

开源社区检索与分析系统

OS-COMMUNITY-EXPLORER

ikim队

目录 / CONTENT

第一部分

背景

BACKGROUND

第二部分

创新性与技术挑战

INNOVATION AND CHALLENGES

第三部分

设计方案

DESIGN SCHEME

第四部分

应用场景

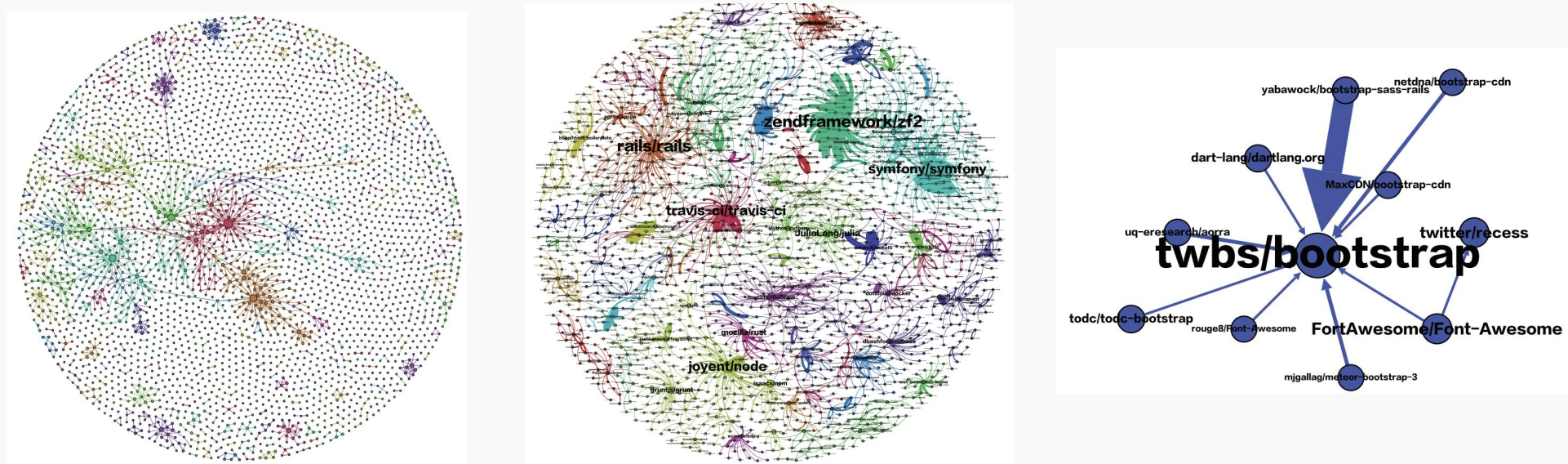
APPLICATION SCENARIOS

第一部分

背景

背景

开源项目往往天然存在社区结构，项目的不同部分由不同的开发者或团队负责，形成了基于任务、兴趣或技术栈的社群。GitHub社区检测旨在通过分析开发者之间的协作关系，识别开源项目中的社群结构。本应用基于GitHub上的贡献数据（如代码提交、拉取请求等日志信息）构建项目网络，提出新的社区检测算法，最后应用社群检测算法发现项目群体，并在页面上实现在线可视化分析。这有助于理解项目的社群结构、项目群中核心项目的观测以及相关仓库发现等。论文^[1]中的可视化应用示例：



[1] Blincoe, K., Harrison, F., Kaur, N., Damian, D. (2019). Reference coupling: An exploration of inter-project technical dependencies and their characteristics within large software ecosystems. Information and Software Technology, 110, 174–189.

第二部分

创新性与技术挑战

创新性

1

交互式图可视化网络

通过项目间的共同参与者构建项目网络，参与包括PR、Issue以及评论等。基于WebGL和力导向布局对项目网络进行可视化展示，支持节点搜索、节点点击交互等。

2

新的社区发现算法

提出新的属性图社区发现算法。针对当前深度图聚类对比学习损失中潜在的假阴样本对的问题，提出动态样本对加权策略，并通过高置信社区样本进一步指导模型训练。对GitHub仓库的标题、标签和描述等文本信息进行编码之后，利用提出算法进行社区发现。

3

项目群社区检索

项目的社区通过节点颜色进行标注，在力导向布局的可视化大图上进行直观展示。支持节点子社区检索功能，包括节点一跳邻居、k-core以及k-truss等社区搜索算法，从而提供给用户基于节点的局部视角。

技术挑战

1

项目网络构建与可视化

当前存在数量众多的开源项目，需要合理筛选出展示的项目，并提取用户交互信息合理地进行边构建。开源项目之间往往具有较多的节点和边，需要对整个网络可视化的布局进行有效考虑。

2

社区发现算法

提出基于聚簇感知的社区发现算法需要在常规的基准数据中进行有效性验证，之后再应用到开源项目网络中。

3

功能集成

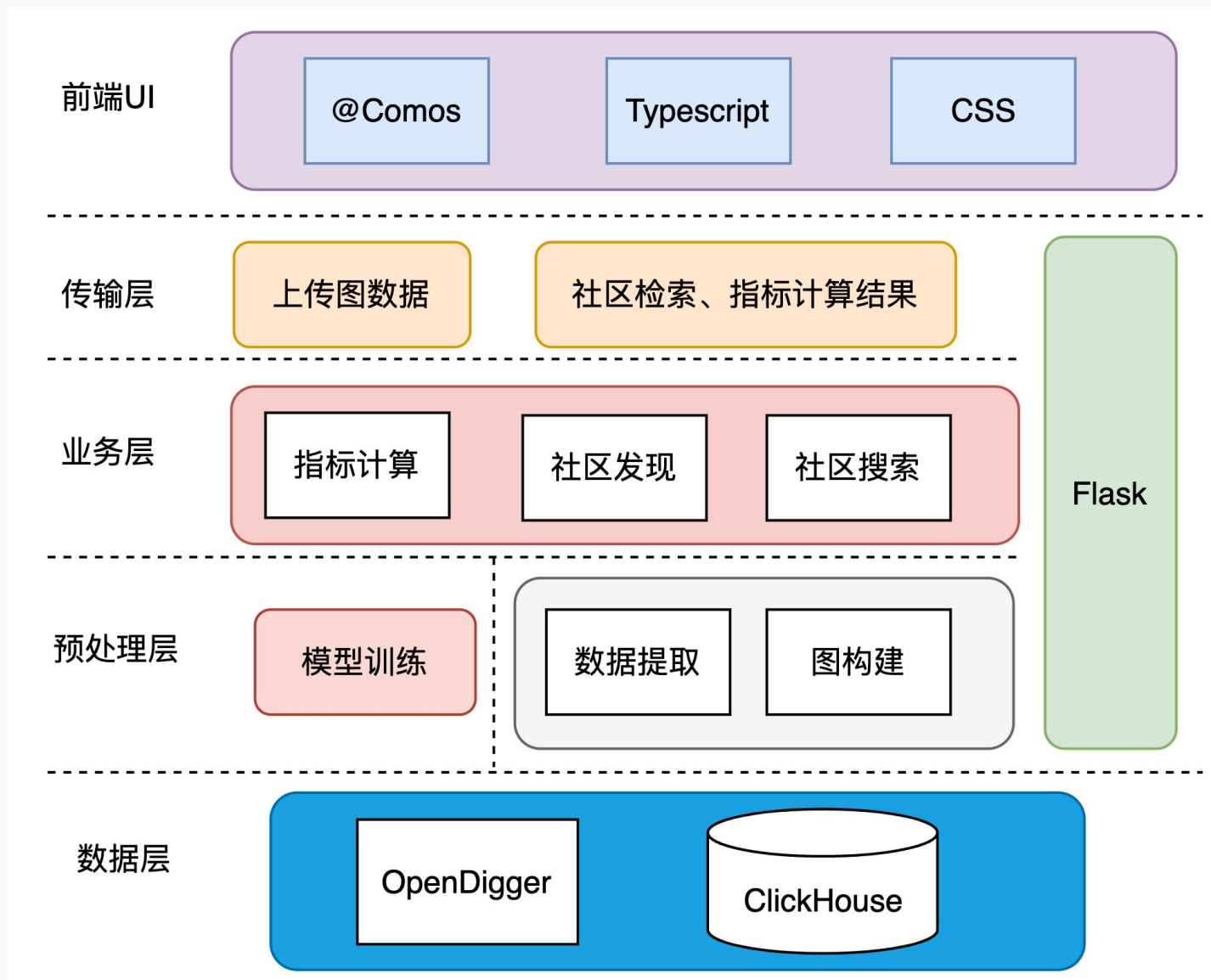
项目不仅仅支持网络的在线动态可视化，还包括社区检索功能、节点网络交互、仓库的分析报表等。需要合理集成功能，保证响应速度，给予用户良好的体验。

第三部分

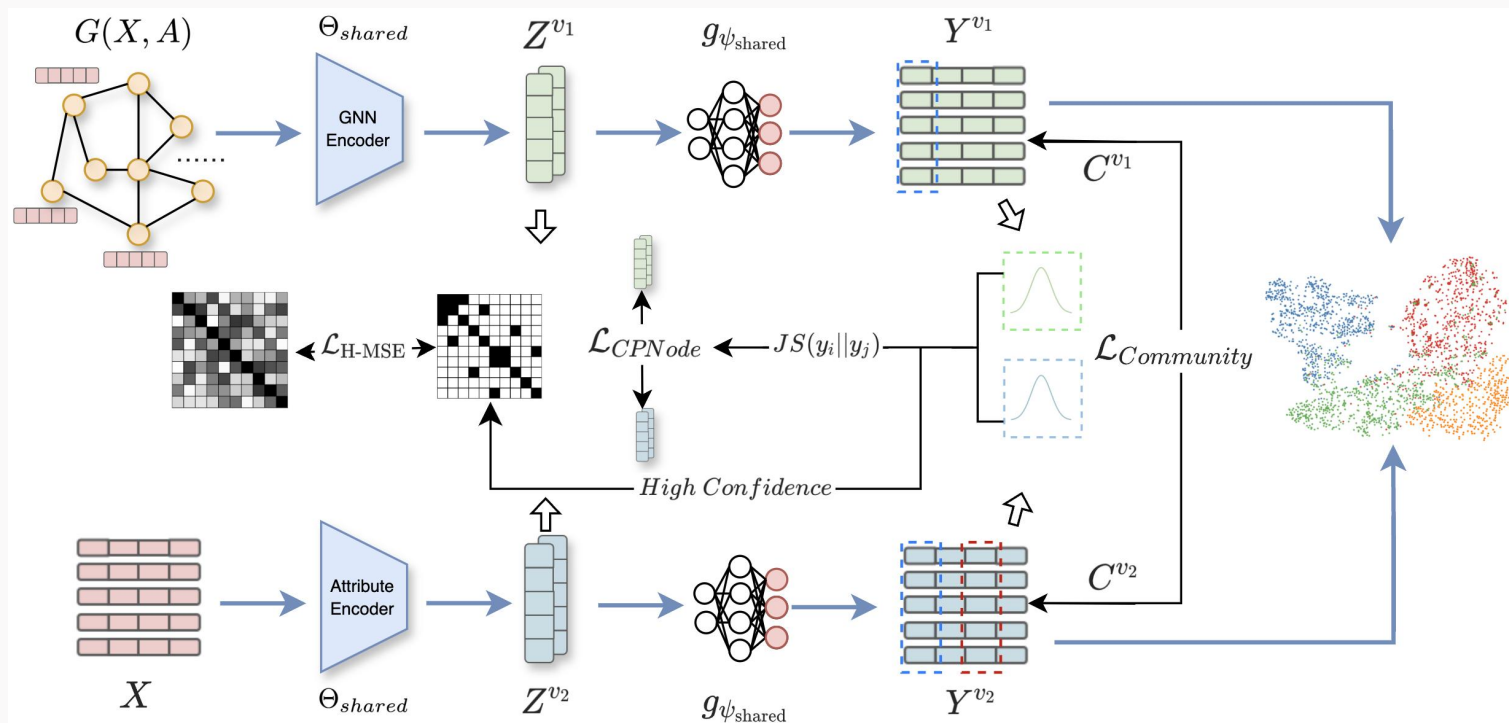
设计方案

整体架构

基于OpenDigger以及ClickHouse的数据进行图构建，并基于构建的图数据对模型进行训练。支持后续的社区发现应用和仓库的指标计算。后端使用Flask，前端使用Typescript以及@Comos大规模图力导向布局可视化库进行展示。



社区发现算法架构



模块一：社区对比模块

对于输入的属性图分别经过两个共享参数编码器得到中间表示后导出节点-社区的样本概率分布，进行社区表示（即概率分布）的对比学习。

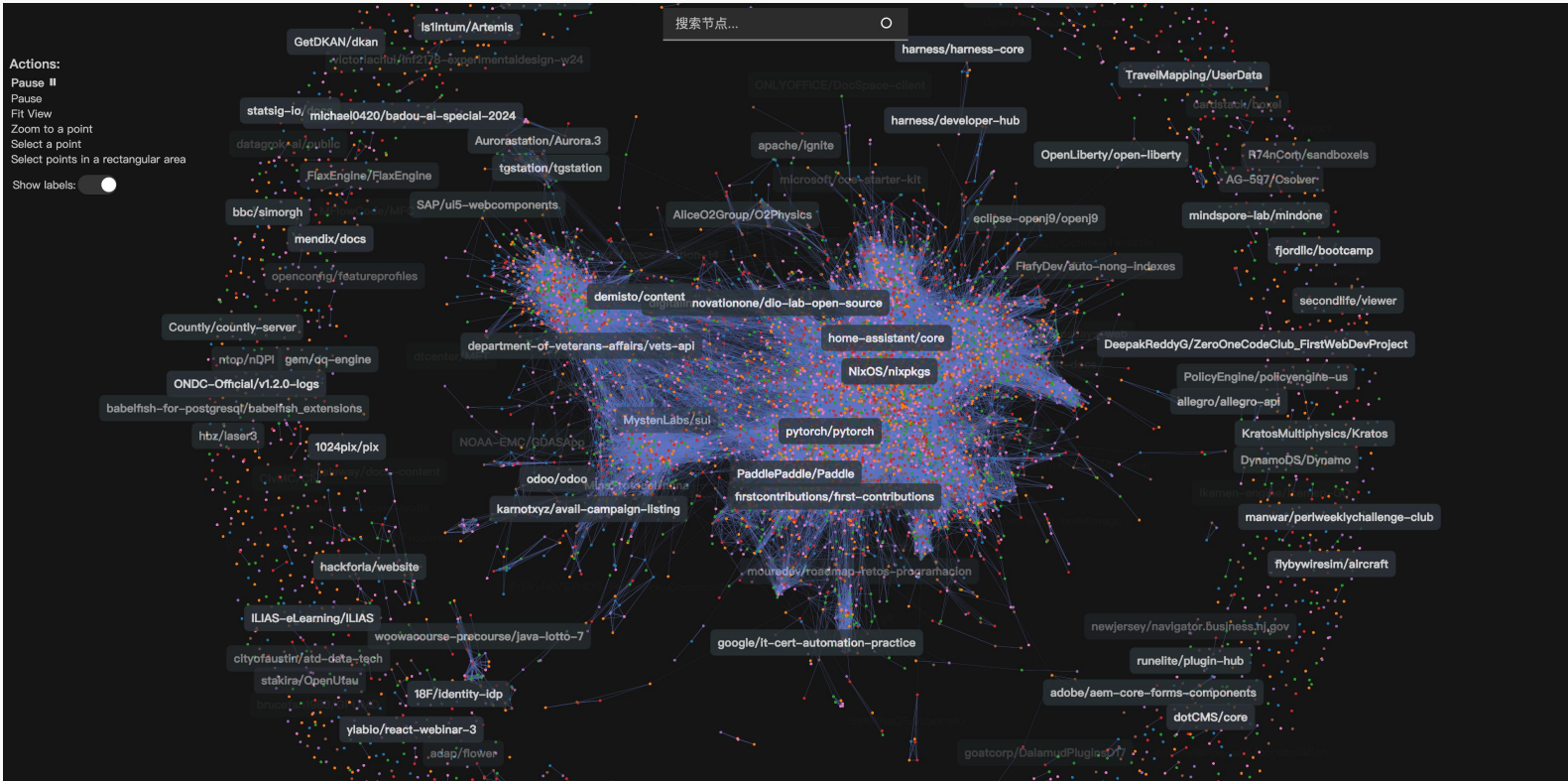
模块二：社区感知的节点对比

通过导出的节点-社区概率分布进一步协同优化中间表示的生成。

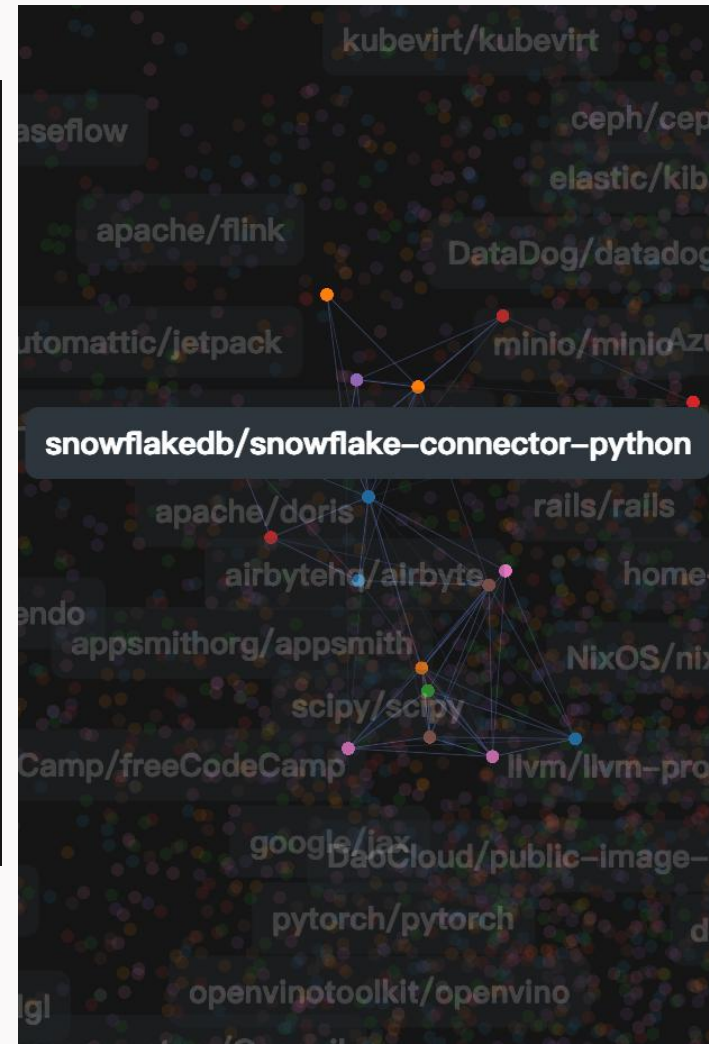
模块三：高置信引导的相关性消减

通过导出的节点-社区概率分布识别潜在的高概率同社区节点，对这一部分节点在表示在学习过程进行优化。

项目预览



后续还包括社区发现节点颜色渲染、搜索节点后的社区搜索结果高亮，以及点击仓库后的报表信息展示。



第四部分

应用场景

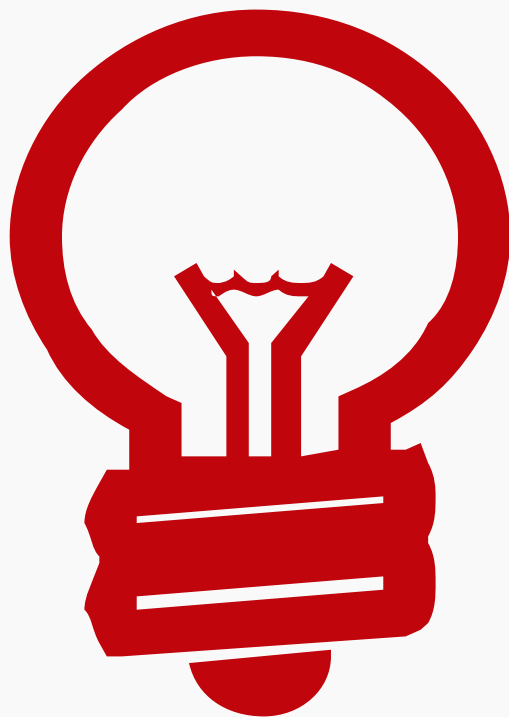
应用场景

用户需要快速探索GitHub的项目，发现其中仓库中的关联以及各个项目的分布情况。

1

对于想要探索一些自己感兴趣方向以及想对其他一些相关仓库做贡献的开发者来说，本项目可以帮助开发者快速找到目标。

3



2

对于仓库的维护和运营者，可以快速了解仓库自己仓库的生态群，同时提升对于组织关联仓库整体情况的感知。

4

对于主流开源项目群之间的关联情况会有直观地展示，通过图可视化快速定位到项目协作群，从而可以用社区的视角来洞察开源。



感谢

THANK YOU