与影响力曲线。从中可以看出，活跃度的波动性较大，特别是在最后横坐标 70 附近的地方，由于部分数据的缺失，出现了断崖的情况，但在影响力曲线中则不存在，也显示了影响力指标的健壮性与稳定性。

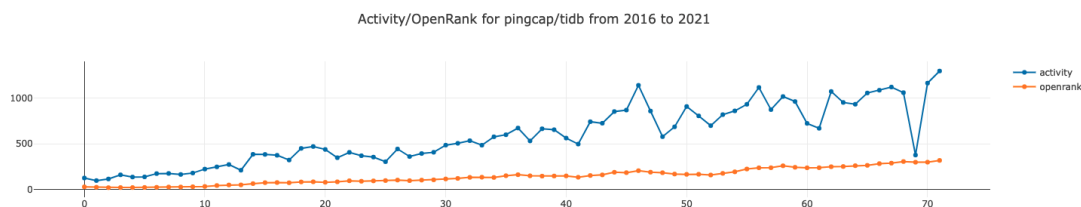


图 4: TiDB 项目在活跃度与影响力下的表现比较

开源象限（OpenQuadrant）是我们提出的第二个分析工具，可以将我们提出的活跃度、影响力指标进行象限组合，或者通过其他指标进行组合，如关注度、流行度等。

这个工具可以非常有效的分析某一个具体领域的开源项目发展情况，例如我们将影响力指标和国际上知名的流行度指标进行组合（来自著名的 DB-Engine 网站），进而可以得出数据库行业细分领域的直观可视化信息。如图 5 所示，横坐标的影响力代表了生产端，而纵坐标的流行度则代表了消费端，能够更加全面的刻画一个开源项目的发展趋势。

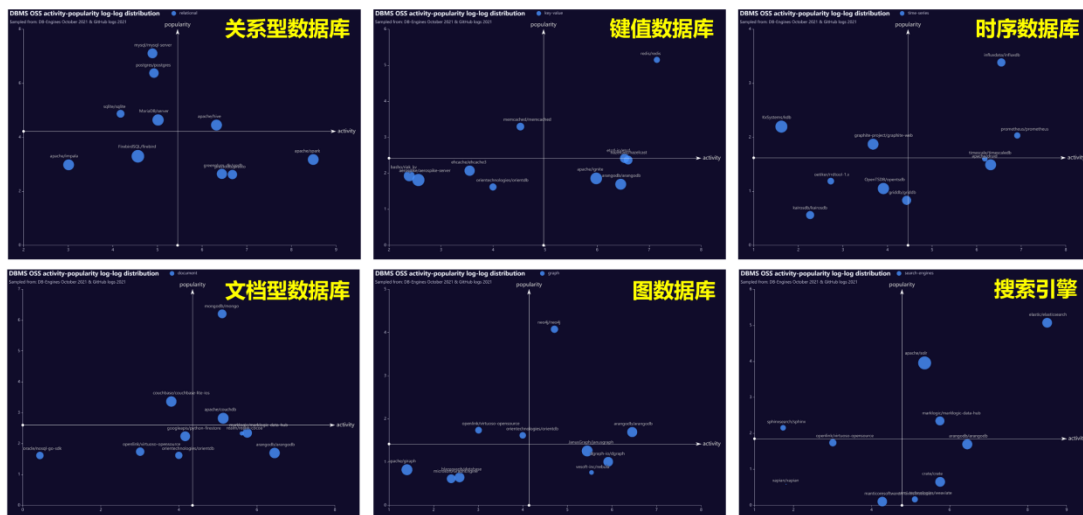


图 5 数据库细分领域的开源象限分析

同样的方法还可以套用到其他领域中，例如如图 6 所示的大数据与云原生领域。

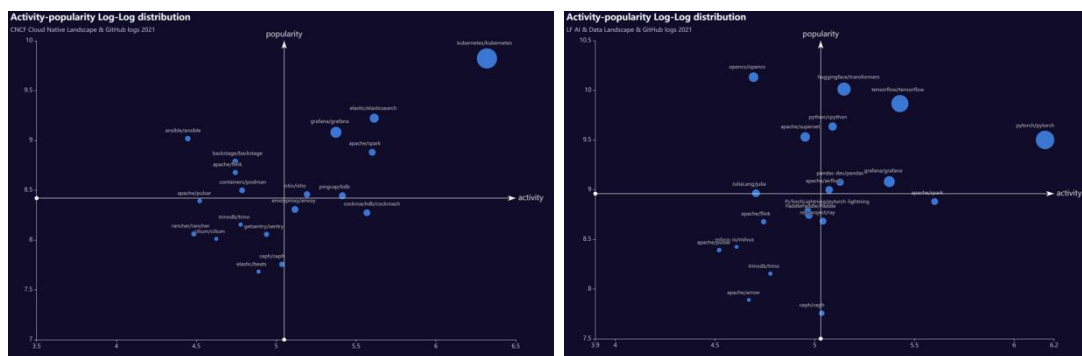


图 6 大数据（左）与云原生（右）领域的开源象限分析

协作影响力基于活跃度构建，但同时又规避了很多活跃度会带来问题，并且利用了开源生态的全域数据所蕴含的一些重要的关联信息。期中带来一个巨大的好处就是对于刷分行为，在这种模型下几乎无法生效。即对自己项目的高活跃的刷分无法带动自己项目的影响力，除非有更多的其他生态项目的开发者在你的项目中活跃。

之前提到的不使用 GitHub Issue 功能的项目，由于其开发者不仅在该项目上，也会在其上下游项目中高度活跃，导致了这些项目相较于活跃度指标表现出了更好的影响力。由于使用了网络关系这种重要的信息，原始的活跃度计算中权重人为指定的影响就会变小。只要还是满足大致的价值判断，即代码贡献大于 review 大于一般问题讨论，则活跃度中权重的变化几乎不会影响到影响力的排名。因为影响力中活跃度虽然是一个基础数据，但协作网络的结构信息使得整个算法具有了更好的稳定性和鲁棒性。

总体而言，基于开源协作网络的项目协作影响力解决了基于统计的活跃度指标中存在的诸多问题，对整个开源生态中项目影响力的评估和洞察提供了一种非常有效的手段。

我们现在正在做的一个最新的指标是价值流网络指标，希望可以在解决协作影响力无法容纳更多数据，从而可以更全面衡量开源生态的同时，也引入一种高可扩展的数学模型，可以在任意时间快速容纳更多元的数据，而不会导致模型的大幅改动。这里就不列出，读者可以参考[8]，等有了实际的落地案例，我们再来详细介绍。

三、从开源研究到开源教育的实践

长期在高校的科研经历告诉我们，好的教育一定是需要将前沿课题研究与产业发展动态进行有机结合的。开源教育是关于开源人才培养的，需要什么样的人是由需求端所决定的，即参与开源生态建设的相关企业、基金会、国际组织等。开源不止于技术，还有战略、治理、运营等重要话题，还需要国际化视野与全局意识，这些对于高校在开源人才上的培养是个巨大挑战，也是目前高校开源教育主要还是侧重要开发技术端的原因，即便是开发技术端，也只有为数不多的几所学校在开设相关课程。

基于上述认识以及我们在开源研究和度量方面的实际经验，我们实验室得以有机会和众多开源组织与开源社区进行深度接触，进而收集到一手信息，整理成可以进行教授的课程内容。开源已经作为一个关键词写进了包括十五规划等多个重量级的政府文件中，备受业界关注。中国开源的发展离不开开源人才的培养，也是我们一直在关注与实践的事情。

我们在华东师范大学所开设的这门开源课程叫做《**开源软件开发与社区治理**》，覆盖面还是非常广泛的，内容也不少，并且我们今天更是会从一个跨学科的角度去思考。我们觉得开源的跨学科属性是一个非常重要的特点，也直接影响着我们对课程的建设，目前的几个跨学科结合点包括开源软件、数据科学以及工商管理，如图 7 所示。

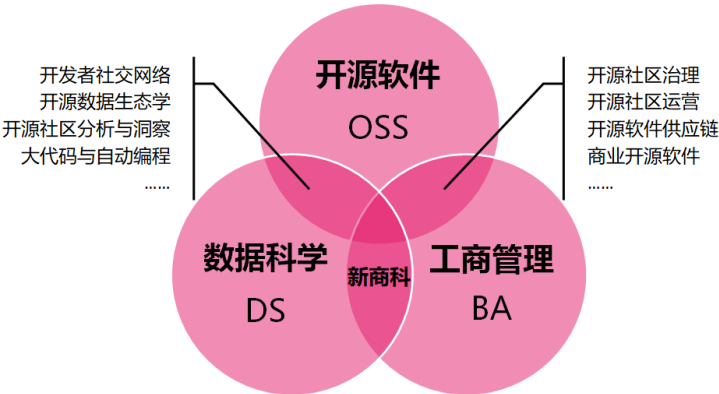


图 7 开源的跨学科研究视角

要建设开源课程，首先要弄清开源究竟包括什么，业界需要什么杨的人才？这样才能准确的设置课程目标。有了目标，就可以来思考教学中究竟应该教授什么？实践什么？因为开源一定是一个实践性比较强的领域，它背后的跨学科理论体系的建设也才刚刚开始。

纵观人类的协作历程，人类的发展实际上就是一个逐渐走向全球化协作的过程，而开源就是在这个趋势下形成的一种专业而先进的协同生产方式。我们认为这是目前全球开源进程

大发展背后最大的一个驱动力，因为任何一个组织都是需要追求诸如社会价值与组织效率这类事情的，背后有着非常深厚的经济学、管理学、社会学原理。

从开源产业发展的角度来看，开源战略、开源治理与开源运营是三个重要关键步骤。而一个组织的开源战略，从一个更普世的意义上来讲，就是要提升这个组织或社区的生产力、创新力、竞争力、影响力。因此，从一个企业的视角来看，我们提出了一个开源参与度模型，包括开源技术使用、开源项目贡献、开源社区运营这三件事，这也就是目前比较流行的一个 OSPO 所需要面对的事情，既包括对内也包括对外的。参照国际上 TODO、欧洲 OSPO 联盟等组织的相关工作，我们在标准院牵头了《开源社区治理与运营》的标准制定工作，并试图提出一些框架来落地为国内的企业与组织服务，如图 8 所示。

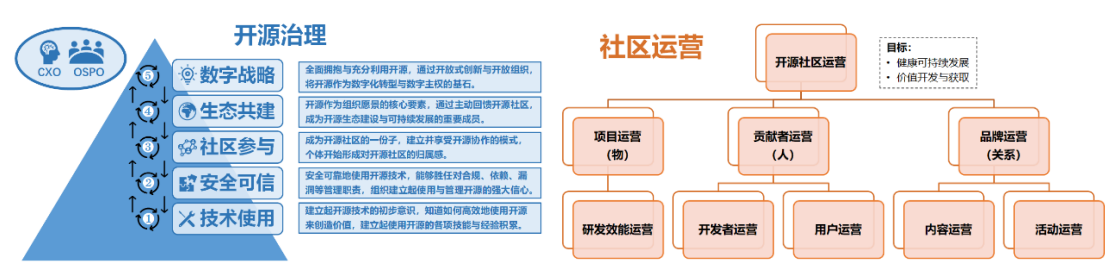


图 8 开源治理与社区运营框架

《开源软件开发与社区治理》这门课借鉴了上述所有的这些经验，同时没有对于特定开源软件或技术的介绍，而是将课程目标定位：为从多学科视角全面认识开源并掌握开源工作流程和一定的开源社区治理分析能力，课程整体内容如图 9 所示。

课次	单元	具体内容	课次	单元	具体内容
01	开源简史与概览	开源启示录、开源的缘起与革命、开源的成熟与机遇、开源究竟是什么？	07	开源治理与社区运营概览	开源战略分析、治理开源项目、运营开源社区
02	多学科视角	历史视角、新制度经济学视角、管理学视角、其他跨学科视角	08	安全可信的开源体系	开源许可证、安全合规与可信的重要性、安全可信合规治理实践
03	软件与开源商业化视角	软件商业简史、Linux 与商业化、开源软件商业化模式、未来开源产业链展望	09	开源社区数据分析	为什么要做数据分析、数据与可视化、开源社区数据分析中的问题以 VSCode 为例
04	软件工程视角看开源	软件工程 1.0、软件工程 2.0、软件工程 3.0、开源软件开发过程	10	开源社区数字化分析洞察	观测、理解、改进、组织与组织管理、开源社区数字化运营开放式组织的未来
05	开源协作与工程全流程	开源协作理解、Git 原理、GitHub Issue 与 PR、深入理解协作流程	11	GitHub 开发者行为数据	GitHub 概览、GitHub 数据结构、GitHub 行为数据分析、异质信息网络
06	DevOps 流程与 CI/CD 构建	DevOps 简介、DevOps 流程实例、CI/CD 构建、CI/CD 实例	12	开源社区度量指标与分析	指标度量、CHAOS 介绍、度量模型、社区分析实例
※	开源协作大作业	Git 协同翻译	※	社区分析大项目	开源社区综合分析项目

图 9 《开源软件开发与社区治理》课程内容结构

本次的课程本身也采用了开源托管的方式，课程的大纲、课件、考核作业提交等均托管于 GitHub[9]，最终所有选课的 63 名研究生同学，我们组织了 6 名助教和 2 名老师全部成为该仓库的贡献者，成功迈出了走向开源世界的第一步。课程中还利用了多款有实验室主导的开源项目作为教学工具[10-13]，取得了非常不错的效果。

而在作业协作方面，我们进行了一个精心的设计，即让同学们直接 fork 原始课程仓库，但只能由组长从组长的仓库提交 PR 到课程仓库，且仅可提交一次 PR，该 PR 中必须包含全

组所有同学提交的 **commit** 记录。这个设计下，同学们的仓库在 **GitHub** 上是对等且没有上下游关系的，迫使同学们必须在 **Git** 层面进行协作，从而可以更好的理解 **Git** 和 **GitHub** 的不同，并且可以深入的理解和实践 **Git** 诞生之初的去中心化设计哲学。

期末考核方面，配合课程第二部分的授课内容，选择让同学们对自己感兴趣的开源社区或项目进行深入的调研分析。分析工作将包含两部分的内容，即定性分析和定量分析。定性分析是对调研的社区或项目的治理规则、社区流程、**CI/CD** 等流程进行深入的调研和整理，而定量分析则是对于开源社区或项目的协作行为数据进行统计分析和可视化，通过这样方式让同学们对于自己的感兴趣的社区有一个全面的认识 and 了解。

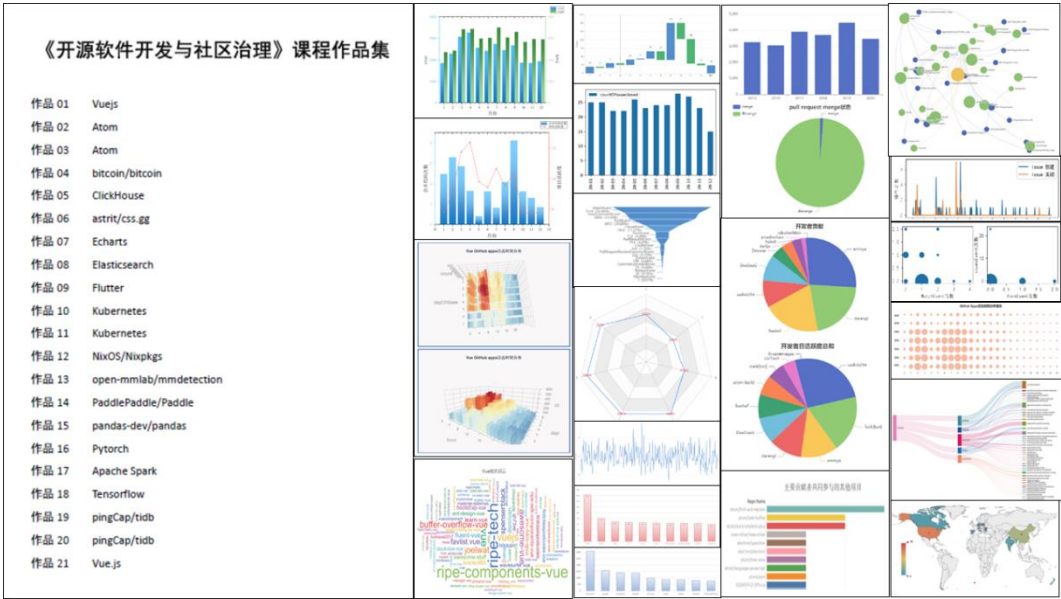


图 10 《开源软件开发与社区治理》课程作品集

4、开源教育的未来发展

可以看出，虽然我们在开源课程的建设方面迈出了第一步，但离整个行业的需求差距还比较大，特别是开源作为上至国家下至企业的战略方向、作为组织的重要创新手段、作为个体的有效职业发展路线等方面，越来越成为一个共识的情况下，针对不同需求层面的开源教育成为一个迫切的需要。

目前，国内的开源教育主要面临着三个方面的挑战：

- 第一，开源人才培养是一种综合性培养模式，并不是单靠传统的教材、课程就可以实现，而是要结合开源社区的具体实践，代码的贡献、动手实践的课程非常重要。并且，开源教育涉及的范围非常广泛，如今的软件工程也正在向开源的软件设计、开发演变。
- 第二，比较大的挑战是师资，教育最重要的就是师资资源，老师首先要懂如何参与开源项目、如何运营开源社区，这样才能更好的为学生服务。目前在这一块也比较稀缺，一方面有开源经验的老师比较少，另一方面开源本身也需要教育模式方面的创新。

- 第三，是学生，需要告诉学生学习开源的目的、价值，并吸引学生学习开源。新的技术从市场需求传递到校园，再到落地成一门课程，这个过程会有滞后，形成一门系统性课程也需要时间周期。

我们认为开源不应只成为专业教育，只放在计算机、信息类的学科下面，因为开源背后还有开源协作、开源精神，有很多思想方面的内容。我希望未来开源教育可以发展成为全民教育，每一个数字时代的公民都应具备开源方面的知识和能力。特别是数字化协作，相信这是和每个人息息相关的，不管你是否从事开发相关的工作，未来都需要用到开源相关的技能，例如我们现在会做一些内容翻译工作，可以基于开源协作来完成，开源教育包含了许多内容，对于每个数字公民来说都是非常重要的。所谓，教人“用”开源，以及用开源“教”人，如图 13 所示。



图 13 开源教育全景图

总而言之，我们倡议开源教育从以下六个方面展开：

- **学技术**：只开源，不商业，支持国创；
- **做贡献**：守规则，重流程，开放协作；
- **搞运营**：搭社区，建生态，共识治理；
- **教学资源**：课程开源，工具开源，平台开源；
- **学习方法**：项目制学习，协作式学习，社会化学习；
- **办学模式**：开放办学，开放思维，开放未来。

结合目前国家在国际开源生态建设方面的诉求，以及在教育改革方面的各项措施，开源教育的几个重点发展方向包括：

- 将我国所主要参与或主导的开源技术与项目，充分的融合到已有的计算机类课程教学内容中，吸引广大师生了解、使用、反馈、贡献这些技术与项目；

- 结合国家特色化示范性软件学院、现代产业学院等计划，大力发展新一代软件产业的专业化人才，培养具备开源素养与技能的各类工程师人才；
- 高校设置跨学科、多学科人才培养项目，特别是积极将融合管理学、经济学、法学、社会学等学科与软件人才培养相结合，重点培养开源战略、开源治理、开源运营等方面的高端急需人才；
- 将开源教育与全民数字素养与技能提升等计划进行结合，将开源与通识教育进行结合，培养具备数字化协作交流、开放创新、国际化视野的新一代数字人才。

参考文献：

1. 《GitHub 2020 数字洞察报告》，<http://oss.x-lab.info/github-insight-report-2020.pdf>
2. 《2021 中国开源发展蓝皮书》，<https://gitcode.net/2021blue-book/china-open-source-blue-book>
3. 《中国十年开源洞察报告》，2021 云栖大会。
4. 《2021 中国开源年度报告》：<https://kaiyuanshe.cn/document/china-os-report-2021/>
5. 中国开源码力榜：<https://opensource.win/>
6. 活跃度指标：<http://blog.frankzhao.cn/how-to-measure-open-source-1/>
7. 影响力指标：<http://blog.frankzhao.cn/how-to-measure-open-source-2/>
8. 价值流网络：<http://blog.frankzhao.cn/how-to-measure-open-source-3/>
9. 开源软件开发与社区治理：<https://github.com/X-lab2017/OSSDevGov2021>
10. OpenLeaderboard：<https://open-leaderboard.x-lab.info/>
11. OpenDigger：<https://github.com/X-lab2017/open-digger>
12. OpenGalaxy：<https://github.com/X-lab2017/open-galaxy>
13. Hypercrx：<https://github.com/hypertrons/hypertrons-crx>