

Researching China's (Online) Politics

A Data Science Toolbox



Christian Goebel

University of Vienna, Department of East Asian Studies
christian.Goebel@univie.ac.at

“open government”: access to valuable data

伽师县法制教育转化学校建设项目（土建一期）一标段

来自: 果招采 (<http://www.gzhaocai.com.cn/>) 标签: 学校 土建

所属地区	新疆	发布时间	2017/4/11	关键词	学校 土建	最近更新1505个项目
------	----	------	-----------	-----	-------	-------------

注册即可查看免费招标信息

[立即免费注册](#)

[立即登录](#)

服务热线: 400-610-9686

如果您已经是会员请先登录



一、招标条件

伽师县法制教育转化学校建设项目（土建一期）一标段已经主发改投资2017【24号批准建设。招标人为伽师县司法局,工程所需资金来源为财政。项目已具备招标条件,现对该项目的伽师县运行公开招标。本次招标计划投资额为10300.00万元,评标办法为经评审的最低投标价法（工程量清单计价方式）。本工程拟采用资格后审方法选择合适的投标报名人参加投标。

二、项目概况与招标范围

1.本次招标项目的建设地点:

2.工程规模:新建约35000平方米法治教育转化学校一所及供水、排水等基础配套设施

3.本公告共划分为1个标段

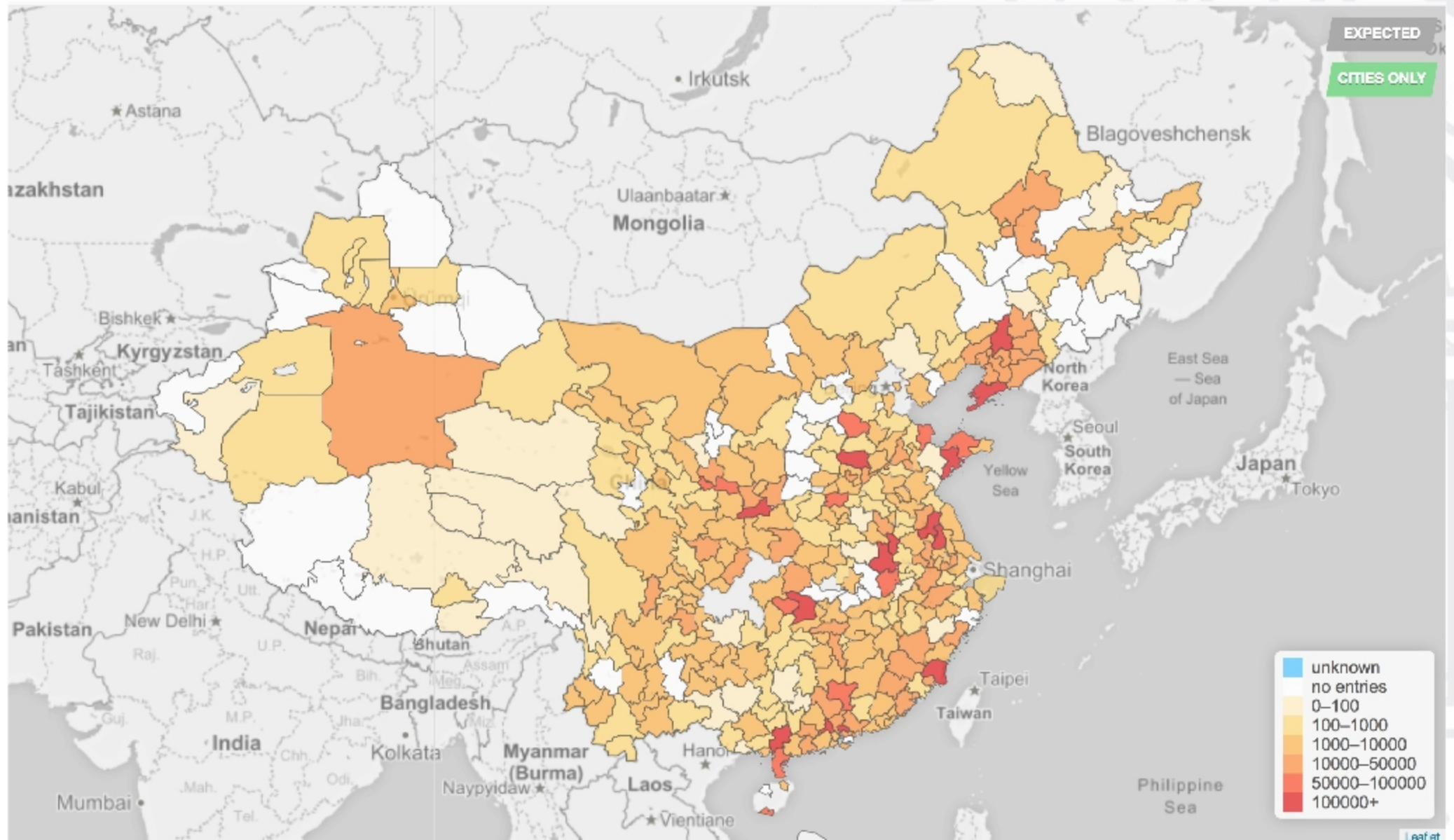
返回顶部

Public feedback used to enhance policy performance

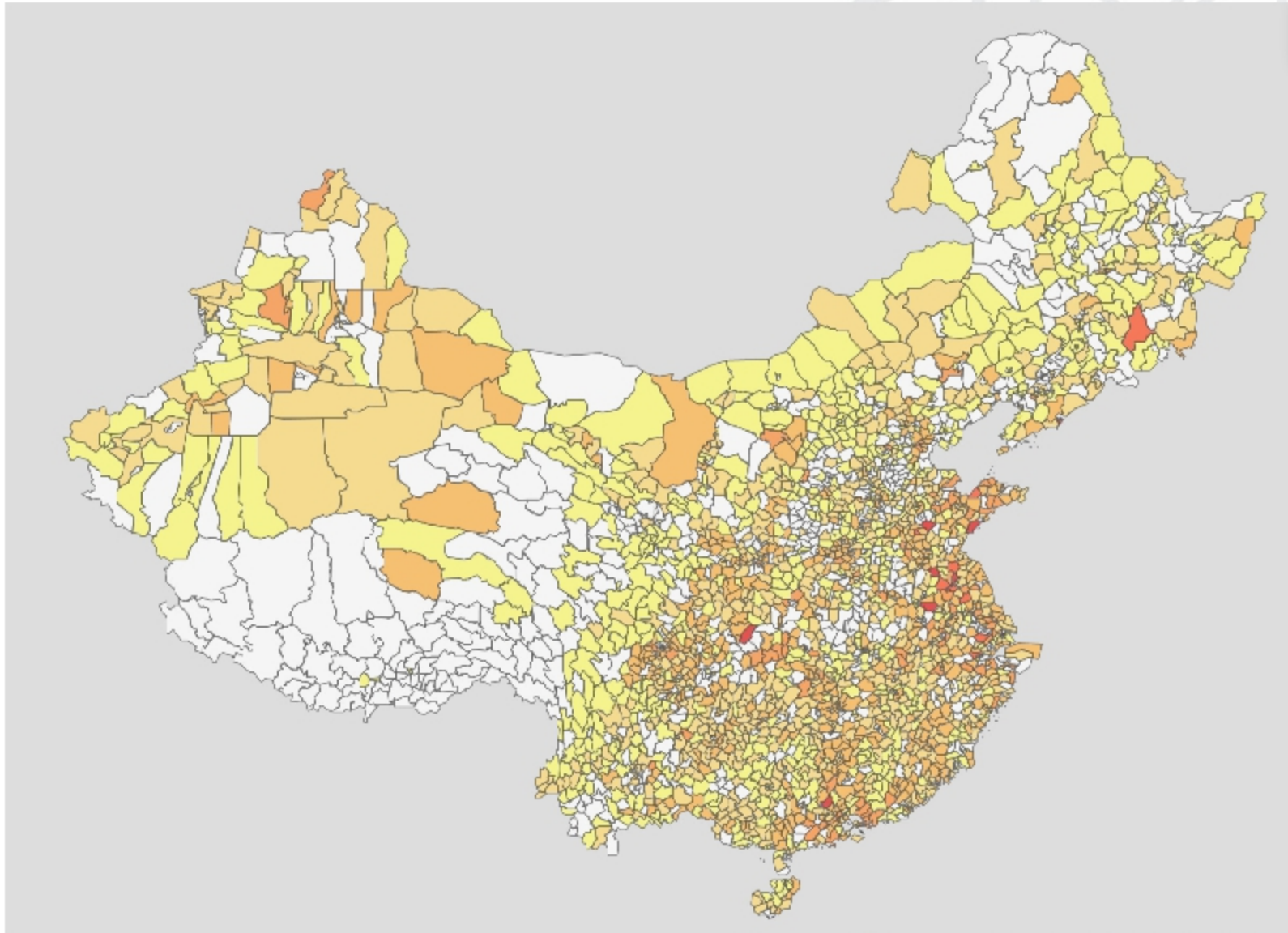
受理编号:	882717	受理时间:	2014-05-20
诉求问题:		办结时间:	2014-06-11
处理状态:	已办结	星级评价:	★★★
办理跟踪信息: 2014-5-19 13:37:25 已进入 处理系统。 2014-5-20 16:26:30 纪委协调办理。 2014-6-11 9:25:56 部门已反馈。 2014-6-11 9:42:05 回访,公示。			
受理单位:	(平均办理速度 42 天 平均办理质量 4.69 分 诉求办理总量 76 件)		
处理结果:	投诉问题一: 诉求人反映该处的征地搬迁工作于2011年5月在南地办事处启动,于2013年底结束,共搬迁居民985户。通过调查,工作人员了解到,在动迁工作中,搬迁工作人员严格按照征地拆迁的有关文件规定认真开展搬迁工作。动迁户房屋面积实施丈量认定工作由市产权处、公证处及区拆迁办人员负责共同丈量,按实际面积进行计算,保护了搬迁当事人的合法权益,确保了搬迁工作的公开、公正、公平,未发现“乱用职权,违法犯法,公开作弊,乱发放国家动迁款,没有按规定公正处理拆迁问题”的问题。 投诉问题二: 该地动迁之前,所属房产部门已经对该区域内部分房屋进行过动迁补偿。办事处搬迁办依据被拆迁户提供的有效法律证件,现场查验房屋,严格执行拆迁政策,进行拆迁补偿工作。并先后采取执法部门介入甄别、动迁户写保证书等措施,严防重复给予拆迁补偿。目前,没有发现重复补偿问题。 投诉问题三: 经过了解,未发现在动迁过场中存在此类问题。 诉问题四: 在搬迁过程中,搬迁办严格按照政府146号令执行,三证户,提供房票到市产权处核实并提供本人户口。二证户,必须提供户口及按政策要求所需要的相关证件,一个房一个户。因此,未发现此类问题。		
回访情况:	2014-6-11 9:11:13 电话回访,诉求人表示不满意。		

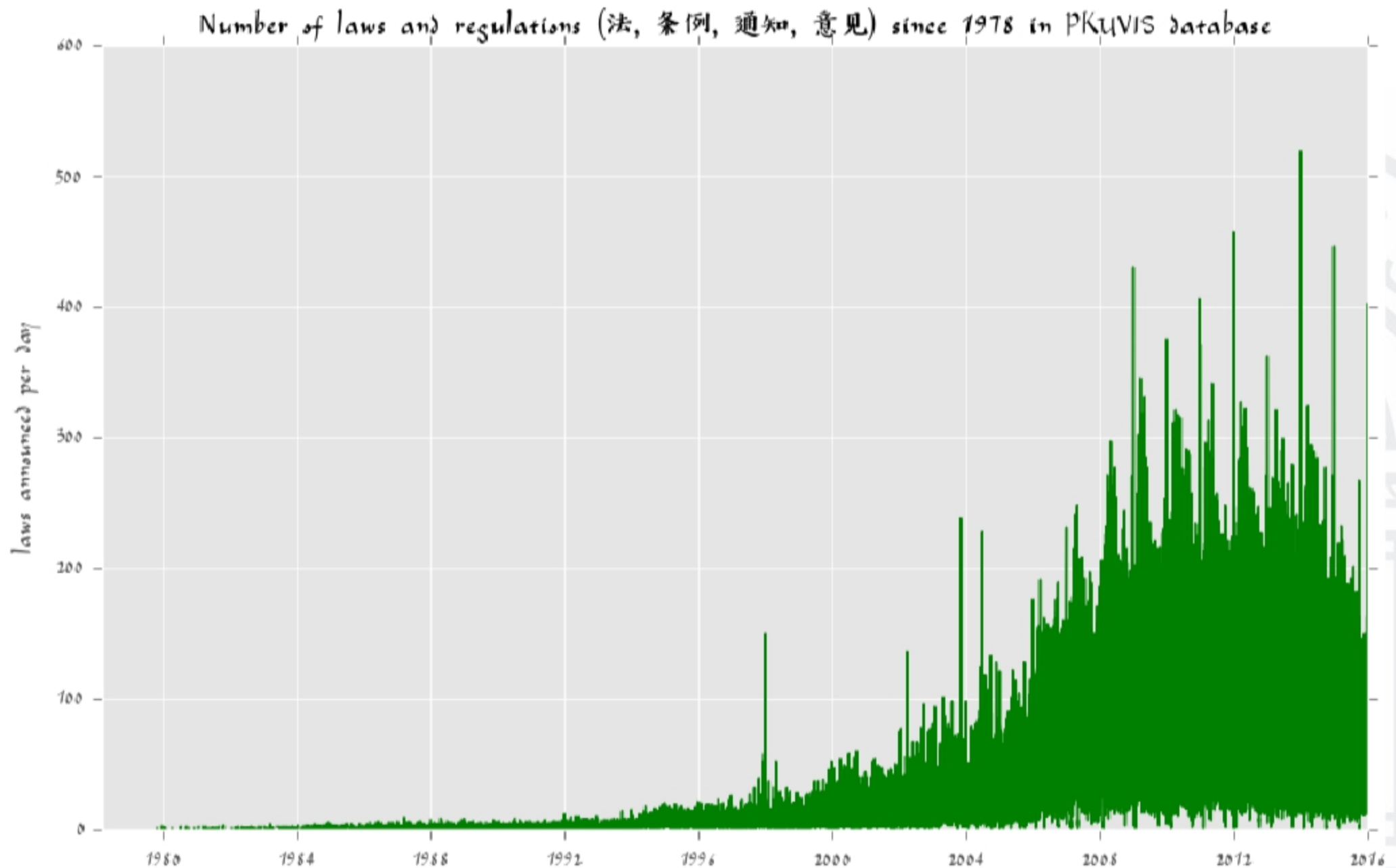
评论类别: ☐ 满意 ☐ 基本满意 ☐ 不满意 ☐ 线索 ☐ 建议

publicly accessible online petitions, prefecture level (2017)



publicly accessible online petitions, county level (2017)





The RESPONSIVENESS project

BACKGROUND:

- increasingly demanding population
 - rising social unrest
 - social media used to document poor service
- => government established online petitioning platforms to provide channel of communication and improve public services

The RESPONSIVENESS project

RESEARCH INTEREST:

- **do platforms make a difference?**
 - > reduction, increase or differentiation of social unrest?
 - > policy changes?
 - > promotion/demotion of local leaders?
- **role of signals?**
 - > how strategic are politicians, petitioners, protesters?
 - > can government “steer” demands of population?

编号	请输入诉求件编号	验证码	搜索	高级搜索
----	----------	-----	----	------

诉求类型

[投诉\(839105\)](#)
[咨询\(382181\)](#)
[举报\(170635\)](#)
[建议\(98165\)](#)
[求助\(4278\)](#)
[感谢\(3915\)](#)

所属区域

[福州市直属\(300167\)](#)
[仓山区\(252804\)](#)
[鼓楼区\(199778\)](#)
[晋安区\(195072\)](#)
[台江区\(128968\)](#)
[闽侯县\(123033\)](#)
[福清市\(82934\)](#)
[连江县\(55279\)](#)
[马尾区\(41394\)](#)
[长乐区\(36383\)](#)
[永泰县\(26606\)](#)
[闽清县\(23764\)](#)
[罗源县\(19555\)](#)
[高新区\(3529\)](#)

[投诉] 东城公寓一楼违规拆墙做公司

尊敬的领导，我属公城公寓业主，上次投诉我小区2号楼一层拆墙违建事宜，贵相关部门上门观看后回复为室内装修，本人不是认同，故再次投诉，并附上照片，，室内装修可以拆外墙，立广告牌吗，这明显是拆了改变了原来功能，改成了公司，请问这可以相关部门的牌照吗，是否违规，请领导定。还我小区安静，舒服的环境。

诉求编号: FJ17092500004 所属区域: 福州市直属 处理情况: 处理完毕 诉求来源: 网站 诉求时间: 2017-10-11

[咨询] 关于福州地下车位“只售不租”的问题

上海市今年出台《关于加强商品住房及其附属地下车库（位）等设施销售监管的通知》沪建房管联[2017]657号），要求“房地产开发企业尚有未出售的车库（位）等附属设施，应当出租给本项目内的业主、租赁人使用，不得以只售不租等理由拒绝出租”，即未出售车位不得“只售不租”。现在福州的小区，基本上按一户...

诉求编号: FJ17092000002 所属区域: 福州市直属 处理情况: 处理完毕 诉求来源: 网站 诉求时间: 2017-10-10

[咨询] 房产加名

你好，我母有一套房子在仓山万达，我自己名下也有一套房产，为了孩子上学，我想在我的母亲的房产上加上我的名字，这需要办理什么手续，要是要做赠予，现在办理加名或赠予需要哪些材料，多少费用，这样房产加名后，孩子属于仓山万达划片学校的片内生吗，这些手续要去哪些部分办理。请回复到邮箱。

诉求编号: FJ17091800003 所属区域: 仓山区 处理情况: 处理完毕 诉求来源: 网站 诉求时间: 2017-09-18

[咨询] 开店需要办哪些证件,要如何办理

我想开一婴儿用品店,现正在找店面,请问要办理哪此开店证件,需要什么材料并去哪里办理?

诉求编号: 071130191725 所属区域: 仓山区 处理情况: 处理完毕 诉求来源: 网站 诉求时间: 2007-11-30

Scrapy: extract data from websites

关于福州地下车位“只售不租”的问题			
诉求编号	FJ17092000002		
诉求标题	关于福州地下车位“只售不租”的问题		
诉求内容	上海市今年出台《关于加强商品住房及其附属地下车库（位）等设施销售监管的通知》沪房地开管字〔2016〕1101号，其中规定：“房地产开发企业尚有未出售的车库（位）等附属设施，应当出租给本项目内的业主、租户。不得拒绝出租”，即未出售车位不得“只售不租”。 现在福州的小区，基本上按一户一车位的比例规划设计，市区的车位价格都在30万以上，业主没有买车位，连车都不让下地下室，微信上甚至有福州某楼盘业主因未购买车位想进小区被拒之门外。 我省是否也能像上海市一样出台这样的惠民政策，让买不起车位的业主，也能够通过租赁解决小区地下停车位停车乱象。		
部门办理	【福州市住房保障和房产管理局】 处理完毕 截止时间：2017-10-13 09:58		
回复意见	您好！根据有关规定，小区车位只提供本小区业主购买或使用，具体的车位使用应符合小区规划要求。我局已结合房地产市场检查和物业小区综合整治，督查开发建设单位和物业企业加强小区车位的规范管理，引导业主合理使用。 福州市住房保障和房产管理局		
诉求人	陈先生	诉求时间	2017-10-10 20:00
事件地点	全省		

The image shows a web browser window with a CSS editor on the left and a live preview on the right. The CSS editor displays a complex layout with a navigation bar, a main content area, and a table. The live preview shows the rendered HTML with the corresponding CSS rules applied. The table in the preview has a blue header and a white body with a red border. The CSS rules are for the table, including background color, border, and text color.

HTML Structure (Left Panel):

```

<!DOCTYPE html>
<html>
<head>
</head>
<body>
<div id="layout" class="layout">
  <div class="wrapper">
    <div id="frame_header_wrapper" class="header_wrapper"></div>
    <div class="nav_wrapper"></div>
    <div id="frame_container" class="container">
      <!--主体-->
      <div class="main">
        <div id="frame_position" class="row clearfix"></div>
        <div class="row clearfix row_box_bg litigation_container">
          ::before
          <div class="litigation_info_box">
            <h1 class="title-info"></h1>
            <div class="table">
              <table id="appeal-nature-0" cellpadding="0" cellspacing="0" border="0">
                <tbody>
                  <tr></tr>
                  <tr></tr>
                  <tr>
                    <td class="ceil-title">诉求内容</td>
                    <td colspan="10">
                      <p>上海市今年出台《关于加强商租住房及其附属地下车库（位）等设施销售监管的通知》沪温房管联〔2017〕657号），要求“房地”开发企业尚有未出售的车库（位）等配套设施，应当出租给本项目内的业主、租赁人使用，不得以只售不租等理由拒绝出租”，即未出售车位不得“只售不租”。
                    </p>
                    </td>
                  </tr>
                </tbody>
              </table>
            </div>
          </div>
        </div>
      </div>
    </div>
  </div>
</body>
</html>

```

CSS Rules (Right Panel):

```

.layout.scss:142
table table tr
td p {
  margin: 7px 0;
}

.layout.scss:5
html, body, div,
dl, dt, dd, ul,
ol, li, h1, h3,
h3, h4, h5, h6,
pre, code, form,
fieldset, legend
input, textarea,
p, blockquote,
th, td, tr, em,
button, img {
  margin: 0;
  padding: 0;
}

Inherited from td
.layout.scss:135
table table tr
td {
  color: #222;
  line-height: 24px;
}

Inherited from body
.layout.scss:43
body {
  font-family:
    "Microsoft
    Yahei",
    serif

```

Scrapy: extract data from websites (and store them in postgres database)

```
# Author: Daniel Lang
# Version: 1.0
import re

from responsiveness.spiders.Rexian12345Spider import Rexian12345Spider

class FuzhouSpider(Rexian12345Spider):
    name = 'fuzhou'
    search_url = 'http://fz12345.fuzhou.gov.cn/callcenter/callSearchDo.do'
    posts_per_page = 40

    def get_log(self, response=None):
        logs = response.xpath("//td[contains(text(), '%s')]/following-sibling::td//tr" % ('转办意见: '))
        log_list = []
        for entry in logs:
            log_row = entry.xpath("td/descendant-or-self::*/text()").extract()
            message = "".join(log_row[1:]).strip()
            while message:
                log = {}
                log["source"] = log_row[0].strip()
                search_expression = r'^\xa0+?[\xa0]([\xa0+).+?. \xa0\((\d{4}-\d{1,2}-\d{1,2} \d{2}):'
                match = re.search(search_expression, message)
                log["destination"] = match.group(1)
                log["time"] = self.to_datetime(match.group(2))
                content = match.group(0)
                message = message.replace(content, "")
                log["content"] = content.strip()
                log_list.append(log)
            return log_list
```

Jieba: segment and POS-tag Chinese text

```
>>>
>>> sentence = "上海市今年出台《关于加强商品住房及其附属地下车库（位）等设施销售监管的通知》沪建房管联[2017]657号），要求“
房地产开发企业尚有未出售的车库（位）等附属设施，应当出租给本项目内的业主、租赁人使用，不得以只售不租等理由拒绝出租”，即未
出售车位不得“只售不租”。"
>>>
>>> print(list(jieba.cut(sentence)))
...
... )
Building prefix dict from the default dictionary ...
[Loading model from cache /var/folders/m2/w1fdw9b5061292b8m1hb27q40000gn/T/jieba.cache
[Loading model cost 0.731 seconds.
Prefix dict has been built succesfully.
['上海市', '今年', '出台', '《', '关于', '加强', '商品', '住房', '及其', '附属', '地下', '车库', '（', '位', '）', '等', '设
施', '销售', '监管', '的', '通知', '》', '沪建', '房管', '联', '[', '2017', ']', '657', '号', '）', '，', '，', '要求', '“
', '房地产', '开发', '企业', '尚有', '未', '出售', '的', '车库', '（', '位', '）', '等', '附属', '设施', '，', '，', '应当', '出
租', '给', '本', '项目', '内', '的', '业主', '、', '，', '租赁', '人', '使用', '，', '，', '不得', '以', '只售', '不租', '，', '，', '等', '理
由', '拒绝', '出租', '”，', '，', '，', '即', '未', '出售', '车位', '不得', '“', '只售', '不租', '”', '，', '。']
>>>
>>> print(list(pseg.cut(sentence)))
[pair('上海市', 'ns'), pair('今年', 't'), pair('出台', 'v'), pair('《', 'x'), pair('关于', 'p'), pair('加强', 'v'), pair('
商品', 'n'), pair('住房', 'n'), pair('及其', 'c'), pair('附属', 'b'), pair('地下', 's'), pair('车库', 'n'), pair('（', 'x')
, pair('位', 'q'), pair('）', 'x'), pair('等', 'u'), pair('设施', 'n'), pair('销售', 'vn'), pair('监管', 'vn'), pair('的',
'uj'), pair('通知', 'v'), pair('》', 'x'), pair('沪建', 'j'), pair('房管', 'j'), pair('联', 'v'), pair('[', 'x'), pair('201
7', 'm'), pair(']', 'x'), pair('657', 'm'), pair('号', 'm'), pair('）', 'x'), pair('，', 'x'), pair('，', 'x'), pair('要求',
'v'), pair('“', 'x'), pair('房地产', 'j'), pair('开发', 'v'), pair('企业', 'n'), pair('尚有', 'v'), pair('未', 'd'), pair(
'出售', 'vn'), pair('的', 'uj'), pair('车库', 'n'), pair('（', 'x'), pair('位', 'q'), pair('）', 'x'), pair('等', 'u'), pai
r('附属', 'b'), pair('设施', 'n'), pair('，', 'x'), pair('应当', 'v'), pair('出租', 'v'), pair('给', 'p'), pair('本', 'r'),
pair('项目', 'n'), pair('内', 'f'), pair('的', 'uj'), pair('业主', 'n'), pair('、', 'x'), pair('租赁', 'v'), pair('人', 'n
'), pair('使用', 'v'), pair('，', 'x'), pair('不得', 'v'), pair('以', 'p'), pair('只售', 'v'), pair('不', 'd'), pair('租',
'v'), pair('，', 'x'), pair('等', 'u'), pair('理由', 'n'), pair('拒绝', 'v'), pair('出租', 'v'), pair('”，', 'x'), pair('，',
'x'), pair('即未', 'd'), pair('出售', 'vn'), pair('车位', 'n'), pair('不得', 'v'), pair('“', 'x'), pair('只', 'd'), pair('
售不租', 'l'), pair('”', 'x'), pair('。', 'x')]
>>> [
```

Latent Dirichlet Allocation (LDA)

- **Mallet (Java)**

- + Hyperoptimisation (“boost” strong topics)
- sub-optimal parallelisation, no distributed computing

- **Structured Topic Models (R)**

- + use metadata to tune document-topic scores
- no parallelisation

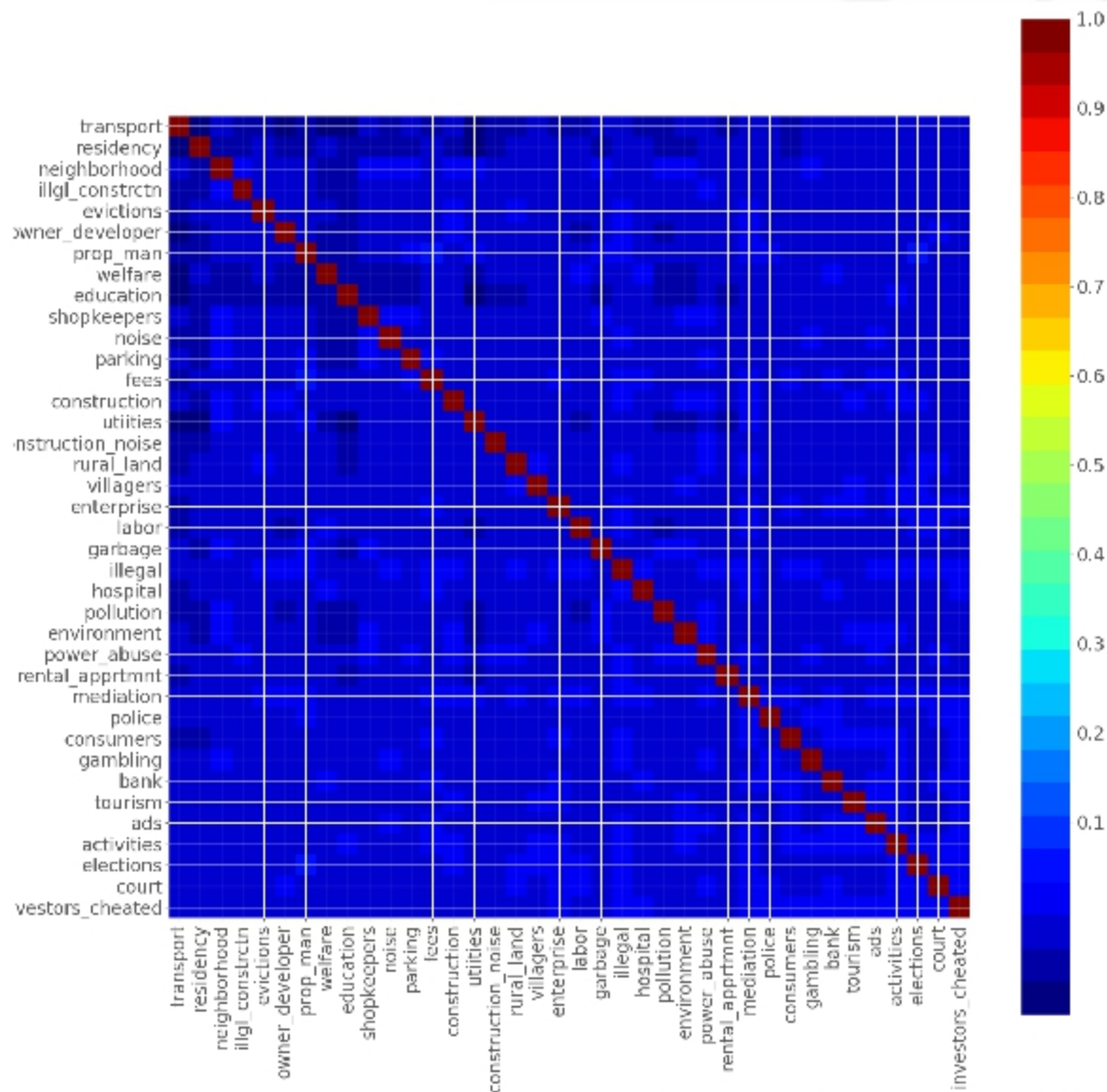
- **PLDA+ (C)**

- + supports distributed computing
- no hyperoptimisation

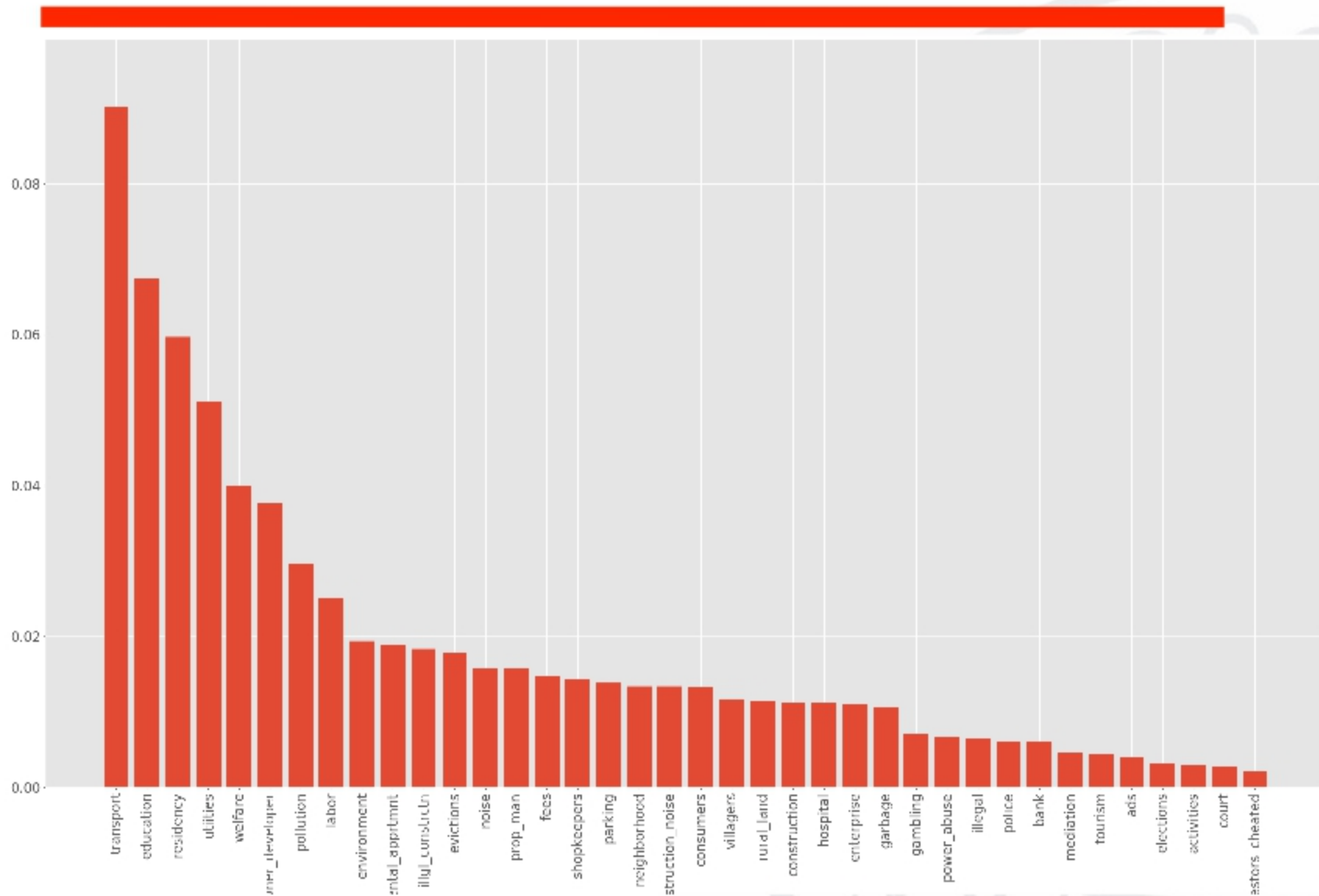
- **gensim (C)**

- + supports parallelisation (but not distributed computing)
- + very good API
- no hyperoptimisation
- less precise than Mallet

Document-topic scores correlation matrix (matplotlib)



K-means, SVM (scikit-learn): Classification



venting steam

“The party committee members above the village- and below the township level do not do their work, they are corrupt and lawless, prey on the people and exchange power for money. The position of chairman of the birth control office costs 200,000 yuan, a temporary employment with the government 300,000 to 600,000 yuan, the office of the chairman of the village committee more than 300,000 yuan. Excuse me, but are the township party committee members also bought, and if so, by whom? Last autumn, Mr. Ou crippled someone, this year on August 3 he was involved in a shootout, and there is more: in the hotel there is prostitution, in the KTV they sell cocaine and heroin, everywhere there is illegal gambling, usury and organised crime!”

Poor bear brunt of Beijing coal cleanup with no heating at -6C

Switch from coal to gas has left residents of towns around Beijing without heating after gas supply falters, reports **Climate Home News**

*Climate Home
News staff*

Mon 4 Dec 2017
16.48 GMT



143



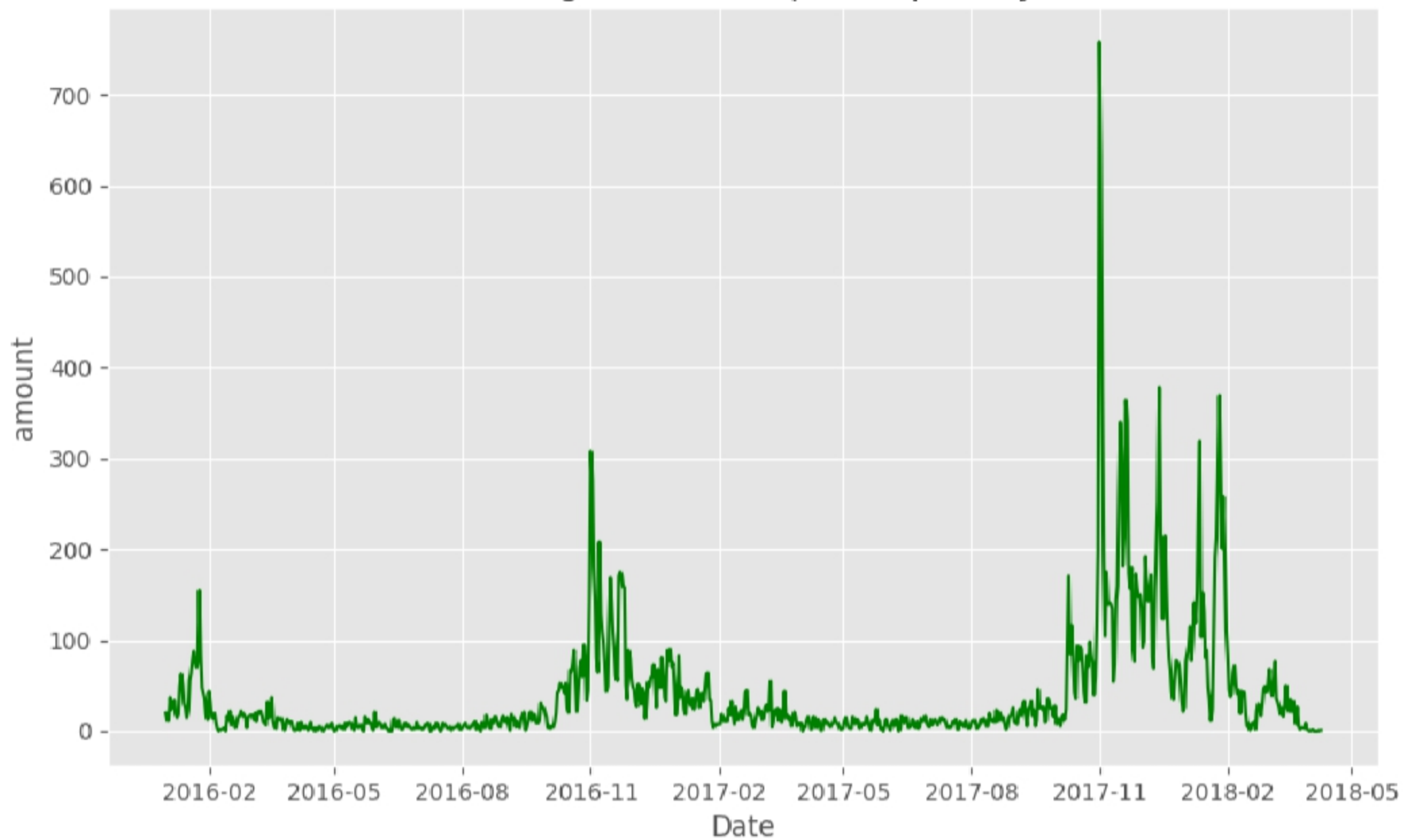
104

🕒 This article is over 6 months old

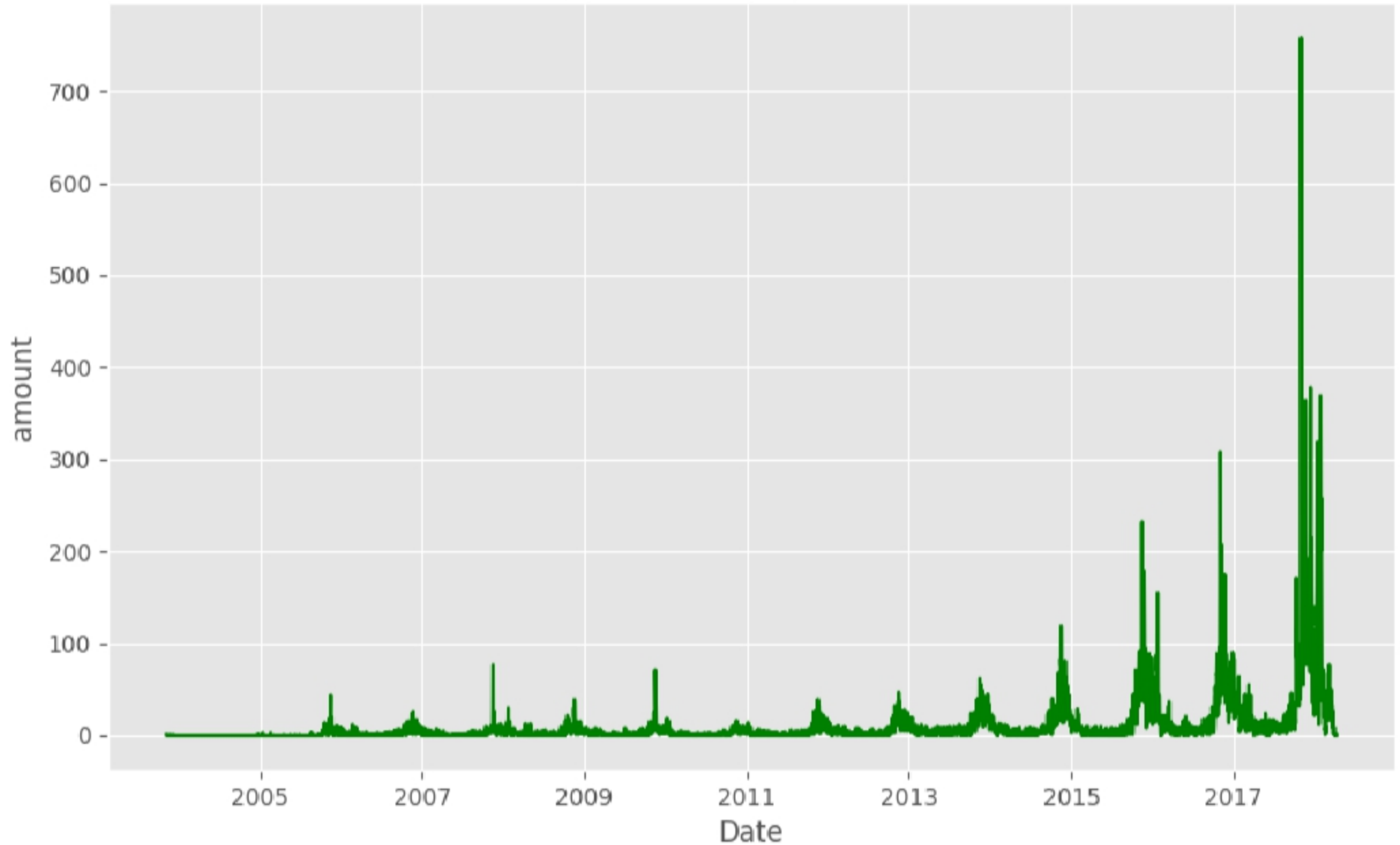


▲ A gas pipeline construction on the outskirts of Beijing. There is a demand-supply gap for gas after coal stoves have been demolished. Photograph: Jason Lee/Reuters

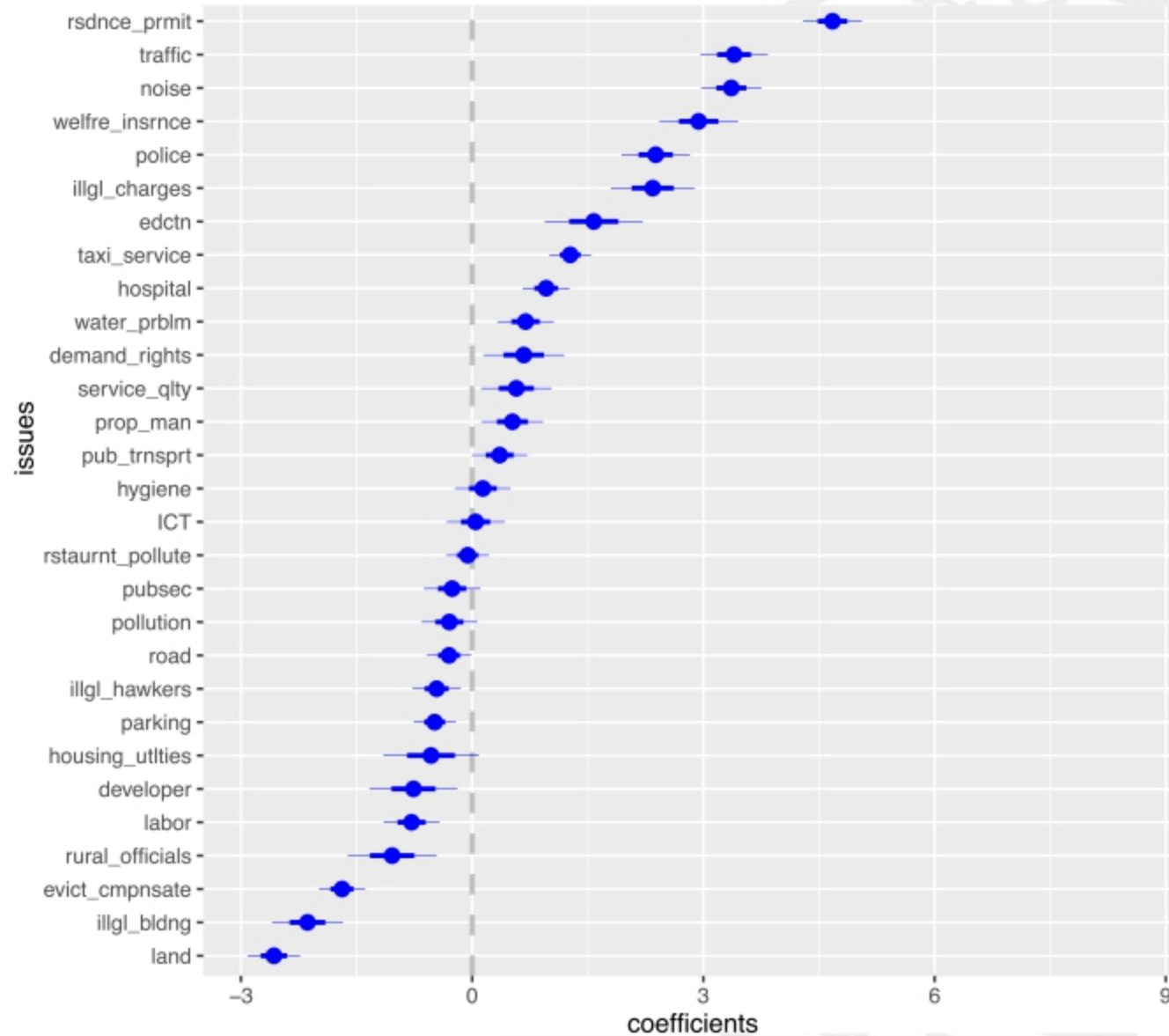
Heating-related complaints per day

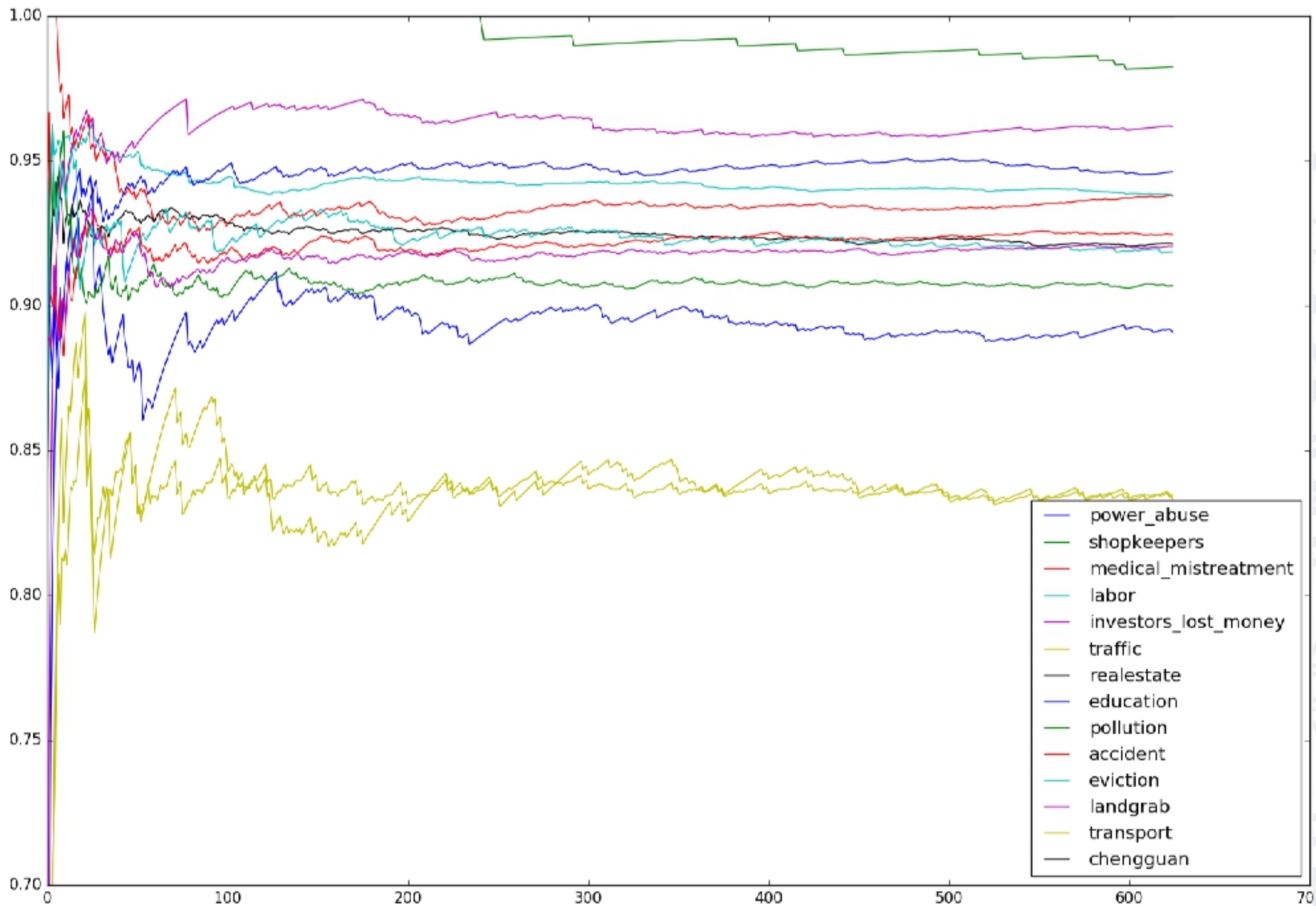


Heating-related complaints per day

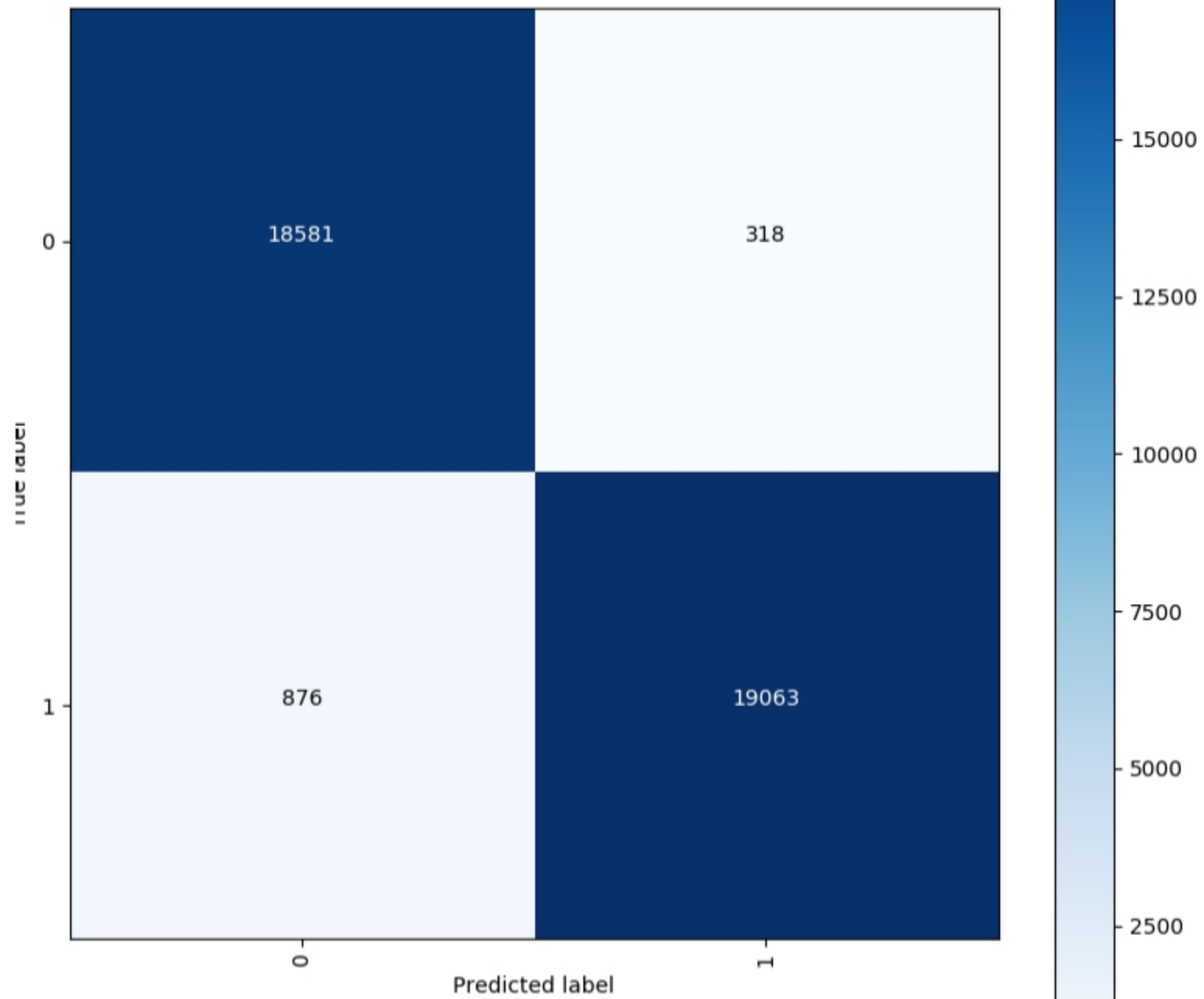


Likelihood of demands being satisfied





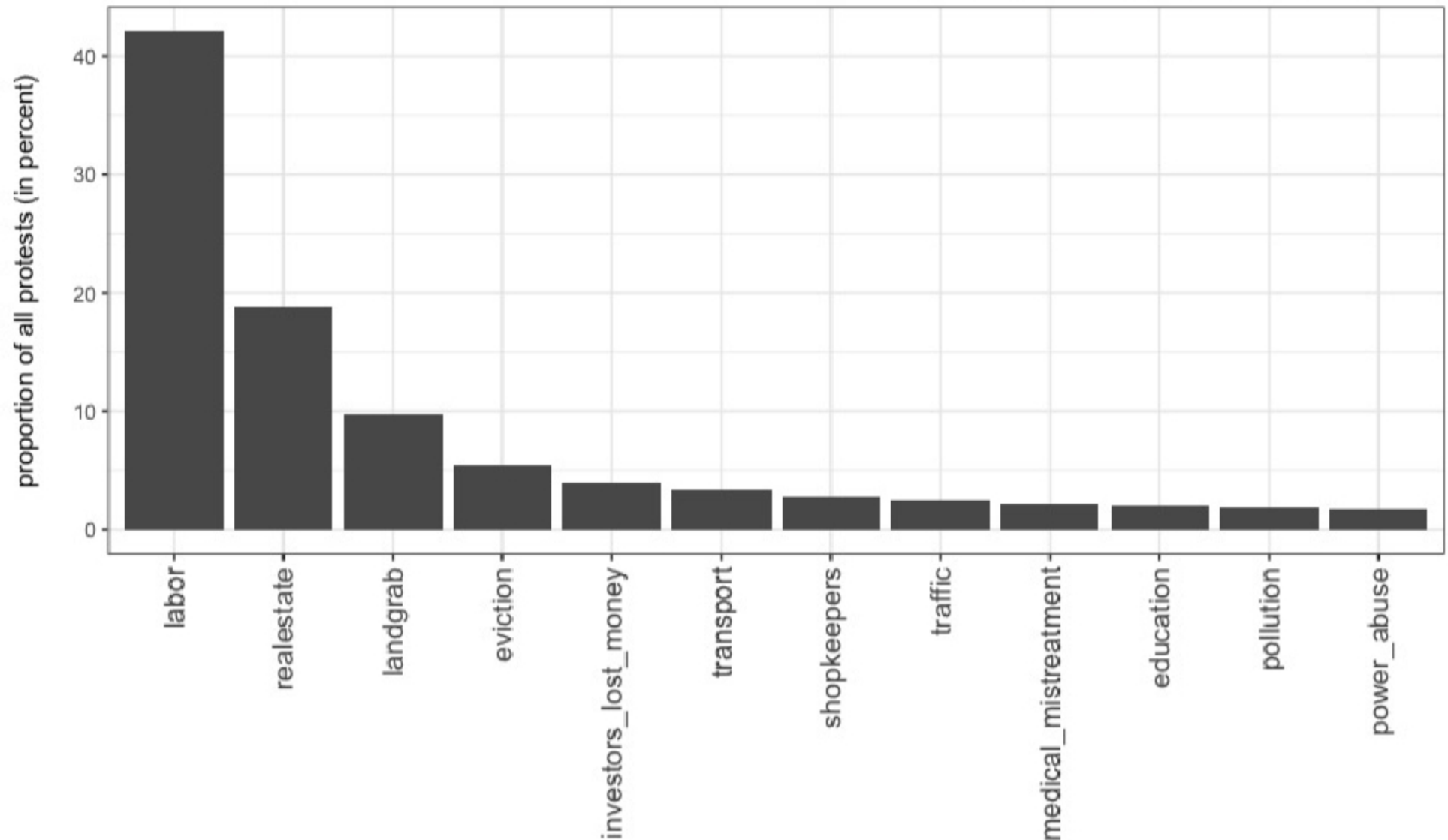
Confusion matrix, without normalization

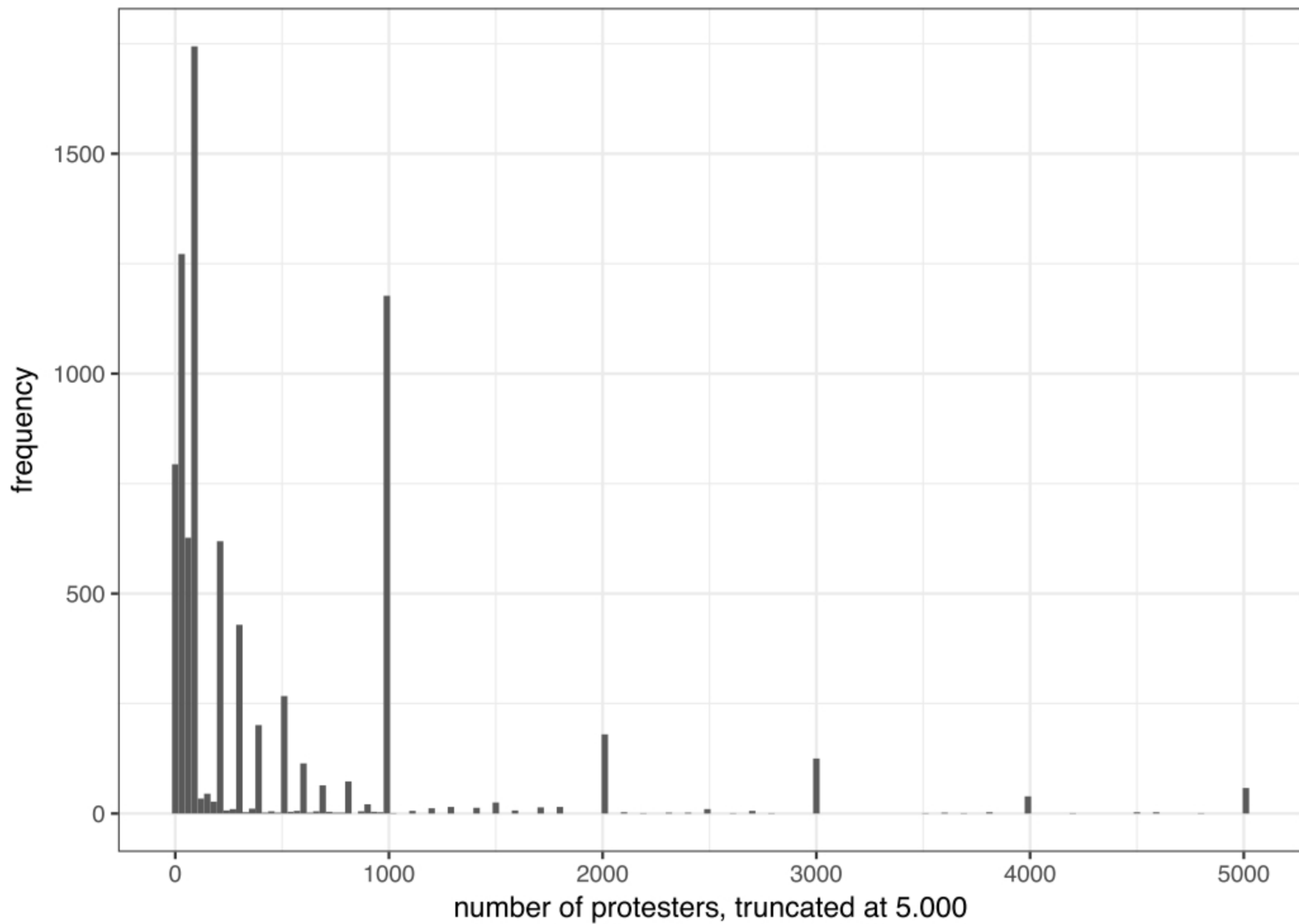


challenge when mining protest data: many misclassifications

- **“Guangzhou welcomed me with a traffic jam”**
- **“a wage of 5000 Yuan in Guangzhou, how can you live like this?”**
- **“Yesterday 6 o'clock in the evening, Mr Zhang left a cell phone in a shop, and then went out on the 30 bus. After he got on, he realized he forgot his mobile phone, but when he wanted to leave the bus, he was stopped by the bus driver, who mistook him for a pickpocket and gave him a beating! What's even more unexpected is that the phone that he left behind was stolen by onlookers”**

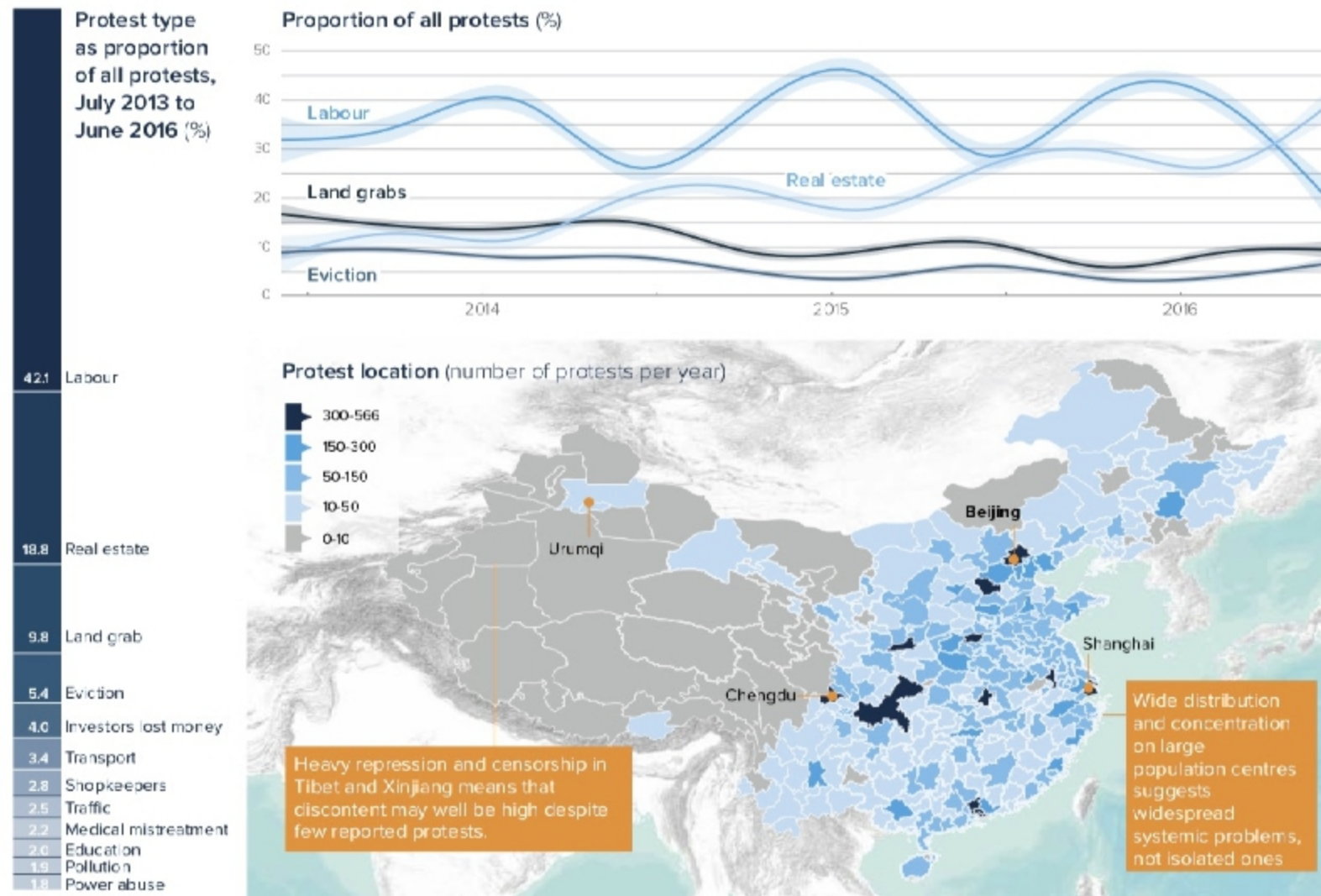
Reason for protest





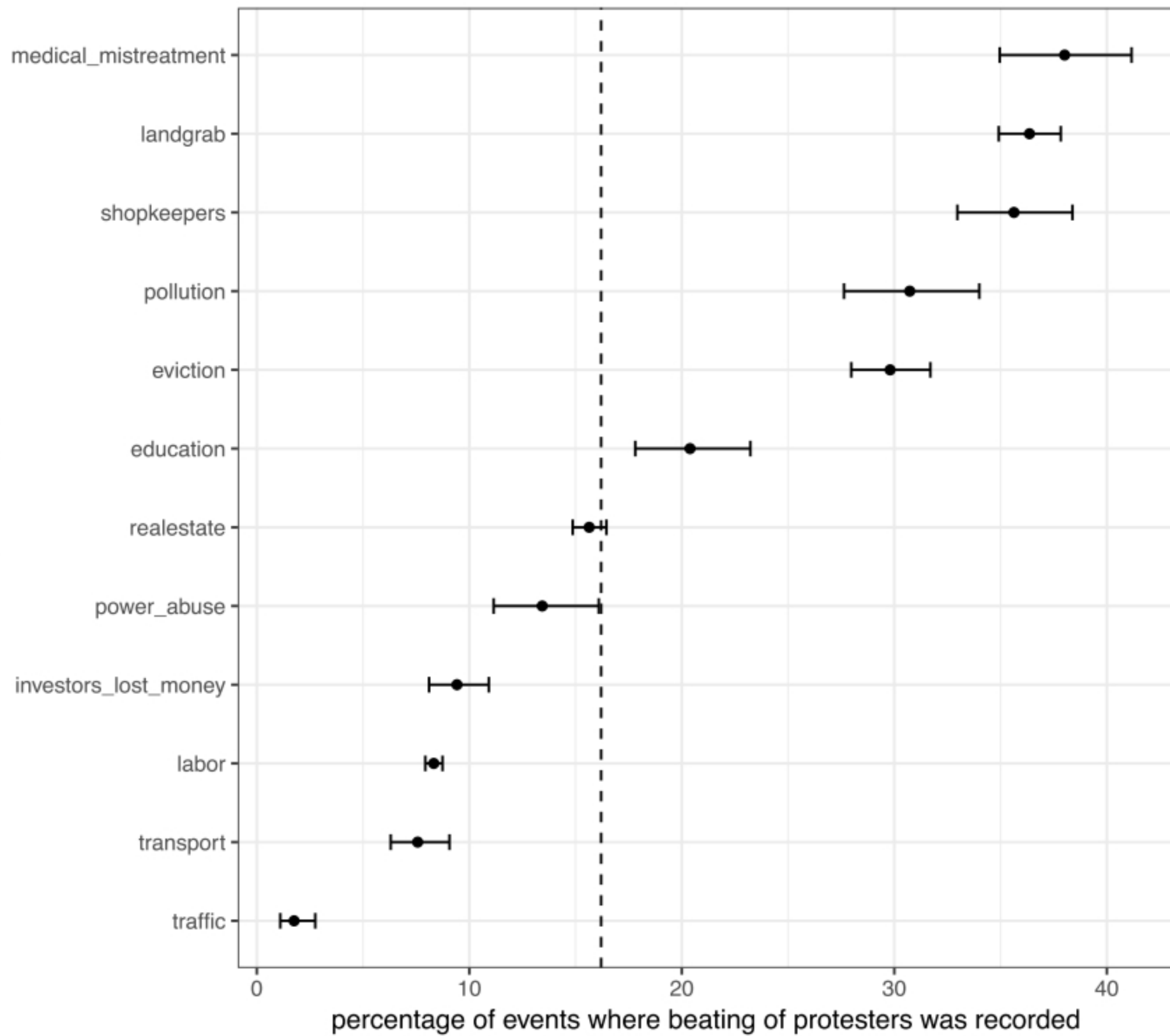
Seasonality of protests

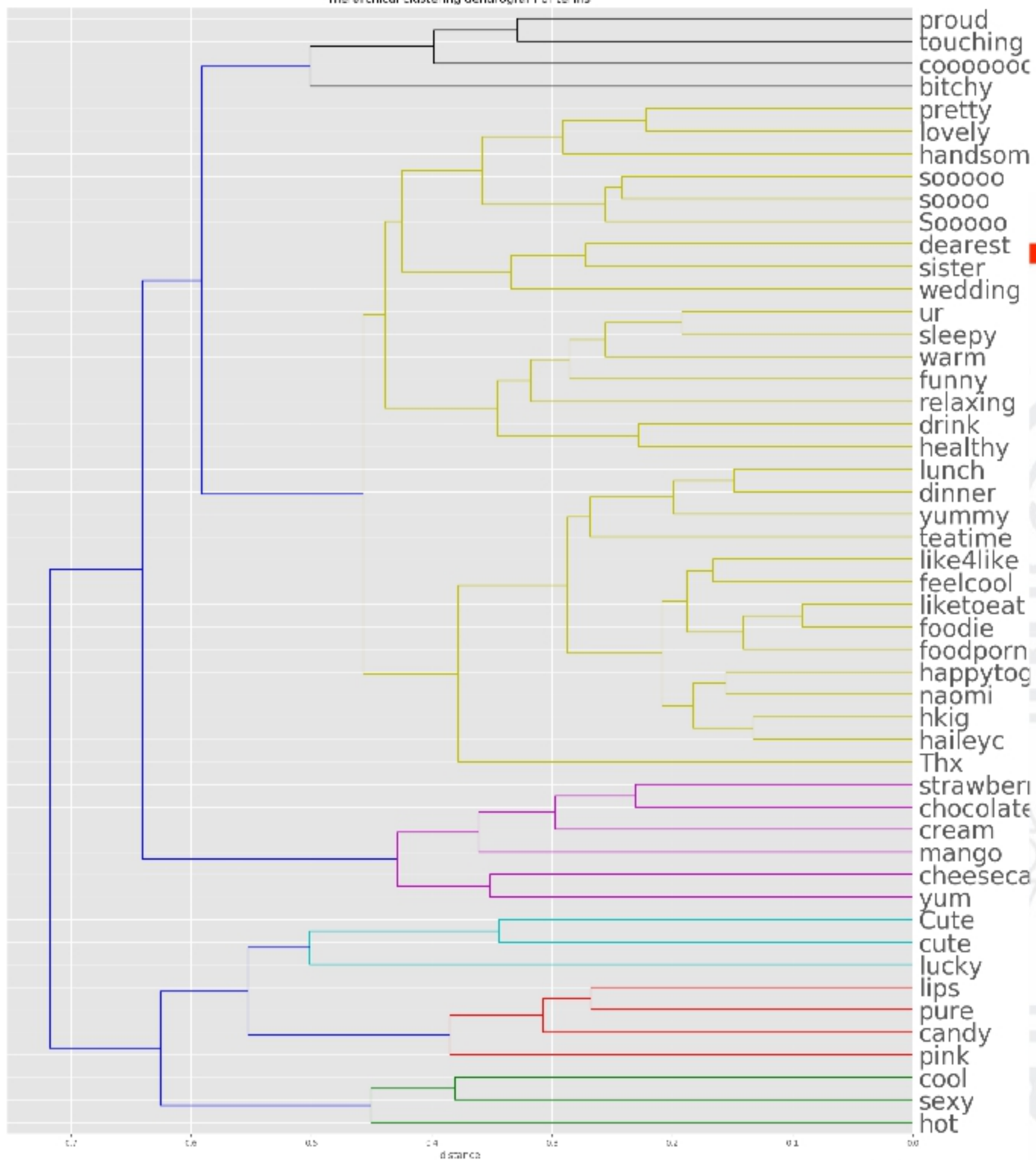
Most protests are about specific financial grievances, and they peak around Chinese New Year holiday



Source: Wickedonna protest dataset, University of Vienna

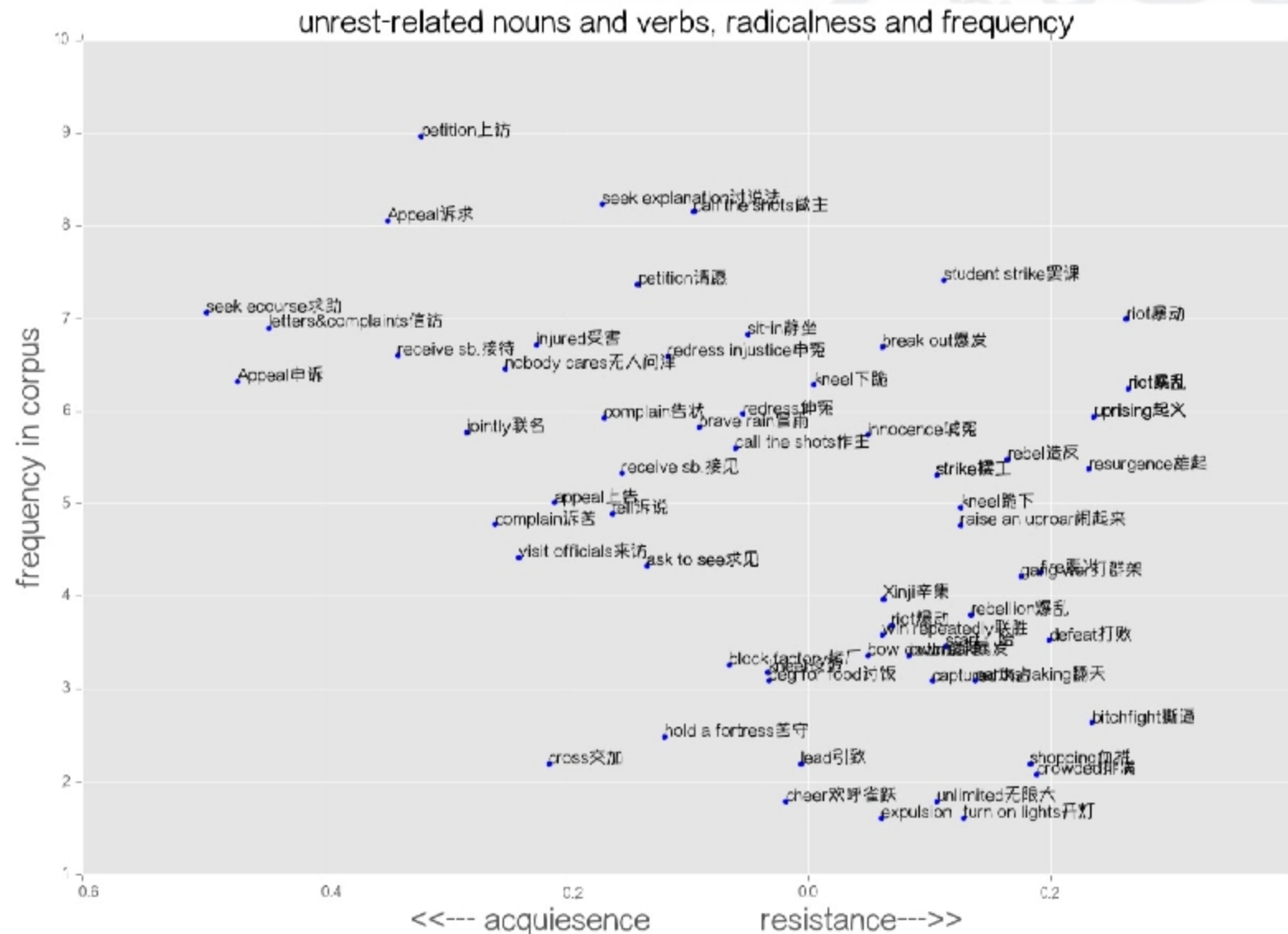
type of protest



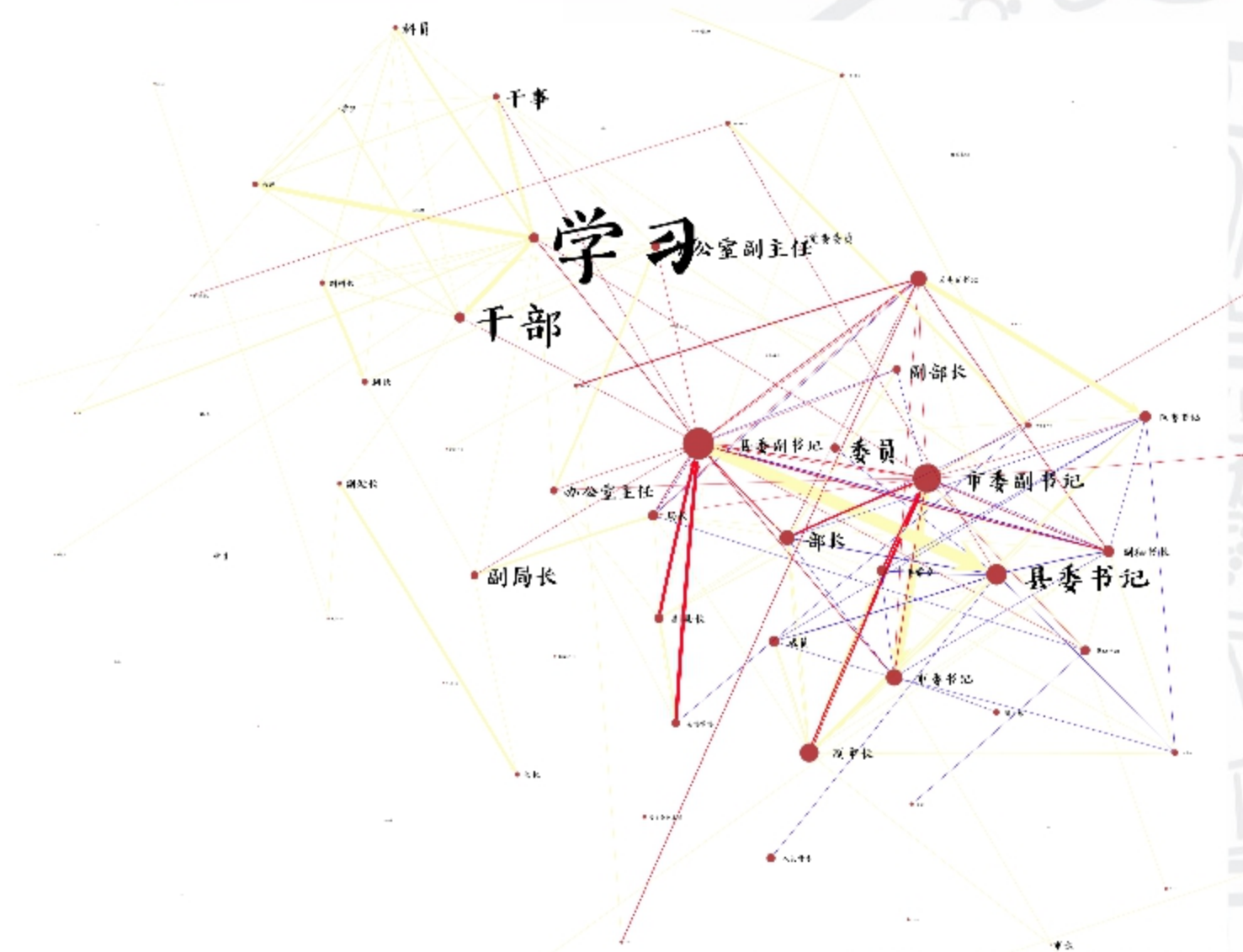


**word
embeddings
(word2vec)**

sklearn: calculate cosine distances



graphtool: modelling networks



The work has just begun...

- **classify protests vs. non-protests correctly**
- **classify online replies correctly**
- **model relationship between complaints, protests, policies, promotions/demotions**
 - bayesian networks?
 - markov networks?
 - reinforcement learning?
 - ...?