



兴趣图谱 与社交图谱

2013. 4. 26 北京

个性化推荐的历史

1992 ~ 2002

电子商务

新闻组

分类浏览

2002 ~ 2012

web
2.0

SNS

广告

兴趣
网络

2012 ~

云计算

移动互联

网络融合



信息

工具/体验

交易

社交图谱



兴趣图谱





可推荐的产品



	用户数	条目数	稀疏性	多样性	时效性	反馈	推荐效果
图书	3,000,000	3,000,000	< 0.5%	高	低	慢	B
电影	5,000,000	100,000	1% ~ 5%	低	中	中	C
唱片	1,500,000	400,000	< 1%	低	低	中	C
小组	5,000,000	200,000	%1	中	中	中	B
人	5,000,000	5,000,000	< 0.5%	高	低	慢	D
文章	500,000	10,000,000	< 0.1%	高	高	快	C
单曲	5,000,000	1,000,000	5% ~ 10%	低	低	快	A
广告	30,000,000	50,000	1%	低	高	中	D



时效性

时效性



模型效果



	冷启动	可解释性	惊喜	时效性	鲁棒性
协同过滤	★	★		★	★ ★
图模型		★	★		★
矩阵分解	★	★	★	★	★
Topic Model	★	★ ★		★	
增强学习	★	★		★ ★	
决策树		★	★	★	★
Boosting	★	★	★	★	★

个性化推荐的收益

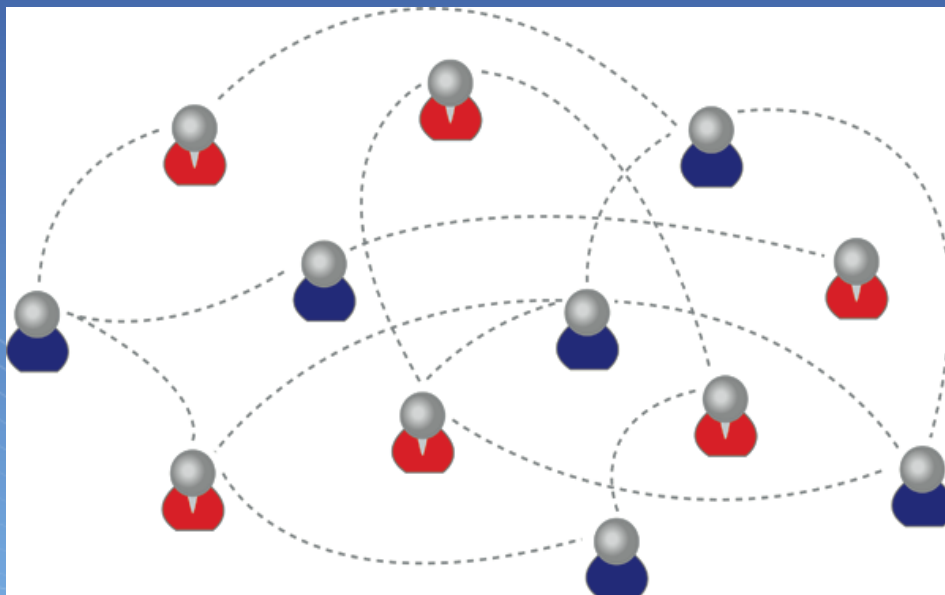
Volume

新用户转化率提升30%
用户粘性提升15%
用户活跃度提升10%





社交图谱



六度空间

150法则

小世界理论

社交图谱的多样性

人与人的关系

语义的复杂性

用人来描述人





兴趣图谱与社交图谱



	人群	粘性	频度	持久性
兴趣图谱	生人网络	中/低	中/低	高
社交图谱	熟人网络	高	高	高/中/低



利用社交网络改进服务



推荐准确率提高10%

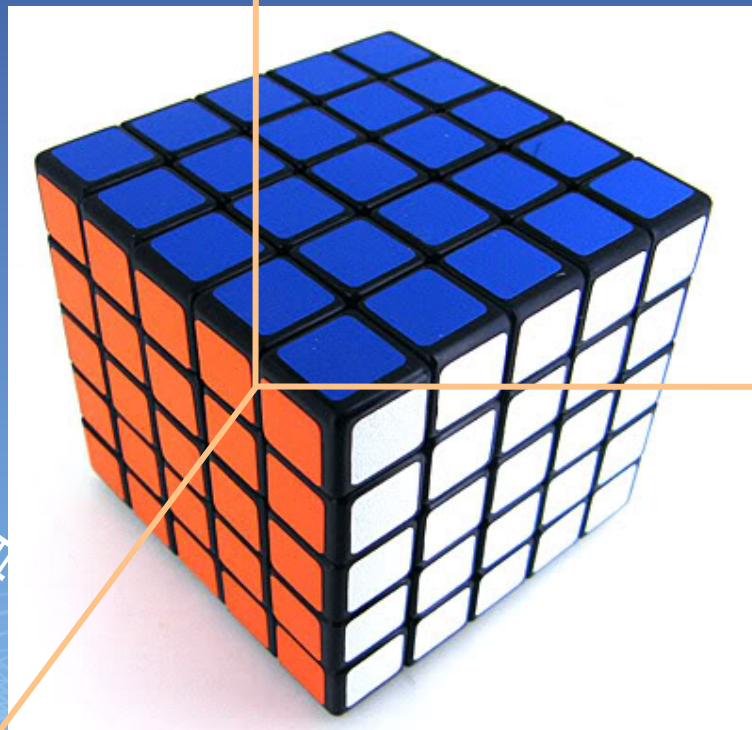
用户粘性增加20%

用户活跃度增加20%

大数据挑战

用户

亿 X 百 X 千



兴趣图谱

社交图谱

大数据挑战

CPU Bound or I/O Bound?



网络的本质是关联

关联意味着
数据规模呈平方量级增长



以迭代为核心的学习算法

对分布式计算和存储提出新要求



下一代推荐系统

Prediction

Forecasting

Recommendation



个性化推荐

前所未有的机会

web 2.0, 云计算, 成熟的技术准备
要么是平台, 要么是平台的一部分



个性化推荐，下一件大事？



Q & A
谢谢