

# Integrating Data Mining and Qualitative Studies in Theory Building

--Taking the Chinese Venture Capital Industry as  
an Example

Source:

Luo, Jar-Der

Tsinghua U. Sociology Dept.

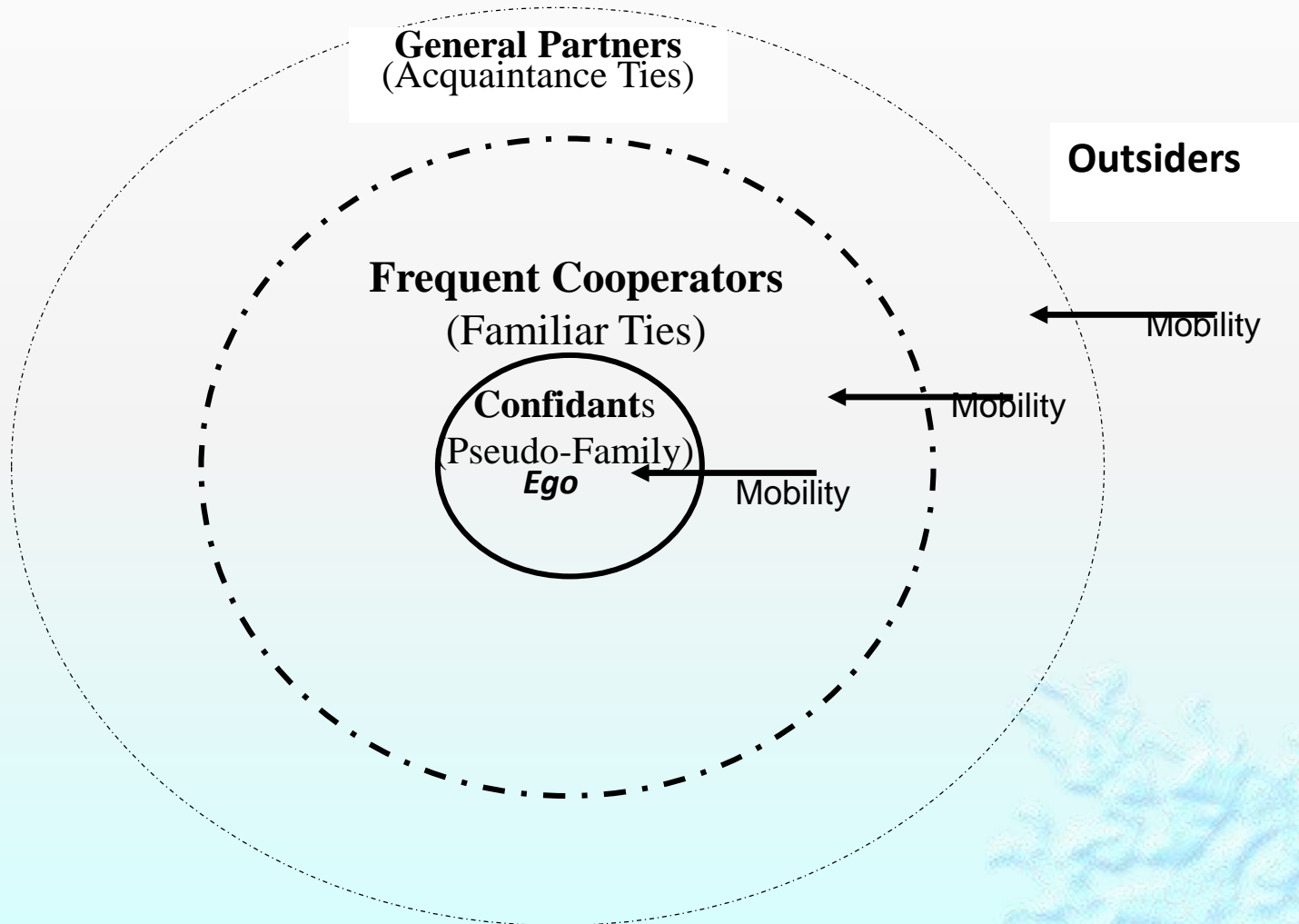


- ◆ How can we integrate data mining, theoretical modeling and theory building?
- ◆ A methodological cycle: data mining → Interpretation → dialogue with theories → theory building → hypotheses testing (in causal models) → finding ground truth
- ◆ 1) → feeding back to the model for data mining
- ◆ 2) Or → building system model → revealing the black box of behavioral process → testing model by real data → finding ground truth
- ◆ → next run of the cycle

# 1. The Origin of the Study: Guanxi Circle

- ◆ 1. A circle is centered on an ego.
- ◆ 2. A circle has with a structure of “differential mode of association”.
- ◆ 3. So, there is a core of the circle, which is indicated by loyal, unbreakable and intimate relationship, while those periphery members outside the core are mainly composed by long-term relations with limited liability in frequent favor exchanges.
- ◆ 4. The boundary of a circle may be open, and thus there is overlapped area among several circles.

# Guanxi circles in the VC Industry



- ◆ As stated above, a guanxi circle is composed of a centered ego's strong ties, including pseudo-family and familiar ties.
- ◆ Summarizing the arguments stated above, the static features of a guanxi circle include the following:
  - ◆ 1. A guanxi circle is centered on an ego.
  - ◆ 2. A guanxi circle has a structure with differential modes of association.
  - ◆ 3. There is an inner core within the guanxi circle, which is indicated by loyal, unbreakable and intimate relationship, while those periphery members outside the core are mainly composed of by long-term relations with limited liability in frequent favor exchanges

- ◆ 4. The boundaries of a guanxi circle are usually open, and thus there is often an overlapping area among several guanxi circles.
- ◆ Thus, I categorize Chinese workers into five types—core members in leaders' guanxi circles, core members in common guanxi circles, periphery members in guanxi circles, bridges, and finally outsiders.



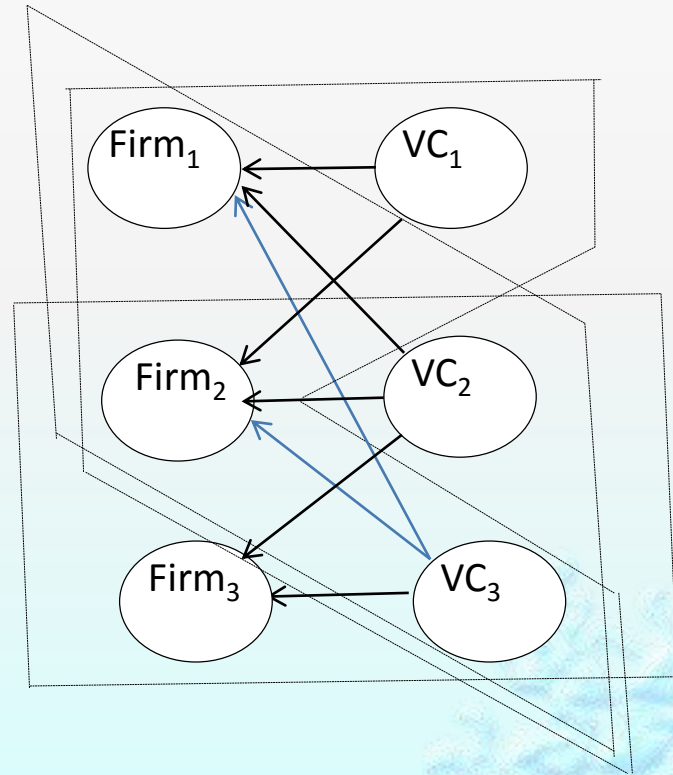
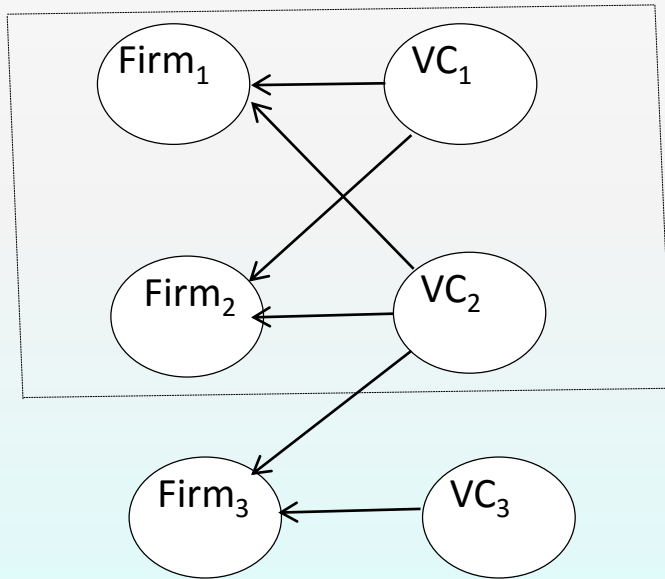
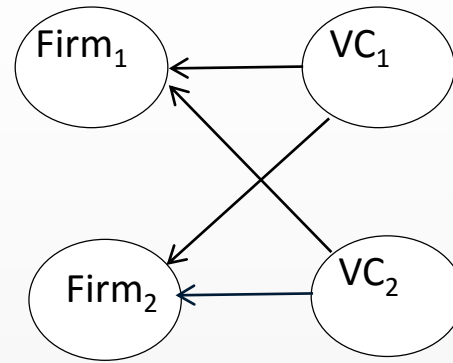
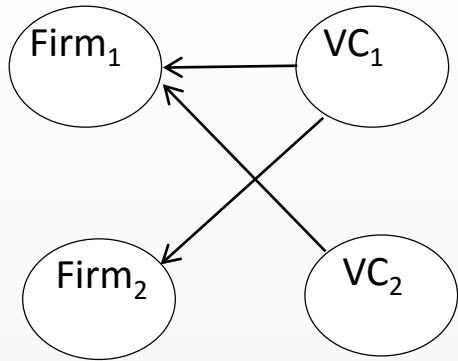
- ◆ Hypothesis 1: Frequent cooperation increases the possibility of new joint investment.
- ◆ Hypothesis 2: Relational distance is convexly associated with the possibility of joint investment.
- ◆ Hypothesis 3: Accumulative advantage is positively associated with the possibility of joint investment.
- ◆ In the regression, all the three hypotheses get supported.

- ◆ In the data mining of link prediction, the following factors are significant predictors:
- ◆ Common neighbors—increase the accuracy of prediction 12%--13%.
- ◆ Similarity of property right.
- ◆ Similarity of industry
- ◆ Following the short trend.
- ◆ All these three predictors increase the accuracy of prediction 6%--7%



## II. The Analytical Results of ERGM Model

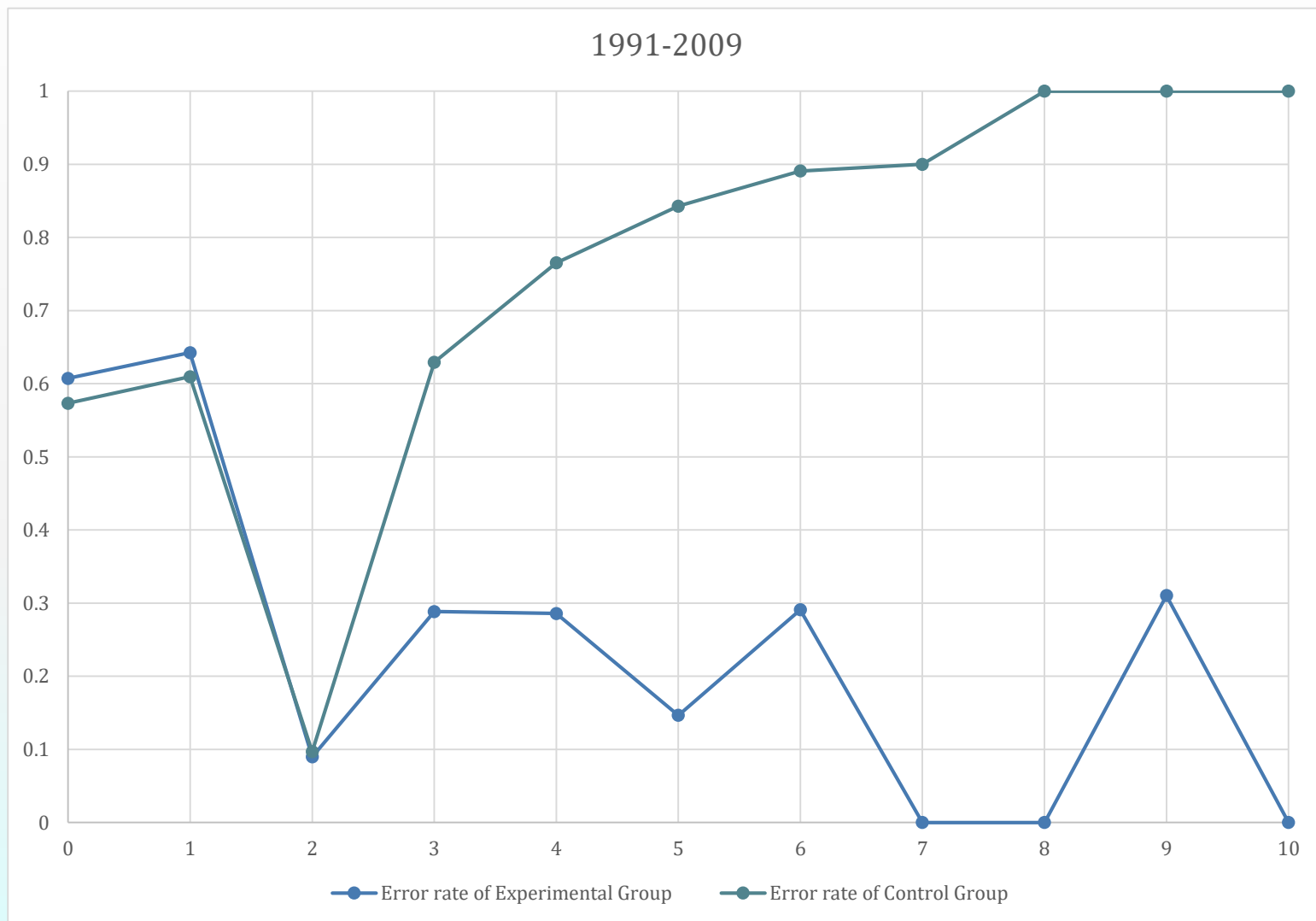
- ◆ In the following we explain the meaning of 3 type of ERGM statistics used in our model:
- ◆ 1. Edge: In the complete network set of ERGM, each edge appears at a probability of 0.5. But in target network, the density of edge is decided by the number of investment records and the number of institutions. So the probability of each edge appears in this network may be much lower than 0.5. We need to consider about the effect of edge. This is measured by statistic "edge" (the number of edges in the network).



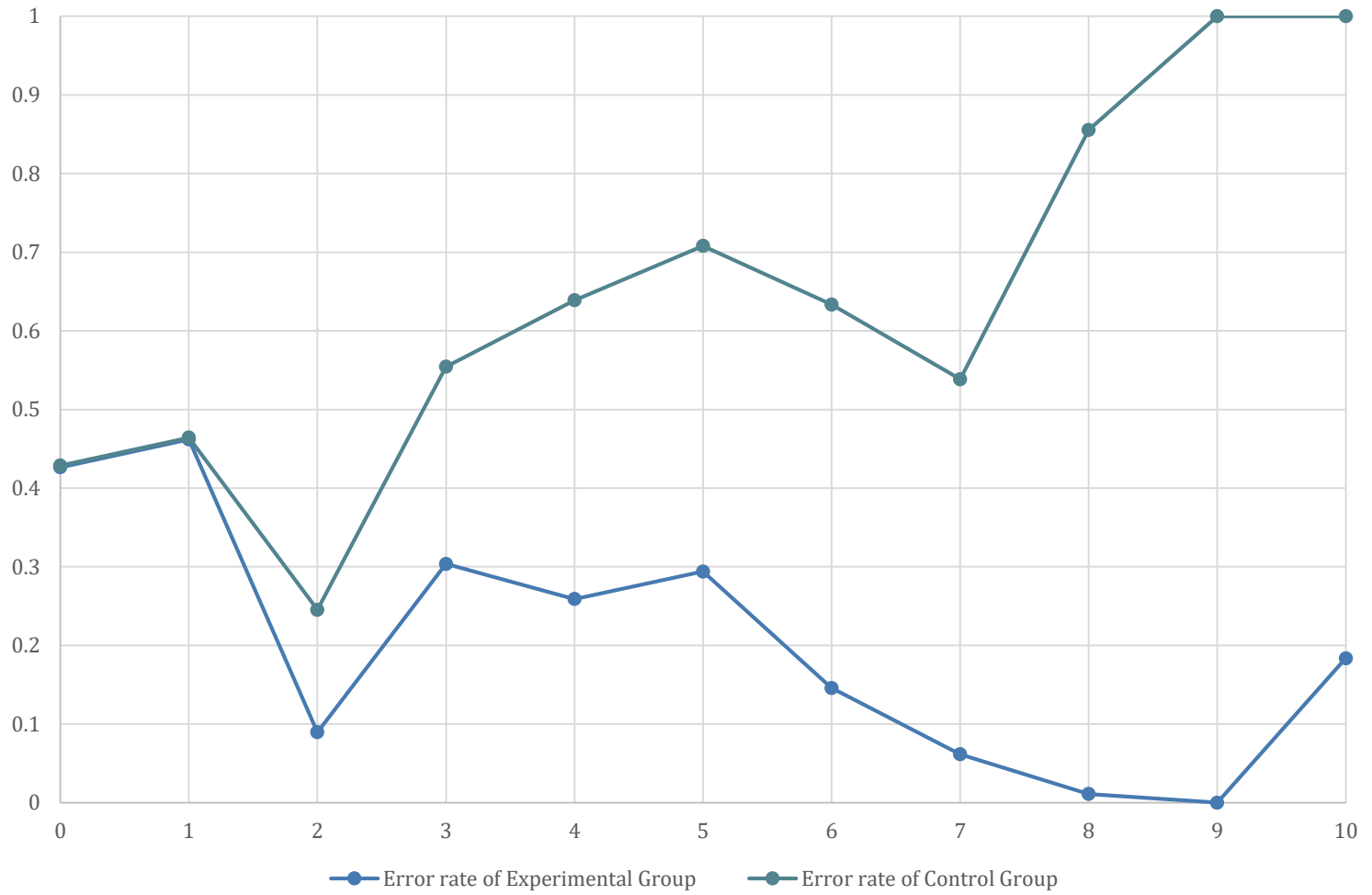
- ◆ 2. Degree: In ERGM, there can be isolated nodes as edges are generated randomly. But as our target network is generated by investment records, so each node has at least one edge, affecting the distribution of degree. We should add some complement statistics to measure the degree character of target network. This can be measured by “b1concurrent” and “b2concurrent” (the number of nodes (b1 means VC and b2 means invested companies) with at least 2 degree), or “kstar2” and “kstar3” (the number of 2-star and 3-star in the network).
- ◆ 3. Quadrangles: open quadrangles is also called “three-path” in network, while closed quadrangles having another name “cycle4”. To prove our hypothesis, the “cycle4” coefficient must compete for the explanation of network structure with all complement shapes mentioned before, i.e. “cycle4” should keep positive whatever kind of compliment shapes we add.

- ◆ We introduce our experimental and control groups here:
- ◆ Target network: the information of original network data
- ◆ Experimental group: the ERGM estimation of target network, with statistics of edges,  $b_1$  concurrent,  $b_2$  concurrent, three-path and cycle4. Making simulated networks based on it.
- ◆ Control group: another ERGM estimation of target network like Experimental group, except that it do not measure closed quadrangles "cycle4".





1991-2011





# Go Back to the Field

- ◆ In addition to providing “surplus” to the centered VCs in a network, new entrants also bring new information, local knowledge and new opportunities into a network. These resources are keys for forming new projects. As a manager of a VC firm, Mr. X, said:
  - ◆ *The familiar partners [those who have joint investment with the interviewee in the past] are hard to negotiate with; sometimes, we disagree with each other due to self-interest, but it is hard to talk straight...we are friends, you know....*
  - ◆ *Cooperation with new VCs for sure is profitable. Money, local knowledge, new information of projects..., but I will not allow them to say too much [new VCs can't manage the project].*
  - ◆ *One side may need to be sacrificed for the other side's interest...in the cooperation of Eastern type [zero-sum game].*

◆ One single transaction is fragile, but a series of investments in the collective actions of a network is robust. As a junior partner of a Chinese VC stated:

◆ *Guanxi is indeed important, ... some projects fail, but it is O.K., since guanxi is there. It [a successful project] is the result of process of collective actions, ... some things are good for guanxi building.*

◆ *I met him [the new syndication partner] several times to look for a new project. But, we didn't have dialogue, ... unfamiliar. Once, we meet in a dinner banquet, [a common good friend] introduce him to me, ... I feel good [for cooperation], ... then we invest jointly.*

- ◆ The center position will attract more partners, who thus are invited into a larger network so as to find more chances to get to know new friends. In return, they bring more good projects and key information in this network.
- ◆ Long-term cooperation breeds trust, and the expectation for future runs of the game ensures cooperative relations continue (Hardin, 2001; Axelrod, 1984).

- ◇ As put by Mr. S, the CEO of a VC, the most important LP (limited partner, or what we say “follower”) of one major player in the VC community:
- ◇ *In our community, it is personal relations, rather than the relation between two funds...He (a famous investor) once called for a joint investment, many (LPs) want to follow....Mr. Shao’s company and Mr. Luo’s company (using two major partners of two VCs as an example) have high frequency (of joint investment), association, intimacy between them. This is a circle, in my understanding.*
- ◇ Mr. S states this as the following:
- ◇ *I live downstairs from Mr. Chu (the CEO of the VC that Mr. S follows)...I even help take care of his apartment when he is on vacation.*

# III The Results of K-shell Analysis:

## Electronics

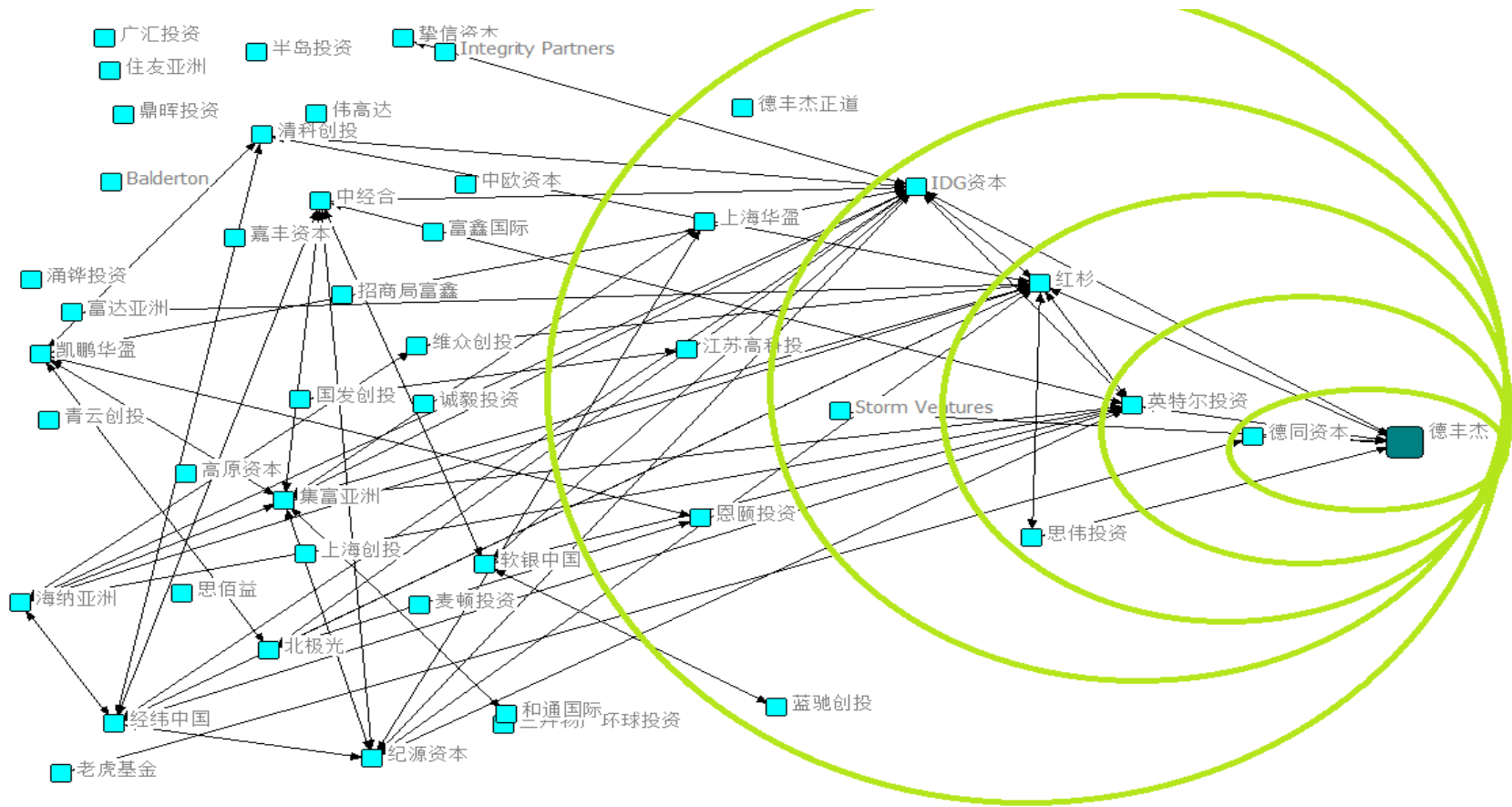
0	华成国际	VC	外资	25	0.29	11	60	今日资本	VC	外资	12	0.590909	11
1	君联资本	VC	China	66	0.14359	11	61	日亚投资	VC	外资	24	0.25	11
2	招商局富鑫	VC	合资	26	0.4	11	62	英特尔投资	VC	外资	71	0.152515	11
4	高通	VC	外资	360	0.250794	11	65	富鑫国际	VC	外资	18	0.464052	11
5	晨兴创投	VC	外资	350	0.294118	11	69	光速创投	VC	外资	32	0.346774	11
6	北极光	VC	外资	450	0.171717	11	78	凯鹏华盈	VC	外资	25	0.24	11
7	红杉	VC	外资	710	0.135614	11	87	祥峰集团	VC	外资	37	0.267267	11
8	IDG资本	VC	外资	100	0.10101	11	98	蓝驰创投	VC	外资	22	0.233766	11
9	经纬中国	VC	外资	310	0.197849	11	105	智基创投	VC	外资	36	0.261905	11
10	启明创投	VC	外资	410	0.209756	11	107	深圳创新投	VC	China	52	0.095023	11
15	赛富投资基 金	VC	外资	290	0.150246	11	110	DCM资本	VC	外资	37	0.168168	11
16	富达亚洲	VC	外资	240	0.217391	11	111	红点投资	VC	外资	30	0.404598	11
21	纪源资本	VC	外资	420	0.247387	11	116	壹普兰	VC	外资	11	1	11
22	德同资本	VC	外资	300	0.195402	11	117	怡和创投	VC	外资	13	0.794872	11
25	高盛	PE	外资	430	0.215947	11	119	德丰杰	VC	外资	33	0.157197	11
26	华登国际	VC	外资	680	0.162862	11	121	海纳亚洲	VC	外资	37	0.207207	11
32	软银中国	VC	外资	420	0.134727	11	125	鼎晖创投	VC	外资	22	0.290043	11
33	集富亚洲	VC	外资	470	0.225717	11	131	LB投资	nan	nan	13	0.75641	11
40	上实投资	PE	外资	300	0.372414	11	139	中经合	VC	外资	26	0.289231	11
50	KTB	VC	外资	340	0.358289	11	160	GeneralCatal yst	VC	外资	19	0.415205	11
53	联创策源	VC	外资	240	0.246377	11							
55	金沙江创投	VC	外资	320	0.171371	11							
59	清科创投	VC	外资	400	0.194872	11							



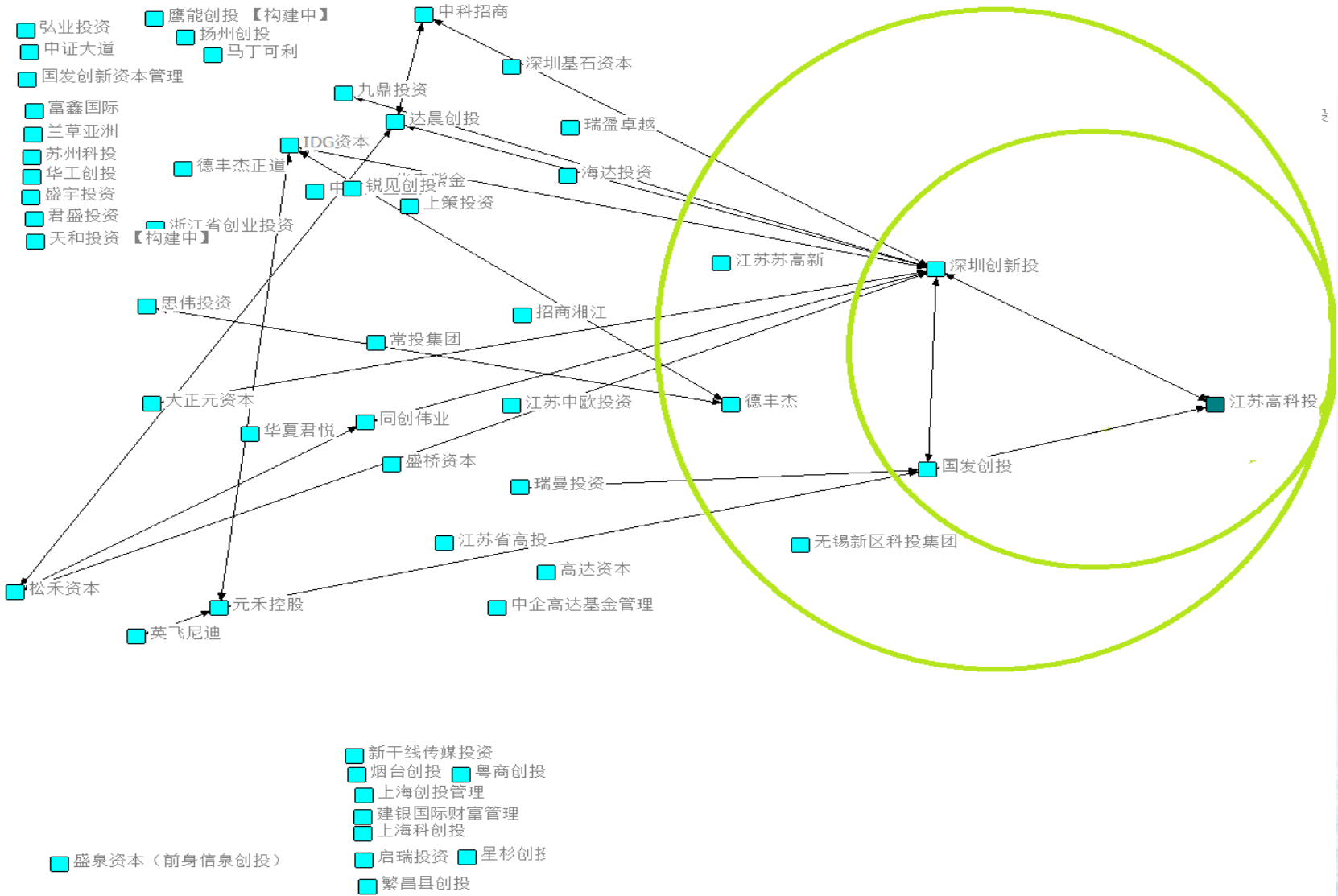
# Go Back to the Field—Delphi

name	VC/PE/other	local/joint/overseas	CC	betweenness	Rank in 829	k-shell 42 in 829	Approval Rate
同创伟业	VC	China	0.055631	154571.7	1	14	4
招商科技	VC	China	0.113846	47052.21	4	14	4
金石投资	Bank	China	0.122807	11220.72	24	10	4
如山投资	VC	China	0.717949	12673.48	20	10	3.5
百瑞创新	VC	China	0.74359	12669.75	21	10	3.5
天创资本	VC	China	0.208791	11343.26	22	10	3.5
深圳创新投	VC	China	0.043774	38595.67	5	12	3
分享投资	VC	China	0.254545	11269.86	23	10	3
九鼎投资	PE	China	0.071301	10631.52	25	10	3
长江国弘	PE	China	0.705128	19898.4	12	11	2.5
天堂硅谷	VC	China	0.095385	22660.06	9	11	2
华睿投资	VC	China	0.123333	22086.75	10	11	2
深圳基石资本	VC	China	0.235294	21424.54	11	11	2
上海创投	VC	China	0.254032	17551.65	15	11	2
深圳高新投	VC	China	0.358974	9382.929	29	9	2
建银国际财富管理	PE	China	0.208791	8632.415	30	9	2





- 毅鸣投资
- 龙脉创投
- WR Hambrecht + Co
- W.R. Hambrecht
- 高特佳
- 普凯投资基金
- SAP Ventures
- 伊藤忠商事
- 和泰投资
- 高正创投
- 慧创志成创投
- 腾讯产业共赢基金
- 盈峰资本
- 天创资本
- UMC Capital
- FPEA 【构建】
- 思科香港



# Next? Bringing Data Mining Back Again.

- ◆ How many types of investors are there?—based on investment frequency and tendency toward syndication.
- ◆ How many types of syndication ties are there?—based on syndication numbers, relational distance and complementarity.
- ◆ How many clusters in the whole VC industry are there?—based on the cliques of “big brothers”.
- ◆ The last and most important one is the dynamic forces behind syndication—the motivations of different types of investors.