News and Networks: Using Text Analytics to Assess Bank Networks During COVID-19 Crisis

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

2022 RiskLab/BoF/ESRB Conference on Systemic Risk Analytics
5-6 May, 2022
Helsinki, Finland

FEDERAL RESERVE BANK OF RICHMOND
Richmond ▪ Baltimore ▪ Charlotte

Motivation

- Studying financial networks is key to understanding:
  - Financial interconnectedness
  - Systemic importance
- Traditionally, bank interdependencies are captured via:
  - Interbank lending data (e.g., Gofman (2011); Afonso, Kovner, and Schoar (2014))
  - Co-movements in market data (e.g., Billio, Getmanzky, Lo, and Pelizzon (2012); Diebold and Yilmaz (2014); Hardle, Wang, Yu (2016))
- Alternatively, one can use text to construct networks: Banks' relationships in the view of public discussion (here, financial news)

This paper

- We study the interconnectedness of large U.S. financial institutions that fall under the Dodd-Frank Act Stress Test (DFAST) umbrella during the events surrounding the stress period related to the COVID-19 pandemic in 2020
- Build upon Rönnqvist and Sarlin (2015, Quantitative Finance) "*text-to-network approach*" and 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

Contribution

- We are the first to study the network among US-based stress tested banks
- We study the network dynamics during time of stress and shed light on the impact of COVID-19 events on the network topology
- We propose using the eigenvector centrality of nodes to rank systemic importance of these financial institutions, and compare it to rankings based on traditional systemic risk measures

Results preview

- Intuitive patterns of DFAST banks networks based on media narrative
  - Similar types of banks are clustered together (e.g., big 6, trusts, credit cards, IHCs)
  - Core-periphery topology (i.e., largest banks clustered together at the center and IHCs at the periphery)
- During periods of stress, we observe:
  - Denser networks, consistent with the literature
  - More connections across different bank groups (i.e., cross-cluster connectivity increases)
  - Connections across big players are quite stable, while connections at the periphery increase
- Text-based eigenvector centrality could serve as a complement to existing traditional systemic risk measures (e.g., by capturing soft information)

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 (18K articles with co-mentions)
- We divide our sample into three parts:
  - Pre-pandemic period (July 2019 through February 2020)
  - High stress period (March through April 2020)
  - Period of a "new normal" (May through September 2020)

Methodology: Network analysis

- We construct weekly co-occurrence network matrices for our sample period:
  - 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

Text2Network

- 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

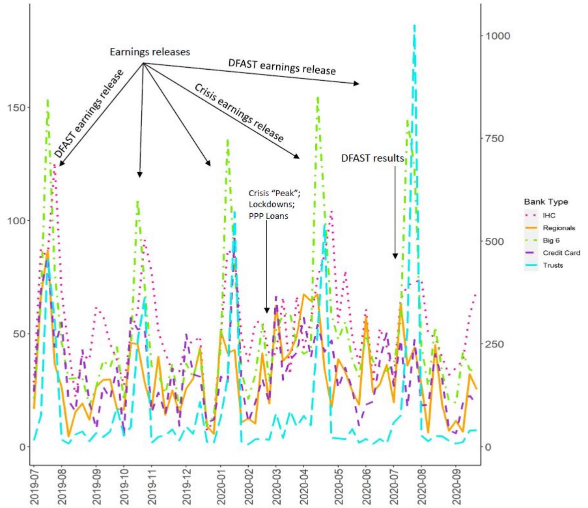
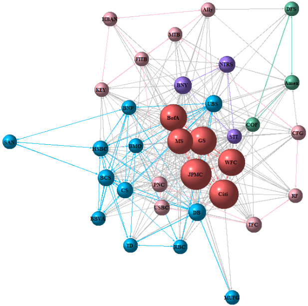
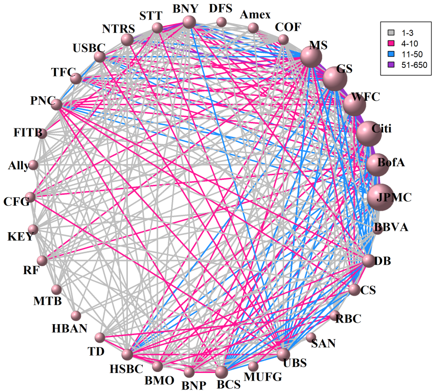


Figure 1: Time series of bank co-occurrences, by bank type
(Big 6 on the right axis)

Network topology graphs



Panel A. Connections & clusters



Panel B. Co-occurrences

Figure 2: Network Graphs: January 2020 Earnings

Heatmaps April Graphs CDF

Network topology comparison

Table 1: Summary statistics: January vs April network matrices

Type	Connections Jan	Connections Apr	Connections %Δ	Co-occurrences Jan	Co-occurrences Apr	Co-occurrences %Δ
Within *Big 6*	12	12	0%	3432	3788	10.3%
Between *Big 6* and Non-*Big 6*	131	141	7.6 %	1069	1218	26.2 %
Within *Regionals*	14	22	57.0%	31	74	138.7 %
Between *Regionals* and Non-*Reg*	98	142	44.9 %	352	526	49.4 %
Within *Trusts*	3	3	0%	67	134	100 %
Between *Trusts* and Non-*Trusts*	58	43	25.8%	364	567	55.8 %
Within *IHC*	38	34	-10.5%	181	104	-42.5 %
Between *IHC* and Non-*IHC*	111	135	21.6 %	684	689	0.7%
Within *CC*	3	3	0%	6	10	66.7 %
Between *CC* and Non-*CC*	32	54	68.8%	73	142	94.5%
Within All Non-*Big 6*	284	358	26.1%	974	1218	25.1%
Total	576	668	16.0%	6544	7696	17.6%

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
- 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: Same as our sample, weekly frequency

Systemic risk rankings: Traditional measures vs EigenC

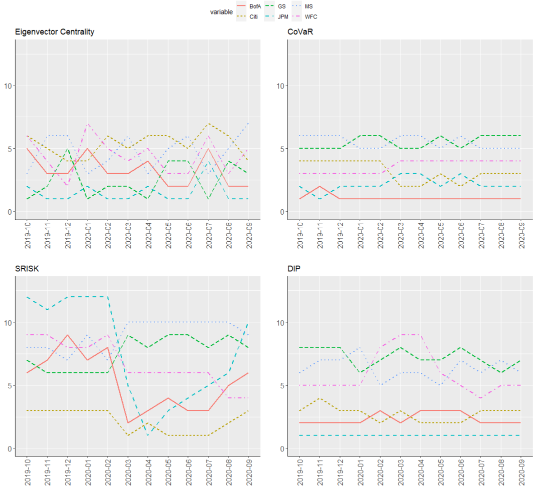


Figure 3: Ranking of *Big 6* Banks (out of 12 LISCC firms): Eigenvector centrality vs traditional measures (SRISK, CoVaR and DIP) - Monthly frequency

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

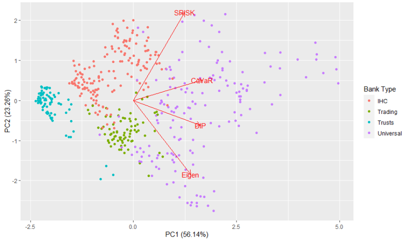
Table 2: PCA loadings & proportion of variance explained

Factor loadings	PC1	PC2	PC3	PC4
Eigenvector centrality	0.42	-0.71	0.56	-0.08
DIP	0.50	-0.34	-0.78	0.17
SRISK	0.53	0.45	0.28	0.66
CoVaR	0.55	0.41	0.01	-0.73

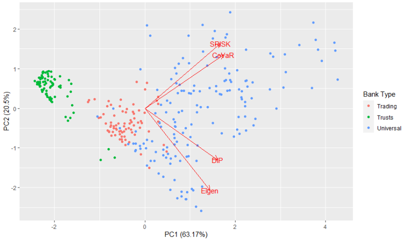
Variance explained	PC1	PC2	PC3	PC4
Proportion of variance	0.61	0.21	0.12	0.06
Cumulative proportion	0.61	0.82	0.94	1.00

PCA w/IHCs Rank Correlations

Systemic risk measures: PCA (cont'd)



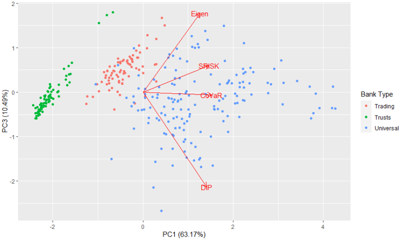
PCA1-PCA2 (with IHCs)



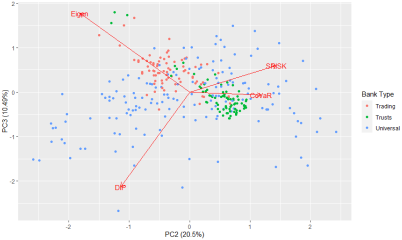
PCA1-PCA2 (without IHCs)

Figure 4: PCA graphs

Systemic risk measures: PCA (cont'd)



PCA1-PCA3 (without IHCs)



PCA2-PCA3 (without IHCs)

Figure 5: PCA graphs

Robustness checks

- Monthly vs weekly eigenvector centrality
- Co-occurrence using select publications: Reuters
- Including IHCs in systemic risk analysis
- 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

Conclusions

- We investigate the interconnectedness of DFAST bank holding companies by analyzing how they are mentioned together in financial news articles in the context of the COVID-19 induced financial crisis
- Text-based networks provide a real time alternative to traditional network approaches with more traceable connections
  - Observed patterns seem intuitive
  - Text narrative can be leveraged to help better understand the observed connections and changes in patterns
  - Network and systemic risk measure can be updated on a frequent basis
  - Allows to study both cross-section and time variation
  - Only public data is needed
- Our PCA analysis suggests that text-based eigenvector centrality offers a complementary measure to existing traditional systemic risk measures

Next steps

- Refine co-occurrence measure by further exploiting the text:
  - Add sentiment
  - Topic analysis of the network connections
- Refine the data pull by removing noisy articles (e.g., articles consisting of mostly tables) or not "news" related (e.g., SEC filings)
- Application of eigenvector centrality to financial data

Thank You!

Appendix

DFAST banks list

Table 3: List of DFAST Bank Holding Companies (BHC)

Bank Type	Bank Name	Symbol
Big 6	Bank of America	BofA
	Citigroup	Citi
	Goldman Sachs	GS
	JPMorgan Chase	JPMC
	Morgan Stanley	MS
	Wells Fargo	WFC
Trusts	BNY Mellon	BNY
	Northern Trust	NTRS
	State Street Corp	STT
Credit Card	American Express	Amex
	Capital One	COF
	Discover Financial	DFS

Bank Type	Bank Name	Symbol
Regionals	Ally Financial	Ally
	Fifth Third Bank	FITB
	Huntington Bank	HBAN
	KeyCorp	KEY
	M&T Bank	MTB
	PNC Group	PNC
	Regions Financial	RF
	Truist	TFC
	US Bancorp	USBC
IHC	BBVA Compass	BBVA
	Bank of Montreal	BMO
	BNP Paribas	BNP
	Barclays Bank	BCS
	Credit Suisse	CS
	Deutsche Bank	DB
	HSBC Bank	HSBC
	MUFG Union	MUFG
	Santander Bank	SAN
	TD Bank	TD
	UBS Group	UBS

Back

News source list

Table 4: 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 BoA.
  - Doc 2: The headquarter of WFC is in SF, and BAC's is in Charlotte.
  - Doc 3: In Q3, WFC was fined \$1.5B for its dealings with JPMC. WFC plans to appeal.

	WFC	BoA	BAC	JPMC
Doc 1	1	1	0	0
Doc 2	1	0	1	0
Doc 3	2	0	0	1

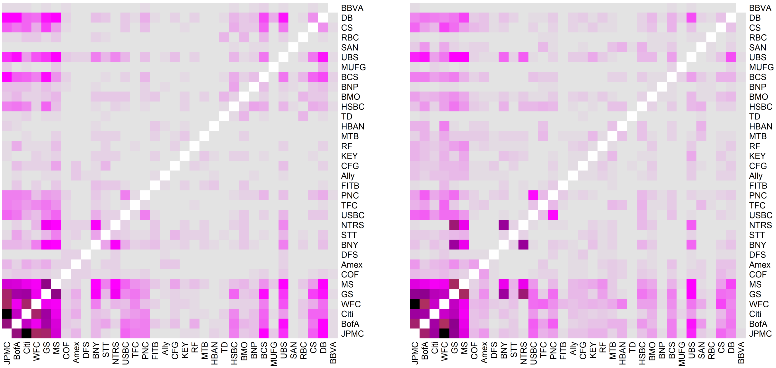
Table 5: Raw term-document matrix: M

	WFC	BAC	JPMC
WFC	3	2	1
BAC	2	2	0
JPMC	1	0	1

Table 6: Co-occurrence matrix: $C = M^T \times M$

Back

Heatmaps

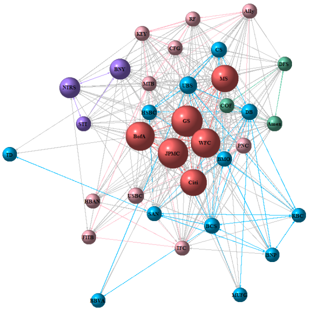


Panel A. January 2020 Earnings

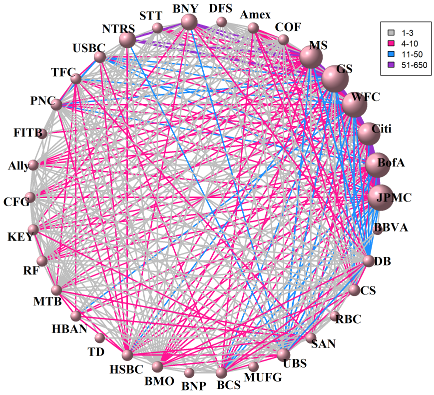
Panel B. April 2020 Earnings

Figure 6: Heatmaps: Pre-crisis vs crisis periods

Network topology graphs



Panel A. Connections & clusters



Panel B. Co-occurrences

Figure 7: Network graphs: April 2020 earnings

Back

CDF: Eigenvector centrality (January earnings week)

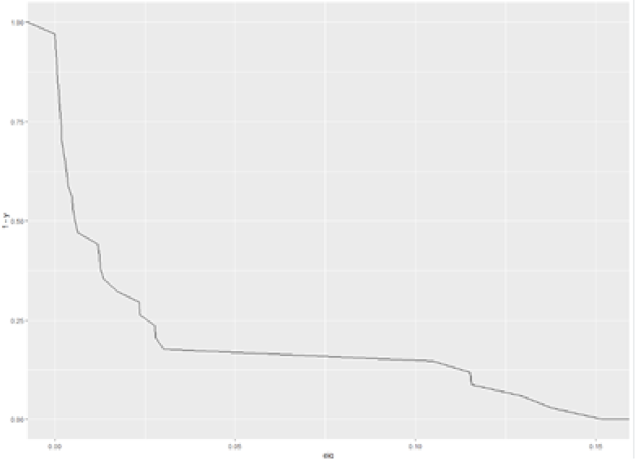


Figure 8: Eigenvector centrality CDF. “January earnings” is defined as the week of January 13, 2020

Back

Systemic risk measures: Brief explanation

- Eigenvector Centrality
    - Measures firm's importance based on network connections
    - 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 rankings: Rank correlations

Table 7: Rank Correlations: January vs April network matrices

January	DIP	SRISK	CoVAR	EIGEN
DIP	1	.50	.84	.21
SRISK	.50	1	.19	.11
COVAR	.84	.19	1	.39
EIGEN	.21	.11	.39	1

April	DIP	SRISK	CoVAR	EIGEN
DIP	1	.79	.66	.24
SRISK	.79	1	.64	.12
COVAR	.64	.66	1	.51
EIGEN	.24	.12	.51	1

Back

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

Table 8: PCA loadings & proportion of variance explained

Factor loadings	PC1	PC2	PC3	PC4
Eigenvector centrality	0.44	0.66	0.42	-0.42
DIP	0.54	0.19	-0.81	0.05
SRISK	0.43	-0.70	0.09	-0.55
CoVaR	0.56	-0.17	0.39	0.71

Variance explained	PC1	PC2	PC3	PC4
Proportion of variance	0.55	0.23	0.12	0.10
Cumulative proportion	0.55	0.78	0.90	1.00

Back