

交易网络中的白富美挖掘

Mining Rich Beauties in Transaction Social Network

天猫商业分析部 - 许亮

天猫 TMALL.COM



纲要

- 淘宝业务和数据架构介绍
- 六度分隔理论
- 交易网络中的社会关系
- 交易网络中的“白富美”买家挖掘
- 交易网络中的高品味商家挖掘



淘宝+天猫业务发展

- 2012年年交易突破**1万亿元**
- 2012年双11成交**8400万笔**，成交金额**191亿**元（天猫132亿，淘宝59亿），平均 58000笔/m



纲要

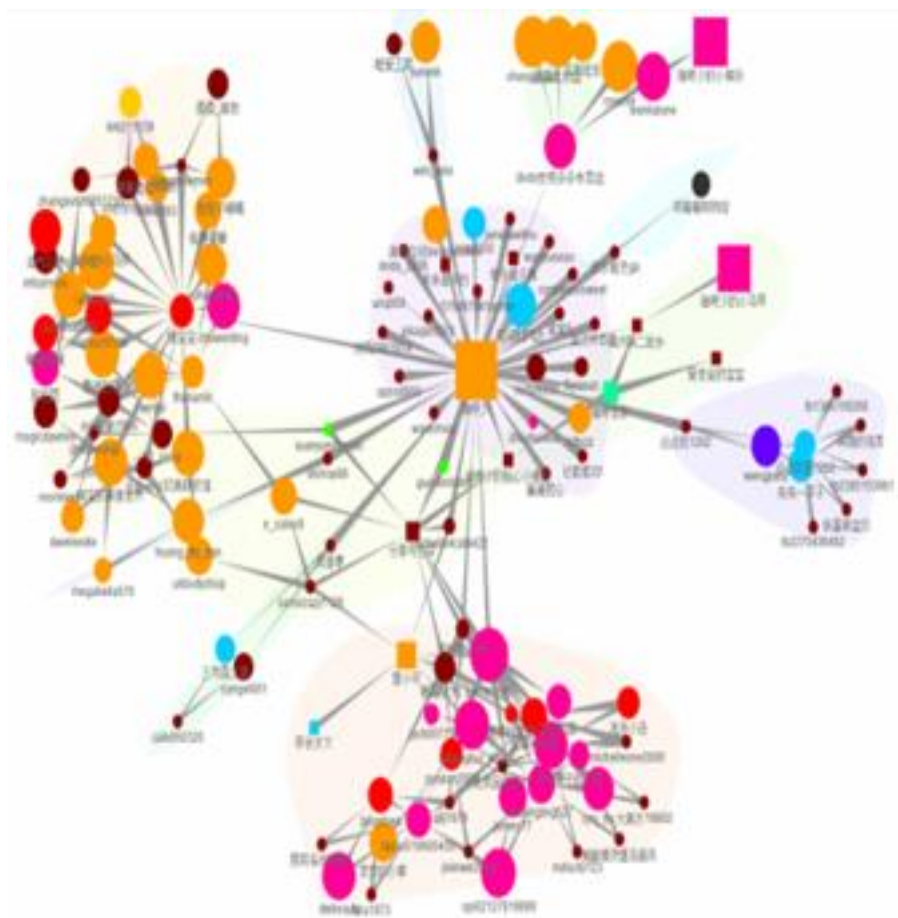
- 淘宝业务和数据架构介绍
- 六度分隔理论
- 交易网络中的社会关系
- 交易网络中的“白富美”买家挖掘
- 交易网络中的高品味商家挖掘



六度分隔理论

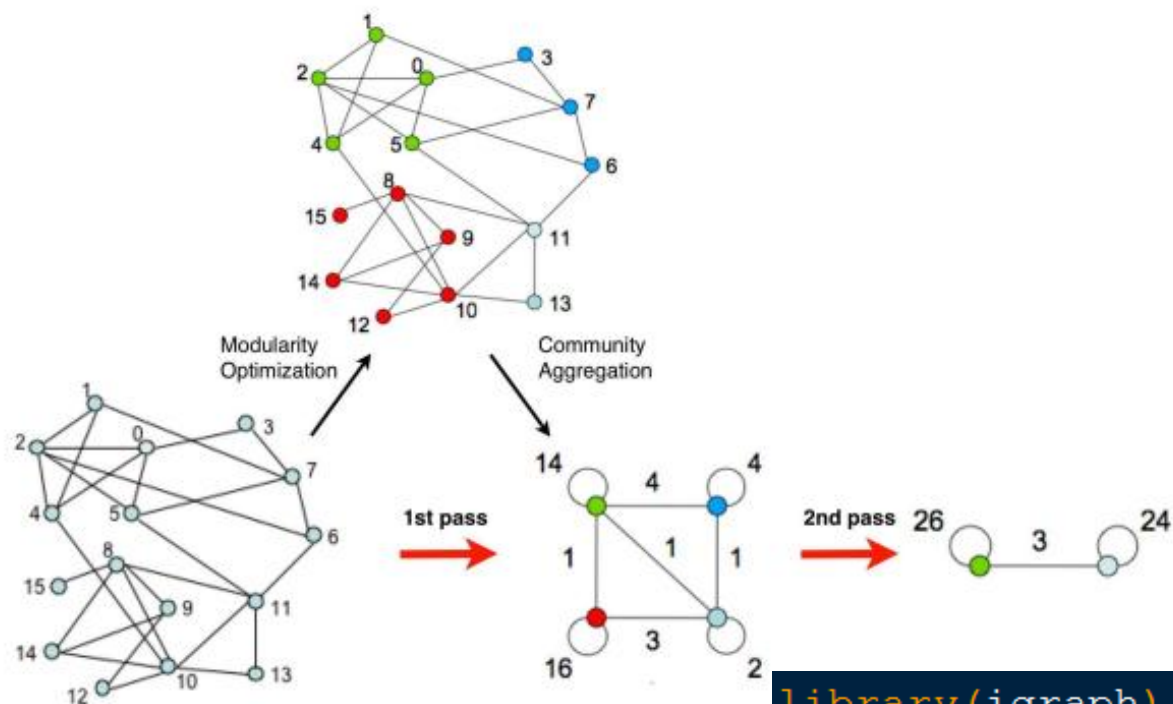
- **小世界现象**（又称**小世界效应**），也称**六度分隔理论**（英文：**Six Degrees of Separation**）。假设世界上所有互不相识的人只需要很少中间人就能建立起联系。
- 后来1967年哈佛大学的心理学教授斯坦利·米尔格拉姆根据这概念做过一次连锁信实验，尝试证明平均只需要6个人就可以联系任何两个互不相识的美国人。

交易网络中的社会关系



淘宝人立方算法

— Community Detection(BGLL算法)



```
library(igraph)
g <- graph.famous("Zachary")
mc <- multilevel.community(g, weights=NA)
plot(mc, g)
```

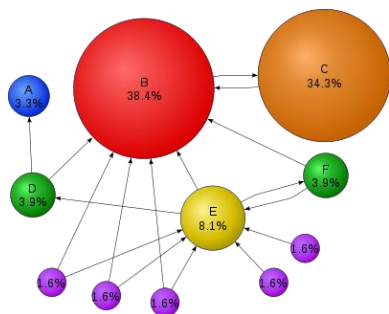
Ref: Vincent D. Blondel, Jean-Loup Guillaume, Renaud Lambiotte, Etienne Lefebvre -- Fast unfolding of communities in large networks

淘宝人立方算法

— Bad Rank (物以类聚, 人以群分)

基本假设: 如果一个网页将其链接指向作弊页面, 则这个网页也很可能是作弊网页

基本假设: 如果一个账号和多个黑名单中的客户进行交易, 则这个账号很有可能有问题



$$\mathbf{R} = \begin{bmatrix} (1-q)/N \\ (1-q)/N \\ \vdots \\ (1-q)/N \end{bmatrix} + q \begin{bmatrix} \ell(p_1, p_1) & \ell(p_1, p_2) & \cdots & \ell(p_1, p_N) \\ \ell(p_2, p_1) & \ddots & & \\ \vdots & & \ell(p_i, p_j) & \\ \ell(p_N, p_1) & & & \ell(p_N, p_N) \end{bmatrix} \mathbf{R}$$

Ref: Zoltan Gyongyi, Hector Garcia-Molina -- Combating Web Spam with TrustRank

纲要

- 淘宝业务和数据架构介绍
- 六度分隔理论
- 交易网络中的社会关系
 - 技术方案
 - **case研究**
- 交易网络中的 “白富美” 买家挖掘
- 交易网络中的高品味商家挖掘

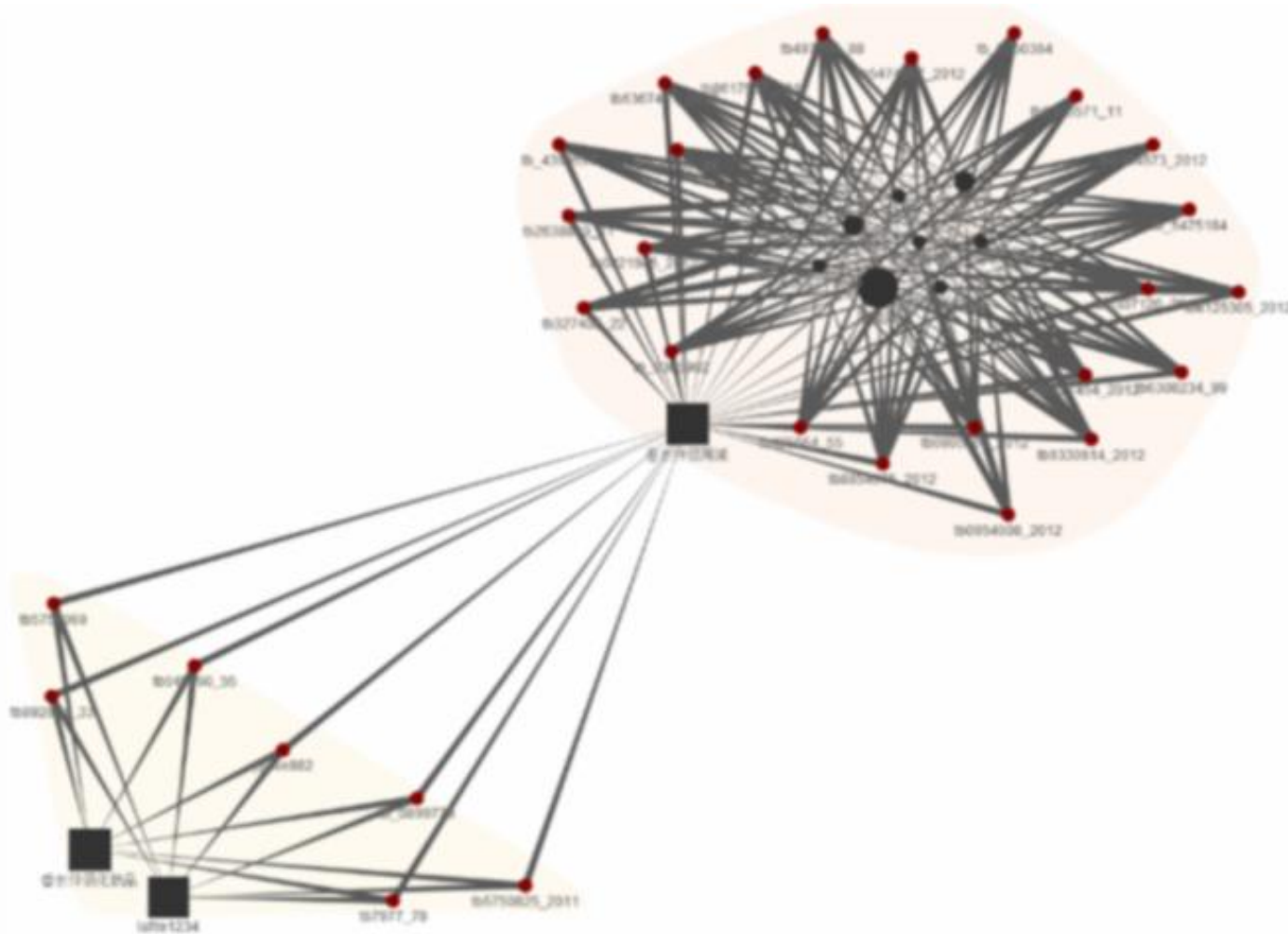
淘宝人立方

-- 供应链研究（供应链利润分析）



淘宝人立方

--挖掘信用炒作



纲要

- 淘宝业务和数据架构介绍
- 六度分隔理论
- 交易网络中的社会关系
- **交易网络中的 “白富美” 买家挖掘**
- 交易网络中的高品味商家挖掘



what else could we do by using social network data ?

“一周新发现”栏目，是找一些平时曝光机会少但有调性的卖家，同时**为这些卖家匹配上对等的买家**



我要建个神秘会场，只能高端用户看到、购买，教育他们:原来天猫还有这么高端的商品

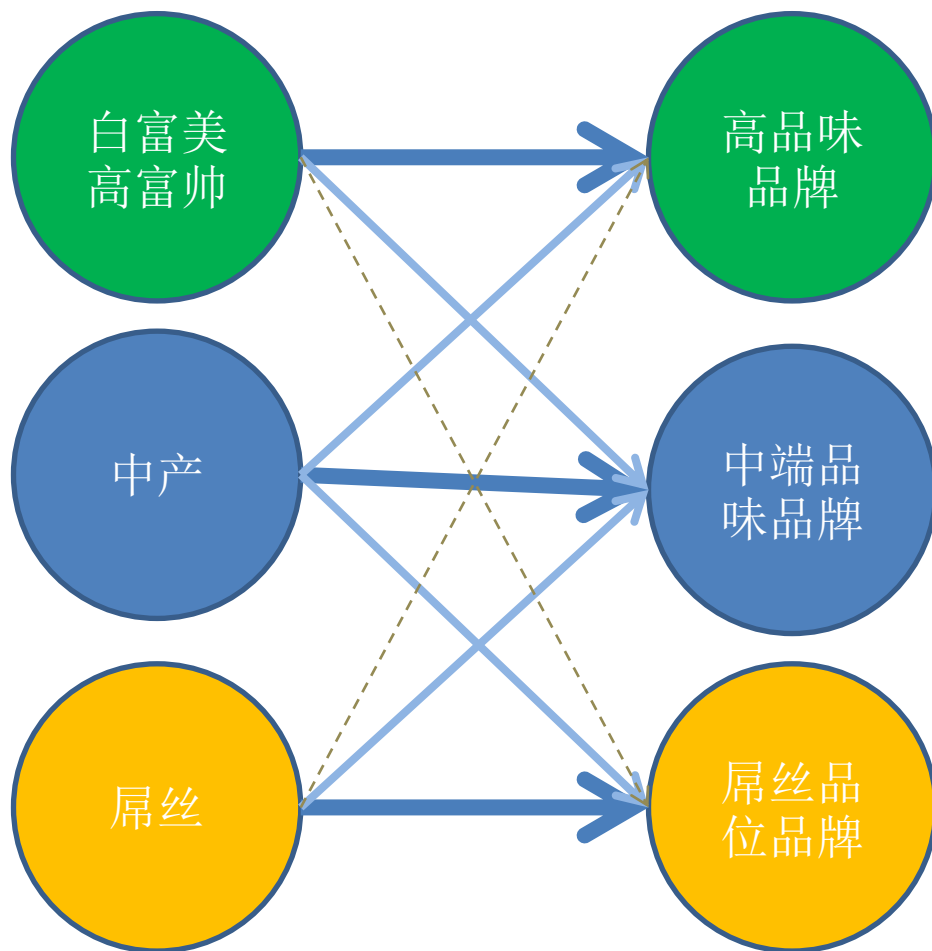
还有我，我们栏目虽然上线有一段时间了，但是流量和转化不是很好，**想定位到真正对这些品牌感兴趣的**用户



交易网络中怎么找到白富美 ???



交易网络中的白富美挖掘



高品位品牌



白富美

交易网络中的白富美挖掘

消费频次

消费档次

品味

交易网络中的白富美挖掘 — 数据标注

设定guideline，抽样各层次用户，进行数据标注（1-5分）

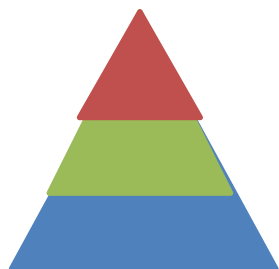
5分：大土豪

4分：土豪，会经常买些奢侈品或者很有品质的装饰等

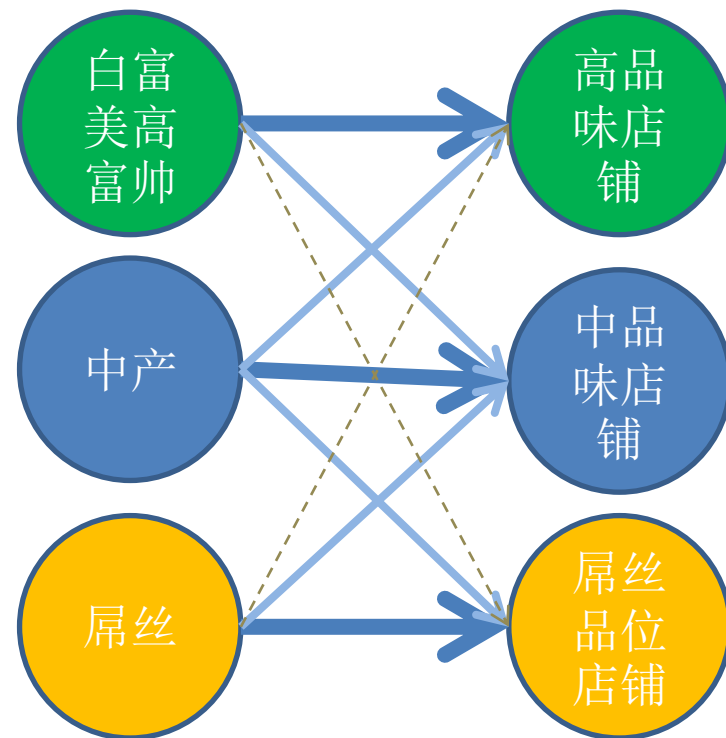
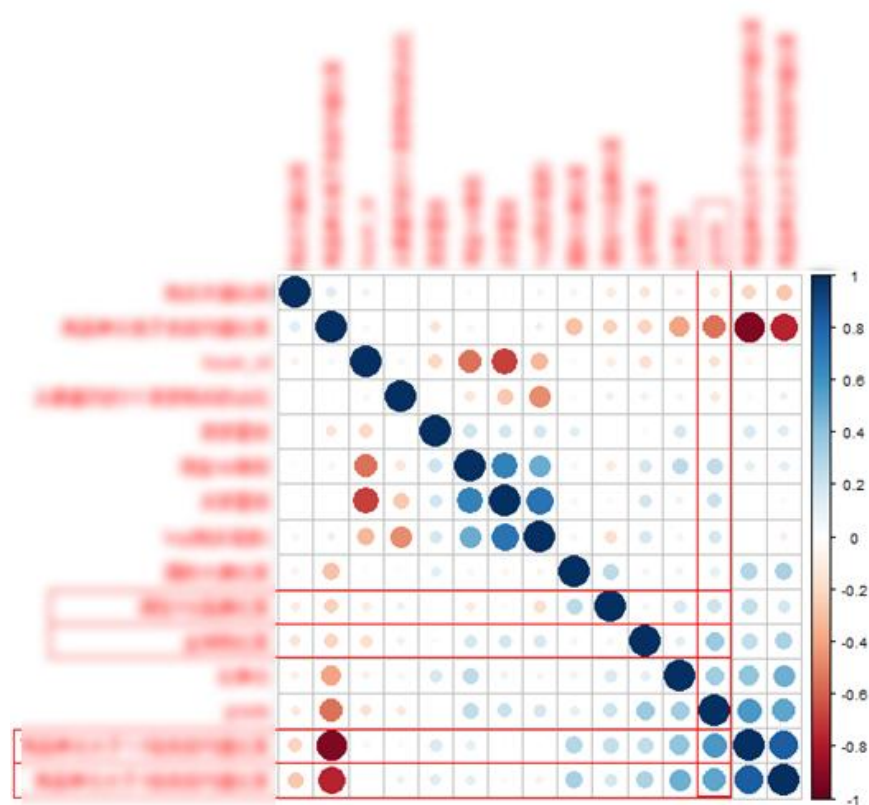
3分：高级白领，其他小土豪，生活品质较高，一年会购买几次奢侈品

2分：小白领，生活品质还不错，但是闲钱不多。

1分：收入很少，量入为出，喜欢买经济实惠的东西。



交易网络中的白富美挖掘—相关性分析



指标两两间的相关性：

```
library(corrplot)
corrplot(cor(dataFrameX),order="hclust")
```

天猫 TMALL.COM

纲要

- 淘宝业务和数据架构介绍
- 六度分隔理论
- 交易网络中的社会关系
- 交易网络中的 “白富美” 买家挖掘
 - 相关性分析
 - **数据训练和测试**
- 交易网络中的高品味商家挖掘



training and testing

将回归问题转化为分类问题再用AdaboostM1算法解决：

$\left\{ \begin{array}{l} \geq 3 \text{ 白富美用户（品质生活）} \\ < 3 \text{ 非白富美用户（小康生活）} \end{array} \right.$

```
library("RWeka")
formula<- Y~.
model_boost<-AdaBoostM1(formula,data=tbml)
evaluate_Weka_classifier(model_boost, numFolds=10,class=TRUE)
```

交叉验证得到：白富美用户(≥ 3 分)的分类准确率和查全率都在85%以上。

类似地：我们将 ≥ 2 分和 $=1$ 分的用户分为两类，准确率和查全率也较高。

≥ 4 分以上数据标注样本量较小，则暂不考虑。

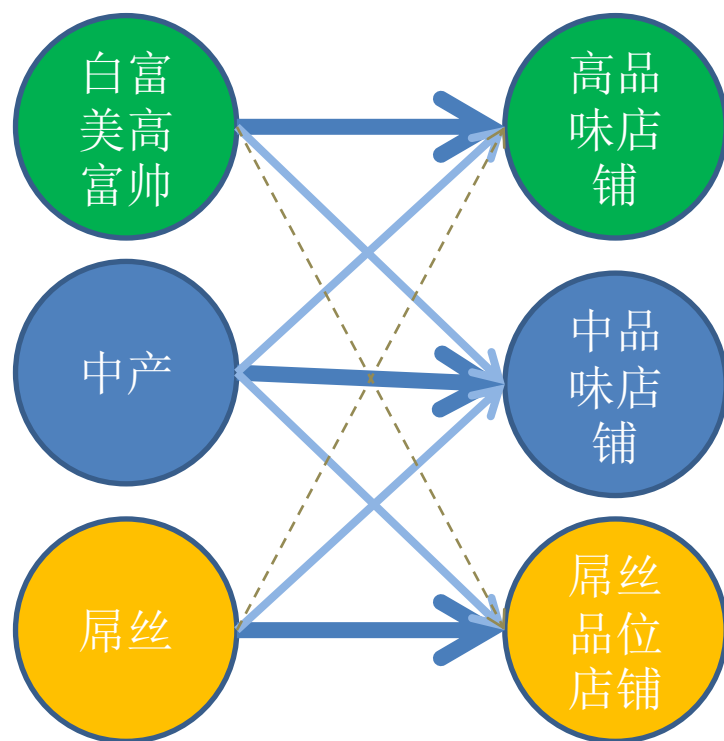
模型效果—效果明显

- 今日最大牌的用户构成：今日最大牌用户与模型匹配度高，并且越接近下单购买环节越明显

使用场景：2013/10/27 和 2013/10/30 两天，在淘宝首页四联通位置定向投放“今日最大牌”

- 投放响应：渠道响应率高端用户对淘宝首页的渠道偏爱不及非高端用户；进一步点击“今日最大牌”素材的比例无明显差异；但浏览转化率、购买转化率高端用户远高于非高端用户(购买转化率是非高端的2-3倍)

通过高端买家来发现高品位的店铺



欧美大牌



白富美



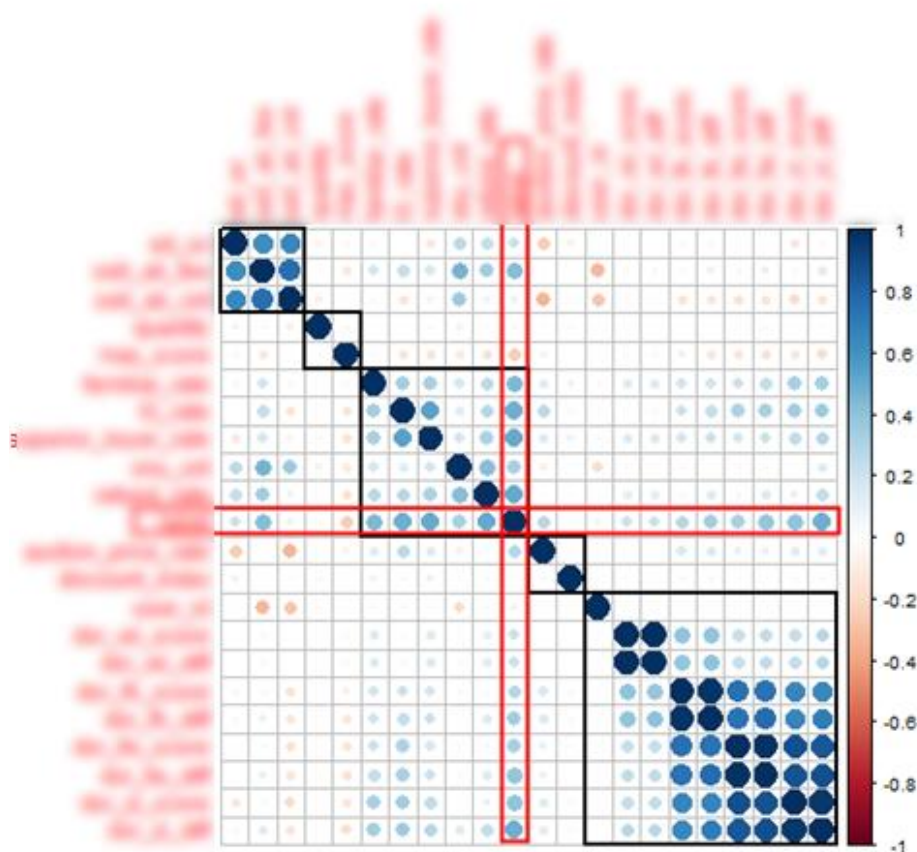
高品位店铺

纲要

- 淘宝业务和数据架构介绍
- 六度分隔理论
- 交易网络中的社会关系
- 交易网络中的“白富美”买家挖掘
- **交易网络中的高品味商家挖掘**



通过高端买家来发现高品位的店铺



重要相关因素有：

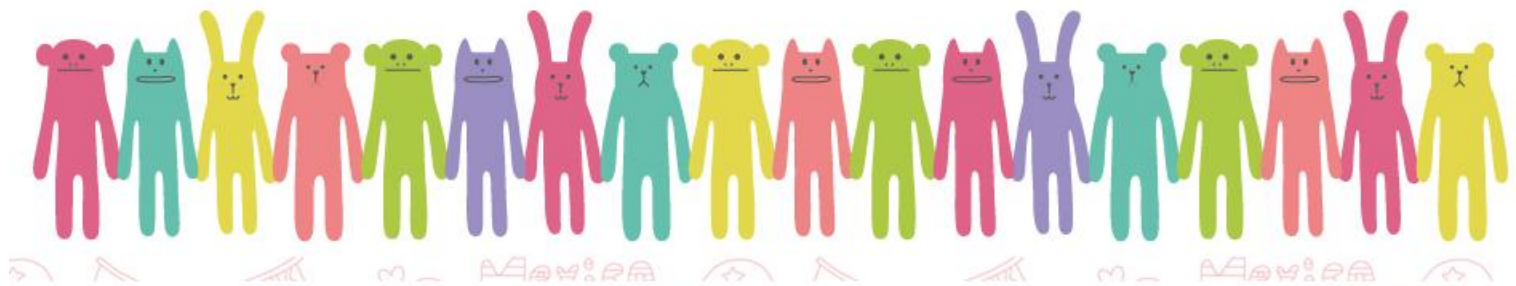
熟客率
买家中t3比例等

通过高端买家来发现高品位的店铺

```
library("RWeka")  
formula<- Y~.  
model_boost<-AdaBoostM1(formula,data=tbml)  
evaluate_Weka_classifier(model_boost, numFolds=10,class=TRUE)
```

Precision	Recall	Class
0.909	0.888	N
0.949	0.959	Y
0.937	0.937	

挖掘出的高品位店铺



CRAFTHOLIC®



挖掘出的高品位店铺



¥ 74000.00

烟波★日本玉川堂 手工锤起铜器 鑲流象嵌
纯铜茶壶 铜壶 铜水壶



¥ 35700.00

烟波★日本玉川堂 手工锤起铜器 葫芦 纯铜
茶壶 茶叶罐 茶具组



¥ 52800.00

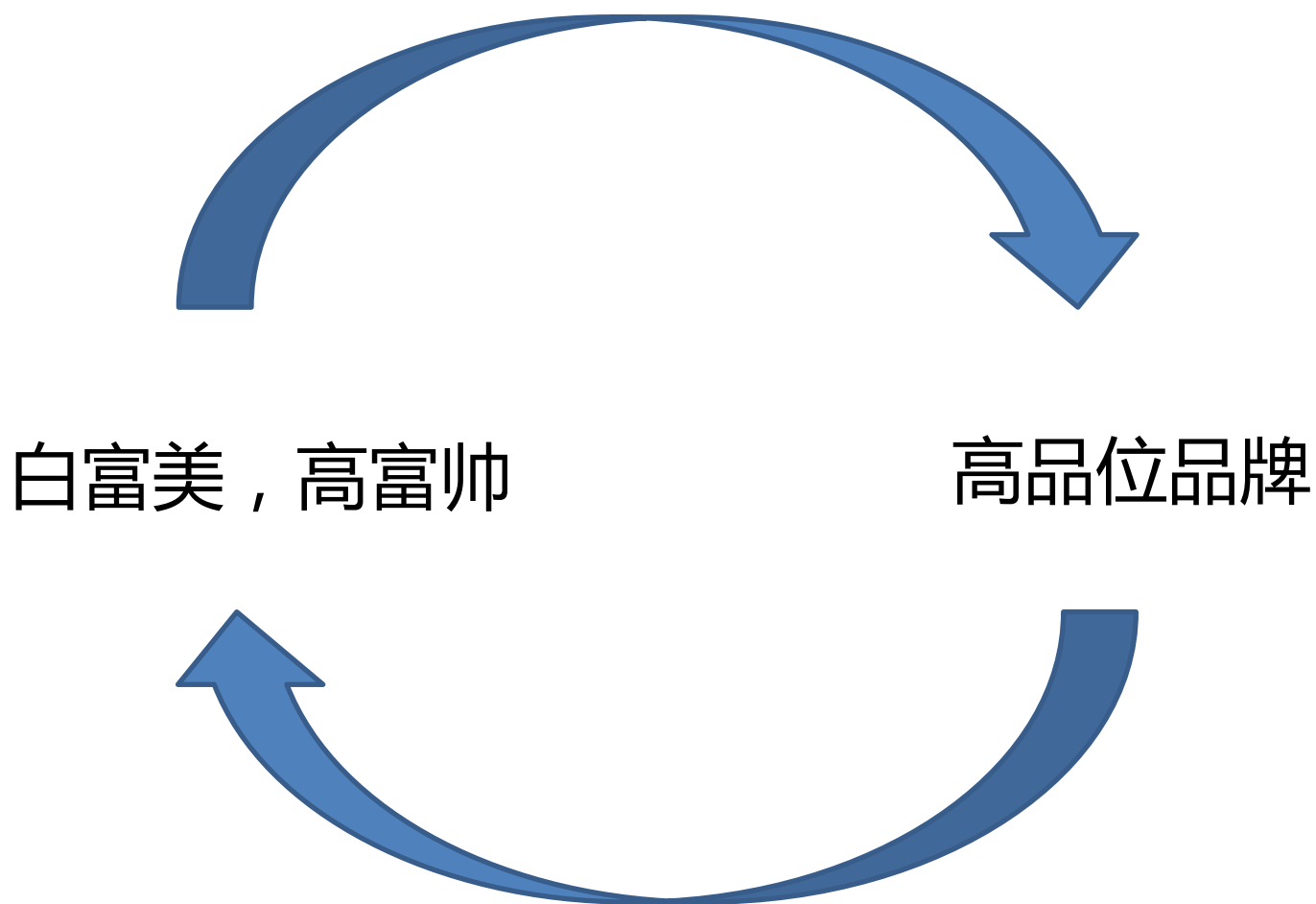
烟波★日本铁壶 云色堂 车轴 银摘 铜盖 铁
瓶 日本老铁壶 铁茶壶



¥ 22080.00

烟波★日本铁壶 清光堂 铜盖银摘 八角松竹
梅 日本老铁壶 铁茶壶

抛砖引玉—利用网络信息协同分层



We are Hiring

发送简历到andy.xul@tmall.com

- 工作年限：硕士3年以上，博士2年以上
- 岗位描述：
 - 爱思考，爱钻研，乐于创新。
 - 信仰并热爱大数据。
 - 数据技术扎实。
-
- 从大数据中挖掘用户本质属性，并分析用户行为和个性化需求。
- 应用先进的统计建模、数据挖掘、机器学习方法建立数据模型解决实际问题，并研发创新方法以解决常规算法不能解决的问题。
- 与业务部门沟通合作，将数据模型应用于实际业务。



We are Hiring

发送简历到andy.xul@tmall.com

岗位要求：（社招）

- 1) 硕士以上学历，博士优先。统计学、计算机科学、数学等相关专业。扎实的统计学、数据挖掘、机器学习理论基础。
- 2) 具有丰富的数据建模实践经验，2年以上相关工作经验。
- 3) 扎实的编程基础，精通至少一门编程语言。熟悉R语言优先。
- 4) 熟练掌握Hive\SQL。
- 5) 拥有海量数据处理经验者优先，熟悉Hadoop优先。
- 6) 具有良好的商业敏感度和优秀的数据分析技能。能够开发创新而实际的分析方法以解决复杂的商业问题。
- 7) 擅长与商业伙伴的交流沟通，具有优秀的报告讲解能力及沟通能力。
- 8) 能够积极创新，乐于面对挑战，负责敬业。
- 9) 优秀的团队合作精神；诚实，勤奋，严谨



We are Hiring



thanks !

- andy.xul@tmall.com
- 微博: 许亮_在路上
- 来往: 剑霸





天猫 TMALL.COM