

战疫先锋APP数据分析： 用户自组织行为的特征与演变

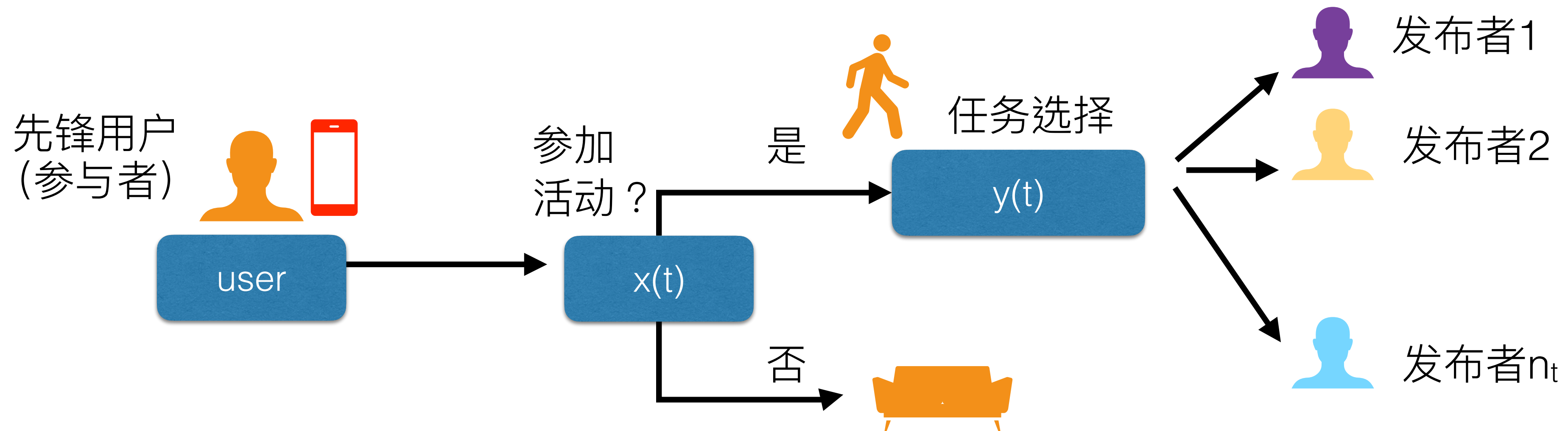
2020/12/9

Overview

- 概述
- 自组织力的演变 （用户-发布者关系）
- 其他自组织行为特征：
 - 任务主题分析
 - 用户-任务关系的模式分析

个人与群体行为

先锋数据提供了个人行为的详细记录：每个用户每天选择参加的志愿活动



研究目的：

- 从大量个人行为数据中，提取群体的行为**共性与特征**。
- 由下至上的新型社会合作机制是否存在**自组织行为**？
- 如何衡量各地区的**自组织力**？影响自组织力的因素有哪些？

什么是自组织行为？

- 自组织行为(self-organizing behavior)：无外在指令的情况下，个体的局部行为形成某种社会秩序的过程
 - **初期系统是无序的**：新平台，新用户，突发事件（疫情）
 - **通过局部相互作用产生有序协调**：用户形成稳定合作关系与分工
 - **开放的系统**：成员不固定，与外界需求与环境互动

<https://zh.wikipedia.org/wiki/自我組織>

von Foerster, Heinz; Pask, Gordon (1961). "A Predictive Model for Self-Organizing Systems, Part I". *Cybernetica*. **3**: 258–300.

如何衡量自组织行为的“自组织力”？

- **组织性**的增加反应为用户行为不确定性的降低
- 通过给定用户的**条件熵**计算用户行为的不确定性

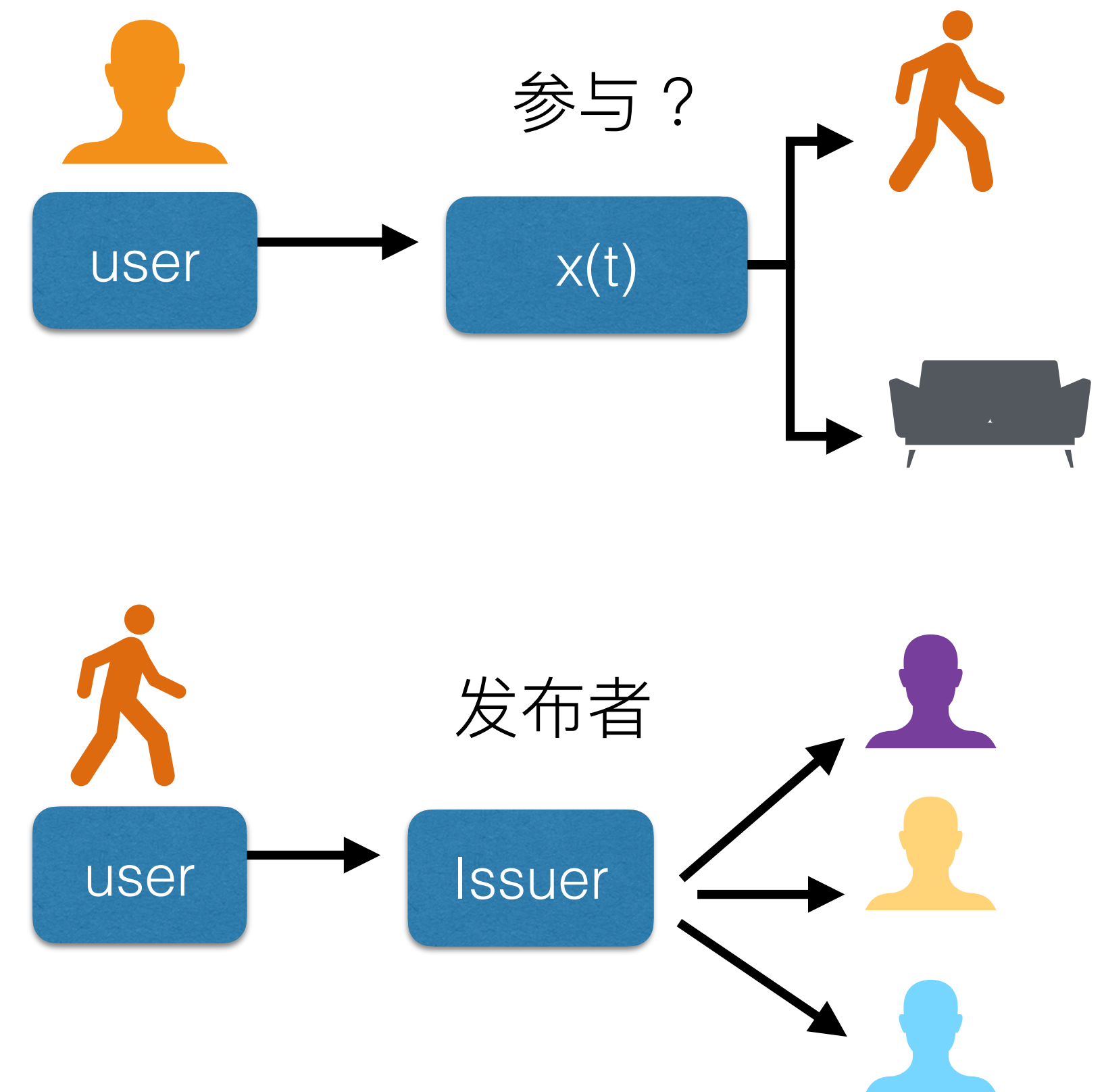
例如：

- 用户参与的不确定性：

$$\frac{H(X|User)}{H_{max}(X|User)}$$

- 任务（发布者）选择的不确定性：(假设X=1)

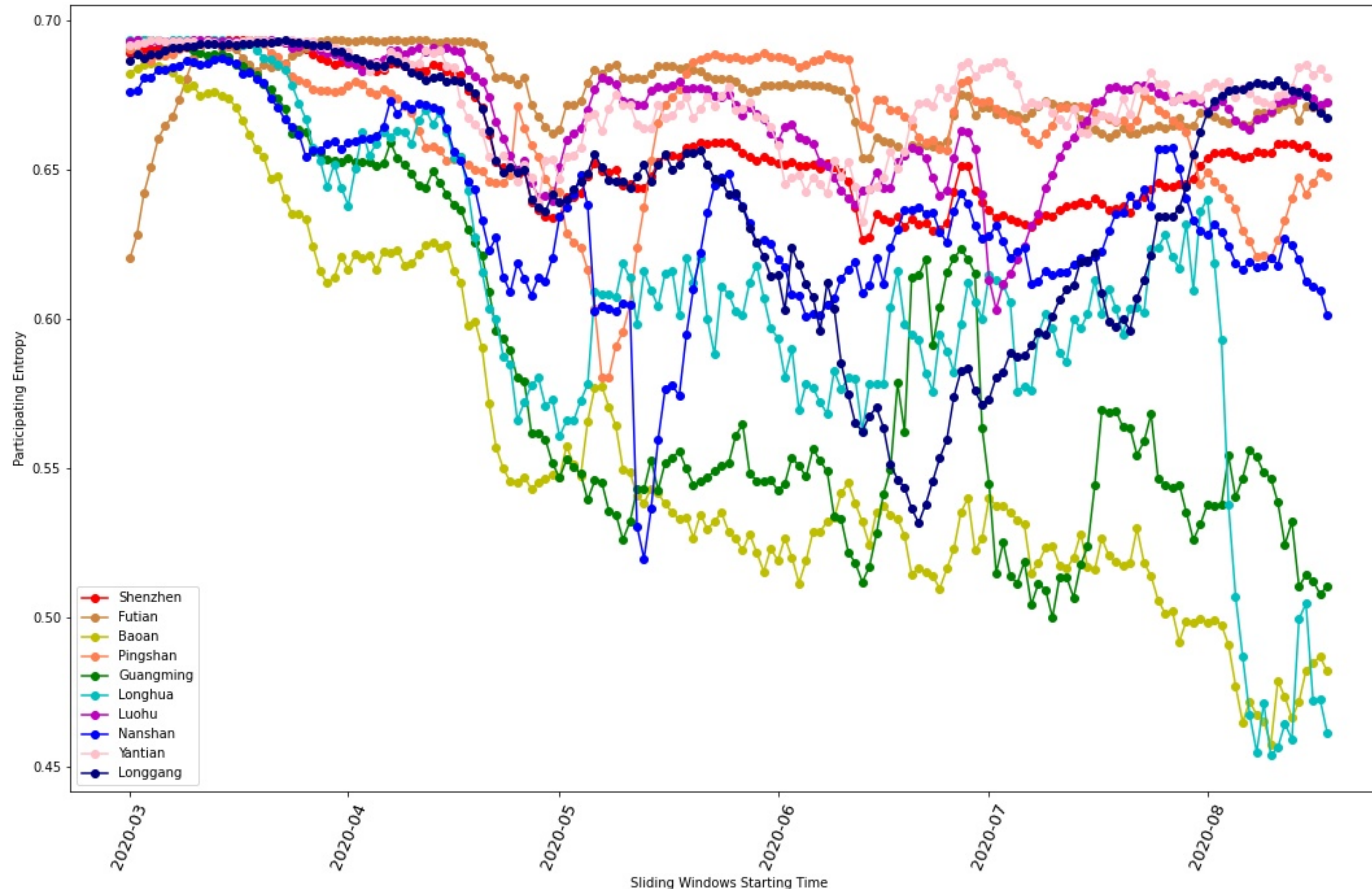
$$\frac{H(Issuer|User)}{H_{max}(Issuer|User)}$$



用户参与的不确定性

时间窗口：2周

对象：参与任务15次以上的用户

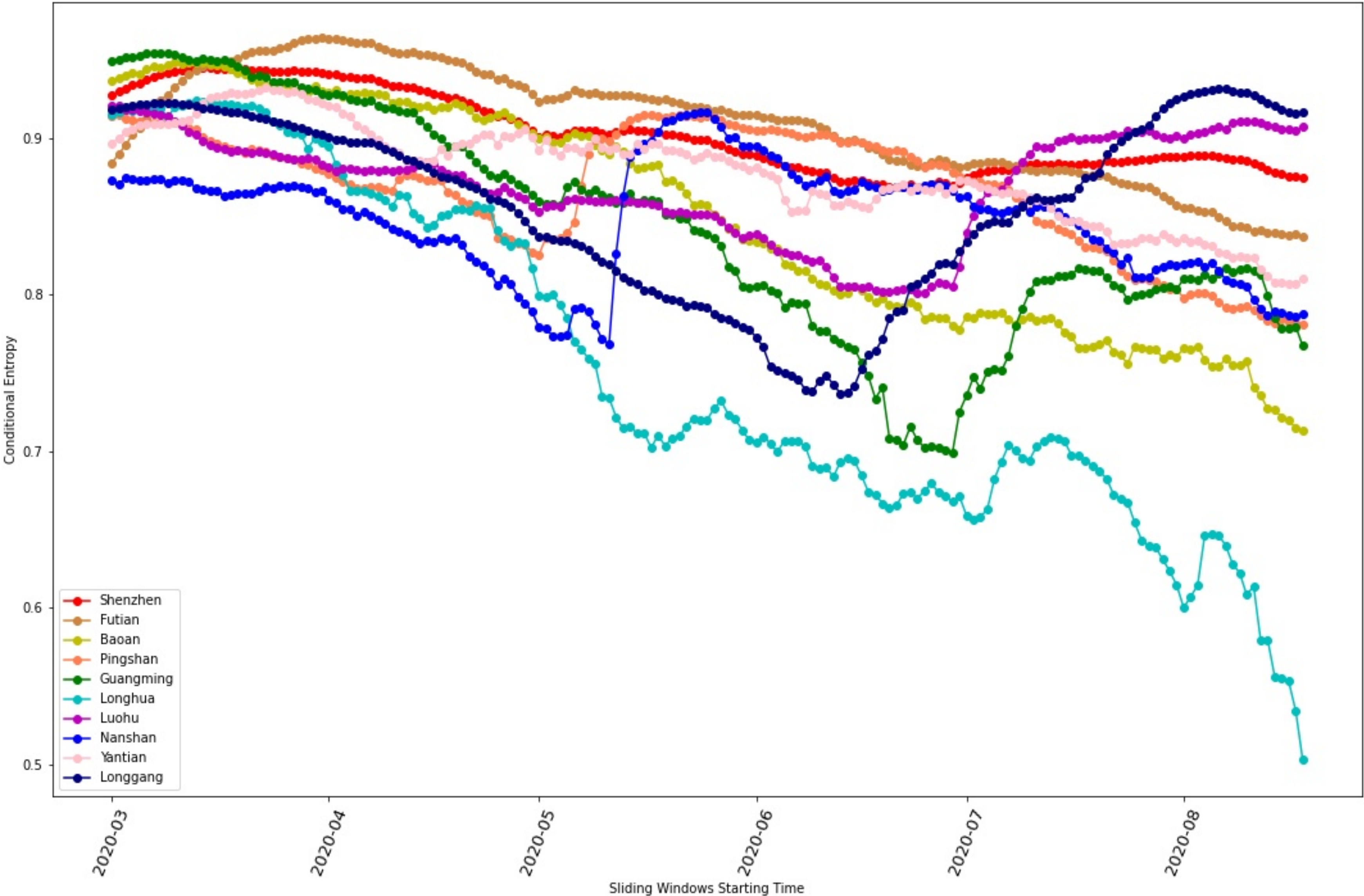


- 大多数区的用户参与的不确定性基本稳定
- 光明、保安持续下降
- 龙华8月份有大幅度下降

发布者选择的不确定性

时间窗口：2周

对象：参与任务15次以上的用户



- 整体呈下降趋势
- 南山、坪山在5月份不确定性上升
- 罗湖、光明、龙岗6-7月份出现上升
- 龙华区8月出现大幅下降

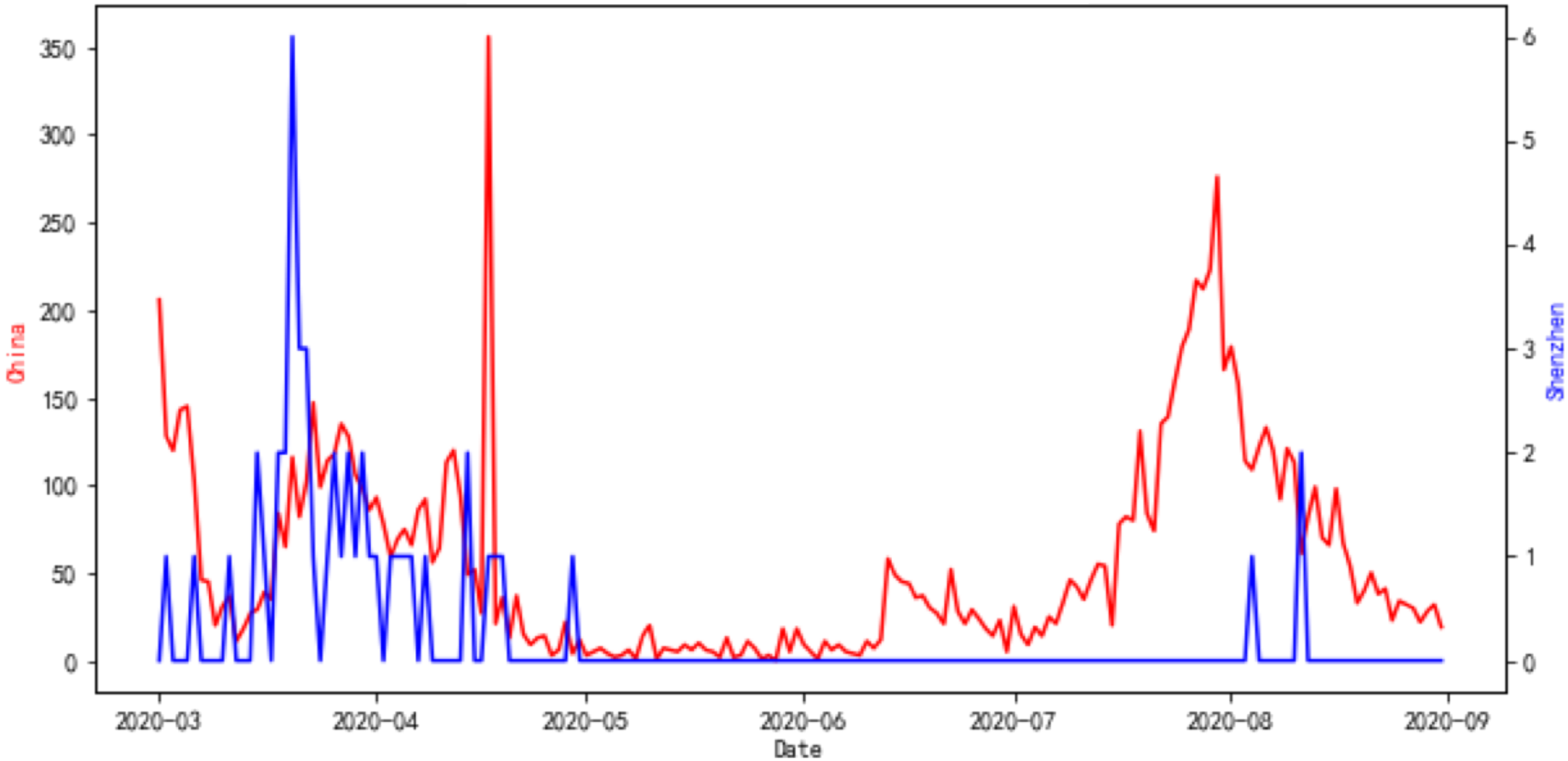
影响自组织力的因素是什么？

影响因素

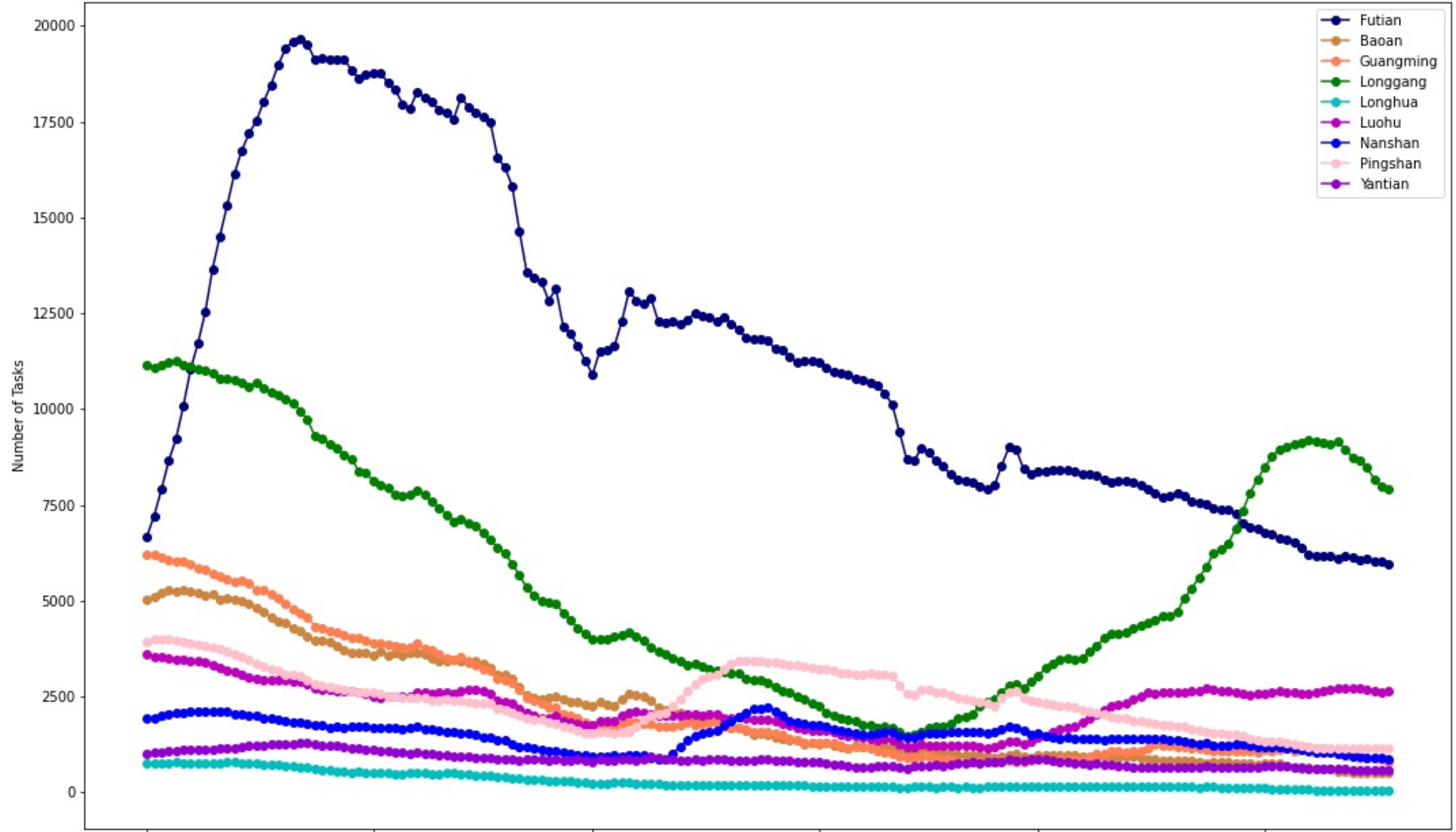
- 疫情：国内疫情动态
- 疫情政策：复工复产、复课
- 社会：五一假期、两会
- 环境：雨天

Event Type	Name	Start time	End time 1	End time 2
Epidemic Outbreak	Covid-19 Cases Increase	2020.3.18	2020.3.21	NaN
Political	International Workers' Day	2020.5.1	2020.5.5	NaN
Epidemic Related Policy	Schools & Colleges Re-opening	2020.4.27	2020.5.9	2020.5.28
Political	The Two Sessions	2020.5.22	2020.5.28	NaN
Environmental Factors	Rainy Weather	2020.6.4	2020.6.14	2020.6.25

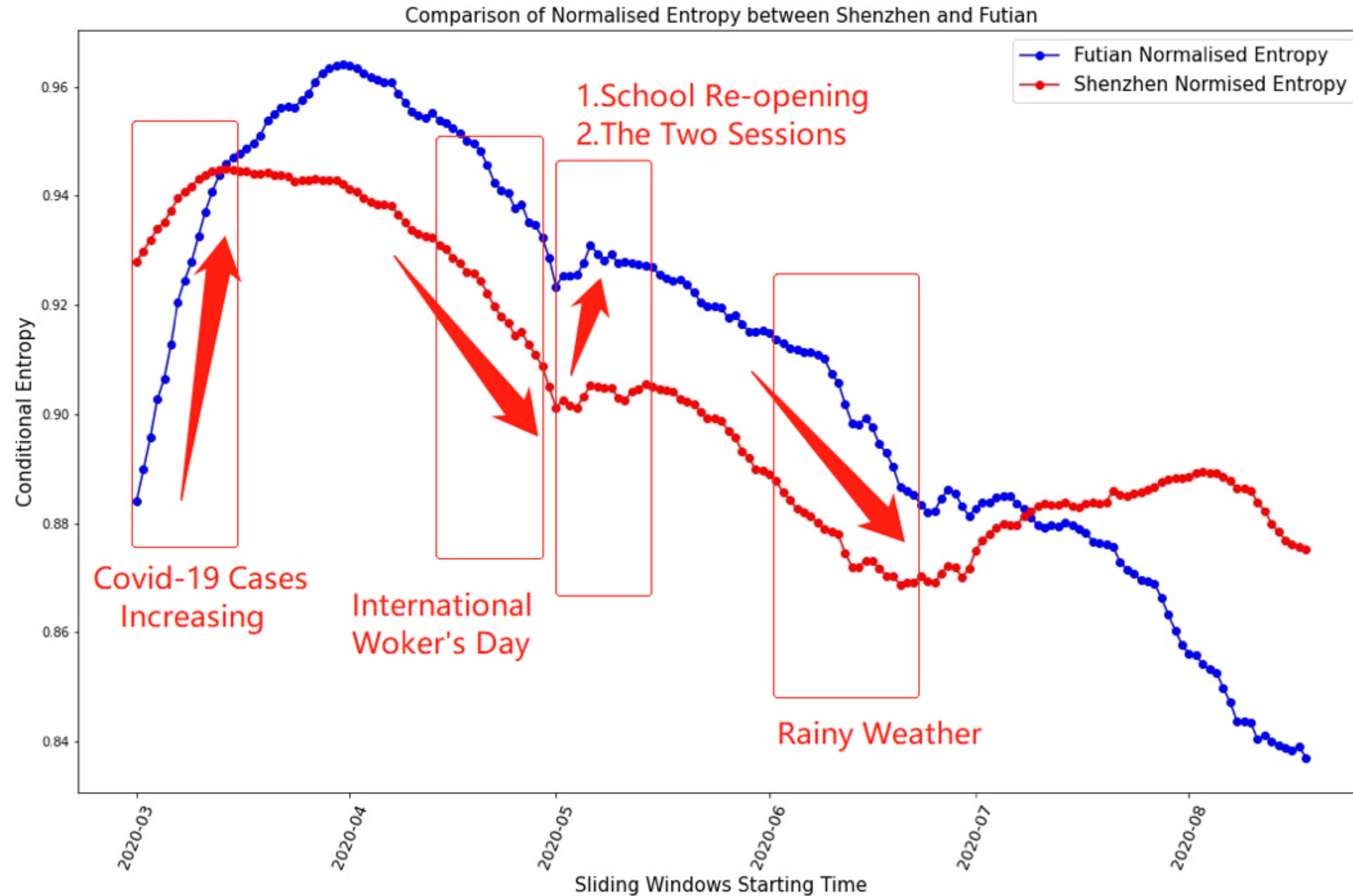
全国与深圳疫情数据



各区任务总数



发布者选择的不确定性（深圳市vs.福田区）



Overview

- 概述
- 自组织力的演变（用户-发布者关系）

其他目标：任务主题、任务模式、任务可预测性

- 其他自组织行为特征：
 - 任务主题分析
 - 用户-任务关系的模式分析

任务主题分析与分类

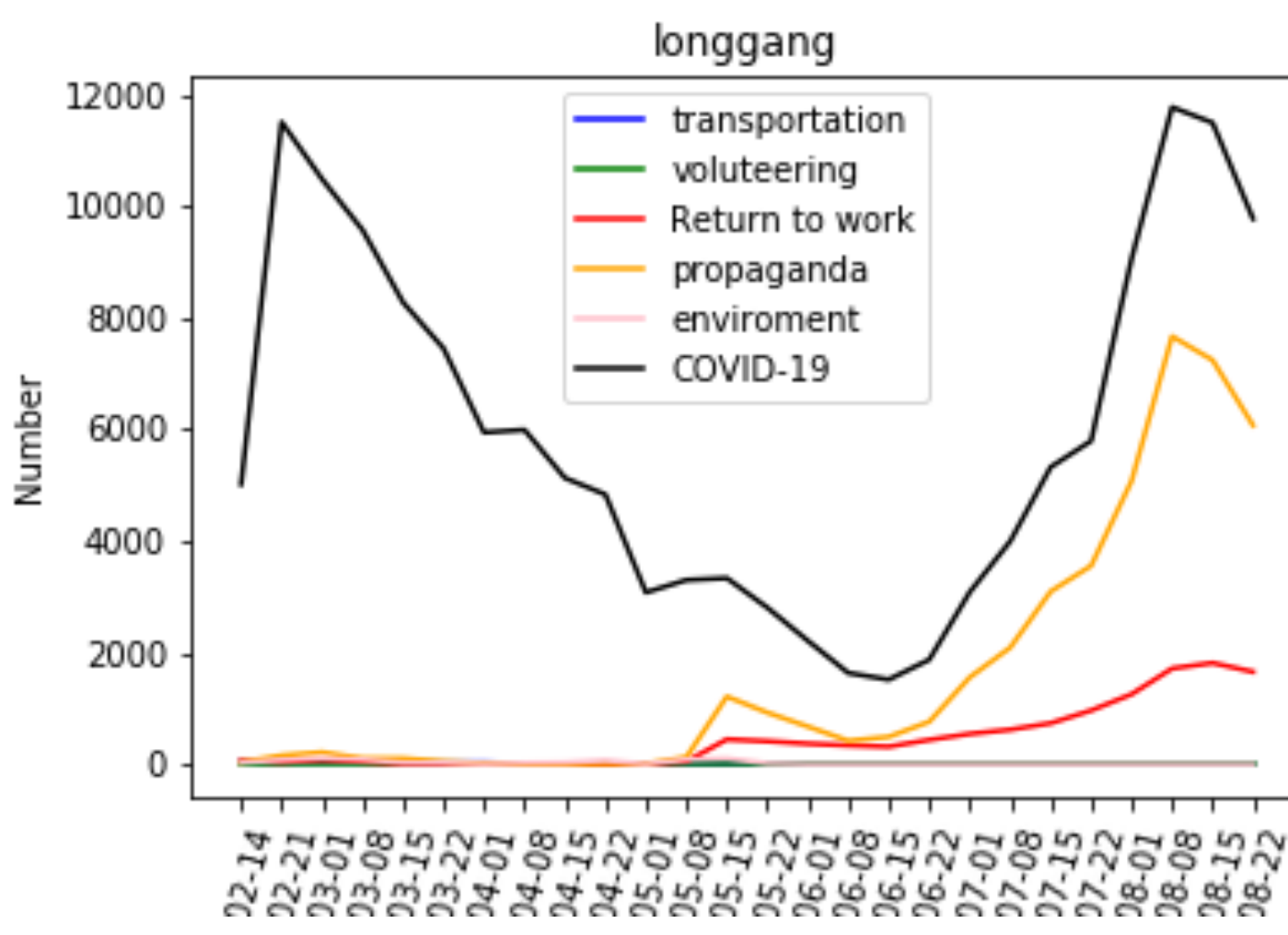
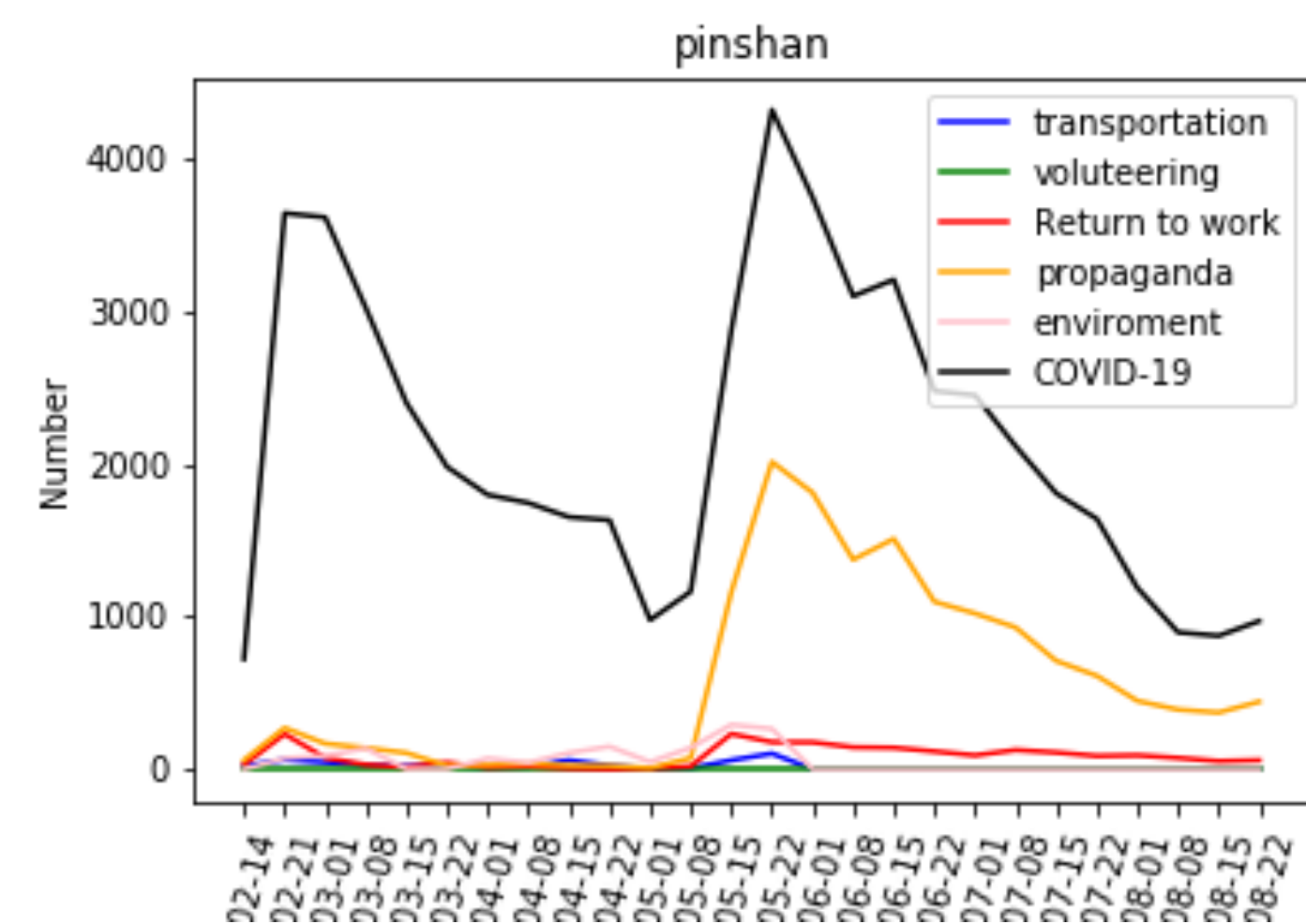
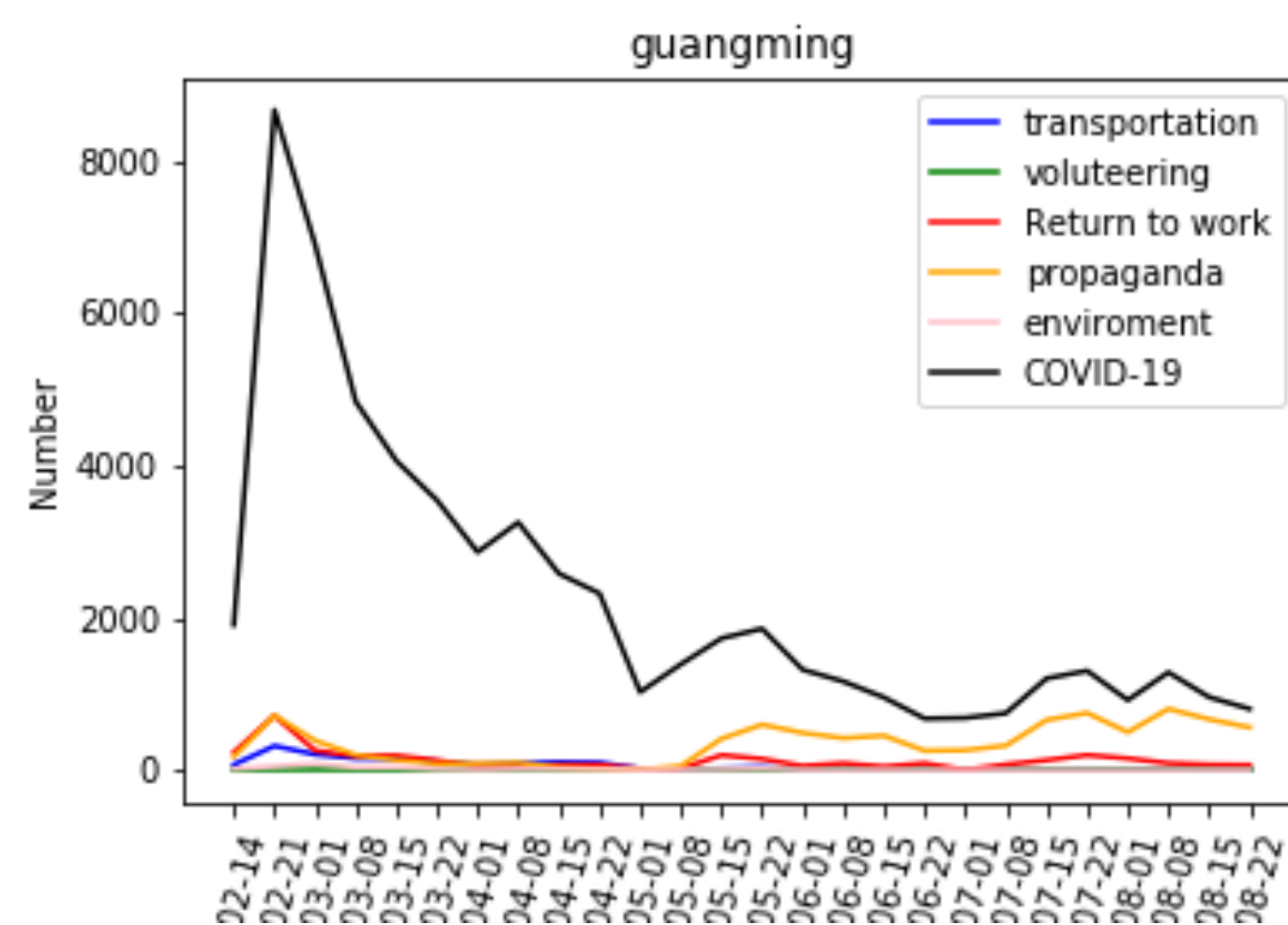
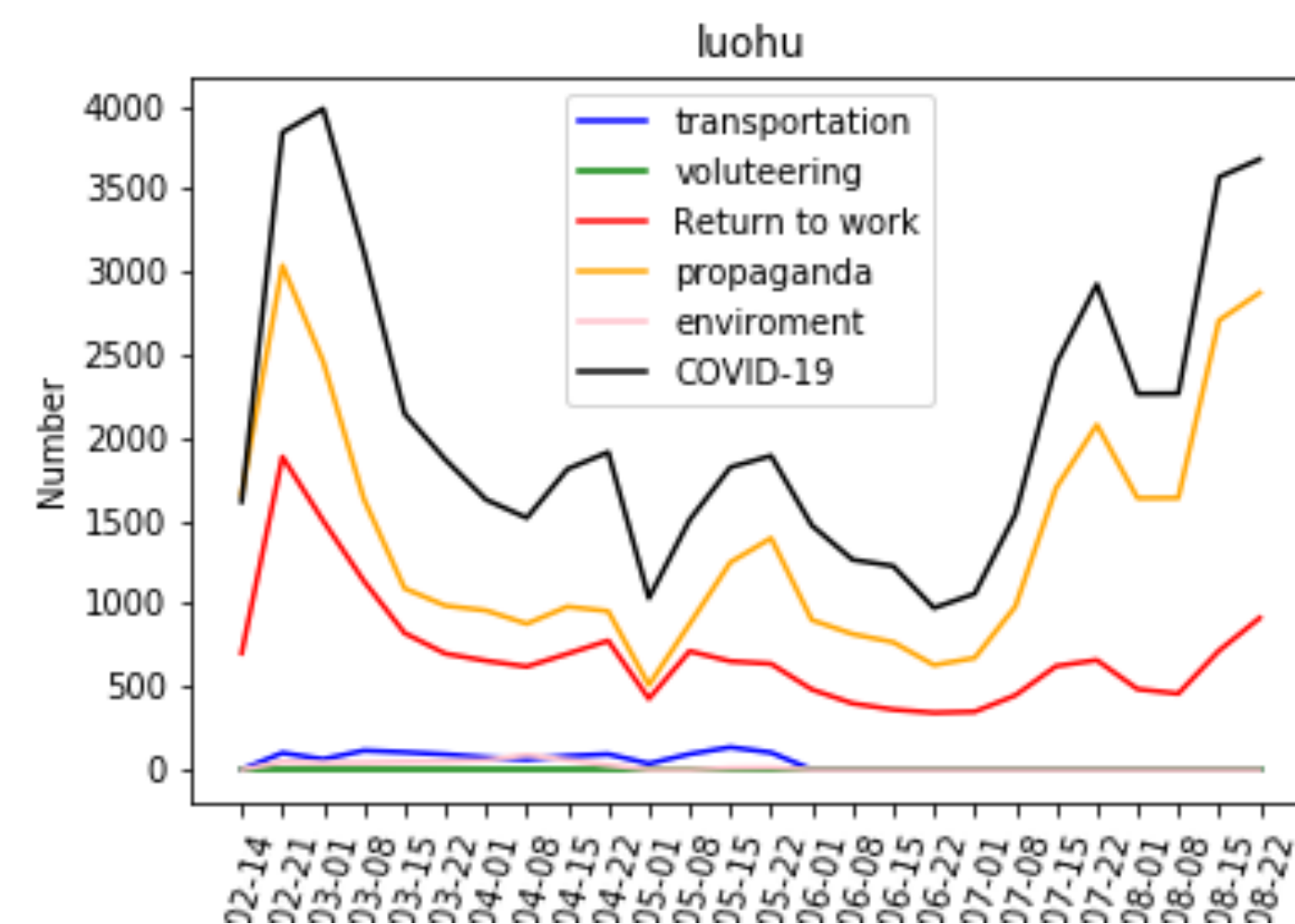
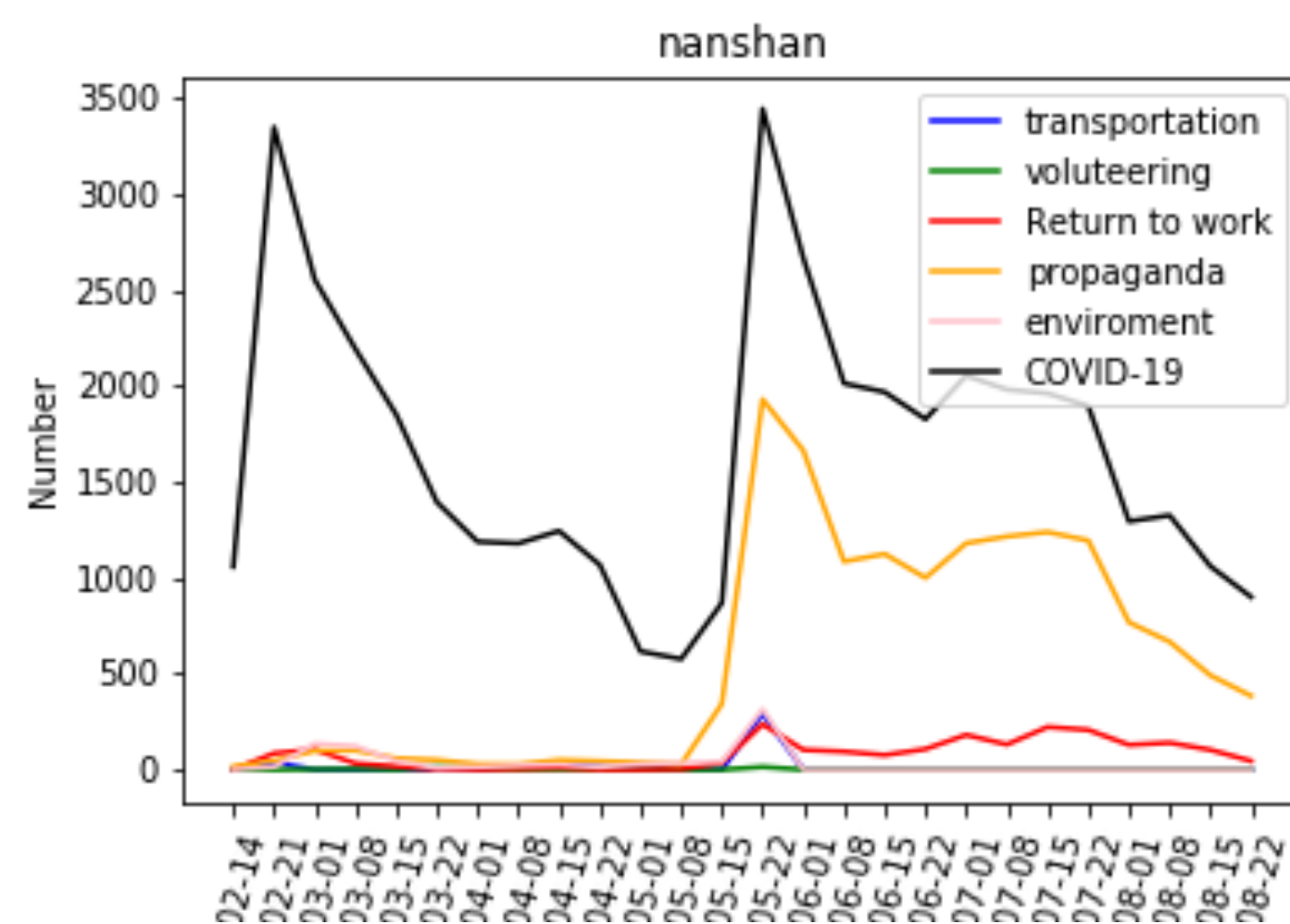
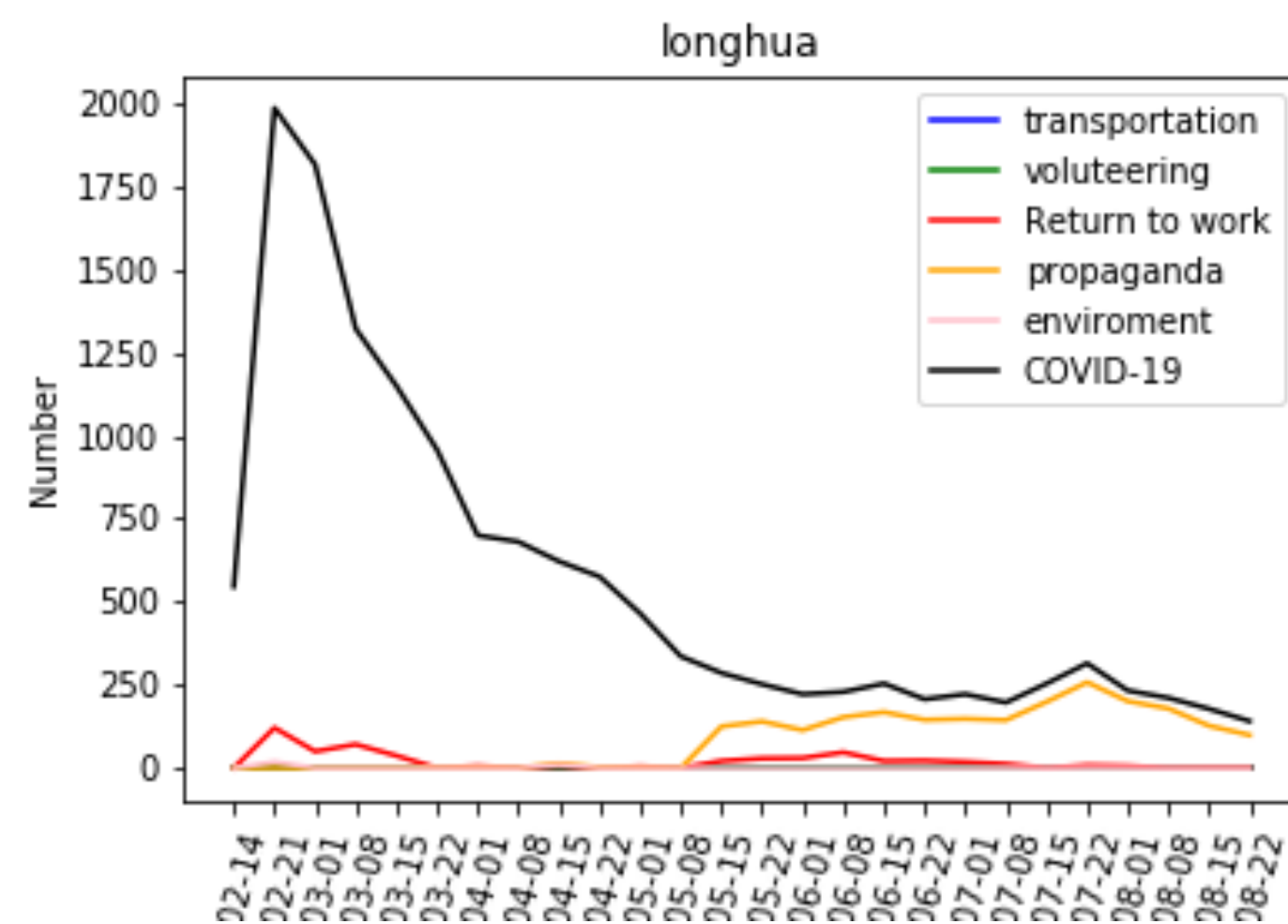
- 基于关键字的任务主题分类

防疫	复工复产	交通安全	宣传	环境建设	公益事业
143561	14951	8747	54456	13906	1441
卡口防疫、 无疫小区、 防疫宣传、 防疫巡查	复工复产、 复工生产检 查、复工复 产巡查	交通引导、 文明劝导、 安全检查	防疫宣传、 政策宣传、 公益宣传、 四点半课 堂、党课	垃圾分类、 市容卫生、 清洁家园	义工义诊、 心理疏导、 技能培训、 扶贫助学、 文体活动、 关爱老人、 走访慰问

- 基于LDA的任务主题分析 (in progress)

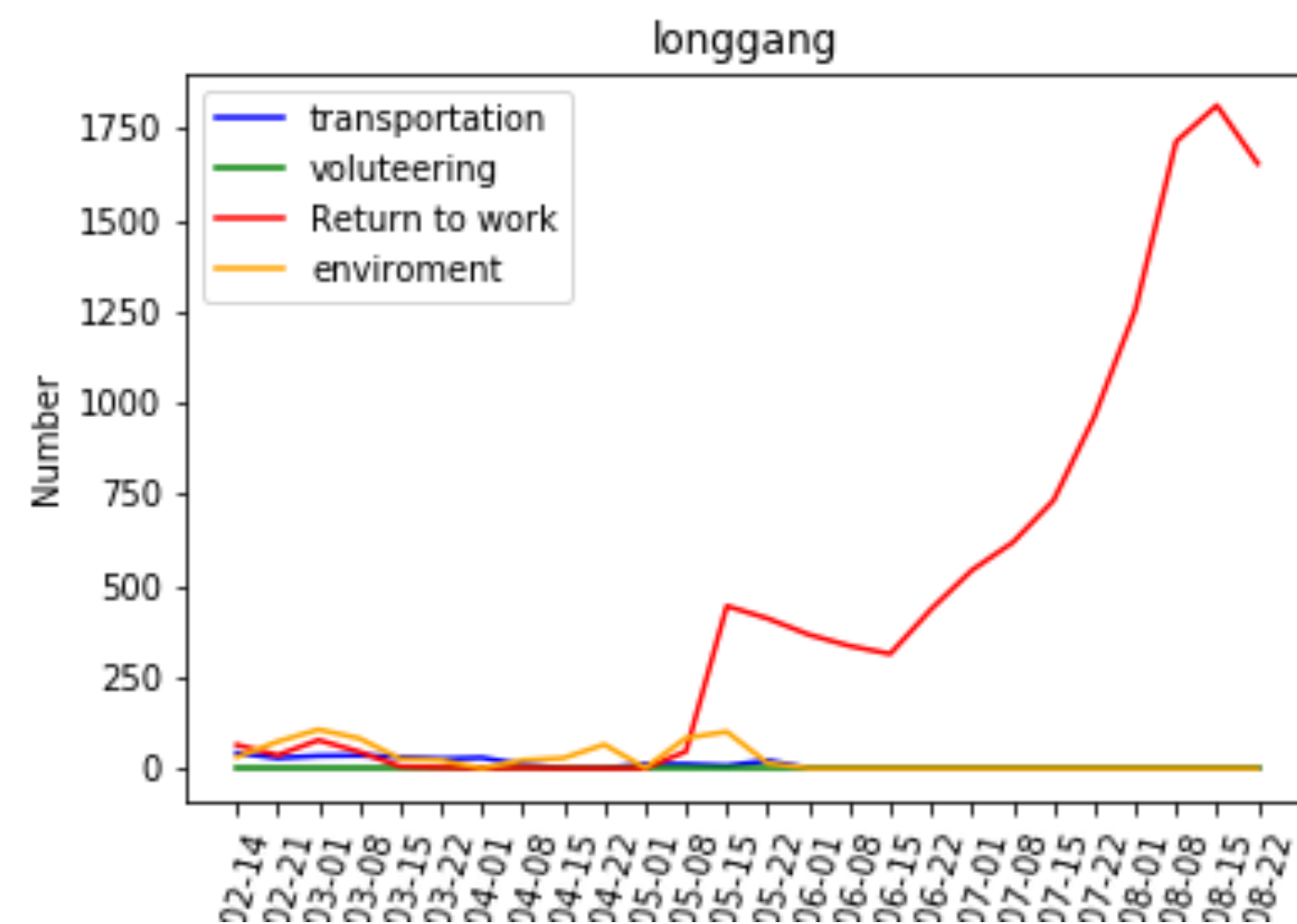
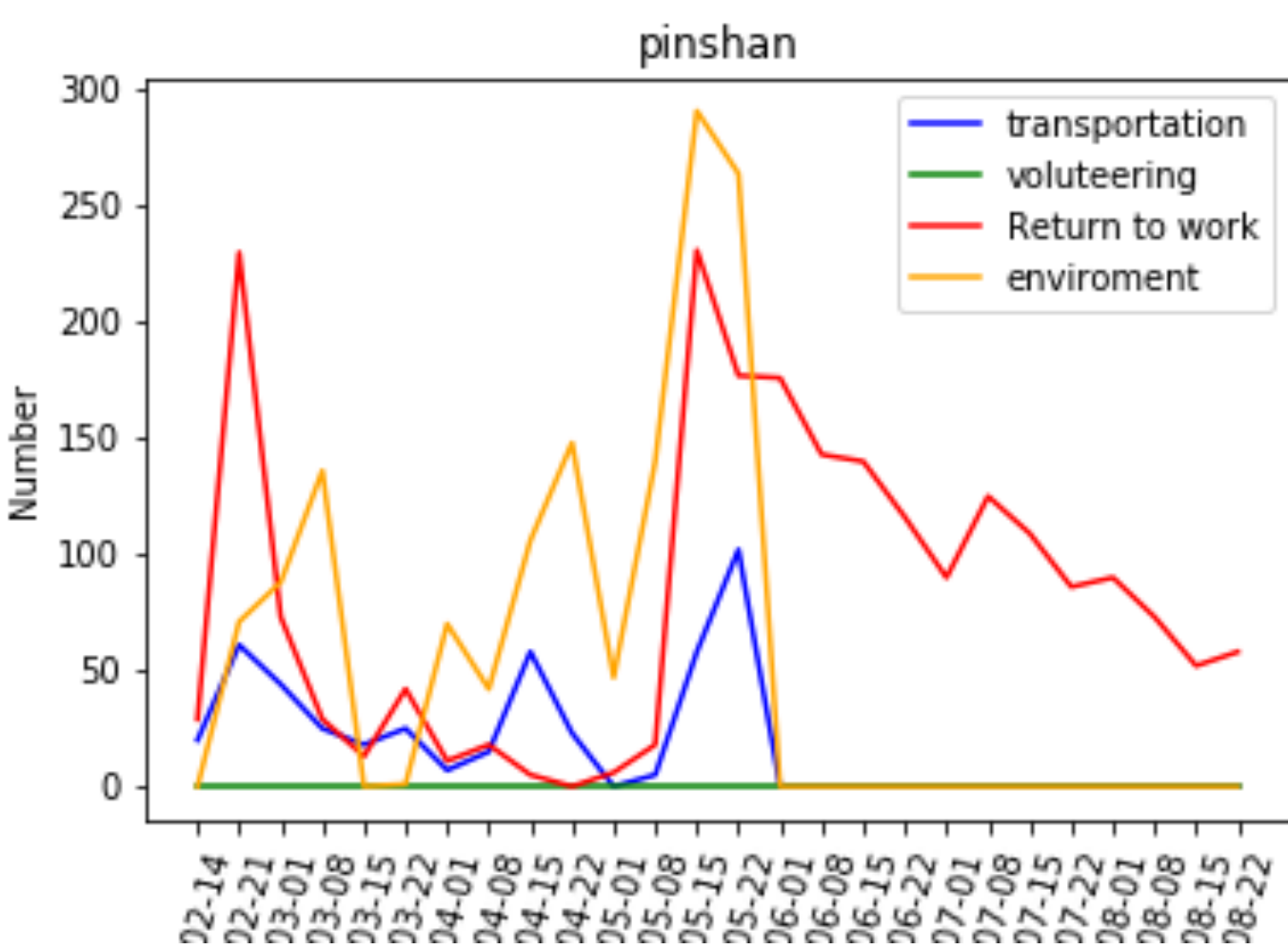
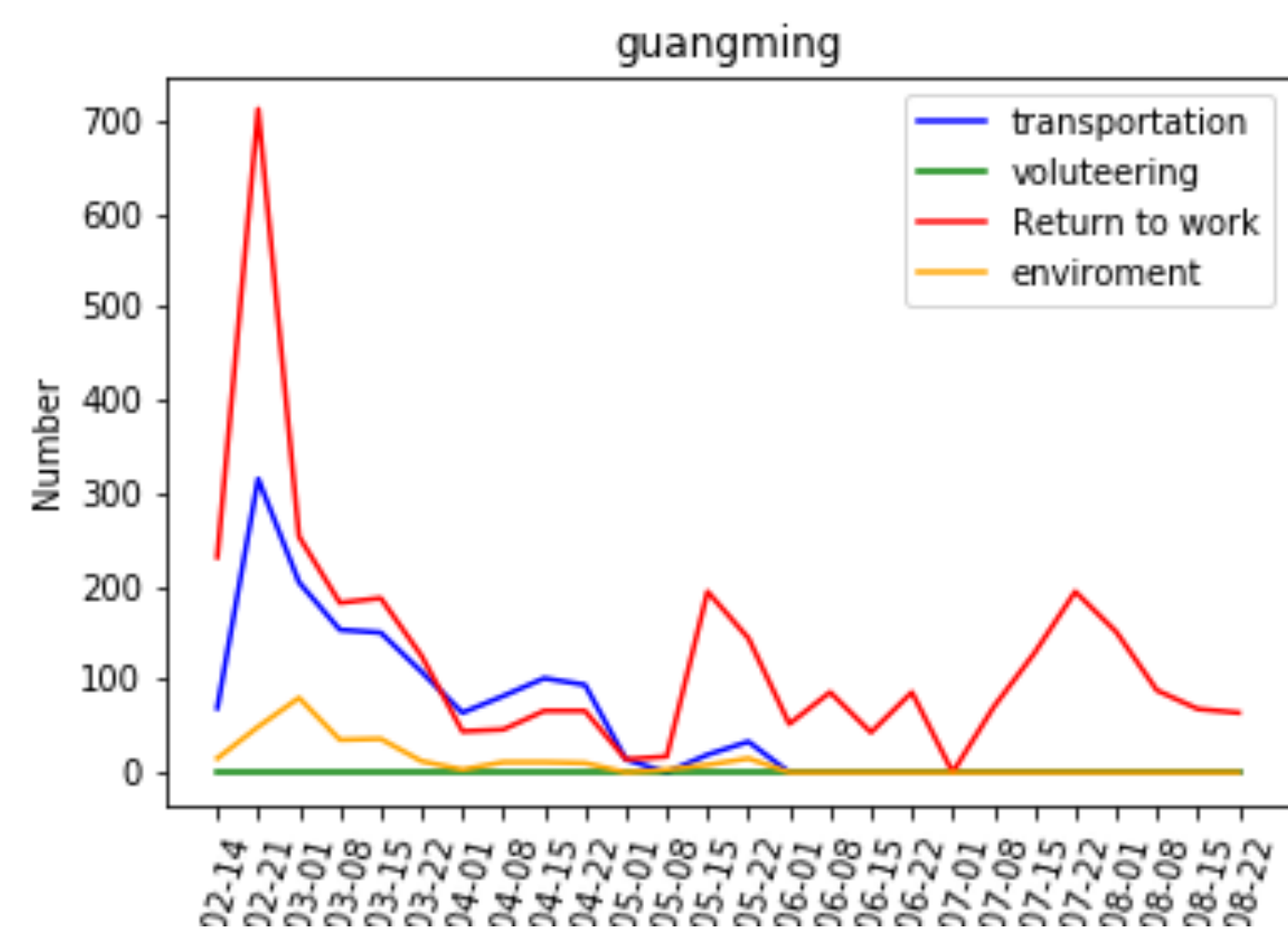
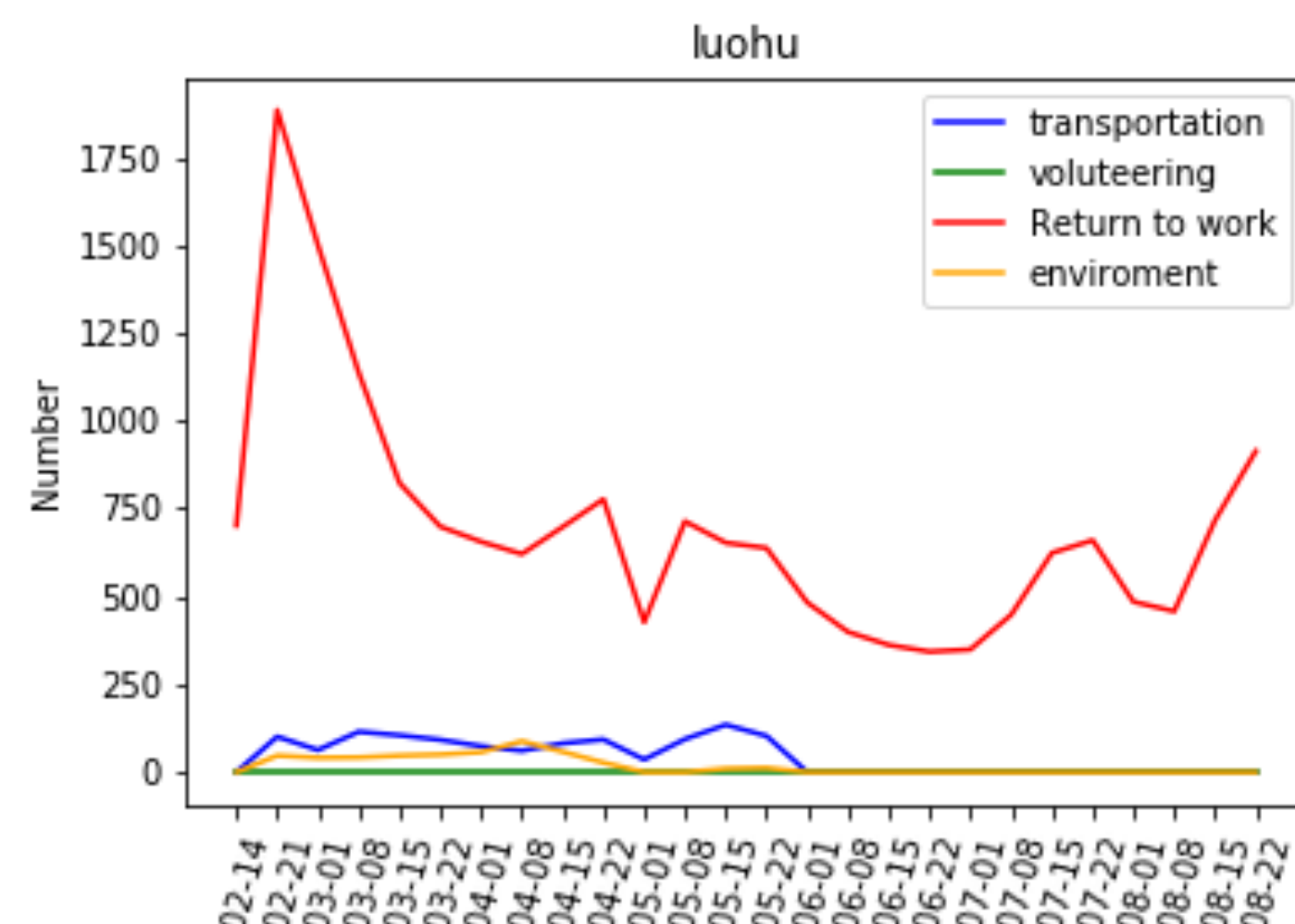
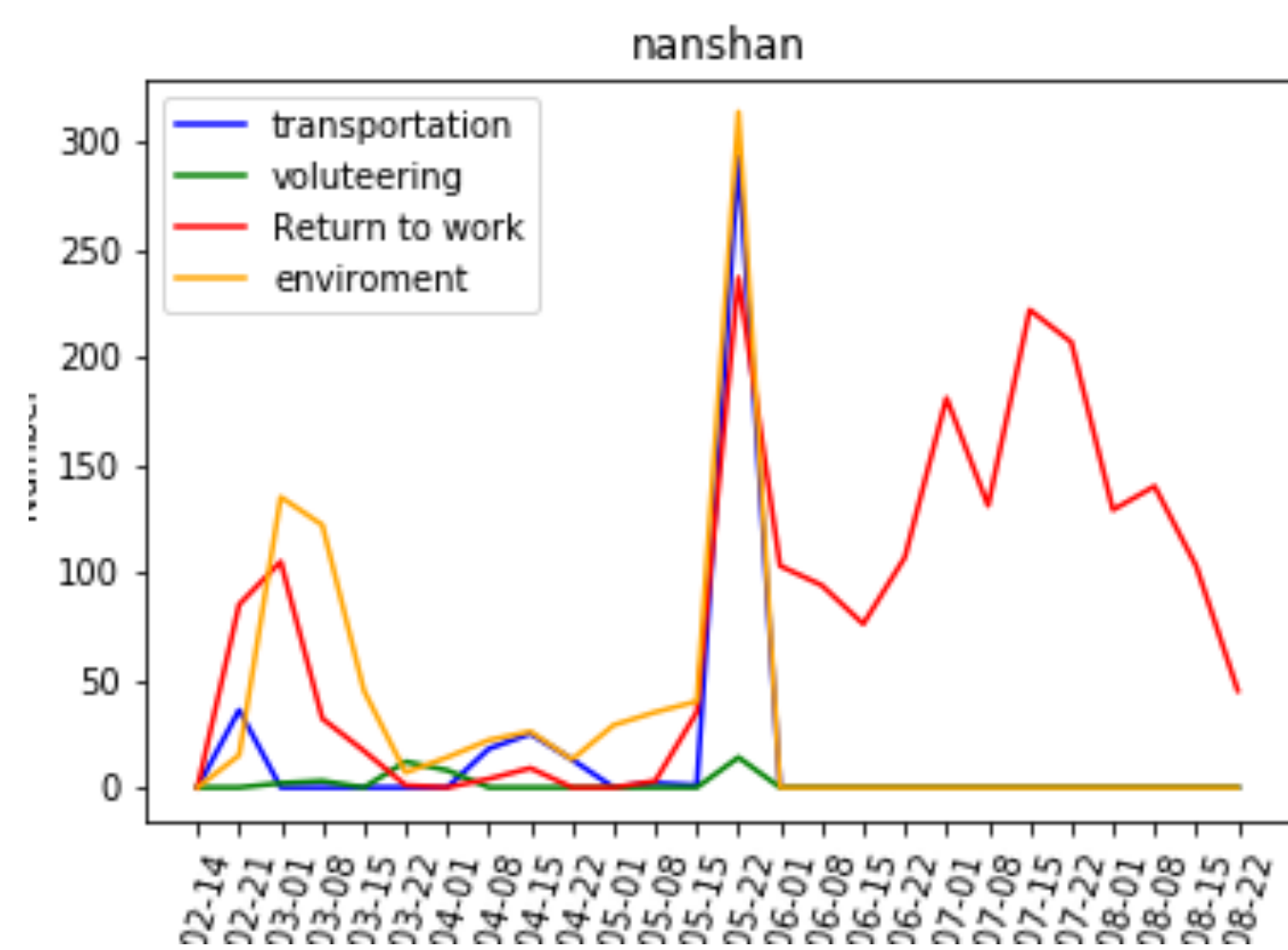
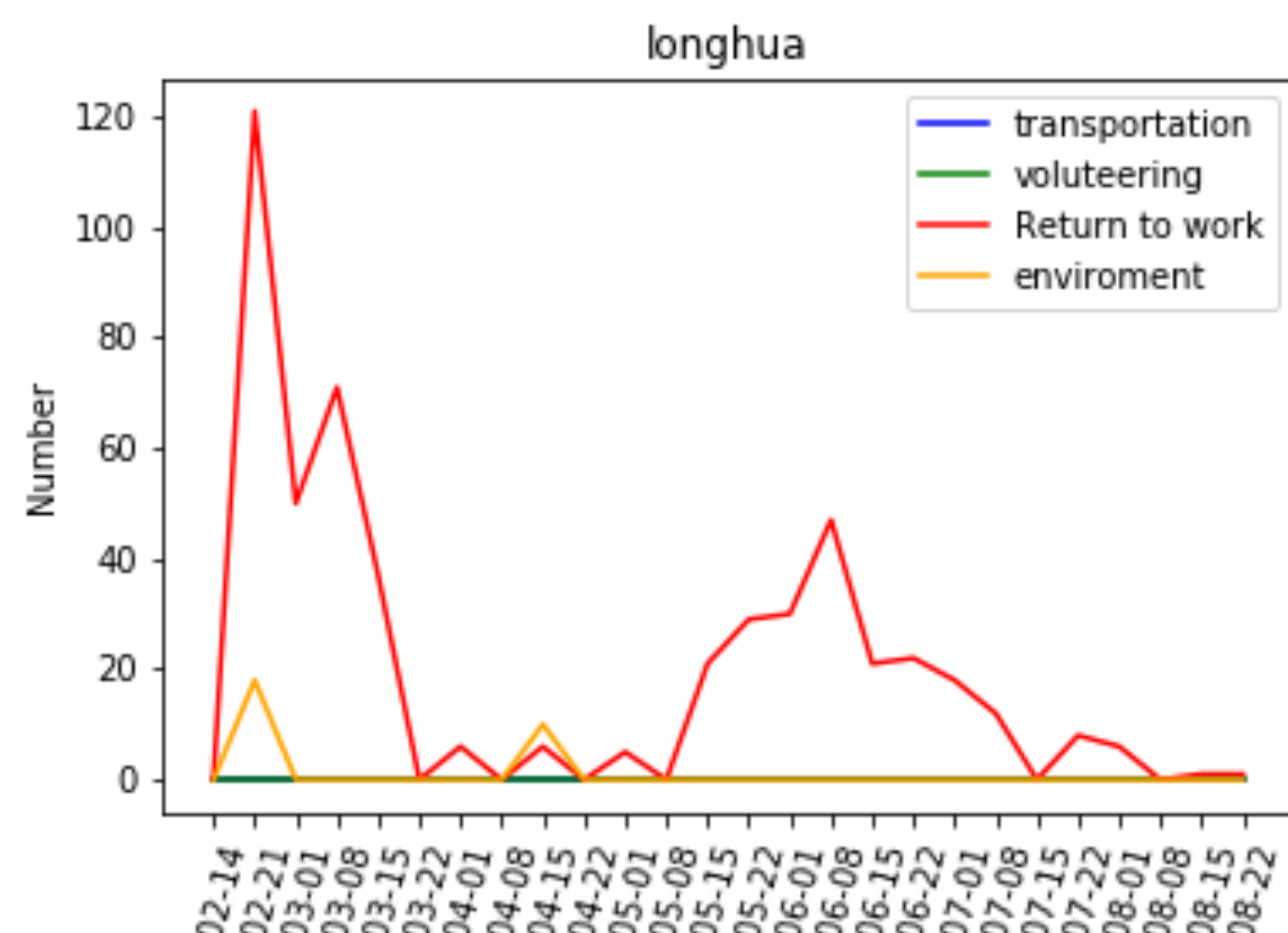
任务主题趋势

时间窗口：1周



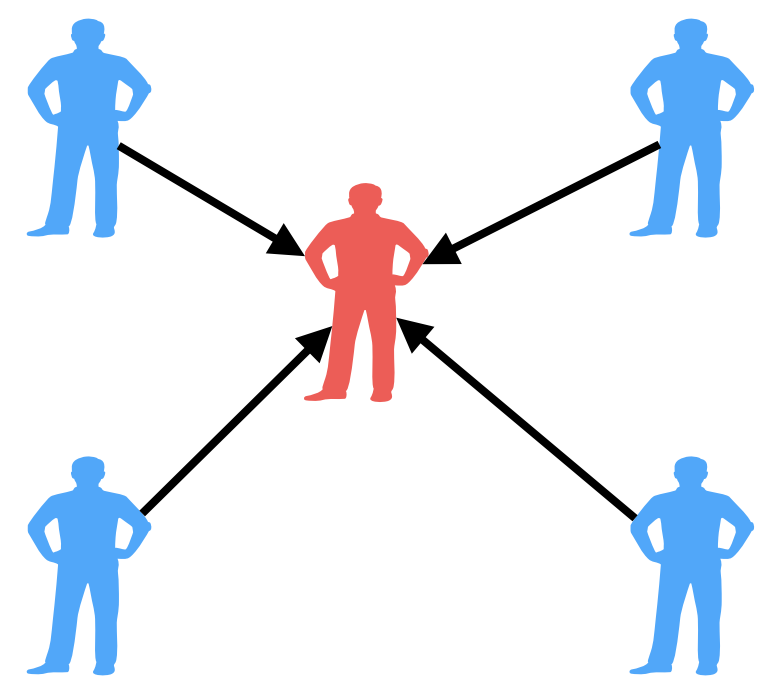
时间窗口：1周

任务主题趋势 (去除top2)

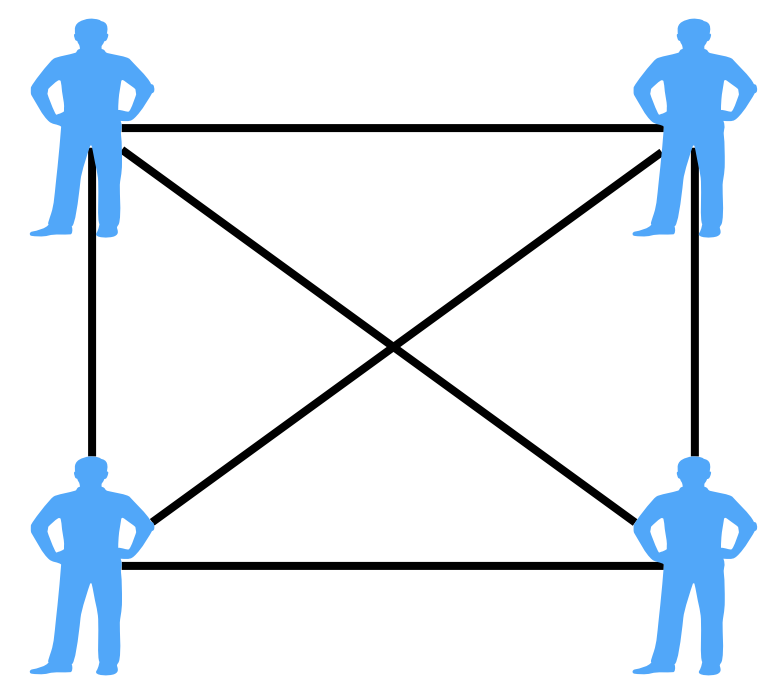


用户合作关系图

- 参与者 - 发布者 (p->i)
 - 有向

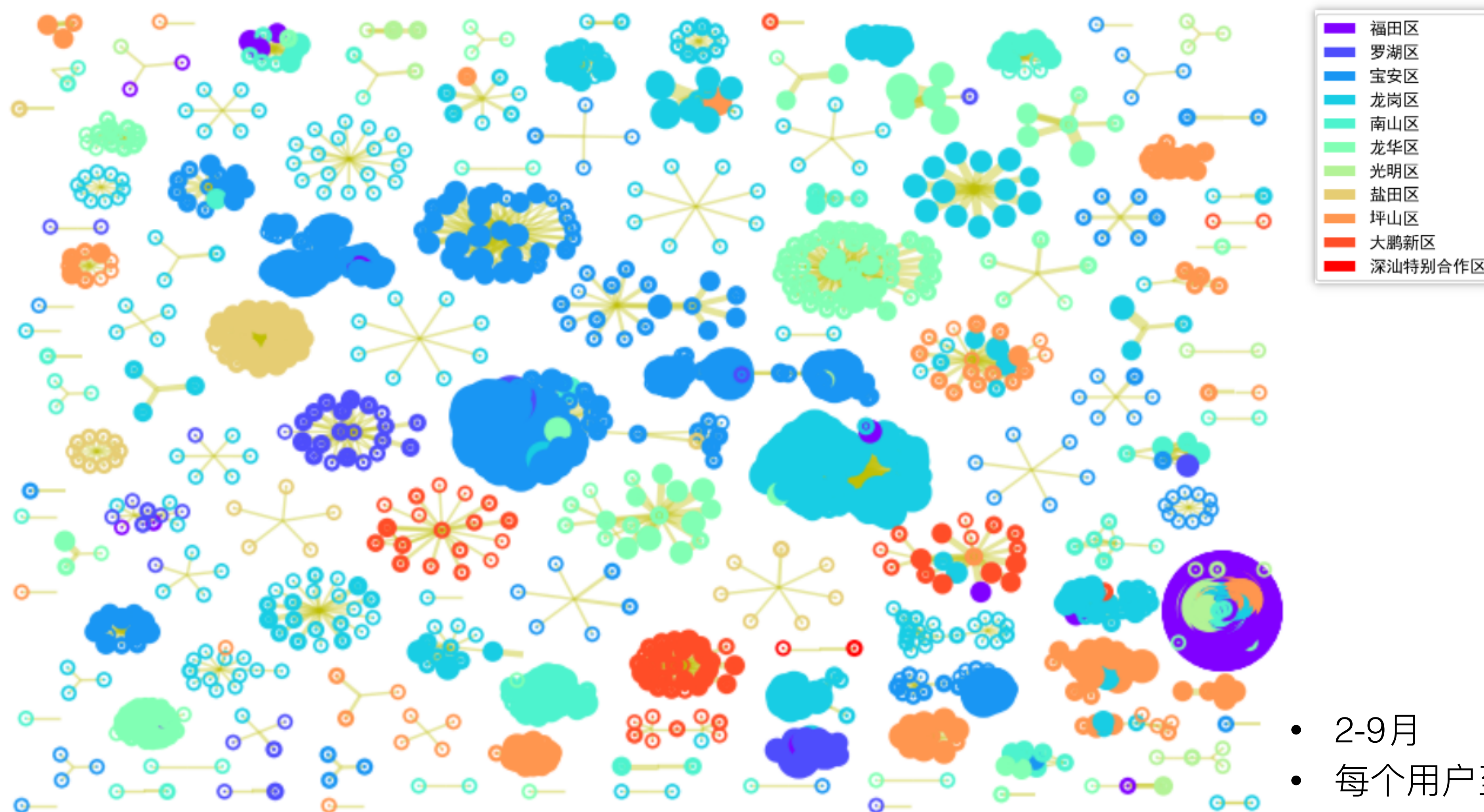


- 参与者 - 参与者 (p->p)
 - 无向



	#of nodes	#of edges	Average Degree	Average Path length	Diameter	Modularity	#of Communities	#of WCC	#of SCC
p→i	72,698	104,745	1.441	1.998	7	0.954	486	449	1232
p-p	69,428	1,030,875	29.696	14.1	25	0.970	1483	1303	-

参与者-发布者图中的用户所属区分布



- 2-9月
- 每个用户至少参加三个任务

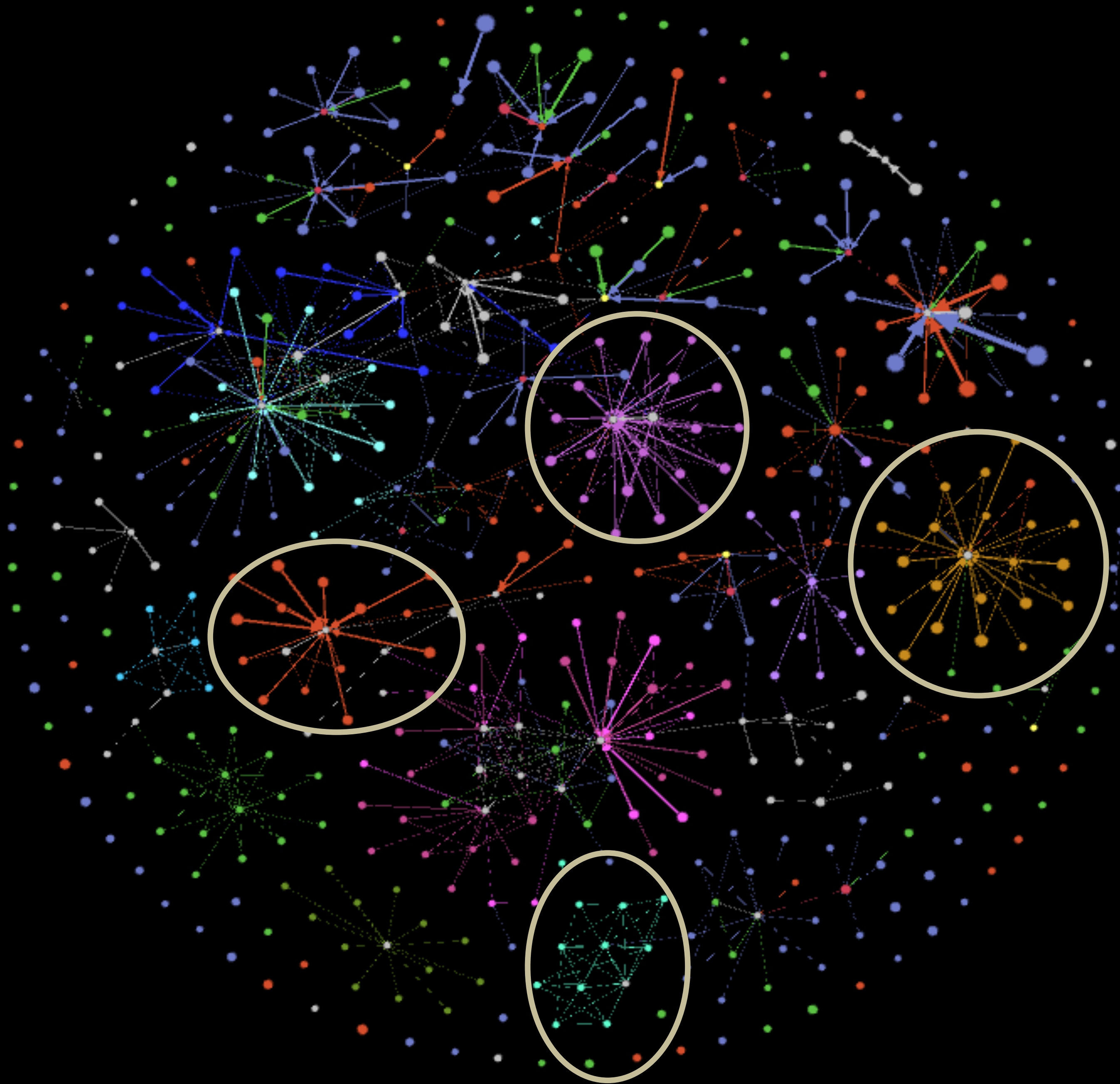
用户-任务关系的模式学习

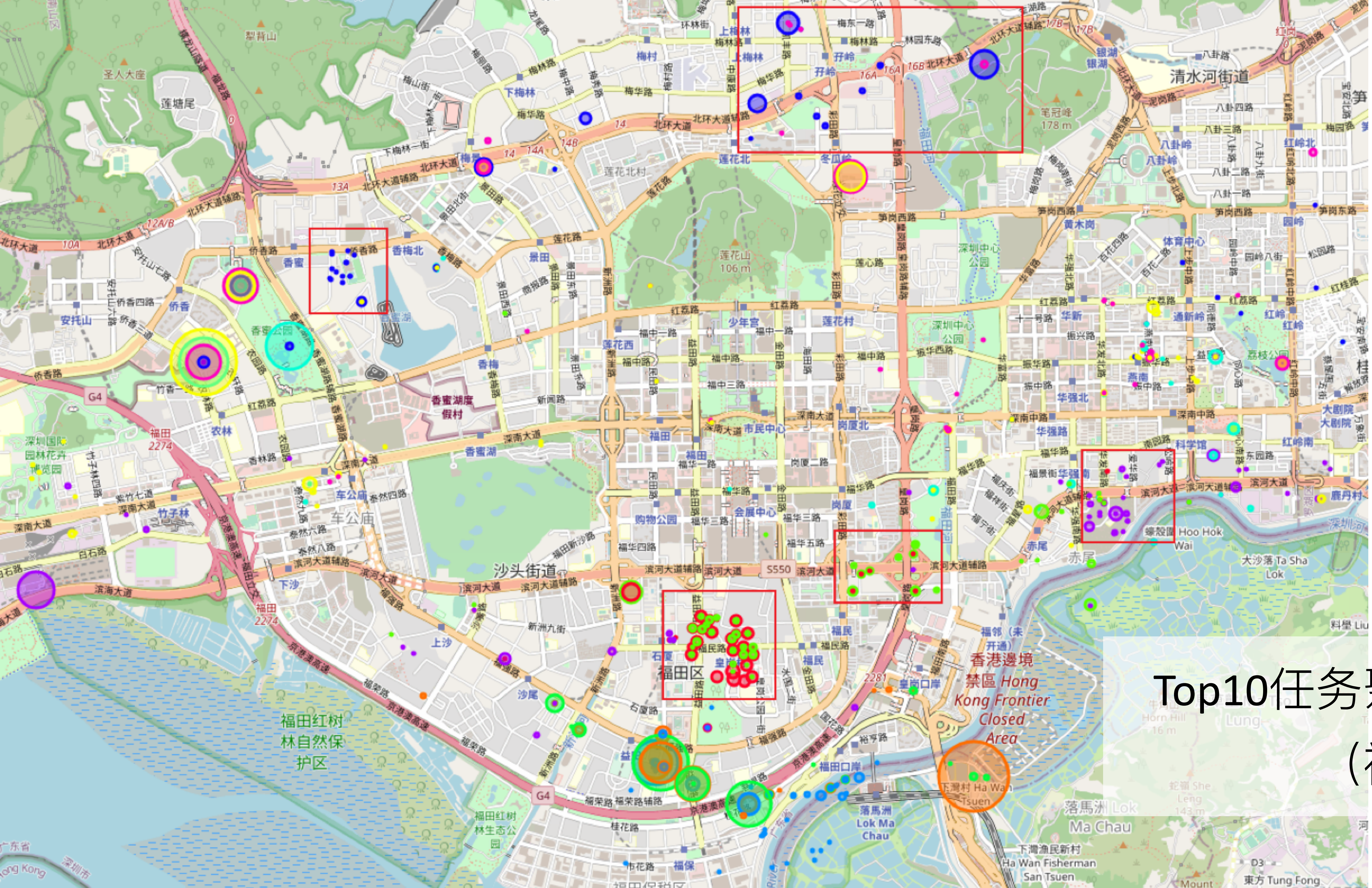
- 用户特征：p-i图中用户节点的Node2Vec embedding
- 任务特征：任务主题、地理位置信息、时间信息、发布者信息
- 目标：通过HGR最大化相关性的特征提取，并对用户与任务进行联合聚类

P-I图中的用户聚类结果

- 相同类别的用户被标记为同一颜色；
- 箭头从参与者P指向发布者I；
- 点越大，该用户P参与任务次数越多；
- 边越粗，用户P参与该发布者I创建的任务次数越多。

可视化结果表明，本方法能够较为准确地识别对相同发布者I的任务感兴趣的参与者，将其映射为相似的向量，以便聚类算法正确分类。





Top10任务聚类结果
(福田区)

下一步工作

- 自组织力的演变
- 其他目标：任务主题、任务模式、任务可预测性
- 自组织力影响因素的量化分析（假设检验、因果分析）
- 自组织成熟度的衡量：与外界环境互动后的自组织（熵减）效率