



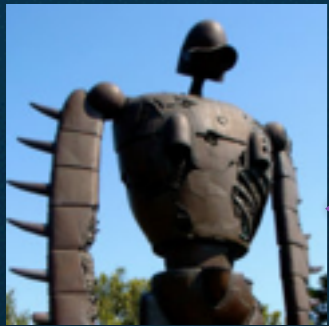
微博大数据 标签体系建设

微博平台及大数据部

杜传赢 @cydu

chuanying@staff.weibo.com

<http://weibo.com/cydu>



微博是什么? 希望微博成为: 获取所有公开信息的一种方式



微博数据多吗?



中国脉搏

- 日活跃用户: 6660W
- 春晚1分钟峰值: 863, 408
- 千亿级的微博
- 千亿级的用户关系
- 百亿级的图片
- 每日万亿级的内部调用量

我们的需求

电影

推荐

广告

美食

搜索

...

什么领域的?
讲了什么?
谁会对他感兴趣?



文怡 V

我儿以前用过几次北医三院的鞣酸软膏，还行。但主要还是得多让孩子晾晾屁股，仅个人经验。谢谢你在此贴下的留言，这个周末，完美了😏。//@小GG走天下：文大美丽，求助啊！娃4个月，最近屁眼红红的，试过好些方法了，都木有效果。咋整???? 跪求回复😞😞



水洋 ★ (林水洋)

@郭蕊 老师推荐的 **Erawan Tea Room** 位於曼谷君悅酒店商場樓2樓，是一家非常好吃的泰餐廳。在普吉島連續喝了6天東陰功後，回到曼谷，這家餐廳吃下來能給人很好的美食體驗，特別是咖喱。普吉島對於泰國還是像極了三亞：糟糕的餐飲和昂貴的物價，好在風景和酒店還是很好的。

| 泰国 บางพลี

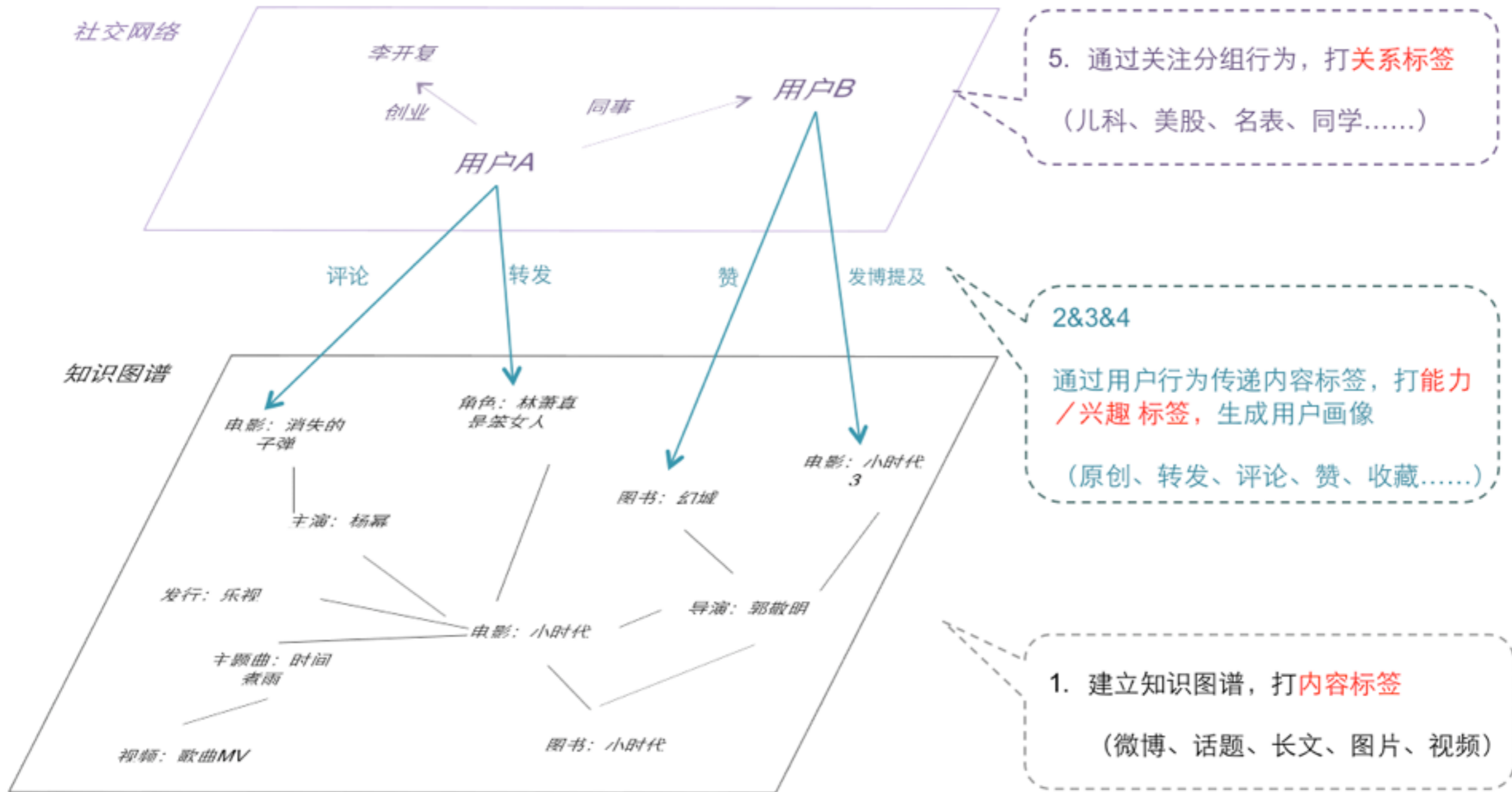
标签化



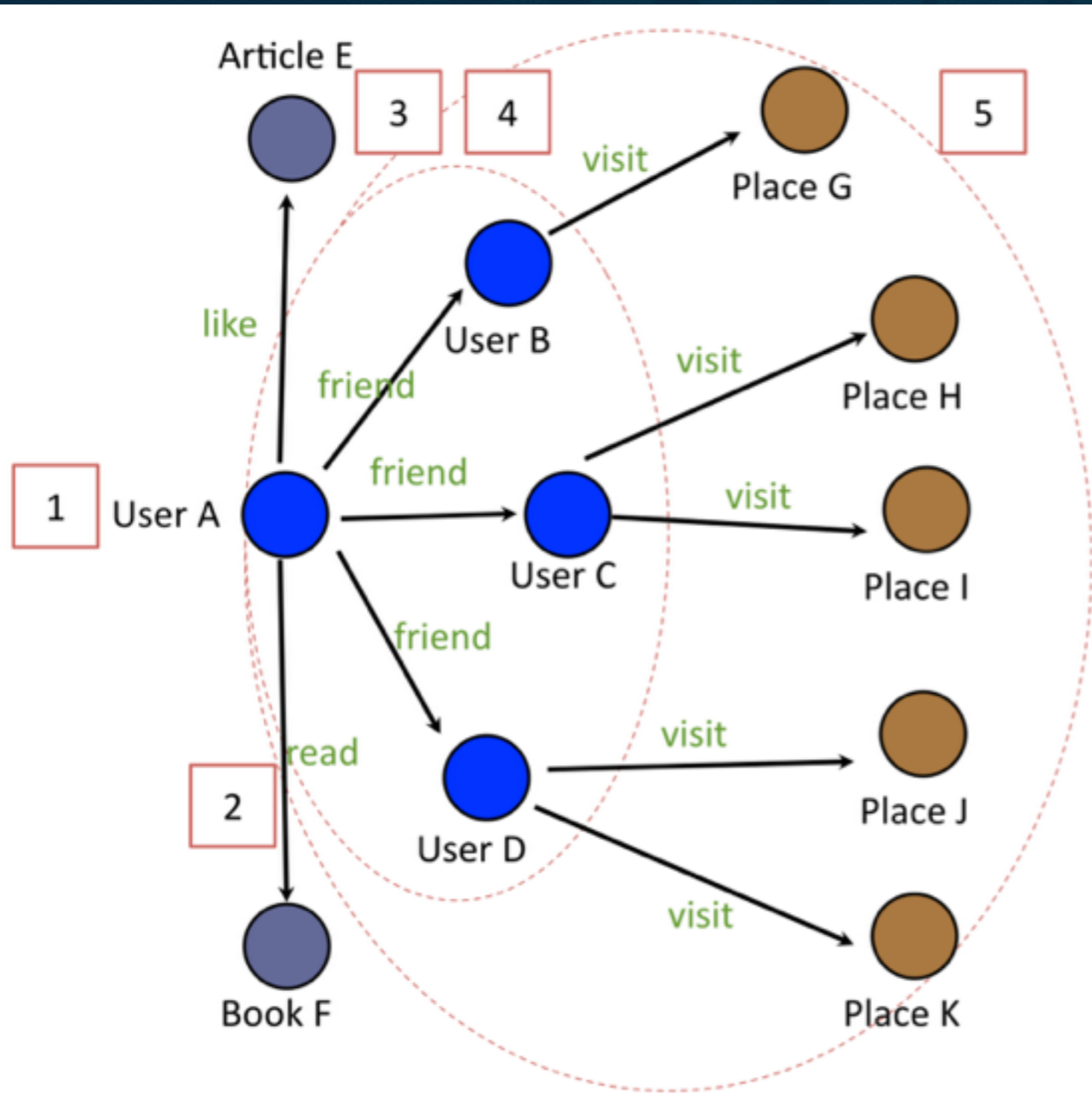
标签分类

- 知识图谱 => #-# / 对象间的关系
- 内容标签 => #对象 / 微博, 图片, 电影, 音乐...
- 能力标签 => @生产者 / 用户, 商户
- 兴趣标签 => @消费者 / 用户
- 关系标签 => @-@ @-# / 对象, 人之间的关系

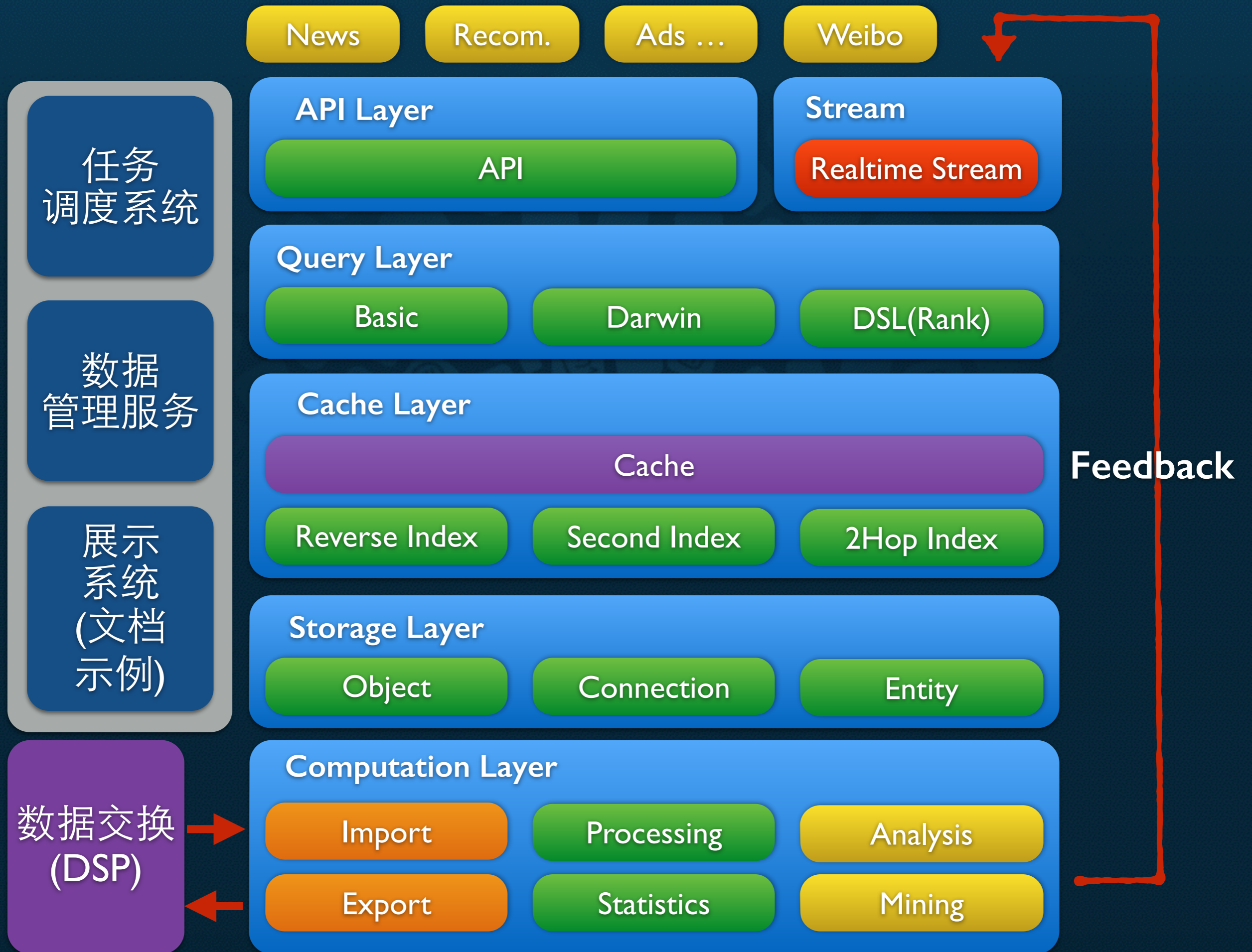
怎么打标签



数据的组织



1. 点
2. 边
3. 一度
4. Top
5. 二度



应用示例

4分钟前 来自手机微电影

提及: 足球国家队:美国队 足球国家队:德国队 足球赛事:世界杯 足球国家队:阿根廷队 足球国家队:尼日利亚队

推荐: 足球:国际足球 足球赛事:足球比赛 足球国家队:巴西世界杯球队 体育:体育 足球:足球 足球赛事:国际足球赛事

Do-peng★: #0元体验世界杯足球赛我押了尼日利亚vs阿根廷、美国vs德国 两场赛事,我押了10元,看好尼日利亚 负、美国 胜,世界杯,来mark一下! <http://t.cn/RvM7Uyr>

4分钟前 来自手机微电影

提及: 足球国家队:美国队 足球国家队:德国队 足球赛事:世界杯 足球国家队:阿根廷队 足球国家队:尼日利亚队

推荐: 足球:国际足球 足球赛事:足球比赛 足球国家队:巴西世界杯球队 体育:体育 足球:足球 足球赛事:国际足球赛事

InternetCompany(12个) 总分: 119

阿里巴巴: 48 facebook: 13 百度: 12 腾讯控股: 11 谷歌: 10

奇虎360: 3 搜狐: 3 优酷土豆: 1 微软: 1

InternetPerson(11个) 总分: 91

马云: 25 曹国伟: 13 周鸿祎: 11 李彦宏: 10 王高飞: 10 刘强东: 3

陈彤: 3 比尔-盖茨: 3

travelfestival(12个) 总分: 80

情人节: 20 圣诞节: 11 青年节: 7 元宵节: 6 国庆节: 6 劳动节: 3

标签	总分	纯度	速率	影响力
足球	17.50	0.15	1.08	59.16
西甲	11.45	0.13	0.33	53.23
英超	11.32	0.13	0.31	52.79
中超	9.56	0.07	0.20	51.46

郭少林 (同事)

刘彬 (同事)

叠石 (同事)

来去之间 (同事)

shijin1 (同事)

童剑 (同事)

关系标签

朋友(53)

震公子... (朋友)

Darrenhp (朋友)

yongqia... (朋友)

yujing_kou (朋友)

熊鹏o (朋友)

echo_ye (朋友)

软件(44)

bluedav... (软件)

纯银V (软件)

陈晓鸣... (软件)



微博推荐



微博找人



电影热议

微博正文

韩寒 5-29 09:05 来自微博 weibo.com

《后会无期》，相聚有时，7月24日上映。



相关微博

“中国好岳父”韩寒耗时三个多月拍摄的《后会无期》日前宣布... 来源于 大众电影

转发 37万 | 评论 11万 | 赞 18万

每日热点聚焦 5-29 09:12 1218

亲自配音是韩寒，听出来了吗？很

转发 | 评论 | 赞

微博正文

郭敬明 昨天 14:15 来自微博电影 +关注

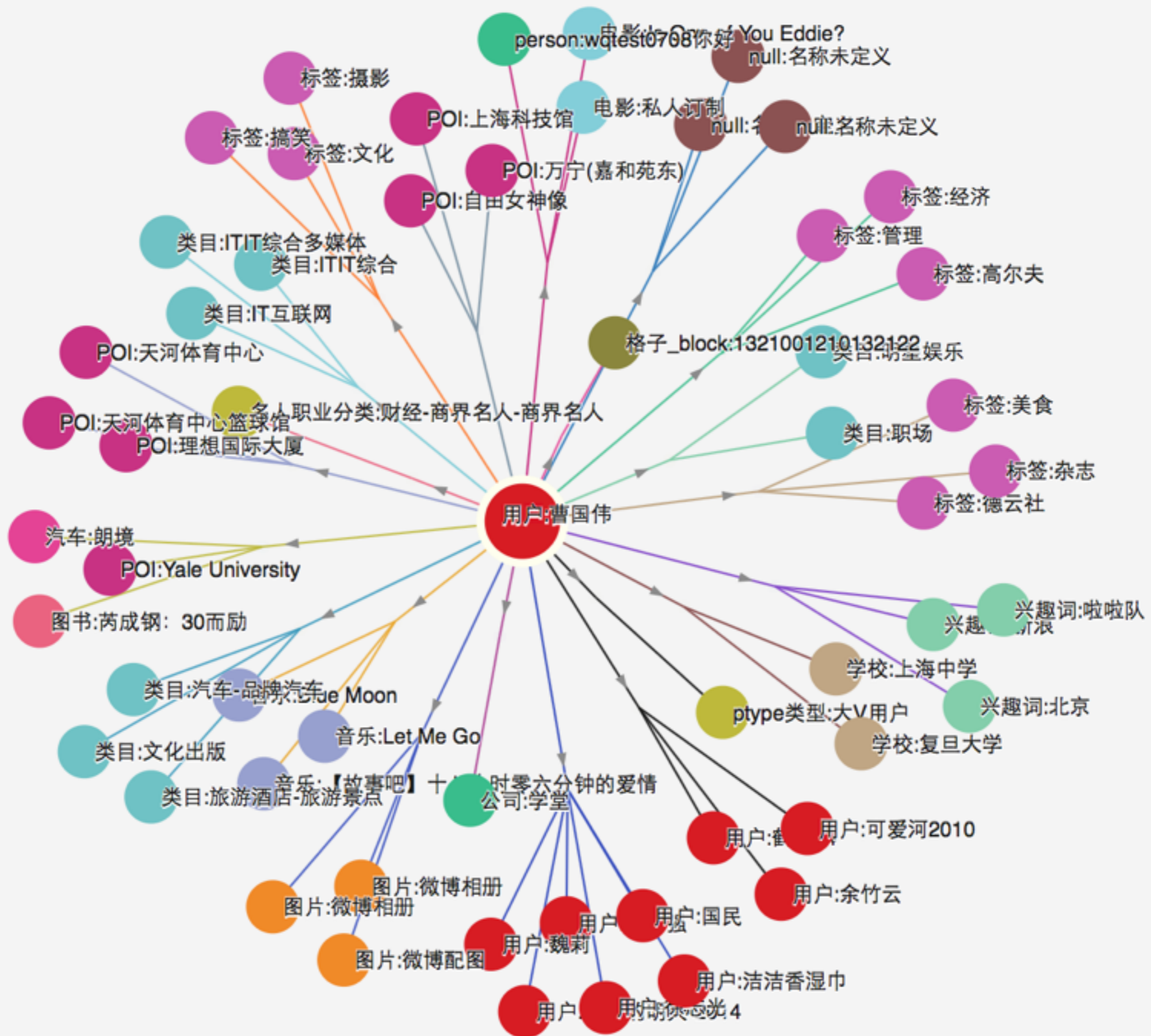
★★★★★怒赞，作为刚入行的新导演，我知道自己还有很多不足。谢谢大家对我的包容和鼓励。但是为了全剧组辛苦付出的主演和主创们，我还是决定怒赞5颗星！谢谢你们，圆了小时代这个梦。7月17日，希望和大家一起分享眼泪和欢笑！

小时代3: 刺金时代 导演: 郭敬明 评分: ★★★★★ 抢票

相关微博

#小时代3#之前跟导演@郭敬明 唧唧歪歪墨墨迹迹的，就知道要... 来源于 乐视网

转发 | 评论 | 赞



对外的合作

- 共建标签体系
- 内容共享
- 商务合作



Thanks!

Big data is like teenage sex:

- everyone talks about it
- nobody really knows how to do it
- everyone thinks everyone else is doing it,
- so everyone claims they are doing it ...