Social Network Modeling in Finance with Text Analytics

The Internet Ecosystem, Investor Sentiment,
and Sociology of Market Microstructure and Investment Managers

Sanjiv Das
Santa Clara University

Papers: http://algo.scu.edu/~sanjivdas/vccomm.pdf
http://algo.scu.edu/~sanjivdas/midaswww2011_FINAL.pdf
http://algo.scu.edu/~sanjivdas/fincom.pdf
http://algo.scu.edu/~sanjivdas/newsmetrics.pdf

Joint work with Jacob Sisk (Thomson-Reuters), Amit Bubna (Indian School of Business), N.R. Prabhala (Univ. of Maryland), Peter Tufano (Harvard), Mike Chen (Intel), Asis Martinez-Jerez (Harvard), Mauricio A. Hernandez, Howard Ho, Georgia Koutrika, Rajasekar Krishnamurthy, Lucian Popa, Ioana R. Stanoi, Shivakumar Vaithyanathan (IBM Almaden Labs).

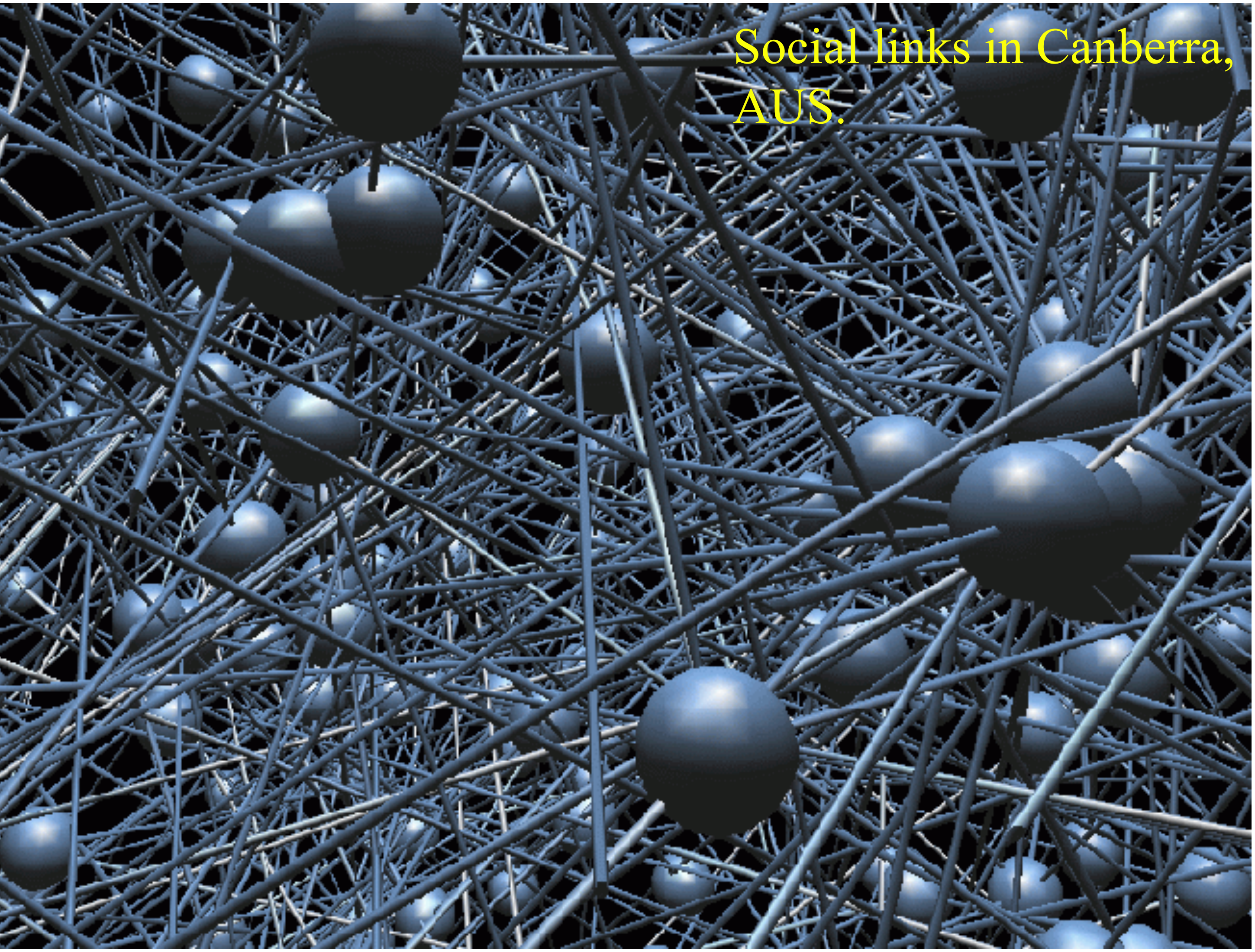
@U of Maryland, April 2013

What is this talk about?

- The structure of the social (random) network over which financial information flows (chat rooms and message boards), and sentiment extraction. And three illustrative applications:
  - The web graph of stock linkages and trading rules.
  - Extracting systemic risk from the web graph of interbank lending relationships.
  - Extracting communities in the venture capital industry.

- If there is time I'll talk about sentiment extraction from web talk.

The interface of finance, statistics, computer science, sociology, psychology

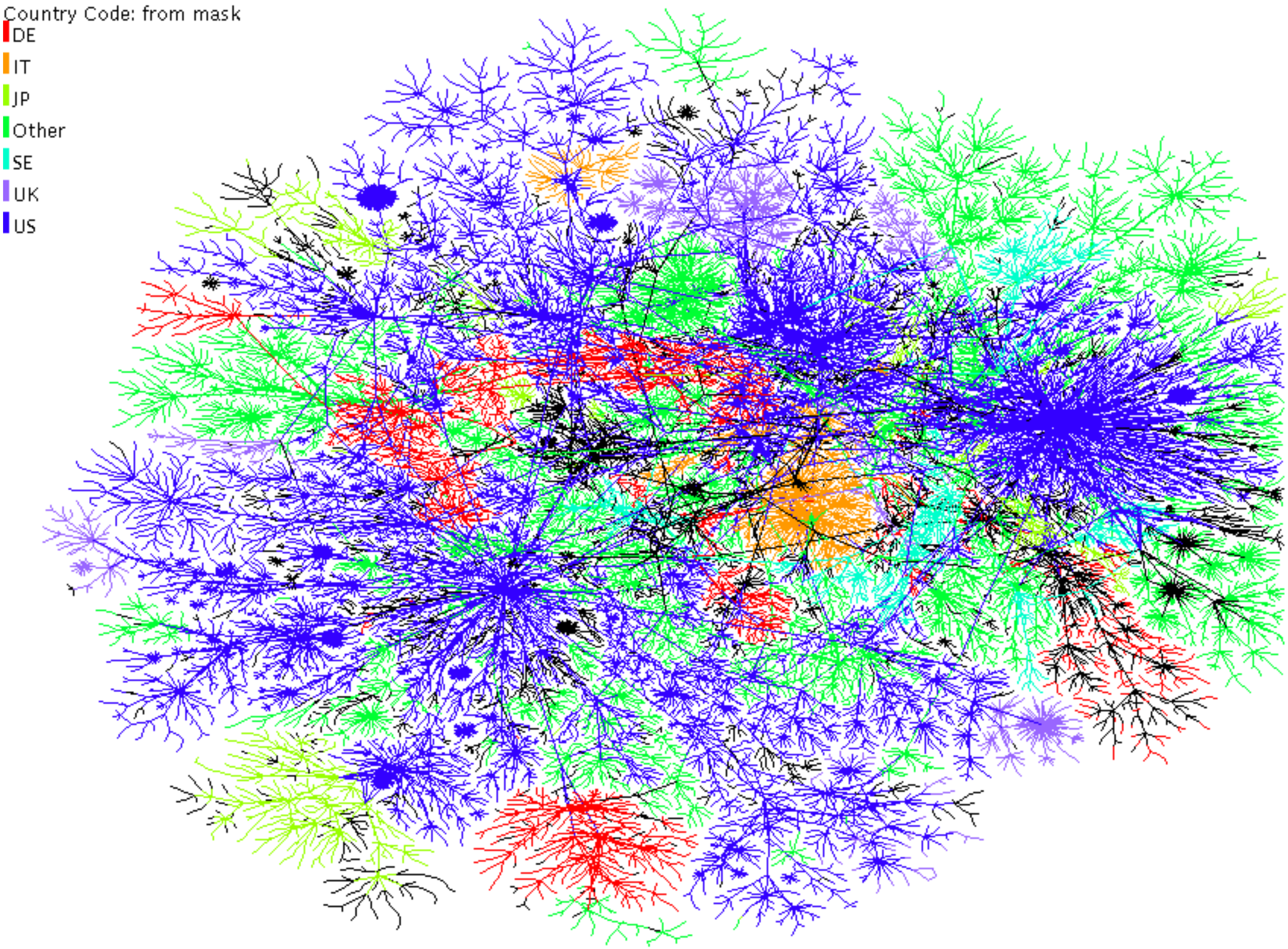


1. Much more information than before.
2. How to quantify and analyze it?

WOM = word of mouth

Finance sentiment is more complex than usual content parsing because….

1. Dual layered (stocks & agents).
2. Multivalent connection definitions (varied criteria).
3. Connection quality also matters.
4. Human behavior is less structured.
5. Content parsing is required.


Country Code: from mask
DE
IT
JP
Other
SE
UK
US

Theoretical Implications

- The spread of a social influence depends on the network graph on which it lies:
  - Is it a regular graph or scale-free? Scale free graphs propagate information faster from “pipelining”
  - How much influence do central nodes have? Graphs propagate faster when central nodes fire more.
  - How does information entering the graph impact sentiment formation? Can we extract sentiment?
  - How does network fragility impact markets?

Network Structure

- Can we figure out the mathematical structure of the network?
- Which structure maximizes social communication, and how is this measured?

  1. Centrality scores (Bonacich).
  2. 6 degrees of freedom (a la Erdos# network) (Milgram).
  3. Power laws (Barabasi, Strogatz, Watts).

Metcalfe's Law: The utility of a network is proportional to the square of the number of users.

Financial Communities

Sanjiv Das & Jacob Sisk (2005)

Journal of Portfolio Management

- Studies the sociological mechanics of the link between stock returns and information.
- Understand how opinions are linked across tickers during small investor discussion based on collective information unit, the financial community. These are clusters of tickers sharing and accessing the same information generators.
- Graph theoretic techniques are used to detect financial communities and to summarize their properties.
- Community stocks display connectedness, and we find that the greater the connectedness in a financial community, the greater the covariance of returns within the community as opposed to that amongst stocks that are not part of a major financial community.
- Highly connected stocks, on average, have lower return variance and higher mean returns.
- Using eigenvector techniques, we detect stocks that are hubs for information flow, using a sociological measure known as centrality. Stocks with high centrality scores tend to have greater average covariance with other stocks than those with low scores.

Figure 1: **Connectedness graph**

The graph presents an example of a connectedness diagram. There are 7 tickers, *A*-*G*, represented by the nodes on the graph. The numbers represent the number of common posters between the ticker's message boards. In this graph, the connection threshold is set to $K = 1$, meaning that two tickers are connected if there is at least one common poster on their message boards. Hence, all connections are valid, irrespective of strength. In this example, we can see that all the stocks form a single community, as they are all linked.

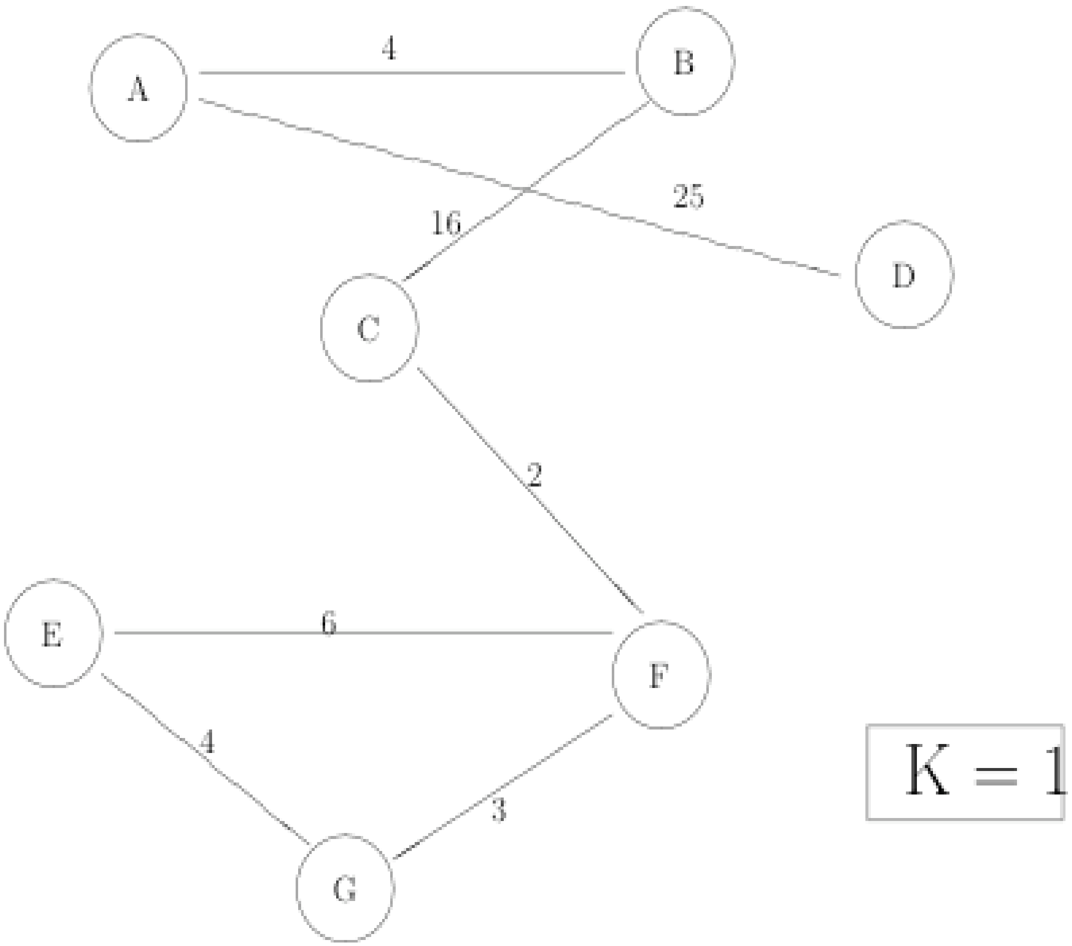
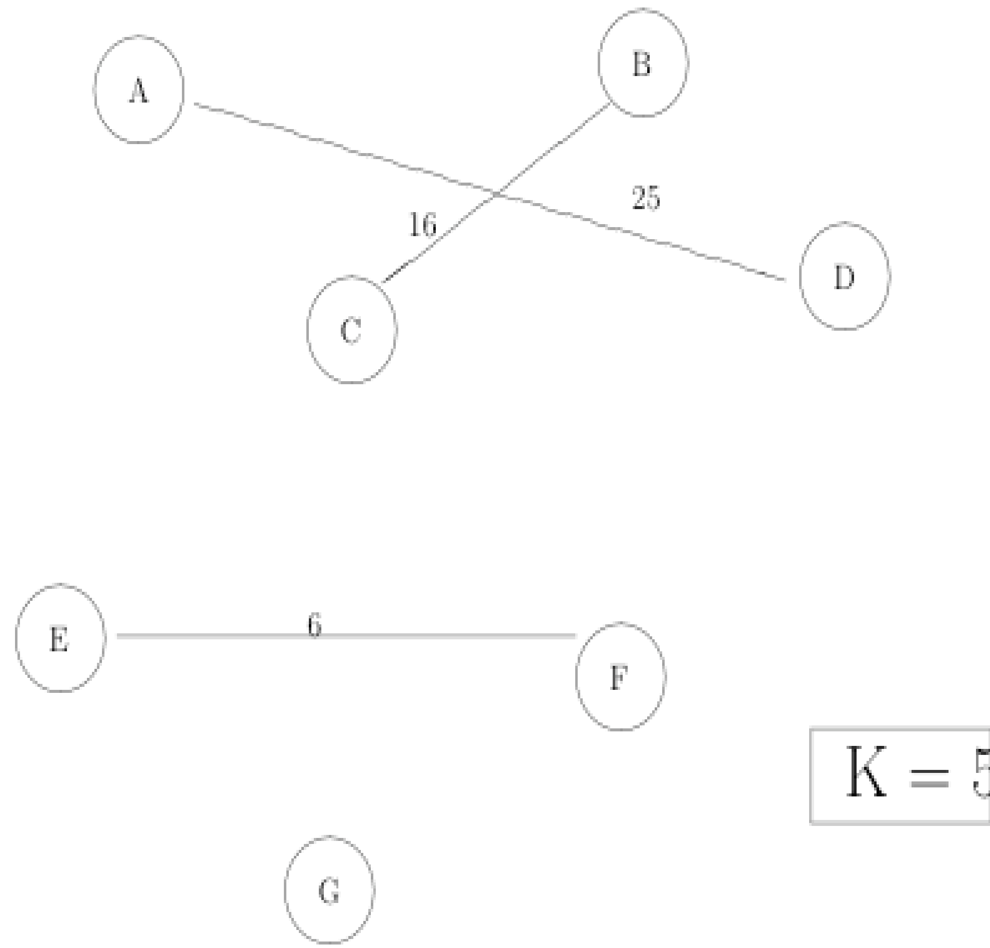


Figure 2: **Connectedness graph with higher threshold**

The graph presents an example of a connectedness diagram with a higher threshold, and should be read in comparison to Figure 1. There are 7 tickers, *A*-*G*, represented by the nodes on the graph. The numbers represent the number of common posters between the ticker's message boards. In this graph, the connection threshold is set to $K = 5$, meaning that two tickers are connected if there are at least 5 common posters on their message boards. Hence, connections are valid, depending on strength. Instead of one large community, we get 4 communities: $\{A, D\}, \{B, C\}, \{E, F\}, \{G\}$. Thus, there are 3 small communities, and one singleton community.



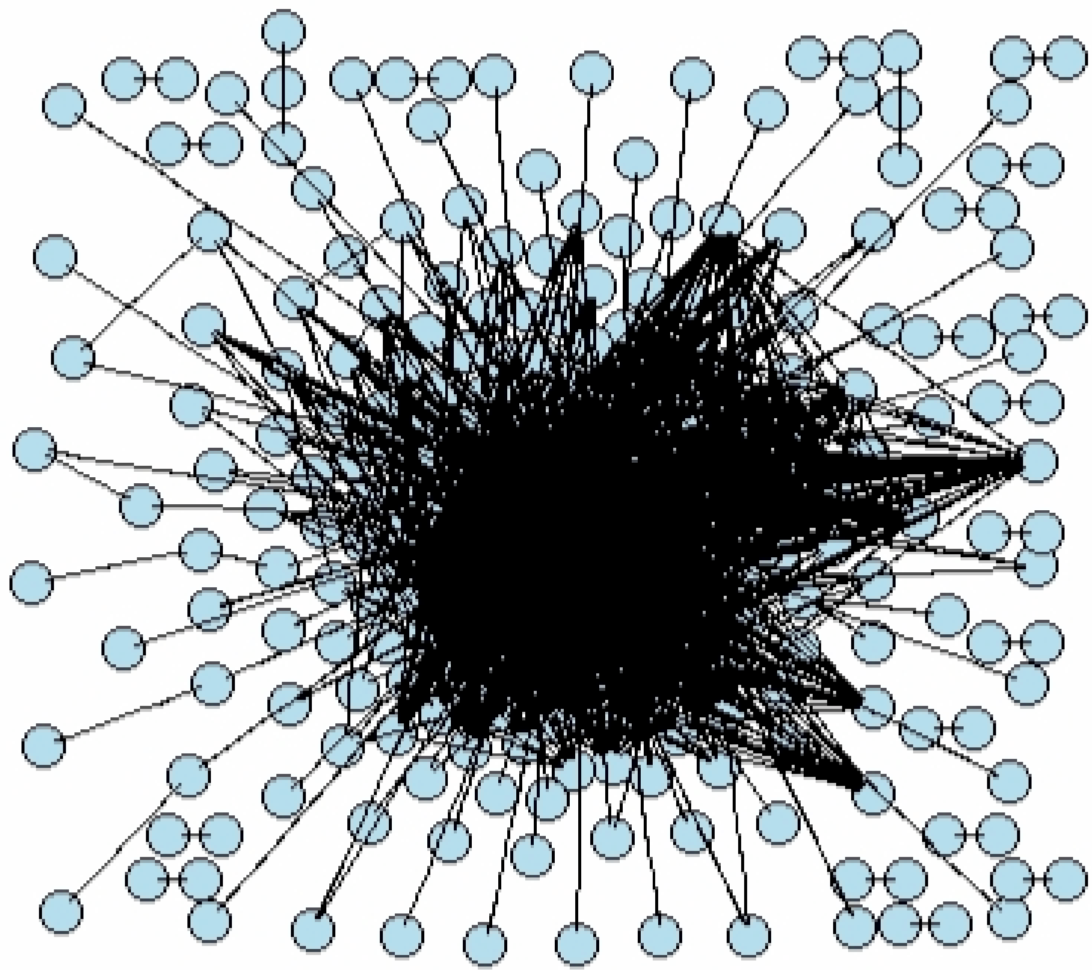
From “Financial Communities” - Evidence of a Power-Law structure

	Overlap Threshold K								
YYYY-MM	1	2	5	10	25	50	100	200	500
2001-01	1779,1	1,45	1,657	1,799	1,745	1,476	1,264	1,104	1,33
		1734,1	2,9	2,13	2,12	2,3	2,4	14,1	
			977,1	3,1	3,1	3,1	30,1		
				534,1	176,1	90,1			
2001-02	1828,1	1,41	1,848	11,038	1,734	1,549	1,259	1,86	1,19
		2,2	2,10	2,12	2,19	2,7	2,2	11,1	2,1
		1783,1	922,1	517,1	3,2	72,1	19,1		
					241,1				
2001-03	1,4	1,119	1,569	1,613	1,506	1,433	1,220	1,90	1,25
	1690,1	2,2	2,12	2,14	2,15	2,8	2,2	9,1	
		5,1	3,1	543,1	3,1	3,1	3,1		
		1566,1	849,1		282,1	68,1	23,1		
2001-04	1,8	1,322	1,668	1,655	1,534	1,373	1,176	1,71	1,17
	1582,1	2,7	2,14	2,15	2,9	2,8	2,3	6,1	
		1254,1	3,1	4,1	3,3	3,2	15,1		
			664,1	412,1	125,1	36,1			

(a,b) : a = size of community, b = number of communities of size a.

Figure 3: **Financial Community Network Graph**

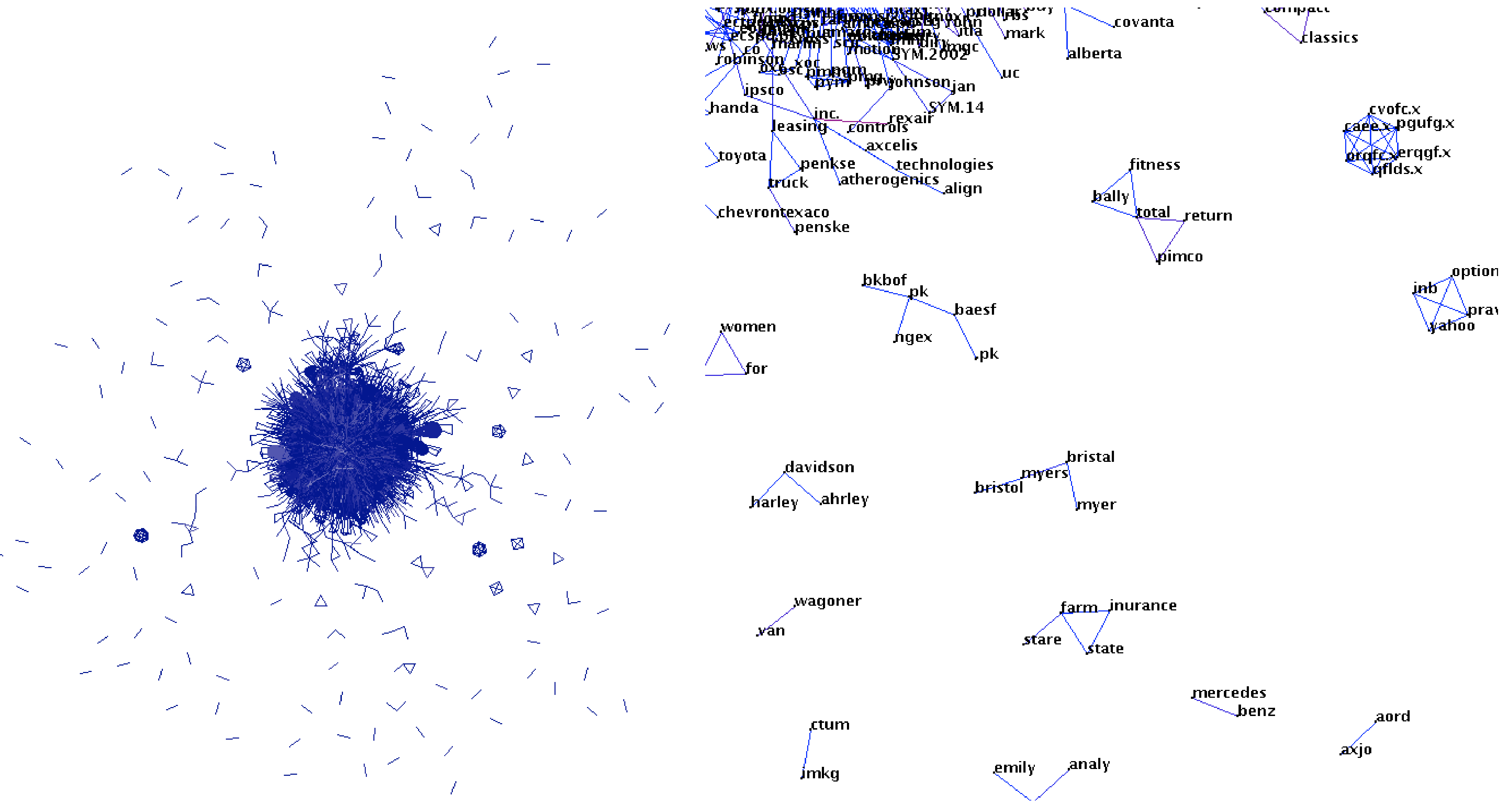
This plot represents all stocks for February 2001 with at least one or more links to others with a connection threshold of $K = 25$. Stocks with no connectedness are not represented on the graph. The plot depicts the structure of one large financial community.



A rendering of a graph of the 6k+ stocks for which someone requested a quote from Yahoo finance. There is an edge between two stocks if someone requested quotes on those stocks at the same time. They are from about 2% of the traffic on Yahoo, on April 1, 2002.

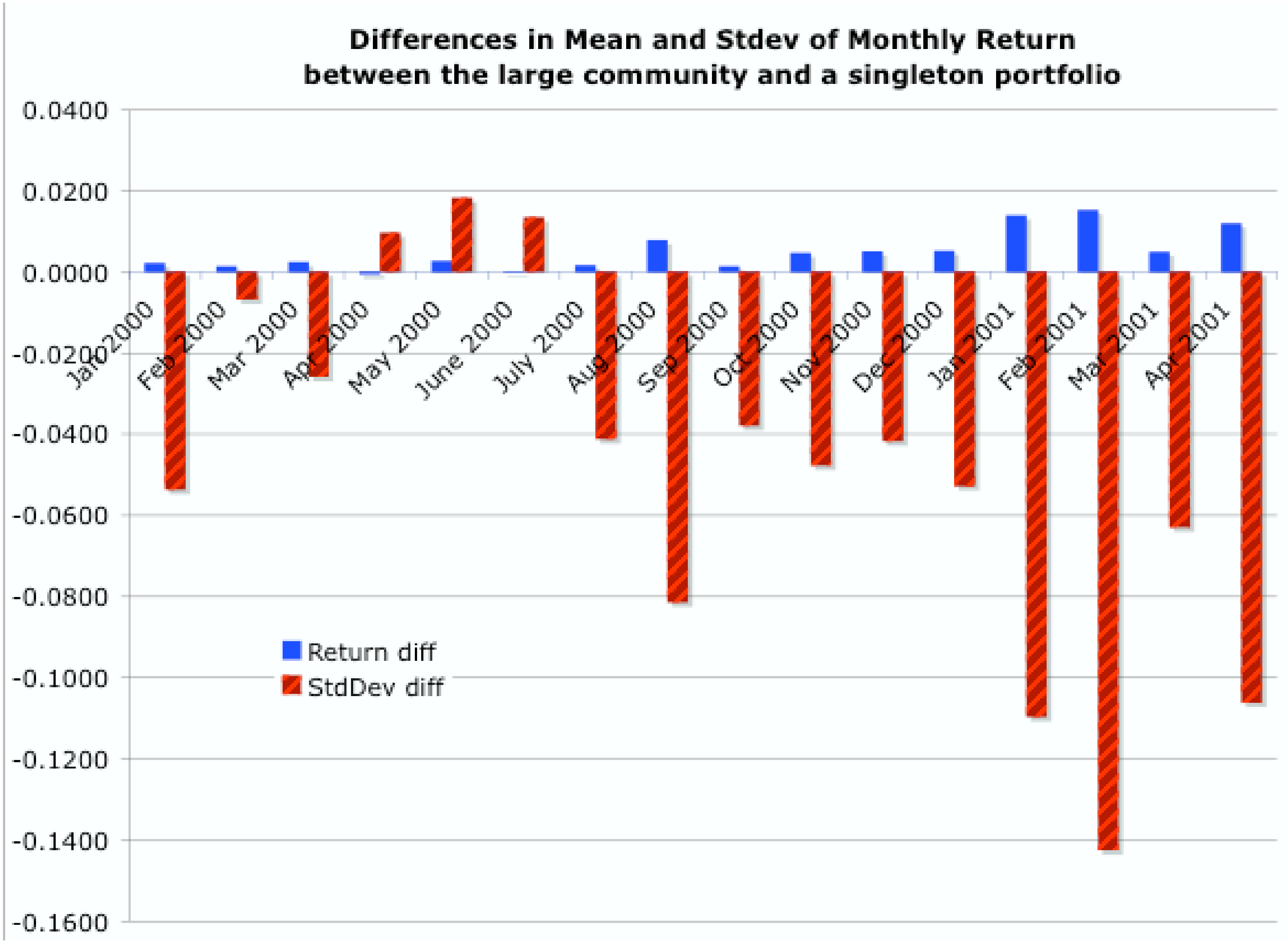
Based on rendering software by: Adai AT, Date SV, Wieland S, Marcotte EM. LGL: creating a map of protein function with an algorithm for visualizing very large biological networks. J Mol Biol. 2004 Jun 25;340(1):179-90.

Courtesy, Jacob Sisk, Yahoo Search Marketing



Communities and Returns

Period	Community			Singletons			Statistics	
YYYY-MM	Mean	Stdev	NoObs	Mean	Stdev	NoObs	Ratio	T-Test
2000-01	0.000141	0.000366	419	0.000064	0.001116	396	2.1997	18.3519
2000-02	0.000077	0.000416	426	0.000055	0.000531	397	1.3869	9.1479
2000-03	0.000187	0.000619	620	0.000160	0.001125	542	1.1650	8.1076
2000-04	0.000303	0.000597	519	0.000165	0.000345	413	1.8331	68.3173
2000-05	0.000148	0.001346	688	0.000082	0.000746	624	1.8086	20.4330
2000-06	0.000036	0.000753	831	0.000031	0.000513	528	1.1646	2.7418
2000-07	0.000047	0.000658	931	0.000050	0.001974	450	0.9514	-0.3844
2000-08	0.000030	0.000395	808	0.000003	0.002775	551	9.3277	3.8002
2000-09	0.000054	0.000365	655	0.000044	0.001272	620	1.2253	3.3068
2000-10	0.000169	0.000719	860	0.000109	0.001968	526	1.5572	11.1434
2000-11	0.000089	0.001532	866	0.000052	0.004031	591	1.7327	3.7871
2000-12	0.000212	0.000920	917	0.000129	0.002256	483	1.6456	12.3207
2001-01	0.000095	0.001077	928	0.000079	0.005708	621	1.1998	1.2023
2001-02	0.000082	0.000402	871	0.000060	0.005928	776	1.3589	1.9920
2001-03	0.000250	0.000583	801	0.000129	0.002142	539	1.9356	21.0984
2001-04	0.000290	0.000571	623	0.000195	0.003774	640	1.4854	11.2060


Differences in Mean and Stdev of Monthly Return
between the large community and a singleton portfolio
0.0400
0.0200
0.0000
-0.0200
-0.0400
-0.0600
-0.0800
-0.1000
-0.1200
-0.1400
-0.1600
Jan 2000
Feb 2000
Mar 2000
Apr 2000
May 2000
June 2000
July 2000
Aug 2000
Sep 2000
Oct 2000
Nov 2000
Dec 2000
Jan 2001
Feb 2001
Mar 2001
Apr 2001
Return diff
StdDev diff

Centrality: "Big Mouth" Effect?

Given m message boards or nodes, we compute an adjacency matrix $\mathbf{A} = \{\mathbf{a_{ij}}\} \in \Re^{\mathbf{m}\times\mathbf{m}}$, where a_{ij} is the information overlap between boards i and j, i.e. the number of common posters on boards i, j. In this setting, we defined connections as existent when the threshold level was $K = 5$. We define $\mathbf{x} \in \Re^{\mathbf{m}}$ as the vector of centrality scores. Since each element of this vector is a function of all the other elements (the centrality of a message board is a function of the centrality of other message boards), we may write the equation system as:

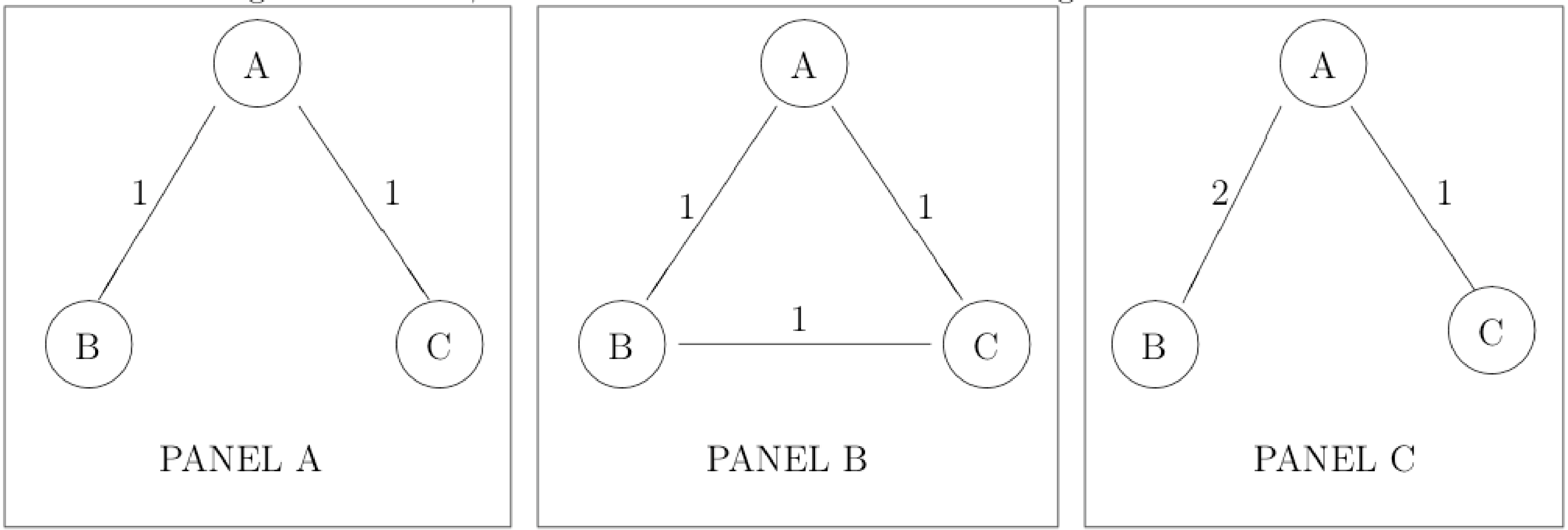
$$\lambda x_i = \sum_{j \neq i} a_{ij} x_j, \quad \forall\, i = 1...m. \tag{1}$$

The parameter λ is a scaling coefficient. If this is written in matrix form, we obtain

$$\lambda \mathbf{x} = \mathbf{A}\mathbf{x}. \tag{2}$$

This equation parallels the definition of an eigensystem. Hence, it has a solution under mild technical conditions. The solution to equation (2) provides a set of m eigenvalues λ, along

Figure 4: **Centrality**
This figure presents the 3 matrices in the centrality section in graphical form. Panel A depicts a hub-and-spoke network. Panel B is a balanced triangular network, and Panel C is an unbalanced triangular network.



$$\begin{bmatrix} 0 & 1 & 1 \\ 1 & 0 & 0 \\ 1 & 0 & 0 \end{bmatrix} \qquad \begin{bmatrix} 0 & 1 & 1 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \end{bmatrix} \qquad \begin{bmatrix} 0 & 2 & 1 \\ 2 & 0 & 0 \\ 1 & 0 & 0 \end{bmatrix}$$

Centrality scores = {0.71, 0.50, 0.50}

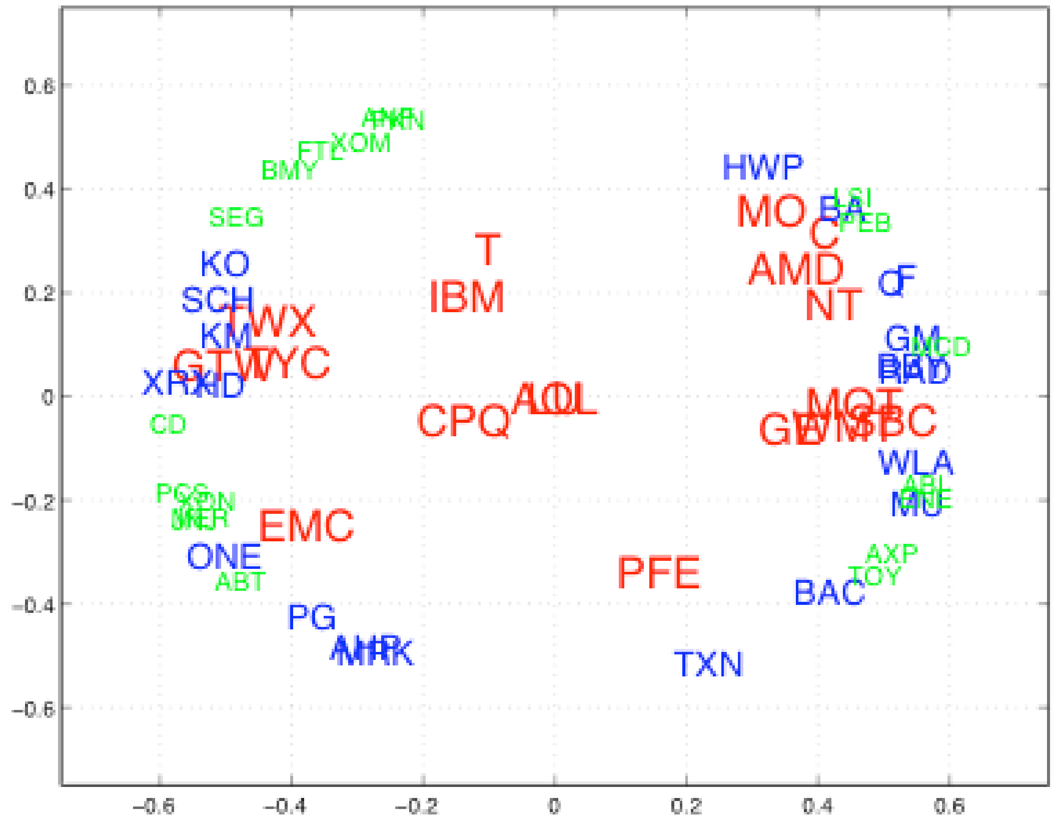
Centrality scores = {0.58, 0.58, 0.58}

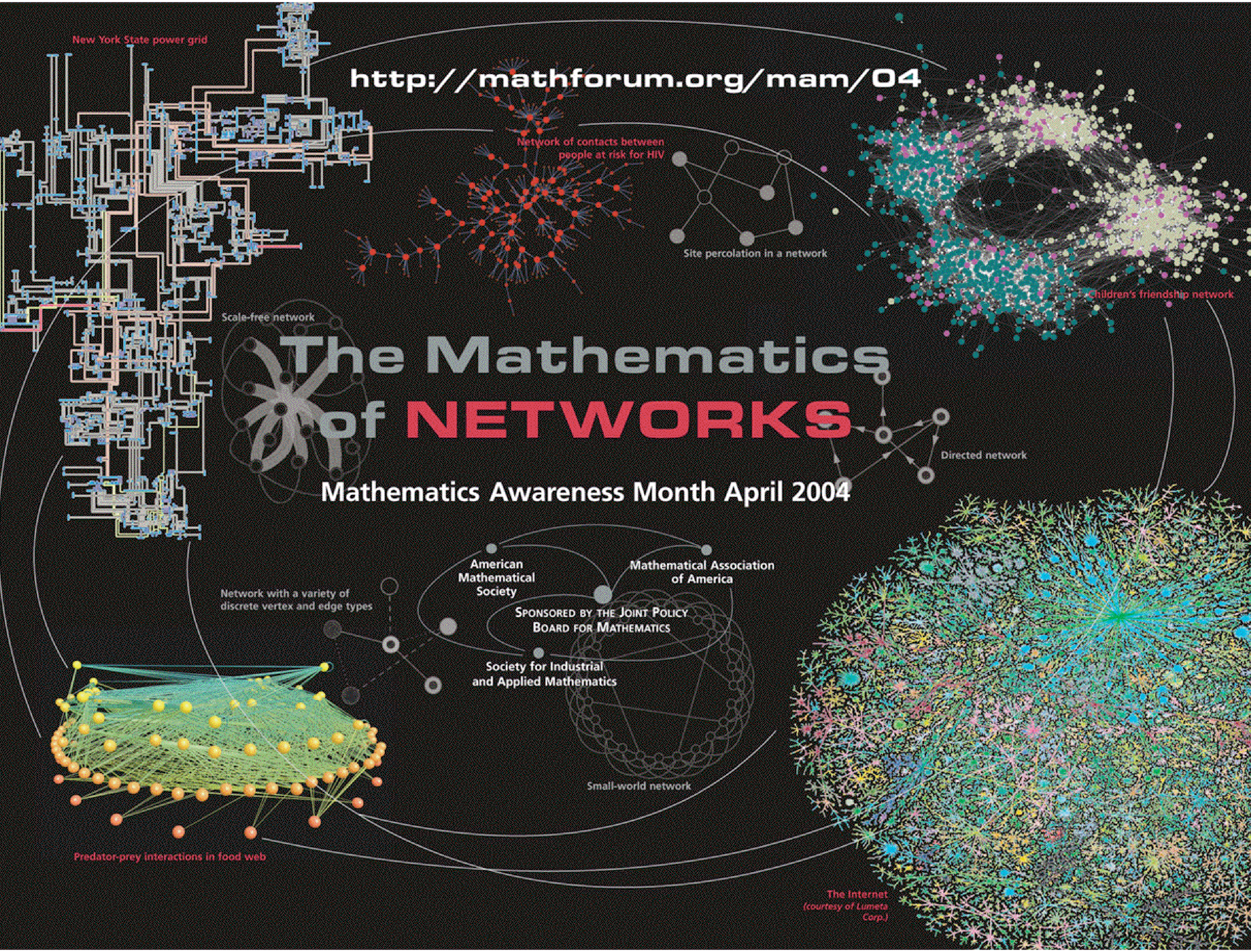
Centrality scores = {0.71, 0.63, 0.32}

Centrality scores = eigenvector for biggest eigenvalue.

Connectedness & Centrality

This plot presents a visual depiction of the connectedness and centrality of individual stock tickers. The location of the ticker on the graph is a function of the degree of connectedness. The most highly connected stocks are placed in the center of the graph, and the distance of the stock from the center reflects declining connectedness. Connectedness is measured by the number of common posters that a stock has with other stocks, and the distances are scaled to reflect the standard deviation of the sample on a (-1,+1) grid. Centrality is reflected by the size of the ticker symbol on the graph, with a large size for the top third of the tickers, medium size for the middle third, and small size for the bottom third. If viewed on screen, the large, medium and small centrality groups are depicted by the colors red, blue and green. To avoid clutter on the graph, connectedness is reflected only for a threshold level of $K = 50$. The plot covers the month of January 2000.




http://mathforum.org/mam/04
The Mathematics
of NETWORKS
Mathematics Awareness Month April 2004
New York State power grid
Network of contacts between people at risk for HIV
Site percolation in a network
Children's friendship network
Scale-free network
Directed network
American Mathematical Society
Mathematical Association of America
Sponsored by the Joint Policy Board for Mathematics
Society for Industrial and Applied Mathematics
Network with a variety of discrete vertex and edge types
Small-world network
Predator-prey interactions in food web
The Internet
(courtesy of Lumeta Corp.)

Systemic Risk & Co-lending Networks

An Application based on IBM's MIDAS System

Joint work with Doug Burdick, Mauricio Hernandez, Howard Ho, Georgia Koutrika, Rajasekar Krishnamurthy, Lucian Popa, Ioana R. Stanoi, & Shivakumar Vaithyanathan (IBM)

See the paper at:
http://algo.scu.edu/~sanjivdas/midas-deb_July2011.pdf
Older version:
http://algo.scu.edu/~sanjivdas/midaswww2011_FINAL.pdf

Overview

- Focus on financial companies that are the domain for systemic risk (SIFIs).
- Extract information from unstructured text (filings).
- Information can be analyzed at the institutional level or aggregated system-wide.
- Applications: Systemic risk metrics; governance.
- Technology: information extraction (IE), entity resolution, mapping and fusion, scalable Hadoop architecture.

Financial Applications

- Aggregate level: Co-lending network analysis to construct a measure of systemic risk.
- Institutional level: (a) key executives and their governance history, (b) insider transactions, (c) relationships of the institution to other entities, such as subsidiaries, cross-holdings, executive connections, and borrowing/ lending activites.

Midas provides an entity view around new sources of data

- Extraction and cleansing of financial entities, their resolution and linkage across multiple sources
- Uncovering non-obvious relationships between financial entities
- Computation of key financial metrics using data extracted from multiple sources of public data
- Information analyzed at the institutional level or aggregated system-wide.

Web Data

News

Blogs

Reviews

Public Data

FDIC Call Data Records

SEC Filings

OTS Thrift Financial Records

Private Data

Hoovers

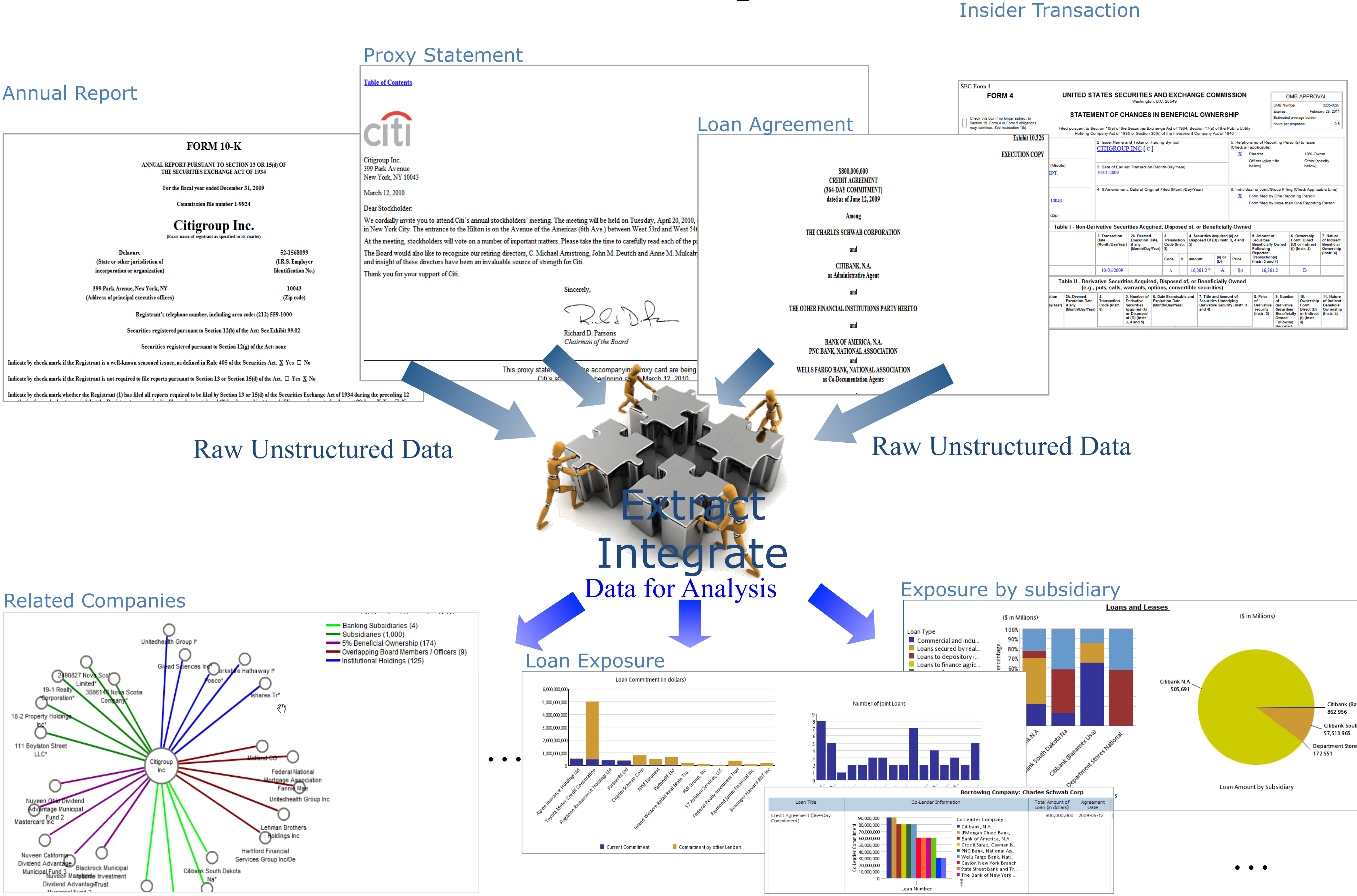
D&B

FINRA

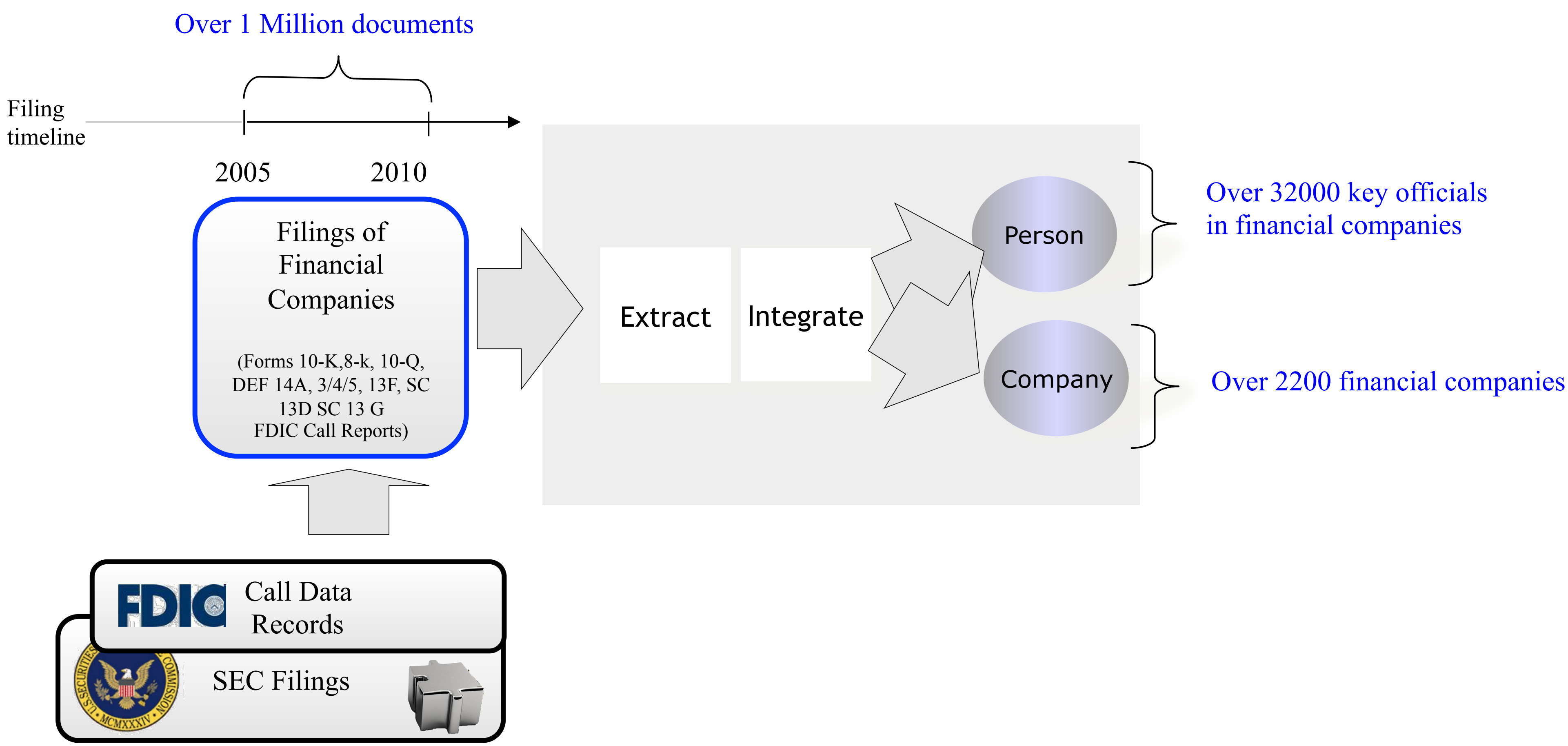
Private Wall Street Journal

- Regulators
- Credit committees
- Investment analysts
- Portfolio managers
- Equity managers

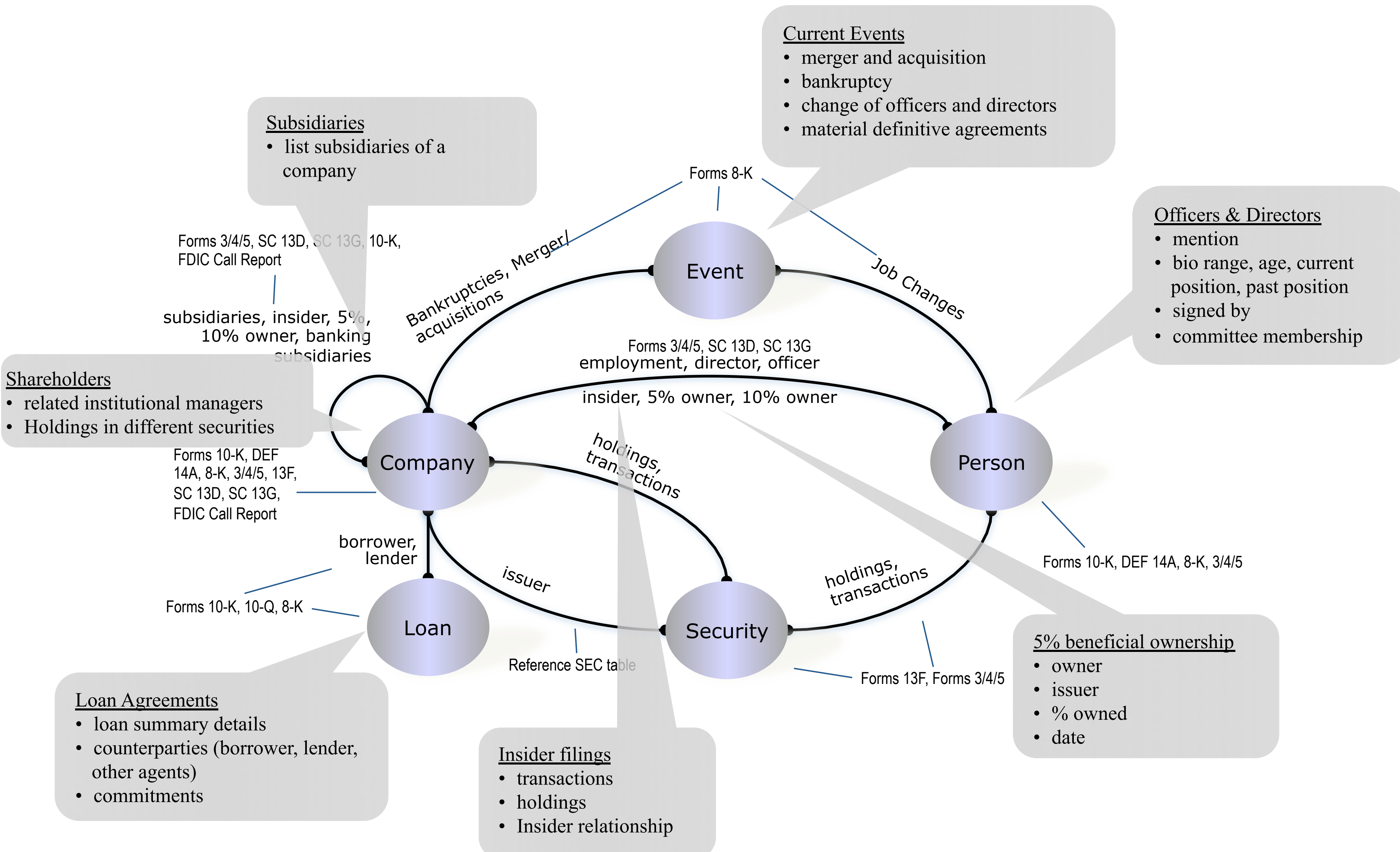
Midas Financial Insights



Example of Midas Financial Insights



Midas provides Analytical Insights into company relationships by exposing information concepts and relationships within extracted concepts



Example Analysis : Extraction of Loan Information Data

Extract and cleanse information from headers, tables main content and signatures

$800,000,000
CREDIT AGREEMENT
(364-DAY COMMITMENT)
dated as of June 12, 2009

Among

THE CHARLES SCHWAB CORPORATION

and

CITIBANK, N.A.
as Administrative Agent

and

THE OTHER FINANCIAL INSTITUTIONS PARTY HERETO

LENDERS' COMMITMENTS

The Charles Schwab Corporation $800,000,000 Credit Agreement (364-Day Commitment) dated as of June 12, 2009.

				Lender Commitment Amount
1.	Citibank, N.A.	1.	$	90,000,000
2.	JPMorgan Chase Bank, N.A.	2.	$	90,000,000
3.	Bank of America, N.A.	3.	$	80,000,000
4.	PNC Bank, National Association	4.	$	80,000,000
5.	Wells Fargo Bank, National Association	5.	$	80,000,000
6.	Credit Suisse, Cayman Islands Branch	6.	$	80,000,000
7.	The Bank of New York Mellon	7.	$	60,000,000
8.	Calyon New York Branch	8.	$	60,000,000
9.	State Street Bank and Trust Company	9.	$	60,000,000
10.	UBS Loan Finance LLC	10.	$	60,000,000
11.	Comerica Bank	11.	$	30,000,000
12.	Lloyds TSB Bank plc	12.	$	30,000,000
		Total	$	800,000,000

Lenders:

CITIBANK, N.A., as Agent and individually as Lender

By: Maureen P. Maroney
Name: Maureen P. Maroney
Title: Vice President

JPMORGAN CHASE BANK, N.A.

By: Catherine Grossman
Name: Catherine Grossman
Title: Vice President

BANK OF AMERICA, N.A.

By: Garfield Johnson
Name: Garfield Johnson
Title: Senior Vice President

Id	Agreement Name	Date	Total Amount
1	Credit Agreement	June 12, 2009	$800,000,000
…			

Loan Information

Id	Company	Role	Commitment
1	Charles Schwab Corporation	Borrower	
1	Citibank, N.A.	Administrative Agent	
1	Citibank, N.A.	Lender	$90,000,000
1	JPMorgan Chase Bank, N.A.	Lender	$90,000,000
1	Bank of America, N.A.	Lender	$80,000,000
…			

Loan Company Information

Notes: Loan Document filed by Charles Schwab Corporation On Aug 6, 2009

Example Analysis : Integration of person information across documents

Do these filings refer to the same person ?

- variability in the person’s name
- lack of a key identifier
- supporting attributes vary depending on the context (form type)

Who Is James Dimon?

position

history

committee membership

Director	Audit	Compensation & Management Development	Corporate Governance & Nominating	Public Responsibility	Risk Policy
Crandall C. Bowles	Member				
Stephen B. Burke		Member	Member		
David M. Cote				Member	Member
James S. Crown				Member	Chair
James Dimon					
Ellen V. Futter				Member	Member

Sincerely,

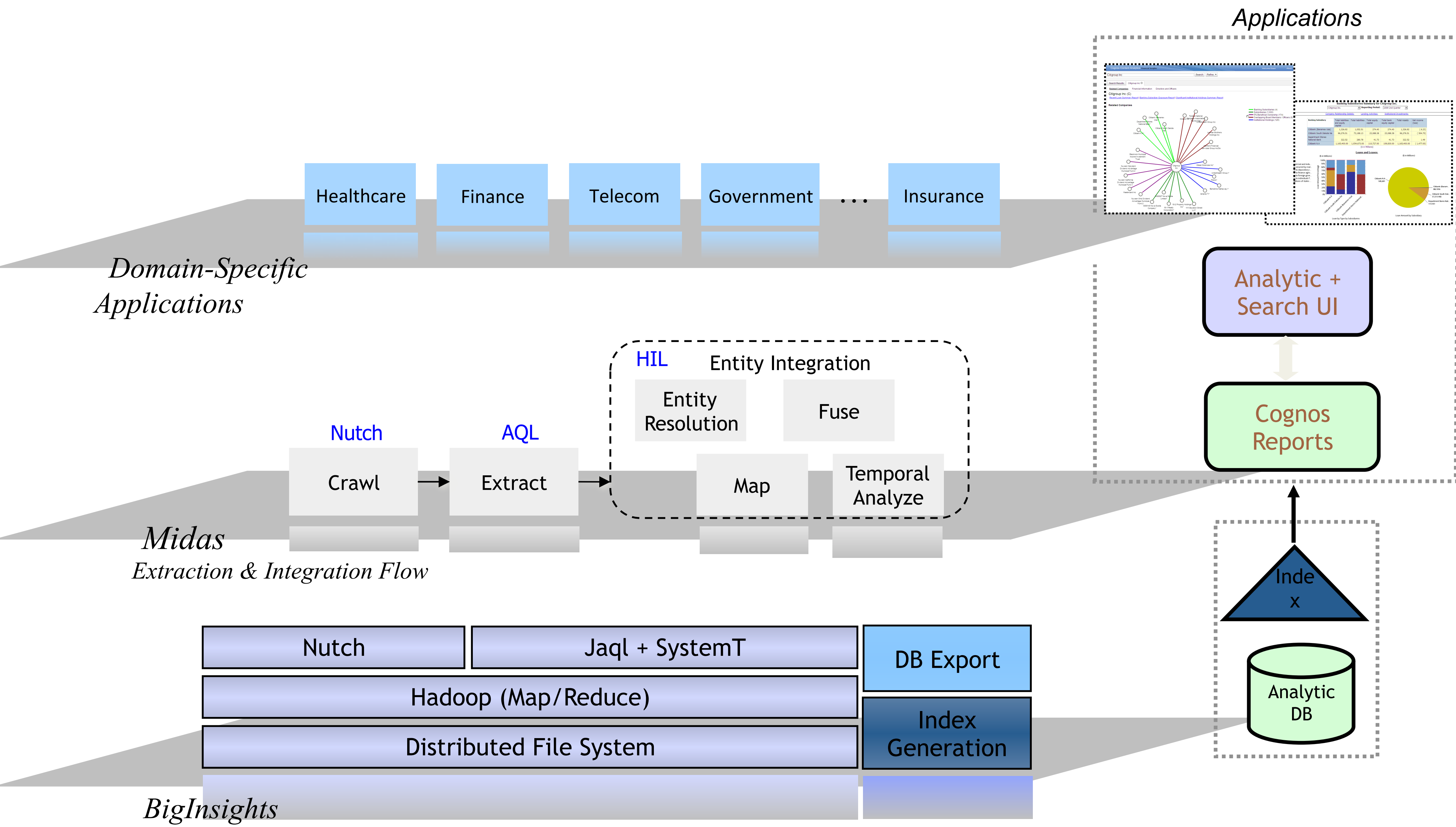
James Dimon
Chairman and Chief Executive O

Name	Grant date	Approval date	Stock awards: Number of shares of stock or units (#) (2)	Option awards: Number of securities underlying options (#)	Exercise price ($/Sh)	Grant date fair value ($)
James Dimon	1/22/2008	1/15/2008	364,048			$14,500,000
	1/22/2008 (3)	1/15/2008		2,000,000 (3)	$39.830	19,868,000 (3)
Michael J. Cavanagh (4)	1/22/2008	1/15/2008	94,151			3,750,000
	1/22/2008	1/15/2008		300,000	39.830	2,980,200
	1/30/2008	N/A		54,271	47.835	360,577
	1/30/2008	N/A		10,625	47.835	91,821
Frank J. Bisignano	1/22/2008	1/15/2008	94,151			3,750,000
	1/22/2008	1/15/2008		300,000	39.830	2,980,200

James Dimon, 53, Chairman and Chief Executive Officer of JPMorgan Chase. Director since 2000.

Mr. Dimon became Chairman of the Board on December 31, 2006, and has been Chief Executive Officer and President since December 31, 2005. He had been President and Chief Operating Officer since JPMorgan Chase’s merger with Bank One Corporation in July 2004. At Bank One he had been Chairman and Chief Executive Officer since March 2000. Mr. Dimon is a graduate of Tufts University and received an MBA from Harvard Business School. He is a director of The College Fund/UNCF and serves on the Board of Directors of The Federal Reserve Bank of New York, The National Center on Addiction and Substance Abuse, Harvard Business School and Catalyst. He is on the Board of Trustees of New York University School of Medicine.

Midas Architecture



Systemic Analysis

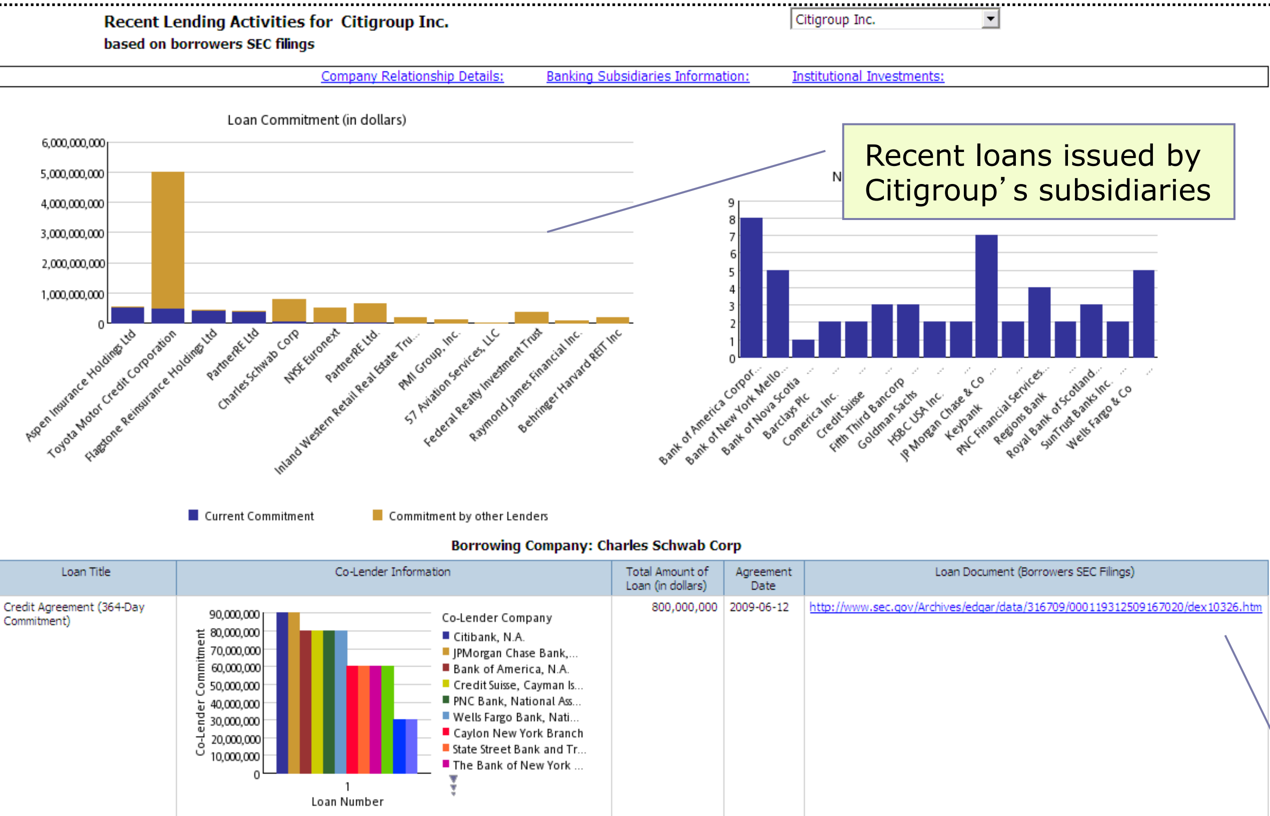
- Definition: the measurement and analysis of relationships across entities with a view to understanding the impact of these relationships on the system as a whole.
- Challenge: requires most or all of the data in the system; therefore, high-quality information extraction and integration is critical.

Systemic Risk

- Current approaches: use stock return correlations (indirect). [Acharya, et al 2010; Adrian and Brunnermeier 2009; Billio, Getmansky, Lo 2010; Kritzman, Li, Page, Rigobon 2010]
- Midas: uses semi-structured archival data from SEC and FDIC to construct a co-lending network; network analysis is then used to determine which banks pose the greatest risk to the system.

Analyze Recent Loan Issued by Counterparty

e.g. Loans issued by Citigroup



Recent loans issued by Citigroup's subsidiaries

Joint loans issued with other lenders

Link to borrower's original SEC filing

Loan details : lender names, commitments

Analyze the Aggregate Lending activity

e.g. drill down on Citigroup loan activity by subsidiaries (from FDIC Call Records)

Banking Subsidiaries Summary for Citigroup Inc.

Citigroup Inc. **Reporting Period:** 2009 2nd Quarter

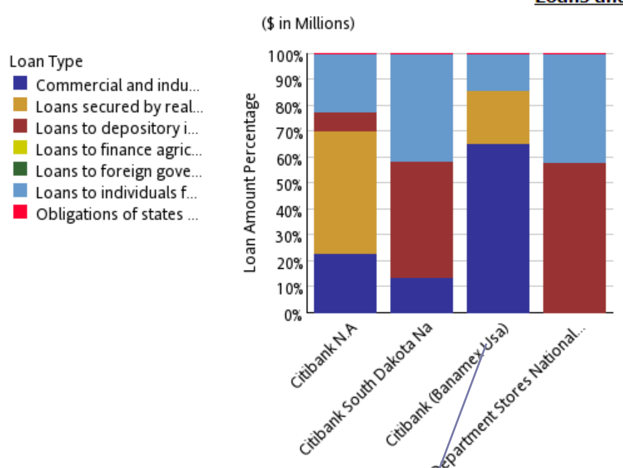
Company Relationship Details: Lending Activities: Institutional Investments:

Banking Subsidiary	Total liabilities and equity capital	Total liabilities	Total equity capital	Total bank equity capital	Total Assets	Net income (loss)
Citibank (Banamex Usa)	1,326.92	1,052.51	274.40	274.40	1,326.92	(6.23)
Citibank South Dakota Na	96,276.51	73,188.13	23,088.38	23,088.38	96,276.51	(504.70)
Department Stores National Bank	322.52	280.78	41.73	41.73	322.52	1.49
Citibank N.A	1,165,400.00	1,054,673.00	110,727.00	109,830.00	1,165,400.00	(1,477.00)

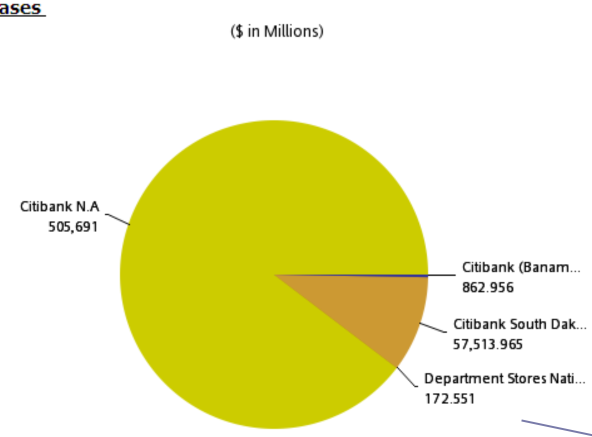
($ in Millions)

Key financial metrics for banking subsidiaries

Loans and Leases



Loan by Type by Subsidiaries



Loan Amount by Subsidiary

Breakdown of overall loans outstanding by subsidiary

Breakdown by type of loan

Co-lending Network

- Definition: a network based on links between banks that lend together.
- Loans used are *not* overnight loans. We look at longer-term lending relationships.
- Lending adjacency matrix: $\mathbf{L} \equiv \{L_{ij}\}, i, j = 1 \ldots N$
- Undirected graph, i.e., symmetric $\mathbf{L} \in R^{N \times N}$
- Total lending impact for each bank: $x_i, i = 1 \ldots N.$

Centrality

- Influence relations are circular:

$$x_i = \sum_{j=1}^{N} L_{ij} x_j, \forall i.$$

$$\mathbf{x} = \mathbf{L} \cdot \mathbf{x}, \text{ where } \mathbf{x} = [x_1, x_2, \ldots, x_N]' \in R^{N \times 1}$$

- Pre-multiply by scalar to get an eigensystem:

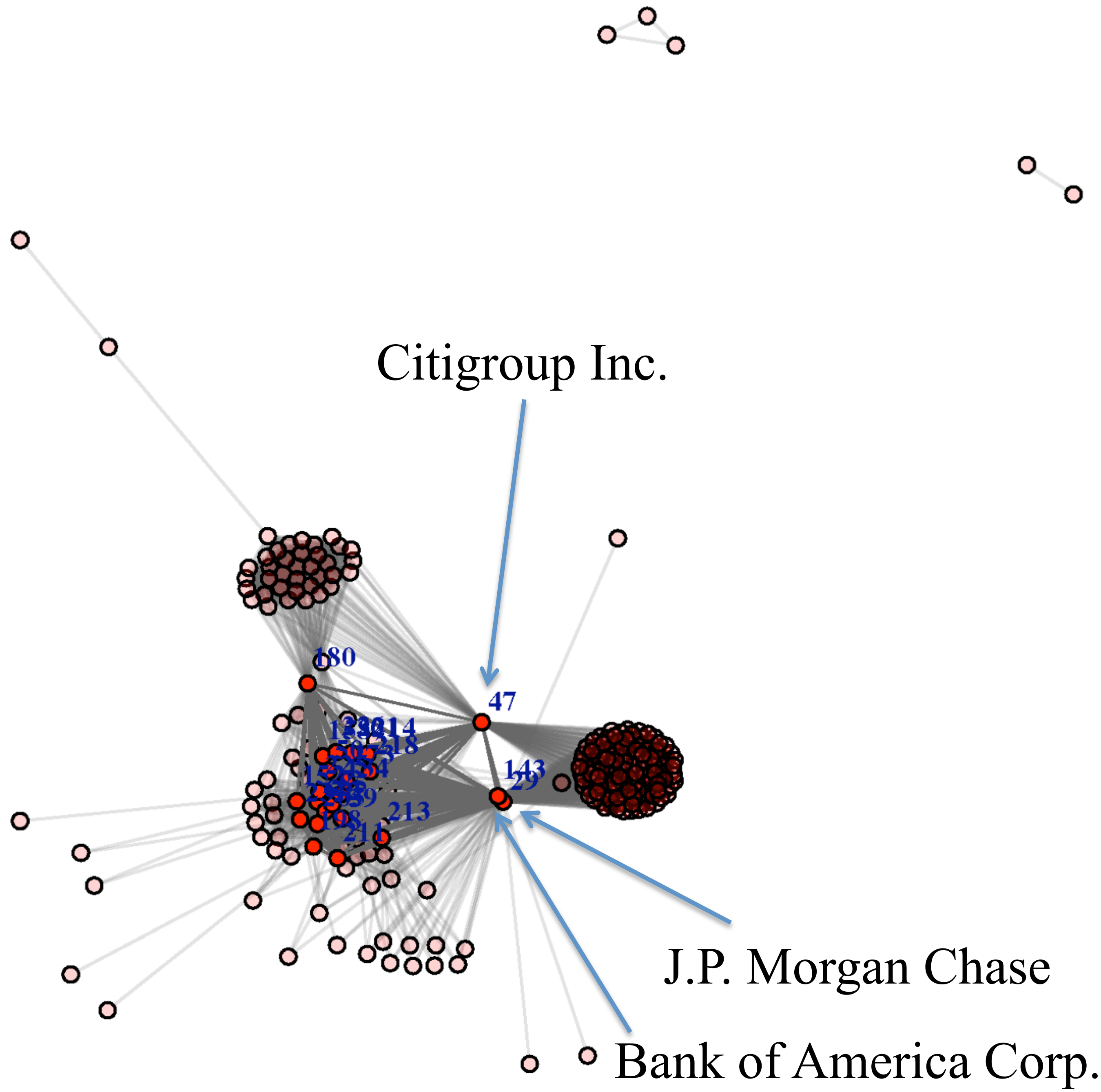
$$\lambda \, \mathbf{x} = \mathbf{L} \cdot \mathbf{x}$$

- Principal eigenvector of this system gives the "centrality" score for a bank.
- This score is a measure of the systemic risk of a bank.

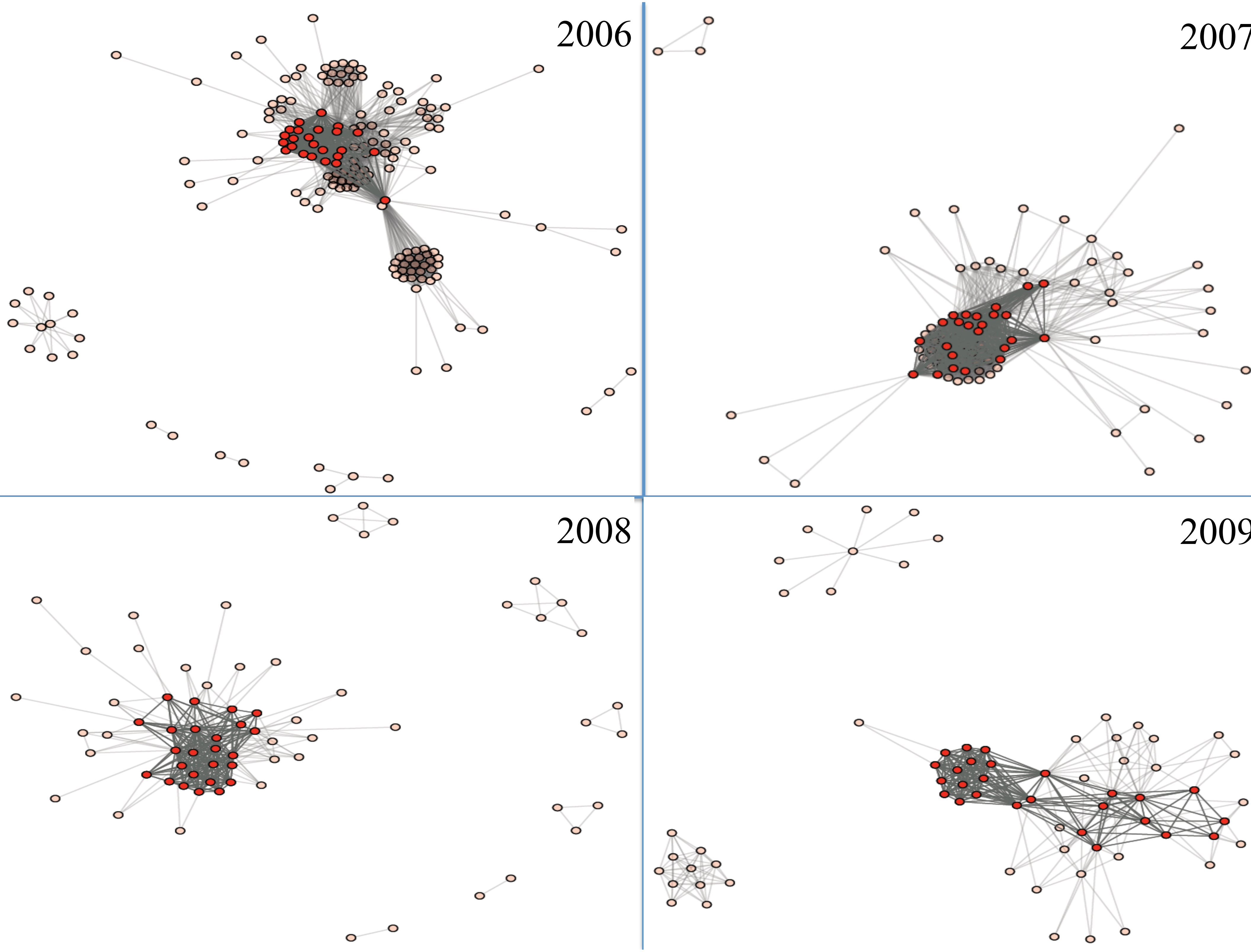
Data

- Five years: 2005—2009.
- Loans between FIs only.
- Filings made with the SEC.
- No overnight loans.
- Example: 364-day bridge loans, longer-term credit arrangement, Libor notes, etc.
- Remove all edge weights < 2 to remove banks that are minimally active. Remove all nodes with no edges. (This is a choice for the regulator.)

2005



Citigroup Inc.
180
47
143
29
213
J.P. Morgan Chase
Bank of America Corp.


2006
2007
2008
2009

Network Fragility

- Definition: how quickly will the failure of any one bank trigger failures across the network?
- Metric: expected degree of neighboring nodes averaged across all nodes.

$E(d^2)/E(d) \equiv R$, where d stands for the degree of a node.

- Neighborhoods are expected to "expand" when $R \geq 2$
- Metric: diameter of the network.

Top 25 banks by systemic risk

Year	#Colending banks	#Coloans	Colending pairs	$R = E(d^2)/E(d)$	Diam.
2005	241	75	10997	137.91	5
2006	171	95	4420	172.45	5
2007	85	49	1793	73.62	4
2008	69	84	681	68.14	4
2009	69	42	598	35.35	4

(Year = 2005)

Node #	Financial Institution	Normalized Centrality
143	J P Morgan Chase & Co.	1.000
29	Bank of America Corp.	0.926
47	Citigroup Inc.	0.639
85	Deutsche Bank Ag New York Branch	0.636
225	Wachovia Bank NA	0.617
235	The Bank of New York	0.573
134	Hsbc Bank USA	0.530
39	Barclays Bank Plc	0.530
152	Keycorp	0.524
241	The Royal Bank of Scotland Plc	0.523
6	Abn Amro Bank N.V.	0.448
173	Merrill Lynch Bank USA	0.374
198	PNC Financial Services Group Inc	0.372
180	Morgan Stanley	0.362
42	Bnp Paribas	0.337
205	Royal Bank of Canada	0.289
236	The Bank of Nova Scotia	0.289
218	U.S. Bank NA	0.284
50	Calyon New York Branch	0.273
158	Lehman Brothers Bank Fsb	0.270
213	Sumitomo Mitsui Banking	0.236
214	Suntrust Banks Inc	0.232
221	UBS Loan Finance Llc	0.221
211	State Street Corp	0.210
228	Wells Fargo Bank NA	0.198

Possible Next Steps

- Analyze the lending network via a directed graph. Power in the system?
- Would the regulators be interested in circulating this measure to banks and the public? At what frequency?
- Are there other domains of supervision in which a similar analysis might be useful? What data does it need? Central information clearing warehouses?

Venture Capital Communities

Joint work with
Amit Bubna (Indian School of Business);
N.R. Prabhala (University of Maryland)

Communities: Multi-Disciplinary Applications

- Biology
  - Metabolic networks of cellular organisms (Duch and Arenas, 2005)
  - Community structure of the human brain (Wu et al, 2011)
  - Compartmentalization of food chain webs (Dunne, 2006)

- Political Science
  - Political preferences through voting patterns (Porter et al, 2007)

- Social interaction
  - Mobile phone and online networks (Porter et al, 2009)
  - Collaboration between scientists (Newman, 2001)

Syndication

- The VC Market
  - 56,000 deals, $146 billion from 1980-1999
  - 39,002 deals, $316 billion from 2000-2010

- Syndication
  - 44% of # deals
  - 66% of amount invested

- There is a large literature showing that syndicate-financed ventures perform better. Some of the performance comes from individual influence - centrality: Hochberg, Ljungqvist, Lu (JF 2007).
- But does team work through repeated interaction play a role? What is the deeper structure of VC syndicates?

- s

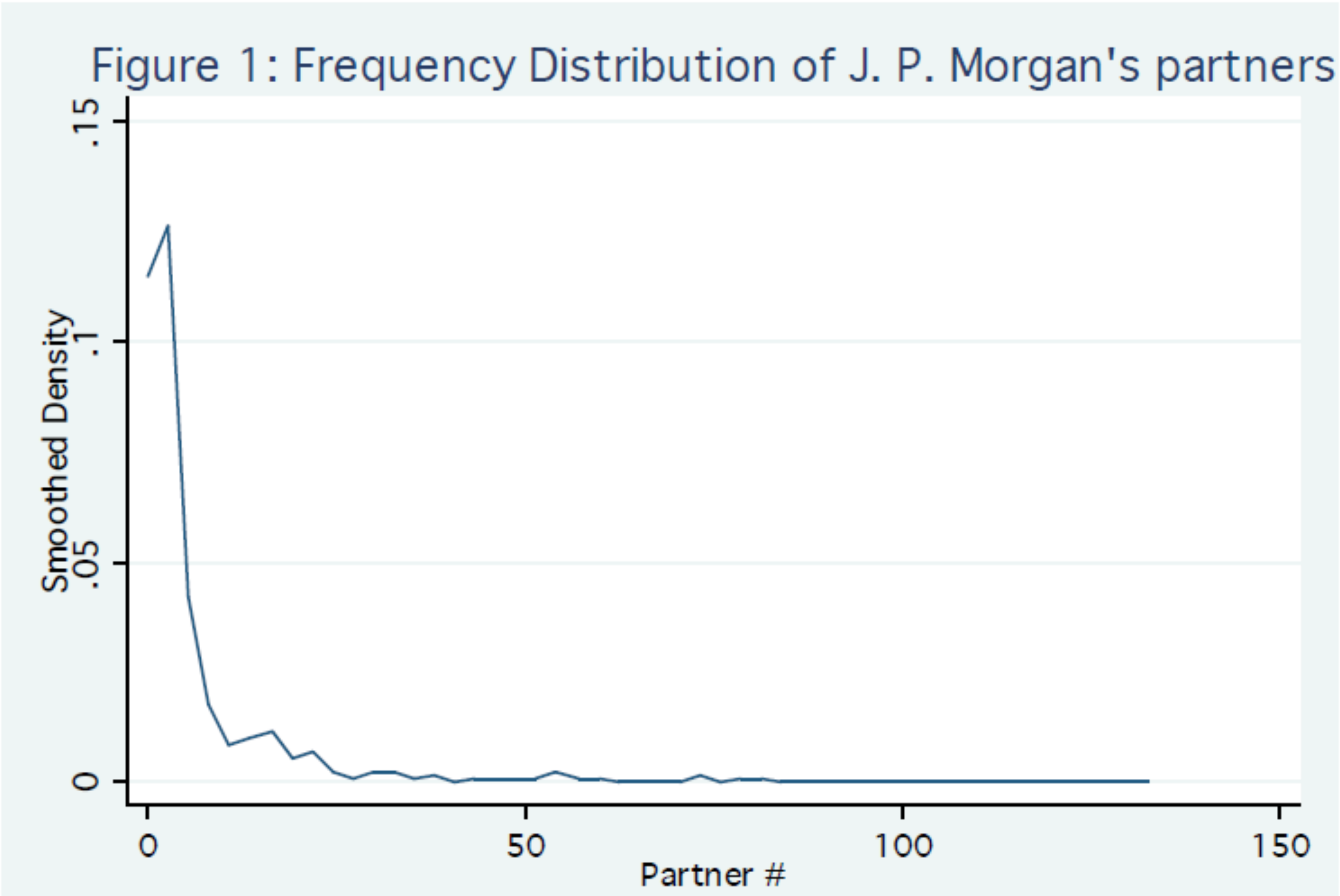
Choosing Syndication Partners

- If partners are chosen at random
  - Spatially diffuse VC network

- If VCs have preferred partners
  - Spatial clustering of VC networks

- We term spatial clusters as VC **communities**.

Example: J. P. Morgan



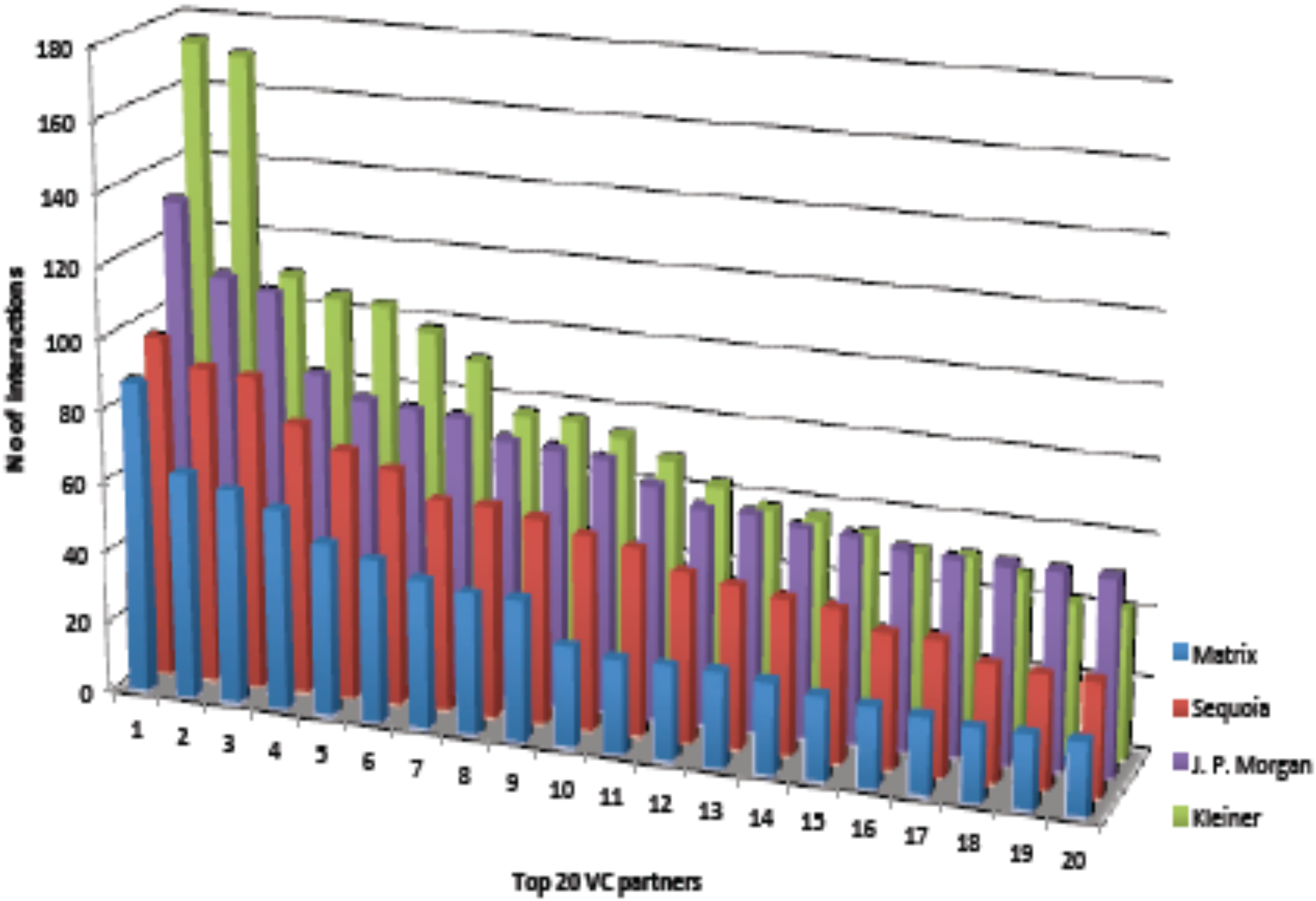


Figure 2: Distribution of the number of interactions of four top firms with their top 20 collaborators.

Why are Communities Important?

- Syndication: acquiring or improving skills

- How does familiarity help?
  - Familiar norms, processes, and people [Gertler (1995), Porter (2000)]
  - Flow of informal knowledge
  - Mitigates incomplete contracting problems, builds trust and enhances reciprocity [Guiso, Sapienza and Zingales (2004); Bottazzi, Da Rin and Hellmann (2011)]

- Pure transaction cost effect
  - Less administrative overheads and paperwork
  - Behavioral affinity for the familiar

- Knowledge spillovers through repeated interactions
  - Acquiring or improving skills
  - Learning facilitated through familiar norms, processes, and people

- Resource sharing without burden of organizational inflexibility

- Economics literature on clustering
  - Krugman (1991), Porter (1998): new organizational paradigm to capture benefits of externalities.
  - Lindsey (2007): VCs blur boundaries between portfolio firms. Communities similarly blur lines between ***VCs***.

Detecting Communities

- Community identification should
  - Accommodate large number of players
  - Not pre-specify the # of communities
  - Allow for VC communities of varying sizes
  - Permit fuzzy boundaries between communities

- This is a computationally hard clustering problem.

- Modularity optimization
  - Modularity – strength of internal ties compared to ties outside (Girvan and Newman, 2003)
  - We implement an agglomerative algorithm
    - "Walktrap" algorithm (Pons and Latapy, 2005)

Community
Mathematical Construct

- Adjacency matrix of a graph A
  - A [*i*,*j*] = n_{ij}
  - n_{ij} = # syndicates involving VC *i* and VC *j*.

- Partition, P, divides A into collections of nodes, P = (P_1, P_2, … P_n)
  - mutually exclusive and collectively exhaustive

- The best community structure maximizes in-community deals relative to the predicted in-community deals, or the *modularity*

$$\text{Modularity } (P) = \sum_{p_n} \sum_{i,j \in P_n} \left[A_{ij} - \frac{k_i k_j}{2m} \right]$$

where, k_i = # syndicates involving VC *i* and *m* = # deals in P_n

Example

Consider a network of five nodes $\{A, B, C, D, E\}$, where the edge weights are as follows: $A : B = 6$, $A : C = 5$, $B : C = 2$, $C : D = 2$, and $D : E = 10$. Assume that a community detection algorithm assigns $\{A, B, C\}$ to one community and $\{D, E\}$ to another, i.e., only two communities. The adjacency matrix for this graph is

$$\{A_{ij}\} = \begin{bmatrix} 0 & 6 & 5 & 0 & 0 \\ 6 & 0 & 2 & 0 & 0 \\ 5 & 2 & 0 & 2 & 0 \\ 0 & 0 & 2 & 0 & 10 \\ 0 & 0 & 0 & 10 & 0 \end{bmatrix}$$

The Kronecker delta matrix that delineates the communities will be

$$\{\delta_{ij}\} = \begin{bmatrix} 1 & 1 & 1 & 0 & 0 \\ 1 & 1 & 1 & 0 & 0 \\ 1 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 & 1 \end{bmatrix}$$

The modularity score is

$$Q = \frac{1}{2m} \sum_{i,j} \left[A_{ij} - \frac{d_i \times d_j}{2m} \right] \cdot \delta_{ij} \qquad (2)$$

where $m = \frac{1}{2}\sum_{ij} A_{ij} = \frac{1}{2}\sum_i d_i$ is the sum of edge weights in the graph, A_{ij} is the (i, j)-th entry in the adjacency matrix, i.e., the weight of the edge between nodes i and j, and $d_i = \sum_j A_{ij}$ is the degree of node i. The function δ_{ij} is Kronecker's delta and takes value 1 when the nodes i and j are from the same community, else takes value zero. The core of the formula comprises the modularity matrix $\left[A_{ij} - \frac{d_i \times d_j}{2m}\right]$ which gives a score that increases when the number of connections within a community exceeds the expected proportion of connections if they are assigned at random depending on the degree of each node. The score takes a value ranging from -1 to $+1$ as it is normalized by dividing by $2m$. When

Quick R

```
> A = matrix(c(0,6,5,0,0,6,0,2,0,0,5,2,0,2,0,0,0,2,0,10,0,0,0,10,0),5,5)
> delta = matrix(c(1,1,1,0,0,1,1,1,0,0,1,1,1,0,0,0,0,0,1,1,0,0,0,1,1),5,5)
> print(Amodularity(A,delta))
[1] 0.4128
```

```
> g = graph.adjacency(A,mode="undirected",weighted=TRUE,diag=FALSE)
```

We then pass this graph to the walktrap algorithm:

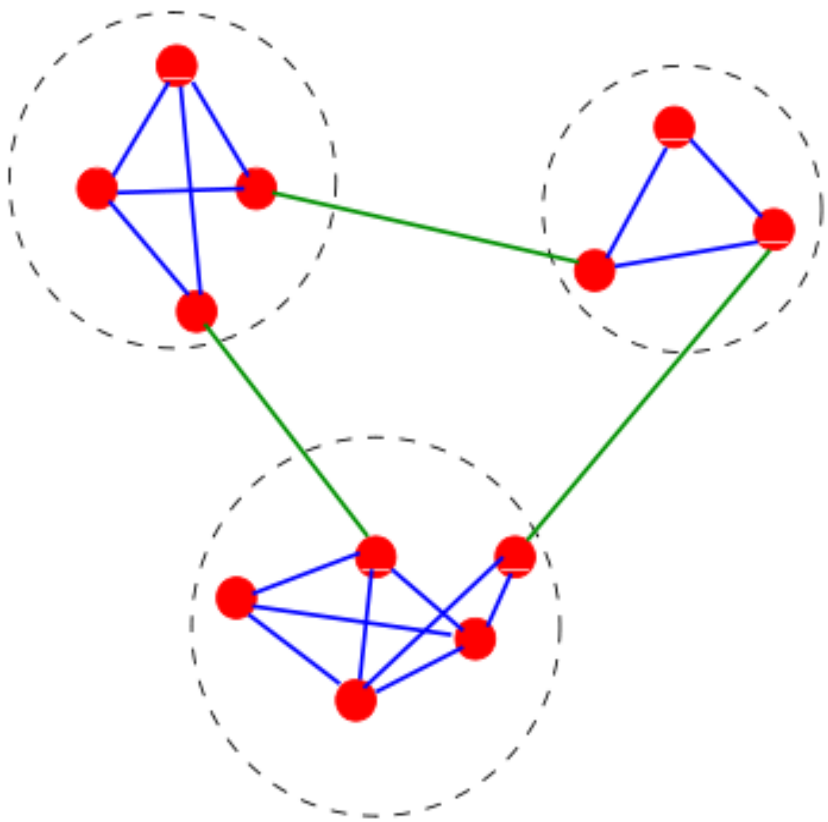
```
> wtc=walktrap.community(g,modularity=TRUE,weights=E(g)$weight)
> res=community.to.membership(g,wtc$merges,steps=3)
```

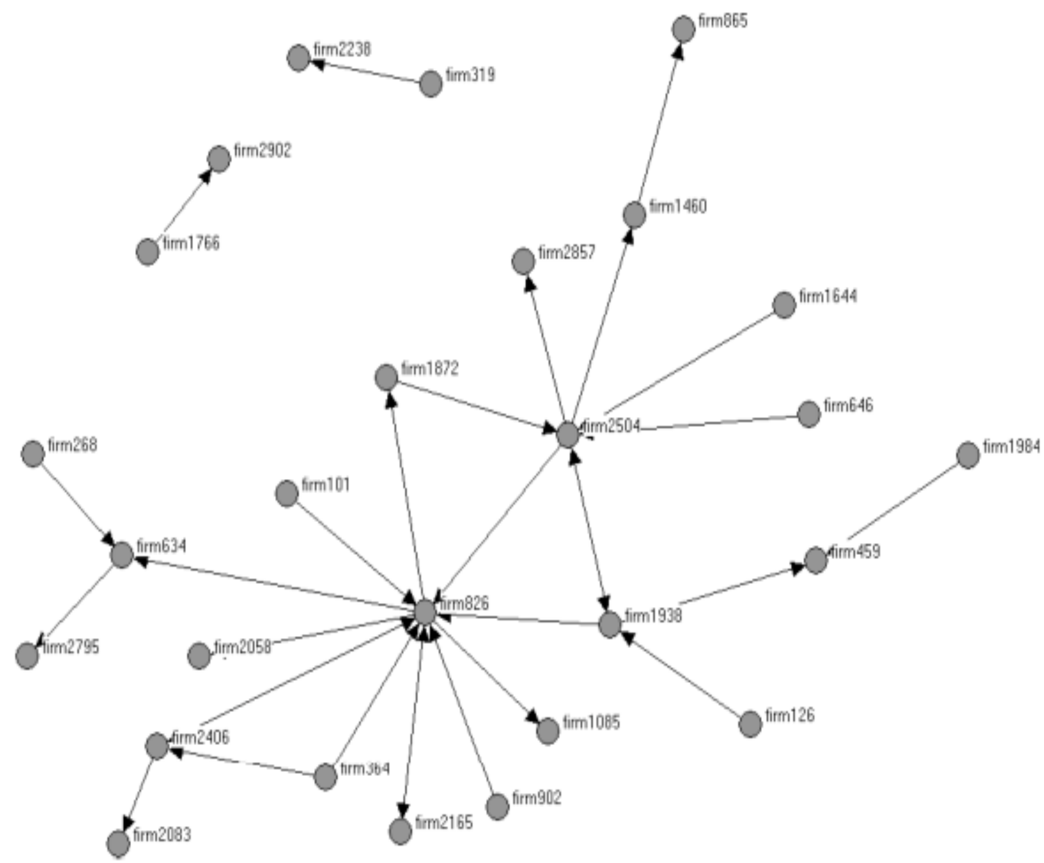
```
> print(res)
$membership
[1] 0 0 0 1 1

$csize
[1] 3 2
```

```
> print(modularity(g,res$membership,weights=E(g)$weight))
[1] 0.4128
```

Community v. Centrality





Communities
- Group-focused concept
- Members learn-by-doing through social interactions.

Centrality
- Hub focused concept
- Resources and skill of central players.

Data

Sources: SDC

- VentureExpert database (VE) - 1980-1999
- Exits data - IPO, M&A: 1980-2010

❑ Level of observation in the VE database:

- Company × Round × Investor

❑ Community identification using VE database:

- Not Individuals, Management or Undisclosed

❑ Filters used in exit analysis:

- U.S. investments
- Investment is not at "Buyout/Acquisition" stage
- Not “Angel or individual” investors

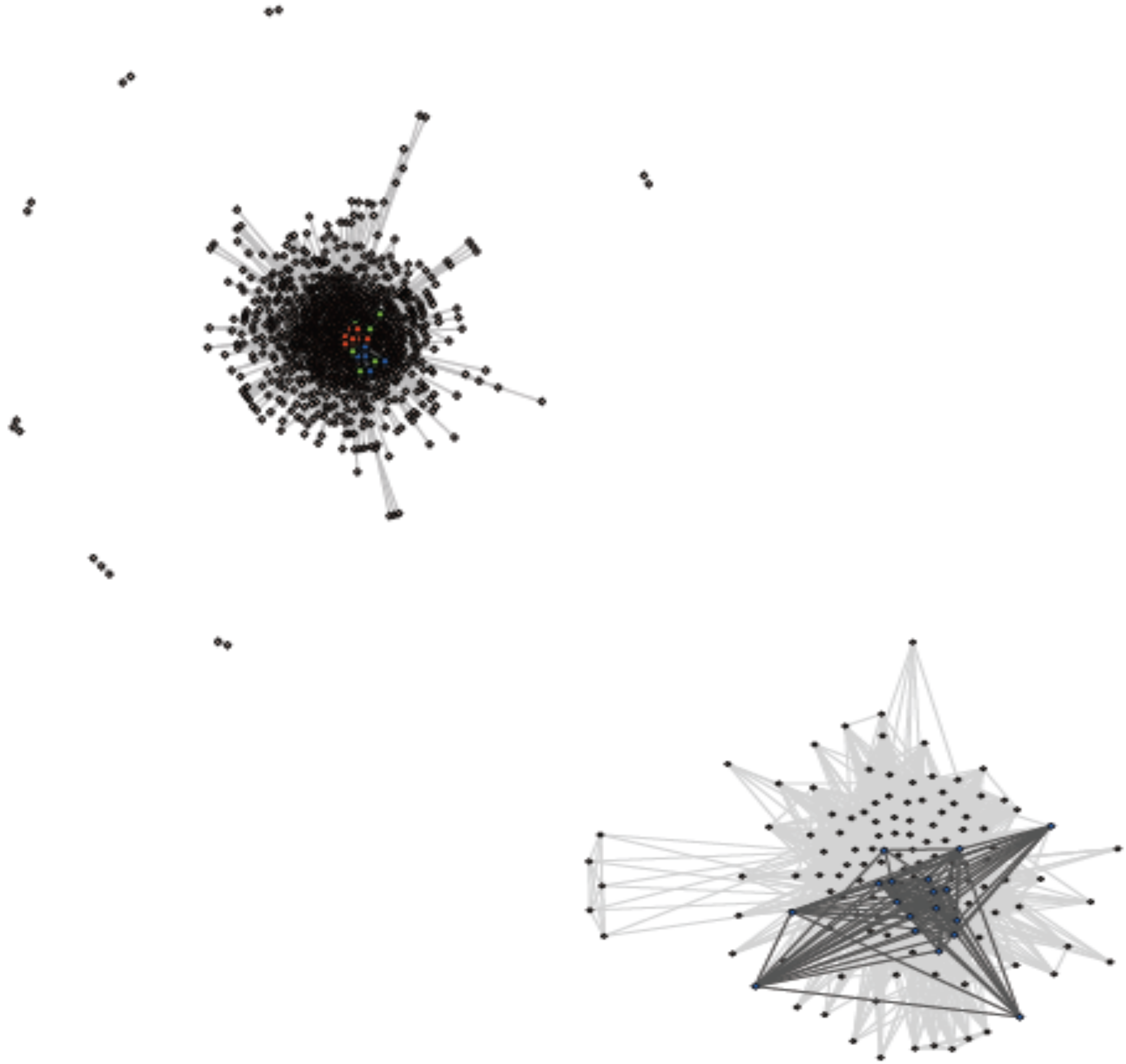


Figure 3: Network graph for connected VCs (1980–84). The upper plot shows the network of all VCs in communities (1180 in all), and blue, green, and red nodes in the center of the network are the VCs in the top three largest communities, respectively. The lower plot shows the network comprised only of the 134 VCs who are members of the 18 communities that have at least five VCs. The darker nodes in the lower plot show the VCs in the largest community.

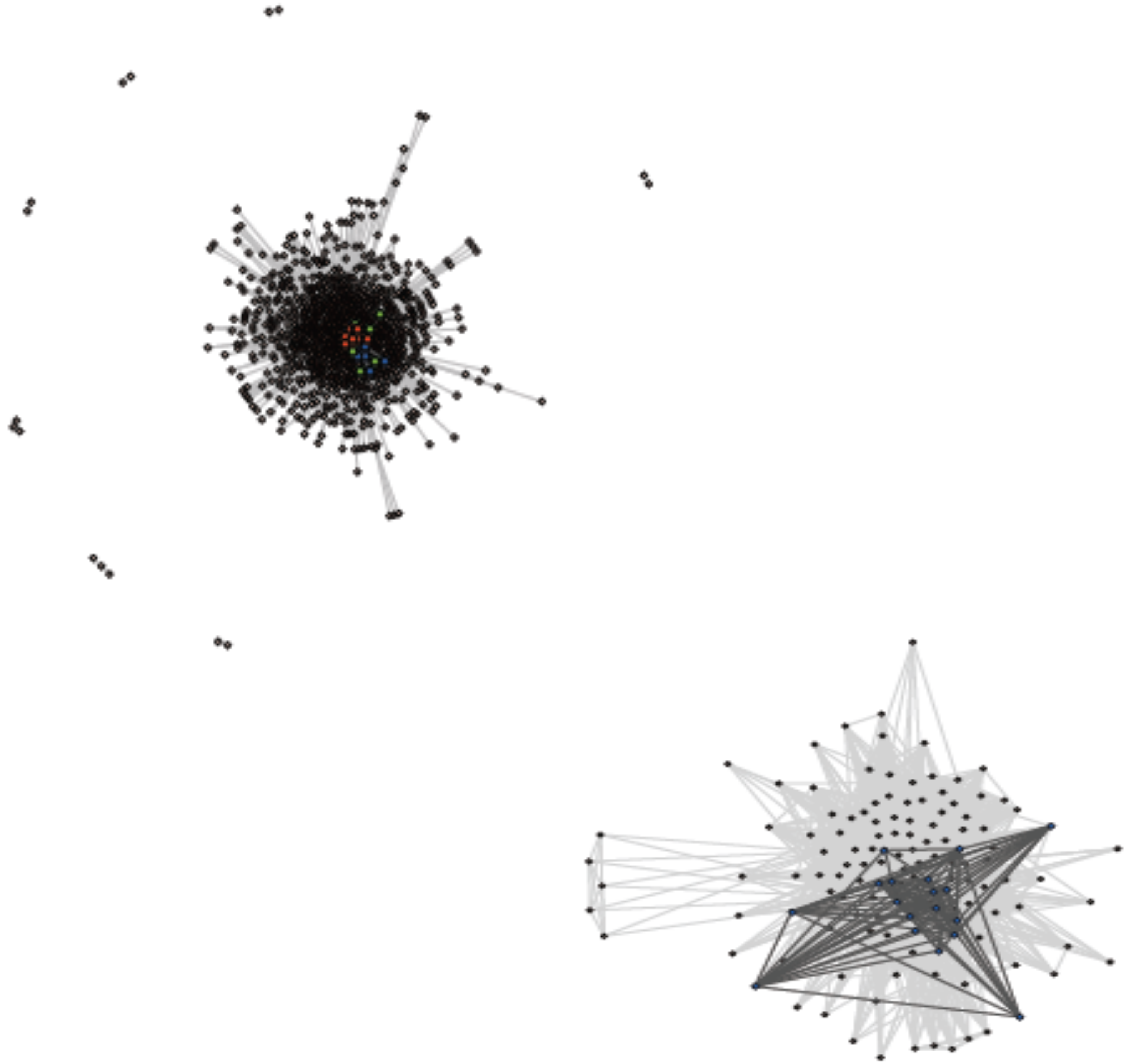


Figure 3: Network graph for connected VCs (1980–84). The upper plot shows the network of all VCs in communities (1180 in all), and blue, green, and red nodes in the center of the network are the VCs in the top three largest communities, respectively. The lower plot shows the network comprised only of the 134 VCs who are members of the 18 communities that have at least five VCs. The darker nodes in the lower plot show the VCs in the largest community.

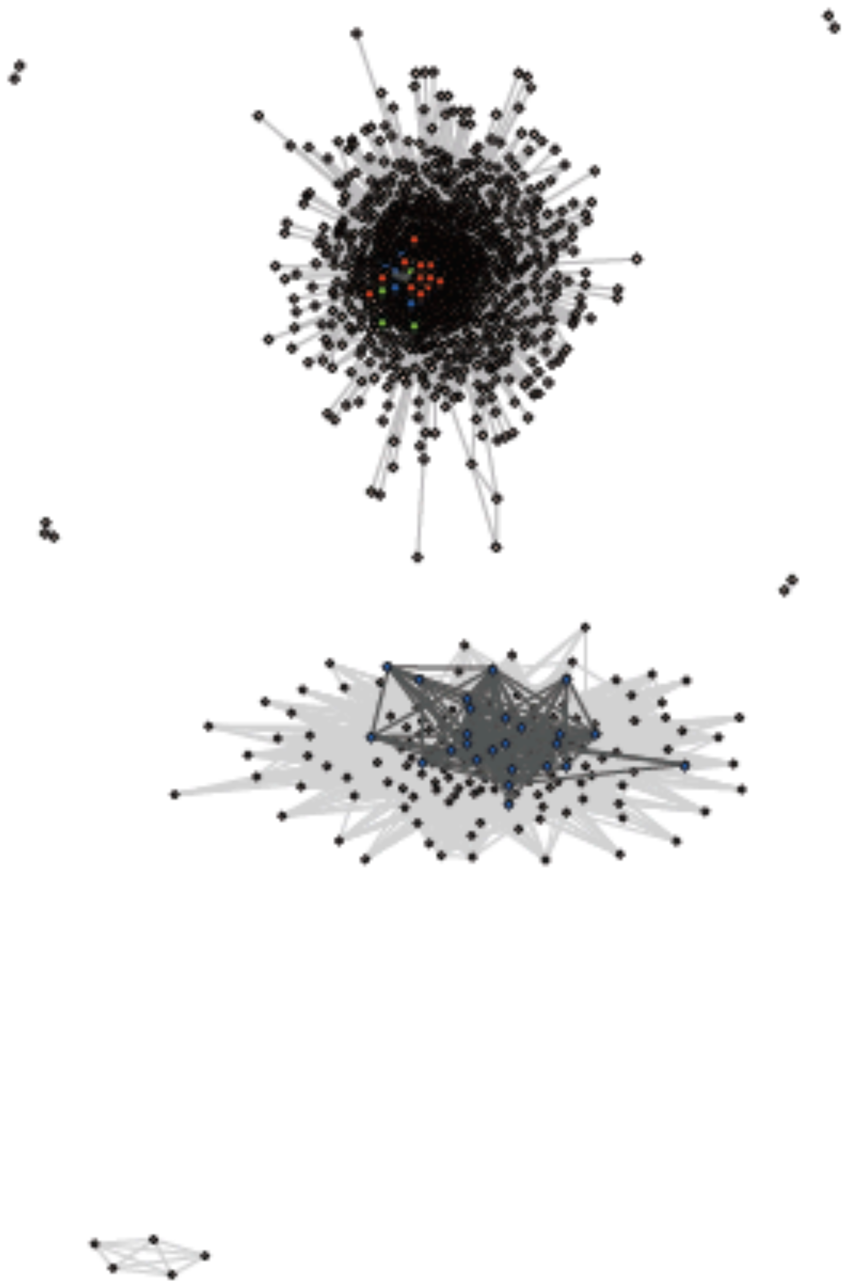
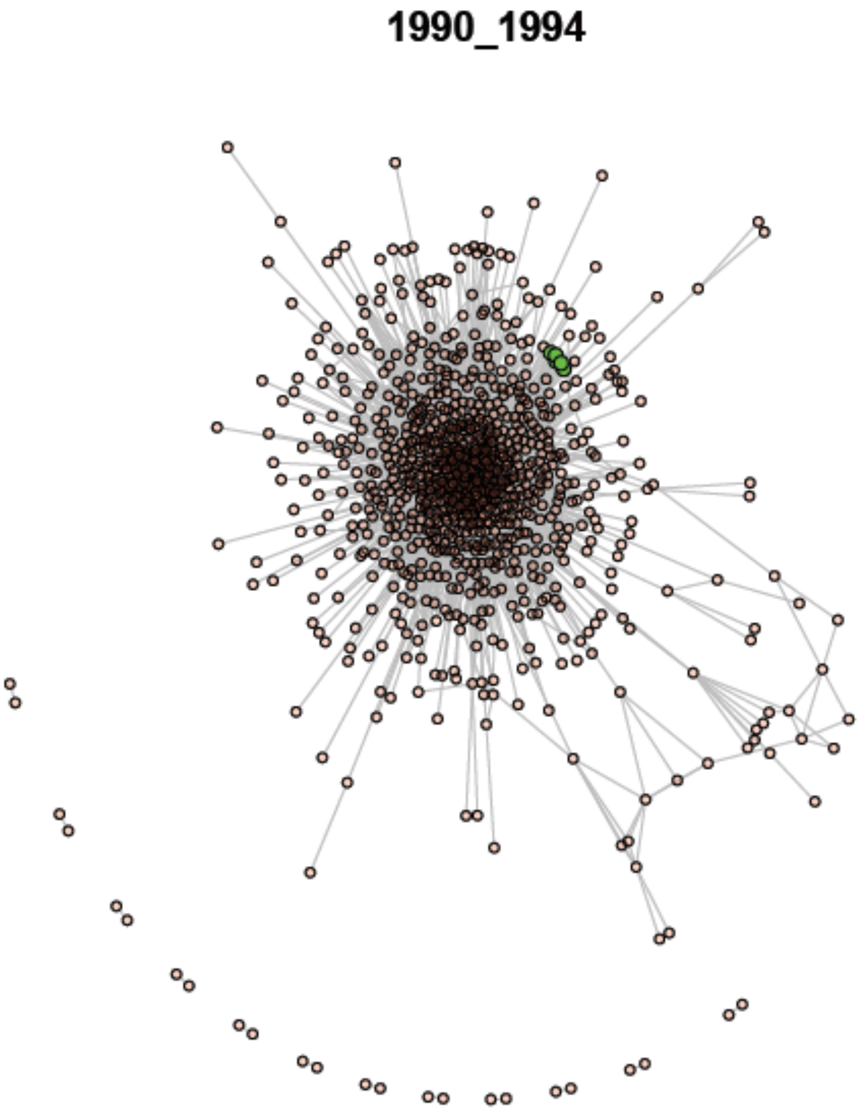
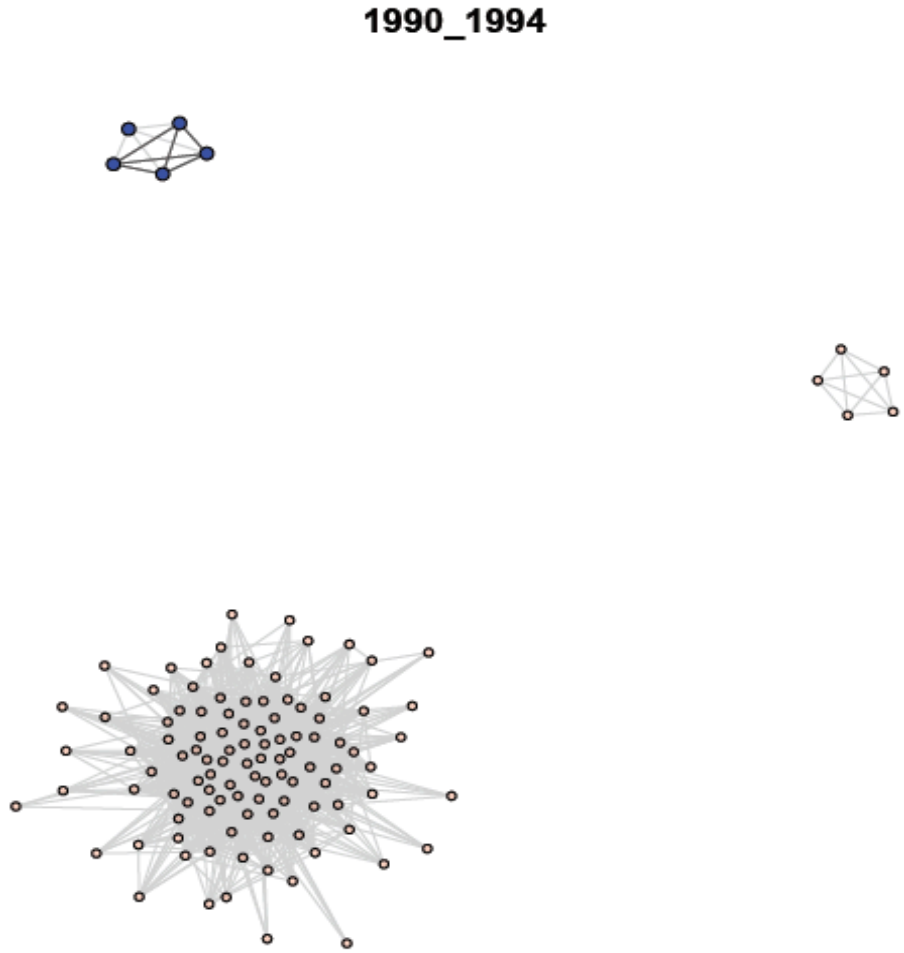


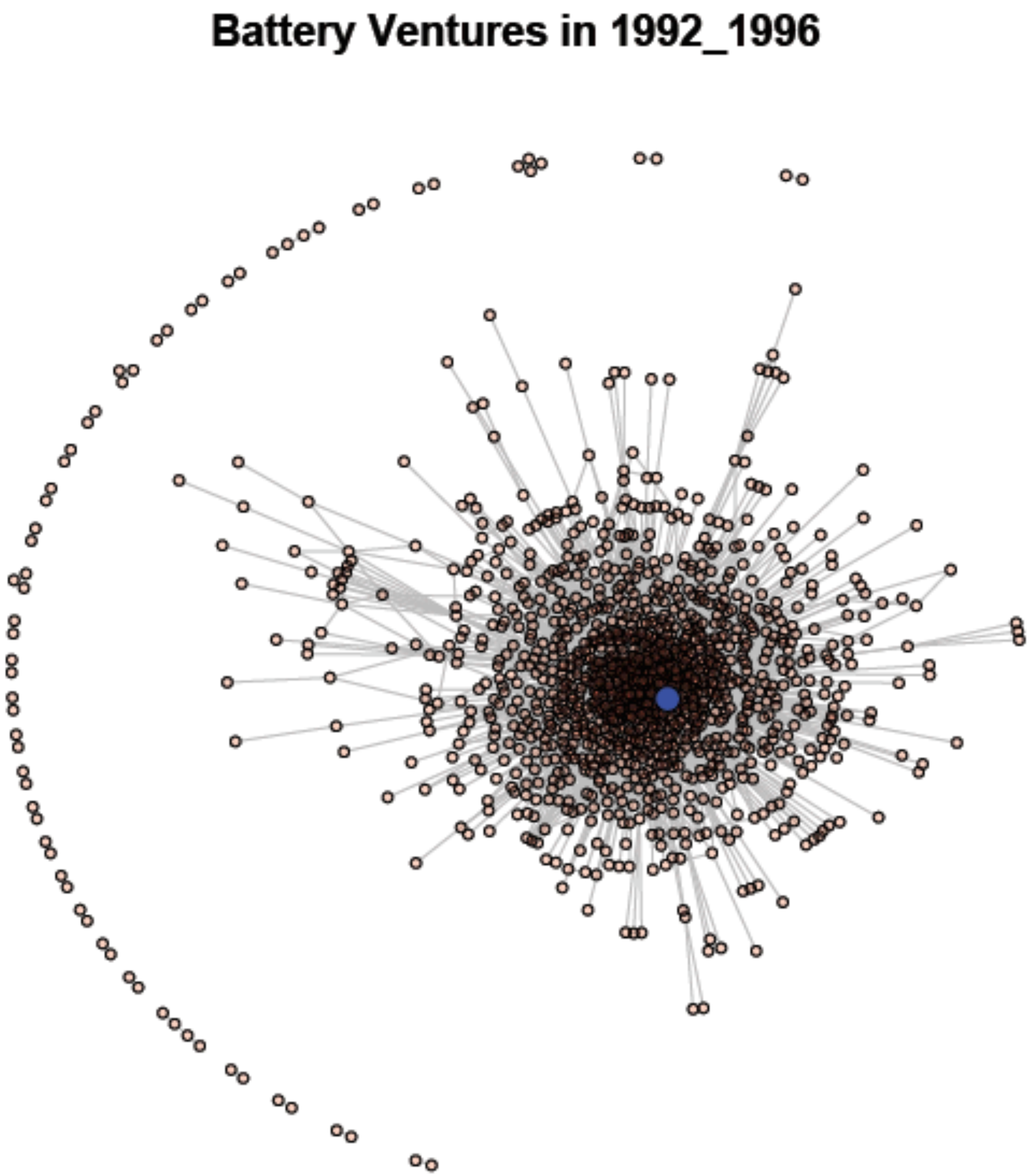
Figure 4: Network graph for connected VCs (1985–89). The upper plot shows the network of all VCs in communities (1295 in all), and blue, green, and red nodes in the center of the network are the VCs in the top three largest communities, respectively. The lower plot shows the network comprised only of the 180 VCs who are members of the 18 communities that have at least five VCs. The darker nodes in the lower plot show the VCs in the largest community. Note the single satellite community at the bottom of the lower plot. Such a community has low centrality.

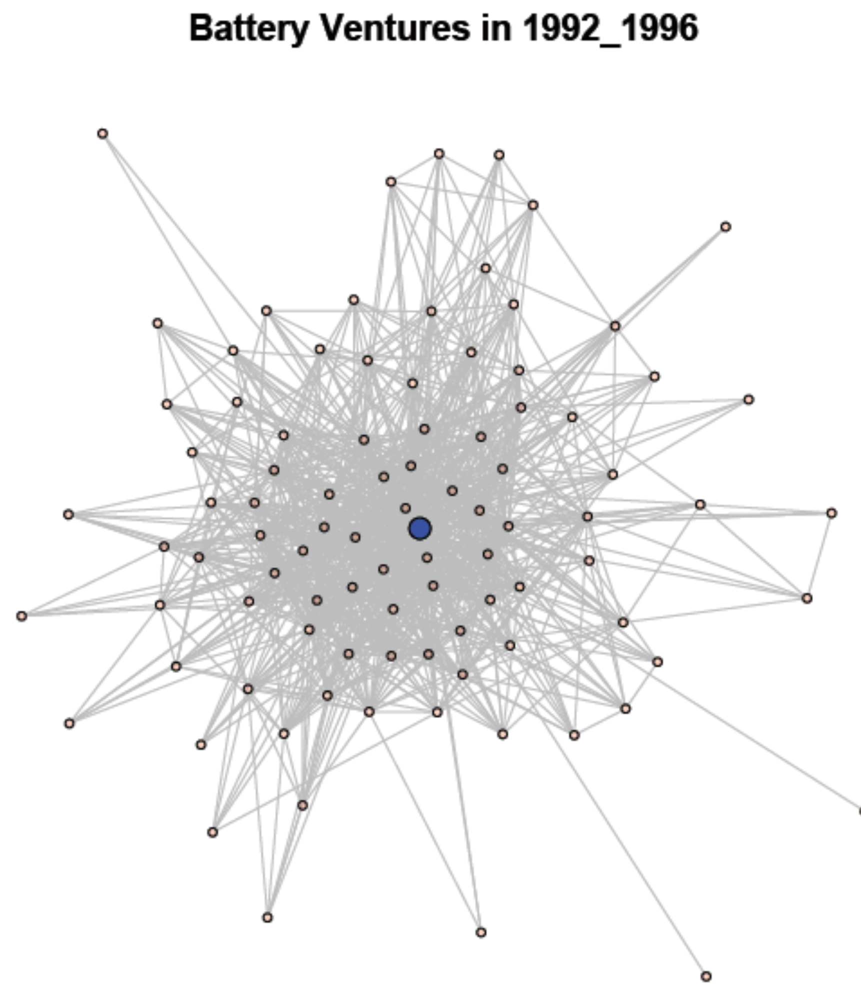
A Low Centrality Community





High Centrality but No Community





Venture Capitalists in the Sample

Variables:	Mean	Median	# Observations
# Rounds	47.98	9.00	1,962
# Companies	21.64	7.00	1,962
Investment per round ($ mm)	1.95	1.06	1,945
% Deals Syndicated	73.62	80.90	1,962
% Early Stage Deals	35.95	33.33	1,962
AUM ($ mm)	128.01	17.50	1,552
Total Investment ($ mm)	59.51	11.05	1,945
Age	9.59	6.00	1,950
# VC firms per MSA	14.24	3.00	127
CA/MA VC	0.35	0.00	1,962

Venture Capitalists in our sample. This table provides descriptive statistics of the 1,962 unique U.S.-based VCs in our database over the entire 20-year period, from 1980 to 1999. Data are from Venture Economics and exclude non-US investments, angel investors, and VC firms focusing on buyouts. We report the number of rounds of financing and the count of portfolio companies a VC invests in. Investment per round is the amount a VC invests in a round. % Deals Syndicated is the number of a VC's syndicated rounds as a percentage of all rounds that a VC invested in. % Early Stage Deals is the number of a VC's investment rounds classified by Venture Economics as early stage as of the round financing date, as a percentage of all Venture Economics deals for the VC between 1980 and 1999. AUM is the sum of the capital under management of a VC in all funds that invested during 1980-1999. Total investment is the sum of a VC's investments over this time period. Age is defined as the difference in the year of the VC's last investment in the period 1980 to 1999 and the VC firm's founding date. # VC firms per MSA is the total number of unique VCs headquartered a metropolitan statistical area (MSA). CA/MA VC is the fraction of all VCs that are headquartered in either California or Massachusetts.

Stability of community status

Window	# Community VCs	After 1 year	After 3 years	After 5 years
1980-1984	134	0.90	0.85	0.77
1981-1985	153	0.96	0.90	0.80
1982-1986	180	0.93	0.80	0.72
1983-1987	177	0.96	0.87	0.77
1984-1988	205	0.87	0.78	0.67
1985-1989	180	0.92	0.83	0.71
1986-1990	169	0.88	0.76	0.69
1987-1991	125	0.88	0.79	0.77
1988-1992	130	0.93	0.78	0.75
1989-1993	111	0.86	0.77	0.71
1990-1994	114	0.89	0.80	0.77
1991-1995	112	0.82	0.80	
1992-1996	146	0.93	0.89	
1993-1997	173	0.90		
1994-1998	246	0.94		
1995-1999	379			

The table provides data on the number of VCs who belong to community clusters in each 5-year window and the fraction of these that remain in a community after 1, 3, and 5 years from the initial window.

Table 3: Sample Communities. This table details venture capitalists that belong to two sample communities, one each for 1985-1989 and 1990-1994. We chose the communities that had Sequoia Capital as a member in both time periods.

Sample community from the period 1985–89:
(1) Arthur Rock & Co., (2) Asset Management Company Venture Capital, (3) Associated Venture Investors (AKA: AVI Capital), (4) Bryan & Edwards, (5) Draper Fisher Jurvetson (FKA: Draper Associates), (6) GT Technology Fund, (7) MedVenture Associates (AKA: MVA), (8) Mohr Davidow Ventures, (9) New Zealand Insurance, (10) Nippon Investment & Finance Co Ltd., (11) OSCCO Ventures, (12) Pacific Venture Partners, (13) Partech International, (14) **Sequoia Capital**, (15) Stanford University, (16) Suez Ventures (FKA: Indosuez Ventures), (17) Technology Venture Investors.

Sample community from the period 1990–94:
(1) Avalon Ventures, (2) Berkeley International Capital Corp., (3) Delphi Ventures, (4) Frazier Healthcare and Technology Ventures(FKA Frazier & Co), (5) Integral Capital Partners, (6) Kleiner Perkins Caufield & Byers, (7) Mayfield Fund, (8) Mohr Davidow Ventures, (9) **Sequoia Capital**, (10) Silicon Graphics, Inc., (11) Stanford University, (12) Technology Investment Fund, Inc., (13) Trinity Capital Partners, (14) Vertex Management Pte, Ltd. (AKA: Vertex Venture Holdings), (15) W.S. Investments.

Stability of community composition

Window 1	Window 2	Community	Bootstrapped Community	p-value
1980-1984	1981-1985	0.188	0.064	0.01^{***}
1981-1985	1982-1986	0.175	0.060	0.01^{***}
1982-1986	1983-1987	0.182	0.056	0.01^{***}
1983-1987	1984-1988	0.217	0.058	0.01^{***}
1984-1988	1985-1989	0.141	0.055	0.01^{***}
1985-1989	1986-1990	0.177	0.052	0.01^{***}
1986-1990	1987-1991	0.155	0.052	0.01^{***}
1987-1991	1988-1992	0.155	0.050	0.01^{***}
1988-1992	1989-1993	0.252	0.055	0.01^{***}
1989-1993	1990-1994	0.123	0.062	0.01^{***}
1990-1994	1991-1995	0.246	0.065	0.01^{***}
1991-1995	1992-1996	0.143	0.055	0.01^{***}
1992-1996	1993-1997	0.128	0.042	0.01^{***}
1993-1997	1994-1998	0.135	0.041	0.01^{***}
1994-1998	1995-1999	0.109	0.042	0.01^{***}

We identify community clusters in a 5-year window and examine whether the communities in the next five year window are similar to the ones in the previous period. We use the Jaccard similarity index which measures the similarity between every pair of communities in the adjacent period, and average it across all non-empty intersections. The Jaccard index is defined as the ratio of the size of the intersection set to the size of the union set. We generate a similar index for simulated communities generated by matching same community sizes and number of communities in each 5-year rolling window as in our sample.

$J(A,B) = \frac{|A \cap B|}{|A \cup B|}$. $\quad JC(A_i, B) = Mean_j(J(A_i, B_j) | J(A_i, B_j) > 0)$

Descriptive Statistics - 1

Variable	Total	Community Round	Not Community Round
	Panel A: Counts By Round		
# Deals	33,924	15,220	18,704
—Round 1	11,018	3,581	7,437
—Round 2	6,881	3,015	3,866
—Round 3	4,784	2,410	2,374
Syndicated	14,897	10,056	4,841
Early stage	12,118	5,472	6,646
Geographical Cluster	16,270	9,607	6,663
Rounds with			
—Geographical Cluster VC	19,678	12,140	7,538
—Corporate VC	3,372	1,923	1,449
—FI VC	7,586	4,415	3,171

Descriptive statistics for 33,924 rounds in 13,541 unique portfolio companies from 1985-1999. A round is a community round if at least one VC firm participating in it comes from a VC community. Communities are detected using a walk trap algorithm applied to syndicated deals over five year windows rolled forward one year at a time. The sample comprises VC deals obtained from Venture Economics but excludes non-US investments, angel investors and VC firms focusing on buyouts. Industry classifications are as per Venture Economics. Exit data are obtained by matching with Thomson Financial IPO and M&A databases.

Descriptive Statistics - 2

Variable	Total	Community Round	Not Community Round
Panel B: Percentage By Venture Economics Industry			
—Biotech	6.8	7.3	6.3
—Commu&Media	12.1	13.3	11.1
—Hardware	7.3	9.0	6.0
—Software	19.8	22.7	17.5
—Semiconductor, Electricals	7.0	7.9	6.3
—Consumer Products	7.8	5.3	9.9
—Industrial, Energy	5.9	3.4	8.0
—Internet	11.0	11.9	10.3
—Medical	13.7	15.0	12.7
—Others	8.5	4.4	11.9

Descriptive Statistics - 3

Variable	Total	Community Round	Not Community Round
Panel C: Round Statistics			
Proceeds ($ million)	4 (1)	5 (2)	3 (1)
# VCs	2.08 (1)	2.89 (2)	1.42 (1)
—in syndicated rounds	3.46 (3)	3.85 (3)	2.64 (2)
—in early stage rounds	1.93 (1)	2.53 (2)	1.43 (1)
—in round 1	1.54 (1)	2.03 (2)	1.31 (1)
—in round 2	2.00 (1)	2.70 (2)	1.45 (1)
—in round 3	2.38 (2)	3.23 (3)	1.52 (1)
PANEL D: Exit			
Rounds with			
—IPO exits	3,828	2,071	1,757
—M&A exits	8,794	4,363	4,431
—Follow-on funding	23,972	11,903	12,069

Characteristics of Same-Community VCs

	Community	Simulated Community	p-value
Age	9.18	8.25	0.01***
AUM	130.40	70.72	0.01***
Centrality	0.08	0.03	0.01***
Industry HHI	0.28	0.48	0.01***
Stage HHI	0.33	0.52	0.01***
Company Region HHI	0.42	0.58	0.01***

The table compares key community characteristics with those of simulated communities generated by matching community sizes and number of communities in each 5-year rolling window. For each community (and simulated community), we generate the mean of the characteristic, and present the average value across communities. *Age* uses the number of years between a VC's last investment in a 5-year window and the founding year of the VC firm. *Assets under management (AUM)*, in $ million, uses the sum of all VC funds that invested during a 5-year period. *Centrality* is based on each VC's eigenvector centrality determined for each 5-year rolling window. For the remaining attributes, we calculate the Herfindahl-Hirschman Index (HHI) as the sum of squared share in each subcategory of the attribute. *Industry HHI* is the Herfindahl index based on the % of a community VC's deals in each industry, while *Stage HHI* is the Herfindahl index based on the % of deals in each stage of investment. *Company Region HHI* is the Herfindahl index based on the % of deals in each geographic region. In unreported tests, we see similar results when we use HHI based on amount invested. The industry, stage and geographic region classifications are those provided by Venture Economics. The last column shows the p-values testing the equality of the means of the community and bootstrapped community characteristics. ***, **, and * denote 1%, 5% and 10% significance, respectively.

Similarity of Within-Community VCs

	Community	Simulated	p-value
	Panel A: Variation in Reach Attributes		
Age	6.86	7.37	0.01***
AUM	142.74	99.52	0.01***
Centrality	0.08	0.05	0.01***
	Panel B: Variation in Functional Styles		
Industry HHI	0.22	0.31	0.01***
Stage HHI	0.21	0.28	0.01***
Company Region HHI	0.21	0.31	0.01***
Industry Variation	0.96	3.20	0.01***
Stage Variation	0.70	2.28	0.01***
Co. Region Variation	0.89	3.65	0.01***
	Panel C: Mean of Community Geographic HHI		
VC MSA HHI	0.35	0.20	0.01***
VC State HHI	0.43	0.24	0.01***
VC Region HHI	0.41	0.25	0.01***
	Panel D: Mean of Community Ownership HHI		
VC Ownership HHI	0.55	0.45	0.01***

The table presents variation in key attributes (in Panels A-B) and mean geographic location HHI (in Panel C) and ownership HHI (in Panel D) of VCs within communities, and compares these to those of simulated communities generated by matching community sizes and number of communities in each 5-year rolling window.

Functional Expertise Similarity of Within-Community VCs

	Community	Simulated	p-value
Industry Rank:			
1	0.16	0.23	0.01^{***}
2	0.13	0.21	0.01^{***}
3	0.12	0.18	0.01^{***}
4	0.11	0.17	0.01^{***}
5 = Others	0.18	0.33	0.01^{***}
Stage Rank:			
1	0.17	0.22	0.01^{***}
2	0.18	0.24	0.01^{***}
3 = Others	0.16	0.28	0.01^{***}
Company Region Rank:			
1	0.20	0.31	0.01^{***}
2	0.10	0.22	0.01^{***}
3	0.11	0.17	0.01^{***}
4	0.07	0.16	0.01^{***}
5 = Others	0.16	0.36	0.01^{***}

We present the mean (across all communities) of the sum of squared deviation of VC's share of deal in some subcategories (based on total $ amount invested in a 5-year rolling window in each of the top 4 industries, top 2 stages, and top 4 company regions, with the remainder share of investment comprising the last subcategory in each). We compare these values to those of simulated communities generated by matching community sizes and number of communities in each 5-year rolling window.

Similarity Across Communities

	Community	Simulated	p-value
	Panel A: Variation in Functional Styles		
Industry HHI	0.14	0.04	0.01***
Stage HHI	0.11	0.05	0.01***
Company Region HHI	0.16	0.07	0.01***
Industry Variation	1.30	0.60	0.01***
Stage Variation	0.63	0.39	0.01***
Company Region Variation	1.40	1.01	0.01***
	Panel B: Variation of Community Geographic HHI		
VC MSA HHI	0.19	0.08	0.01***
VC State HHI	0.20	0.09	0.01***
VC Region HHI	0.19	0.09	0.01***
	Panel C: Variation of Community Ownership HHI		
VC Ownership HHI	0.19	0.14	0.01***

The table presents across community variation in (average) key VC attributes (in Panel A), in geographic location HHI (in Panel B) and in ownership HHI (in Panel C) of VCs within communities, and compares these to those of simulated communities generated by matching community sizes and number of communities in each 5-year rolling window.

Success through next round financing or exit

	Round1	Round2	Round3
	(1)	(2)	(3)
Community	0.093**	0.192***	0.033
Early Stage	0.299***	0.280***	0.271***
Company Geographical Cluster	0.090**	0.039	0.142**
AUM_Round	0.179***	0.073***	0.106***
Corporate VC	-0.066	0.026	0.131
FI VC	-0.124***	-0.081	0.019
Syndicated	0.515***	0.558***	0.589***
IPO Rate	-0.267***	-0.556***	-0.194
Centrality	-0.068***	0.021	0.125**
VC Geographical Cluster	0.066*	0.029	-0.116
Experience	-0.102***	-0.088**	-0.125***
Early Stage Focus	0.320***	0.734***	0.705**
Industry Focus	0.082	0.086	0.139
# Observations	9,328	4,262	3,105

The table reports the estimates of a probit model in which the dependent variable is 1.0 if there is a successful exit (IPO or merger) or a follow-on financing round within 10 years of the investment round and 0 otherwise. See Appendix B for a description of the independent variables. All specifications include year and industry fixed effects, which are not reported for brevity. The sample comprises VC deals obtained from Venture Economics but excludes non-US investments, angel investors and VC firms focusing on buyouts.

Time to exit and probability of exit.

	Cox	Probit	Competing Hazards		
			IPO	Round 1	Round 2
	(1)	(2)	(3)	(4)	(5)
Community	1.089***	0.043*	1.116**	1.095**	0.950
Early Stage	0.911***	-0.037**	0.849***	1.425***	1.375***
Company Geographical Cluster	1.057**	0.038**	1.060	1.078**	0.959
AUM_Round	1.088***	0.057***	1.130***	1.151***	1.048*
Corporate VC	1.320***	0.202***	1.503***	0.835***	0.978
FI VC	1.083***	0.056***	1.191***	0.897***	0.900*
Syndicated	1.318***	0.211***	1.311***	1.386***	1.305***
IPO Rate	1.084	0.063	1.145	0.692***	0.648**
Centrality	0.998	0.006	0.983	0.943***	1.032
VC Geographical Cluster	1.039	0.026	1.075	1.011	1.000
Experience	0.958***	-0.035***	1.002	0.919***	0.946*
Early Stage Focus	1.043	0.008	0.546***	1.894***	1.850***
Industry Focus	1.090	0.062	1.542**	1.155	1.040
# Observations	23,977	24,864	23,977	9,037	4,108

Specification (1) reports the estimates of a Cox proportional hazards model. The dependent variable is the number of days from financing to the earlier of exit (IPO or merger) or April 30, 2010. Specification (2) reports the estimates of a probit model in which the dependent variable is 1.0 if there is an exit (IPO or merger) within 10 years of the investment round and 0 otherwise. Specifications (3)-(5) report estimates of a competing hazards model where the event of interest is exit only through an IPO (Specification (3)), IPO or follow on financing after round 1 (Specification (4)) or after round 2 (Specification (5)). A merger is the competing risk in the competing hazards models. See Appendix B for a description of the independent variables.

Robustness of Within-Communty Similarity

	Community	Simulated	p-value
Panel A: Only First Time Deals in Each 5-Year Window			
Industry HHI	0.20	0.31	0.01***
Stage HHI	0.18	0.27	0.01***
Company Region HHI	0.21	0.31	0.01***
Industry Variation	0.96	3.20	0.01***
Stage Variation	0.86	2.48	0.01***
Company Region Variation	0.89	3.64	0.01***
Panel B: Community detected based only on First Round Syndicates			
Age	7.58	7.22	0.10*
AUM	154.80	96.54	0.01***
Centrality	0.08	0.04	0.01***
Industry HHI	0.17	0.31	0.01***
Stage HHI	0.17	0.27	0.01***
Company Region HHI	0.14	0.31	0.01***
Industry Variation	0.56	2.43	0.01***
Stage Variation	0.34	1.76	0.01***
Company Region Variation	0.49	2.79	0.01***

This table provides two robustness tests of similarity among VCs within communities. Panel A considers a community VC's first investment in each portfolio company for determining % of deals in each subcategory of attributes, and uses it to determine within-comunity HHI variation as well as variation between subcategories. Panel B uses an alternative basis for communities, namely the first round of syndications rather than all rounds used in our analysis so far. Using these alternative communities, we determine within-community HHI variation and variation between subcategories. Given the alternative community, we additionally present the standard deviation of the reach variables only in Panel B.

Summary

- Communities are detectable in the VC galaxy.
- VC communities display homophily on some attributes and heterogeneity on others.
- Syndicates compete through differentiation and specialization rather than generalized skills relevant to young firm financing.
- Repeated interaction provides benefits to portfolio companies over and above individual VC influence.
- Community backed ventures are more likely to exit successfully.
- Our results are consistent with learning-by-doing or incomplete contracting models of VC investing in which familiarity aids learning and enhances trust and reciprocity.