

SACC

卓越 5周年 变迁

SequeMedia
盛拓传媒

IT168
www.it168.com

ChinaUnix

ITPUB

2013中国系统架构师大会

SYSTEM ARCHITECT CONFERENCE CHINA 2013

大数据下的IT架构变迁

构建人人网社会化推荐引擎

邓雄

xiong.deng@renren-inc.com

人人网 应用研究中心

关于我

- 人人网应用研究中心负责人 技术经理
- 曾任职百度商务搜索部鳳巢广告平台 高级研发工程师
- 曾受邀IBM Ireland Research Center演讲数据挖掘相关Topics
- 帝国理工(英) 数据挖掘 PhD



内容提要

- 社会化推荐系统面临的重要挑战
- 社交图谱与社会化推荐系统
- 人人网社会化推荐系统架构
- 人人网社会化推荐主要模型及算法
- 人人网社交图谱挖掘的其他进展
- 社会化推荐的未来





Johnny

VIP 1 46天

我常用的

新鲜事

与我相关 5

相册 20+

好友

分享

公共主页

状态

我的分组

特别关注

好友内容 20+

大学同学 14

普通朋友 1

中小学同学

家人亲戚 2

发状态

传照片

发分享

评电影

写日志

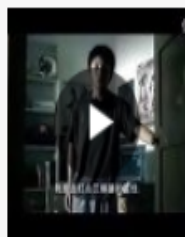
你正在干嘛？

人人话题：盘点人人网摄影圈上那些大牛们，快来膜拜！



施云雯Margaret：o//人人公益：家，以爱为圆心，幸福为半径的一个圆...

...

新加坡感人公益广告
来自：韩荣斌

今天 00:27 通过人人Windows Phone发布

赞 分享



施云雯Margaret：o

2013-08-01 00:27

评论...



殷利锋alexander：仅仅几千字建国前讲到现在，是纪实，但看起来像小说，活着//张学斌：//刘承泽：//张俊：当年读的时候只读出个政治学的味道//孙浩然：卧槽这篇文章不能更赞了，集历史学政治学经济学于一身啊，可做政治经济经典分析案例了，十分满分我打11分草草真的不能更赞了//林丝言：分享日志

最近来访 3886



好友推荐



米存仓

+ 好友



沈小虎

+ 好友



于琦

+ 好友



张扬

+ 好友

携手巨星，为音乐而创



我要参加

+ 分享

好友生日 今天 | 08-02 | 08-03 | 更多



林淇莎



下载人人桌面



好友列表

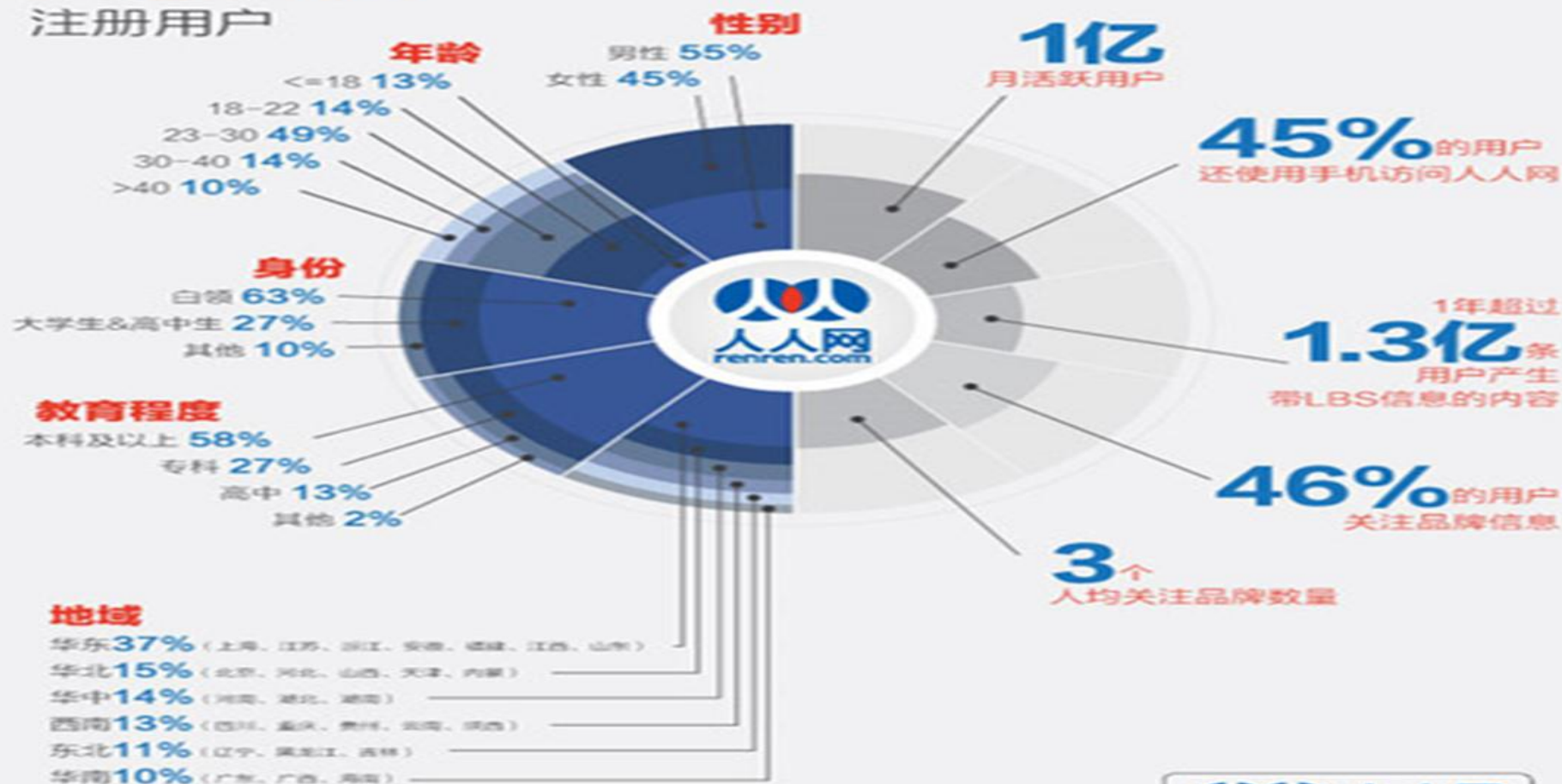
NEW

The Real renren.com

你所不知道的人人网

2.2亿

注册用户





Johnny

VIP 1 46天

我常用的



发状态



传照片



发分享



评电影



写日志

你正在干嘛？

热点内容发现、预测

最近来访 3886

导航浏览
优化

- 好友内容 20
- 大学同学 14
- 普通朋友 1
- 中小学同学
- 家人亲戚 2

新鲜事排序

+

用户关系建模

今天 00:27 通过人人网

赞 分享



施云雯Margaret

2013-08-01 00:27

评论...



殷利锋alexander : 仅仅几千字建国前讲到现在,是纪实,但看起来像小说,活着//张学斌://刘承泽://张俊:当年读的时候只读出个政治学的味道//孙浩然:卧槽这篇文章不能更赞了,集历史学政治学经济学于一身啊,可做政治经济经典分析案例了,十分满分我打11分草草草真的不能更赞了//林丝言:分享日志

好友推荐

米存仓

沈小虎

于琦

张扬

+ 好友

+ 好友

+ 好友

+ 好友

广告排序优化

立即参与

赢取太阳马戏团JAY-Z不朽巡演门票

JAY-Z

我要参加

+ 分享

好友生日

今天 08-02 08-03 更多



林淇莎



下载人人桌面

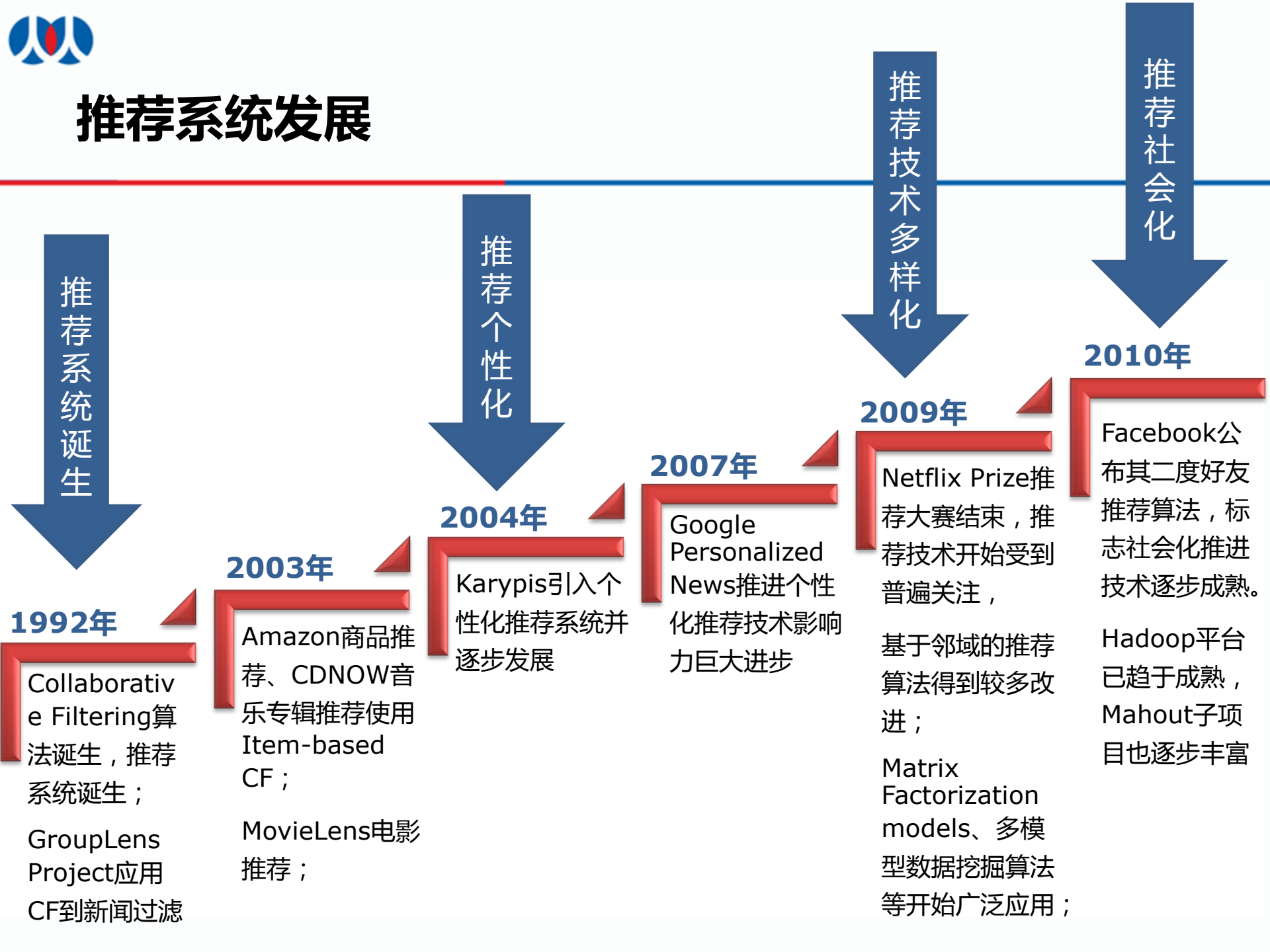


好友列表

NEW



推荐系统发展



社会化推荐系统

- 背景：信息爆炸、信息过载
 - ✓ 1分钟互联网产生多少数据？
 - 48小时新视频@Youtube
 - 2000000次搜索请求@Google
 - 684478分享消息@Facebook
 - 100000条微博@Twitter
 - 3600张照片@Instagram

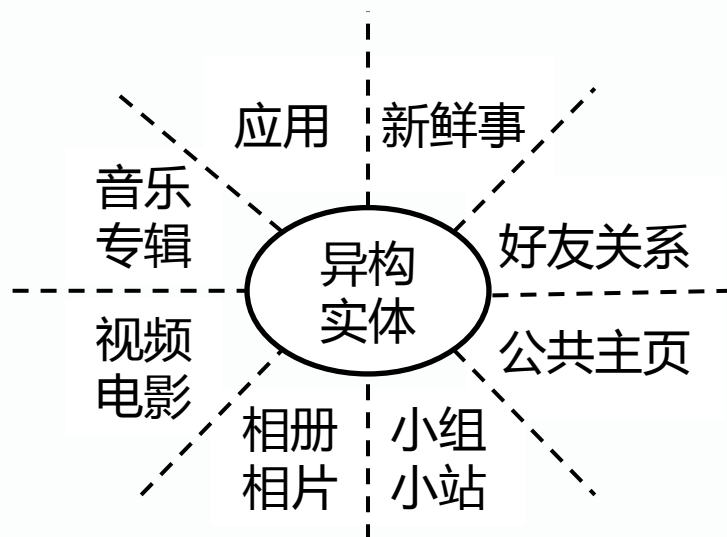
*“We are moving from an **Information Age** to the **Recommendation Age**. ”*
– “The Long Tail” by Chris Anderson

- 推荐系统：给用户推荐符合其喜好的信息
 - ✓ 好友推荐，商品推荐，日志推荐，视频推荐，App推荐,广告推荐
 - ✓ Amazon, Facebook , Google, Netflix, Youtube, Apple...



社会化推荐系统

- 社交网络的信息有什么特殊之处？
 - ✓ 信息特性：多样化，网状化，异构化实体
- 异构复杂数据



社会化推荐系统

- 社交网络的信息有什么特殊之处？
 - ✓ 信息特性：多样化，网状化，异构化实体
 - ✓ 社交特性：社交关系是一种特殊的信息



社会化推荐系统

- 社交网络的信息有什么特殊之处？
 - ✓ 信息特性：多样化，网状化，异构化实体
 - ✓ 社交特性：社交关系是一种特殊的信息
- Social Graph

“The global mapping of everybody and how they're related. ”

– Brad Fitzpatrick



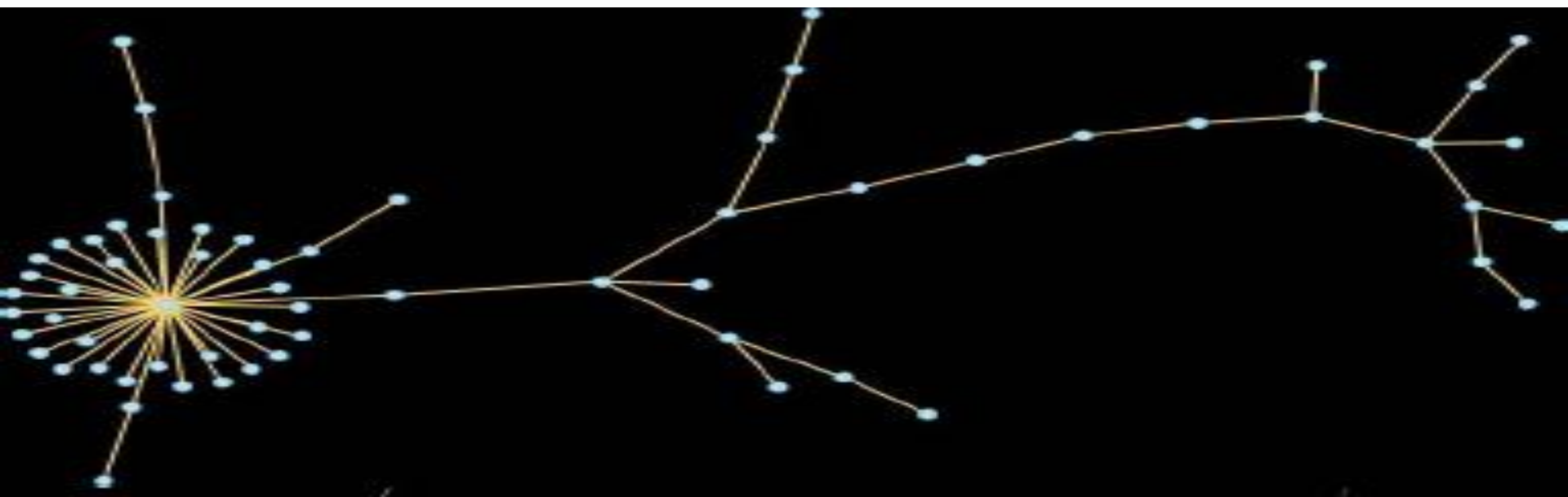
社会化推荐系统

- 社交网络的信息有什么特殊之处？
 - ✓ 信息特性：多样化，网状化，异构化实体
 - ✓ 社交特性：社交关系是一种特殊的信息
 - ✓ 流动性：信息在社交网络中动态流动，加速增长，充满噪声



社会化推荐系统

- 社交网络的信息有什么特殊之处？
 - ✓ 信息特性：多样化，网状化，异构化实体
 - ✓ 社交特性：社交关系是一种特殊的信息
 - ✓ 流动性：信息在社交网络中动态流动，加速增长，充满噪声
- Cascades



社会化推荐

好友推荐



唐伟

+ 好友



李聪

+ 好友



刘健

+ 好友



张栋杰

+ 好友

找公司同事

完善资料找好友



陆韵

广汽丰田有限.

+ 好友



林智

广汽丰田有限.

+ 好友



邓大为

广汽丰田有限.

+ 好友



李国荣

广汽丰田有限.

+ 好友



高俊鹏

广汽丰田有限.

+ 好友



庾健民

广汽丰田有限.

+ 好友

推荐加入本组的好友



崔硕朋

+ 好友



王尔晴

+ 好友



潘朋奇

+ 好友



魏斌 Bin

+ 好友



丁意龙

+ 好友

推荐



彭湃

+ 加为好友

查看资料(90个共同好友)

性别：女生

家乡：北京 房山区

城市：北京市

共同好友：丁伟彬 刘启荣

李婧 mika...♥ 李勇 林颖璐 柳煜 马晖

吴璐 瑞华华♥小蜜枣 赵秀丽 等90人



谢叶山

+ 加为好友

查看资料(28个共同好友)

1



人人网 我们猜您可能认识这些人



蔡蕊

30个共同好友



刘耿耿

30个共同好友



熊玮

30个共同好友



亢爱迪

30个共同好友



马春阳

30个共同好友

还有10位好友

9分钟前

推荐



单身男女俱乐部

+ 加入小组

赵朝 瑞华华 等3位好友已加入



数码单反DSLR俱乐部

+ 加入小组

黄晶 等1位好友已加入

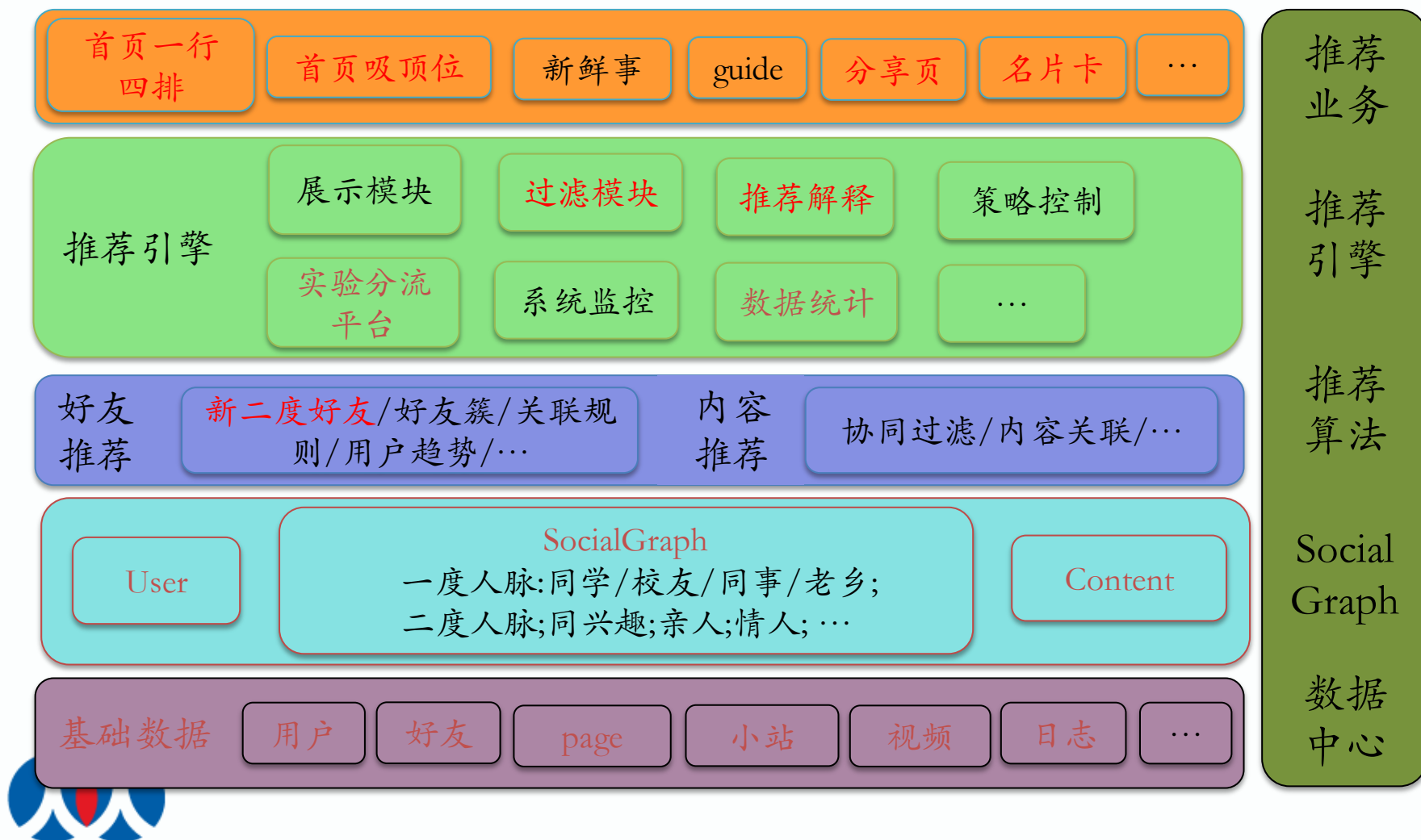
热门推荐



王源 上传于2013年8月23日



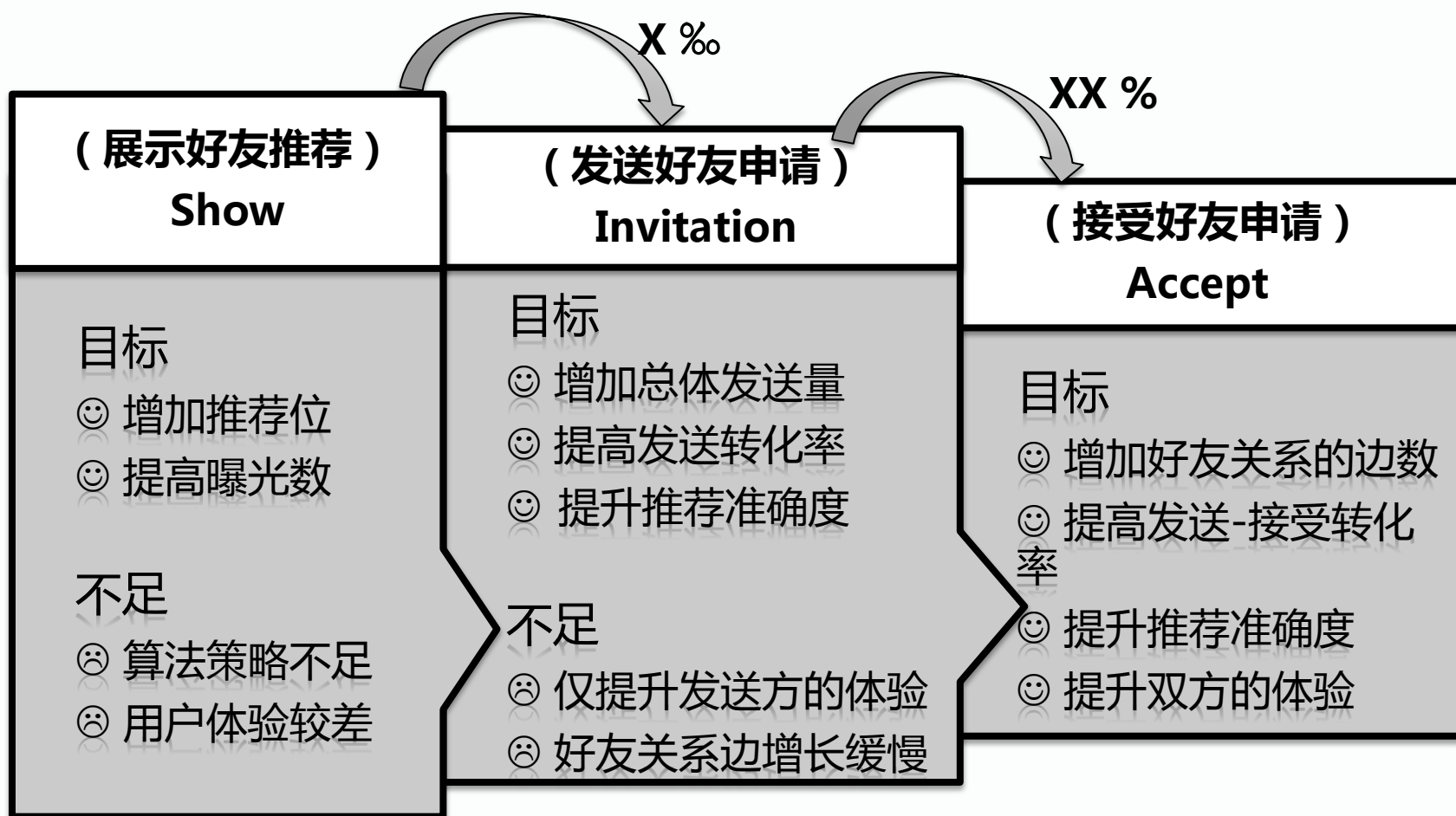
社会化推荐引擎



Friend Recommendation

- Challenges

- ✓ Two-step link prediction (bi-directional con.)



Friend Recommendation

- **Challenges**

- ✓ Two-step link prediction
- ✓ Reduce redundant shows of recommended friends between multiple rec. places

好友推荐



米存仓
+ 好友



沈小虎
+ 好友



于琦
+ 好友



张扬
+ 好友

1



人人网 我们猜您可能认识这些人



蔡蕊
30个共同好友



刘耿耿
30个共同好友



熊玮
30个共同好友



亢爱迪
30个共同好友



马春阳
30个共同好友

4

还有10位好友

9分钟前

推荐



彭湃

+ 加为好友

查看资料(90个共同好友)

性别：女生

家乡：北京 房山区

城市：北京市

共同好友：丁伟彬 刘启荣

李婧 mika...♥ 李勇 林颖璐 柳煜 马晖

吴疆 瑞华华 小蜜枣 赵秀丽 等90人



谢叶山

+ 加为好友

查看资料(28个共同好友)

1. Friends of Friends

• Challenges

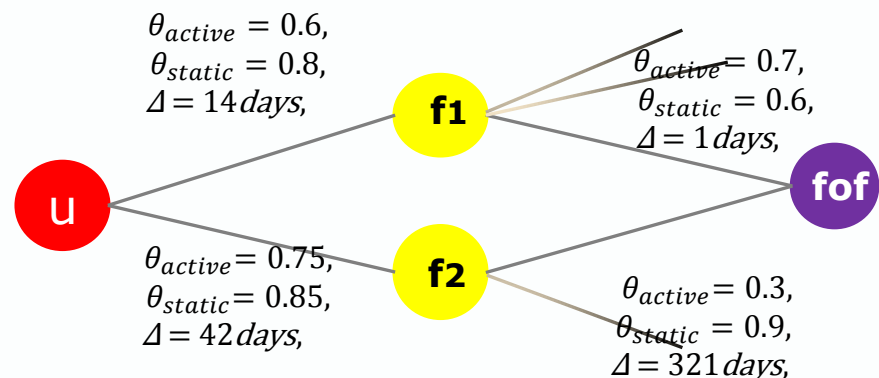
- ✓ 怎样推荐潜在最佳二度好友？
- ✓ 平均每个用户二度好友数在4万
- ✓ 2亿用户*4万=8万亿
- ✓ 网状海量数据：**200TB**
- ✓ 日更新数据：**10亿次**

• Computing paradigm

- ✓ 建立好友关系的可能性：有10个共同好友是有1个共同好友为12倍
- ✓ HDFS/HBase/MapReduce

• The system

- ✓ 线下计算：全量更新4.5小时/次；增量1小时/次
- ✓ 线上更新：月活跃用户每天2次，日登用户每天4次



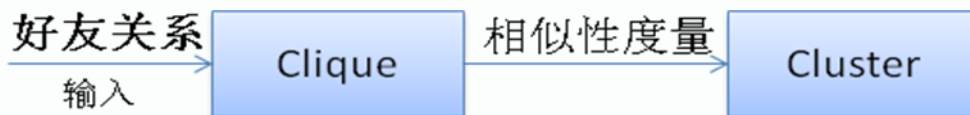
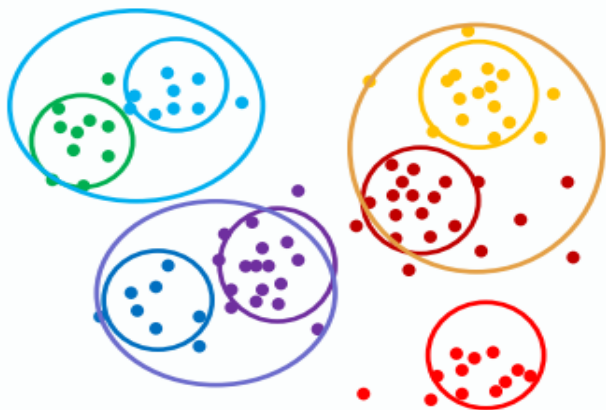
$$v(fof) = \sum_{fi} F(\delta_{u,fi}, \delta_{fi,fof}, friends_{fi}, \theta_{static}, \theta_{active})$$

• Results

- ✓ 好友申请发送率提高近100%
- ✓ 首页推荐位发送申请数提高近70%

2. 好友簇算法

- 算法基础：
 - ✓ 社会网络拓扑结构中存在大量的强关系紧密结构
 - ✓ 将用户分成簇，同簇的成员之间具备相近的社交关系
- 好友簇模型：
 - ✓ 最大团，层次聚类
- 效果：
 - ✓ 加好友转化率提高122%
 - ✓ 首页推荐位加好友数提高近89%



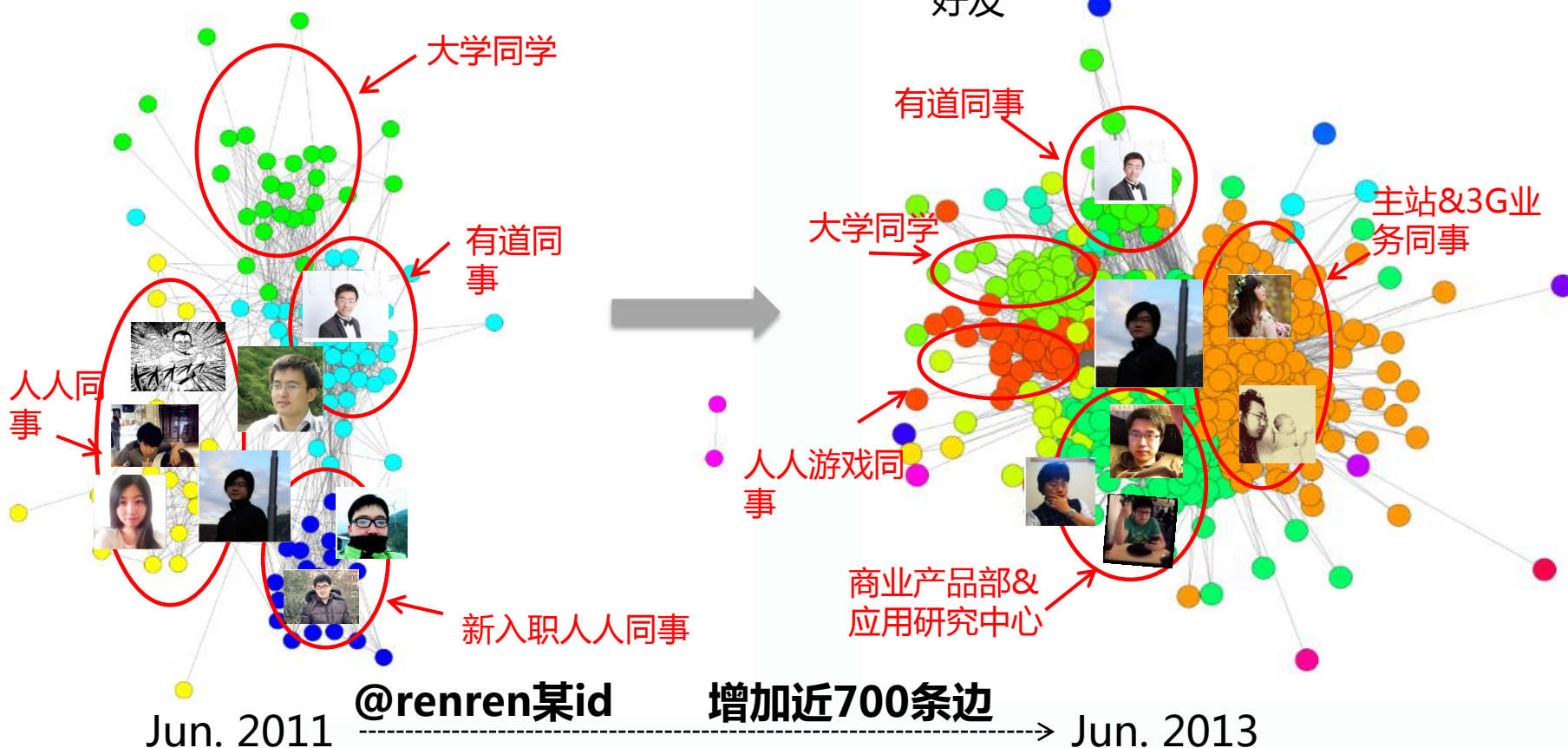
3. Predicting Evolution Trends of Friend Links

• Motivation

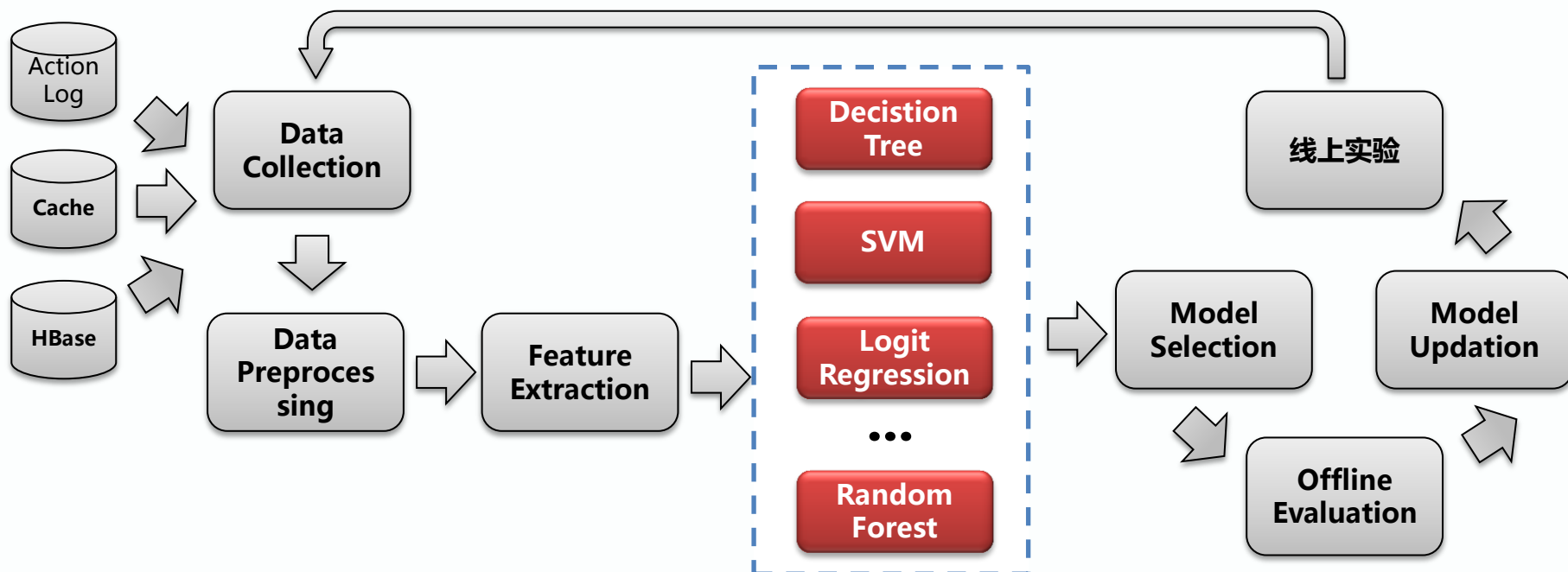
- ✓ 好友最大扩充的方向未必是当前时间最大的群组

• The model

- ✓ 发现用户对不同群组的倾向性
- ✓ 根据用户反馈及时调整倾向性
- ✓ 为用户推送当前时刻最想扩充的群组对应的好友



4. Estimating Acceptance Probability of Friend Invitations



多数据源整合生成特征

- Action/Business LOG
- Cache
- MySql
- Hbase

多模型组合提高精度

- Boosting
- Random Forest

离线评估/模型更新自动化

- MAE/RMSE
- AUC
- 训练、评测、更新自动化

线上实验

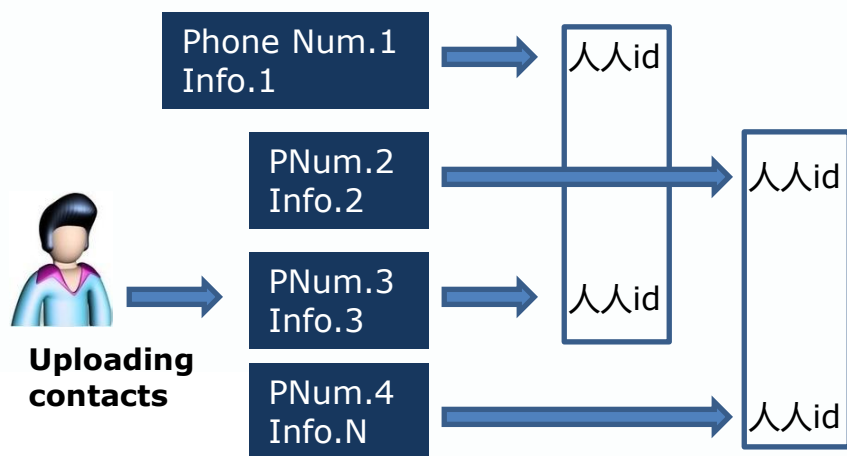
- A/B test
- 实时报表反馈

5. Merging Contact Graph into Social Graph

• Challenges:

- ✓ Recommending users with smartphone contacts in **real time** ?
- ✓ Integrating multiple social graphs as **widely** as possible ?

A1 : Building multiple indices , recommending by searching



A2 : Matching Info. instead of phone numbers only

The Algorithm2

1) 构建<手机号, 姓名>列表

<手机号1, 姓名1>
<手机号2, 姓名2>
<手机号3, 姓名3>
<手机号1, 姓名4>
<手机号1, 姓名5>
<手机号3, 姓名6>
<手机号2, 姓名7>

→

<手机号1, 姓名1|次数, 姓名4|次数, , >
<手机号2, 姓名2|次数, 姓名7|次数, , >

2) 匹配姓名, 构建<手机号, 人人id>索引。

<id1, 姓名1>
<id2, 姓名2>
<id3, 姓名3>
<id4, 姓名4>
<id5, 姓名5>

matching

<手机号1, 姓名1>
<手机号2, 姓名2>
<手机号3, 姓名3>
<手机号4, 姓名4>
<手机号5, 姓名5>

人人网一度及二度好友

手机通讯录

3) 将<手机号, 人人id>索引用于推荐

人人网社会化图谱挖掘的其他进展

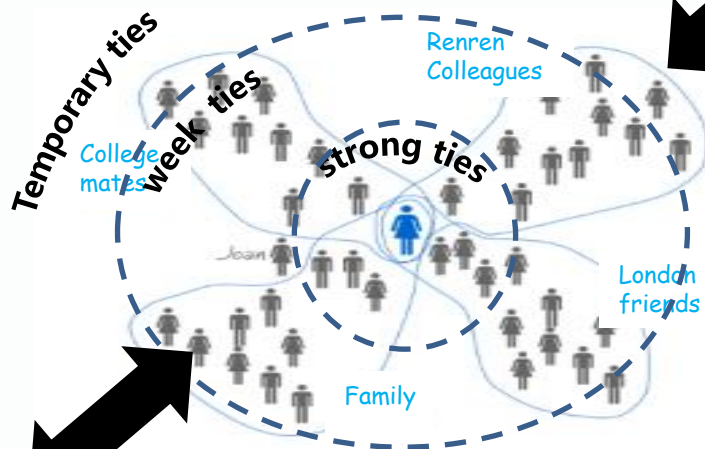
Interest Discovery

Products : 人人头条、
横插内容新鲜事推荐



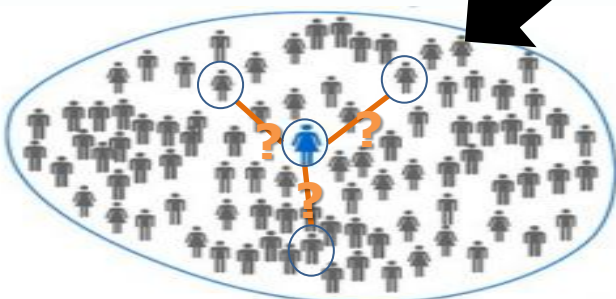
Friend Grouping

Products: 群聊讨论组推荐、好友分组



Friend Rec.

Products: 可能认识的人、
附近的人



好友智能分组及命名算法；
群/讨论组推荐

二度好友模型、基于加好友趋势预估模型；接受率模型；
通讯录合并算法；
好友亲密度模型

1. Friend Strength – Why Useful ?

- Friend ranking
- UGC ranking

687 好友

管理



共同好友 (36)

查看全部



徐瑞航



黄晶



史颖华



朱允铭



邹宇



马洁馨

发起聊天

+ 创建

- ▶ 群 1
- ▶ 讨论组 11
- ▶ 最近联系 20
- ▼ 在线好友 274



小黄鸡



朱小强



晏芳



梁小隼



曹森 **VIP 6**: 发现08,09年未登录天数有点多, 求购买补登陆卡渠道。心理价位不超过10块一张。🐼

昨天 17:37

👍 赞(1) ➡ 分享(1)

👍 1个人觉得赞!



张振宇: 同求 能不能 连续续上给个VIP

2013-08-01 17:45

评论...



Johnny Deng : 哇晒! //曾珍://苏杨: 分享照片

人人(NYSE:RENN)

Renren Inc.

\$3.68 +0.350 (+10.51%)

2013-8-01 11:50:00 (北京时间)
实时行情

盘后交易 0.00 (0.00%)

今开: 3.37

昨收: 3.33

最高: 3.68

最低: 3.37

成交量: 15.55万/186.00万

52周最高: 4.39

52周最低: 2.52

每股收益: 0.20

每股净资产: 2.92

总股本: 3.00亿

总市值: 19.97亿

换手率: 10.00%

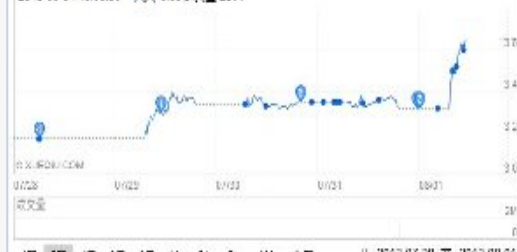
股票对比 - (净值单位: 美金)

历史行情

实时行情

技术分析

2013-08-01 10:00:00 *人入 3.05 成交量 2001



1. Friend Strength – the Method

- What matter?
 - ✓ 交互行为
 - ✓ 用户资料
 - ✓ 用户主动排序
 - 特别好友
 - 新鲜事黑白名单
- The method
 - ✓ 交互建模
 - 以交互周期为单位增量计算
 - 不同行为赋予不同权重
 - 区分交互对象，人重要于内容
 - 区分短时间连续交互行为与间隔交互行为
 - ✓ 静态资料建模
 - ✓ 资料相似度
 - ✓ 共同好友



The Result (某id)

• Top 15 Strong Renren Friends

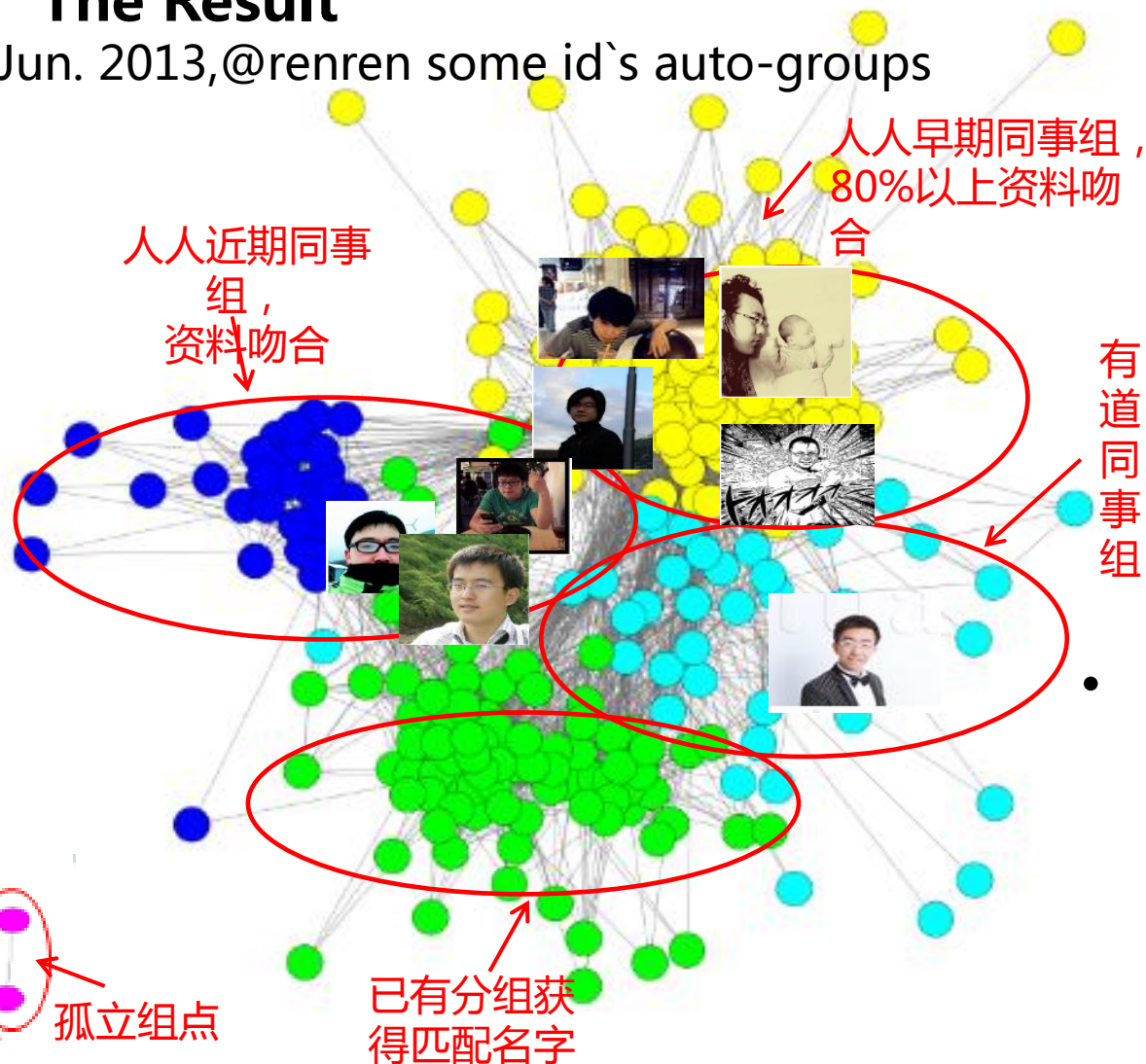
- | | | |
|----|-----------|-----------|
| ① | 224548194 | 金玲 |
| ② | 142525482 | 汤成信? |
| ③ | 231318925 | 任今涛 |
| ④ | 236254775 | 王洋 |
| ⑤ | 103113664 | 陈一舟 |
| ⑥ | 200000130 | 刘健 |
| ⑦ | 225017732 | 唐智晖 happy |
| ⑧ | 237146562 | 吴疆 |
| ⑨ | 200000032 | 庄宏斌 |
| ⑩ | 79512511 | 许朝军 |
| 11 | 200301672 | 李旭 |
| 12 | 230393642 | 张新泳 |
| 13 | 224322196 | 裴圣愚 |
| 14 | 200000053 | 刘启荣 |
| 15 | 200000108 | 唐炜 |
| 16 | 221372767 | 金鹏 |



2. Friend Grouping

- **The Result**

Jun. 2013, @renren some id's auto-groups



- **Problems**

- ✓ Groups are fundamental to cascades and privacies.
- ✓ However, to create groups manually is inconvenient for many, especially for users of many friends.
- ✓ How to evaluate results automatically

- **The model**

- ✓ Improving Walktrap to group friends
- ✓ Filtering, matching, merging, naming groups based on users' info. and existed groups.

Feedback



曲莉莉: 人人智能分组居然都能把我的教学班和行政班同学完好无误的分到两个组。给数据挖掘跪了

2013-07-25 10:58

分享



王丹妮.caj: 人人好友的智能分组真是高霸上吊炸天!! 连我08年洛杉矶游学团, 10年夏威夷夏令营, 10年北大模联.....都能认出来= =.....

2013-07-04 17:40

分享



张竞月: 好久不上人人, 发现新出了一个智能分组功能, 分别将我的小学, 初中, 高中, 大学同学分开, 看看里面的人, 好想问问你们过的还好吗?

2013-07-14 20:46

分享



朱晓倩: 一个互联网产品做到用户需求的前面是令人感动的, 曾经淘宝给我感动过, 一个功能想都没想过, 但是当我看到之后, 真心觉得做到我的心里。。。转自曹淼: 多年梦想, 一朝实现! 🎉🎉 转自Johnny Deng: 顶! 好友分组下一步在自动命名和自动分组的结合、自动分组的精准度、自动分组的可视化显示上还需要再持续优化和突破。转自谭译泽: 继续努力👊 转自尹文斌



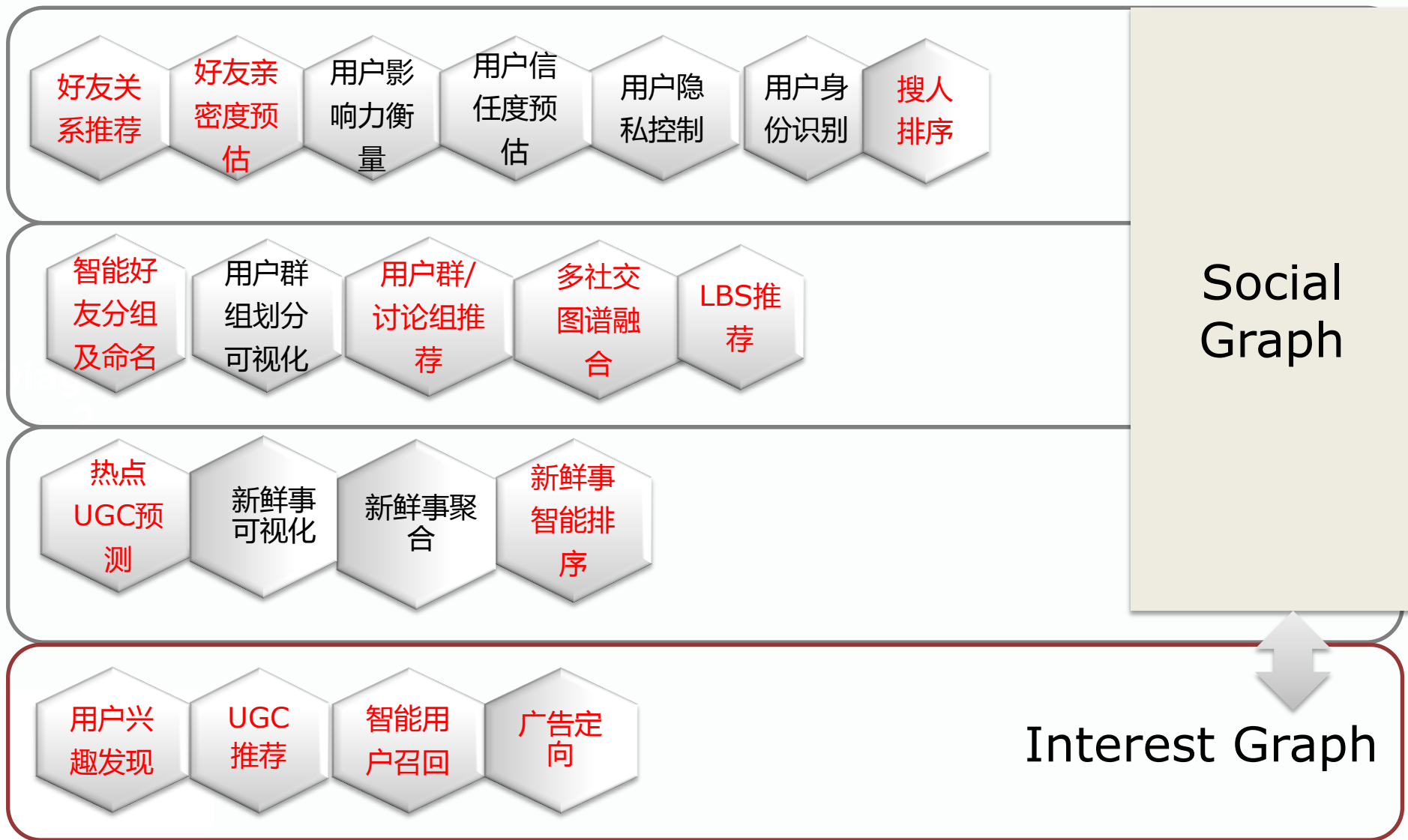
邓捷.mmmap: 我了个大C, 人人现在提供好友智能分组了, 非常精准.....原来人人公司的程序员不是盖的啊.....

2013-07-07 18:07

分享



Mining Social Graph: a Preview

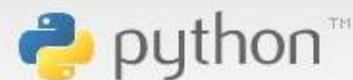


我们用到的技术



Protobuf

Breakpad



STL



APACHE
HBASE

NGINXTM

我们的主要研究方向

方向	内容
推荐引擎	研发面向超大规模数据的多异构实体推荐框架及算法
机器学习、数据挖掘	整合并改进机器学习前沿关键技术成果，诸如分布式、并行机器学习、图学习、文本挖掘、数据仓库
社交网络	研究SNS的社会化生态圈，包括：发现SNS中高复杂度、高维拓扑结构；识别SNS中社会化信息及关系的扩散原理和方式；理解SNS实体的社会化属性演化规律等。
自然语言处理	研究文本分类和聚类、信息检索和过滤、信息抽取、问答系统、机器翻译、新信息检测等前沿技术等
应用云计算及分布式计算架构	开展基于Hadoop、Hbase、Pig、Hive等分布式框架的应用计算研究
多媒体处理	图像、视频、音频等的智能识别、有效分析及高效传输等

Thanks!

SequeMedia
盛拓传媒

IT168.com
www.it168.com

ChinaUnix.net

ITPUB

We are hiring...

renren.research.zp@renren-inc.com

社交关系维度	如何衡量？
Trust	I trust this person
Friendship	I consider this person a friend
Interaction Frequency	How often are you in touch with the person?
Relationship Duration	How long have you known the person
Influence	The person has great impact on me
Reputation	The person is reputable

