# (2) 数据可视化

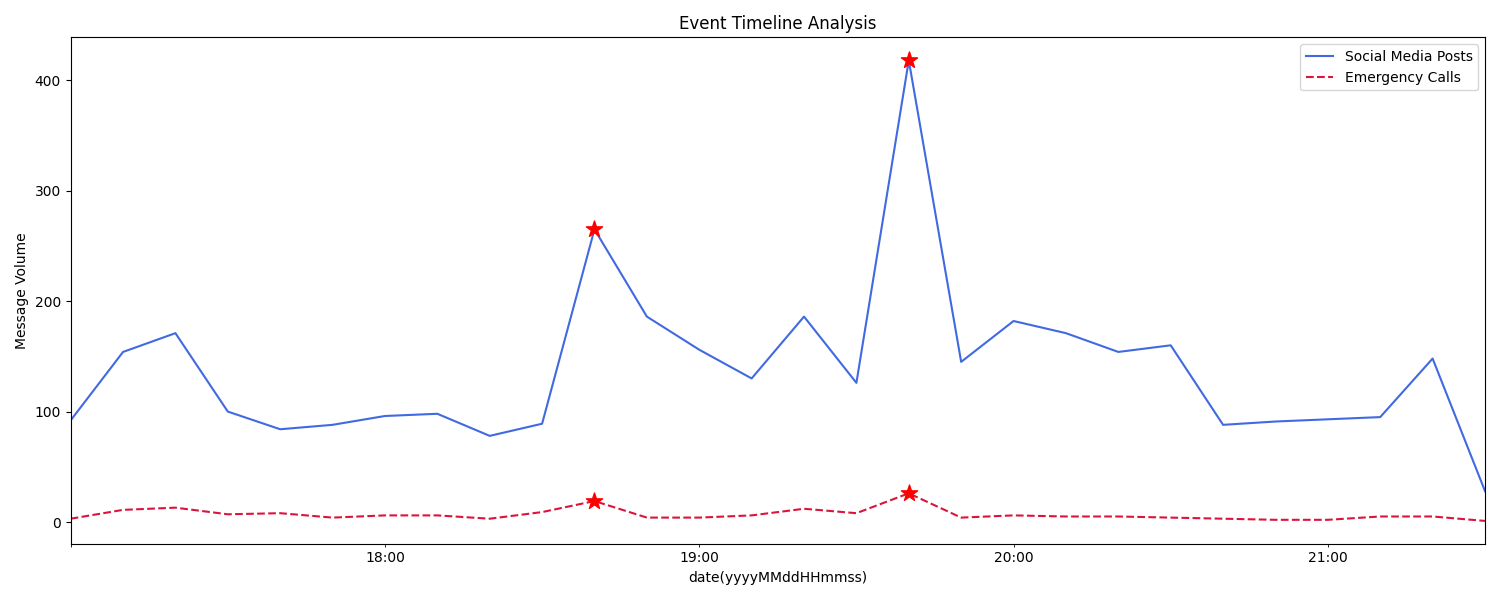
两类数据：

mbdata：社交媒体信息（microblog）

ccdata：紧急调度中心通话记录（call center）

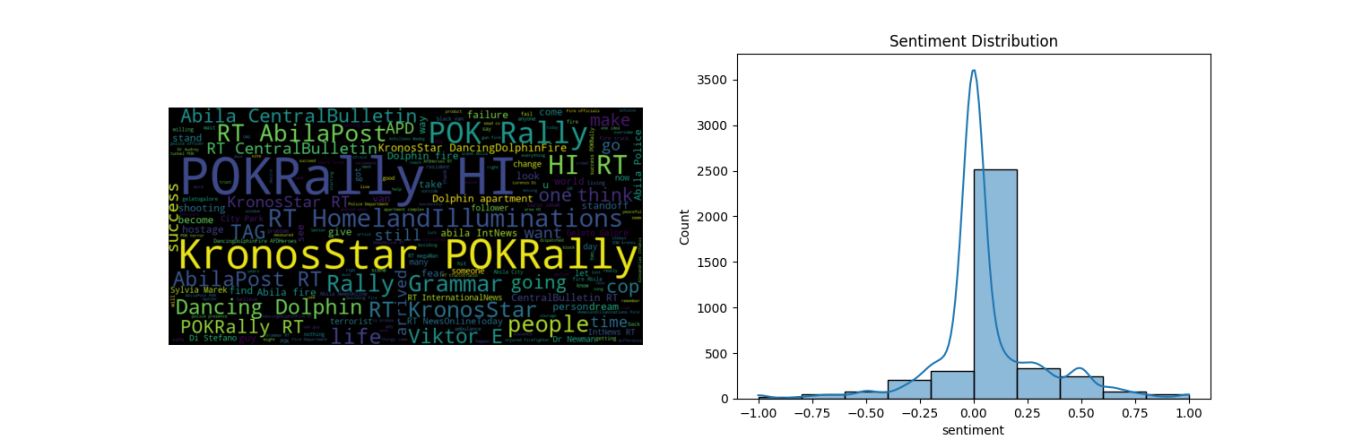
# 事件时间序列变化图

对mbdata数据和ccdata数据，按10分钟间隔统计消息量，并标记消息突然增多的关键时间点（三倍标准差作为异常阈值检测消息突发点），检测结果显示2014-01-23 18:40:00的10min之内消息突发量为284，2014-01-23 19:40:00的10min之内消息突发量为444。图中已标为红星。



# 关键词词云与情感分析

对mbdata数据制作词云图，并使用TextBlob计算每条消息的情感极性（-1到1）进行情感分析，由图可知“POKPally”、“HI”、“KronosStar”等出现次数很多；正负情绪message数量相对平衡。



# 风险等级划分

对ccdata进行风险等级划分，对高风险和中风险message进行关键词匹配，其余归类为低风险，具体关键词如下：

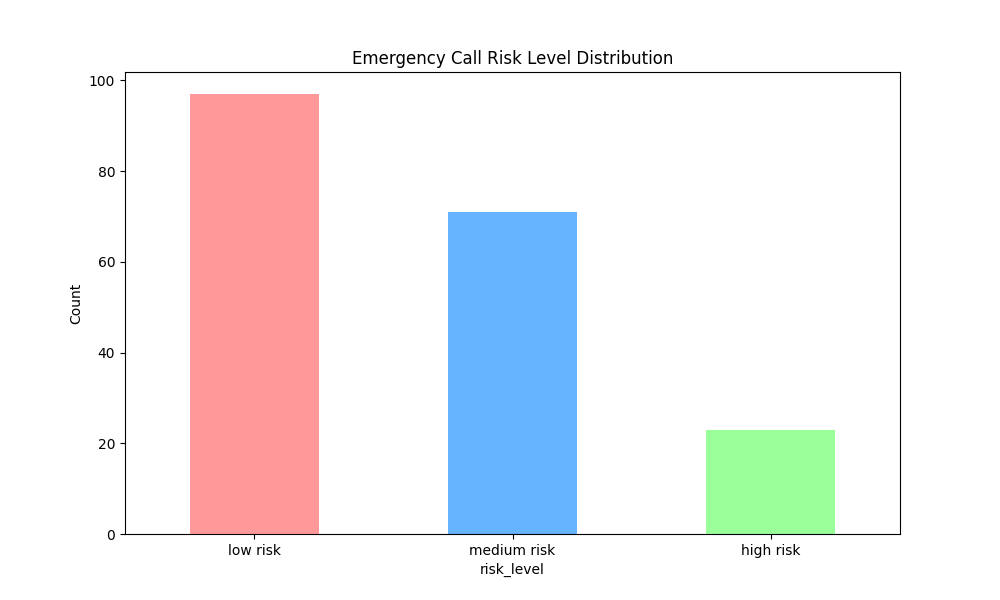
# 高风险：涉及暴力犯罪或重大灾害

high\_risk = r'\b(assault|shooting|hostage|riot|fire|bomb|explosion|armed|stabbing|active shooter|homicide|injured officer|building fire|officer down)\b'

# 中风险：财产犯罪或潜在危险

med\_risk = r'\b(disturbance|alarm|crime scene|investigation|vandalism|theft|suspicious|break-in|burglary|dwelling of interest|subject stop|shots fired)\b'

由图可得，低风险警报message最多，中风险其次，通常与财产犯罪或潜在危险有关，高风险message最少但社会危害程度最高，通常涉及暴力犯罪或重大灾害。



# 用户活跃度分析

识别高频发帖用户，并绘制top10用户的条形图。

mbdata平台的数据来源主要分为两类：一是来自具有公信力的社交账号（包括政府机构、新闻媒体及事件目击者），二是来自闲聊内容、垃圾信息及商业广告账号。这两类信息均存在转发行为，可通过账号可信度进行区分。

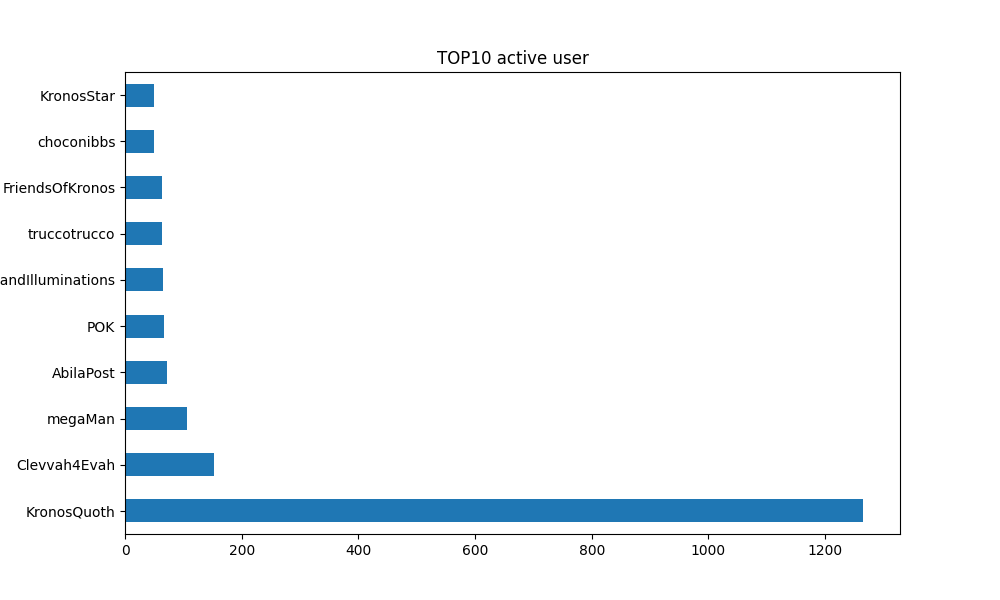
在可信账号类别中：

* 政府机构账号如"AbilaFireDept"（阿比拉消防局）和"AbilaPoliceDepartment"（阿比拉警察局）会发布权威公告，例如："阿比拉消防局宣布扩大疏散区域#AFD"。
* 新闻媒体账号如"AbilaPost"和"InternationalNews"则报道事件动态，例如："预计POK集会将吸引超过1000人"。
* 目击者账号会提供现场实时信息。例如，账号"anaregent"在集会现场发布："今晚到场人数很多。这里肯定有2000人！"；账号"truccotrucco"、"Simon\_Hamaeth"和"roger\_roger"疑似Gelato Galore事件的人质，其中"truccotrucco"称："我为什么没留在集会上？我被困在这里了，面包车就在门外"。

与之形成鲜明对比的是低可信度账号：

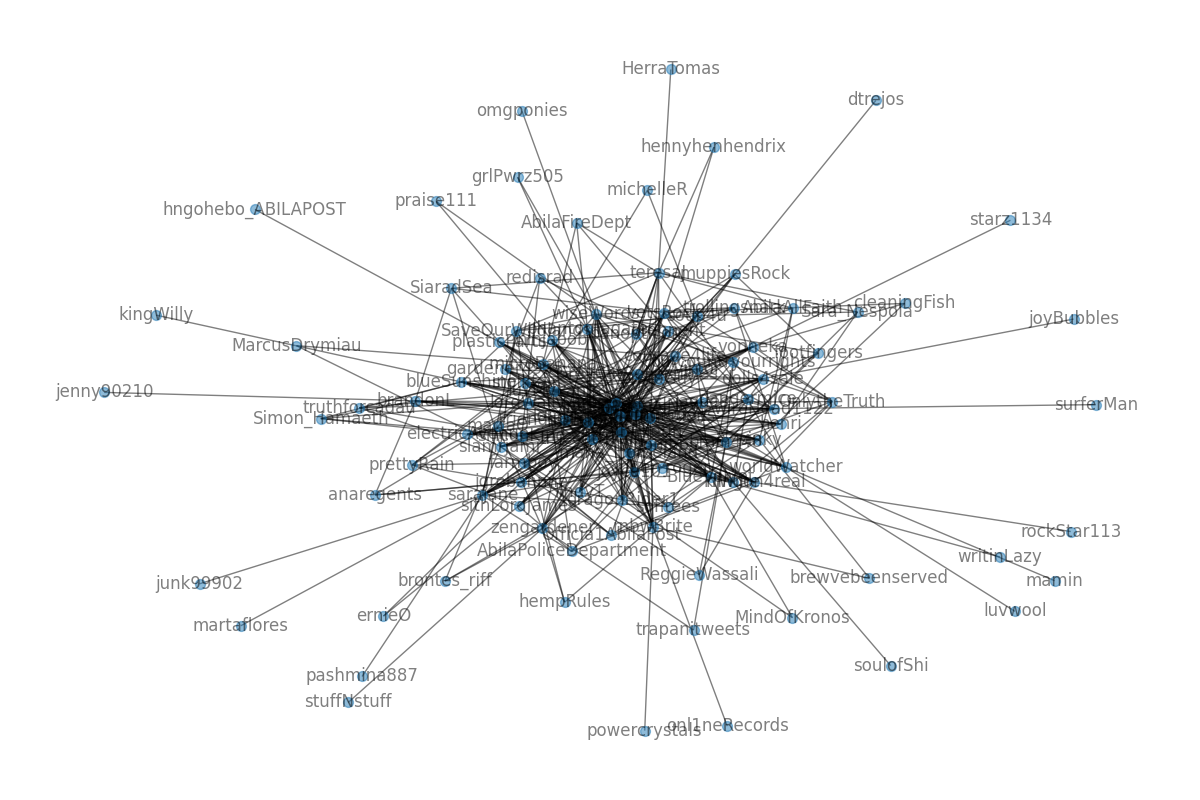
* 垃圾广告账号如"junkman377"和"junkman995"仅发布商业推广，例如："你的信用评级如何？需要提升吗？"。
* 无意义内容账号则发布随机言论，如"Clevvah4Evah"讨论语法问题，"KronosQuoth"频繁发布"如果你正在经历地狱，那就继续前进"等泛泛之谈。

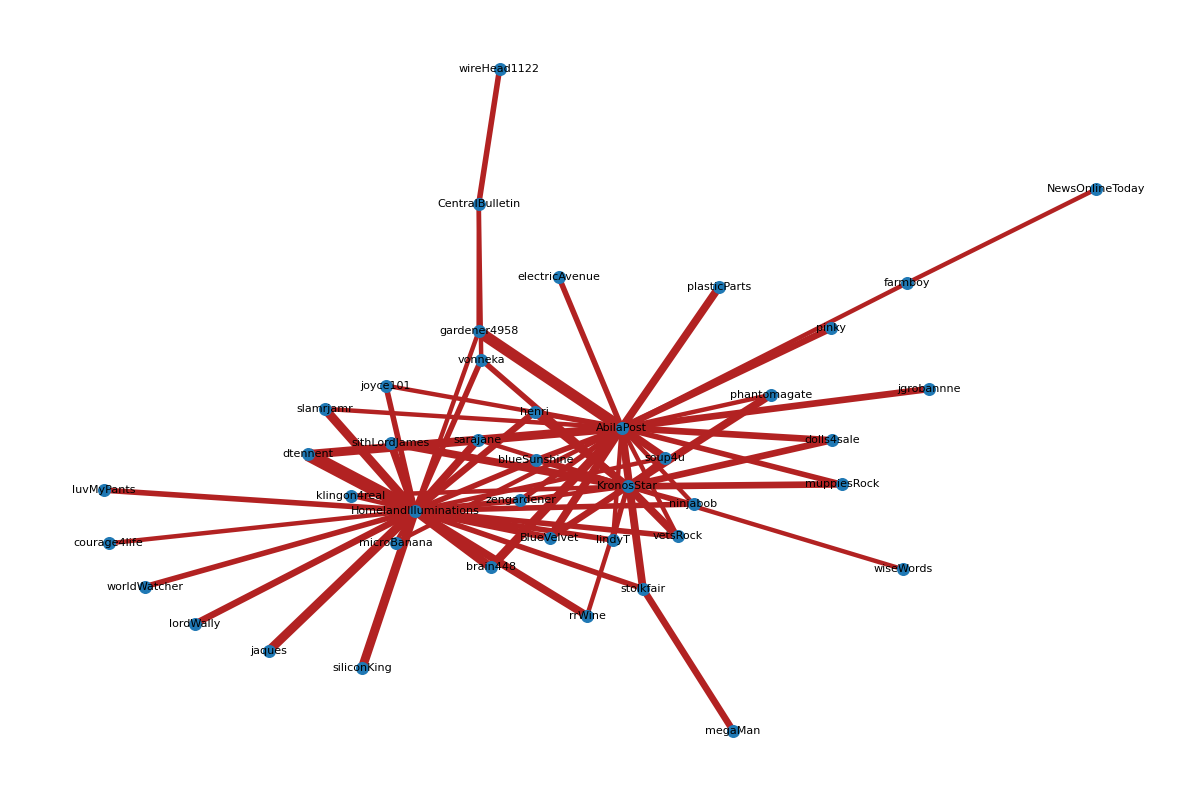
由图可以发现："KronosQuoth"的发帖频率显著高于其他账号，是最活跃的发布者。



# 构建账号转发网络

根据转发的发送者和接收者来构建转发网络，绘制原始完整网络图。通过统计转发频次，绘制绘制高频子图。比如。例如，Cevvah4Evah、POK、FriendsOfKronos 和 Victor - E倾向于经常互相转发帖子。





# 筛选有效消息与广告消息

* 筛选来源于权威账号、type为ccdata、包含关键词的message，一共627条数据。
* 通过匹配message中是否含有“http”“.kronos”“click here”“followme”来识别出广告message，一共226条数据。