

# 理论、数据与预测模型 的三角对话

罗家德

清华大学社会学系与  
公共管理学院合聘教授

# 计算社会科学新方法论

数据挖掘

提供议题方向  
提供扎根真相

进行推论  
预测新“事实”

理论发展  
定性访谈  
定量调查

建立预测模型  
(网络动态模型)

指导建立模型

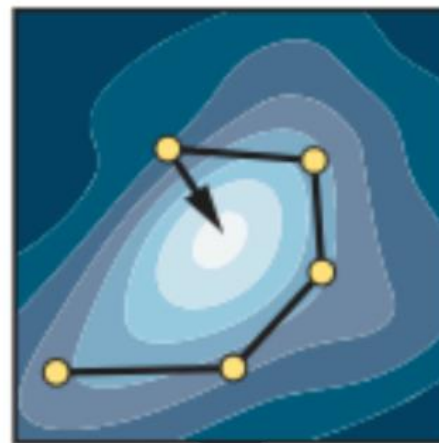
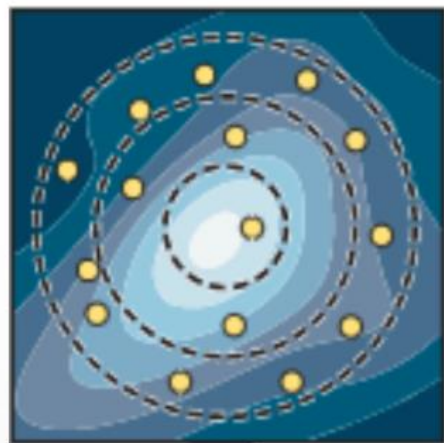
# 太阳底下无新鲜之事

- 上世纪八十，九十年代一系列统计学顶尖的社会统计学者，如Donald Rubin, Cliff Clogg, Smith Herbert, Paul Holland以及他们的统计学盟友 David Freeman等人发起的讨论「统计如何才能得到因果推论」，以反对那时理论粗糙却大搞结构化资料库data mining的研究，比如path model。在因果模型与统计分析不断对话下，终于建立了今天我们常见的「主流定量论文范式」——明确地社会科学议题，穷尽的相关理论回顾，逻辑完整的理论推演得到的假设，导出因果模型，好的研究设计，在时间或「空间」上找出中介或干扰变量，因此有了清楚的模型指定，包括变量、测量方法、因果机制、因果函数型式、资料来源等等。时隔三十几年一模一样的历史又发生在大数据的data mining身上。

# 三角对话过程

- 一、大数据内容分析-----x---扎根真相（ground truth）
- 二、数据挖掘-----x---社会调查
- 三、依据理论计算指标并建立弱模型（weak model）-----x---实地资料验证
- 四、往复交叉验证
- 五、模型确定与推论→逐渐完善模型，做理论的演绎和推理，从而作预测
- 六、在理论指导下进行应用

# 从随机抽样到最适化抽样



Random  
Samples



Optimized  
Samples

Experimental Design

as

Algorithm Design

## 3.1 问题的提出以及研究设计

- 本研究的主要问题是计算**关系强度**，刻画中国人**人脉圈**可分为几层。
- 从西方**邓巴圈的理论**以及中国本土**人情、关系和面子理论**出发，对中国人人脉圈层进行猜想
- **关系强度**是社会关系网络中的重要指标，不同的关系强度往往会产生不同的互动逻辑，从而影响关系网络的信息和资源的流动。同时关系强度对于研究个体寻找机会和优势、计算社会资本的重要指标。格兰诺维特将关系强度划分为强关系和弱关系，但是并没有一个具体的数学上的计算方法和模型为强弱关系的划分提供一个清晰的界限。

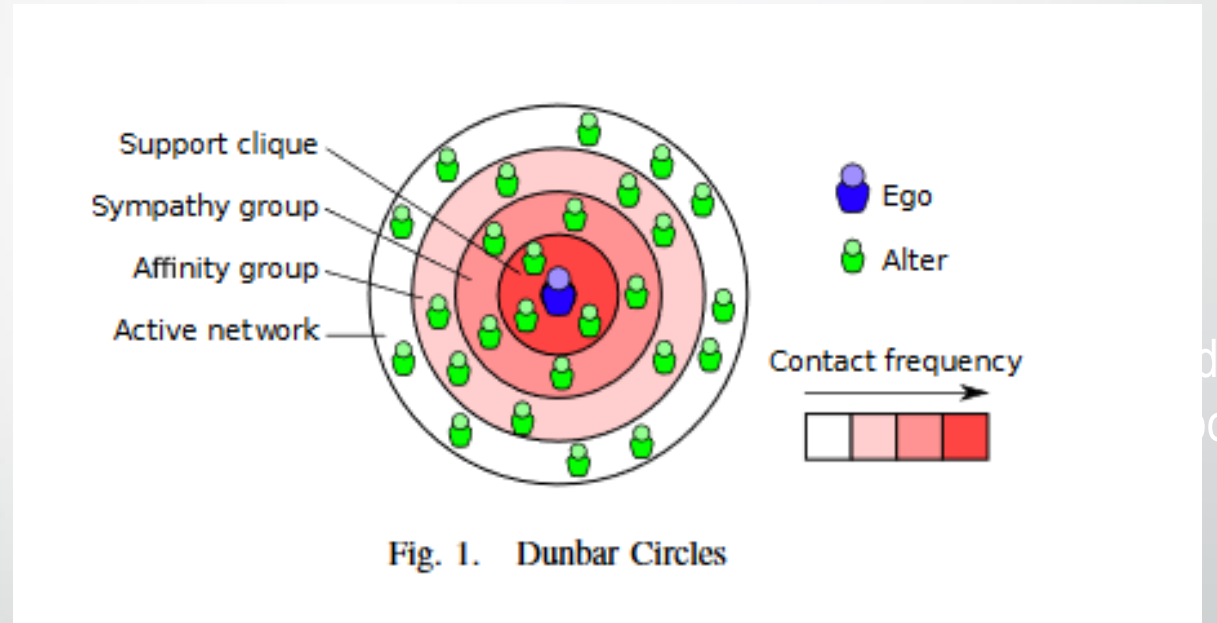
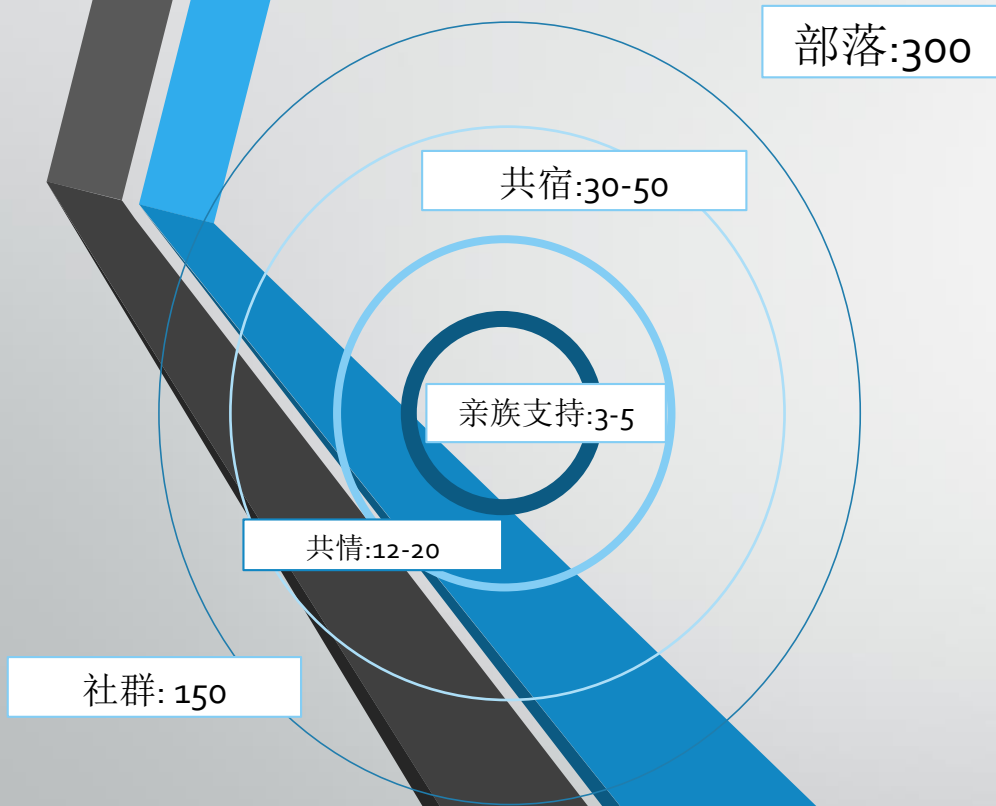
# 邓巴圈的定义

邓巴圈（Dunbar circle）即指存在于社会网络中的每一个行动者所能够直接连接到的其他人形成的圈子（active group），或者说人与人之间因为直接的互动而形成的网络（Durban,1993; Hill & Durban,2003）。

邓巴数：5-15-50-150-300

- 最内层（5）暗示了一个**亲族支持群体**（support clique），它内部的成员往往是网络核心（ego）的姻血亲成员，或者能提供核心（ego）直接的意见和物质帮助的成员
- 第二层（15）的成员数目与社会心理学中的**共情群体成员数**（sympathy group）暗合（Buys,Larson,1979）。这一层的成员将提供给核心情感上的纾解（Dunbar,Spoors,1995; Sutcliffe,Dunbar,2012）；
- 第三层（50）与人类学意义上的**共宿群体**（overnight camp group）在数量关系上相关，意指由多个共情群体相结合，一组保证成员安全且具有分工的群体（affinity group），因此他们也具有格兰诺维特意义上的信息交互作用。
- 第四层（150）意指一个**社群或族系**（community or clan），是核心（ego）所直接认知的社会群体。
- 第五层（300）意指**部落群体**（Tribe），是指可以被共同文化或非成文的规制约束的群体。

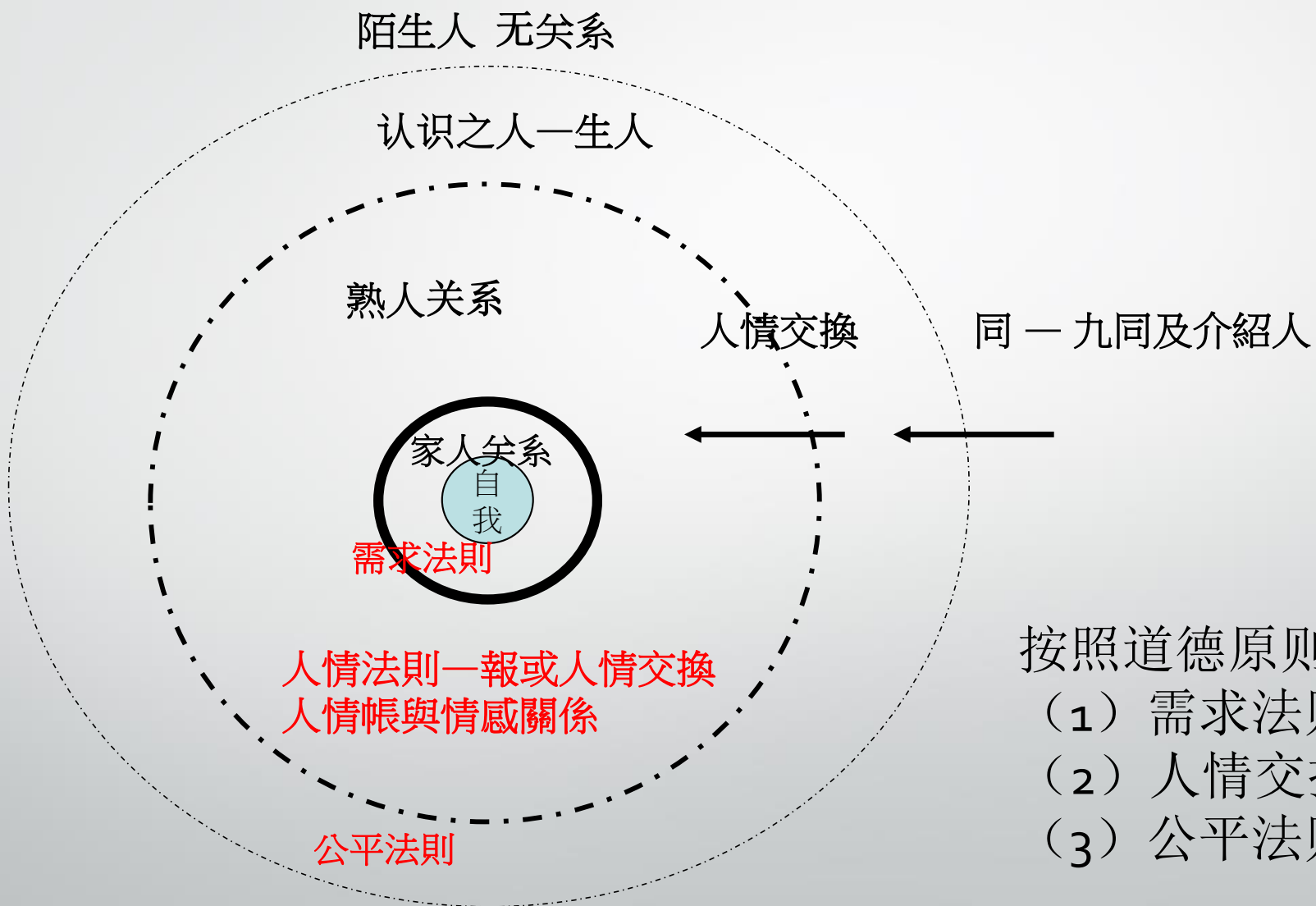
# 邓巴圈的定义



Arnaboldi V, Conti M, Passarella A, et al. Analysis of ego network structure in online social networks[C]//Privacy, security, risk and trust (PASSAT), 2012 international conference on and 2012 international confernece on social computing (SocialCom). IEEE, 2012: 31-40.



# 中国人的人脉圈：人情、关系、面子理论



按照道德原则进行划分：

- (1) 需求法则
- (2) 人情交换法则
- (3) 公平法则三层

## 3.2 理论指导的扎根真相——问卷调查

由于邓巴圈的定义过于学术，同时涉及过多人类学和进化心理学的内容，为了研究目的，我们基于中国本土理论重新阐释了邓巴圈：

圈层/关系远近	问卷描述
家人、拟似家人	在生活中，我们身边会有一些与我们最亲密的，常常是家人或者被我们看做是家人的极为亲密的好友，大多在5人以内
亲密熟人	还会有一些人，我们认为他们是我们的铁哥们或闺蜜，或者经常联系的亲戚，但不如家人与我们那样亲密
一般熟人	也会存在一些人，我们愿与其长期交往，可能相互欠人情，相互帮忙的关系
潜在熟人	还会有一些人，我们尽管认识他们，现在会联系，但不一定与其长期交往，也并不会互相欠人情
认识之人	我们存过他们的联系方式，但也不打算与其长期交往，以后也几乎不会联系

# 收集扎根真相 (label)

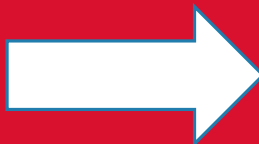
经过问卷调查，收集关系强度的Ground Truth，开始对邓巴的理论进行验证。

	次序	时间	地点	样本量	目的
针对社交软件A	第1次调查	2017. 04. 19- 2017. 05. 21	上海市; 北京市	96名用户的1749条 Tie	作为基础标识数据用于数据挖掘和训练模型
	第2次调查	2017. 10. 01- 2017. 10. 12	北京市	29名用户的580条 Tie	作为验证数据集，主要收集A软件活跃用户信息，并用以防止过拟合的问题
	第3次调查	2017. 11. 05- 2017. 11. 16	北京市	37名用户的736条 Tie	针对最外圈分类准确率极低的情况进行样本补充，同时利用定性手段寻找有效的特征。
针对社交软件B	第4次调查	2018. 8. 1- 2018. 8. 6	北京市; 上海市; 广州市	66名用户1654条tie	作为社交软件B研究的主要的训练和测试集

### 3.3 计算大数据指标

#### 格兰诺维特 强关系 弱关系理论

关系久暂  
联系频率  
互惠程度  
亲密程度



指标类型	变量含义
点对点互动指标	非工作时间互动标准差
	工作时间互动标准差
	工作时间点对点发消息总频率
	非工作时间点对点发消息总频率
	点对点发消息总频率
	点对点互动时间标准差
	点对点发消息相对频率
	点对点电话、视频相对频率
	点对点电话、视频频率
点对点红包、转账频率	
共同群互动	共同群数
	共同群互动频率
共同邻居	点对点共同邻居数
成为好友时间	点对点成为好友时间
结为好友次数	点对点结为好友次数
见面频次	同出现在一个城市的一个区的频次
相似性指标	年龄相似性
	性别相似性

## 3.4 建立人脉圈分类模型

- 1、**重复邓巴实验**：将这些关系链视为无标签的tie，对其直接进行聚类
- 2、**建立有监督分类弱模型**：根据收集的标签建立有监督的人脉分类弱模型
- 3、**尝试建立四层分类模型**：根据实际数据以及理论猜想，除了3、5层模型，建立4层模型
- 4、**校准模型**：根据分类结果，建立预测值与实际值分类矩阵，对模型进行校准

准确率的计算方式是聚类后，针对特定的类簇，我们将该类簇中的标签最多的设为该类的标签

1. 邓巴的方法（点对点发消息相对频率）进行K-Means聚类

模型	类别	聚类指标	准确率
<b>K-Means</b>	<b>5</b>	<b>点对点相对发消息频率</b>	<b>0.329</b>

2. 多个指标（发消息频率、共同好友、共同群等综合征指标）进行K-Means聚类

测试集比例	模型	类别	训练集准确率	测试集准确率	每一类准确率
<b>0.1</b>	<b>K-Means</b>	<b>5</b>	<b>0.339</b>	<b>0.388</b>	0.000_0.258_0.200_0.000_0.000

结论：多个指标进行K-Means聚类，准确率的提高说明其他指标对于用户关系强度同样存在预测能力。

3. 基于理论建立人脉圈分类模型

测试集比例	模型	类别	训练集准确率	测试集准确率	每一类准确率
<b>0.1</b>	<b>SVM</b>	<b>5</b>	<b>0.475</b>	<b>0.424</b>	0.000_0.367_0.357_0.375_0.889

结论：基于理论指导的有监督的分类算法，准确率有了一定提高，可见基于理论指导的人脉圈的分类模型更具有推广意义。

## 3.4.2 建立有监督分类弱模型并进行修正

### 建立有监督分类模型

SVM  
Logistic  
决策树  
随机森林  
GBDT

几次修正的准确率对比

	5层模型	4层模型	3层模型
随机分类模型	0.2	0.25	0.33
初步模型	0.4988	0.5112	0.6832
第一次修正	0.4579	0.6838	0.6900
第二次修正	0.4957	0.7058	0.7016
第三次修正	0.5	0.7327	0.7207
<b>最新一轮修正</b>	<b>0.543</b>	<b>0.7748</b>	<b>0.8079</b>

4_1	第4、5层一类
4_2	第3、4层一类,
4_3	第2、3层一类
4_4	第1、2层一类
3_1	第4、5层一类, 第3、4层一类
3_2	第1、2层一类, 第4、5层一类
3_3	第3、4、5层一类
3_4	第2、3、4层一类
3_5	第1、2层一类, 第3、4层一类
3_6	第1、2、3层一类

其中:

4层模型准确率最高: 家人、拟似家人 | 亲密熟人 | 一般熟人 | 潜在熟人和认识之人  
3层模型准确率最高: 家人、拟似家人 | 亲密熟人 | 一般熟人、潜在熟人和认识之人

# 模型修正中理论与数据的对话

5层模型实际值与预测值的分布矩阵

真实值 \ 预测值	最外层	第四层	第三层	第二层	最内层
最外层	122	1	13	2	2
第四层	30	36	4	1	0
第三层	44	6	13	6	3
第二层	29	7	13	17	7
最内层	25	2	6	3	10



建立分布矩阵，看分错误样本的特征，进行理论与数据的对话。

对于tie两端的活跃用户，模型会更精准  
住在一起和不住在一起的家人具有大的差异，加入位置信息对于准确率提高有较大帮助



## 总结

经过几轮修正，可以得到如下结论：

(1) 3、4、5层模型相对于随机分类模型准确率都有很大的提升，这说明经过这样不断的三角对话，模型越来越接近最优值；

(2) 从最后准确率的对比上看，4层模型的准确率最高，因此4层模型在本研究的情境下最接近最优值，对于扎根真相的解释力最强；

(3) 在修正过程中4层模型的准确率相对于随机分类模型提升最为显著（ $52.48\%$ (4层)  $>$   $50.44\%$  (3层)  $>$   $34.3\%$ (5层)），4层模型是三者中最优的模型，4层模型在此情境下最具解释力。

随着预测精度越来越高，会不会暂时的结论有所改变？这有待更多轮的三角对话。

# 案例、风险投资产业网的动态变化模型

## 1. Big Data on Internet

- [http://i.ifeng.com/news/news?aid=32901285&vt=5&ch=rj\\_bd\\_me&mid=](http://i.ifeng.com/news/news?aid=32901285&vt=5&ch=rj_bd_me&mid=)
- **DCM联合IDG向B2C网站优雅100投资1000万美元**
- 2012-03-12 08:46 中国风险投资网
- 【导读】家纺B2C电子商务网站文雅100创始人陈腾华称，其曾经完成第一轮1000万美圆融资，由IDG和DCM结合停止投资。他以为家纺业B2C年销售额应当在将来两三年内会有大幅增长，也希望文雅100做到家纺B2C的第一。
- 家纺B2C电子商务网站文雅100创始人陈腾华称，其曾经完成第一轮1000万美圆融资，由IDG和DCM结合停止投资。  
陈腾华表示，本轮资金将主要用户招募员工、做品类优化以及产品开发方面。  
“文雅100”( [uya100.com](http://uya100.com) )定位于在线的时髦家居生活馆，目前仅仅上线2个月左右，主要销售自有品牌家纺类产品以及其他品牌类的床品、毛巾浴巾等家纺用品。  
谈及为何离任并进入家纺B2C时，陈腾华表示这个市场范围每年可以到达7000-8000亿元，目前B2C行业还没有质量家居生活的领头者。在传统市场又存在制造分散、品牌集中度分散等问题，因而B2C在这方面就存在时机。  
值得注意的是，除了作为平台出卖其他品牌家纺，文雅100还推出了自有品牌Toscaso。陈腾华以为，垂直类B2C的一个开展趋向就是自有品牌，由于无论是制造、本钱还是质量和渠道都可以有效控制，并且具有较高利润。  
陈腾华表示固然有自在品牌，但不会做线下实体展现店。他以为家纺业B2C年销售额应当在将来两三年内会有大幅增长，也希望文雅100做到家纺B2C的第一。

## 2. Structure Data Base Collected from Big Data

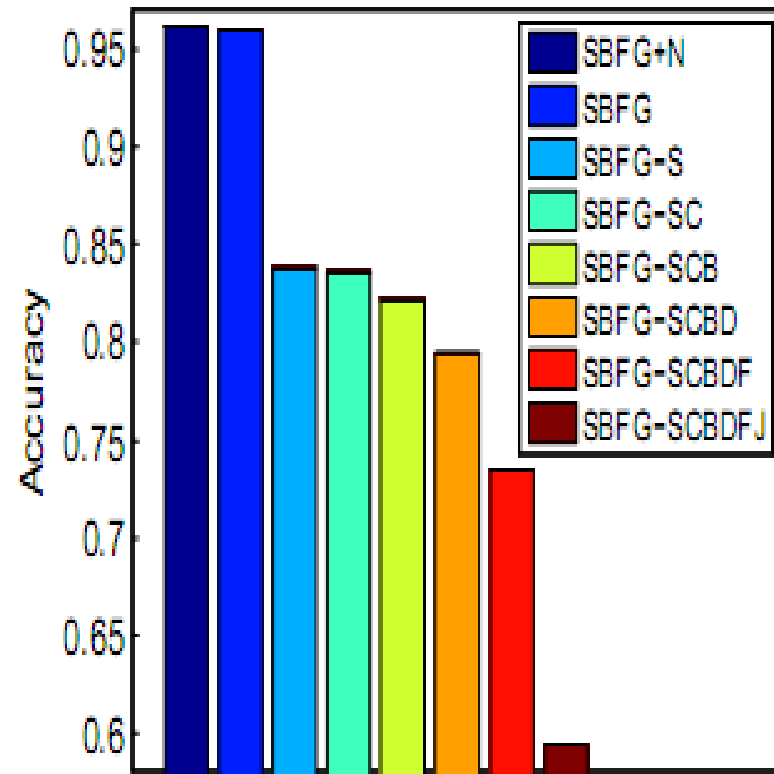
投资机构	投资机构	获投企业	获投企业	行业代码	行业描述	国家	省份	城市
IDG资本	123	慧聪国际	1674	H	批发和B2B	中国	北京	北京市
IDG资本	123	慧聪国际	1674	H	批发和B2B	中国	北京	北京市
IDG资本	123	慧聪国际	1674	H	批发和B2B	中国	北京	北京市
达晨创投	385	金银岛网	1916	H	批发和B2B	中国	北京	北京市
德丰杰	406	点视传媒	744	H	批发和B2B	中国	上海	上海市
富达亚洲	483	阿里巴巴	137	H	批发和B2B	中国	浙江	杭州市
富达亚洲	483	阿里巴巴	137	H	批发和B2B	中国	浙江	杭州市
富达亚洲	483	阿里巴巴	137	H	批发和B2B	中国	浙江	杭州市
高盛	500	阿里巴巴	137	H	批发和B2B	中国	浙江	杭州市
和利资本	616	卖买提	2560	H	批发和B2B	中国	上海	上海市
红杉	652	点视传媒	744	H	批发和B2B	中国	上海	上海市

# I. 资料挖掘：预测变量

Feature	No.	Ord.*	Name	Short description
Static feature	01	43	latitudeMax	Larger value of latitude
	02	34	latitudeMin	Smaller value of latitude
	03	26	latitudeSingle	Single value of latitude
	04	33	latitudeDiff	Difference of latitude
	05	43	longitudeMax	Larger value of longitude
	06	31	longitudeMin	Smaller value of longitude
	07	16	longitudeSingle	Single value of longitude
	08	23	longitudeDiff	Difference of longitude
	09	20	absoluteDistance	Straight-line distance
	10	17	timeZoneDiff	Difference of time zone
	11	80	sameCity	Are two VCs in the same city?
	12	01	sameCountry	Are two VCs in the same country?
	13	14	ethnicitySim	Ethnicity similarity**
	14	39	languageSim	Language similarity**
	15	29	religionSim	Religion similarity**
	16	06	investorCombination	Combination of investor type
	17	36	sameCVCField	Are two company VCs of the same field?
Dynamic domain feature	18	08	fieldsMax	Larger value of #field
	19	07	fieldsMin	Smaller value of #field
	20	09	fieldsSingle	Single value of #field
	21	61	fieldsDiff	Difference of #field
	22	77	fieldsSum	Sum of #field
	23	10	fieldsJaccard	Jaccard similarity of fields
	24	23	shortTrendMax	Larger value of short trend***
	25	13	shortTrendMin	Smaller value of short trend
	26	18	shortTrendSingle	Single value of short trend
	27	23	longTrendMax	Larger value of long trend
	28	39	longTrendMin	Smaller value of long trend
	29	12	longTrendSingle	Single value of long trend
	30	36	firstInvestYearMax	Larger value of first year of investment
	31	20	firstInvestYearMin	Smaller value of first year of investment
	32	81	firstInvestYearSingle	Single value of first year of investment
	33	61	firstInvestYearDiff	Difference of first year of investment

Dynamic  
topology  
feature


34	04	distanceBefore	Shortest distance of two VCs
35	39	degreeMax	Larger value of degree
36	52	degreeMin	Smaller value of degree
37	20	degreeSingle	Single value of degree
38	47	degreeDiff	Difference of degree
39	69	degreeSum	Sum of degree
40	36	shConstraintMax	Larger value of structural hole constraint****
41	56	shConstraintMin	Smaller value of structural hole constraint
42	34	shConstraintSingle	Single value of structural hole constraint
43	73	shConstraintDiff	Difference of structural hole constraint
44	67	shConstraintSum	Sum of structural hole constraint
45	55	shConstraintMaxEgo	Larger value of structural hole constraint of ego net
46	52	shConstraintMinEgo	Smaller value of structural hole constraint of ego net
47	43	shConstraintSingleEgo	Single value of structural hole constraint of ego net
48	71	shConstraintDiffEgo	Difference of structural hole constraint of ego net
49	73	shConstraintSumEgo	Sum of structural hole constraint of ego net
50	03	betweennessMax	Larger value of betweenness
51	11	betweennessMin	Smaller value of betweenness
52	05	betweennessSingle	Single value of betweenness
53	76	betweennessDiff	Difference of betweenness
54	42	betweennessSum	Sum of betweenness
55	49	betweennessMaxEgo	Larger value of betweenness of ego net
56	19	betweennessMinEgo	Smaller value of betweenness of ego net
57	26	betweennessSingleEgo	Single value of betweenness of ego net
58	65	betweennessDiffEgo	Difference of betweenness of ego net
59	67	betweennessSumEgo	Sum of betweenness of ego net
60	30	densityMaxEgo	Larger value of ego density
61	56	densityMinEgo	Smaller value of ego density
62	43	densitySingleEgo	Single value of ego density
63	78	densityDiffEgo	Difference of ego density
64	61	densitySumEgo	Sum of ego density
65	49	firstNeighborsMax	Larger value of #neighbor
66	56	firstNeighborsMin	Smaller value of #neighbor
67	26	firstNeighborsSingle	Single value of #neighbor
68	47	firstNeighborsDiff	Difference of #neighbor
69	72	firstNeighborsSum	Sum of #neighbor
70	02	firstCommonNeighbors	#common neighbor
71	65	secondNeighborsMax	Larger value of #secondary neighbor
72	52	secondNeighborsMin	Smaller value of #secondary neighbor
73	15	secondNeighborsSingle	Single value of #secondary neighbor
74	69	secondNeighborsDiff	Difference of #secondary neighbor
75	56	secondNeighborsSum	Sum of #secondary neighbor
76	31	secondCommonNeighbors	#common secondary neighbor
77	51	clusterCoefficientMax	Larger value of clustering coefficient
78	73	clusterCoefficientMin	Smaller value of clustering coefficient
79	56	clusterCoefficientSingle	Single value of clustering coefficient
80	79	clusterCoefficientDiff	Difference of clustering coefficient
81	61	clusterCoefficientSum	Sum of clustering coefficient



## 十大预测变量

- 相同国别
- 共同邻居数
- 中介中心性
- 距离
- 相同产权
- 投资领域数量
- 相同的投资领域数量

Fig. 18. Feature contribution analysis for the case of 2014. SBFG stands for the proposed method with the top 10 features. The plus mark denotes additional features besides the top 10 features, and the minus mark denotes features that are excluded from the top 10 features.



# 加入理论的动态模型

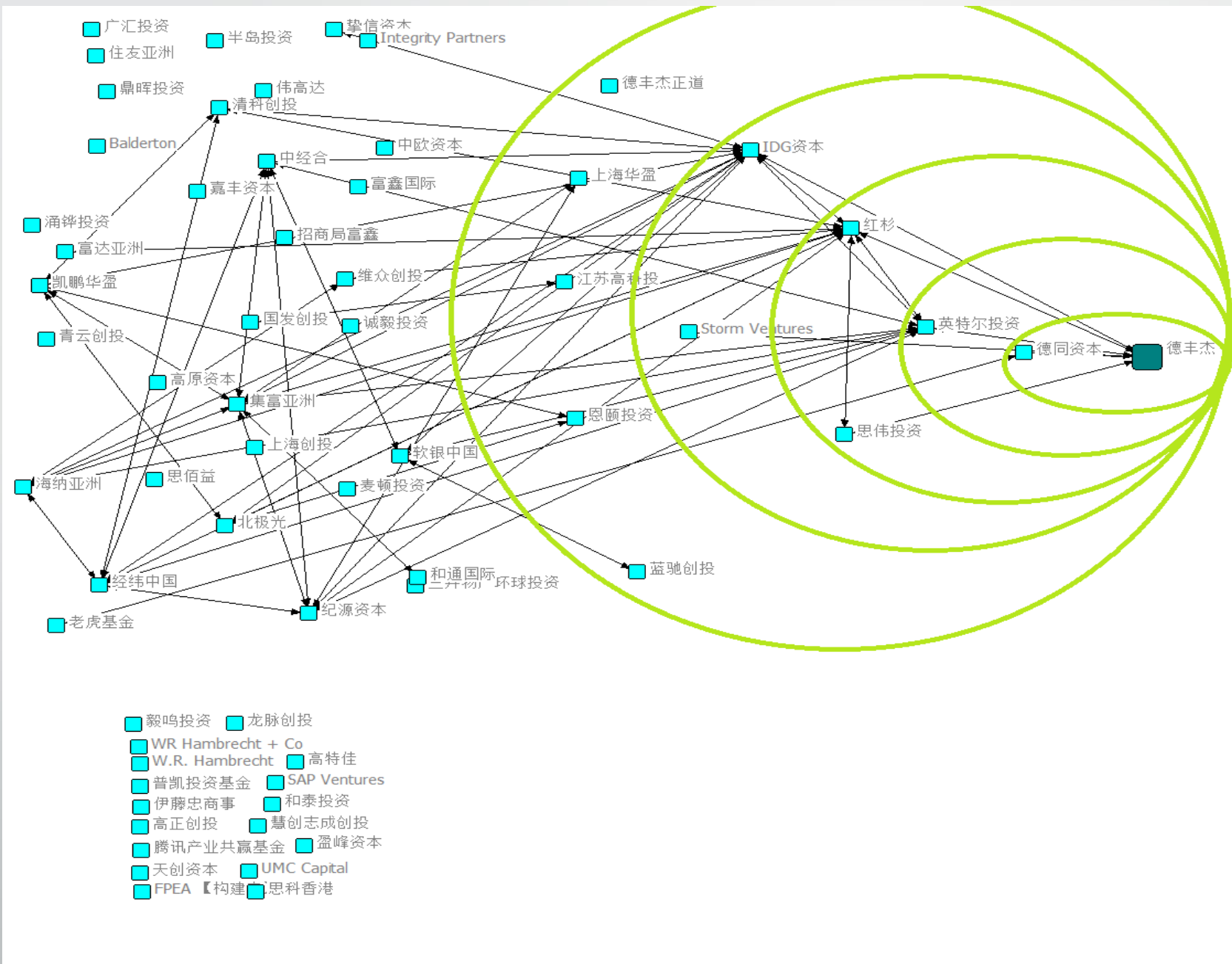
# 为什么要抱团？

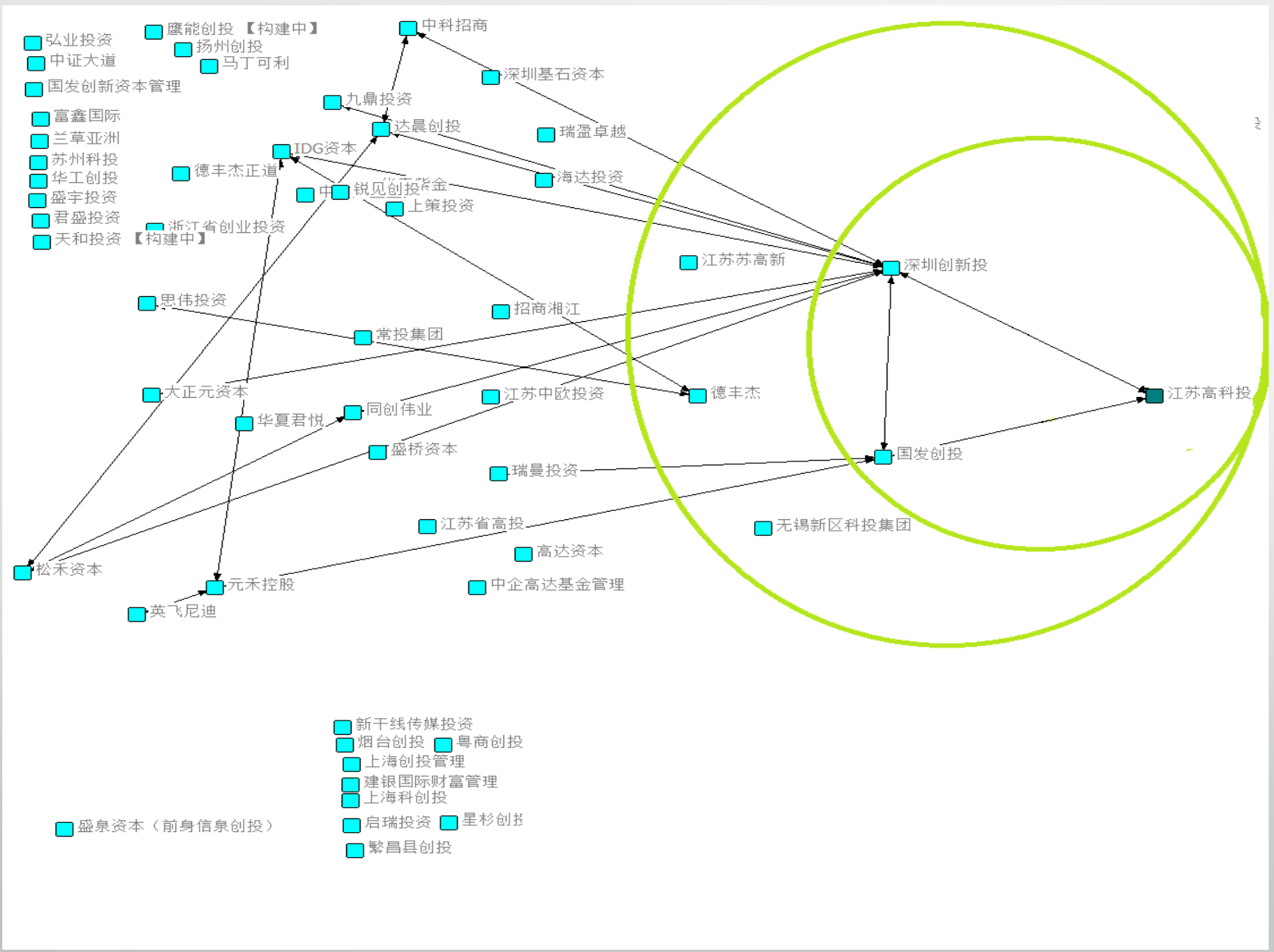
- “关系很重要”；“拿项目纯粹就在拼关系”；“都是靠关系相互介绍的”……
- “其实（联合投资）之前见过几次。都是跑去看（别人介绍的）项目遇到的，但是没说过话，不熟。后来有一次饭局，遇到了。（有共同的朋友）一聊，觉得可以（合作），就一起投了。”
- “关系很重要！初期联系企业与投资……（项目失败了）没关系，关系在那里了……（成功是一个）理智集体过程，……很多（事）在社交（关系）上事半功倍。”



## 大哥带小弟

- 在具体的投资实践中，常常出现“大哥带小弟”的现象。所谓大哥带小弟，指的是，当与政府关系强、拿资源多的投资机构进行投资时，小弟用钱来买关系，用钱来打入圈子中。一位国有投资机构的负责人L介绍这种大哥带小弟的逻辑：
- “大哥和小弟的区别嘛，大哥有声誉，拿投资业绩来支撑在圈内的声誉……大哥喜欢小弟因为小弟有钱，而且常常能够成为炮灰，而小弟喜欢大哥是可以通过融（拿钱）来拿声誉，一合作，就可以直接把大哥投资好的声誉拿走了……而且小弟可以跟着大哥去圈，再慢慢爬到圈子中心，就可能自己成为大哥，比如Q公司的老总就是自己搞了活动，忽悠了好多大佬，他们天天混在一块儿，后来他就自己拿着一点钱，跟着大佬们投资项目了，积累了投资的好业绩，然后不知不觉也变成一个大佬了……好多项目你给别人一说，都是好业绩啊，但是其实可能你只是一个跟投，一个小弟，不过市场不管你这些，看到你那些（好）业绩，会慢慢认可你……小弟这样（跟着大哥）在圈子里爬得更快。”



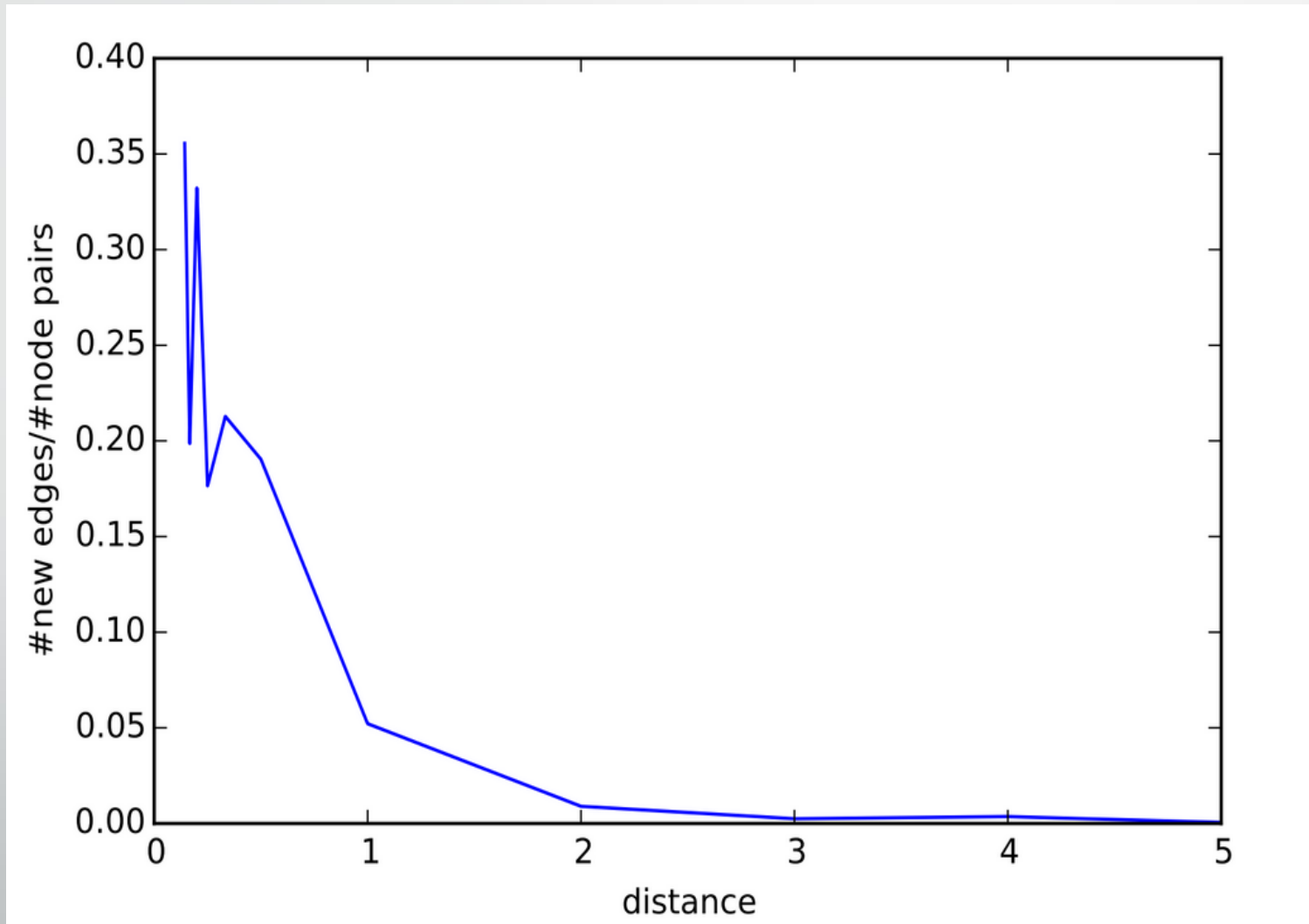


# 基于圈子理论建立的假设

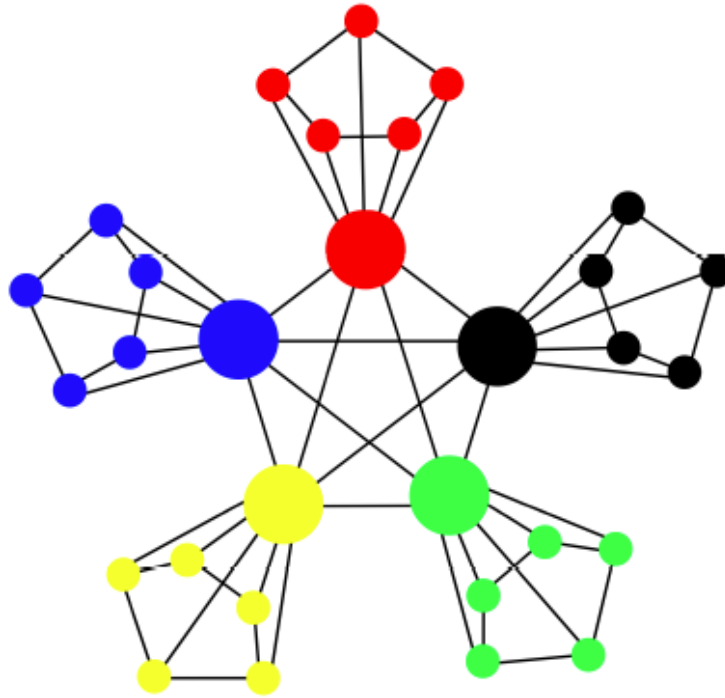
- 关系嵌入性理论
- Hypothesis 1: 过去多次合作增加了未来联合投资的可能性。
- 结构嵌入性理论
- Hypothesis 2: 关系距离减低了未来联合投资的可能性。
- Hypothesis 3: 没有共同朋友降低了联合投资的可能性
- 结构位置理论
- Hypothesis 4: 新加入者增加了联合投资的可能性
- Hypothesis 5: 风投公司的中心性高增加联合投资的可能性。

	模型一	模型二	模型三	模型四	模型五
联合投资次数		0.42***			0.27***
关系距离			2.65***		1.87***
有共同朋友				0.57***	0.39***
度中心性优势	0.01***	0.01***	0.01***	0.01***	0.01***
纳新趋势	-0.1	-0.06	-0.1	-0.02	-0.03
行业相似度	2.17***	1.95***	1.78***	1.76***	1.47***
常数项	-0.27	-0.24	-0.29	-0.3	-0.28
ADJUSTED R <sup>2</sup>	0.036	0.049	0.053	0.04	0.06

# Probability of Syndication and Distance



# Small world with an elite-clique



*Figure 1.* A toy example illustrating the small-world with an elite-clique property of the VC network. There are some major players (the nodes located in the center with a bigger node sizes.), each of whom leads a small group of followers (nodes of the same color but smaller node size), and they often play the bridging roles to connect some of the small cliques.

# 网络动态建模

- 两类点：投资者类与被投资者类
- 初始状态下75个投资者（2000年资料）
- 初始状态下有375个被投资者
- 每个投资者都有两个属性：投资力度F、联合倾向C
- 所有投资者类点的投资力度分成三类，以十四年来平均值设定参数：低投资倾向（例如每轮有0.25的概率会进行投资）、中投资倾向（例如每轮会进行0.8次投资）、高投资倾向（例如每轮会进行5次投资），三类点的数量各占三分之一。
- 所有投资者类点的联合倾向分为三类：低联合倾向者每笔投资有0.3个联合投资伙伴、中等联合倾向有0.6个联合投资伙伴、高联合倾向有1个联合投资伙伴，），三类点的数量各占三分之一。

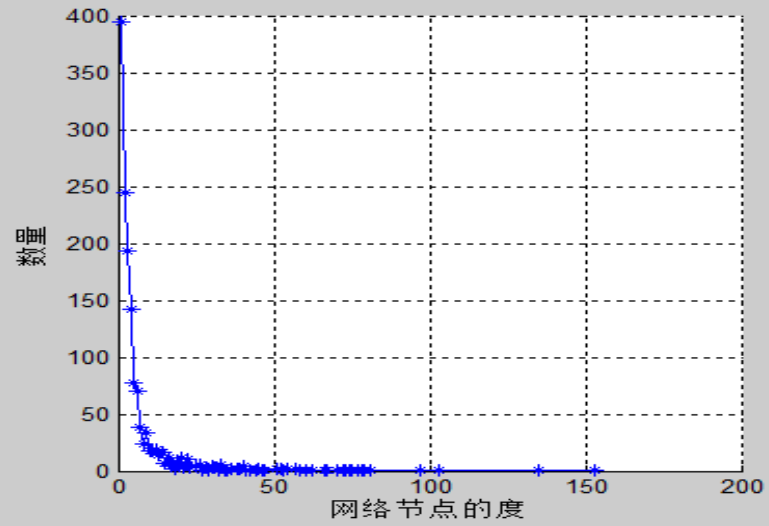


- 每个回合由以下几部分组成：增长点期、一轮投资期、二轮投资期
- **【增长点期】**
- 投资者增长30%
- 被投资者增长30%
- **【基础动态网模型】**
- 所有投资者按照其投资力度随机一个或多个被投资者进行连接，若成功连接，则定义该点进行了一次“投资”
- 若有两个或以上的投资者类点同时选择了同一个被投资者，则定义这两个投资者进行了一次联合

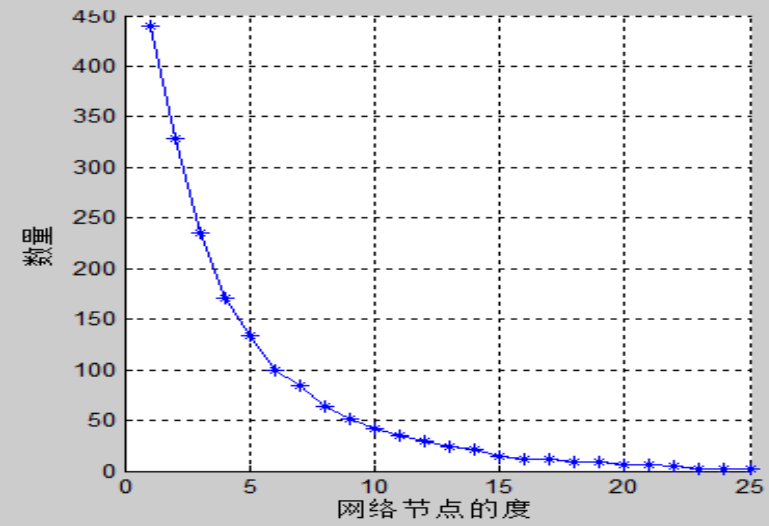
- 【加入圈子理论的动态网】
- 加入下述三个选择伙伴的条件
- 模型二，加入
- 一、过去合作次数越多合作机率越高。
- 模型三，加入
- 二、三步距离之外合作机率为零。
- 三、距离越近合作机率越高。

# 度分布图

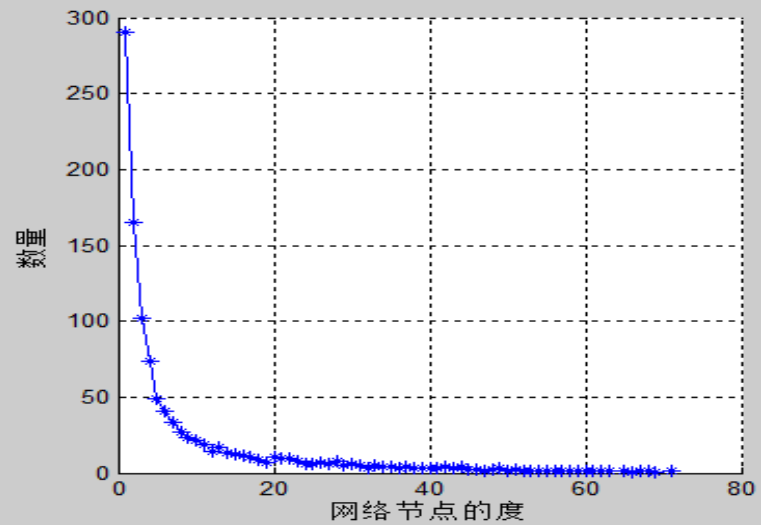
真实数据联合投资网络节点度分布



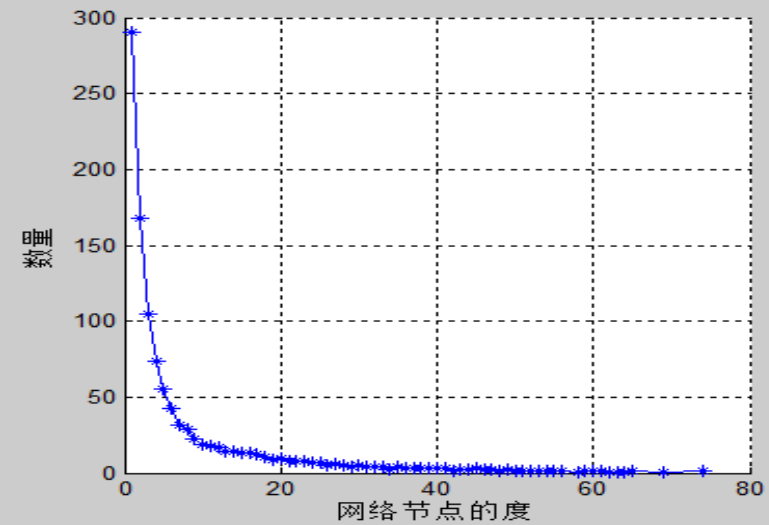
模型1网络节点的度分布



模型2网络节点的度分布

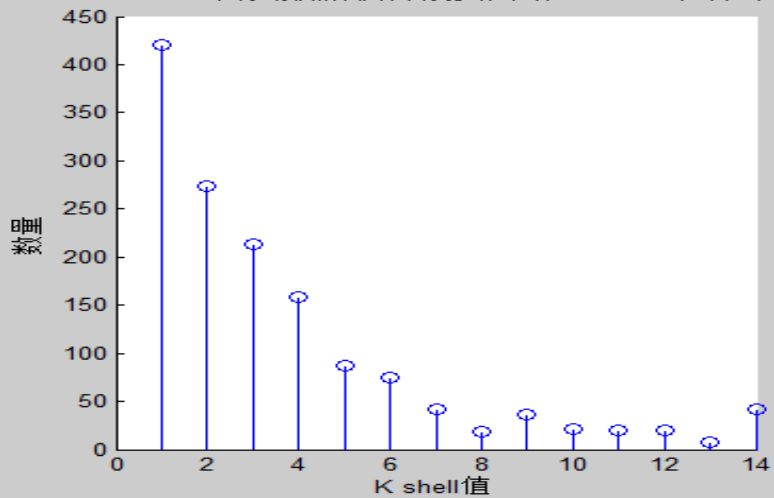


模型3网络节点的度分布

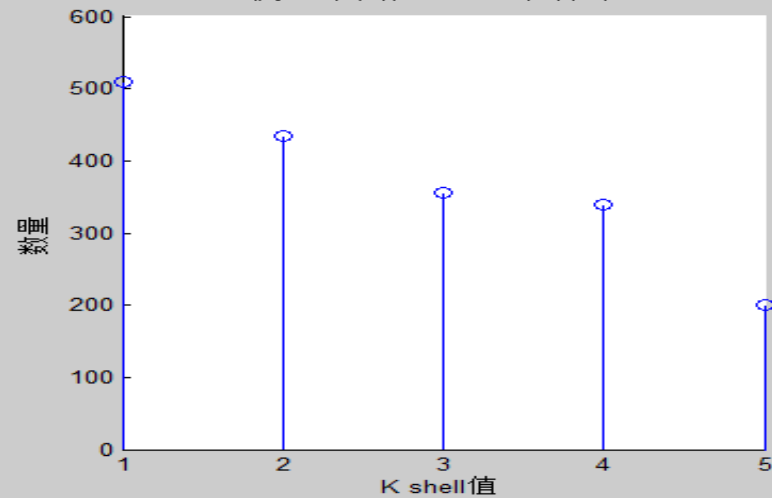


# K-Shell分布图

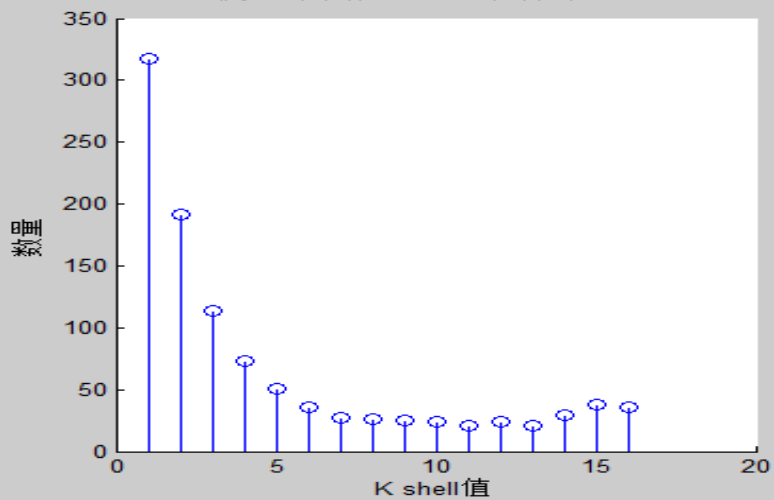
真实数据联合投资网络K-shell值分布



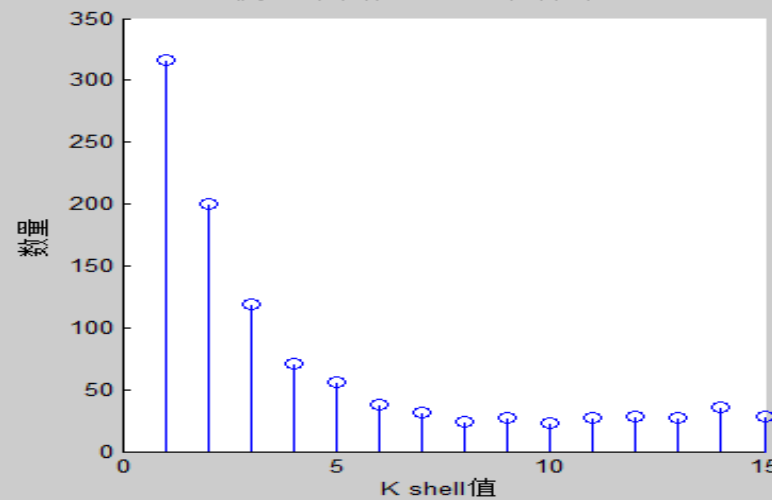
模型1网络K-shell值分布



模型2网络K-shell值分布

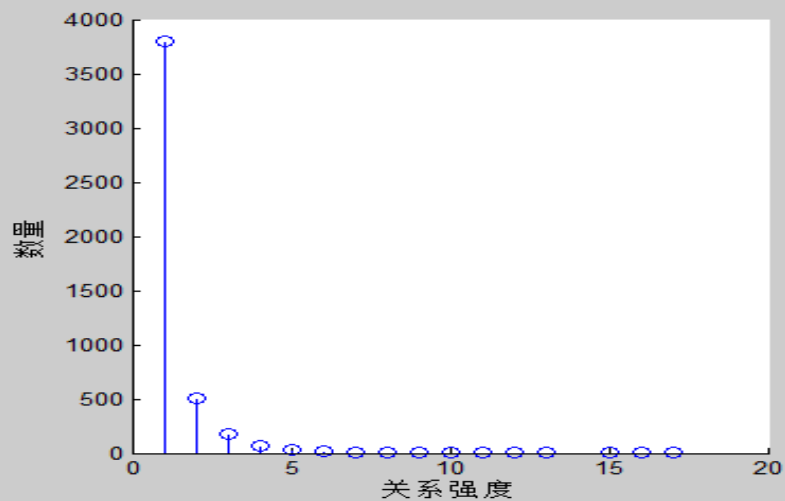


模型3网络K-shell值分布

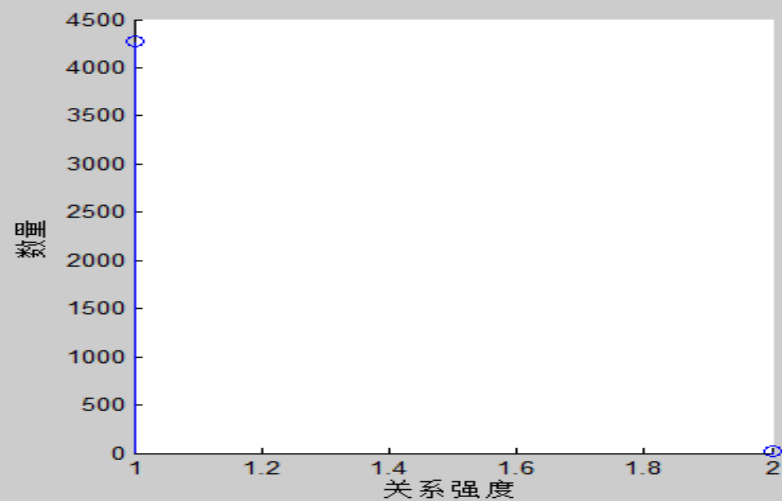


# 关系强度分布图

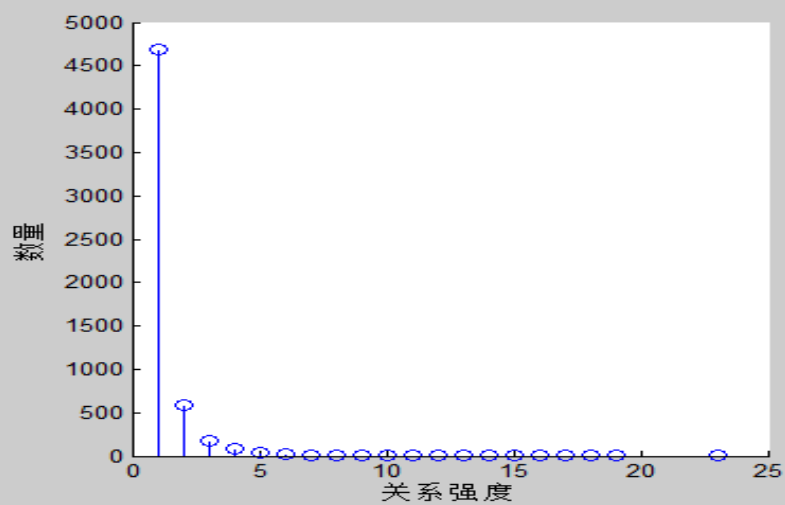
真实数据联合投资网络关系强度分布



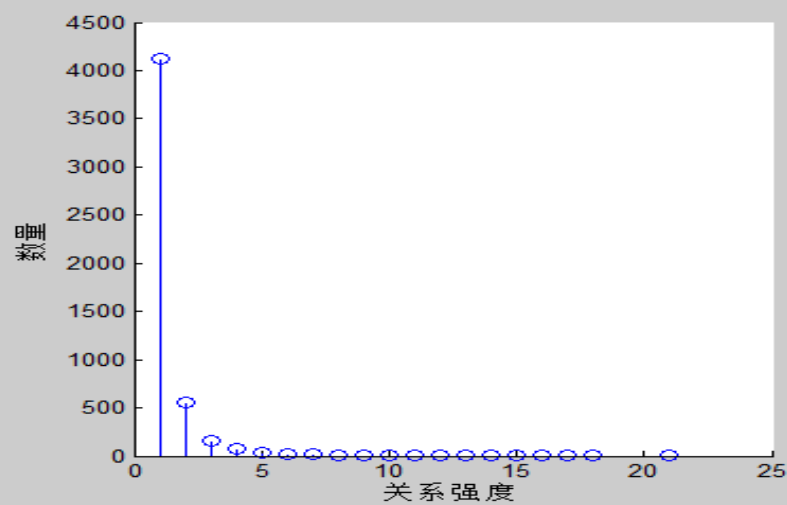
模型1网络关系强度分布



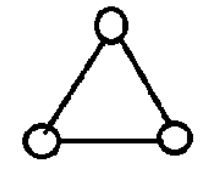
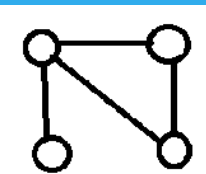
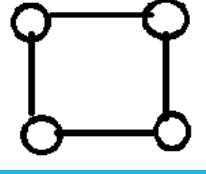
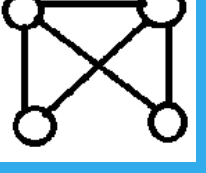
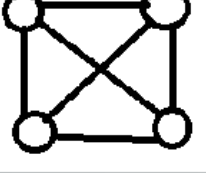
模型2网络关系强度分布



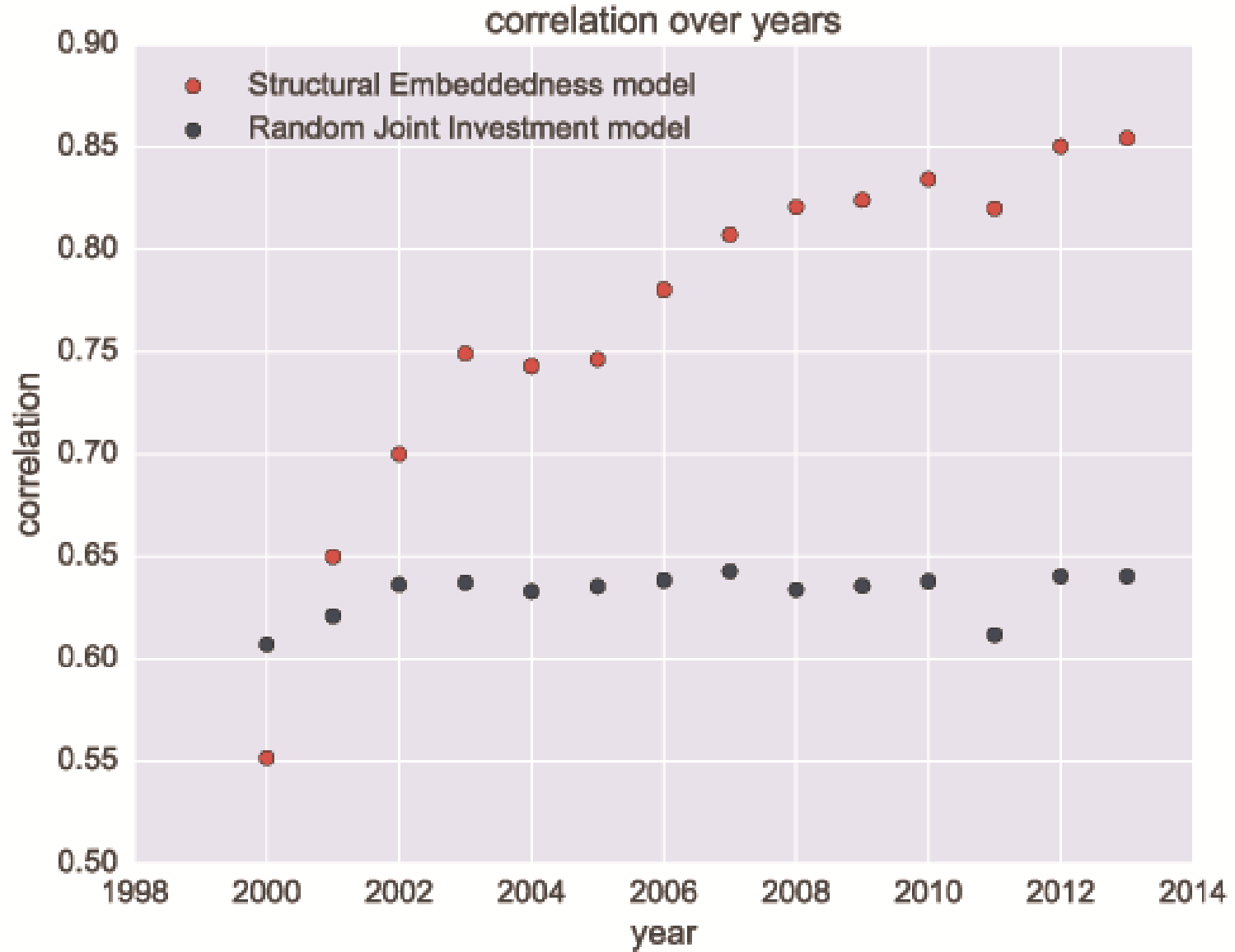
模型3网络关系强度分布



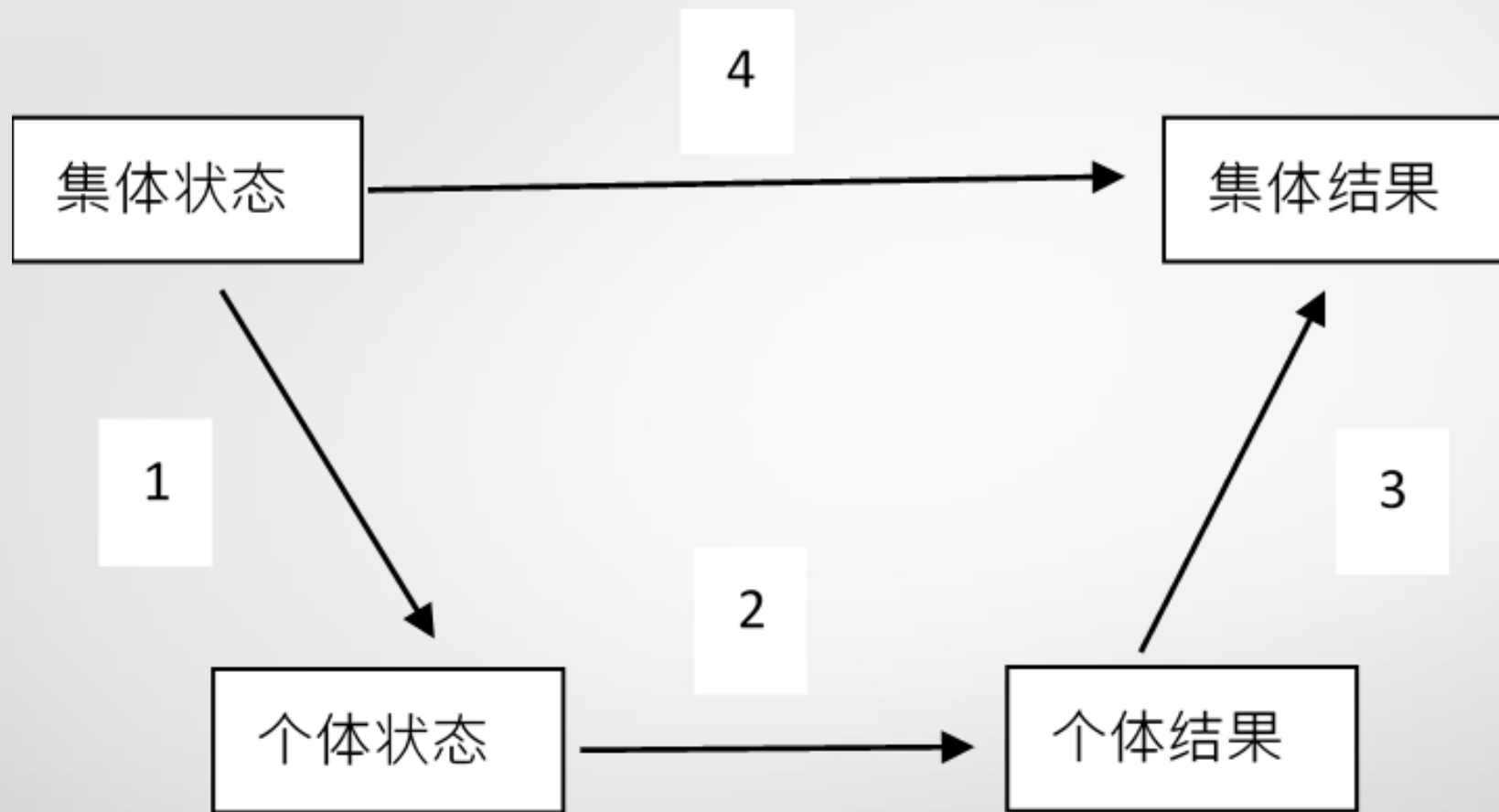
## Motif 数量比较

模体类型	真实数据	模型1	模型2	模型3
	7668	634	6646	7566
	633815	14642	441325	469410
	0	551	0	32126
	58765	127	51026	53060
	10258	50	5041	5359

Network Name	Real Network Property			Structural Embeddedness		Random Investment	
	elite	follower	all nodes	elite	follower	elite	follower
nodes category	elite	follower	all nodes	elite	follower	elite	follower
degree	18.216	3.650	6.440	12.397	2.583	0.765	0.765
k-shell	12.000	2.330	3.678	7.516	1.018	0.519	0.519
betweenness	0.014	0.002	0.008	0.019	0	0	0
investment frequency	47.310	4.090	8.500	75.572	4.450	0.765	0.765
elite-clique density	0.492	0.003	0.004	0.248	0.001	0	0
elite-clique EI index	123	0.750	-	124	0.500	0	0
elite-follower cave density	0.100	-	-	0.100	-	0.341	-
caves' EI index	25	-	-	50	-	226	-







James Coleman的模型

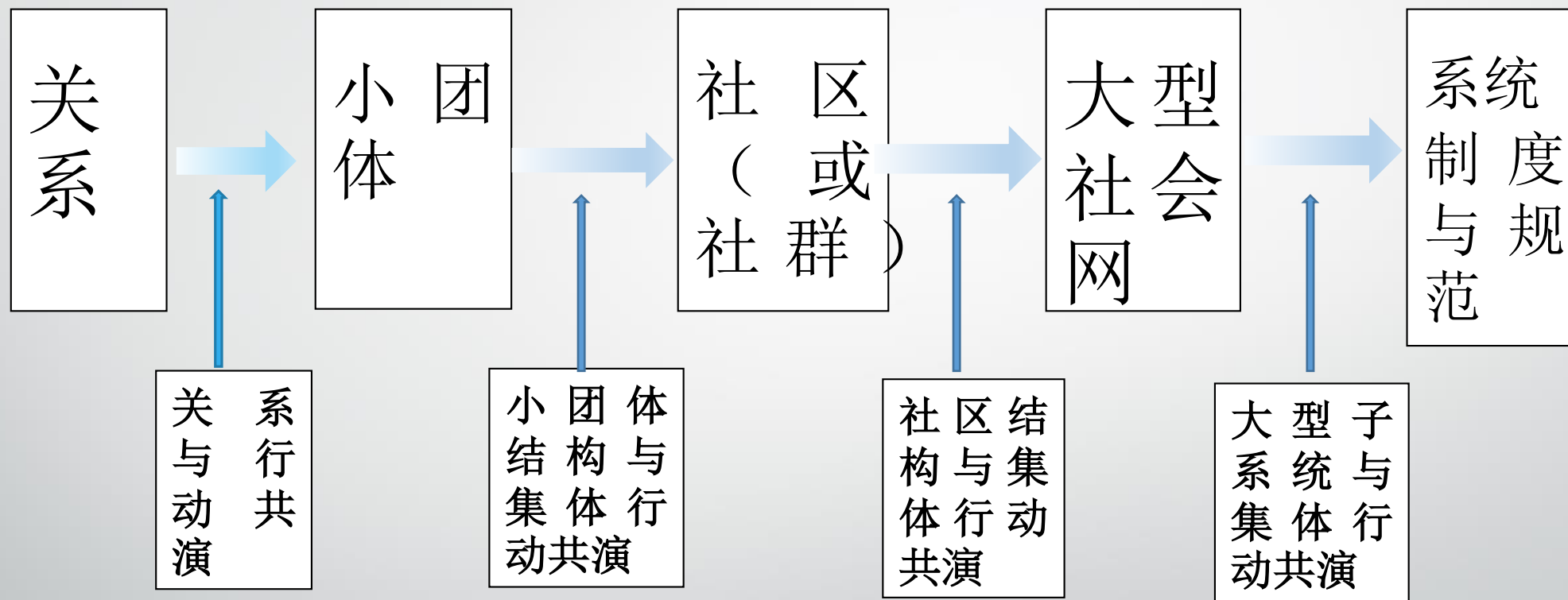
# 一个复杂思维的解释

- 解决上图中第一和第三的问题。
- 加入四个要素 a. 权力与信任关系 b. 小团体 c. 较大复杂网络的结构与演化 d. 制度与文化网络的共同演化。
- 有权力掌握者就要有权力屈从者，这些人间的关系与网络如何展现？所以格兰诺维特探讨了结构洞的概念，不管个人拥用怎样的权力特质，这是一种权力的结构性表现，也就是两群人间的中介者拥有了信息垄断与鱼蚌相争渔翁得利的利益，但其权力成立的前题又包括下述几项，两个分立的圈子各自拥有足堪权力剥削的资源，两个圈子有沟通的需要，且依赖中介者的沟通，最后，中介者还能操控两个圈子的分分合合，合了，桥就不需要了，分远了，圈子各自孤立，桥就断了，中介者的权力都不复存在。

# 复杂社会系统

- 系统观
- 引入社会连带（关系）
- 结构与行动的共演
- 自适应——多元「势」
- 多元势的动态平衡
- 动态与演化
- 搭一条微观与宏观间的桥
- 反化约主义
- 网络结构
- 非常态与系统突变
- 自组织
- 涌现现象
- 节——预测并掌握势的拐点

# Granovetter提出的理论架构



# 大数据为理论带来新的范式

- 因果模型 vs. 系统模型
- 化约思维 vs. 复杂思维
- 静态或比较静态研究 vs. 动态研究
- 常态系统 vs. 系统转型
- 线性因果 vs. 涌现现象
- 新范式：理论+实证，大型理论+中层理论+理论模型，定性研究+定量研究，结构化数据+大数据
- 共同解决理论中的复杂性、动态性、网络、涌现以及非常态系统的非线性演化问题。