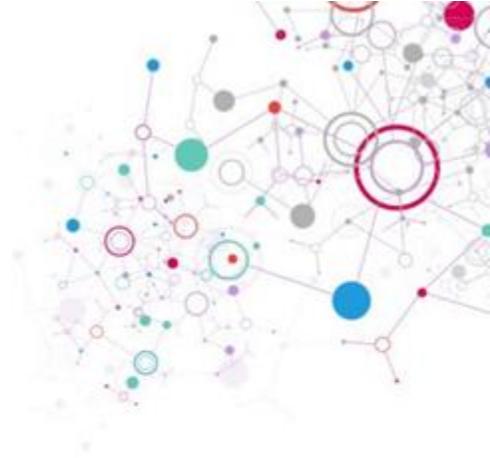




面向开源创新的贡献评价与激励机制

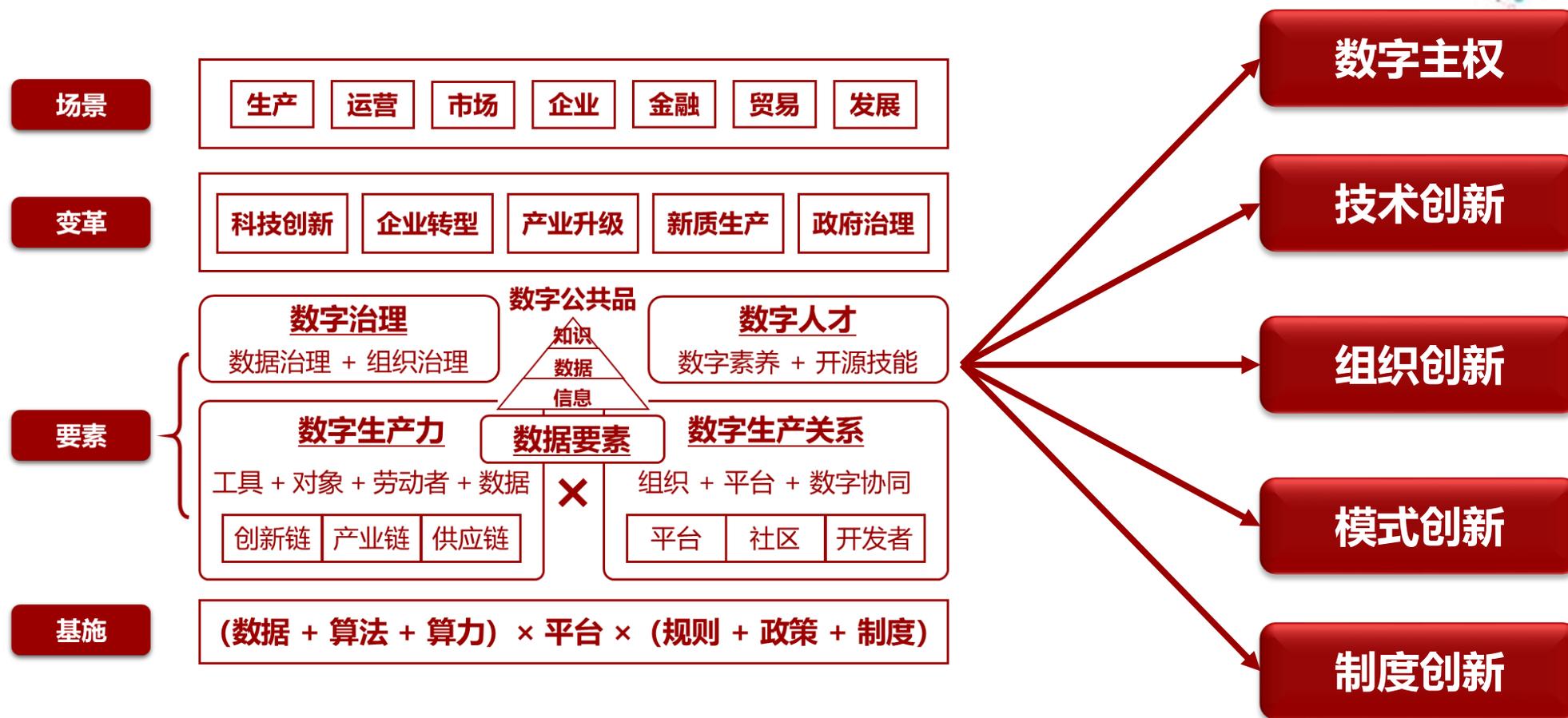
关键技术研究

华东师范大学
X-lab 开放实验室
2024 年 6 月

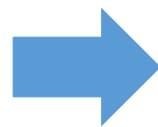
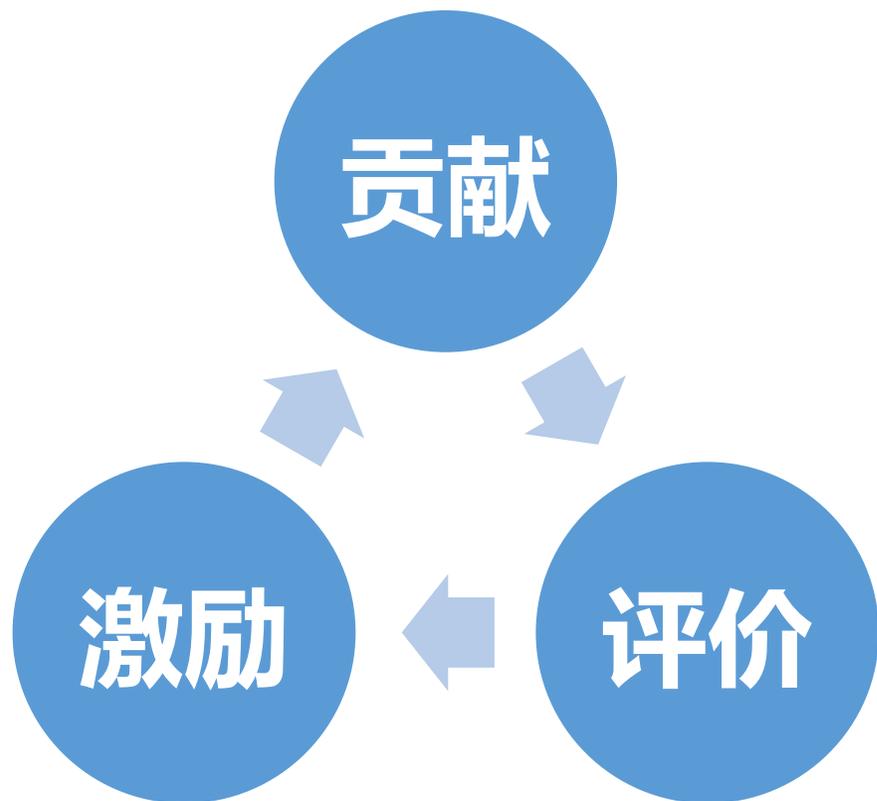


背景介绍

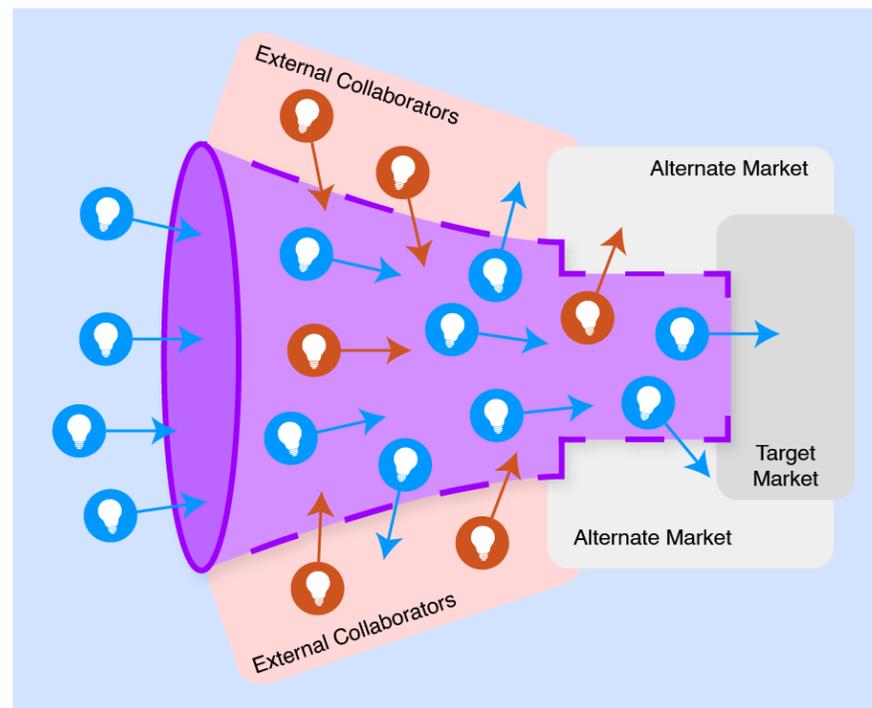
数字经济时代的开源创新



评价与激励是开源创新的“牛鼻子”工程



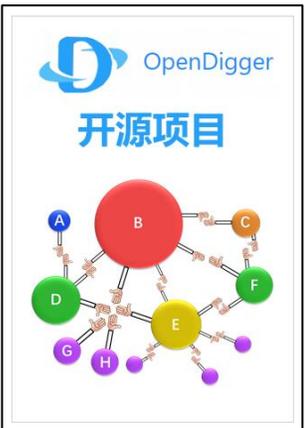
网络化开放式创新



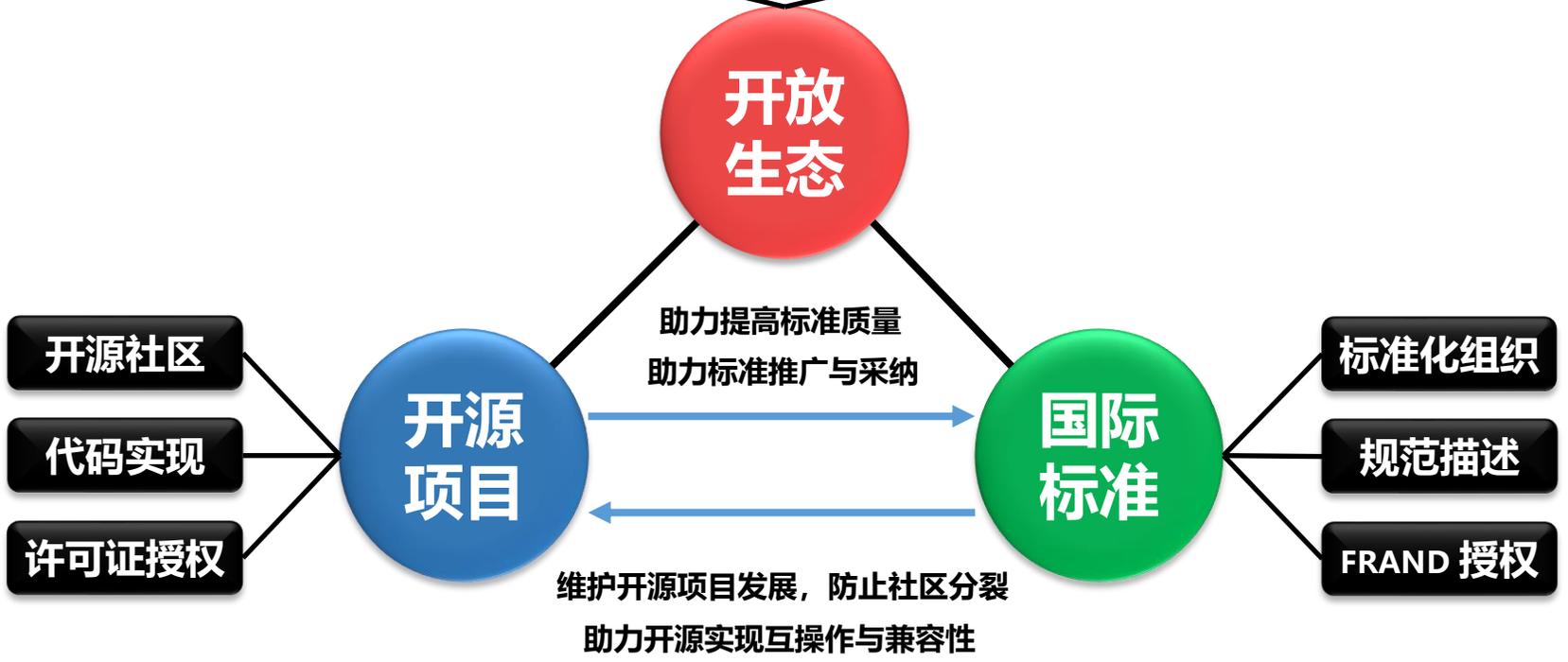
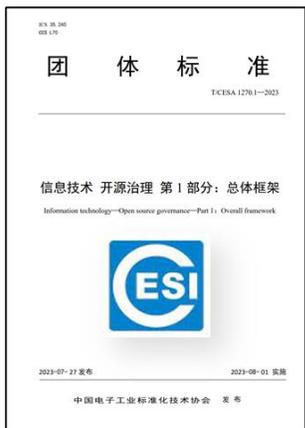
战略：开源与标准协同共建开放技术和产业生态



标准开源化



开源标准化



成果 1：开源知识体系与标准

生态洞察

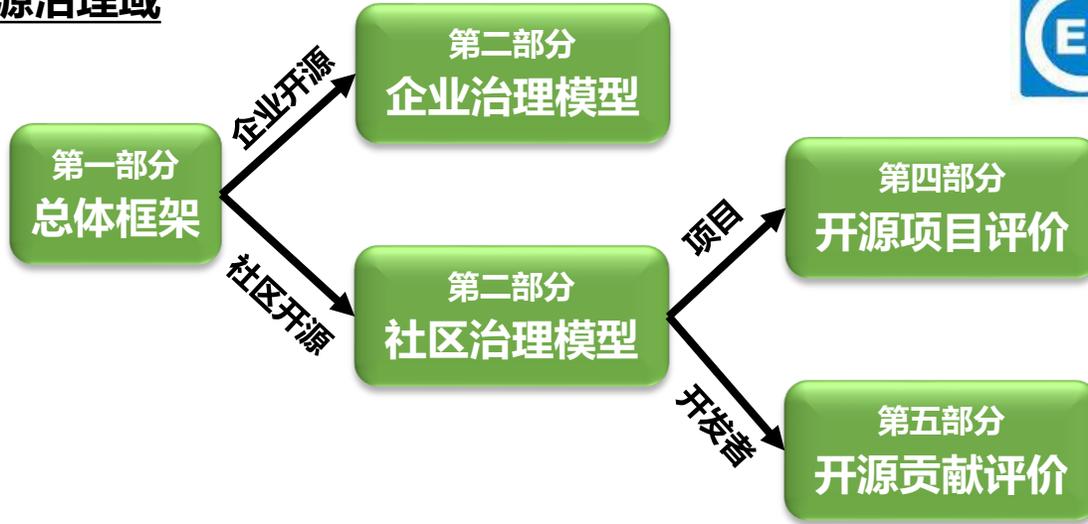
项目选型

OSPO治理

人才评价

社区激励

开源治理域

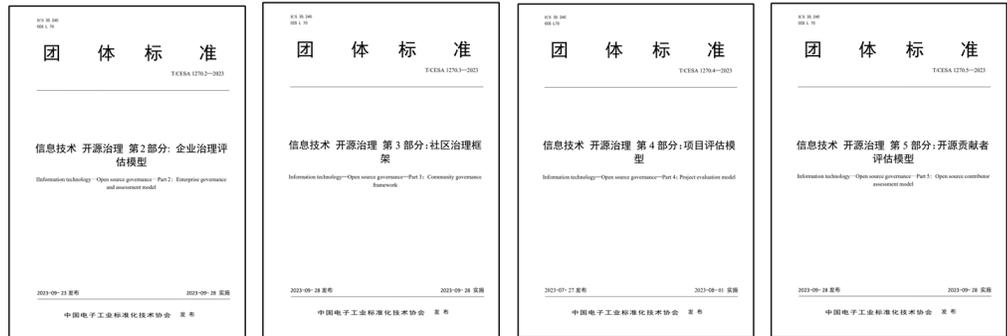
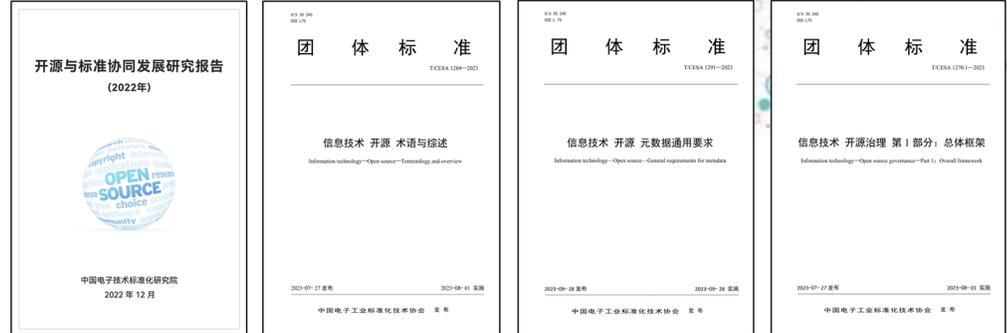


开源规则域

开源术语

开源许可证

开源元数据



前言

本文件按照 GB/T 1.1—2020《标准化工作导则 第1部分：标准化文件的结构和起草规则》的规定起草。

本文件是 T/CECSA 1270《信息技术 开源治理》的第1部分。T/CECSA 1270 已经发布了以下部分：

——第1部分：总体框架；

——第4部分：项目评估模型。

本文件由华东师范大学提出。

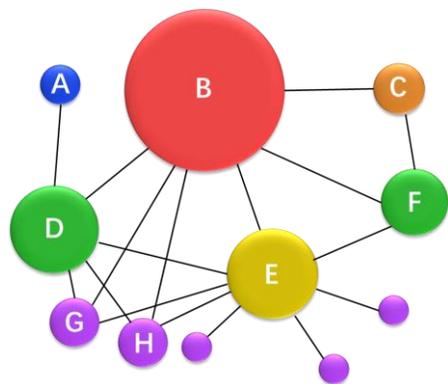
本文件起草单位：华东师范大学、中国电子技术标准化研究院、开放原子开源基金会、中兴通讯股份有限公司、浪潮电子信息产业股份有限公司、浪潮云信息技术股份有限公司、北京百度网讯科技有限公司、普元信息技术股份有限公司、上海计算机软件技术开发中心、阿里云计算有限公司、浙江九州云信息科技有限公司、蚂蚁集团集团股份有限公司、北京华胜天成科技股份有限公司、中移（苏州）软件技术有限公司、苏州旋极七彩信息科技有限公司、上海交通大学、中国软件评测中心（工业和信息化部软件与集成电路促进中心）、东软集团股份有限公司、中移系统集成有限公司、深圳市金耀天云计算股份有限公司、杭州星云科技集团有限公司、中国移动通信研究院、中国移动安信息技术有限公司、北京宝兰迪软件股份有限公司、北京中电协信息技术有限公司。

本文件主要起草人：王伟、杨丽霞、李成双、张百林、王荷舒、于秀明、李明、江燕、黄先芝、马红伟、顾伟、孟庆余、王超、周光明、周鹏、付欣伟、章津楠、郭智慧、梁朝、钱岭、胡吉兵、金耀辉、赵赫、文礼、边思康、王旭、吴琦、袁斌、袁薇、梁大功、杨佳丽、孙康健、林科、庠敏越、郁志强、杨静、孙福洲、魏兰芳、刘伟、詹年科、高家祺、黄静、万里鹏飞、黄雷宇、黄浩东、郭雪雯、赵海玲、张康杰、杨程好、付辉、彭哲、王耀、王媛媛、魏飞铃、刘紫君、张天、李任平、苏广峰、沈斌、葛建新、周灵斌、陈庆坤、饶雪、王廷升。

- 20 多家企业
- 10 余个科研院所
- 70 多位编制专家
- 近 3 年时间

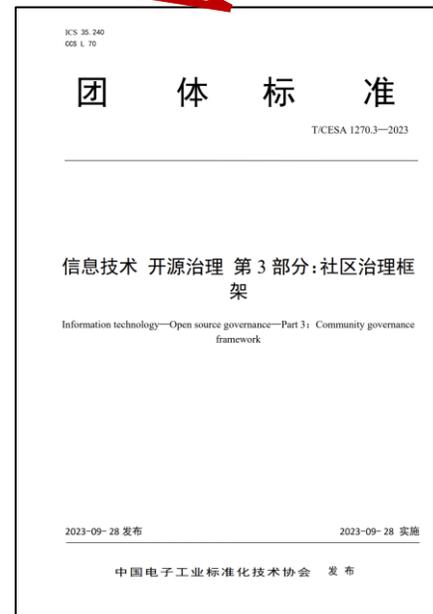
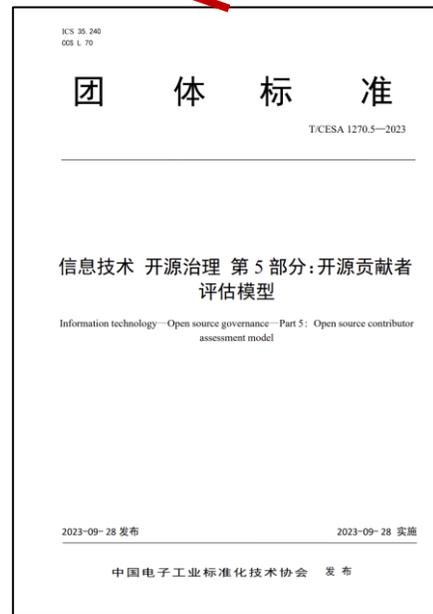
成果 2: OpenRank 理论体系与统一模型

- 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**

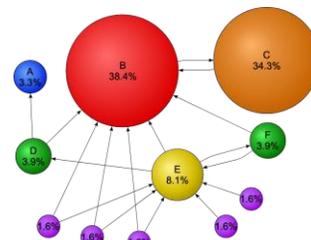


• 建立 OpenRank 理论体系与统一模型来度量影响力与贡献度

- 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)



OpenRank 成果的国际认可

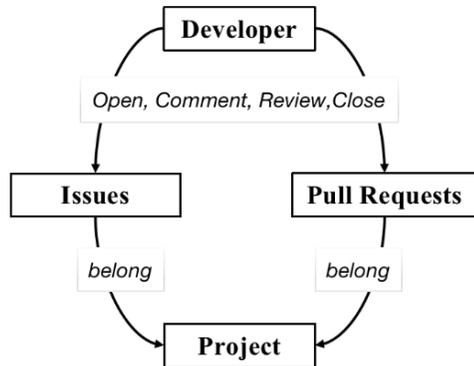


- 社区大于代码
- 开放式交流
- 通过贡献获得权威性

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

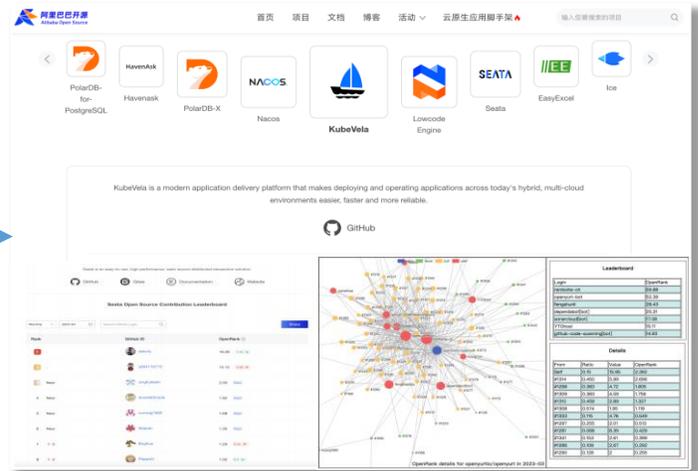


OpenRank Model

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

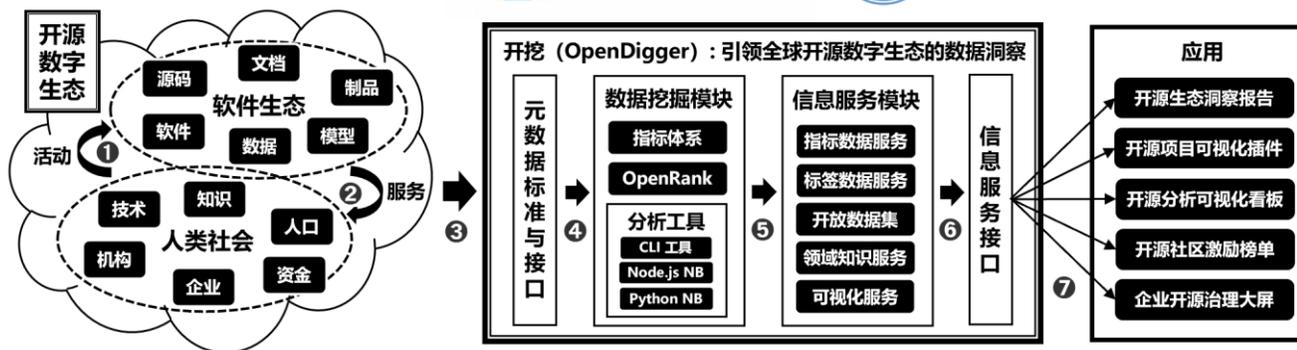
- OpenRank Leaderboard 能够:
- 有效地反映开源开发者的贡献
 - 激励开发者更积极地参与和贡献
 - 对开源项目的合作行为产生了积极影响



成果 3: OpenDigger 开源治理标准的开源实现



Mulan Community



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/CHAOSS] technical_fork
- [LF/CHAOSS] issues_new
- [LF/CHAOSS] issues_closed
- [LF/CHAOSS] code_change_lines_add
- [LF/CHAOSS] code_change_lines_rem
- [LF/CHAOSS] code_change_lines_sum
- [LF/CHAOSS] change_requests
- [LF/CHAOSS] change_requests_accept
- [LF/CHAOSS] change_requests_reviews
- [LF/CHAOSS] bus_factor

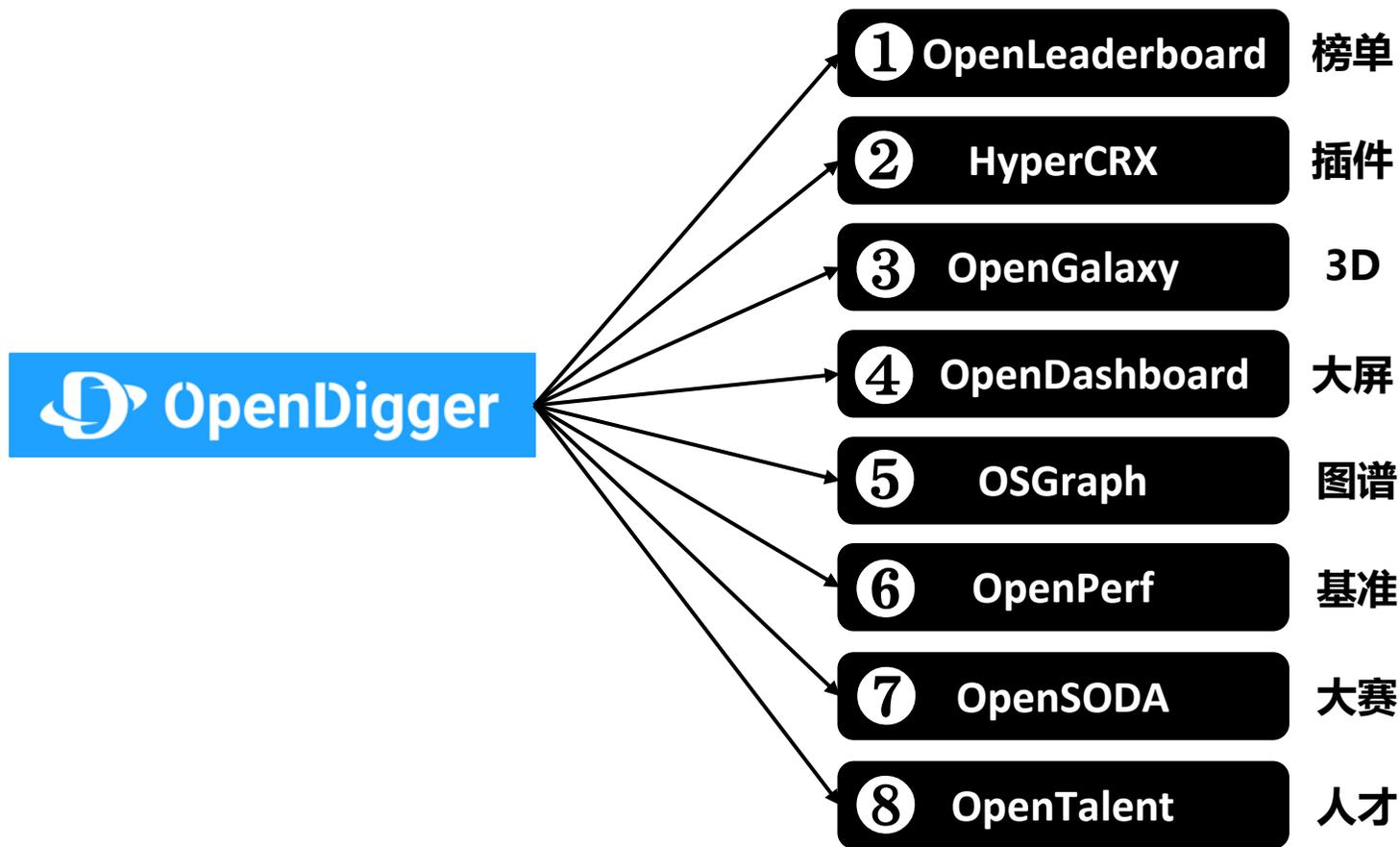
For users

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

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

成果 4: OpenDigger 生态项目群



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	OpenRank
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	6300.00	100.00
9	VMWare	5000.00	100.00
10	Ant group	4900.00	100.00

1 OpenLeaderboard

Activity charts and metrics for the hypertrons/hypertrons-crx repository, including Issue Open/Close/Comment Events, Activity, Fork Events, Open PR, Merge, Review Comment Events, Participants, Star Events, and Merged Addition & Deletion Code Lines.

2 HyperCRX

Project Correlation Network and Active Developer Collaboration Network visualizations for the hypertrons/hypertrons-crx repository, showing relationships between projects and active developers.

3 OpenGalaxy

OpenGalaxy interface showing a star field visualization of projects and a list of contributors with their activity metrics.

4 OSGraph

Project Contribution Graph visualization showing relationships between projects and contributors.

项目贡献图谱

Project Ecosystem Graph visualization showing relationships between projects and contributors.

项目生态图谱

Project Community Graph visualization showing relationships between projects and contributors.

项目社区图谱

Development Activity Graph visualization showing relationships between projects and contributors.

开发活动图谱

TuGraph / 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 comparison tool and a search bar for repository names. Below these is a '单点搜索' (Single Search) section with a search bar and a '历史记录' (History) section showing '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

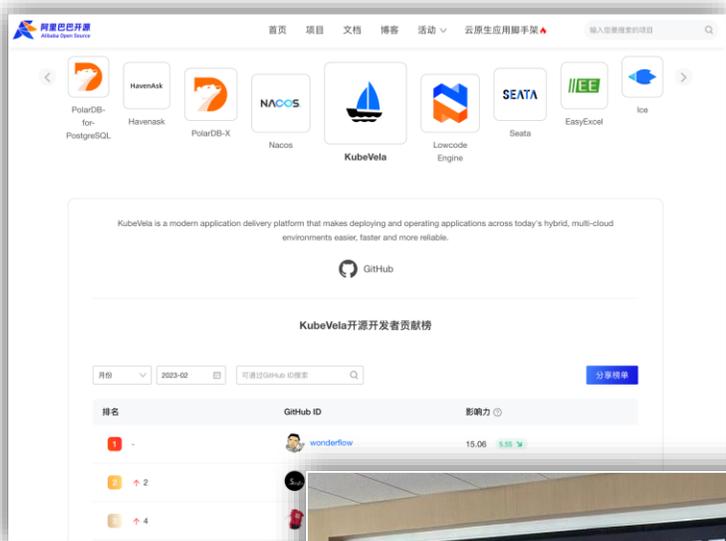
¹ East China Normal University, Shanghai, China
hwr0577@gmail.com, xiaoya@stu.ecnu.edu.cn, ayzhou@sei.ecnu.edu.cn,

DASFAA 2024
Gifu Japan



成果 5：成果转化与应用服务

阿里巴巴开源开发者贡献榜



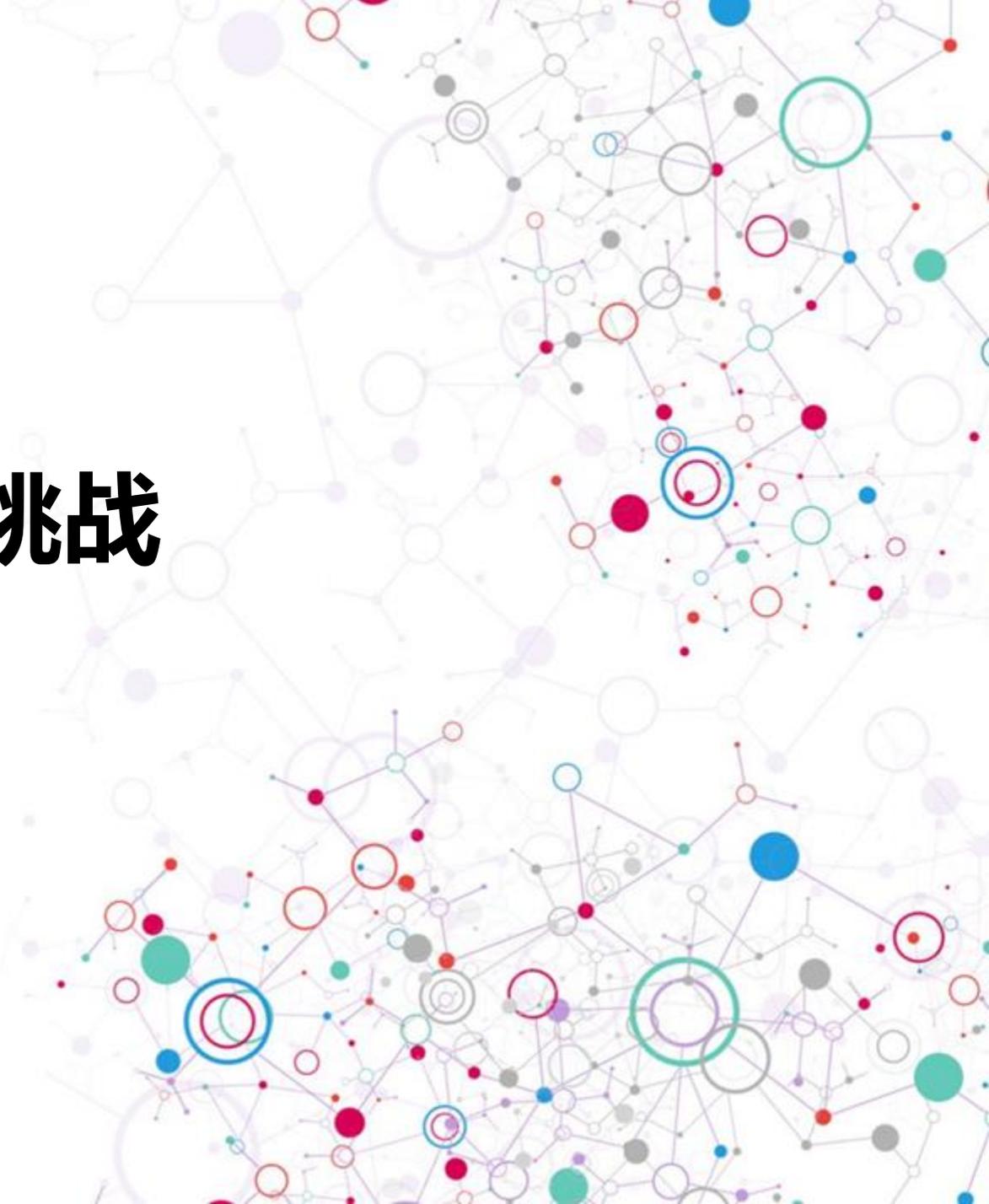
蚂蚁 / 字节开源治理可观测系统



开放原子全球开源协作全景图



问题与挑战



CCF-蚂蚁科研基金「图计算专项」



蚂蚁技术研究院
ANT RESEARCH

- 1 CCF-蚂蚁科研基金「图计算专项」整体介绍
宣讲嘉宾：洪春涛 蚂蚁集团 | 图计算负责人
- 2 CCF-蚂蚁科研基金「图计算专项」课题解析
宣讲嘉宾：洪春涛 蚂蚁集团 | 图计算负责人
陈琨 蚂蚁技术研究院图计算实验室 | 研究员
刘永超 蚂蚁集团 | 图计算高级技术专家
- 3 CCF-蚂蚁科研基金「图计算专项」课题申报Q&A

徐碧璘-蚂蚁集团的屏幕共享

洪春涛
陈琨-蚂...
刘永超
徐...

前沿科技探索与研发

在数字化浪潮下，数据要素价值日益重要，发掘前沿数字科技，深挖底层数据价值，让数据高效安全地流动并创造价值，是一条全新的科技起跑线，也是蚂蚁集团数实融合持续发力点。

63% 204.6亿元
18,678件 167项

成立蚂蚁技术研究院

2021年4月，蚂蚁技术研究院正式成立。面向数字化、智能化未来，瞄准世界科技前沿，推进关键核心技术攻关，促进“产学研用”深度融合，为做强、做大、做优数字技术贡献一份力量。

研究目标：
以通过图灵测试为目标
的通用人工智能
高精度真实世界的AIGC
加速100000倍的高
隐私保护系统

围绕三大技术领域，下设六大实验室：

- AI
交互智能实验室
聚焦图神经网络模型研究，开发通用人工智能算法架构，包括内容生成、多模态理解、数字人技术等人机交互关键技术。
- 数据库实验室
针对时代的金融级分布式HTAP数据库关键技术。
- 图计算实验室
高性能实时图计算，图数据库和图智能。
- 隐私计算
密码学实验室
探索下一代隐私计算技术的研究，包括不限于安全多方计算、同态加密、差分隐私等。
- 计算系统实验室
致力于以数据驱动技术设计实现高性能的先进计算系统。
- 程序设计与编译基础设施实验室
致力于以数据驱动技术设计实现高性能的先进计算系统。包括面向任意安全模型的先进计算系统、AI加速、分布式AI计算架构等。

图隐私计算

- 图计算已经在人工智能、计算科学、生物信息、金融科技、社交网络等领域发挥着举足轻重的作用
- 未来：融合多方数据，补全缺失联系，建模万物互联
- 获得全面视角，洞悉深层关系；充分挖掘图数据的价值和图计算的潜力
- 另一方面：图数据可能包含个人的身份、关系、偏好和行为等高度敏感信息，必须保证数据隐私安全

徐碧璘-蚂蚁集团的屏幕共享

基于隐私保护多方图学习的异常检测或社群挖掘

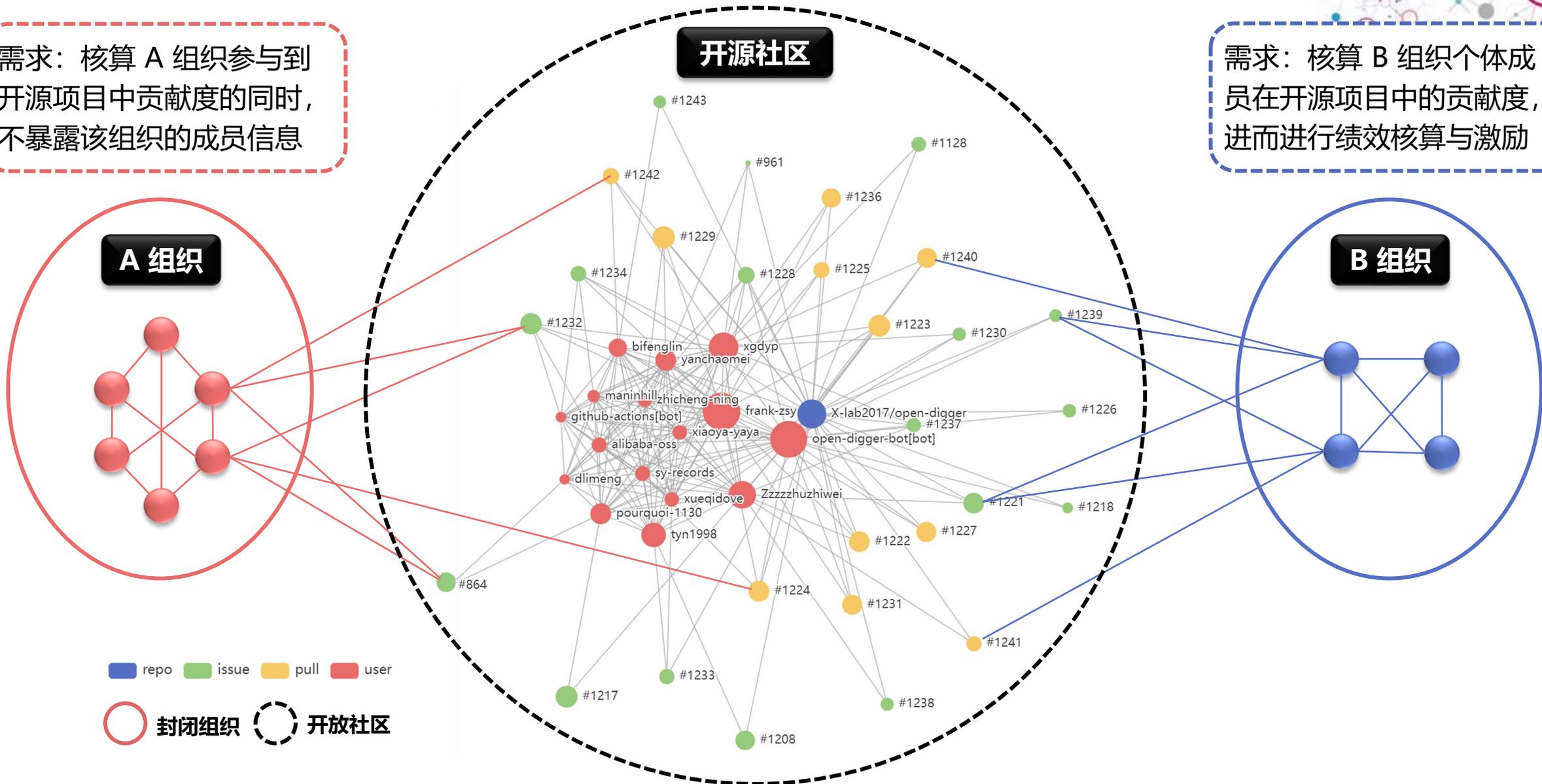
- 图学习已经成为异常检测或社群挖掘等领域重要的基础技术，并得到广泛应用
- 本课题旨在研究在保护数据隐私的前提下，采用融合多方图数据的图学习进行异常检测或者社群挖掘的系统或算法
- 综合运用同态加密、多方安全计算、差分隐私或联邦学习等技术
- 1) 产出1~2篇蚂蚁认可的高质量顶级学术会议(CCF-A)论文；2) 申请专利1项以上；3) 交付1套完整的代码和文档。

徐碧璘-蚂蚁集团的屏幕共享

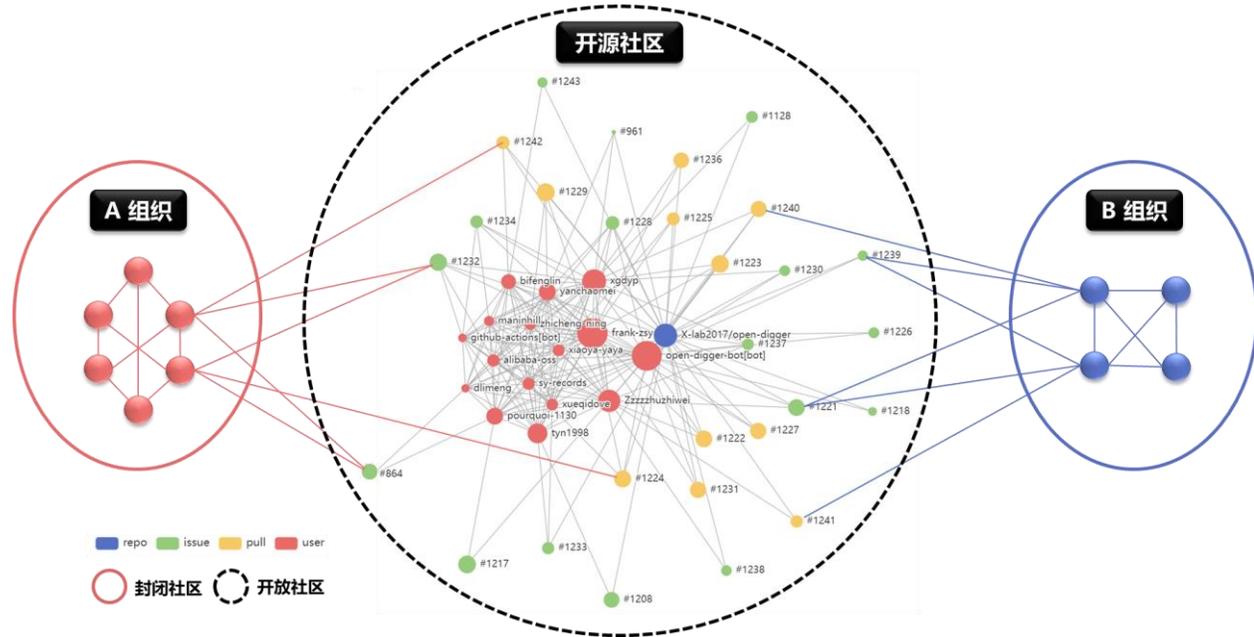
一个由开放社区与封闭组织组成的协作网络

需求：核算 A 组织参与到开源项目中贡献度的同时，不暴露该组织的成员信息

需求：核算 B 组织个体成员在开源项目中的贡献度，进而进行绩效核算与激励

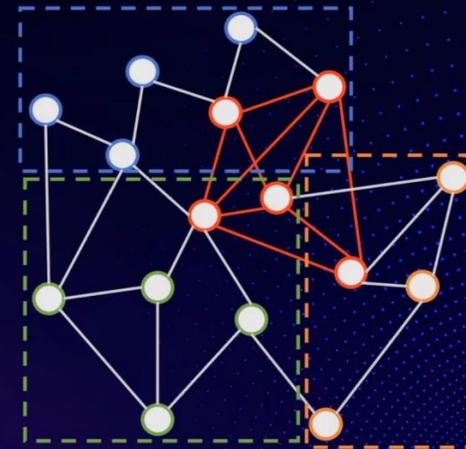


申请指南：基于隐私保护多方图学习的异常检测、社群挖掘与影响力评估问题



基于隐私保护多方图学习的异常检测或社群挖掘

- 图学习已经成为异常检测或社群挖掘等领域重要的基础技术，并得到广泛应用
- 本课题旨在研究在保护数据隐私的前提下，采用融合多方图数据的图学习进行异常检测或者社群挖掘的系统或算法
- 综合运用同态加密、多方安全计算、差分隐私或联邦学习等技术
- 1) 产出1~2篇蚂蚁认可的高质量顶级学术会议(CCF-A)论文；2) 申请专利1项以上；3) 交付1套完整的代码和文档。



异常检测

- 恶意刷分 (Star、PR、Issue 等)、集团勾结

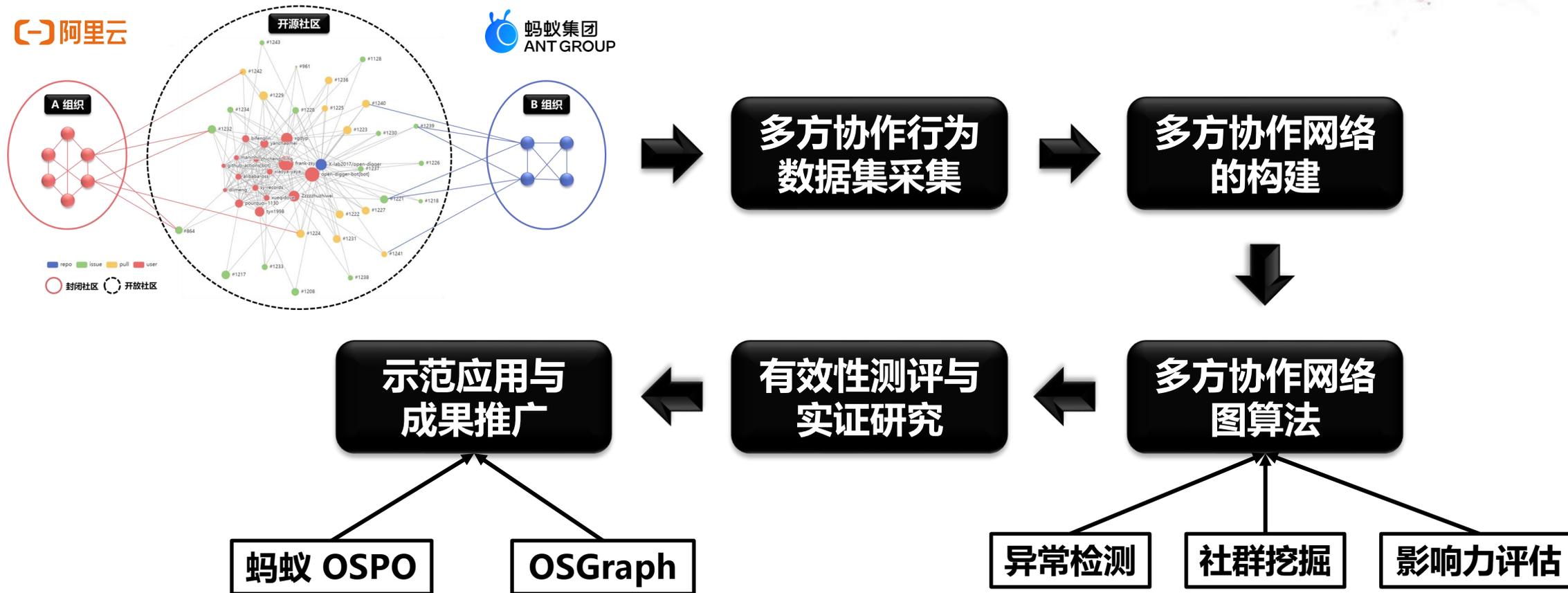
社群挖掘

- 开源共同体识别、新兴技术领域识别

影响力评估与最大化

- OpenRank 在隐私保护多方图上的扩展
- 开发者影响力最大化

技术路线与方案



Thanks

