News and Networks: Using Financial News Coverage to Measure Bank Interconnectedness

Sophia Kazinnik, Daniela Scidá, John Wu (FRB Richmond)
Cooper Killen (FRB Chicago)

CEBRA 2023
July 7, 2023

FEDERAL RESERVE BANK OF RICHMOND
Richmond ▪ Baltimore ▪ Charlotte

Motivation

- Studying financial networks is key to understanding:
    - Financial interconnectedness
    - Systemic importance

- Traditional methods to capture financial interconnectedness rely on either:
    - Data that explicitly documents bank connectivity, like inter-bank lending data (e.g., Gofman (2011); Afonso, Kovner, and Schoar (2014))
    - Econometric models that infer connectivity from co-movements in market data (e.g., Billio, Getmanzky, Lo, and Pelizzon (2012); Diebold and Yilmaz (2014); Hardle, Wang, Yu (2016))

Motivation (cont'ed)

- Constructing a real-time interconnectedness measure is methodologically challenging:
  - Data availability
  - Interpretability
- To overcome these challenges, our paper exploits an alternative way of measuring financial interconnections in real-time by studying banks' relationships in the context of financial news
  - Showcases banks' relationships in the view of public discussion
  - Captures soft aspects of financial interconnectedness (e.g., market sentiment, trust, reputation, competitive dynamics, etc) that are not captured by traditional structured data sources
  - Provides a clear narrative to each observed connection
  - Does not limit ex-ante the nature of the links to a specific transaction, relationship, and/or aspect of interconnectedness

This paper

- We propose the use of financial news to study the interconnectedness among the largest U.S.-based financial institutions:
  - Dodd-Frank Act Stress Test (DFAST) participants
  - Both during normal and stress times (i.e., COVID-19 peak of stress)

- We use a"*text-to-network approach*" to construct weekly network matrices based on co-mentioning of banks in news

- Financial connections should be broadly understood as resulting from any financial link (positive or negative) from news that translate into two banks being co-mentioned. For example,
  - *"BofA followed JPMC and WFC in increasing loss reserves and reported a significant drop in Q1 profits"*
  - *"Citi is focusing its corporate banking on tech, healthcare and startup companies, not attempting to compete with JPMC, BofA and GS in the small to mid-sized business market"*

What does it mean that two banks are mentioned in an article?

- Financial news reflects market dynamics and information flows
- The fact that two banks are mentioned together in financial news suggests some level of interdependence or correlation between their activities, even if the correlation is inverse
- The co-occurrence of banks in news articles is therefore indicative of shared economic factors or events that influence their operations
- Co-mention weights help with relevance
- March 2023 banking turmoil offers some good recent examples (e.g., SVB and Signature Bank)

Contribution

- We contribute both empirically and methodologically
- Financial Networks
  - We are the first to study the network among U.S.-based stress tested banks
  - We deliver an interconnectedness approach in real-time with a clear associated narrative
  - We provide powerful visualization tools and showcase how to use the narrative to understand the observed connectivity patterns
- Systemic Risk
  - We introduce a novel systemic risk measure based on negative sentiment eigenvector centrality
  - We rank systemic importance of these financial institutions according to our metric, and study complementarities with traditional systemic risk measures

Some Related Literature

- Interbank lending and balance sheet measures used for capturing bank interconnectedness (Gofman (2011), Schoar et al. (2014), Greenwood, Augustin, and Thesmar (2015)).
- Interconnectedness examined through shared holdings of equities, debt, and liabilities (Elliott, Golub, and Jackson (2014)).
- Systemic risk and financial interconnectedness measured through the contagion channel and shared default risk (Jackson and Pernoud (2021)).
- Market co-movement data used to capture financial interconnectedness (Billio, Getmansky, Lo, and Pelizzon (2012), Diebold and Yilmaz (2014)).
- VaR and other market measures used to determine financial connectivity (Hautsch, Schaumburg, and Schienle (2015), Barigozzi and Brownlees (2019)).
- Text analysis used to convert qualitative information contained in news stories and corporate announcements into quantifiable financial metrics (Boudoukh, Feldman, Kogan, and Richardson (2013), Baker, Bloom, and Davis (2016), Shapiro, Sudhof, and Wilson (2020), Calomiris and Mamaysky (2019)).
- Recent literature on the relationship between news article sentiment and market reaction during the COVID-19 period (Costola, Hinz, Nofer, and Pelizzon (2023), Mamaysky (2020), Baker, Bloom, Davis, Kost, Sammon, and Viratyosin (2020)).

Results preview

- Intuitive patterns of DFAST banks networks based on media narrative
  - Similar types of banks are clustered together (e.g., big 6, trusts, etc)
  - Core-periphery topology with the largest banks at the center
- During periods of stress, we observe:
  - Stronger network ties, consistent with the literature
  - More connections across regional banks and less across IHCs
  - Connections across big players are quite stable, yet stronger
- Text-based eigenvector centrality complements popular systemic risk measures:
  - Provides intuitive and more stable rankings
  - Adds to the information set (e.g., by capturing soft information)
  - The text-narrative can help understand observed changes
- Eigenvector centrality correlates with firm size and business mix (trading in particular) and shows useful in explaining movements in financial variables, such as banks' cumulative abnormal returns

Data: News articles

- We derive our financial interconnectedness measure from financial news articles:
    - Dow Jones Factiva Analytics database
    - All articles on DFAST banks from top financial news sources from 07/01/2019 - 09/30/2020 DFAST Banks Sources
    - Around 70K articles in total and 18K articles with co-mentions (after data cleaning we are left with 49K and 11.4K, respectively)

Methodology: Network analysis

- We construct weekly co-occurrence network matrices: Text2Network
  - Connections are captured by non-zero co-occurrences between every bank-pair
  - Weights are given by co-occurrence values, which measure the importance of each connection
- We use *eigenvector centrality* to determine centrally positioned nodes
  - It weighs both the importance of own (i.e., direct) and neighbors (i.e., indirect) connections → quality besides quantity of connections matters

Co-occurrence across time

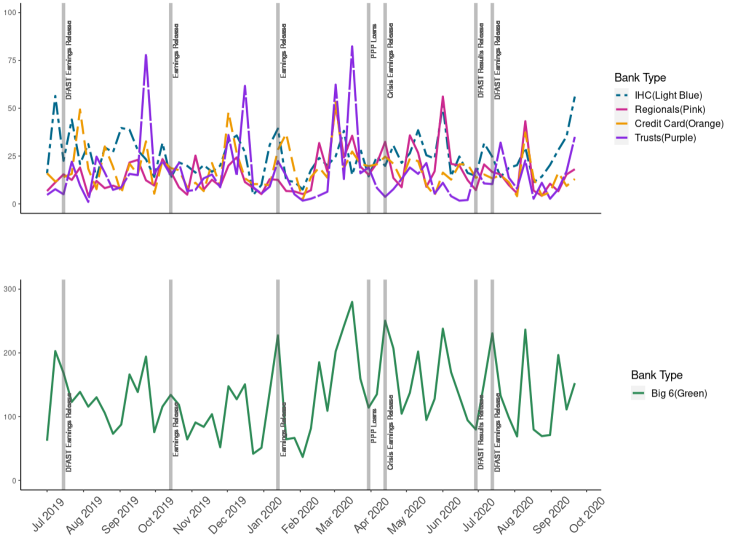
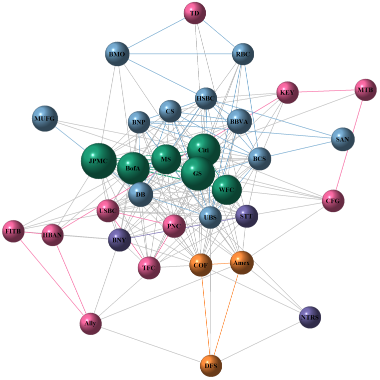
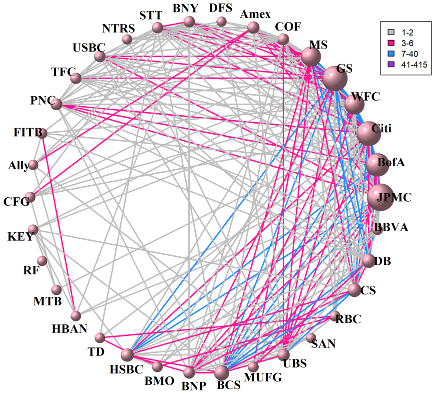


Figure 1: Time series of weekly bank co-occurrences, by bank type

Network topology graphs



Panel A. Connections & clusters

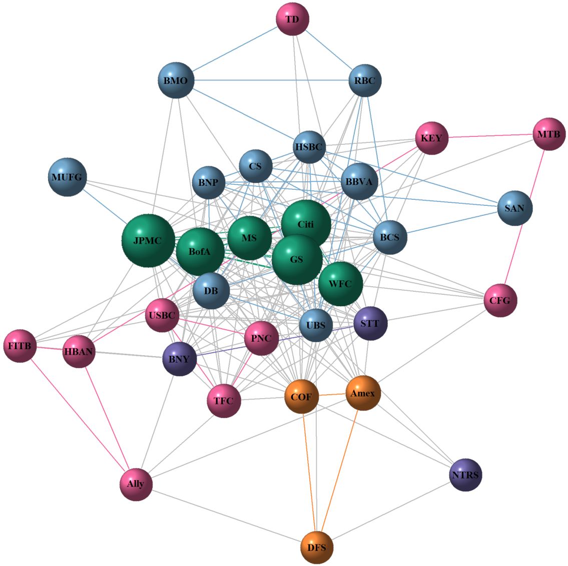


Panel B. Co-occurrences

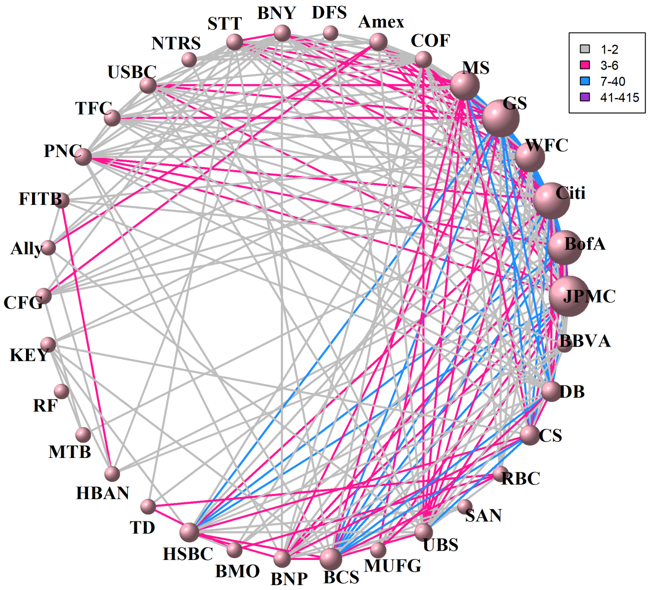
Figure 2: Network Graphs: January 2020 Earnings. Panel A: Nodes are colored by bank type (Big 6 (green), CC (orange), Trusts (purple), Regionals (Pink), IHCs (light blue)) and links color correspond to within bank type connections. Panel B: Nodes are sorted by bank type and link colors correspond to co-occurrence counts (i.e., connections weights).

Heatmaps April Graphs

Network topology graphs (zoom)



Network topology graphs (zoom)



Network topology comparison

Table 1: Summary statistics: January vs April network matrices

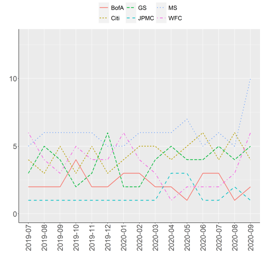
	Connections			Co-occurrences		
Type	**Jan**	**Apr**	%Δ	**Jan**	**Apr**	%Δ
Within *Big 6*	15	15	0%	504	588	16.7%
Between *Big 6* and Non-*Big 6*	103	97	-5.8%	357	327	-8.4%
Within *Regionals*	9	12	33.3%	11	49	345%
Between *Regionals* and Non-*Reg*	58	69	19.0%	103	226	119%
Within *Trusts*	1	0	-100%	3	0	-100%
Between *Trusts* and Non-*Trusts*	34	6	-82.4%	61	11	-82.0%
Within *IHC*	30	12	-60.0%	103	26	-74.8%
Between *IHC* and Non-*IHC*	77	63	-18.2%	266	186	-30.1%
Within *CC*	3	1	-66.7%	6	3	-50%
Between *CC* and Non-*CC*	32	29	9.4%	73	68	-6.8%
Within All Non-*Big 6*	184	120	-34.8%	392	320	-18.4%
Total	210	172	-18.1%	1057	1075	1.7%
Average Path Length	1.51	1.41	-6.6%			
Clustering Coefficient	0.70	0.77	10.0%			

Note: January Earnings is 13 - 19, 2020; April Earnings is 13 - 19, 2020. Connections is the number of links and co-occurrences is the number of co-mentions in articles (weight of connections). Clustering coefficient is calculated as the transitivity or connectivity of a network and average path length is the mean shortest path between two nodes.

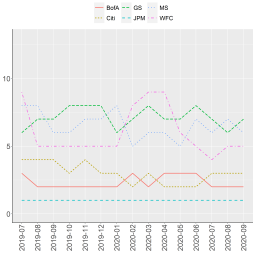
Systemic risk measures: Setup

- Goal: Compare our text-based eigenvector centrality to traditional systemic risk measures. Two main questions:
  - Do they rank systemic importance of financial institutions similarly?
  - Do they capture similar sources of information?
- Text-based eigenvector centrality: Based on news articles that convey negative sentiment
- Comparison measures: SRISK, DIP, CoVaR Defs.
- Data source: Research and Statistic Department, BOG
- Financial institutions: 12 LISCC firms (subset of DFAST banks)
  - U.S. banks: BofA, Citi, JPMC, WFC, GS, MS, BNY, STT
  - IHCs: BCS, CS, DB, UBS (*no longer LISCC as of 2021*)
- Period: October 2019 to September 2020 (1-year), monthly frequency (averages for DIP, SRISK, and CoVaR)

Systemic risk rankings: Traditional measures vs Eigen

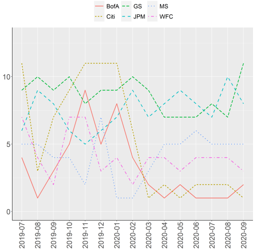


a. Eigenvector Centrality



b. DIP

c. SRISK



d. CoVaR

Figure 3: Ranking of *Big 6* Banks (out of 12 LISCC firms): Eigen vs traditional measures - Monthly frequency

Systemic risk measures: Principal Component Analysis (PCA) - LISCC firms w/ IHCs

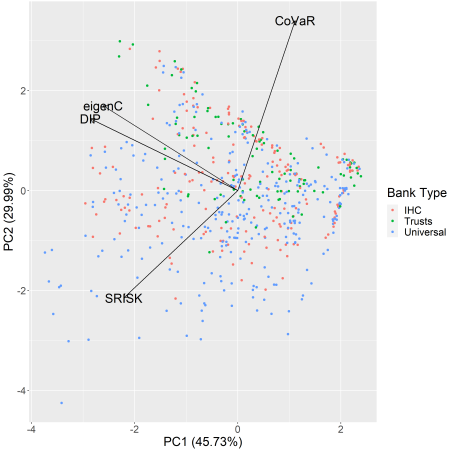
Table 2: PCA loadings & proportion of variance explained

Factor Loadings	PC1	PC2	PC3	PC4
Eigen	-0.57	0.37	-0.47	0.57
DIP	-0.62	0.31	0.15	-0.70
SRISK	-0.48	-0.47	0.65	0.36
CoVaR	0.24	0.74	0.58	0.24

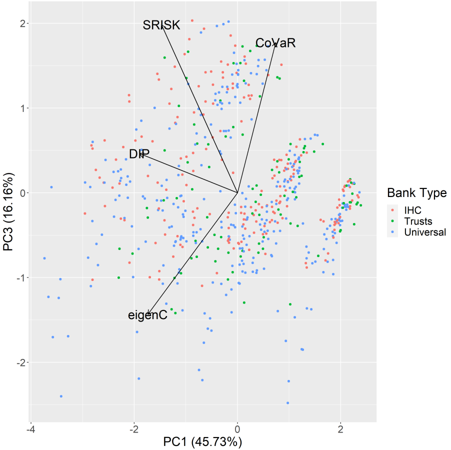
Variance Explained	PC1	PC2	PC3	PC4
Proportion of Variance	0.46	0.30	0.16	0.08
Cumulative Proportion	0.46	0.76	0.92	1.00

Correlations

Systemic risk measures: PCA (cont'd)



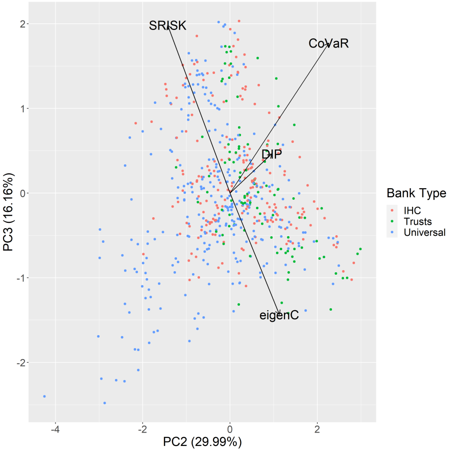
a. PC1 vs PC2



b. PC1 vs PC3

Figure 4: PCA graphs

Systemic risk measures: PCA (cont'd)



c. PC2 vs PC3

Figure 5: PCA graphs

Eigenvector centrality and firms characteristics

Table 3: Eigenvector centrality and firm characteristics

	Dependent variable: Eigen (quarterly)		
	(1)	(2)	(3)
Log(Total Assets)	0.1040***	0.0921***	0.0698***
	(0.0070)	(0.0084)	(0.0141)
Log(Trading Assets)	0.0030**	0.0041***	0.0044***
	(0.0012)	(0.0012)	(0.0012)
ROA	-0.0021*	-0.0022*	-0.0020
	(0.0012)	(0.0012)	(0.0013)
Trading income/TA		11.58***	3.241
		(2.826)	(3.767)
Non-interest income/TA		-1.626**	-1.073
		(0.7160)	(0.6692)
Interest income/TA		1.053**	1.002**
		(0.4576)	(0.4261)
Top 5 trading firms			0.1086**
			(0.0452)
Constant	-1.961***	-1.803***	-1.374***
	(0.1292)	(0.1581)	(0.2691)
Adj. R^2	0.72	0.77	0.79
Num. obs.	164	164	164

Note: $^{*}p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$

Eigenvector centrality and CAR

Table 4: Weekly CAR Quantile Regression (10th Percentile)

	Dependent variable: CAR		
	(1)	(2)	(3)
Eigen	0.029**	0.014**	0.015**
	(0.012)	(0.006)	(0.008)
Asset Growth	−0.001**	−0.0004***	−0.0004**
	(0.0003)	(0.0001)	(0.0002)
Market Cap	0.052***	0.031***	0.026***
	(0.005)	(0.004)	(0.008)
Retained Earnings	0.022	0.005	0.008
	(0.028)	(0.020)	(0.022)
ROA	0.002***	0.0003	0.0002
	(0.0005)	(0.0004)	(0.0004)
Stock Price		0.00005**	0.00004**
		(0.00002)	(0.00002)
Trading Volume ($B)		−0.041***	−0.014***
		(0.003)	(0.005)
VIX			−0.001***
			(0.0002)
Constant	−0.065***	−0.005	−0.002
	(0.003)	(0.004)	(0.004)
Num. Obs.	2,122	2,122	2,122

Note: $^{*}p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$

Eigenvector centrality and CAR (cont'd)

Table 5: Weekly CAR Quantile Regression (90th Percentile)

	Dependent variable: CAR		
	(1)	(2)	(3)
Eigen	−0.011	−0.007	−0.006
	(0.016)	(0.012)	(0.012)
Asset Growth	−0.0004	−0.0004*	−0.0004
	(0.0003)	(0.0002)	(0.0002)
Market Cap	−0.044***	−0.023**	−0.023**
	(0.009)	(0.011)	(0.011)
Retained Earnings	−0.018	−0.001	−0.0003
	(0.029)	(0.031)	(0.031)
ROA	−0.001	−0.001	−0.001
	(0.001)	(0.001)	(0.001)
Stock Price		−0.00002	−0.00002
		(0.00003)	(0.00004)
Trading Volume ($B)		0.028***	0.028***
		(0.004)	(0.007)
VIX			0.00003
			(0.0003)
Constant	0.054***	0.013**	0.013*
	(0.004)	(0.006)	(0.006)
Num. Obs.	2,122	2,122	2,122

Note: $^{*}p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$

Robustness checks

- Monthly vs weekly eigenvector centrality
- Co-occurrence using select publications: Reuters
- Systemic risk and CAR analysis using Eigenvector centrality based on all articles (regardless of sentiment)
- Manual classification of articles of our two key weeks (January and April 2020):
    - Assess accuracy of co-occurrence
    - Further investigate narrative behind connections
    - In particular, better understand drivers of new connections (or differences) during stress

Conclusion

- Financial news is a rich source of data: No longer just serving to daily inform ourselves
- Lots of information and research goes behind financial news articles
- Hard to know everything discussed in the news on all institutions:
    - In the absence of stress, when not paying attention
    - During stress, when the amount of news is just too large to the human mind to summarize
- Real time news-based networks can help regulators:
    - Summarize the giant amount of information in the news
    - Assess potential contagion channels in real time and systemic importance
    - Learn and investigate links of interest (e.g., unexpected links) in normal times

Thank You!

Appendix

DFAST banks list

Table 6: List of DFAST Bank Holding Companies (BHC)

Bank Type	N	Bank Name	Symbol
Big 6	1	Bank of America	BofA
	2	Citigroup	Citi
	3	Goldman Sachs	GS
	4	JPMorgan Chase	JPMC
	5	Morgan Stanley	MS
	6	Wells Fargo	WFC
Trusts	7	BNY Mellon	BNY
	8	Northern Trust	NTRS
	9	State Street Corp	STT
Credit Card	10	American Express	Amex
	11	Capital One	COF
	12	Discover Financial	DFS

Bank Type	N	Bank Name	Symbol
Regionals	13	Ally Financial	Ally
	14	Citizens Financial Group	CFG
	15	Fifth Third Bank	FITB
	16	Huntington Bank	HBAN
	17	KeyCorp	KEY
	18	M&T Bank	MTB
	19	PNC Group	PNC
	20	Regions Financial	RF
	21	Truist	TFC
	22	US Bancorp	USBC
IHC	23	BBVA Compass	BBVA
	24	Bank of Montreal	BMO
	25	BNP Paribas	BNP
	26	Barclays Bank	BCS
	27	Credit Suisse	CS
	28	Deutsche Bank	DB
	29	HSBC Bank	HSBC
	30	MUFG Union	MUFG
	31	Royal Bank of Canada	RBC
	32	Santander Bank	SAN
	33	TD Bank	TD
	34	UBS Group	UBS

Back

News source list

Table 7: List of news source groups from Factiva Analytics

Code	Name	Notable Examples
TDJW	Dow Jones Newswire	Dow Jones Institutions News
TMNB	Major News and Business Sources	CNN, NY Times, Charlotte Observer
TPRW	Press Release Wires	Business Wires, Nasdaq/Globenewswire
TRTW	Reuters Newswires	Reuters News
SFWSJ	Wall Street Journal Sources	The Wall Street Journal
IBNK	Banking/Credit Sources	American Banker, Financial Times
IFINAL	Financial Services Sources	The Economist, MarketWatch

Back

Methodology: From text to network

- Look at the co-occurrences of entity names in a given news article
- Example: Assume we have the following documents (i.e., news article) in our corpus:
  - Doc 1: Acme Corp banks with both WFC and BofA.
  - Doc 2: The headquarter of WFC is in SF, and BofA's is in Charlotte.
  - Doc 3: In Q3, WFC was fined $1.5B for its dealings with JPMC. WFC plans to appeal.

	WFC	BofA	JPMC
Doc 1	1	1	0
Doc 2	1	1	0
Doc 3	2	0	1

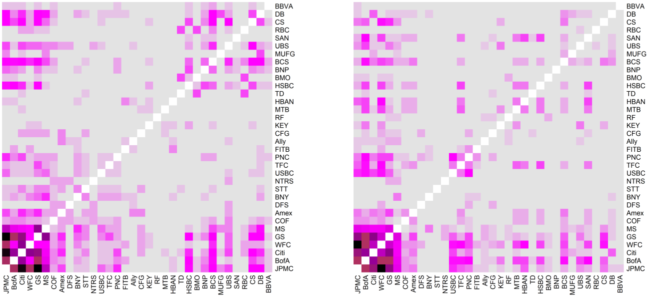
Table 8: Raw term-document matrix: M

	WFC	BofA	JPMC
WFC	3	2	1
BofA	2	2	0
JPMC	1	0	1

Table 9: Co-occurrence matrix: $C = M^T \times M$

Back

Heatmaps

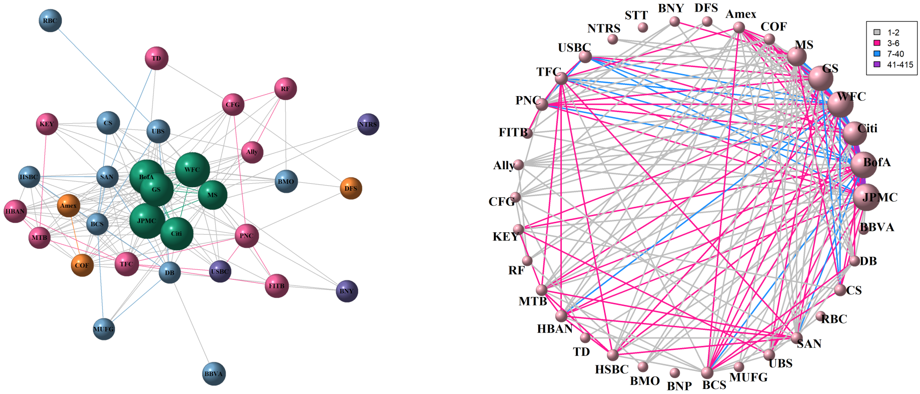


Panel A. January 2020 Earnings   Panel B. April 2020 Earnings

Figure 6: Heatmaps: Pre-crisis vs crisis periods

Back

Network topology graphs



Panel A. Connections & clusters

Panel B. Co-occurrences

Figure 7: Network Graphs: April 2020 Earnings. Panel A: Nodes are colored by bank type (Big 6 (green), CC (orange), Trusts (purple), Regionals (Pink), IHCs (light blue)) and links color correspond to within bank type connections. Panel B: Nodes are sorted by bank type and link colors correspond to co-occurrence counts (i.e., connections weights).

Back

Systemic risk measures: Brief explanation

- Eigenvector Centrality
  - Measures firm's importance based on network connections with associated negative sentiment
  - Financial news text based; captures traditional financial data and soft information
- DIP (Distress Insurance Premium)
  - Measures the expected credit loss that equal or exceed a minimum share of the sector's total liabilities
  - Based on bank size, default probability (from CDS spreads), and asset return correlations
- SRISK
  - Measures a banks' systemic vulnerability as expected capital shortfall conditional on a large market downturn
  - E(CS) is based on required capital given a bank's assets minus a bank's market equity
- CoVaR
  - Measures the spillovers to the whole financial network based on one distressed bank
  - Stock return-based measure

Back

Systemic risk measures: Correlations

Table 10: Correlation matrix based on the 12 LISCC banks as of 2020

	Eigen	DIP	SRISK	CoVaR
Eigen	1.00	0.61	0.16	-0.06
DIP	0.61	1.00	0.34	0.01
SRISK	0.16	0.34	1.00	-0.35
CoVaR	-0.06	0.01	-0.35	1.00

Back