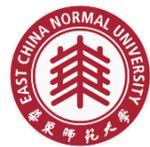


从 OpenDigger 看木兰开源社区的 高水平开放之路



王伟

华东师范大学

X-lab 开放实验室

2024 年 7 月

目录

01

协同布局

02

协同发展

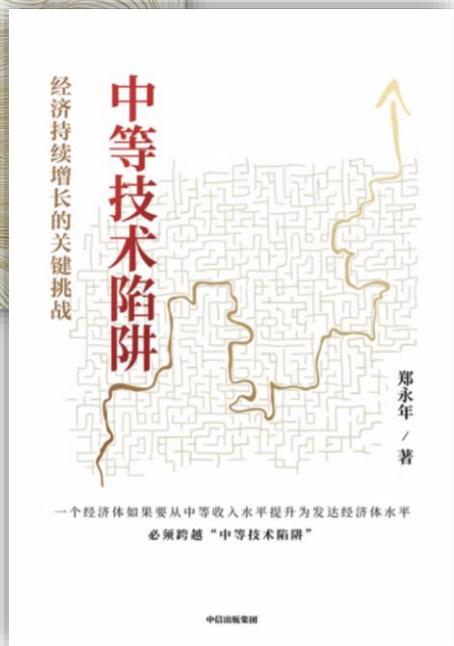
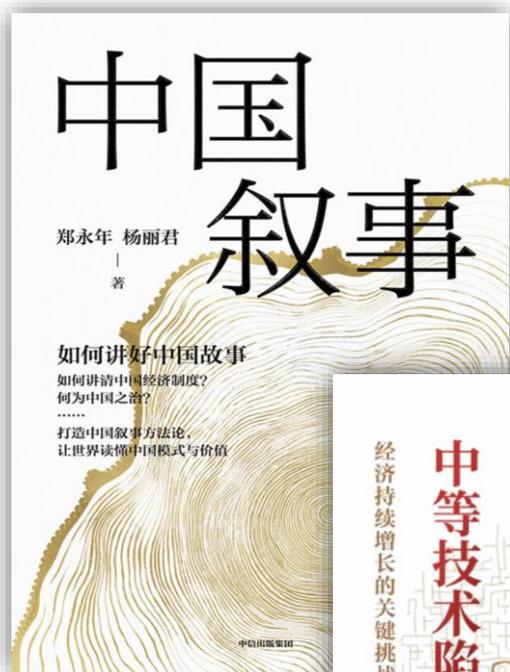
03

协同未来

高水平开放 = 开源标准化 + 标准开源化

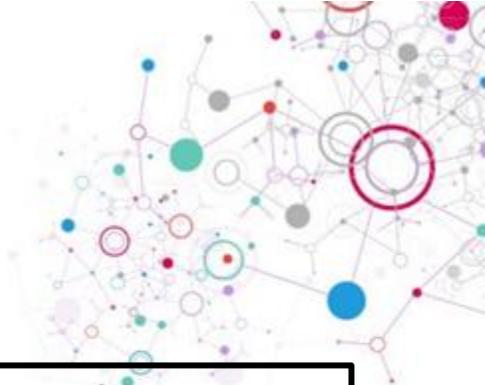


国家自主发展、经济持续增长的关键挑战



- **知识体系**是最重要的软实力。中国文明复兴的关键就是**建立中国自己的知识体系**。
 - 路径：**了解自己、走出去、解释并改造世界**。
 - 核心目标：确立以**中国实践经验**为基础的**中国知识体系**。
 - 平台是谁建立的不重要，重要的是**为谁所用**。
-
- 今天的中国处于**第三次开放（高水平开放）**，即以**规则、规制、标准和**管理为核心的**制度型开放**。
 - 突破“中等技术陷阱”的三个条件：
 - 一批有能力进行**基础科学研究**的大学与科研机构
 - 一批有能力把基础科研转化成**应用技术**的企业或机构
 - 一个开放的**创新平台与金融系统**

开源知识体系与标准



生态洞察

项目选型

OSPO治理

人才评价

社区激励

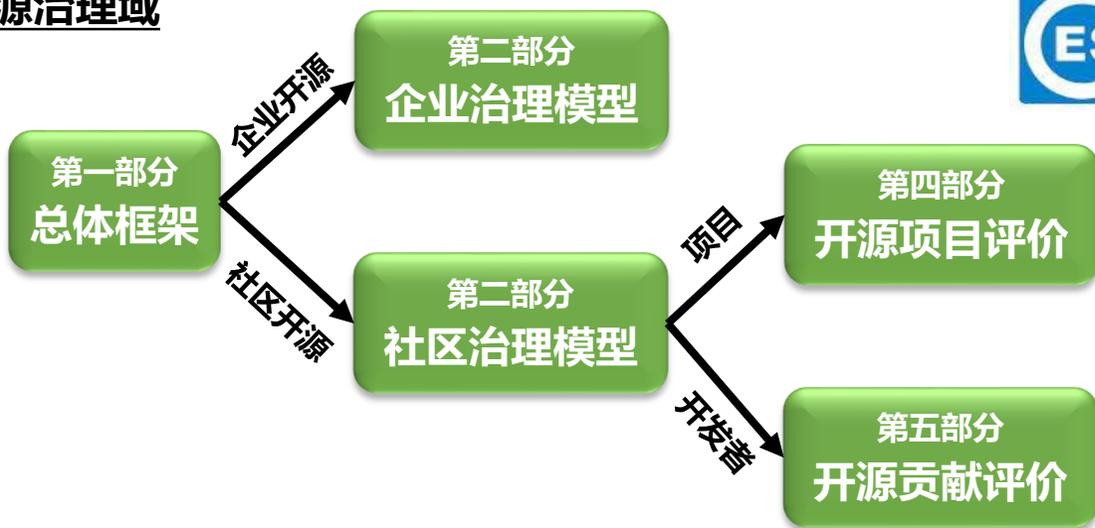
• 原创的知识体系

- 木兰开源许可证
- 开源术语、开源元数据
- 开源治理模型与评价指标

• 开放的制度设计

- 开源的标准化 + 标准的开源化
- 开源生态与开源社区
- 开放创新平台
- 开放金融系统

开源治理域



开源规则域

开源术语

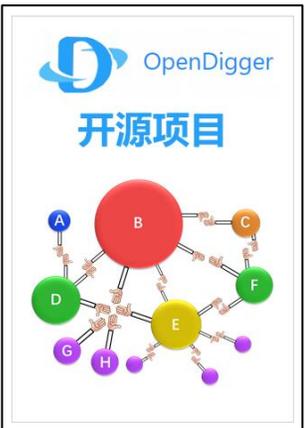
开源许可证

开源元数据

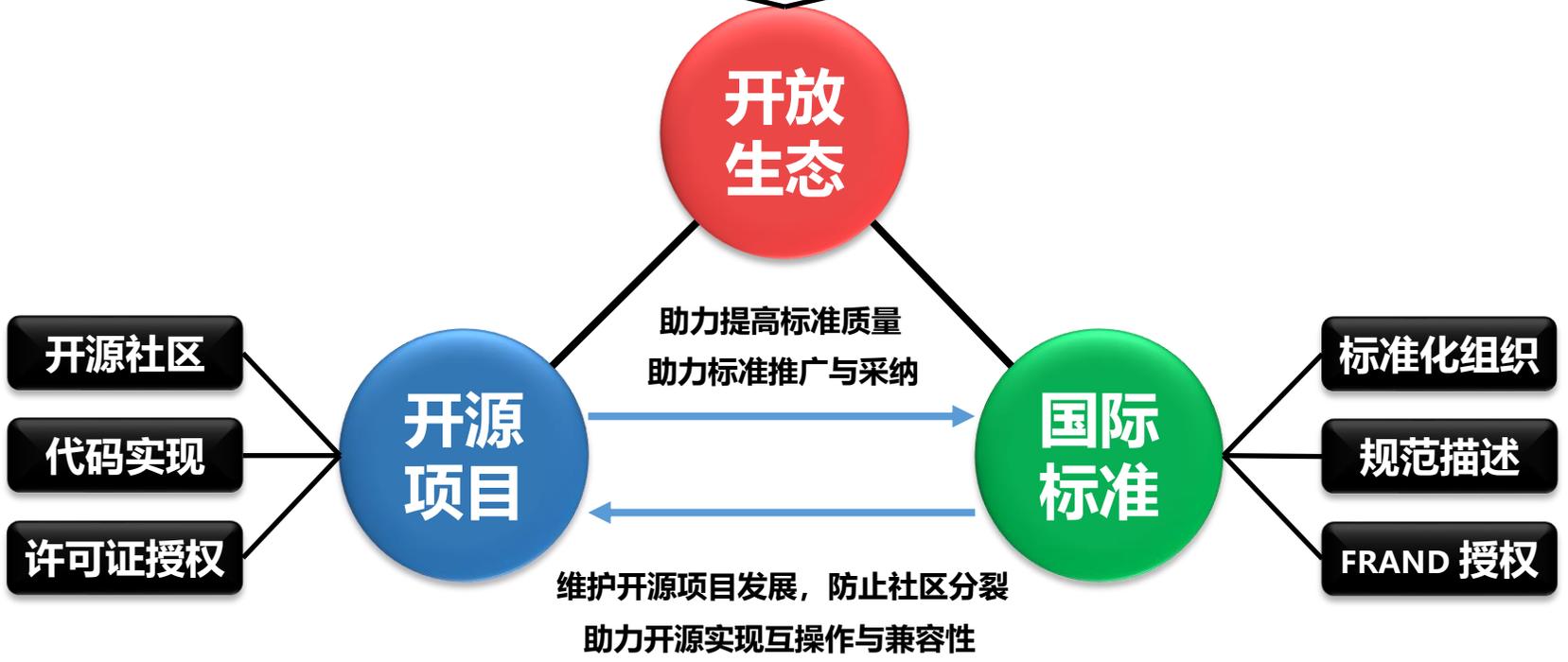
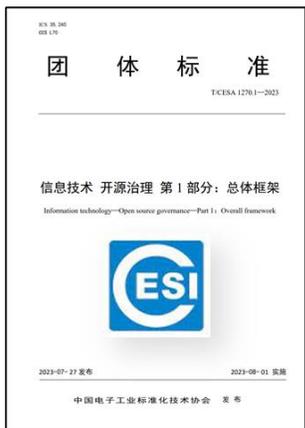
开源与标准协同共建开放技术和产业生态



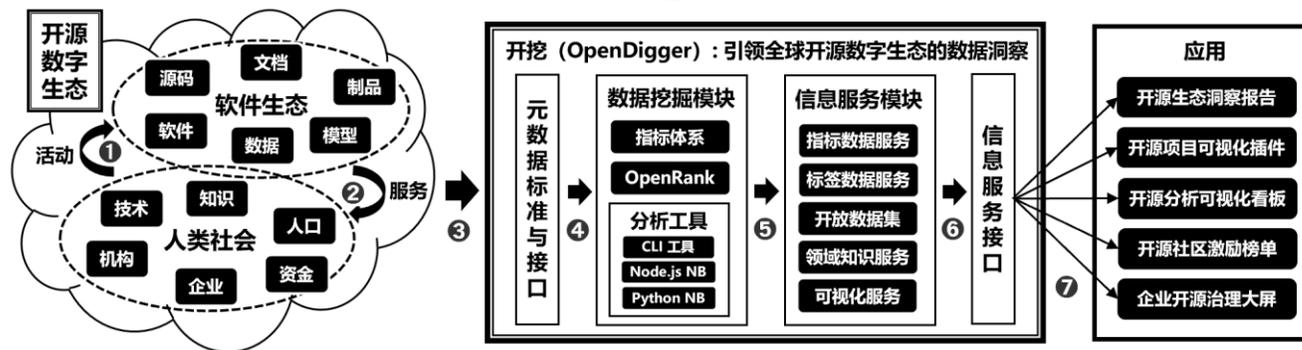
标准开源化



开源标准化



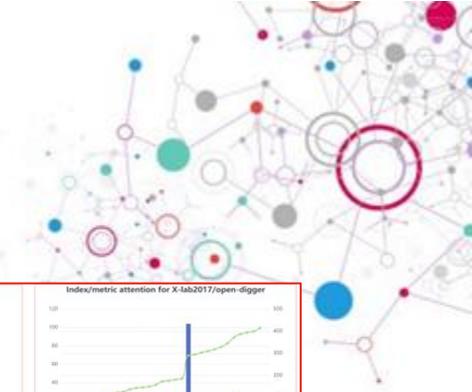
OpenDigger: 开源标准的开源实现



OpenDigger 是一个面向开源协作数字生态的一站式数据挖掘与信息服务项目，目标是构建开源领域的数字基础设施，促进开源生态的持续发展。所支持的开源领域数据包括 GitHub 和 Gitee 行为日志数据、制品库数据、安全漏洞数据、社区问答数据等。包括 Hypercrx、OpenLeaderboard、OpenGalaxy 等均是其下游开源项目。

- GitHub 行为日志数 58 亿条
- Gitee 行为日志数 1600 万条
- NPM/PyPI 等制品库数据 620 万条
- CVE 安全漏洞数据 16 万条
- StackOverflow 问答帖 2500 万条
- 标签类数据, 其中 GitHub 组织数量 413 个, 涉及仓库数 89427 个

OpenDigger 的指标与度量体系



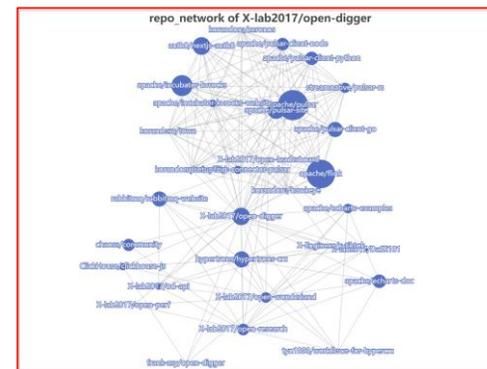
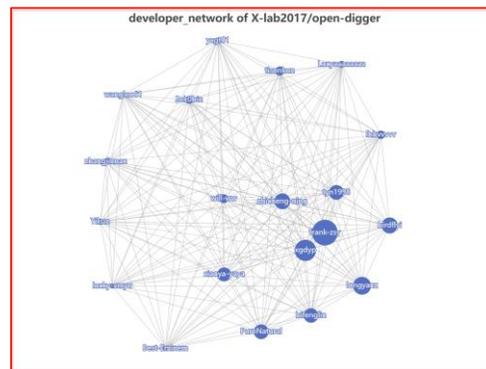
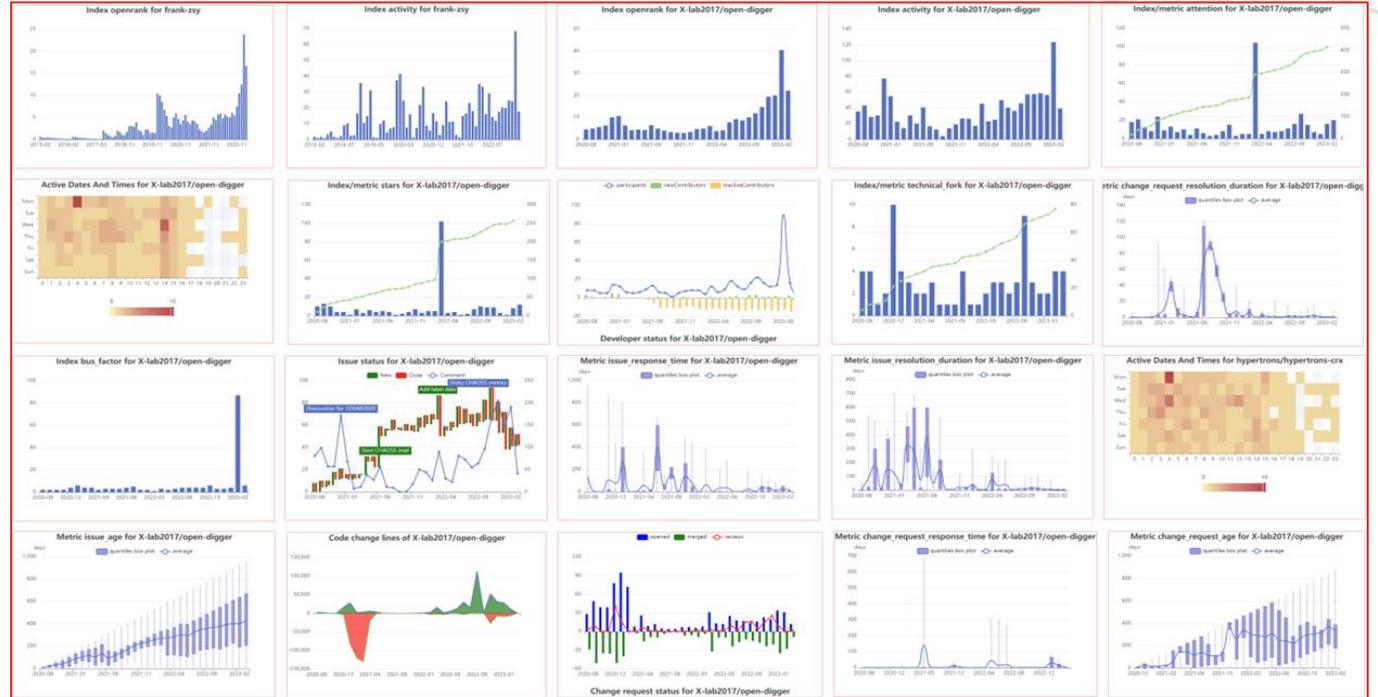
指标

For repos

- [标准院/X-lab] activity 评价
- [标准院/X-lab] **openrank** 评价
- [标准院/X-lab] attention
- [标准院/X-lab] stars
- [标准院/X-lab] issue_comments
- [标准院/X-lab] participants
- [LF/CHA OSS] technical_fork
- [LF/CHA OSS] issues_new
- [LF/CHA OSS] issues_closed
- [LF/CHA OSS] code_change_lines_add
- [LF/CHA OSS] code_change_lines_rem
- [LF/CHA OSS] code_change_lines_sum
- [LF/CHA OSS] change_requests
- [LF/CHA OSS] change_requests_accept
- [LF/CHA OSS] change_requests_reviews
- [LF/CHA OSS] bus_factor

For users

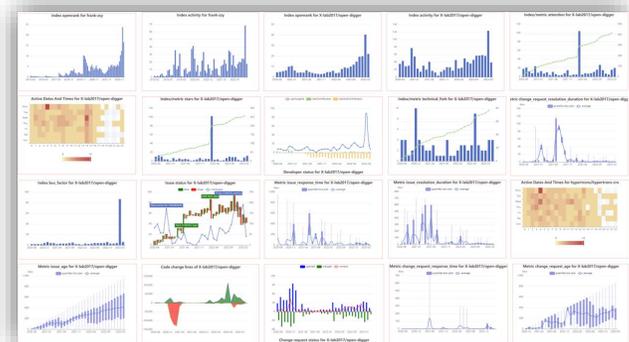
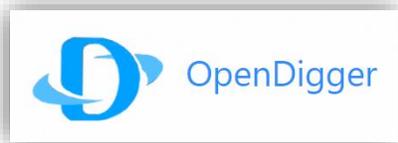
- [标准院/X-lab] activity 评价
- [标准院/X-lab] openrank



原创：北极星指标 OpenRank



开发者行为数据



指标与模型

The screenshot displays the OpenDigger interface for the 'X-lab Project OpenRank Network'. It features a central network graph with nodes representing developers and their interactions. To the left is a code editor with HTML, CSS, and JS code. To the right are two data tables: 'Leaderboard' and 'Details'.

Login	OpenRank
will-ww	44.91
TieWay59	42.82
zhicheng-ning	16.65
Liuyushii	7.64
abcdabcd3899	6.53
JinkaiXu	5.94
TW-githubs	5.74

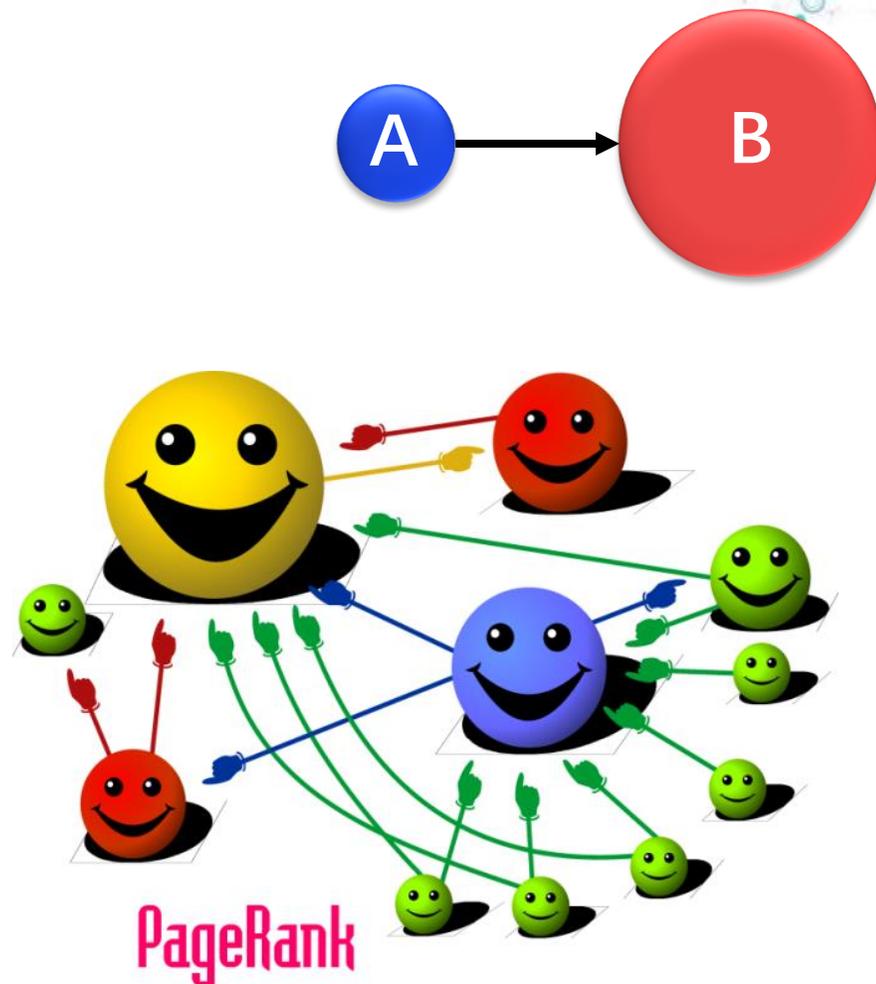
From	Ratio	Value	OpenRank
Self	0.15	35.87	5.380
#62	0.383	4.81	1.84
#70	0.383	4.32	1.652
#45	0.765	1.84	1.408
#64	0.765	1.82	1.392
#110	0.765	1.82	1.392
#52	0.402	3.4	1.366
#46	0.427	3.05	1.304
#48	0.637	2.04	1.3
#12	0.765	1.68	1.285
#59	0.408	2.94	1.2
#69	0.574	2.03	1.165

- 统计指标存在明显的观察者效应 (#Star、#Dev、#PR)
- 统计指标难以有效促进协作 (考核什么, 得到什么)
- 网络模型有良好的价值导向 (度量与行为影响一体化, 制度设计融入算法)

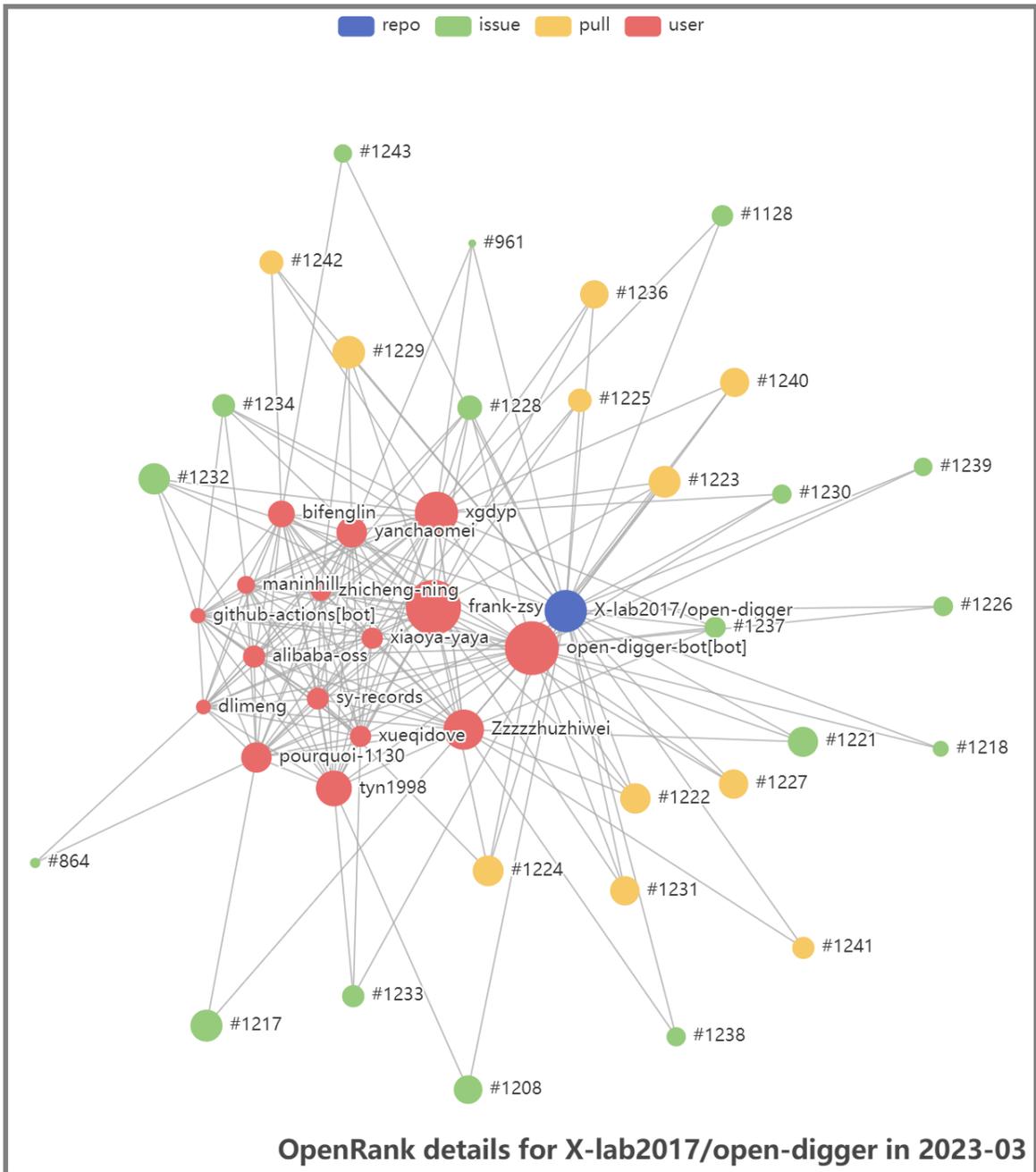
PageRank 的原理

• PageRank 理论

- 若**网页 A** 上有连接到**网页 B** 的链接，说明 A 认为 B 有链接价值，是一个“重要”网页
- 因此，一个网页的**重要性（权威性）** 由两个因素决定：
 - 该网页的**导入链接的数量**
 - 以及这些**导入链接的重要性**



OpenRank 价值网络



Leaderboard

Login	OpenRank
frank-zsy	26.79
open-digger-bot[bot]	24.95
xgdyp	12.64
Zzzzhuzhiwei	10.46
tyn1998	7.74
yanchaomei	5.34
pourquoi-1130	5.3
bifenglin	4.05

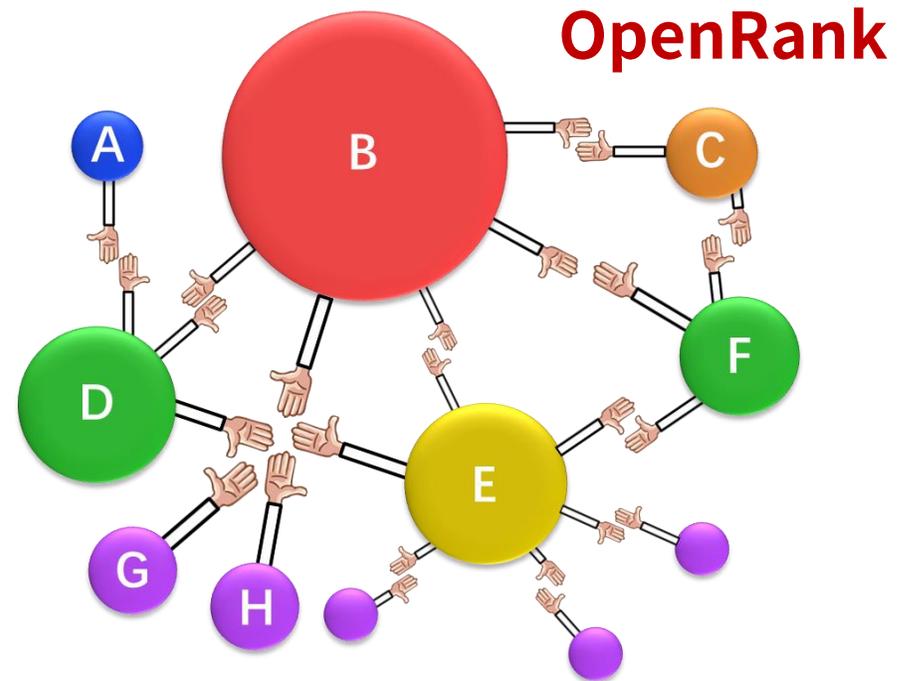
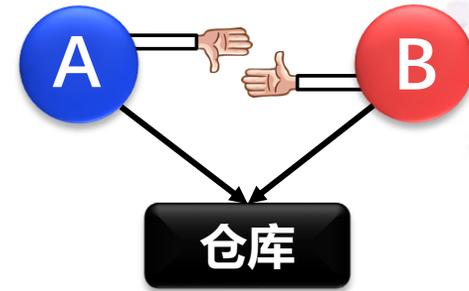
Details

From	Ratio	Value	OpenRank
Self	0.15	81.79	12.269
#1223	0.383	5.89	2.253
#1229	0.340	6.23	2.118
#1227	0.383	4.91	1.878
#1236	0.383	4.58	1.752
#1225	0.383	3.17	1.213
#1232	0.191	5.66	1.082
#1224	0.191	5.45	1.042
#1222	0.191	5.3	1.014
#1231	0.191	4.91	0.939
#1228	0.164	3.45	0.566
#1234	0.055	3	0.164
open-digger-bot[bot]	0.006	24.95	0.141
xqdyp	0.006	12.64	0.072

OpenRank 的原理：协作网络

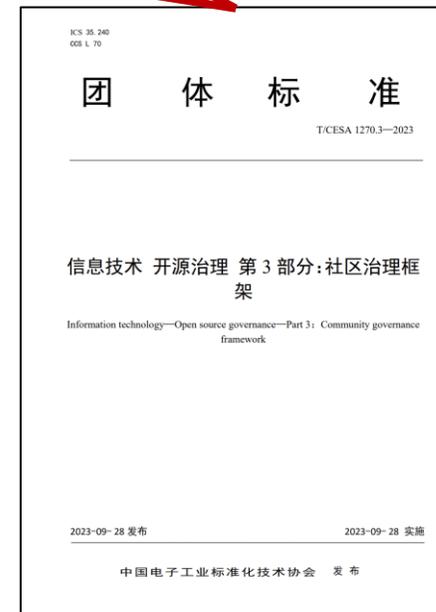
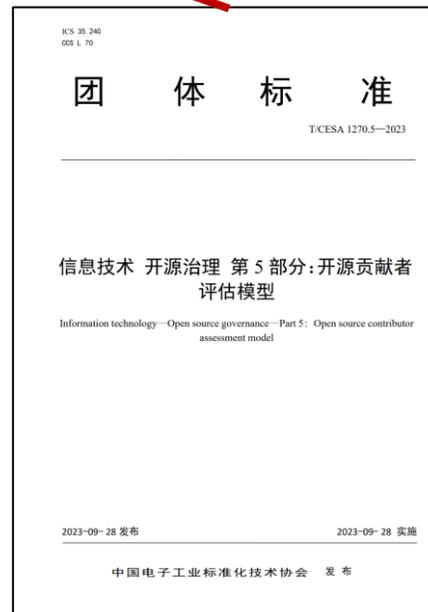
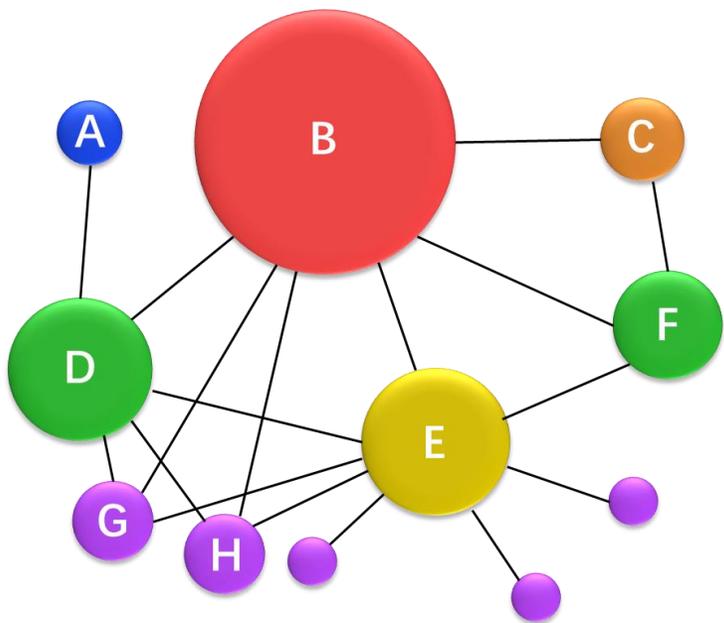
• OpenRank 理论（开发者协作网络）

- 若**开发者 A** 和**开发者 B** 在某个仓库上有过协作，说明彼此之间有共同的**价值认可**，且同时认为该**仓库是有价值的**
- 因此，一个**开发者的价值**由两方面因素决定：
 - 同该开发者协作过的**开发者数量**以及这些**开发者的价值**
 - 该开发者活跃过的**仓库数量**以及这些**仓库的价值**
- 同时，一个**仓库的价值**也由两方面因素决定：
 - 在该仓库上活跃过的**开发者数量**
 - 以及这些**开发者的价值**

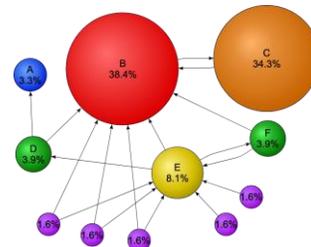


OpenRank 的标准化：价值单元、影响力与贡献度

- 一个开源项目所包含的所有仓库价值加总 → **开源项目影响力 (认可度)**
- 一个开源项目内的 Issue 和 PR → **价值单元**
- 开发者的价值 → **开发者贡献度**



OpenRank 实证研究的学术界认可



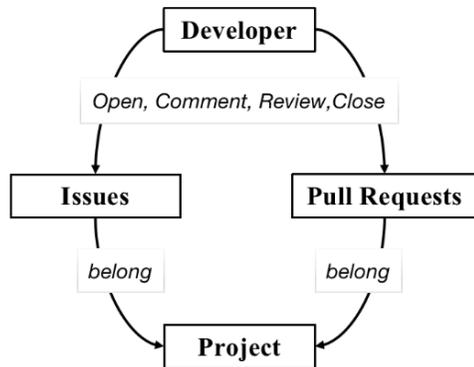
- 社区大于代码
- 开放式交流
- 通过贡献获得权威性

The Apache Way

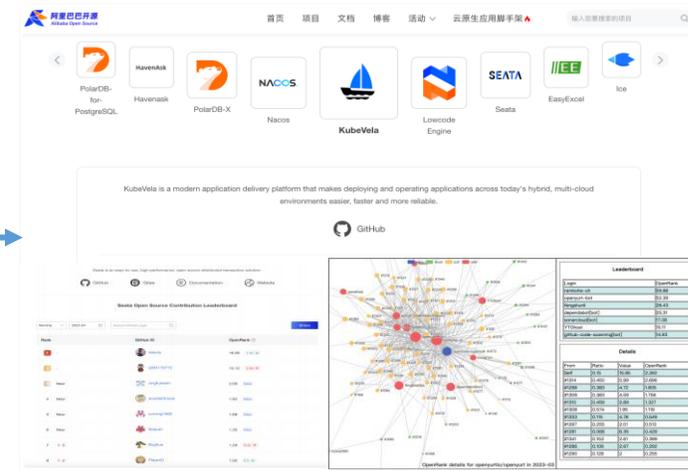
OpenRank

- 影响力较大的项目会吸引更多的人参与协作
- 更多有影响力的人参与的项目，影响力会增大

OpenRank Model



OpenRank Leaderboard



价值主张:

- 开发者倾向于与其他高价值开发者合作
- 如果有更多的开发者参与，问题/合并请求会更有价值
- 高价值开发者倾向于处理高价值的问题/合并请求
- 得到价值认可的开发者往往拥有更多的合作机会

OpenRank Leaderboard 能够:

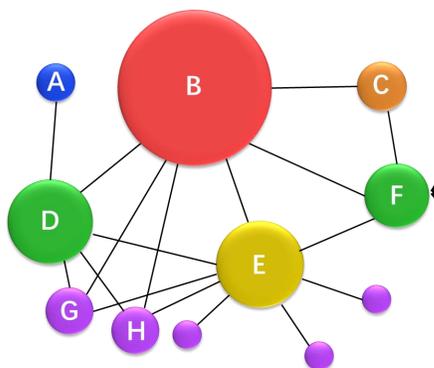
- 有效地反映开源开发者的贡献
- 激励开发者更积极地参与和贡献
- 对开源项目的合作行为产生了积极影响

OpenRank 的扩展：异质复杂网络

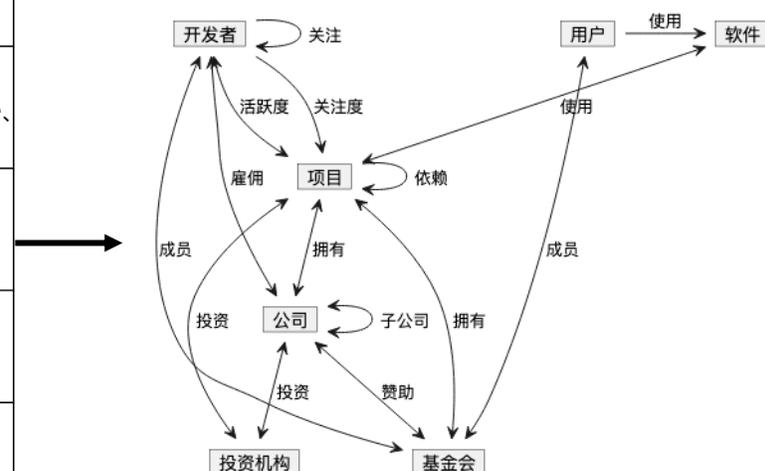


开源组件的全生命周期（供应链视角）

全域生态视角

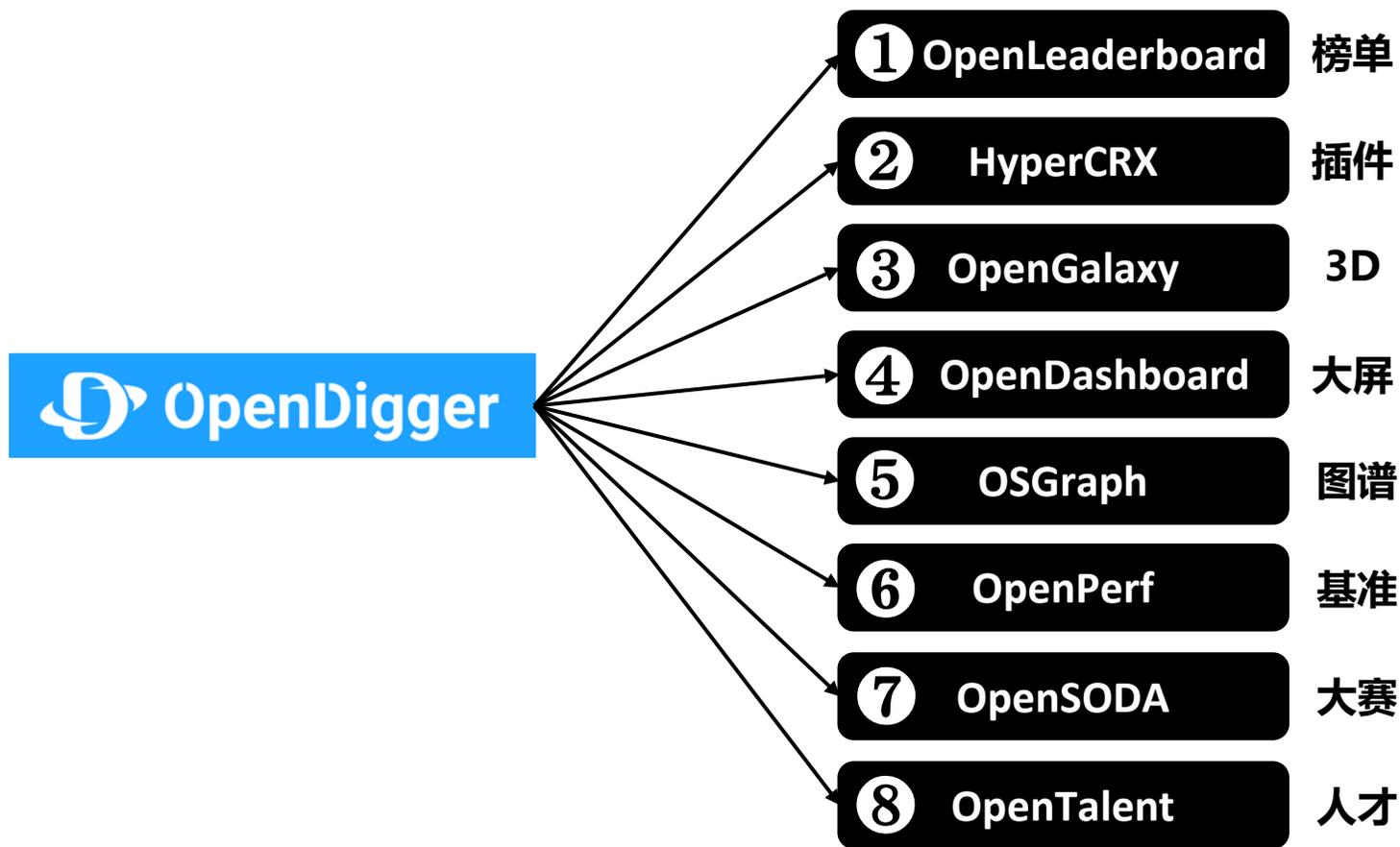


主要阶段	关键活动	相关概念	(图) 数据元件	(图) 网络构建
生产 (Produce)	数字产品（组件）的研发与生产等活动	开发、构建、测试、集成	开发协作网络	数字产品开发过程中开发者在协作单元（Issue、PR 等）上的开发行为
发布 (Release)	数字产品（组件）的交付与发布，以及推广宣传等活动	交付、传播、营销、宣传	社区关系网络	开发者在社区开发协作过程中的社交行为形成的关系网络
使用 (Apply)	数字产品（组件）的组装、部署、应用与服务等活动	组装、部署、调用、服务	组件依赖网络	数字制品复用过程中形成的供应关系网络
维护 (Maintain)	数字产品（组件）的升级、维护与演化等活动	更新、升级、演化、变更	交叉引用网络	通过项目之间的交叉引用，构建数字生态系统层面的技术依赖关系



- **群体博弈与制度性创新**
- **更多组织关系与价值流模型**
- **实体间价值流动与实体价值判断**
- **形成有效的竞合关系**

OpenDigger 生态项目群



1 OpenLeaderboard



https://open-leaderboard.x-lab.info/ ● 最近更新于 2023年12月12日

目标: 企业 项目 开发者 指数: 活跃度 OpenRank

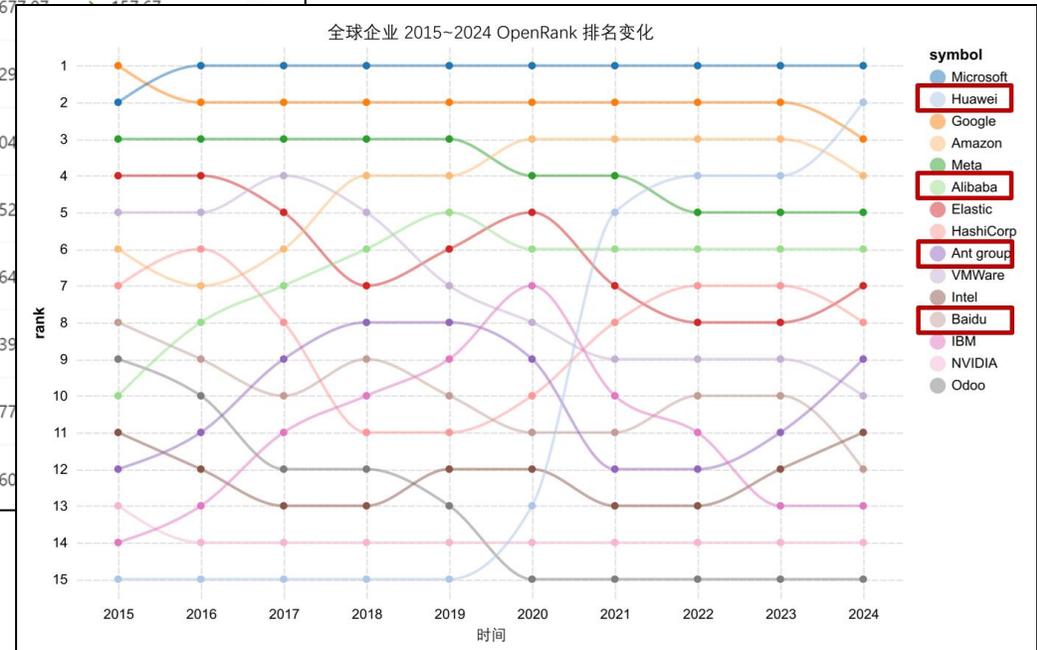
地区: 中国 全球 时间: Month 2023/11 详情:

排名	企业	活跃度
1	Microsoft	59520.25 ↘ 398.85
2	Huawei	25456.9 ↗ 11375.43
3	Google	23733.75 ↘ 1381.13
4	Amazon	22783.52 ↘ 3538.49
5	Meta	9866.89 ↘ 1024.67
6	Alibaba	7760.73 ↗ 740.76
7	Elastic	6582.63 ↘ 145.92
8	HashiCorp	6395.06 ↘ 555.06
9	VMWare	5201.1 ↘ 266
10	Ant group	4920.02 ↗ 455.95

目标: 企业 项目 开发者 指数: 活跃度 OpenRank

地区: 中国 全球 时间: Month 2023/11

排名	企业	OpenRank
1	Microsoft	41150.65 ↗ 390.86
2	Google	15249.99 ↘ 151.65
3	Amazon	12677.97 ↘ 157.57
4	Huawei	825.00 ↗ 1.00
5	Meta	704.00 ↘ 1.00
6	Elastic	552.00 - 1.00
7	Alibaba	464.00 - 1.00
8	HashiCorp	439.00 - 1.00
9	VMWare	377.00 - 1.00
10	Intel	360.00 - 1.00



Powered by OpenDigger Mulan Community

2022年6月北美 Open Source Summit 上正式发布~

② HyperCRX



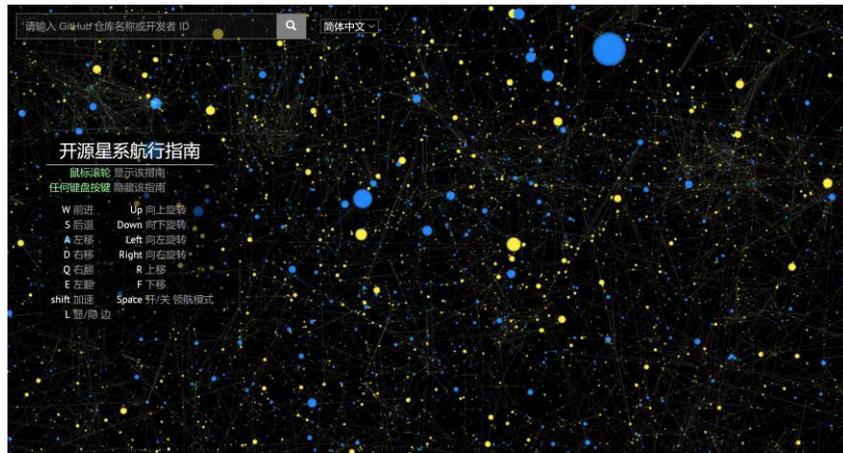
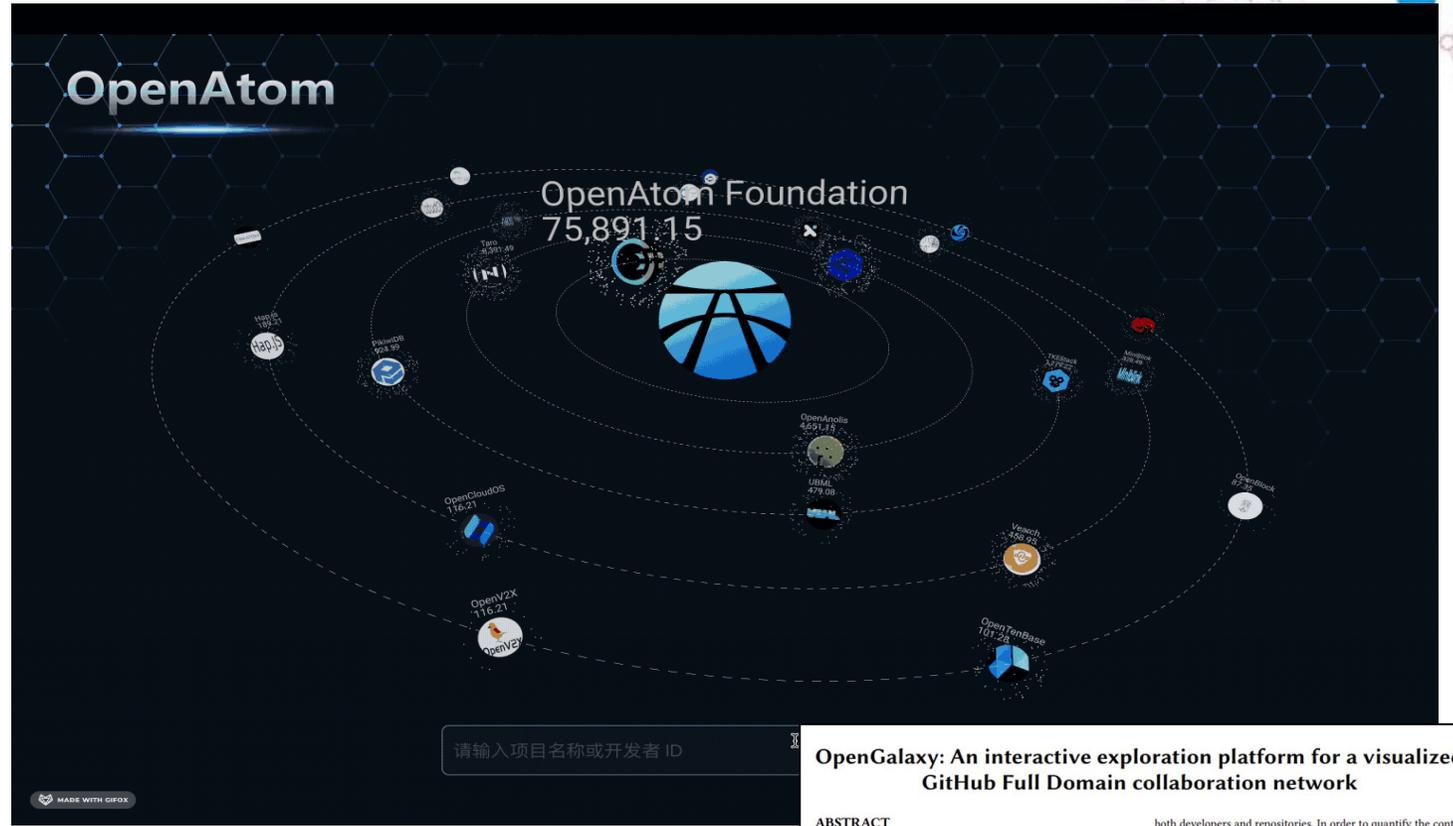
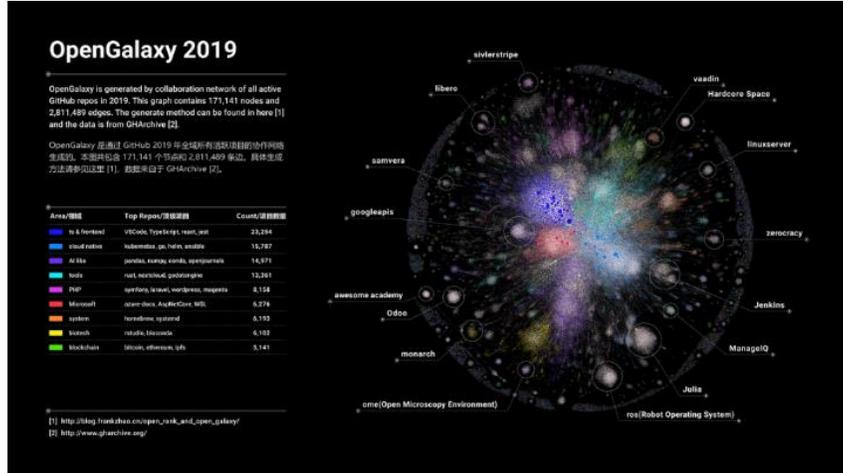
File/Folder	Description	Age
github	refactor: DevelopNetwork -> developer-networks (#638)	8 months ago
.husky	ci: add pre-commit and run prettier in it (#386)	last year
assets	[refactor] manifest v2 -> manifest v3 (#274)	last year
publish	chore[release]: v1.9.3	last month
scripts	feat: media player controllers for contributor activity racin...	2 months ago
src	fix: update mininterval for xAxis in multiple charts (#781)	4 days ago
utils	refactor: reconstruct option management (#659)	6 months ago
.babelrc	[refactor] manifest v2 -> manifest v3 (#274)	last year
.github/babelrc	[fix] prettier on windows OS (#332)	last year
.gitignore	feat: add Repo Activity Racing Bar (#580)	3 months ago
.nvmrc	feat: use the dominant color of a user's avatars as the colo...	2 months ago
.prettierrc	[refactor] manifest v2 -> manifest v3 (#274)	last year
.prettierrc	add source code	2 years ago
CODE_OF_CONDUCT.md	Refactor project using React. (#21)	3 years ago
CONTRIBUTING.md	docs: revise grammatical errors and typos of MD (#674)	6 months ago
INSTALLATION.md	feat: media player controllers for contributor activity racin...	2 months ago
INSTALLATION.zh-CN.md	feat: media player controllers for contributor activity racin...	2 months ago
LICENSE	Refactor project using React. (#21)	3 years ago
NOTICE	docs: add a NOTICE file to apply the LICENSE to our work...	10 months ago
README.md	feat: media player controllers for contributor activity racin...	2 months ago

hypertrons/hypertrons-crx
984 users, 967 users
首届 Microsoft Edge 大赛冠军
DaSE Data Science Engineering, X-lab 开放实验室

Hypercrx: A browser extension for insights into GitHub projects and developers
Anonymous Author(s)

ICPC 2024 The 32nd IEEE/ACM International Conference on Program Comprehension
Lisbon, Portugal. April 15-16, 2024

③ OpenGalaxy



OpenGalaxy: An interactive exploration platform for a visualized GitHub Full Domain collaboration network

ABSTRACT
In this work, we introduce OpenGalaxy - an interactive exploration platform for a visualized GitHub Full Domain collaboration network based on 3D force-oriented layouts. We first collected GitHub domain-wide log data, built a developer-repository heterogeneous collaboration network, calculated both the Influence value for each repository and the activity value for each developer, finally per-

both developers and repositories. In order to quantify the contributions of developers, we introduced an activity metric, and for measuring the project's influence, we employed a weighted PageRank algorithm. Finally, a 3D force-guided layout was implemented to effectively visualize this intricate collaborative network. Furthermore, we implemented an interactive 3D game-like mobile mode to enrich user exploration. This tool not only supports dynamic

ICPC 2024
The 32nd IEEE/ACM International Conference on Program Comprehension
Lisbon, Portugal, April 15-16, 2024

5 OSGraph



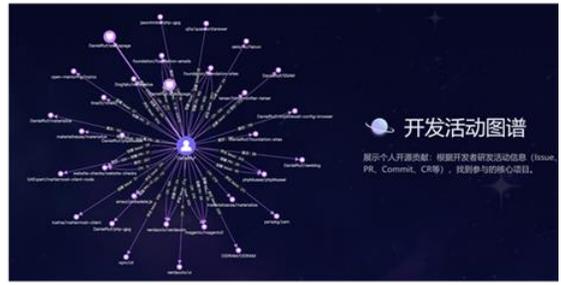
The screenshot shows the OSGraph web interface with four main analysis modules: '领域发展史' (Domain Development History), '生态对比' (Ecosystem Comparison), '用户与仓库' (Users and Repositories), and '热门技术' (Popular Technologies). Each module includes a timeline or graph visualization and a search input field. Below these modules is a search bar with a 'Repo' dropdown and a search button. A '历史记录' (History) section shows 'X-lab2017/oss101'.

OSGraph: A Data Visualization Insight Platform for Open Source Community

Wenrui Huang¹[0009-0009-8645-3193], Xiaoya Xia¹, Aoying Zhou¹, Xuan Zhou¹, Wei Wang¹, Shengyu Zhao², Zhiyong Wang³, and Sikang Bian³

¹ East China Normal University, Shanghai, China
hwr0577@gmail.com, xiaoya@stu.ecnu.edu.cn, ayzhou@sei.ecnu.edu.cn,





OpenDigger Timeline

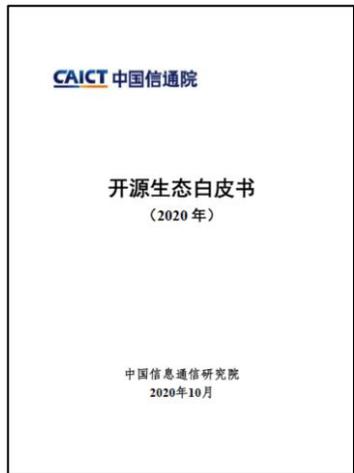


- 报告与论文类
- 工具与应用类
- 大屏与看板类
- 模型与算法类

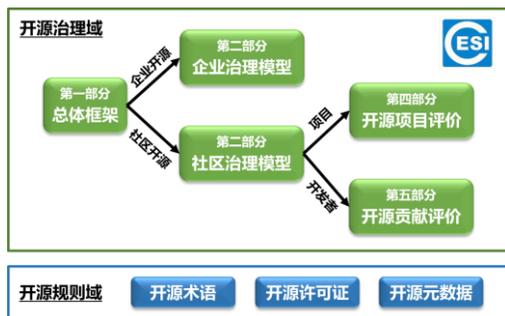




采纳 OpenDigger 数据与 OpenRank 指标的各类报告



OpenDigger 以及木兰开源社区的未来



- 标准与项目的**持续迭代**
- 标准与项目的**持续国际化**
- 标准与项目的**持续应用推广**



- **知识体系**是最重要的软实力。
- 中国文明复兴的关键就是**建立中国自己的知识体系**。
- 路径：**了解自己、走出去、解释并改造世界**。
- 平台是谁建立的不重要，重要的是**为谁所用**。

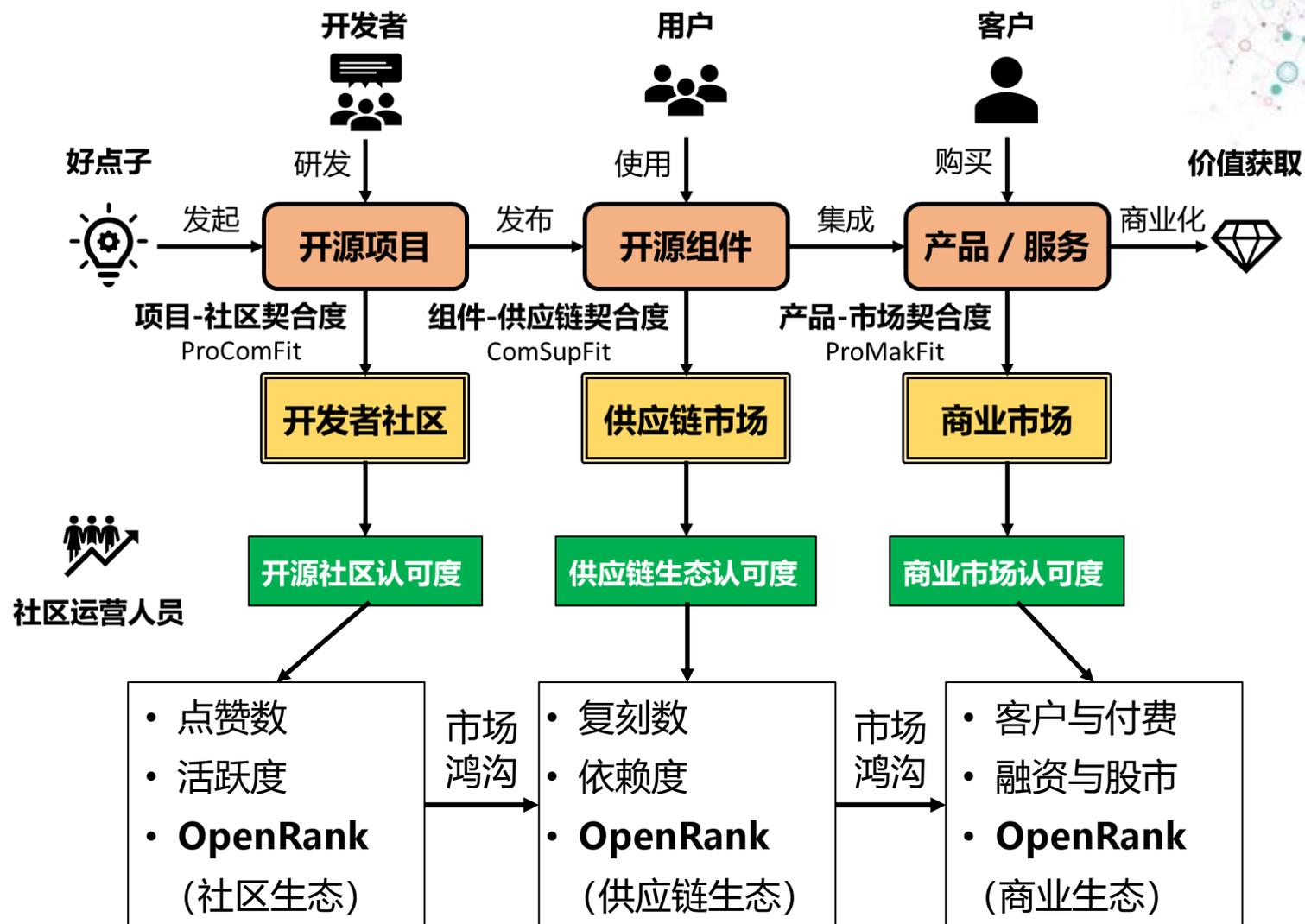
- 今天的中国处于**第三次开放（高水平开放）**，即以**规则、规制、标准和管理**为核心的**制度型开放**。
- 突破“中等技术陷阱”的三个条件：
 - 一批有能力进行**基础科学研究**的大学与科研机构
 - 一批有能力把基础科研转化成**应用技术**的企业或机构
 - 一个开放的**创新平台与金融系统**



- 理论研究，木兰研究院
- 应用技术转化，木兰孵化器
- 开放创新平台，Git-X
- 开放金融系统，开源激励

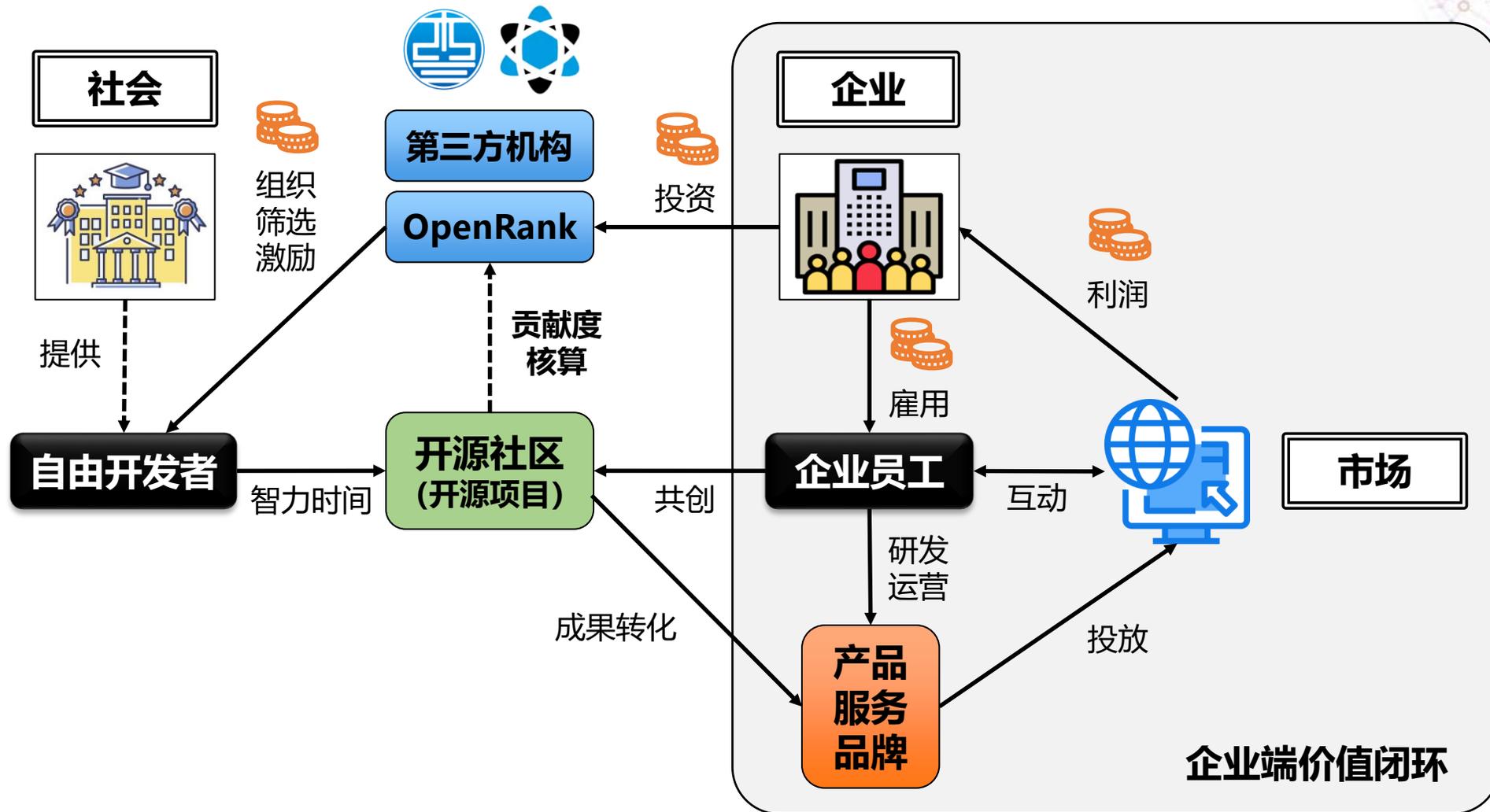
OpenRank 在商业开源上面的应用

数字经济视角下的
开源软件商业化模型



鸿沟理论对商业开源
发展的意义

OpenRank 经济系统



Thanks

