Irving Fisher Committee on
Central Bank Statistics

BANK FOR INTERNATIONAL SETTLEMENTS

IFC – Bank Indonesia International Workshop and Seminar on *"Big Data for Central Bank Policies / Building Pathways for Policy Making with Big Data"*

Bali, Indonesia, 23-26 July 2018

Big data and FinRisk[1]

Sanjiv R. Das,

Santa Clara University

[1] This presentation was prepared for the meeting. The views expressed are those of the author and do not necessarily reflect the views of the BIS, the IFC or the central banks and other institutions represented at the meeting.

Big Data and FinRisk

Sanjiv R. Das
Santa Clara University

Bank of Indonesia and BIS/IFC
International Conference on Big Data
Bali, Indonesia, July 26, 2018

Outline

1. Overview of challenges in Big Data for Finance
2. Financial networks for Systemic Risk Measurement.
    a. USA
    b. India
3. Zero-revelation prediction of bank malaise.

Research Challenges in Financial Data Modeling and Analysis

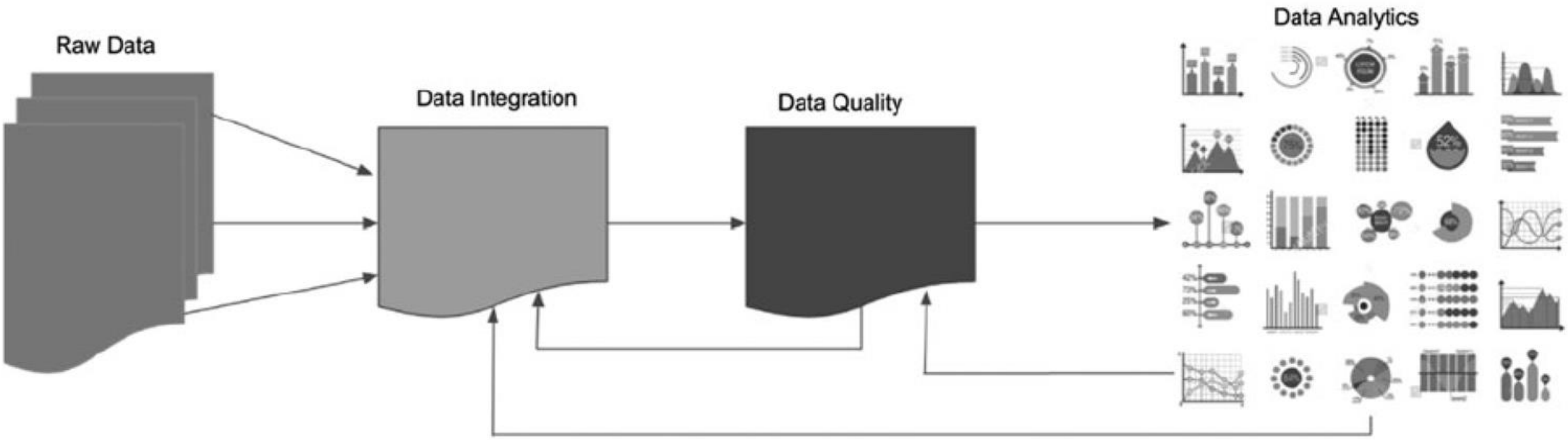
Lewis Alexander[1], Sanjiv R. Das[2,*] Zachary Ives[3] H.V. Jagadish[4] and Claire Monteleoni[5]

http://srdas.github.io/Papers/big.2016.0074_FINAL.pdf

Abstract
Significant research challenges must be addressed in the cleaning, transformation, integration, modeling, and analytics of Big Data sources for finance. This article surveys the progress made so far in this direction and obstacles yet to be overcome. These are issues that are of interest to data-driven financial institutions in both corporate finance and consumer finance. These challenges are also of interest to the legal profession as well as to regulators. The discussion is relevant to technology firms that support the growing field of FinTech.

Big Data
Volume 5, Number 3, 2017

DOI: 10.1089/big.2016.0074



Staging Template

Areas	*Type of issues/problems*			
	Level 1: Curation at the unit level within a firm	*Level 2: Curation and aggregation at the firm level*	*Level 3: Curation and aggregation at the system level*	*Across levels: Quality issues (privacy, veracity, etc.)*
Data integration				
Standards	□	○	○	□
Application-specific tools	□	○	○	○
Text mining tools	□	□	○	○
Data quality management				
BSBS239 (14 principles, 4 areas)	□	○	○	□
Errors in recording, extraction, entity-matching, interpretation	□	□	○	○
Data timeliness: Nowcasting	□	○	○	□
Data manipulation	□	□	○	○
Data analytics				
Feature selection	□	□	□	○
Model selection	□	□	□	○
Online learning	□	○	○	○
AI and deep learning	□	○	○	○
Systemic risk	□	□	□	□
Consumer finance	□	□	□	□
Text analytics	□	□	□	□
High frequency trading	□	—	○	○
Blockchains	□	—	○	○
Cybersecurity	□	□	□	○

Codes: ○ represents nascent solutions; □ represents work underway, not fully developed; and empty cells represent that decent progress has been made.

Systemic Analysis

The Dodd-Frank Act (2010) and Basel III regulations characterize a systemically risky FI as one that is

1. Large;
2. Complex;
3. Interconnected;
4. Critical, i.e., provides hard to substitute services to the economy.

The DFA does not provide quantification guidance.

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.

Attributes of Systemic Risk Measures

Systemic risk is an attribute of the economic system and not that of a single entity. Its measurement should have two important features:

1. Quantifiability (Aggregation): must be measurable on an ongoing basis.

1. Decomposability (Attribution): Aggregate system-wide risk must be broken down into additive risk contributions from all entities in the system.

Financial institutions that make large risk contributions to system-wide risk are deemed “systemically important.”

An Extensive Literature

References

Abbass, P., C. Brownlees, C. Hans, and N. Podlich (2016, April
nectedness: What does the market really know? *Journal of*
1–12.

Acemoglu, D., A. Ozdaglar, and A. Tahbaz-Salehi (2015). Syst
in financial networks. *American Economic Review 105*, 564

Acharya, V., R. Engle, and M. Richardson (2012). Capital sho
to ranking and regulating systemic risks. *American Econom*

Acharya, V., L. Pedersen, T. Philippon, and M. Richardson (
suring systemic risk. *Review of Financial Studies 30*(1), 2
New York University.

Acharya, V., P. Schnabl, and G. Suarez (2013). Securitization
Journal of Financial Economics 109, 515–536.

Acharya, V. V., J. A. C. Santos, and T. Yorulmazer (2010, New
Bank of New York, August)). Systemic risk and deposit
Economic Policy Review.

Adrian, T. and M. K. Brunnermeier (2016). Covar. *Am*
view 106(7), 1705–1741.

Adrian, T. and H. Shin (2010). Liquidity and leverage. *Jour*
mediation 19, 418–437.

Ahern, K. R. (2013). Network centrality and the cross-section o
ing Paper, USC-Marshall School of Business.

Allen, F. and E. Carletti (2013). What is systemic risk? *Jou*
and Banking 45, 121–127.

Allen, L., T. Bali, and Y. Tang (2012). Does systemic risk
predict future economic downturns? *Review of Financial St*

Anand, K., P. Gai, S. Kapadia, and S. Brennan (2013). A netv
system resilience. *Journal of Economic Behavior and Orga*

Avdjiev, S., M. Chui, and H. Shin (2014). Non-financial corpo
market economies and capital flows. *BIS Quarterly Review*.

Avramidis, P. and F. Pasiouras (2015). Calculating systemic
model approach. *Journal of Financial Stability 16*, 138–150

Benoit, S., J. Colliard, C. Hurlin, and C. Perignon (2017). \
survey on systemic risk. *Review of Finance 21*(1), 109–152.

Betz, F., N. Hautsch, T. A. Peltonend, and M. Schienle (2016). Syster
in the european banking and sovereign network. *Journal of Finan*
206–224.

Bianchiy, D., M. Billio, R. Casarinz, and G. Massimo (2015). Mo
and systemic risk. Working Paper, University of Warwick.

Billio, M., M. Getmansky, D. Gray, A. Lo, R. Merton, and L. P
Sovereign, bank and insurance credit spreads: Connectedness and s
Working Paper, International Monetary Fund.

Billio, M., M. Getmansky, A. W. Lo, and L. Pelizzon (2012a). Econ
of connectedness and systemic risk in the finance and insurance se
Financial Economics 104(3), 535–559.

Bisias, D., M. Flood, A. Lo, and S. Valavanis (2012). A survey
analytics. *Annual Review of Financial Economics 4*, 255–296.

Black, L., R. Correa, X. Huang, and H. Zhou (2016). The syster
pean banks during the financial and sovereign debt crises. *Journa*
Finance 63, 107–125.

Bluhm, M. and J. P. Krahnen (2014). Systemic risk in an interco
system with endogenous asset markets. *Journal of Financial Stab*

Bonacich, P. (1987). Power and centrality: A family of measures. *A*
of Sociology 92(5), 1170–1182.

Bonacich, P. and P. Lloyd (2001, July). Eigenvector-like measures
asymmetric relations. *Social Networks 23*(3), 191–201.

Borri, N. (2017). Local currency systemic risk. Working Paper, ssrn

Brownlees, T. and R. Engle (2015). Srisk: A conditional capital sl
systemic risk measurement. Working Paper, New York University

Brunetti, C., J. H. Harris, S. Mankad, and G. Michailidis (2015).
ness in the interbank market. Finance and Economics Discussion
Governors of the Federal Reserve System.

Brunnermeier, M. (2009). Deciphering the liquidity and credit crun
Journal of Economic Perspectives 23, 77–100.

Brunnermeier, M. and L. Pedersen (2009). Market liquidity and f
Review of Financial Studies 22(6), 2201–2238.

Chan-Lau, J., M. A. Espinosa-Vega, K. Giesecke, and J. Solé (2009
systemic implications of financial linkages. IMF Global Financial S
Vol. 2.

ChanLau, J. A., C. Chuang, J. Duan, and W. Sun (2016, May).
and systemic risk via forwardlooking partial default correlation
IMF.

Colliard, J.-E., T. Foucault, and P. Hoffmann (2017). Interbank
mented otc market. Working Paper, European Central Bank.

Covitz, D., N. Liang, and G. Suarez (2013). The evolution of
Collapse of the asset-backed commercial paper market. *Journ*
815–848.

Das, S. R. (2016). Matrix metrics: Network-based systemic risk s
Alternative Investments 18(4), 33–51.

Das, S. R., S. R. Kim, and D. N. Ostrov (2017). Dynamic syste
Working Paper, Santa Clara University.

Das, S. R. and J. Sisk (2005). Financial communities. *Journal of*
ment 31(4), 112–133.

De Bandt, O. and P. Hartmann (2000). Systemic risk: A survey
European Central Bank.

Demirer, M., F. X. Diebold, L. Liu, and K. Yilmaz (2017). Estin
network connectedness. *Journal of Applied Econometrics*.

Diebold, F. and K. Ylmaz (2014). On the network topology of va
tions: Measuring the connectedness of financial firms. *Journal of*
119–134.

Duan, J.-C. and W. Miao (2016). Default correlations and lar
analysis. *Journal of Business & Economic Statistics 34*(4), 536

Duarte, F. and T. M. Eisenbach (2015). Fire-sale spillovers and syst
Paper, Federal Reserve Bank.

Elliott, M., B. Golub, and M. Jackson (2014). Financial netwo
American Economic Review 104, 3115–3153.

Freeman, L. (1977). A set of measures of centrality based on betw
try 40, 35–41.

Gabrieli, S. and C.-P. Georg (2014). A network view on interba
Working Paper, Banque de France.

Gale, D. M. and S. Kariv (2007). Financial networks. *American*
Papers and Proceedings.

Giglio, S., B. Kelly, and S. Pruitt (2016). Systemic risk and the m
empirical evaluation. *Journal of Financial Economics 119*(3), 4

Gobat, J., T. Barnhill, A. Jobst, T. Kisinbay, H. Oura, T. S
(2011). How to address the systemic part of liquidity ris

Goodhart, C. (2009, August). Liquidity management. Jack
ity and Macroeconomic Policy Symposium, Federal Res

Gorton, G. and A. Metrick (2012). Securitized banking and
of Financial Economics 104, 425–451.

Hanson, S., A. Kashyap, and J. Stein (2011). A macropruden
regulation. *Journal of Economic Perspectives 25*, 3–28.

Härdle, W. K., W. Wang, and L. Yuc (2016). Tenet: Tail-
Journal of Econometrics 192(2), 499–513.

Hautsch, N., J. Schaumburg, and M. Schienle (2015). Finan
contributions. *Review of Finance 19*, 685–738.

Huang, X., H. Zhou, and H. Zhu (2012). Systemic risk
Financial Services Research 42, 55–83.

Karolyi, A., J. Sedunov, and A. Taboada (2016). Cross-b
temic risk. Working Paper, Cornell University.

Kitwiwattanachai, C. (2015). Learning network structur
from cds data. Working Paper, University of Connectic

Laeven, L., L. Ratnovski, and H. Tong (2016). Bank size,
Some international evidence. *Journal of Banking and F*

Lehar, A. (2005). Measuring systemic risk: A risk manage
Banking and Finance 29, 2577–2603.

Li, J. and G. Zinna (2014). On bank credit risk: Systemic
for the united states and united kingdom. *Journal of F*
Analysis 5/6, 1403–1442.

Liang, N. (2013). Systemic risk monitoring and financial st
Credit and Banking 45, 129–135.

Liu, S., C. Wu, C.-Y. Yeh, and W. Yoo (2015). What d
evidence from the us state cds market. Working Paper,

Markose, S., S. Giansante, and A. Shaghaghi (2012). 'T
financial network of us cds market: Topological fragility
of Economic Behavior and Organization 83, 627–646.

Merton, R. C. (1973). Theory of rational option pricing.
and Management Science 4, 141–183.

Nier, E., J. Yang, T. Yorulmazer, and A. Alentorn (2007). Network models and financial stability. *Journal of Economic Dynamics and Control 31*, 2033–2060.

Nucera, F., B. Schwaab, S. J. Koopman, and A. Lucas (2016). The information in systemic risk rankings. *Journal of Empirical Finance 38*, 461–475.

Oh, D. H. and A. J. Patton (2016). Time-varying systemic risk: Evidence from a dynamic copula model of cds spreads. *Journal of Business and Economic Statistics forthcoming*.

Pagano, M. S. and J. Sedunov (2016). A comprehensive approach to measuring the relation between systemic risk exposure and sovereign debt. *Journal of Financial Stability 23*, 62–78.

Perotti, E. and J. Suarez (2009). Liquidity risk charges as a macroprudential tool. CEPR Policy Insight.

Poledna, S., J. L. Molina-Borboad, S. Martínez-Jaramillod, M. van der Leije, and S. Thurner (2015). The multi-layer network nature of systemic risk and its implications for the costs of financial crises. *Journal of Financial Stability 20*, 70–81.

Saldías, M. (2013). Systemic risk analysis using forward-looking distance-to-default series. *Journal of Financial Stability 9*, 498–517.

Schwarcz, S. (2008). Systemic risk. *Georgetown Law Journal 97*, 193–249.

Sedunov, J. (2016). What is the systemic risk exposure of financial institutions? *Journal of Financial Stability 24*, 71–87.

Sensoya, A. (2017). Firm size, ownership structure, and systematic liquidity risk: The case of an emerging market. *Journal of Financial Stability forthcoming*.

Silva, W., H. Kimura, and A. Sobreiro (2017). An analysis of the literature on systemic financial risk: A survey. *Journal of Financial Stability 28*, 91–114.

Tasca, P., P. Mavrodiev, and F. Schweitzer (2014). Quantifying the impact of leveraging and diversification on systemic risk. *Journal of Financial Stability 15*, 43–52.

Billio, Getmansky, Lo, Pelizzon (2012)

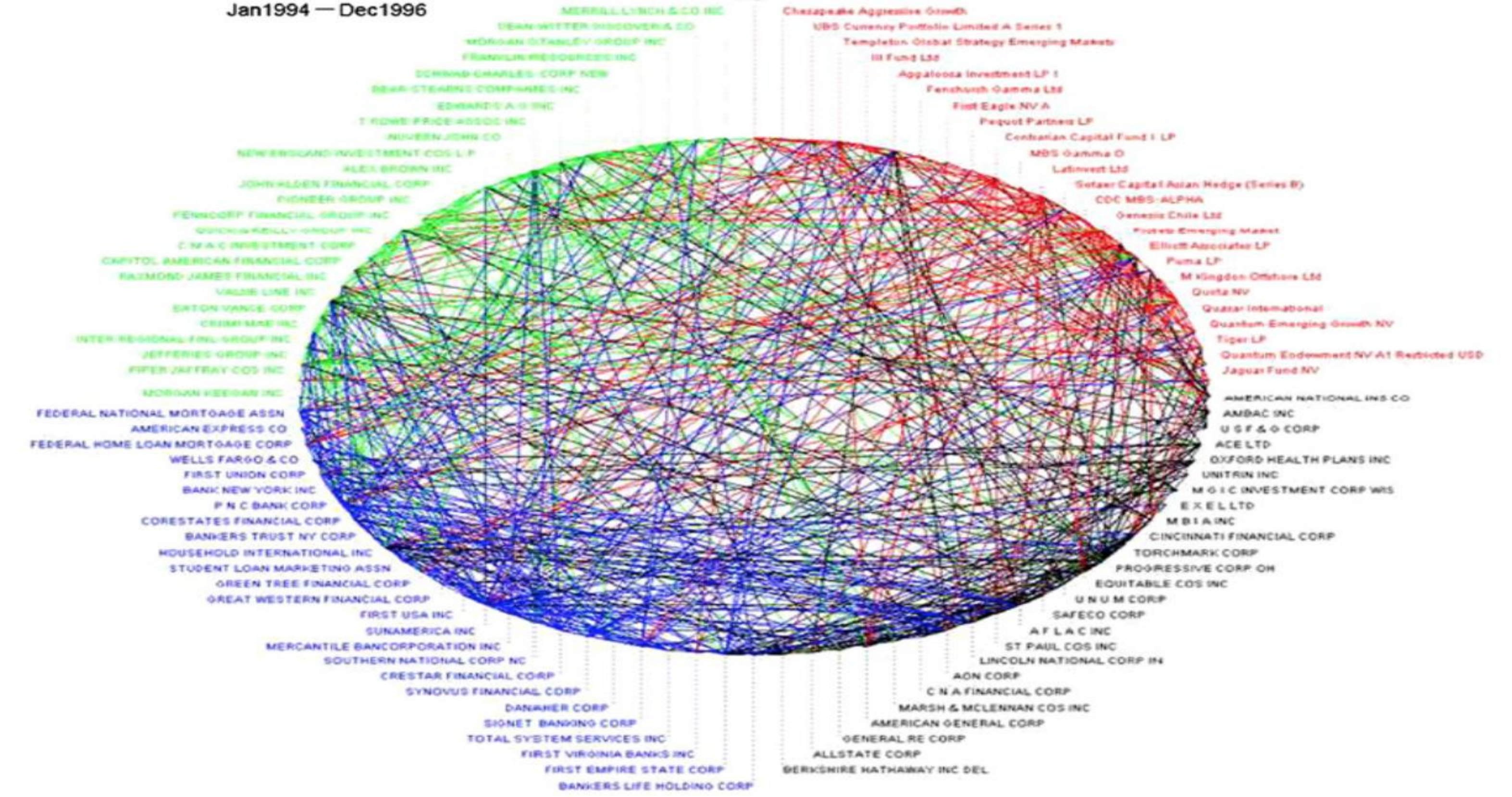
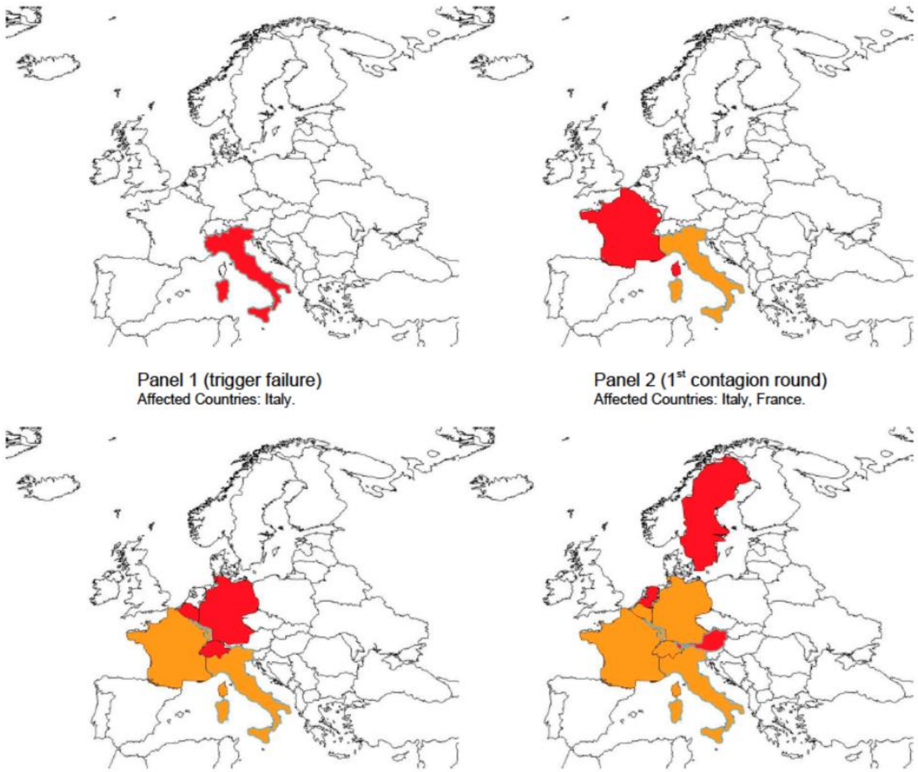


Fig. 2. Network diagram of linear Granger-causality relationships that are statistically significant at the 5% level among the monthly returns of the 25 largest (in terms of average market cap and AUM) banks, broker/dealers, insurers, and hedge funds over January 1994 to December 1996. The type of institution causing the relationship is indicated by color: green for broker/dealers, red for hedge funds, black for insurers, and blue for banks. Granger-causality relationships are estimated including autoregressive terms and filtering out heteroskedasticity with a GARCH(1,1) model.

Contagion Networks (Espinosa-Vega & Sole, IMF 2010)



Panel 1 (trigger failure)
Affected Countries: Italy.

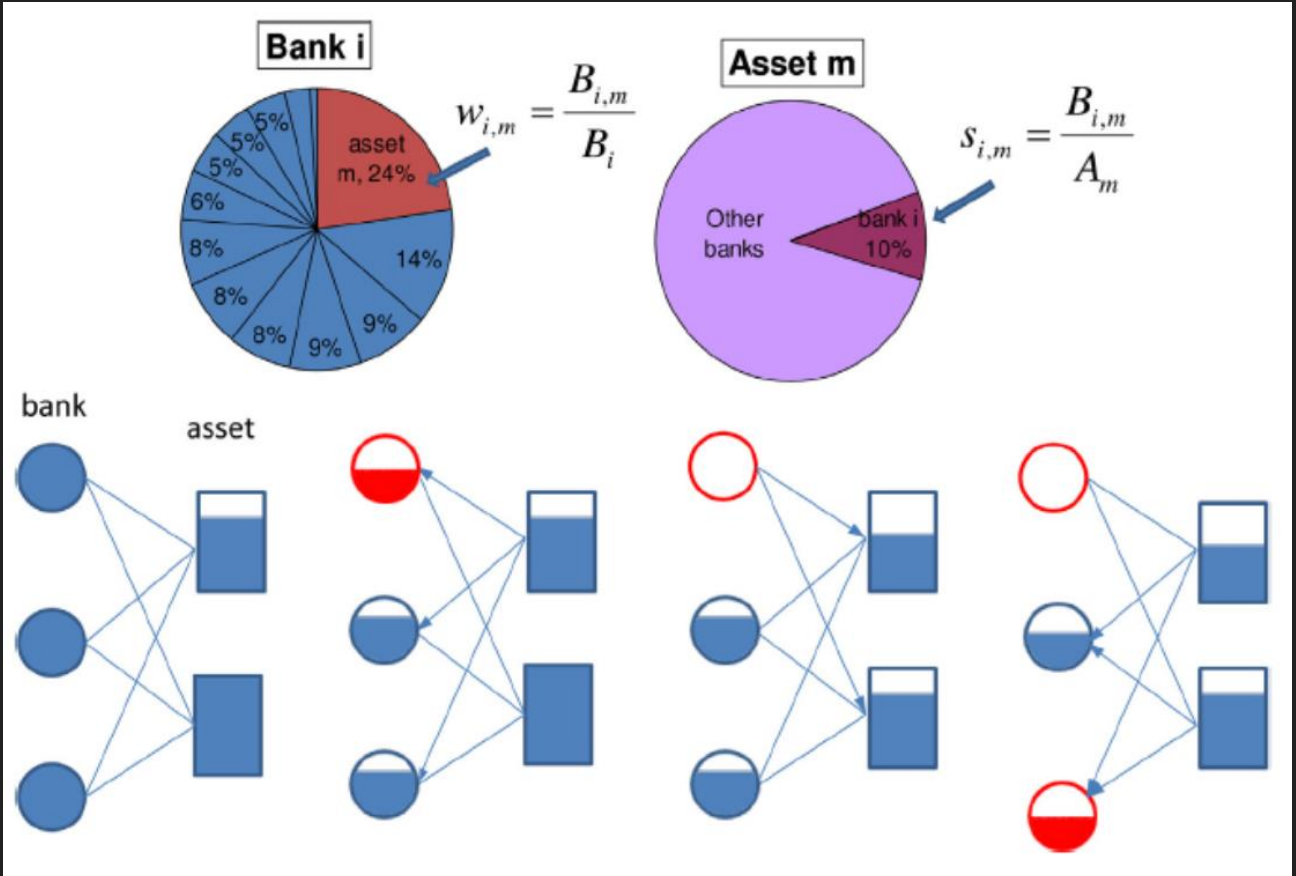
Panel 2 (1st contagion round)
Affected Countries: Italy, France.

Panel 3 (2nd contagion round)
Affected Countries: Italy, France, Belgium, Germany, Switzerland.

Panel 4 (final round)
Affected Countries: Italy, France, Belgium, Germany, Switzerland, Austria, Sweden, Netherlands.

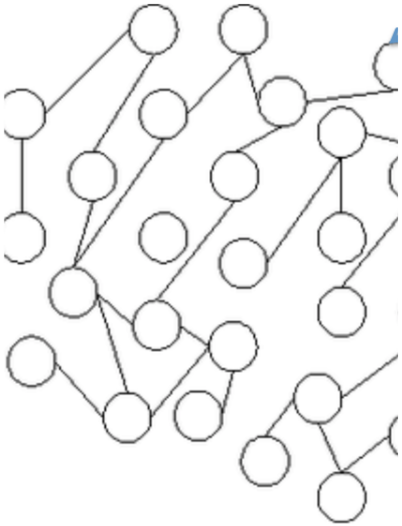
Bivalent Networks

Levy-Carciente, Kennet, Avakian, Stanley, Havlin, JBF 2015



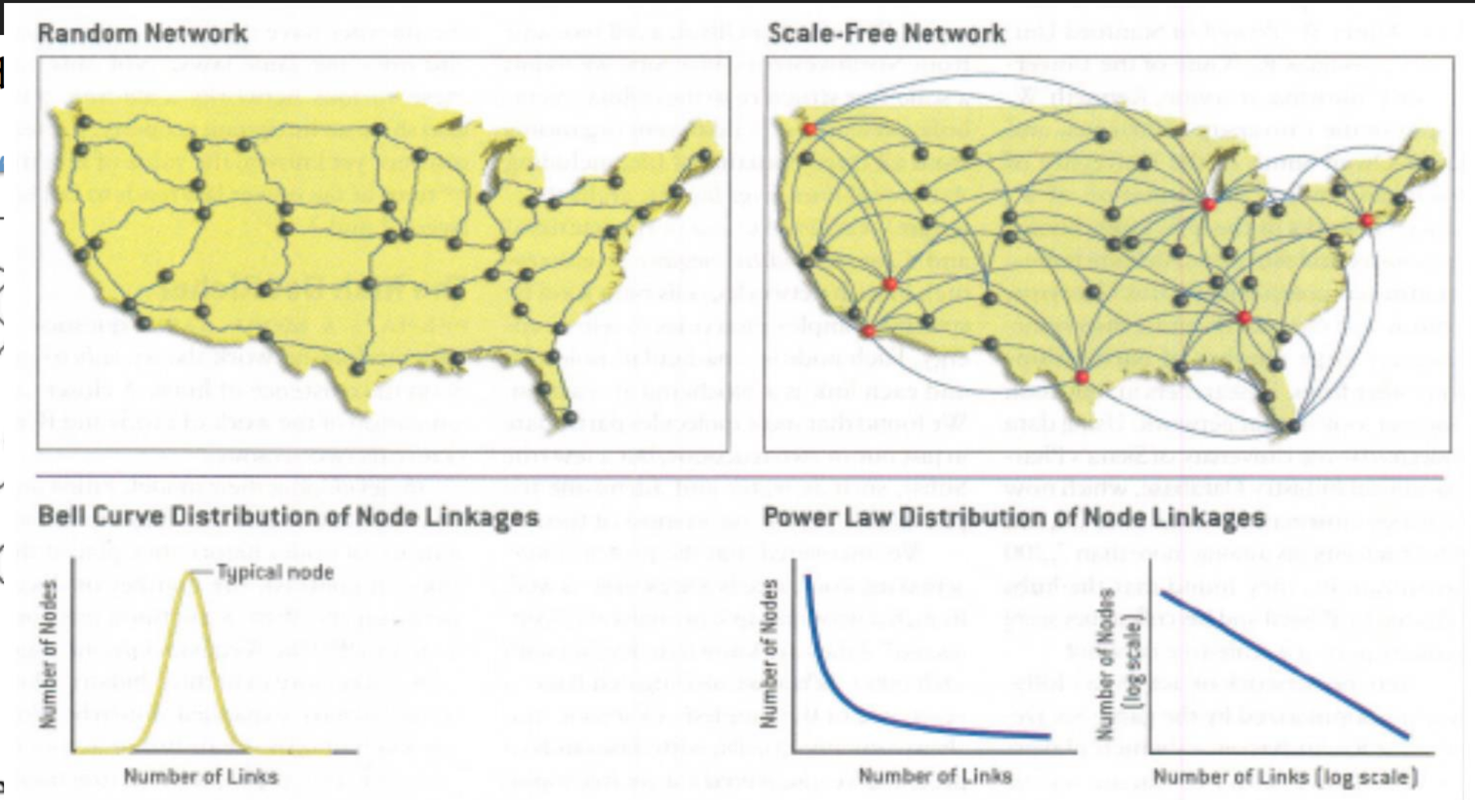
Graph Theory

Nod



(a) Random ne

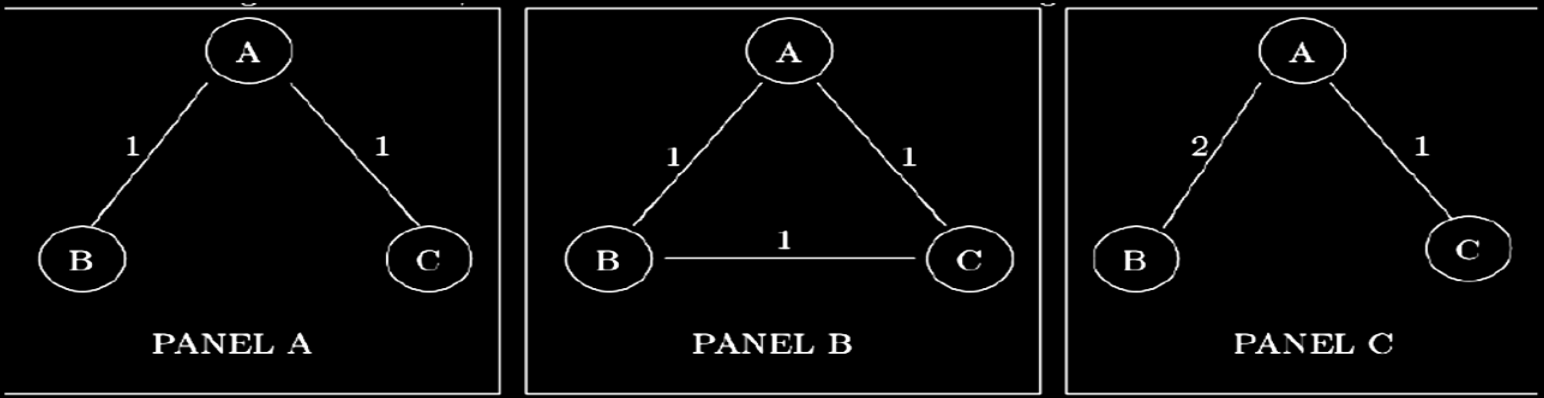
$f(d) \sim N($



Barabasi, Sciam, May 2003

Centrality (Bonacich 1987)

- Similar to 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.



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

Centrality scores = {0.71, 0.50, 0.50}

Centrality scores = {0.58, 0.58, 0.58}

Centrality scores = {0.71, 0.63, 0.32}

Diameter

Longest sh

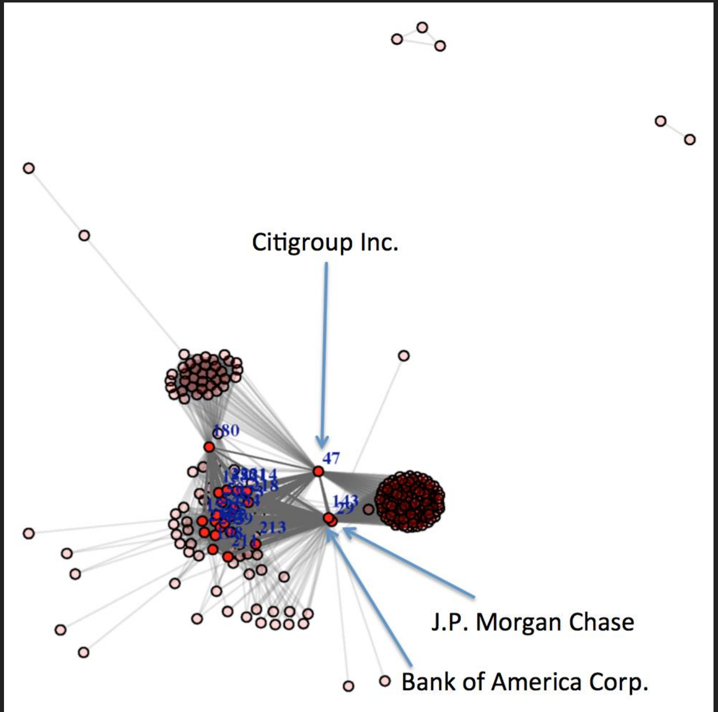
The diame

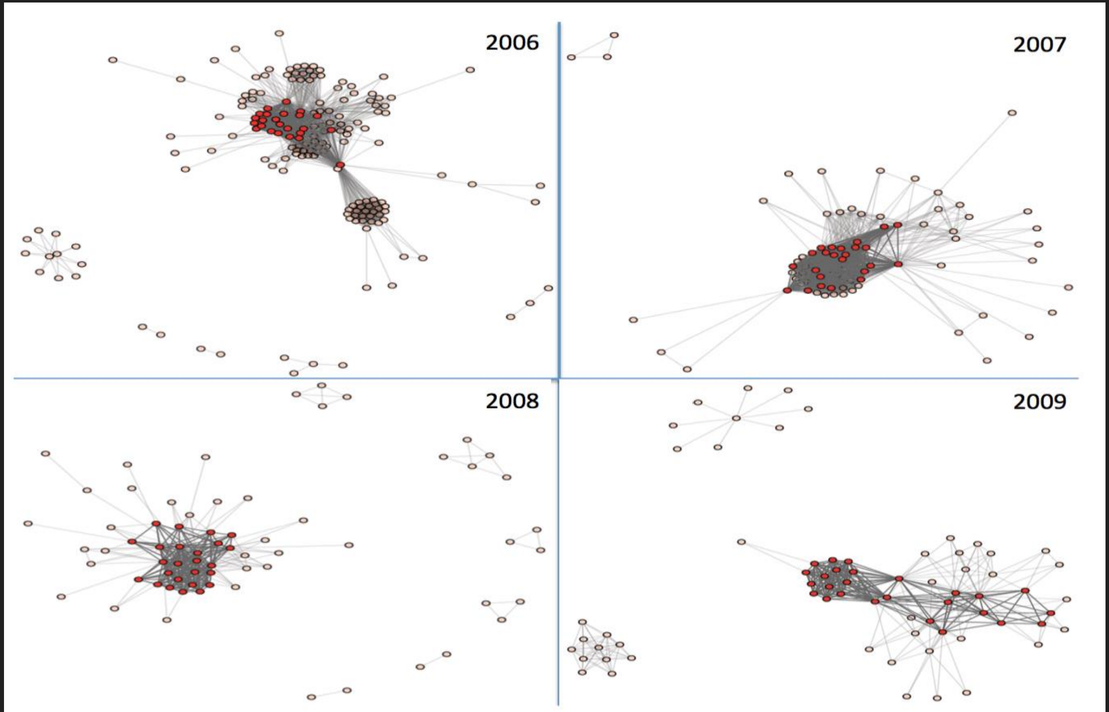
Fragility

- 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.
- Similar to a normalized Herfindahl Index.
- Fragility of the sample network $= 20$

Interbank Loan Networks (U.S.)

“Extracting, Linking and Integrating Data from Public Sources: A Financial Case Study,” (2011), (Douglas Burdick, Sanjiv Das, Mauricio A. Hernandez, Howard Ho, Georgia Koutrika, Rajasekar Krishnamurthy, Lucian Popa, Ioana Stanoi, Shivakumar Vaithyanathan), *IEEE Data Engineering Bulletin*, 34(3), 60-67.





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

One Score for Systemic Risk

$$S = \frac{1}{n}\sqrt{C^\top \cdot A \cdot C} \geq 0$$

banks (normalization across time)

Adjacency matrix

A(i,j) in (0,1)
A(i,i) = 1

Vector of credit risk scores {PD, rating, etc}. Higher = more risk

C(i) > 0

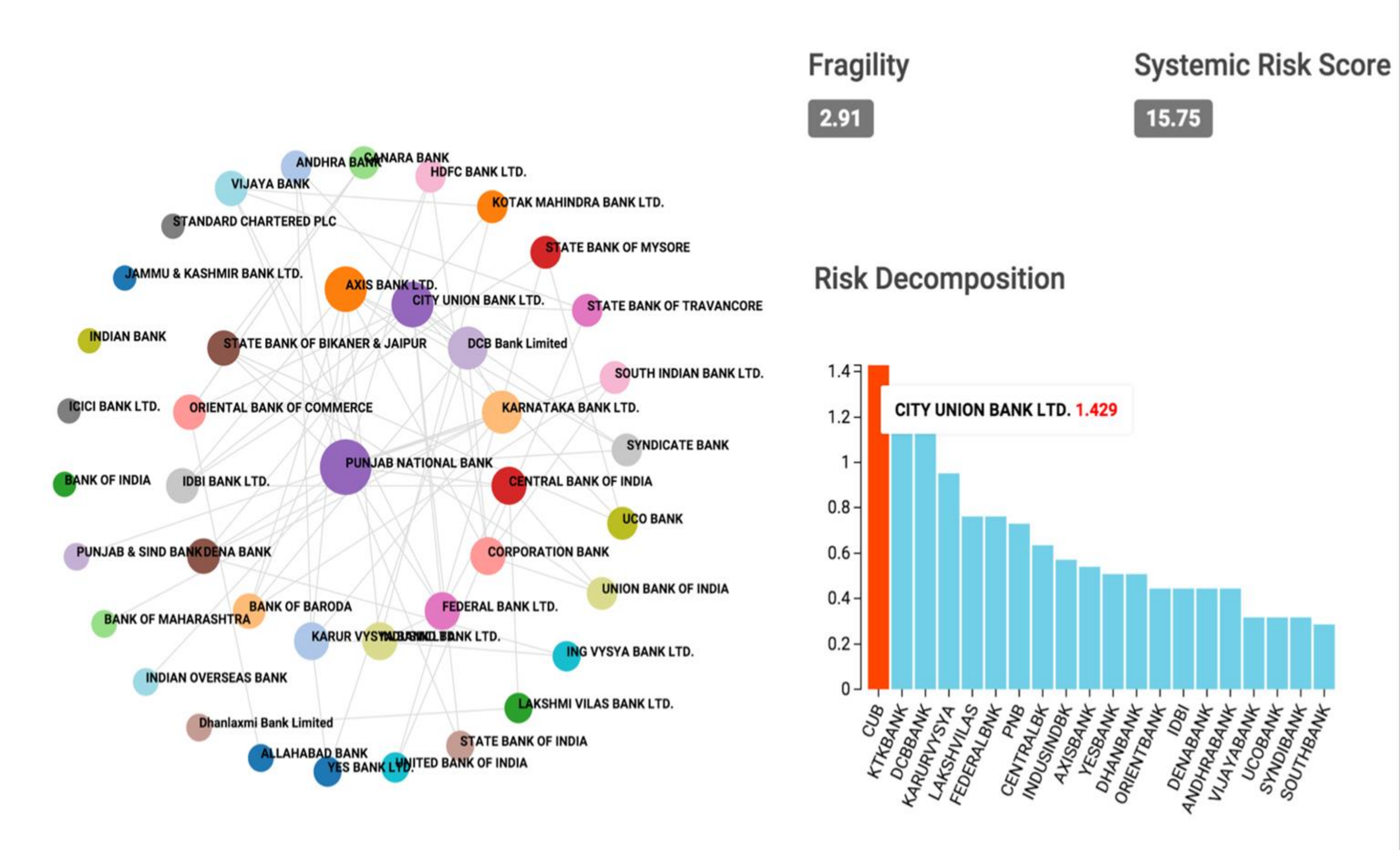
S(C,A) is linear homogenous in *C*

Apply Euler's Formula

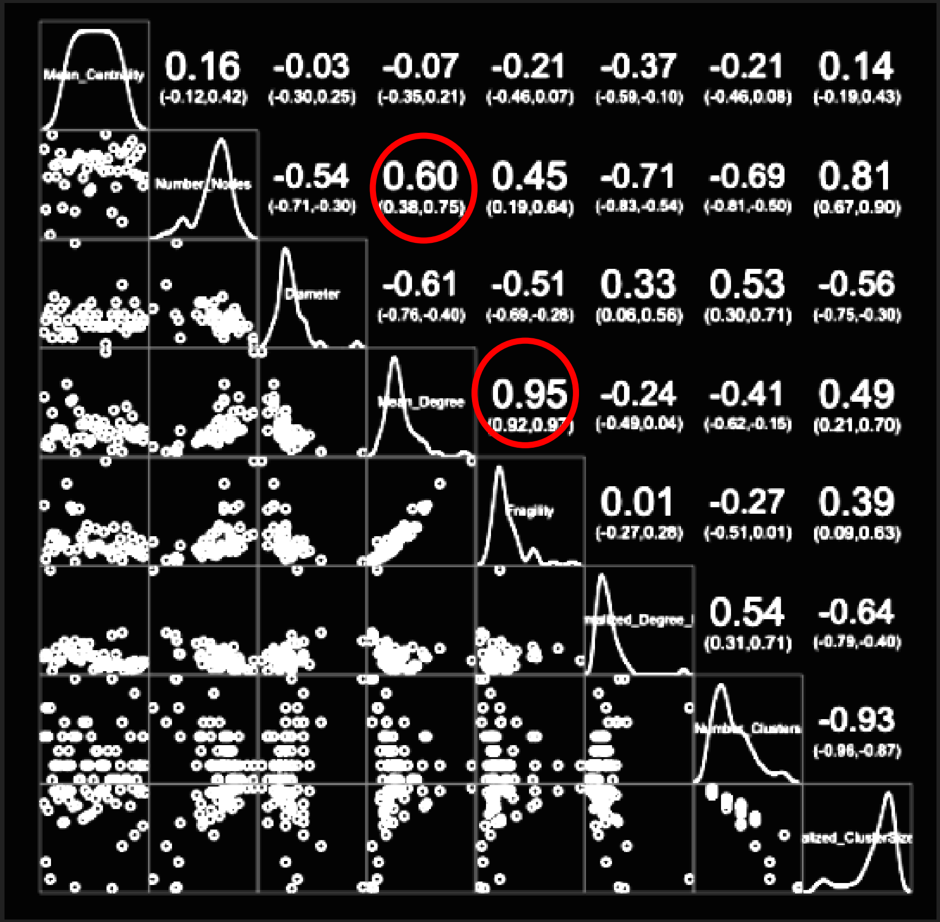
$$S = \frac{\partial S}{\partial C_1} C_1 + \frac{\partial S}{\partial C_2} C_2 + \ldots + \frac{\partial S}{\partial C_n} C_n = \sum_{i=1}^{n} \frac{\partial S}{\partial C_i} C_i$$

Risk Contribution

First iteration : India



Correlations



Mean Centrality

Number of Nodes

Diameter

Mean Degree

Fragility

Normalized degree Herfindahl Index

Number of Clusters

Normalized cluster size Herfindahl

Probabilities of Default (PDs)

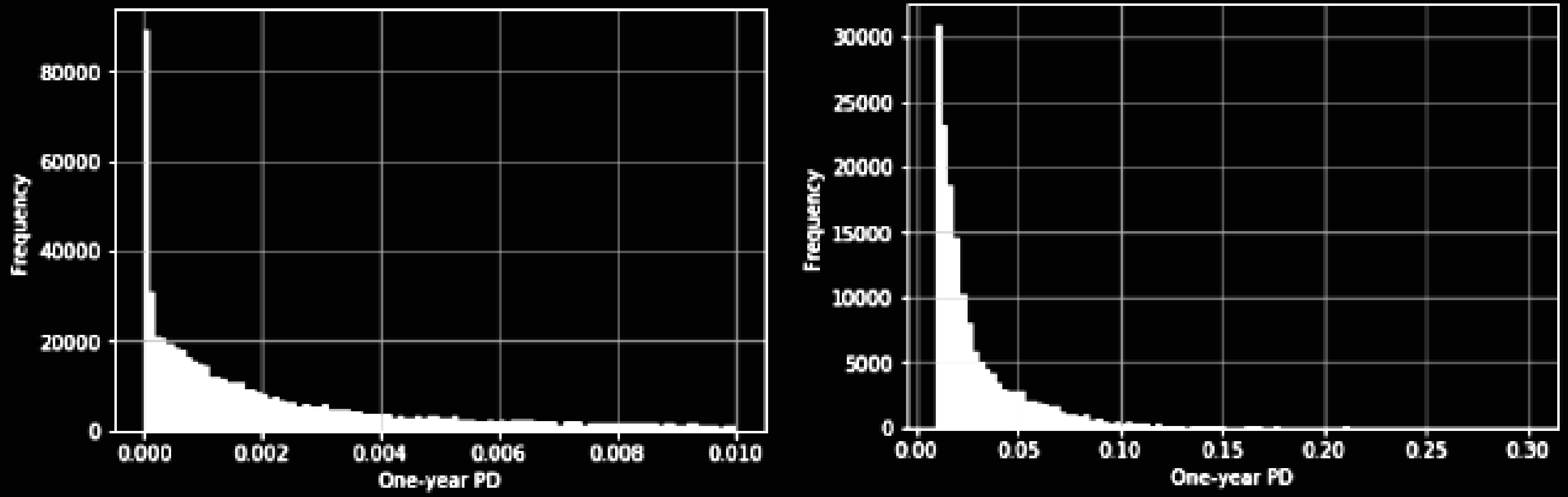


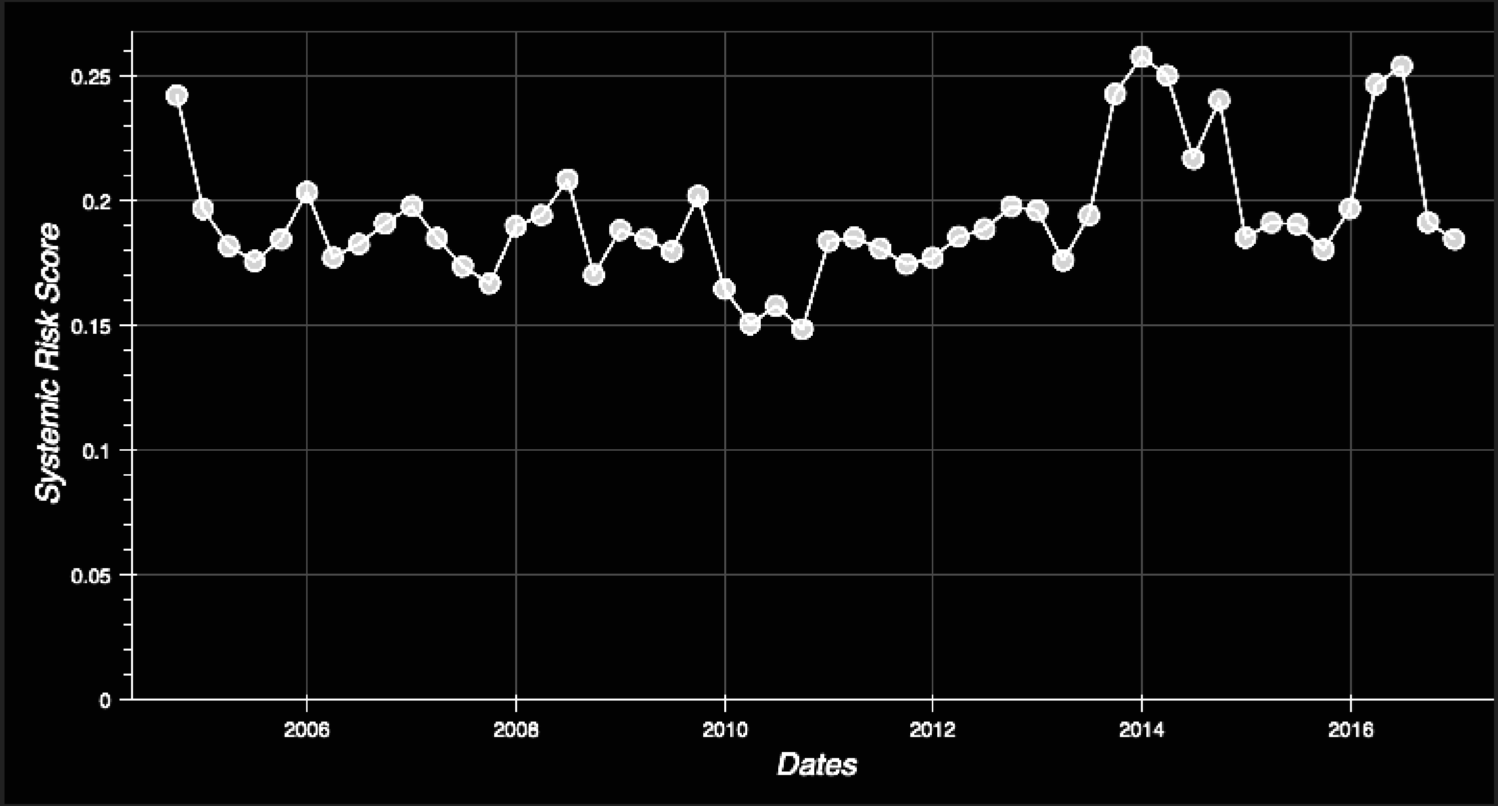
Figure 11: Distribution of PDs of all Indian FIs from 2004 to 2016. The first plot is the histogram of PDs that lie in the interval $(0, 0.01)$, and the second in the interval $(0.01, 0.30)$.

Highest PD = 26.36%

$C = 1 + 30\ PD$

Since $PD < 0.30$, C lies in $(0,10)$

Systemic Risk Score (*S*)



Correlation of PDs and S = 69.7%

Risk Contributions of top 20 banks

	2005-Q1		2016-Q1	
	Bank Name	Risk Decomp	Bank Name	Risk Decomp
1	PRIME SECURITIES	2.705139	BANK OF MAHARASHTRA	2.222866
2	STATE BANK OF INDIA	2.476634	UCO BANK	1.698109
3	UCO BANK	2.438924	POWER FINANCE	1.437113
4	CORPORATION BANK	1.882045	UNITED BANK OF INDIA	1.410672
5	GIC HOUSING FINANCE	1.771204	STATE BK.OF BIN.& JAIPUR SUSP - SUSP.15/03/17	1.388539
6	I N G VYSYA BANK SUSP - SUSP.15/04/15	1.696898	DENA BANK	1.343904
7	UNION BANK OF INDIA	1.607279	STATE BANK OF INDIA	1.335314
8	IFCI	1.597618	INDIAN OVERSEAS BANK	1.331388
9	SUNDARAM FINANCE	1.569000	BANK OF TRAVANCORE SUSP - SUSP.15/03/17	1.309907
10	P N B GILTS	1.492469	CIL SECURITIES	1.282169
11	DHANLAXMI BANK	1.328556	COMFORT COMMOTRADE	1.137495
12	JAMMU & KASHMIR BANK	1.322932	BANK OF BARODA	1.093183
13	INDIABULLS FINL.SVS. SUSP - SUSP.18/03/13	1.215547	ANDHRA BANK	1.066791
14	DEWAN HOUSING FINANCE	1.198211	DEWAN HOUSING FINANCE	0.994385
15	ALMOND GLOBAL SECURITIES	1.195593	ORIENTAL BK.OF COMMERCE	0.917884
16	DENA BANK	1.194755	JAGSONPAL FIN.& LSG.	0.917517
17	ANDHRA BANK	1.193921	ELIXIR CAPITAL	0.873306
18	INDUSIND BANK	1.163923	MAHA.& MAHA.FINL.SVS.	0.871946
19	MARGO FINANCE	1.163827	CUBICAL FINANCIAL SVS.	0.855089
20	UNITED CREDIT	1.148539	VAX HOUSING FINANCE	0.852056
	TOTAL	31.36301	TOTAL	24.33963

Explaining quarterly systemic risk

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Constant	0.1580***	0.1498***	0.2685	−0.0112	−0.0112	0.2730	0.2730
	(26.93)	(9.57)	(1.41)	(−0.04)	(−0.04)	(1.45)	(1.45)
Mean PD	3.8253***		5.2279***	5.0884***	5.0884***	5.2666***	5.2666***
	(6.73)		(18.85)	(9.50)	(9.50)	(10.45)	(10.45)
Mean Degree		0.0041*	0.0134***	0.0065	0.0065	0.0130**	0.0130**
		(2.30)	(3.58)	(1.58)	(1.58)	(2.76)	(2.76)
Degree HHI		6.4870*	5.7260*	4.3454	4.3454	6.2504**	6.2504**
		(2.42)	(2.55)	(2.01)	(2.01)	(3.00)	(3.00)
Mean Bet. Centrality			−0.0001***	−0.0001**	−0.0001**	−0.0001*	−0.0001*
			(−4.65)	(−3.05)	(−3.05)	(−2.67)	(−2.67)
Diameter			0.0003	0.0002	0.0002	−0.0002	−0.0002
			(0.50)	(0.31)	(0.31)	(−0.26)	(−0.26)
Fragility			−0.0039	0.0004	0.0004	−0.0034	−0.0034
			(−1.72)	(0.15)	(0.15)	(−1.26)	(−1.26)
Num. Clusters			−0.0030	−0.0014	−0.0014	−0.0033	−0.0033
			(−1.17)	(−0.52)	(−0.52)	(−1.35)	(−1.35)
Cluster HHI			−0.1429	−0.0110	−0.0110	−0.2230	−0.2230
			(−0.76)	(−0.05)	(−0.05)	(−1.20)	(−1.20)
Median Log(Assets)				0.0046	0.0046		
				(1.04)	(1.04)		
Median Log(Market Cap)						0.0040*	0.0040*
						(2.65)	(2.65)
Median Loans/Assets				0.0001	0.0001	0.0122	0.0122
				(0.01)	(0.01)	(0.97)	(0.97)
Median Loans/Deposits				0.0564	0.0564	−0.0035	−0.0035
				(0.59)	(0.59)	(−0.05)	(−0.05)
Median Debt/Assets				0.1058	0.1058		
				(1.29)	(1.29)		
Median Debt/Equity						0.1224*	0.1224*
						(2.16)	(2.16)
Median Debt/Capital				−0.0000	−0.0000	0.0002	0.0002
				(−0.04)	(−0.04)	(0.89)	(0.89)
Median ROA				0.0015	0.0015		
				(1.64)	(1.64)		
Median ROE						0.0001	0.0001
						(0.06)	(0.06)
Median Market/Book				0.0096	0.0096	−0.0020	−0.0020
				(1.23)	(1.23)	(−0.22)	(−0.22)
Observations	50	50	50	50	50	50	50
R^2	0.485	0.160	0.923	0.948	0.948	0.955	0.955
Adjusted R^2	0.475	0.124	0.908	0.925	0.925	0.935	0.935

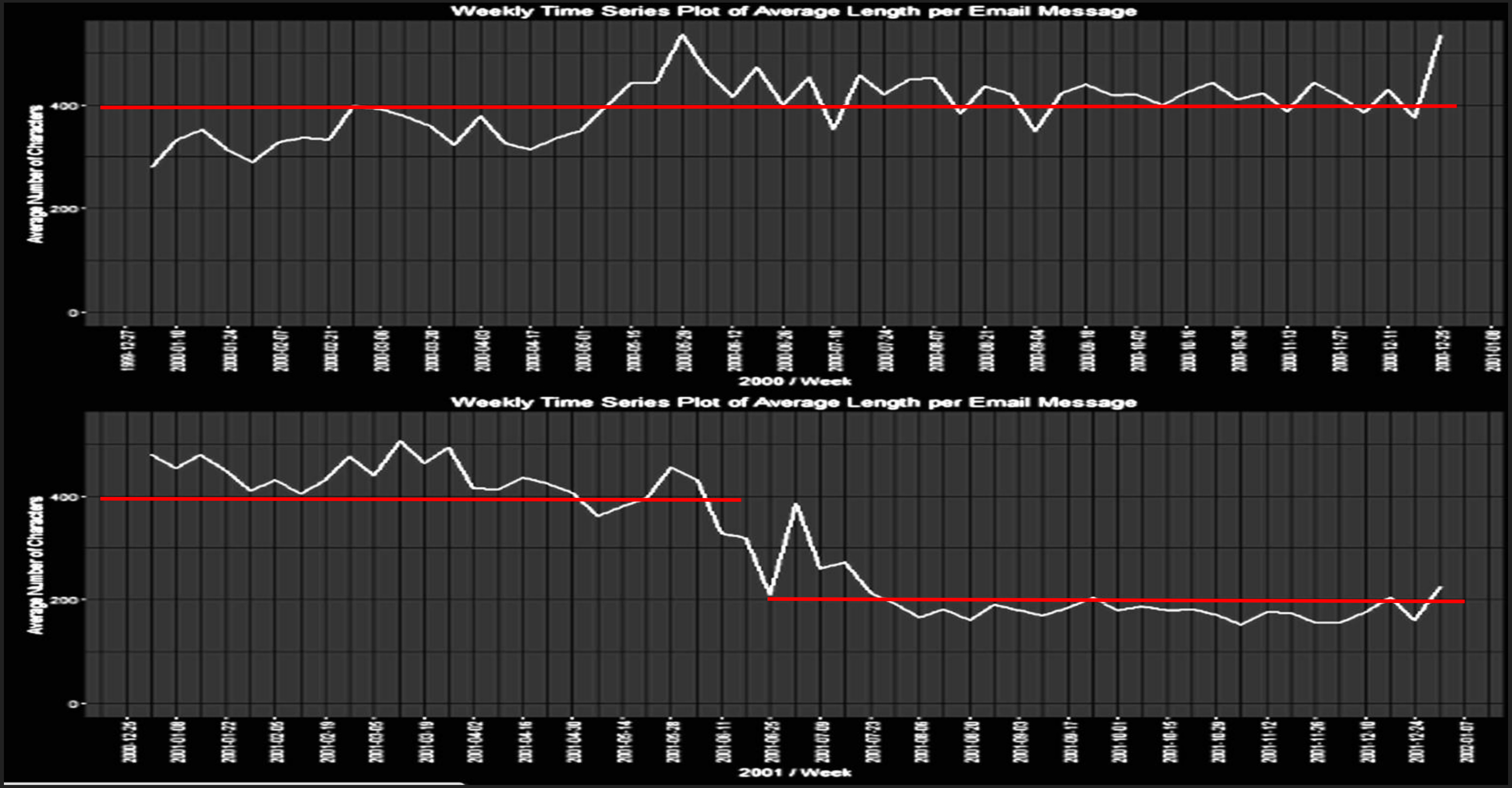
t statistics in parentheses
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Text, Sentiment, and RegTech

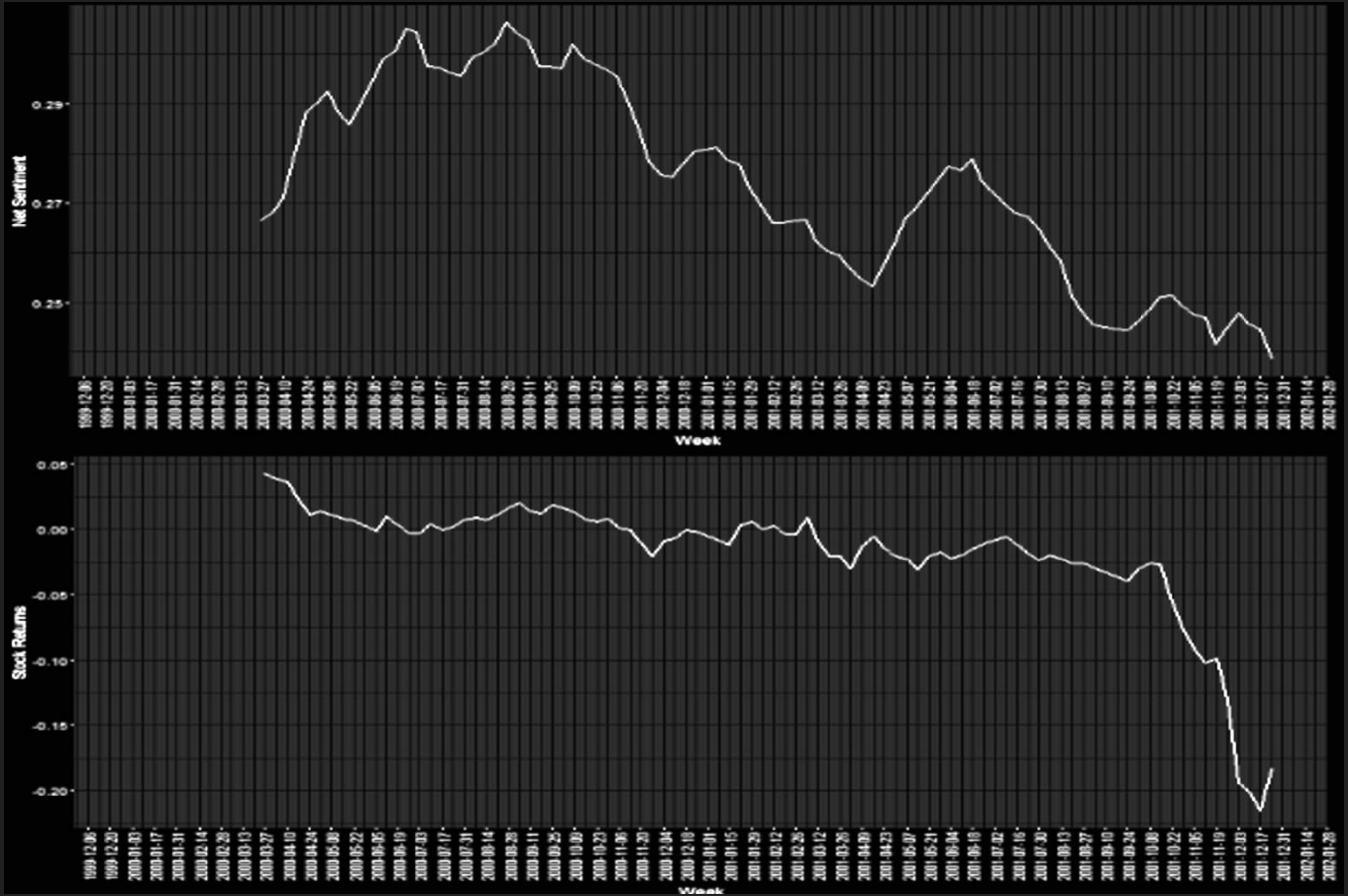
Zero-Revelation Linguistic Regulation: Detecting Risk Through Corporate Emails and News (Das, Kim, Kothari 2016)

- Financials are often delayed indicators of corporate quality.
- Internal discussion may be used as an early warning system for upcoming corporate malaise.
- Emails have the potential to predict such events.
- Software can analyze vast quantities of textual data not amenable to human processing.
- Corporate senior management may also use these analyses to better predict and manage impending crisis for their firms.
- The approach requires zero revelation of emails.

Enron: Email Length



Enron: Sentiment and Returns



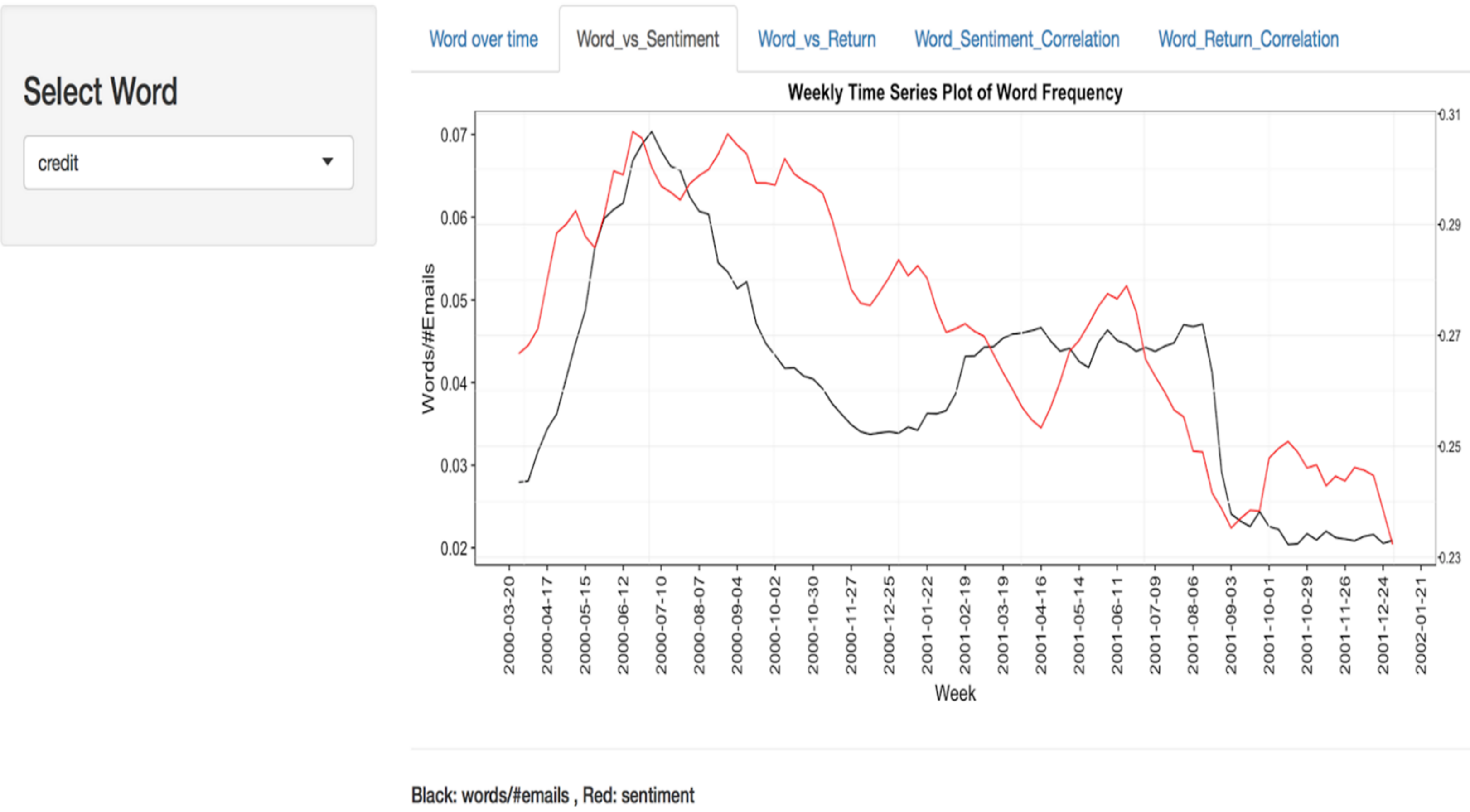
Enron: Returns and Characteristics

Variable	Coefficient Estimate (*t*-statistic)			
	(1)	(2)	(3)	(4)
MA Net Sentiment $_t$	XXX*** (XXX)	0.575 (0.63)	2.330*** (3.14)	-1.397 (-1.25)
MA Email Length $_t$		0.584*** (2.97)		1.046*** (4.19)
MA Total Emails $_t$			-0.004 (-0.10)	-0.131*** (-2.83)
Intercept		-0.406* (-1.93)	-0.671*** (-3.08)	0.117 (0.43)
Adjusted *R*-squared	XXX		0.09	0.24
Number of observations	88	88	88	88

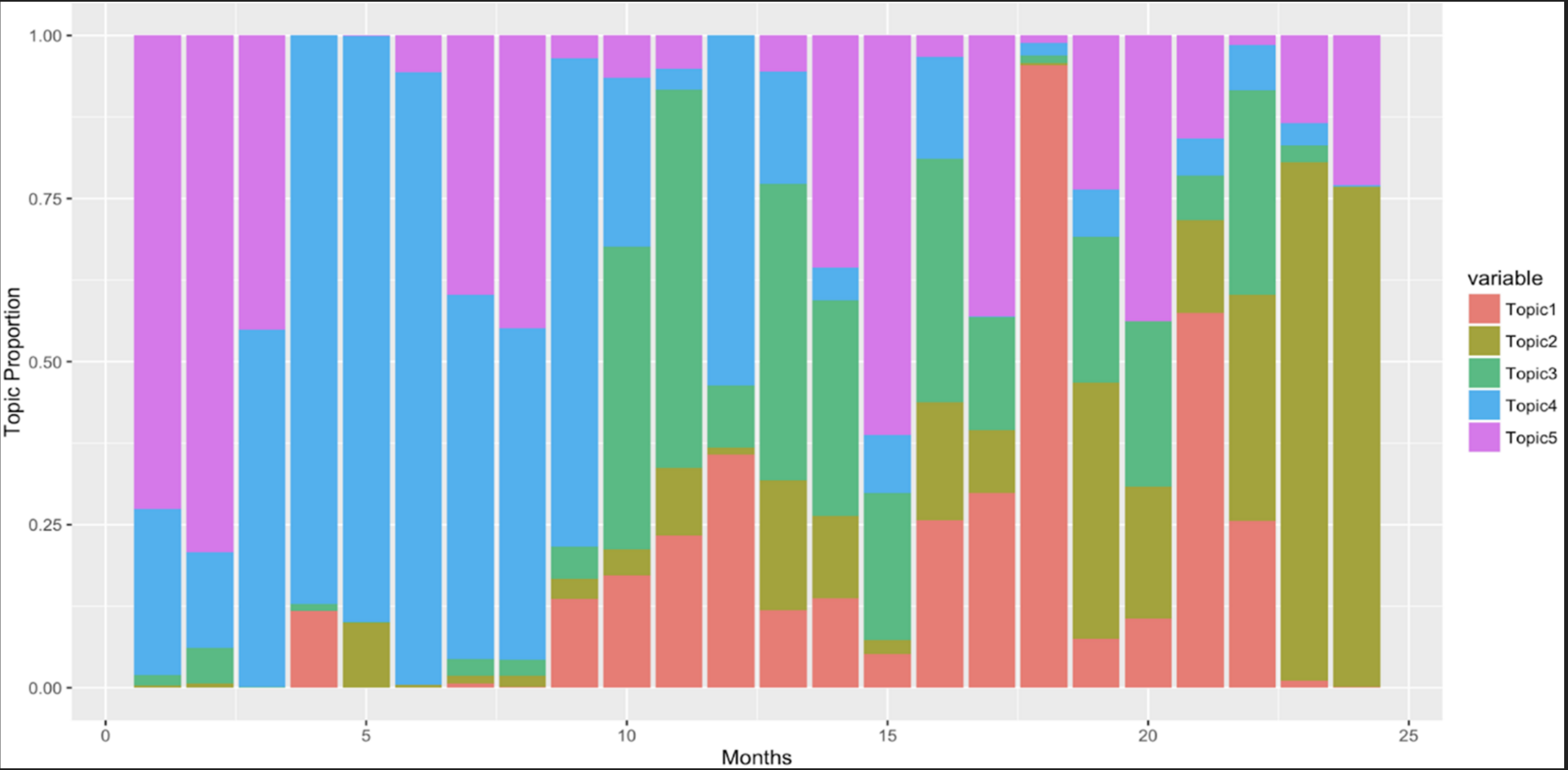
Enron: Returns and Characteristics

Variable	Coefficient Estimate (*t*-statistic)			
	(1)	(2)	(3)	(4)
MA Net Sentiment $_t$	XXX***	0.575	2.330***	-1.397
	(XXX)	(0.63)	(3.14)	(-1.25)
MA Email Length $_t$		0.584***		1.046***
		(2.97)		(4.19)
MA Total Emails $_t$			-0.004	-0.131***
			(-0.10)	(-2.83)
Intercept		-0.406*	-0.671***	0.117
		(-1.93)	(-3.08)	(0.43)
Adjusted *R*-squared	XXX		0.09	0.24
Number of observations	88	88	88	88

Enron: WordPlay

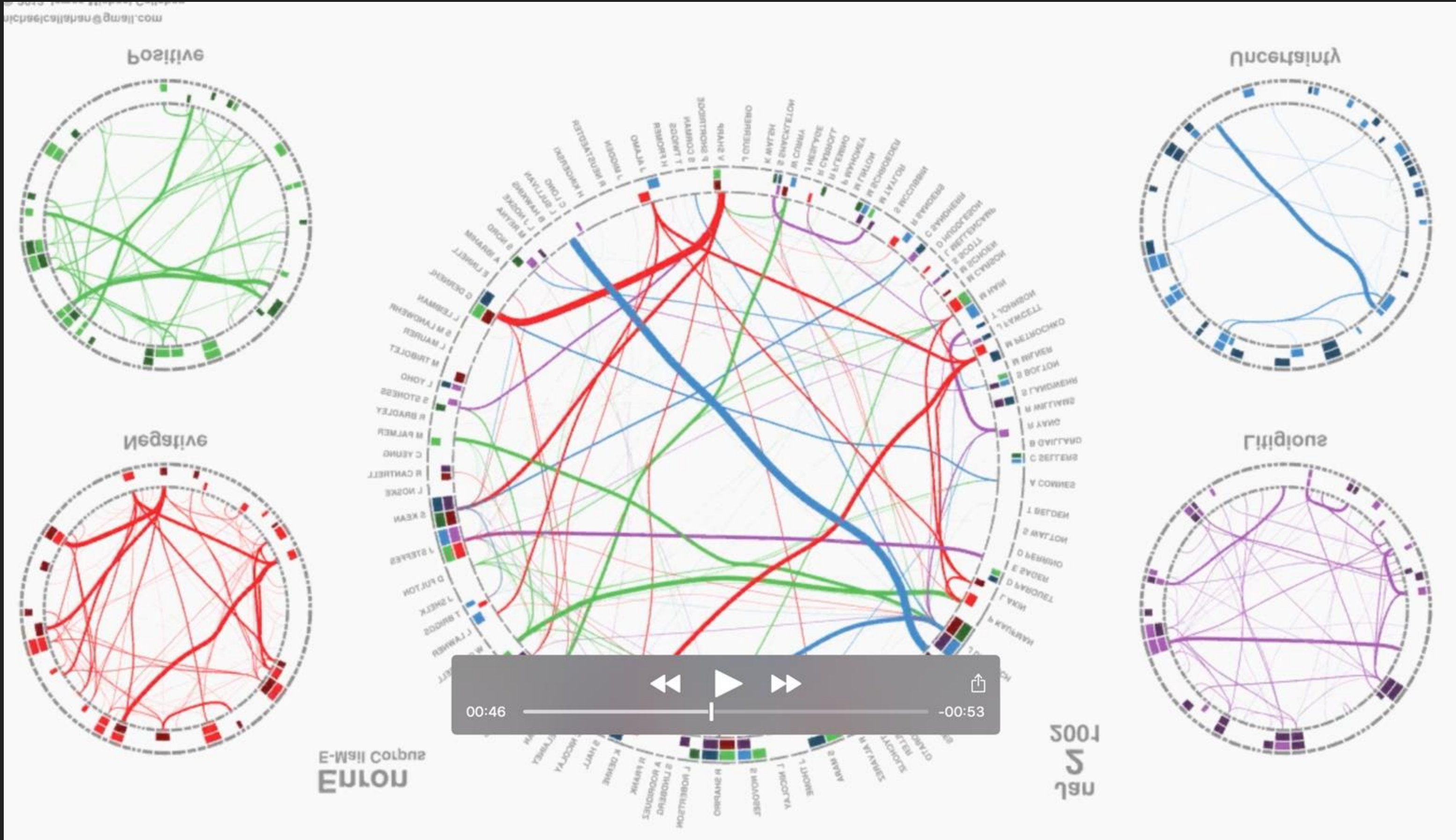


Enron: Topic Analysis



Enron Movie (by Jim Callahan)

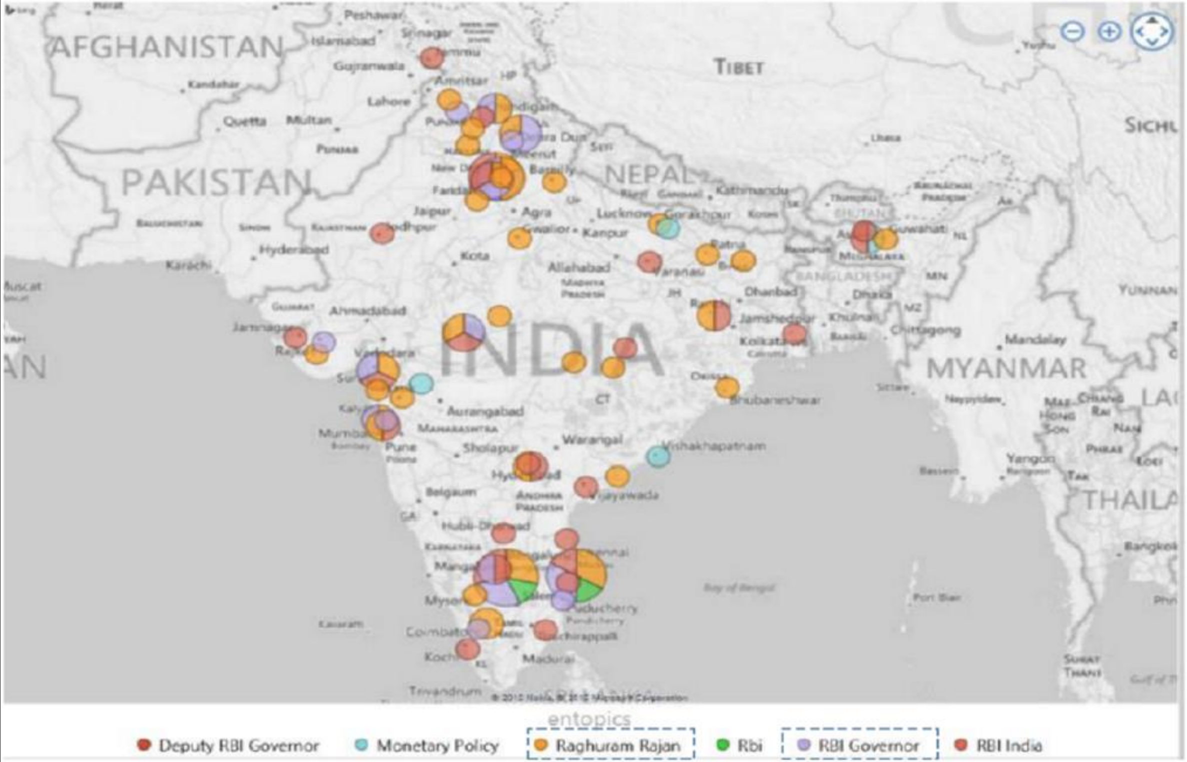
http://srdas.github.io/Presentations/JimCallahan_enron-sm.mov



India: Topic Analysis

Conversations across India and around RBI

topycs



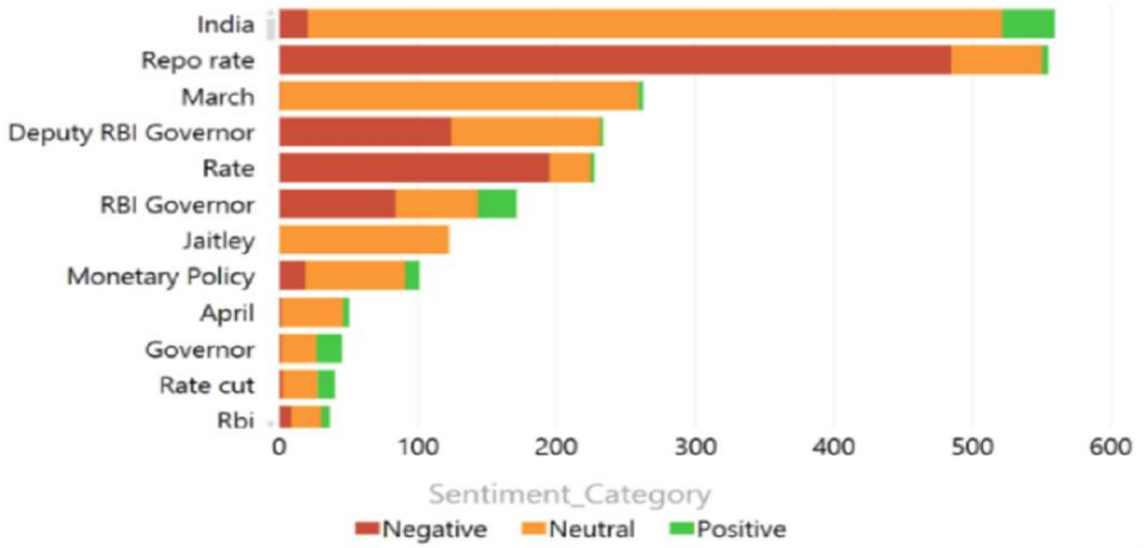
- Conversations across India on RBI, its people and the monetary policy
- Governor features in many conversations across both rural and urban areas
- Some conversations specifically around monetary policy

- Bubbles show split of conversations around Deputy RBI Governor, Monetary Policy, Raghuram Rajan, RBI and RBI Governor.
- Based on count of unique conversations
- Date Range: 1st – 14th April, 2015

India: Topic Analysis

Top Topics along with RBI

topycs



text

"@NDTVProfit: RBI unlikely to change repo rate at policy review smlion

"Digging India's RBI Out of Morass of Debt" by on

"Financial stability is like Pornography. You can't define it but when you see it you know it" - D Subbarao (RBI Governor)

"I was disappointed by the fiscal relaxation." Ex-RBI Governor on India's budget and growth:

"Rajan is perfect, he explains complex economic," PM Modi on RBI governor.

"RBI Conference" shows up as trending topic in India at rank 10

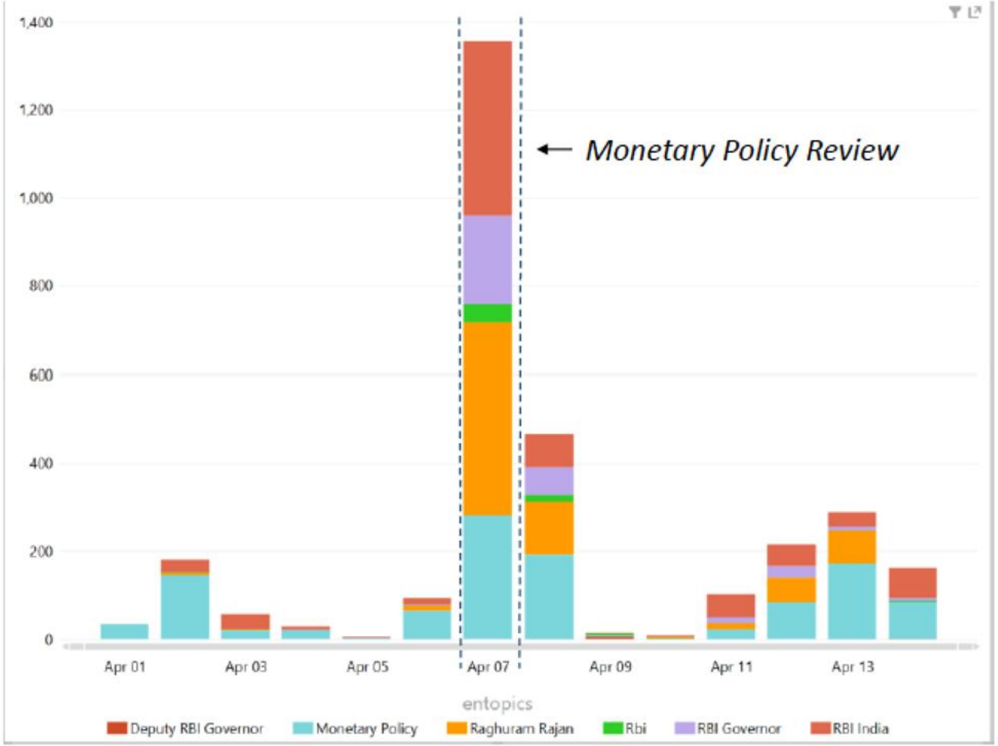
- Repo rate evokes negative sentiment as people don't expect it to be changed
- Repo rate, rate cut and monetary policy are discussed frequently with RBI

- Vertical Axis – Topics of Discussion
- Horizontal Axis – Count of Unique Conversations
- Date – 25th March - 14th of April
- Colors represent sentiment for conversation, Negative – Red, Neutral – Orange, Positive - Green

Monetary Policy (India)

Largest number of conversations around the date of the policy review

topycs



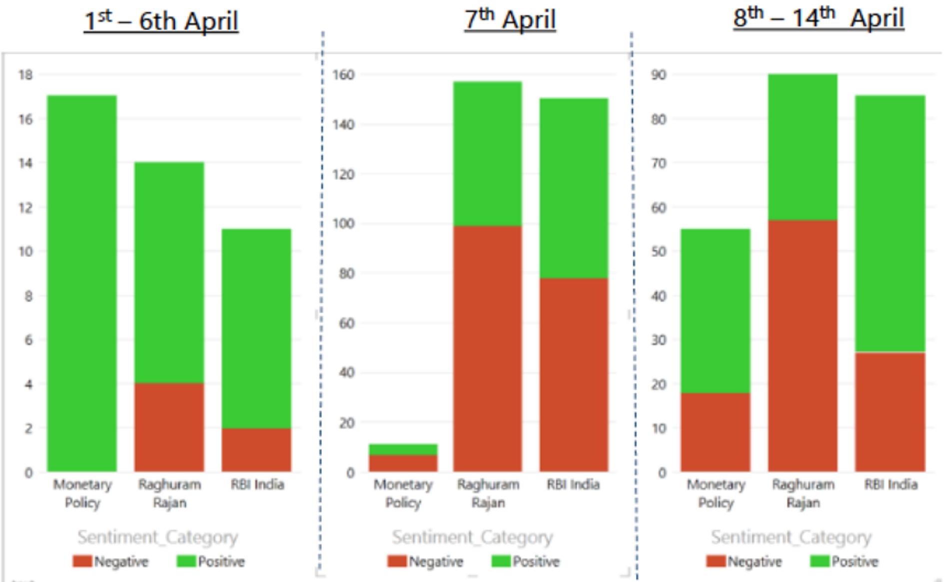
- Largest number of conversations on 7th April and the day after
- People are talking about the monetary policy prior and after the review
- RBI and Governor show up only around the review

- Vertical Axis – Count of Unique Conversations
- Horizontal Axis – Dates between 1st – 14th April
- Colors represent topics of conversation

India

Sentiment around key topics before, during and after the review

topycs



text

"@IndiaToday: SBI, ICICI, HDFC Bank pay heed to Raghuram Rajan's call, cut base rate trending

"@timesofindia: Top 3 banks cut rates after RBIs tough talk" what about you?

"Effectively, monetary policy transmission may move from market forces to fiat which w... - NewsInTweetsIndia

"Effectively, monetary policy transmission may move from market forces to fiat which would be regressive."

"India Inc disappointed with RBI's move on policy rates - SME Times"

"Making monetary policy more potent"

"RBI Conference" shows up as trending topic in India at rank 10

- Conversations positive around monetary policy prior to review, but show more negativity after the review
- More conversations around Raghuram Rajan and RBI India on day of and after the review than prior to the review

- Vertical Axis – Count of Unique Conversations
- Horizontal Axis – Topics of Conversation
- Date – 1st – 14th April
- Colors represent sentiment for conversation, Negative – Red, Positive - Green

Key Principles in Using Big Data for Finance

- Using Theory to develop models to apply to Big Data.
- Questions/problems are primary, data is secondary, in the success of FinTech ventures.
- Simplicity, transparency of models fosters implementability.
- Analytics per se is multidisciplinary.
- Disparate data is the norm.
- Significant investment in hardware and talent.

Thank You!!

http://srdas.github.io/Papers/India.pdf
http://srdas.github.io/Papers/JAI_Das_issue.pdf
http://srdas.github.io/Papers/EnronZeroRev.pdf
http://srdas.github.io/Papers/dyn_syst.pdf