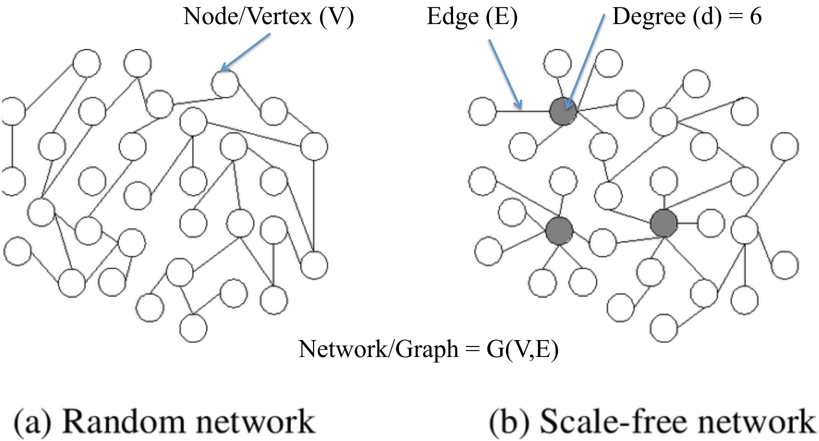
Inside the Matrix: Modeling the Network Effect

Sanjiv R. Das[1]
Santa Clara University

@RFInance
Chicago
May 2015

[1] http://algo.scu.edu/ sanjivdas/risknet.pdf

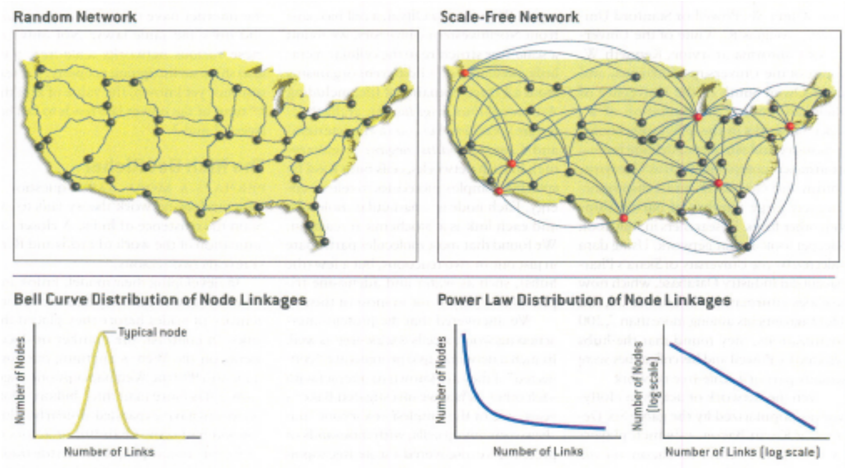
Graph Theory: Network Types



$$f(d) \sim N(\mu, \sigma^2)$$

$$f(d) = d^{-\alpha}, \quad 2 < \alpha < 3$$

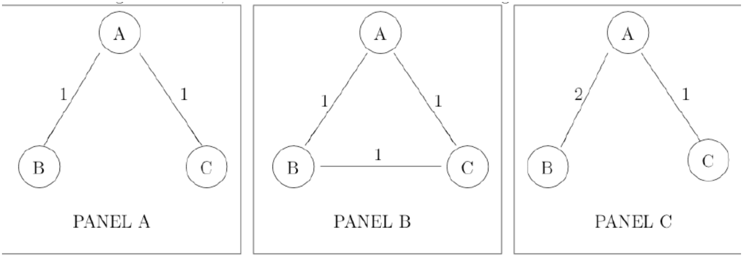
Random vs Scale-Free Graphs



Barabasi, Sciam, May 2003

Centrality (Bonacich 1987)

- Also known as `PageRank` by Google.
- Adjacency matrix: $A_{ij} \in \mathcal{R}^{N \times N}$
- Influence: $x_i = \sum_{j=1}^{N} A_{ij} x_j$
- $\lambda \mathbf{x} = \mathbf{A} \cdot \mathbf{x}$
- Centrality is the eigenvector $\mathbf{x}$ corresponding to the largest eigenvalue.



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

Centrality scores = {0.71, 0.50, 0.50}

Centrality scores = {0.58, 0.58, 0.58}

Centrality scores = {0.71, 0.63, 0.32}

Fragility

- Definition: how quickly will the failure of any one node trigger failures across the network? Is network malaise likely to spread or be locally contained?
- Metric:

$$R = \frac{E(d^2)}{E(d)},$$

  where d is node degree.
- Fragile if $R > 2$.
- Fragility of the sample network $= 20$

What is Systemic Analysis?

1. 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.
2. Challenge: requires most or all of the data in the system; therefore, high-quality information extraction and integration is critical.

Midas Project: Overview

Joint work with IBM Almaden[2]

- 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.

---

[2] "Extracting, Linking and Integrating Data from Public Sources: A Financial Case Study," (2011), (with Douglas Burdick, Mauricio A. Hernandez, Howard Ho, Georgia Koutrika, Rajasekar Krishnamurthy, Lucian Popa, Ioana Stanoi, Shivakumar Vaithyanathan), *IEEE Data Engineering Bulletin*, 34(3), 60-67. [Proceedings WWW2010, April 26-30, 2010, Raleigh, North Carolina.]

Entity View

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

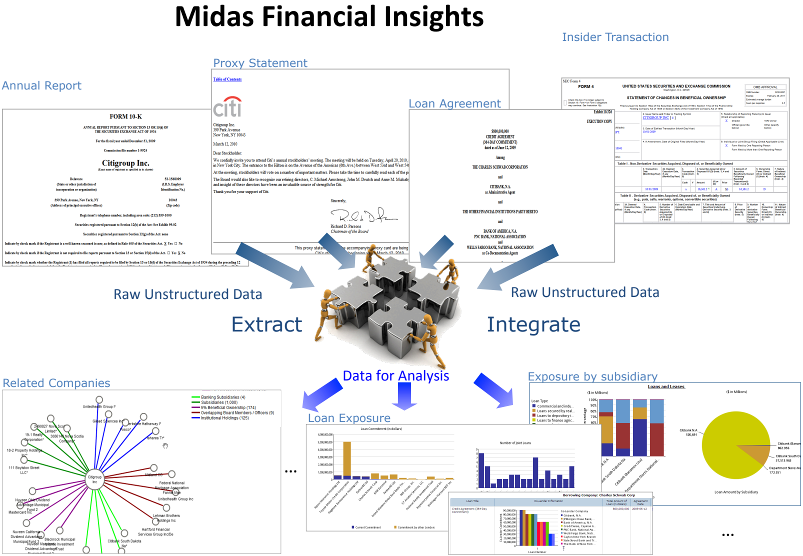
Private Data

Hoovers | D&B | FINRA

Private Wall Street Journal

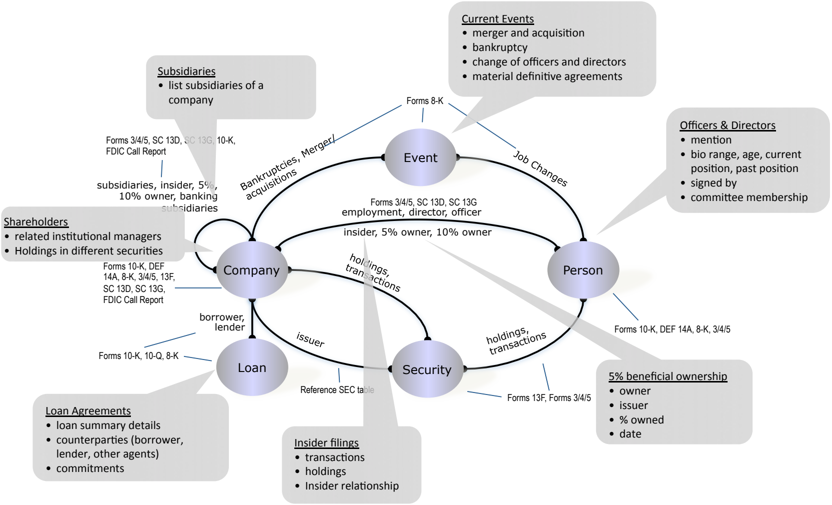
- Regulators
- Credit committees
- Investment analysts
- Portfolio managers
- Equity managers

Input & Output



Data

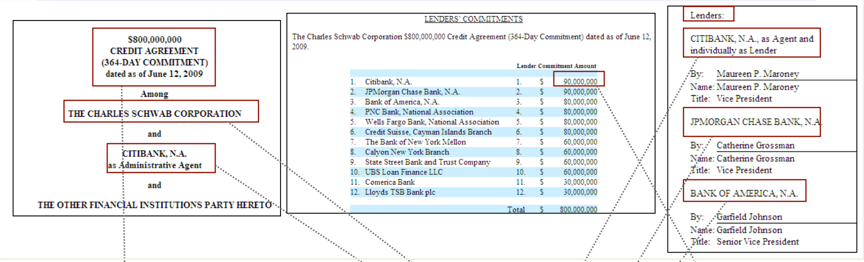
Midas provides Analytical Insights into company relationships by exposing information concepts and relationships within extracted concepts



Loan Extraction

Example Analysis : Extraction of Loan Information Data

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



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

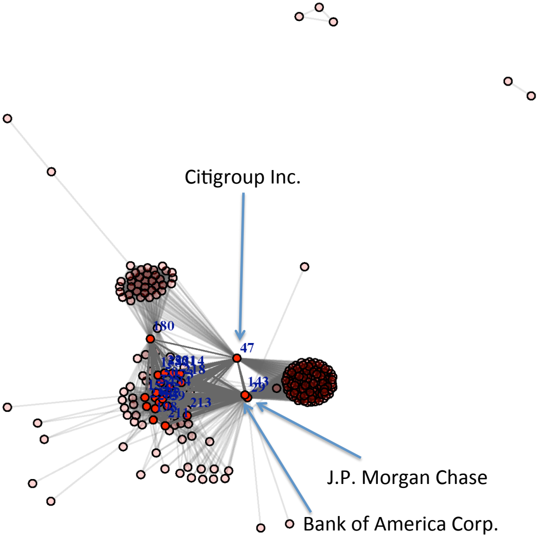
Loan Information

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
...			

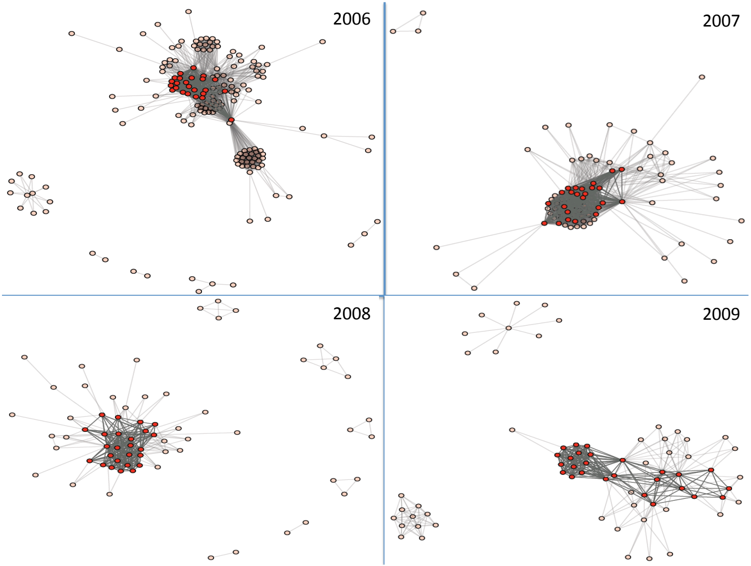
Loan Company Information

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

Loan Network 2005



Loan Network 2006–2009



Systemically Important Financial Institutions (SIFIs)

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

Risk Networks: Definitions and Risk Score

- Assume n nodes, i.e., firms, or "assets."
- Let $E \in R^{n \times n}$ be a well-defined adjacency matrix. This quantifies the influence of each node on another.
- E may be portrayed as a directed graph, i.e., $E_{ij} \neq E_{ji}$. $E_{jj} = 1$; $E_{ij} \in \{0, 1\}$.
- C is a $(n \times 1)$ risk vector that defines the risk score for each asset.
- We define the "risk score" as

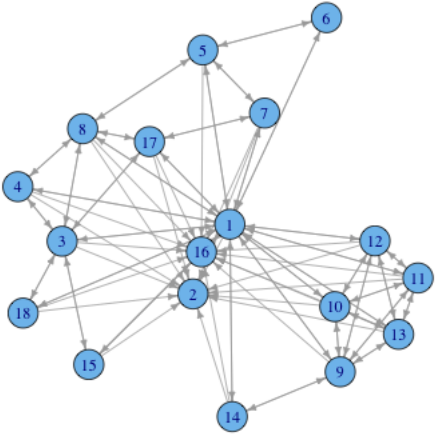
$$S = \sqrt{C^\top \ E \ C}$$

- $S(C, E)$ is linear homogenous in C.

Example

Risk vector C: 0 0 1 2 2 2 2 2 1 0 2 2 2 2 1 0 1 1

Risk Score: $S = 11.62$

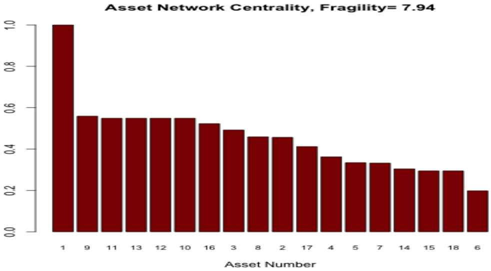


Example: Adjacency Matrix

	[,1]	[,2]	[,3]	[,4]	[,5]	[,6]	[,7]	[,8]	[,9]	[,10]	[,11]	[,12]	[,13]	[,14]	[,15]	[,16]	[,17]	[,18]
[1,]	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
[2,]	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
[3,]	1	1	1	1	0	0	0	1	0	0	0	0	0	0	1	1	1	1
[4,]	1	1	1	1	0	0	0	1	0	0	0	0	0	0	0	1	0	0
[5,]	1	0	0	0	1	1	1	1	0	0	0	0	0	0	0	1	0	0
[6,]	1	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0
[7,]	1	1	0	0	1	0	1	0	0	0	0	0	0	0	0	1	1	0
[8,]	1	1	1	1	1	0	0	1	0	0	0	0	0	0	0	1	1	0
[9,]	1	0	0	0	0	0	0	0	1	1	1	1	1	1	0	1	0	0
[10,]	1	1	0	0	0	0	0	0	1	1	1	1	1	0	0	1	0	0
[11,]	1	1	0	0	0	0	0	0	1	1	1	1	1	0	0	1	0	0
[12,]	1	1	0	0	0	0	0	0	1	1	1	1	1	0	0	1	0	0
[13,]	1	1	0	0	0	0	0	0	1	1	1	1	1	0	0	1	0	0
[14,]	1	1	0	0	0	0	0	0	1	0	0	0	0	1	0	1	0	0
[15,]	1	1	1	0	0	0	0	0	0	0	0	0	0	0	1	1	0	0
[16,]	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0
[17,]	1	1	1	0	0	0	1	1	0	0	0	0	0	0	0	1	1	0
[18,]	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	1	0	1

Centrality and Fragility

- Centrality is the principal eigenvector x of dimension $(n \times 1)$ such that for scalar λ: $\lambda\, x = E\, x$
- Plot:



- Fragility: for each node with degree d_j, fragility is the score given by

$$E(d^2)/E(d)$$

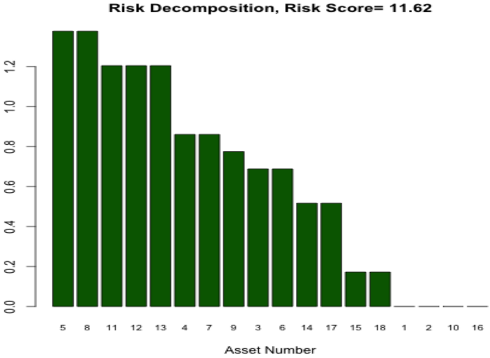
Increasing values imply a more fragile network.

Risk Decomposition

1. Exploits the homogeneity of degree one property of S.
2. Risk decomposition (using Euler's formula):

$$S = \frac{\partial S}{\partial C_1} C_1 + \frac{\partial S}{\partial C_2} C_2 + \ldots + \frac{\partial S}{\partial C_n} C_n$$

3. Plot:

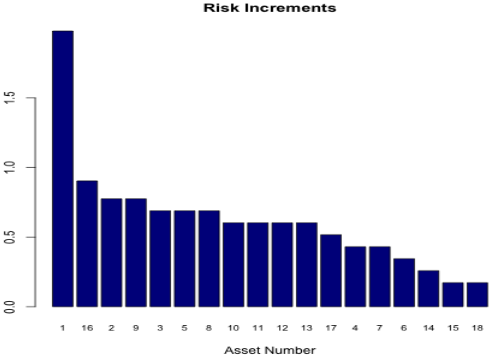


Risk Increments

- Increments are simply:

$$I_j = \frac{\partial S}{\partial C_j}, \quad \forall j$$

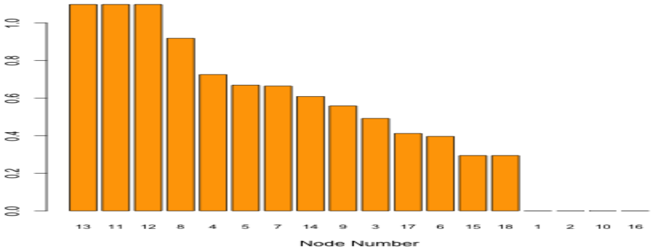
- Plot:



Criticality

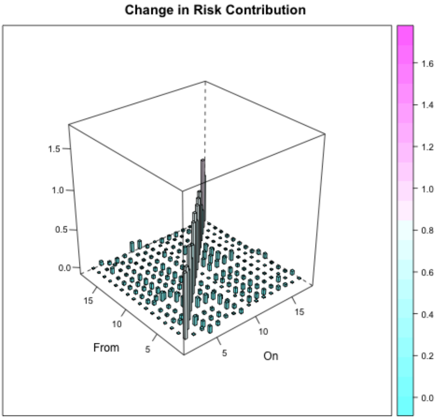
Definition: "Criticality" is compromise-weighted centrality. This new measure is defined as $y = C \times x$ where $y, C, x \in \mathcal{R}^n$. Note that this is an element-wise multiplication of vectors C and x.

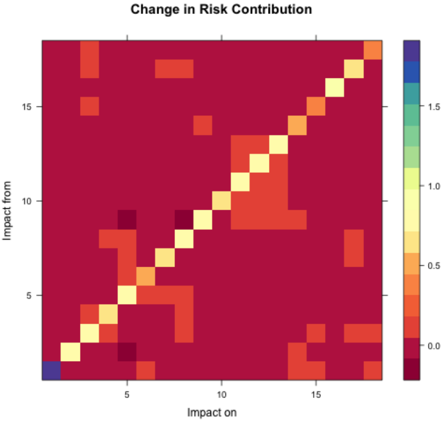
- Critical nodes need immediate attention, either because they are heavily compromised or they are of high centrality, or both.
- It offers a way for regulators to prioritize their attention to critical financial institutions, and pre-empt systemic risk from blowing up.



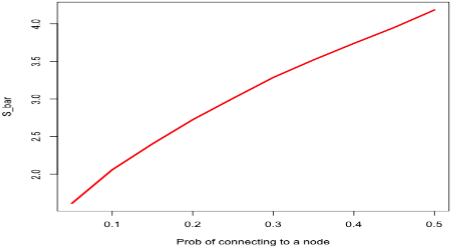
Cross Risk

Is the spill over risk from node i to node j material?



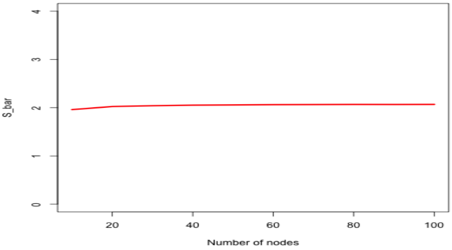


Risk Scaling



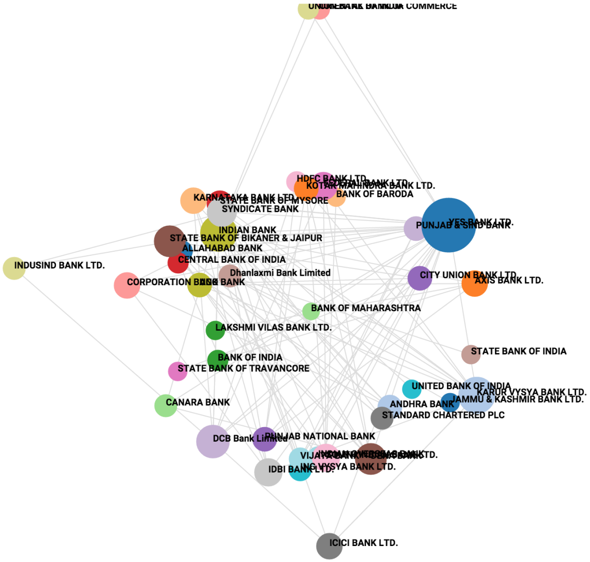
The increase in normalized risk score $\bar{S}$ as the number of connections per node increases. The plot shows how the risk score increases as the probability of two nodes being bilaterally connected increases from 5% to 50%. For each level of bilateral probability a random network is generated for 50 nodes. A compromise vector is also generated with equally likely values $\{0, 1, 2\}$. This is repeated 100 times and the mean risk score across 100 simulations is plotted on the y-axis against the bilateral probability on the x-axis.

Too Big To Fail?

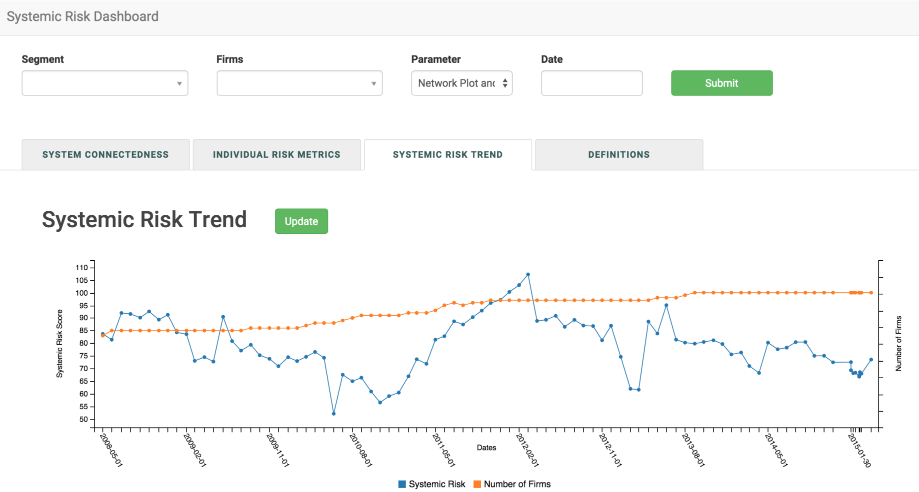


Change in normalized risk score $\bar{S}$ as the number of nodes increases, while keeping the average number of connections between nodes constant. A compromise vector is also generated with equally likely values $\{0, 1, 2\}$. This is repeated 5000 times for each fixed number of nodes and the mean risk score across 5000 simulations is plotted on the y-axis against the number of nodes on the x-axis.

Systemic Risk in Indian Banks



Systemic Risk in India over time



Risk Decomposition in India

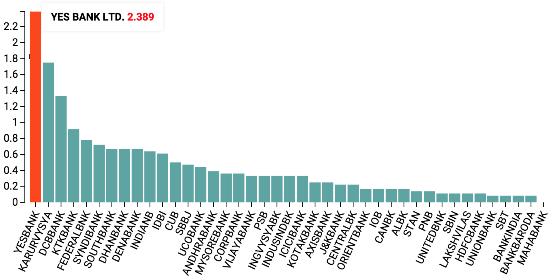
Fragility

7.39

Systemic Risk Score

18

Risk Decomposition



Risk Increment