

Financial Knowledge Graph

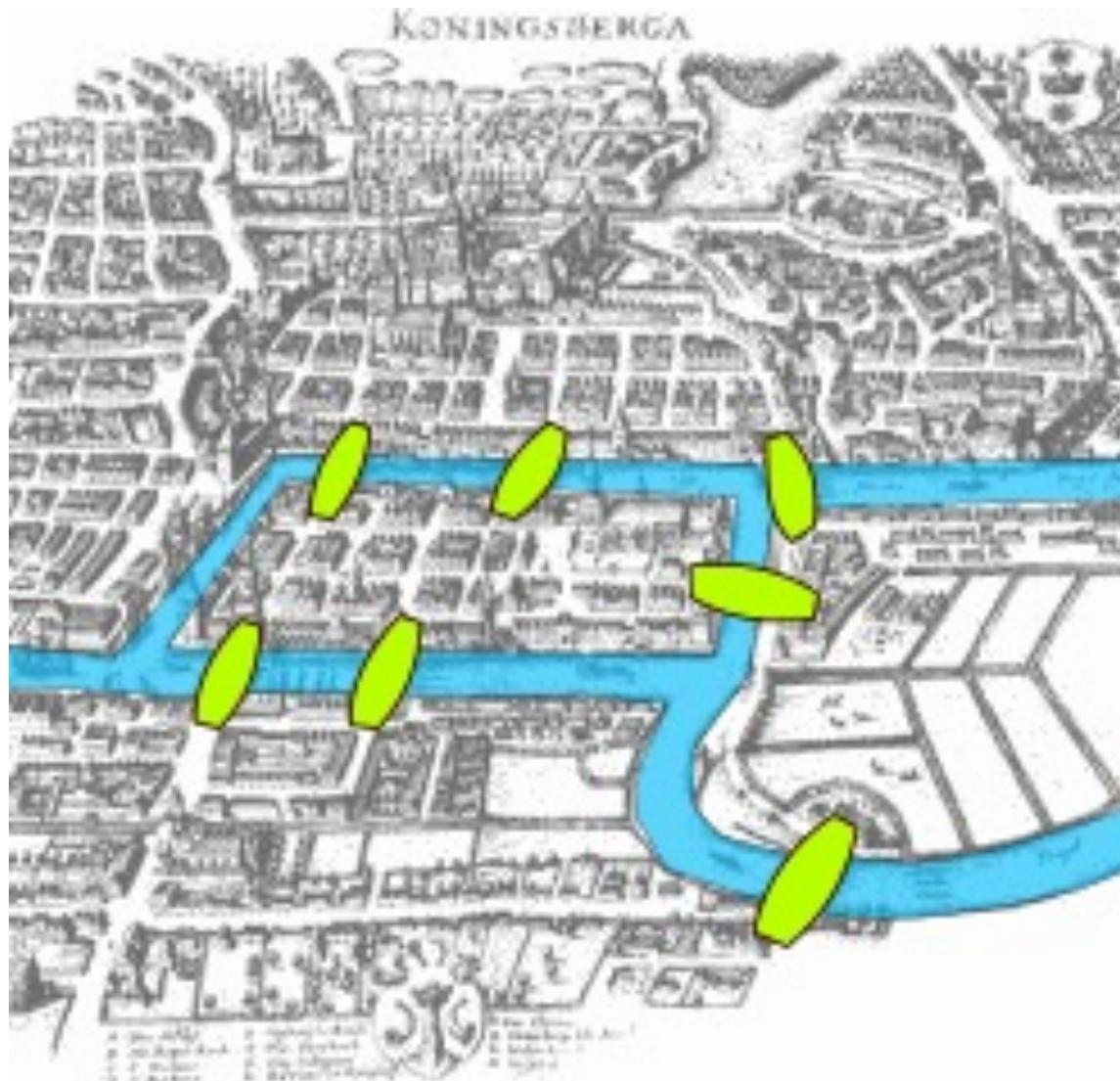
A use case for systemic risk

Agenda

- Introduction (page 2)
- Graph theory and Knowledge graph Brief introduction (Page 4-6)
- Financial Knowledge Graph with Volatility and Liquidity (Page 7-9)
- Financial Knowledge Graph with NLP(Page 10-17)
- Other Applications (Page 18)

What is Graph?

- Origin: Seven Bridges:
 - Leonhard Euler
 - 1736, in Königsberg, Prussia (modern-day Kaliningrad, Russia.)
 - Question: Can we take a walk through the city that would cross each of the seven bridges only once?



Graph Theory

A well-developed theory with the following **terminology**:

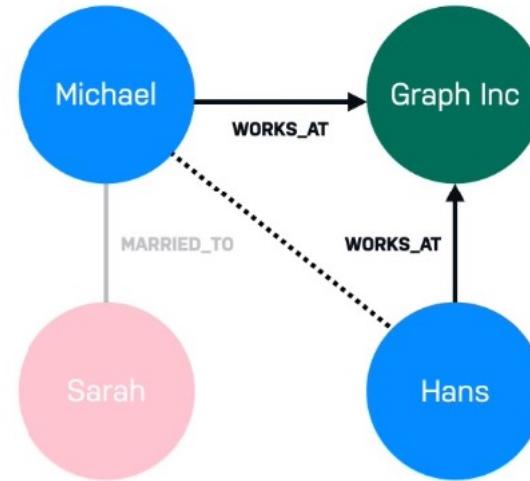
Nodes (Vertices)

- Representation of an entity (e.g., person, product, etc.)
- Degree: number of connections of a node.

Edges (relationships): connections hierarchy etc.

Label: the coloring of the node

- Types: Directed Graph vs Undirected Graph
- Weighted vs. Unweighted
- DAG: Directed Acyclic Graph



Types of graphs

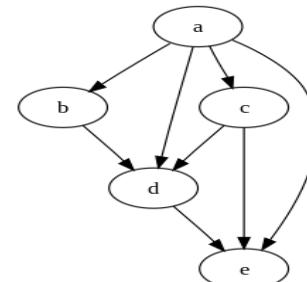
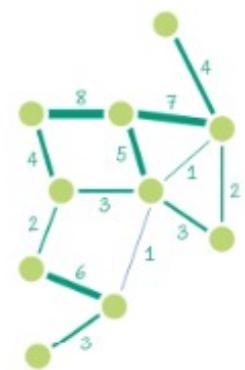
undirected

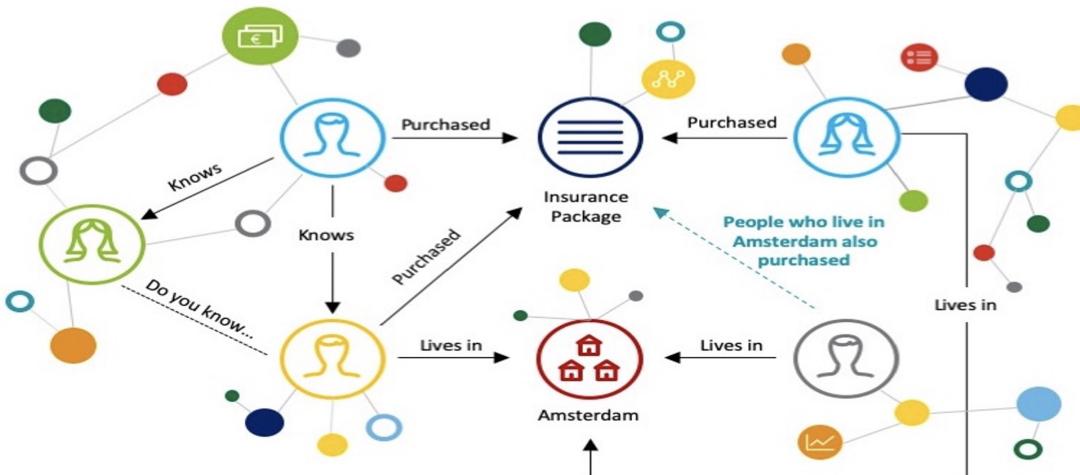


directed



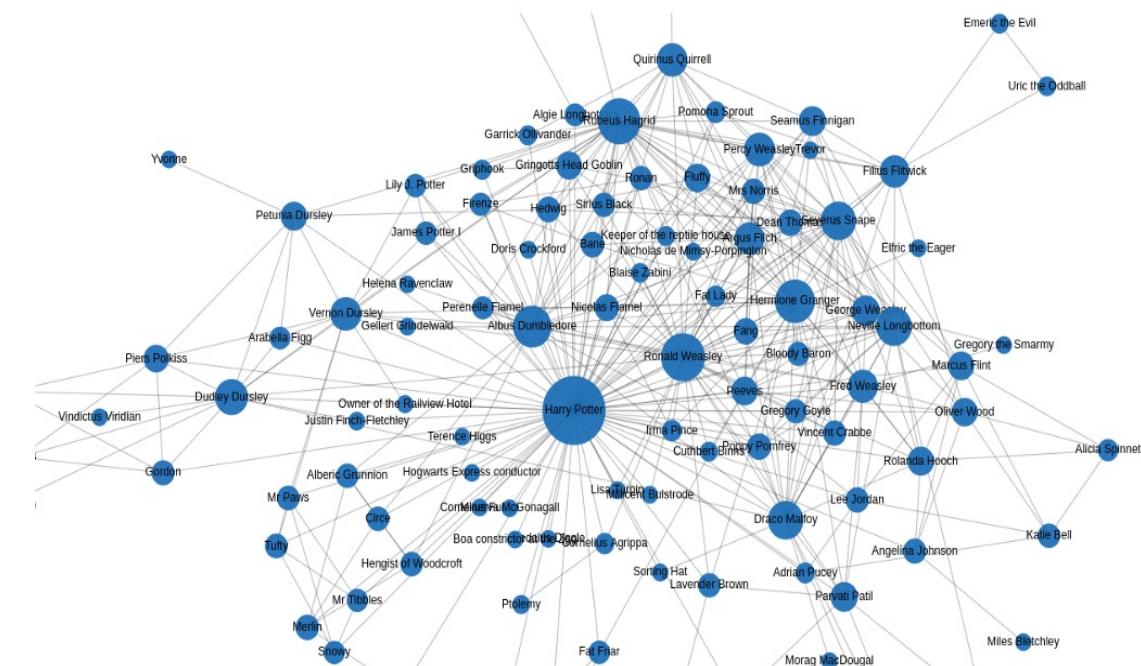
weighted





Knowledge Graphs provide context-driven recommendations.

Fig 1 Deloitte-Financial Knowledge Graph



Harry Potter KG(source: [neo4j](#))

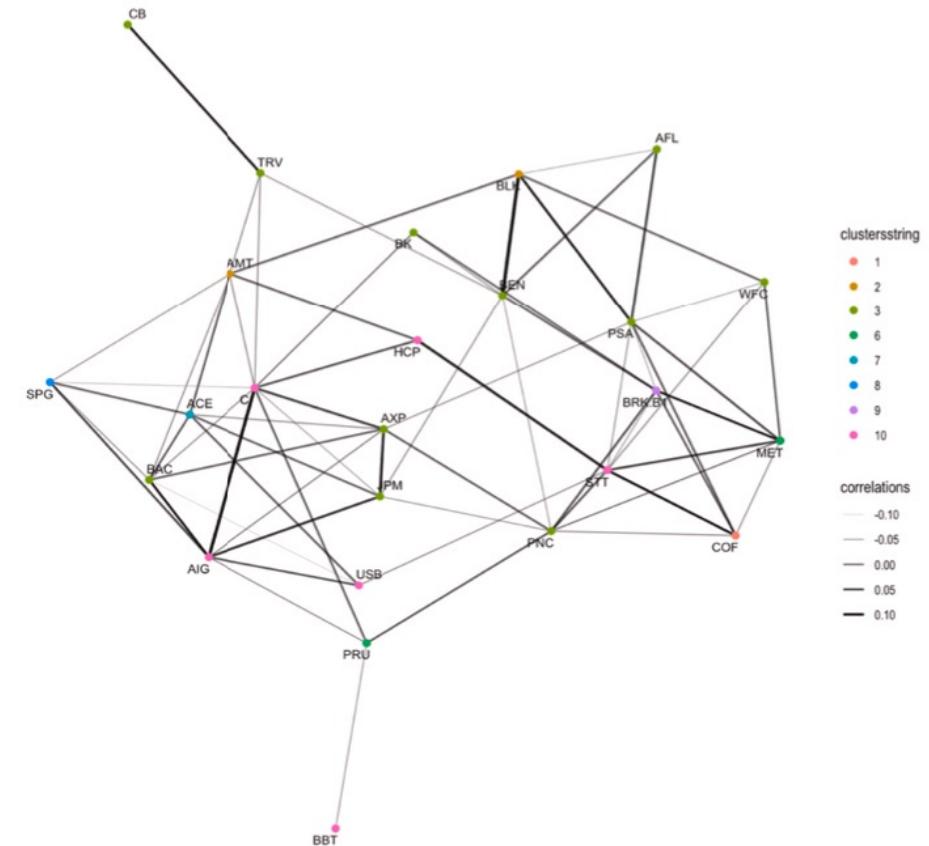
Knowledge Graph(KG)

- A structured representation of **knowledge** with graph form including **nodes** (entities), **edges** (relationships between entities), and **properties** (attributes of nodes) that can be visualized
- Application: widely used in NLP space, in the NLP space, in translational embedding model space ([TransE](#), [DistMult](#)), to build graphical database and recommendation system(Fig 1)
- Popular Knowledge Graph(Harry Potter characters)
 - ❖ Harry Potter 4 series Books are the knowledge source; the main character is the **node**
 - ❖ **Edge:** Frequency with which Hermione Granger and Ron Weasley jointly mentioned
 - ❖ Harry Potter is the most connected node and hence the most important character
- Broadly speaking, a graph with “**knowledge**”
- “**Knowledge**” refers to the connection/Edge(**financial market variable based metrics**)

Financial KG for Systemic Risk

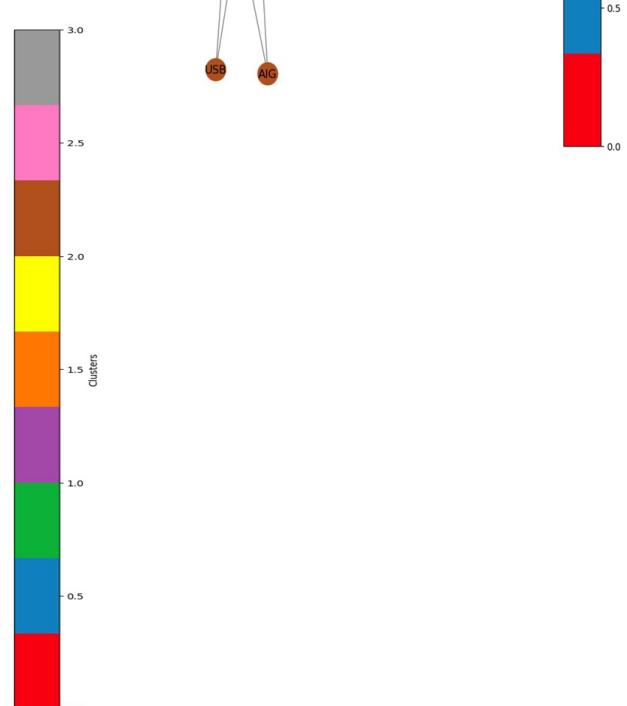
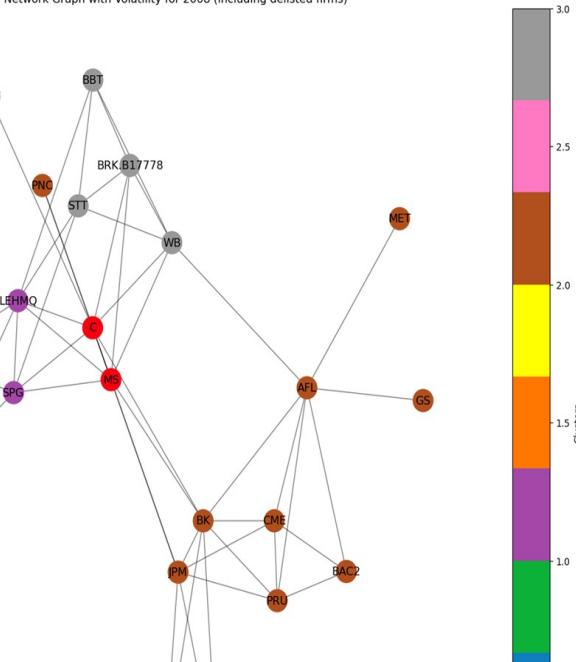
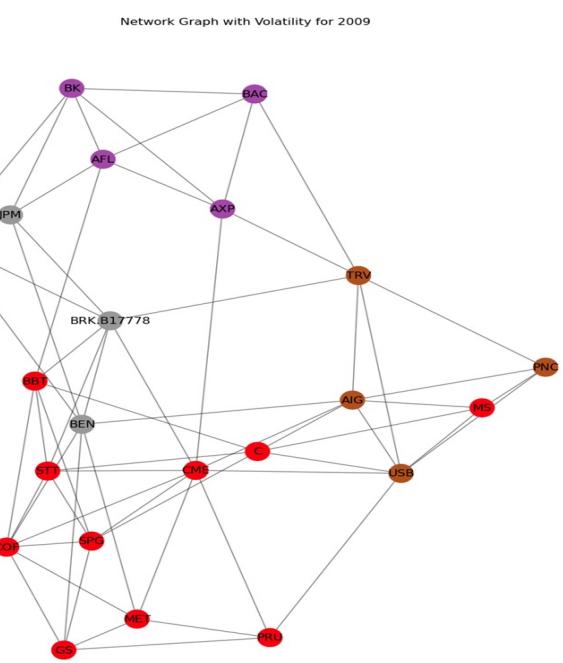
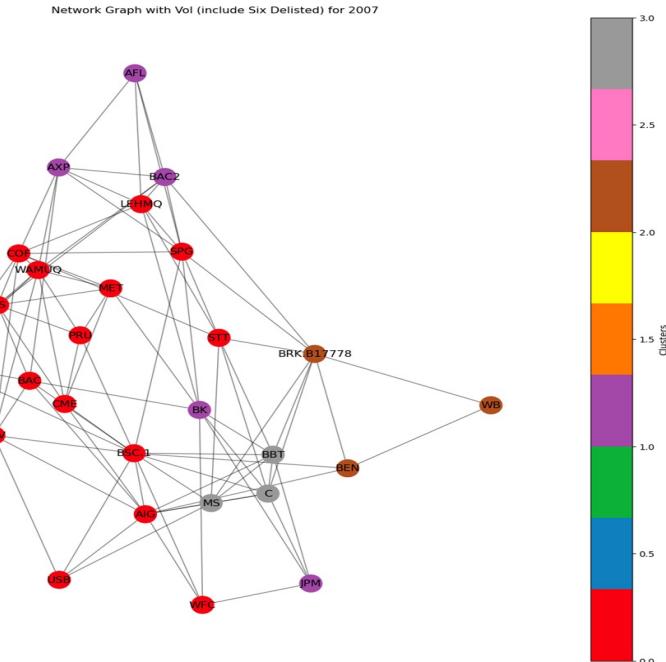
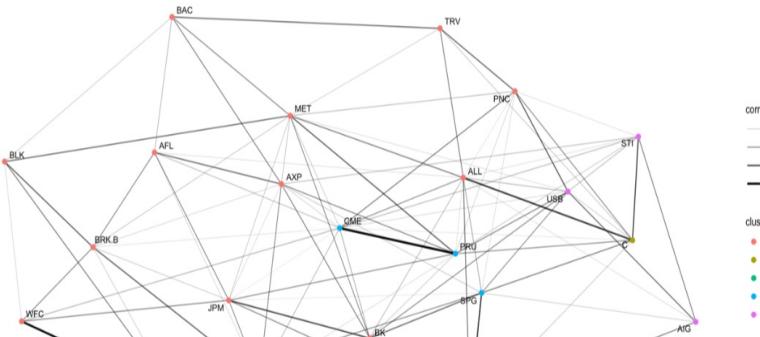
- Source Data: ~200 Financial Institutions with liquidity and equity metrics (Monthly, 2006-2013)
- Node: Top 25 Financial Firms selected based on Market Cap
- Metrics: market implied Volatility of the financial firm and liquidity discount
- Edge: *Partial Correlation (Appendix A)* or Conditional Dependency of the Implied volatility and liquidity Discount
- Threshold for partial correlation applied to trim highly correlated pairs
- Thickness of the Edge: strength of the correlation
- Label: Clustering analysis of 25 financial firms with k-nearest neighbors (KNN) based on market cap

(a) Full Sample (2006-2012), Top 25 with 0.03 Cutoff



Financial Knowledge Graph (Volatility in Financial Crisis 2007-2009)

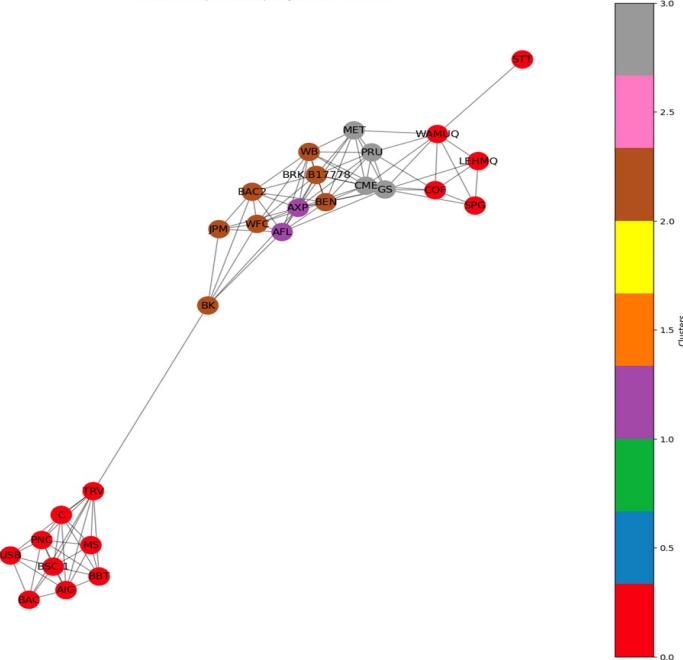
- Spring Network Graph of market-implied volatility
- CME Group as the clearing house sits in the center of the graph



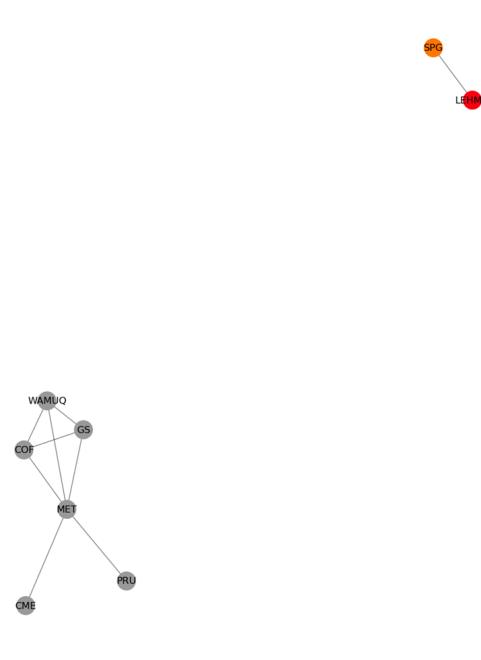
Financial Knowledge Graph with Liquidity Metrics (for illustration purpose)

- Liquidity Discount: measurement of compression for the Asset price driven by lack of liquidity during Financial Crisis ([Appendix B](#))
- Empirical Observation: Liquidity is a better measurement as it can reflect the hidden connections. Volatility is more connected due to the nature of crisis.
- Observations:
 - Before the crisis, retail-heavy banks had similar liquidity discounts
 - The partial correlation of liquidity is extremely low during 2008, except for Lehman and SPG

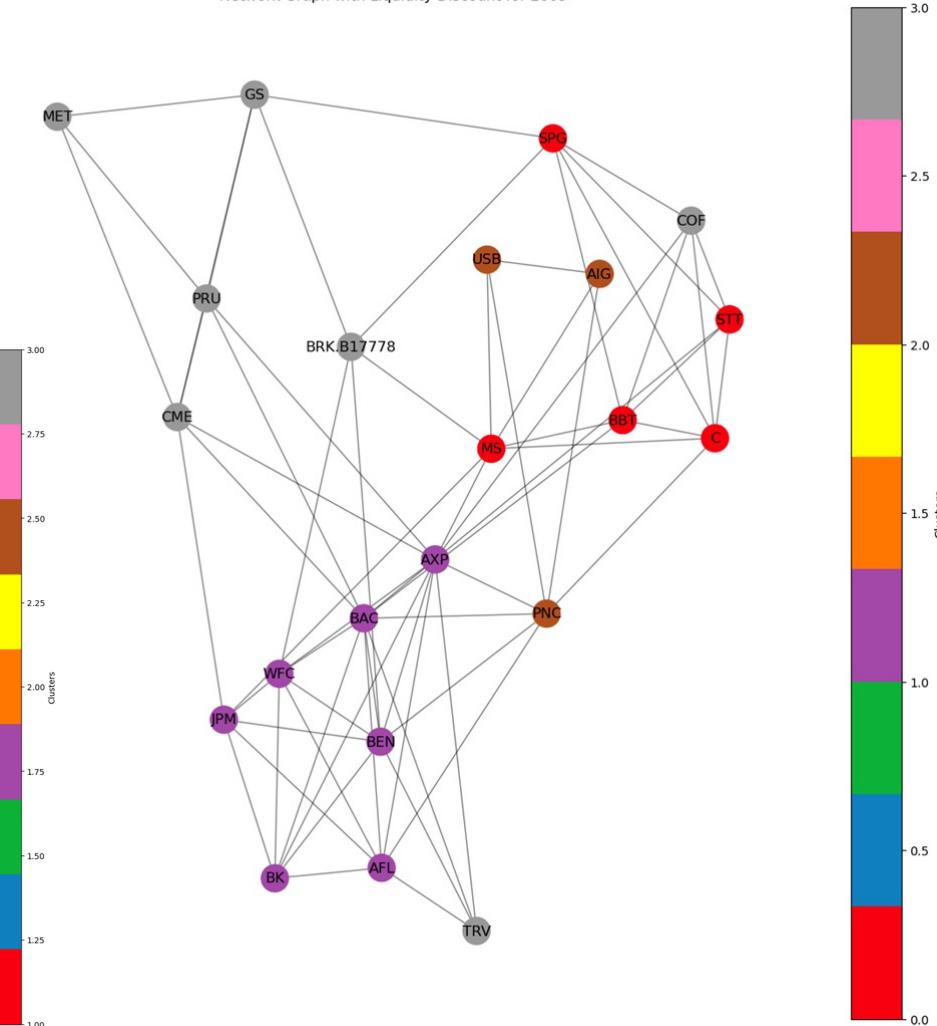
Network Graph with Liquidity Discount for 2007



Network Graph with Liquidity Discount for 2008 (including delisted firms)



Network Graph with Liquidity Discount for 2009



Financial Knowledge Graph with NLP (Working Paper submitted)

- Data Source: Textual News from LexisNexis; news must be written in English
- News Sample: [Appendix D-D.1](#)
- Combining Knowledge from news with the financial KG.

Methodology:

- Clean up the textual data(remove stopping word, normalized text, stemming etc.)
- Pick 20 firms align with Financial KG, add additional troubled firms(Lehman, AIG, Bear etc.)
- KG1: Draw KG similar to Harry Potter KG with Spacy extract tags and count frequency ([Appendix D.1](#))
- KG2: Leverage **OpenAI** Embedding model convert text into embedding vectors(Ada-002, txtEmbedding-3 Large): [Appendix C.1](#)
- Draw KG based on the embedding vector with a machine learning approach(RNN)

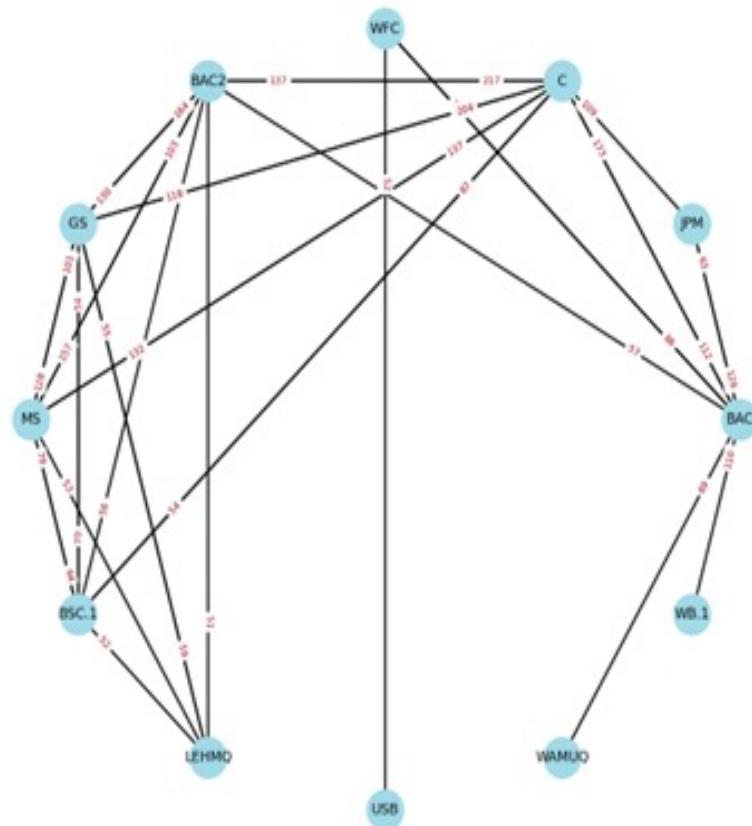
Number of organizations	27
Number of news	31,533
Number of sentences	2,286,011
Number of words	25,263,309
Number of characters	165,656,691

Ticker	Name	Number
AFL	Aflac	1
AXP	American Express	2
BAC	Bank of America	3
BBT	BB & T (Truist)	4
BEN	Franklin Resources	5
BK	Bank of NY Mellon	6
BRK	Berkshire Hathaway	7
C	Citi Group	8
CME	CME Group	9
COF	Capital One	10
GS	Goldman Sachs	11
JPM	JP Morgan	12
MET	MetLife	13
MS	Morgan Stanley	14
PNC	PNC Bank	15
PRU	Preduntial	16
SPG	Simon Property	17
STT	State Street	18
TRV	Travelers Group	19
USB	US BankCorp	20
WFC	Wells Fargo	21
AIG	AIG	22
BSC.1 (BSC)	Bear Sterns	23
LEHMQ (LEH)	Lehman Brothers	24
BAC2 (MER)	Merrill Lynch	25
WAMUMQ (WAMU)	Washington Mutual	26
WB.1 (WB)	Wachovia	27

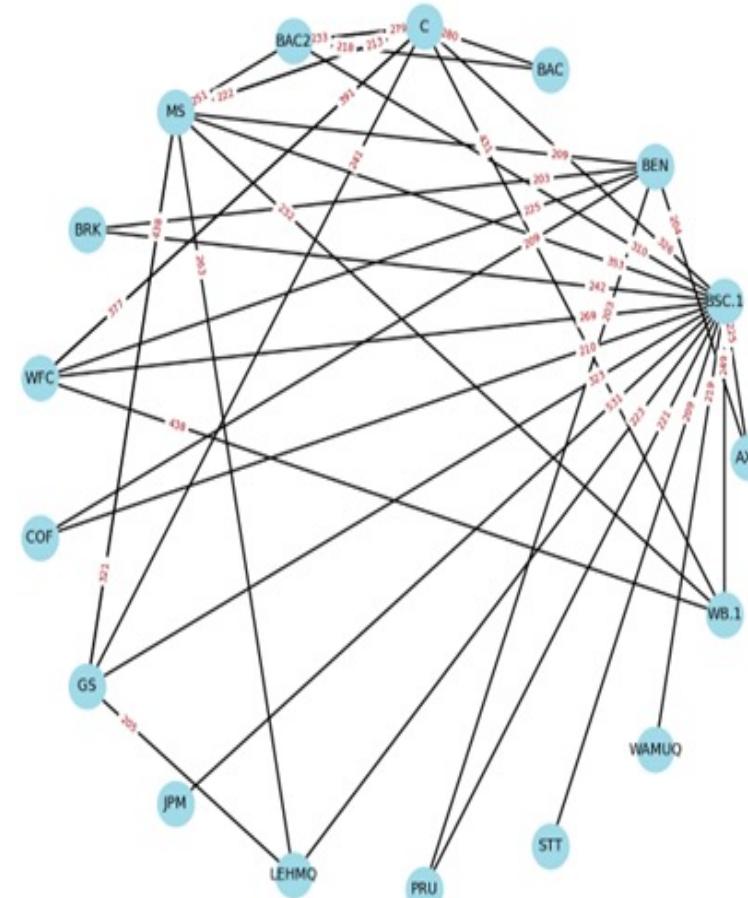
Financial KG with News Tags(Similar to the harry potter KG)

2007-2009

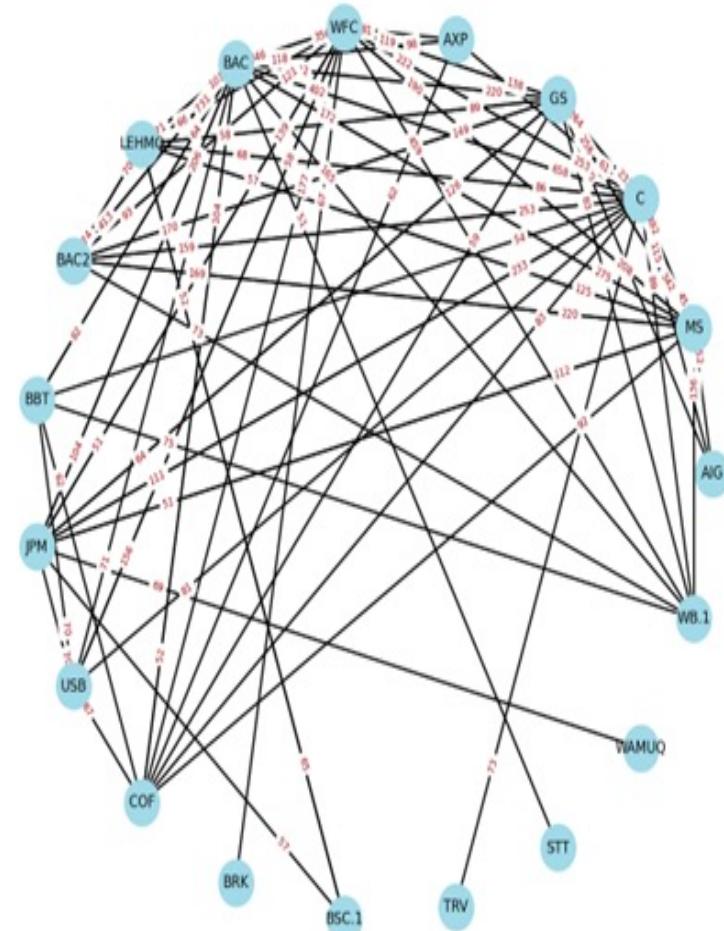
2007 News Network Graph Including Six Firms (with Interconnection > 50)



2008 News Network Graph (with Interconnection > 200)



2009 News Network Graph (with Interconnection > 50)



Some empirical Observations

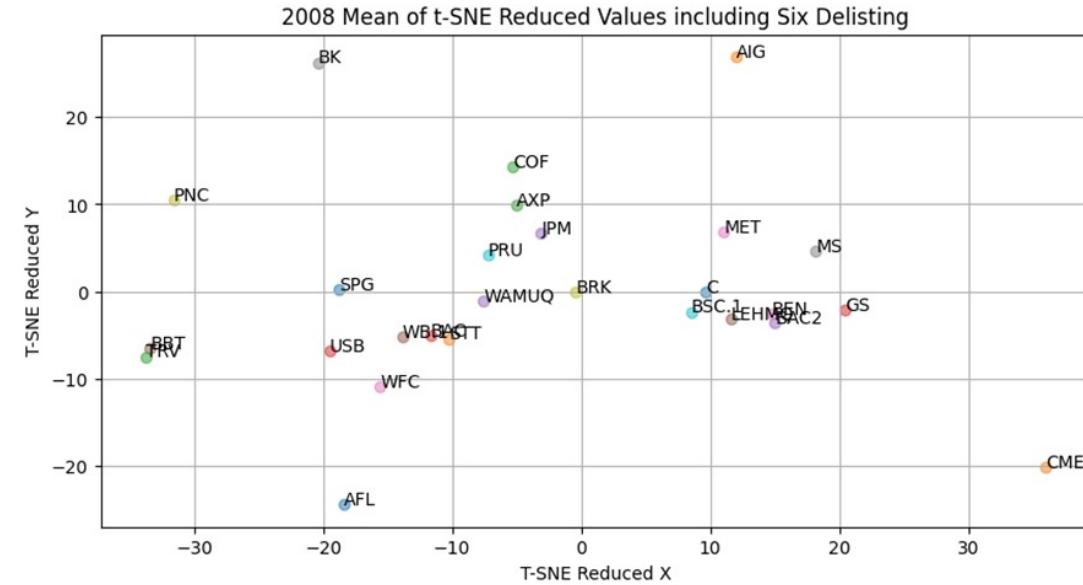
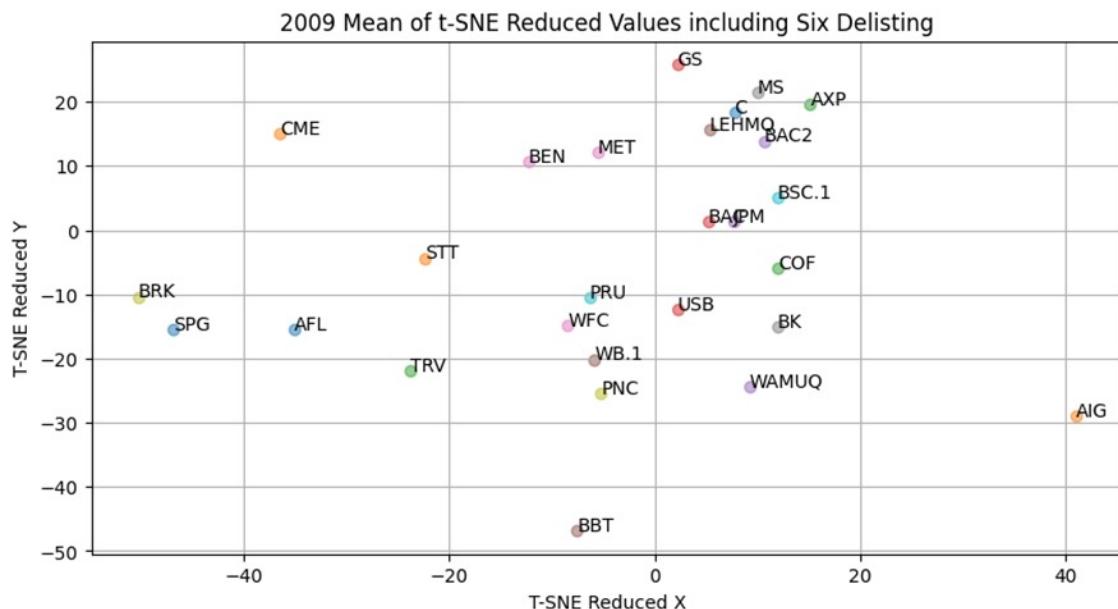
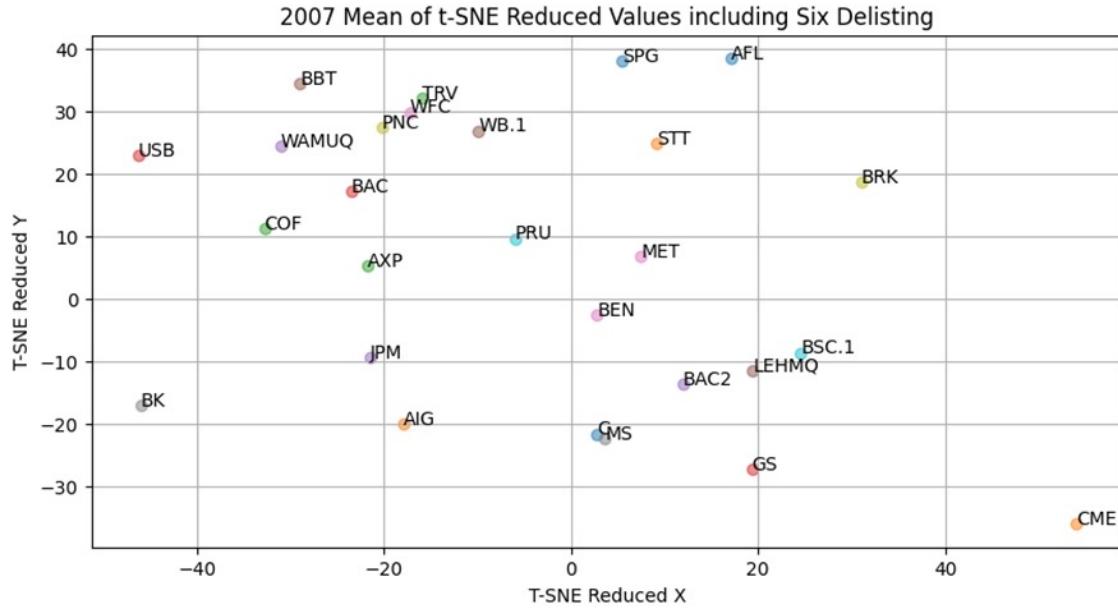
Observations:

- In 2007, bankrupt firms: Lehman (LEHMQ) , Bear Sterns (BSC.1), Washington Mutual (MAMUQ), and Wachovia (WB.1), had very few connections to the other major banks
- most connected firms are Merrill Lynch (BAC2), Bank of America (BAC) and Citi Group (C)
- In 2008, more firms were connected and mentioned together in the news, particularly for Bear Stearns.
- The network structure was more complex in 2008

Conclusions:

- Before the crisis, it was hard to imply who was more vulnerable to the crisis.
- The network is more complex post-crisis than pre-crisis due to the media impact.

Embedding Data study

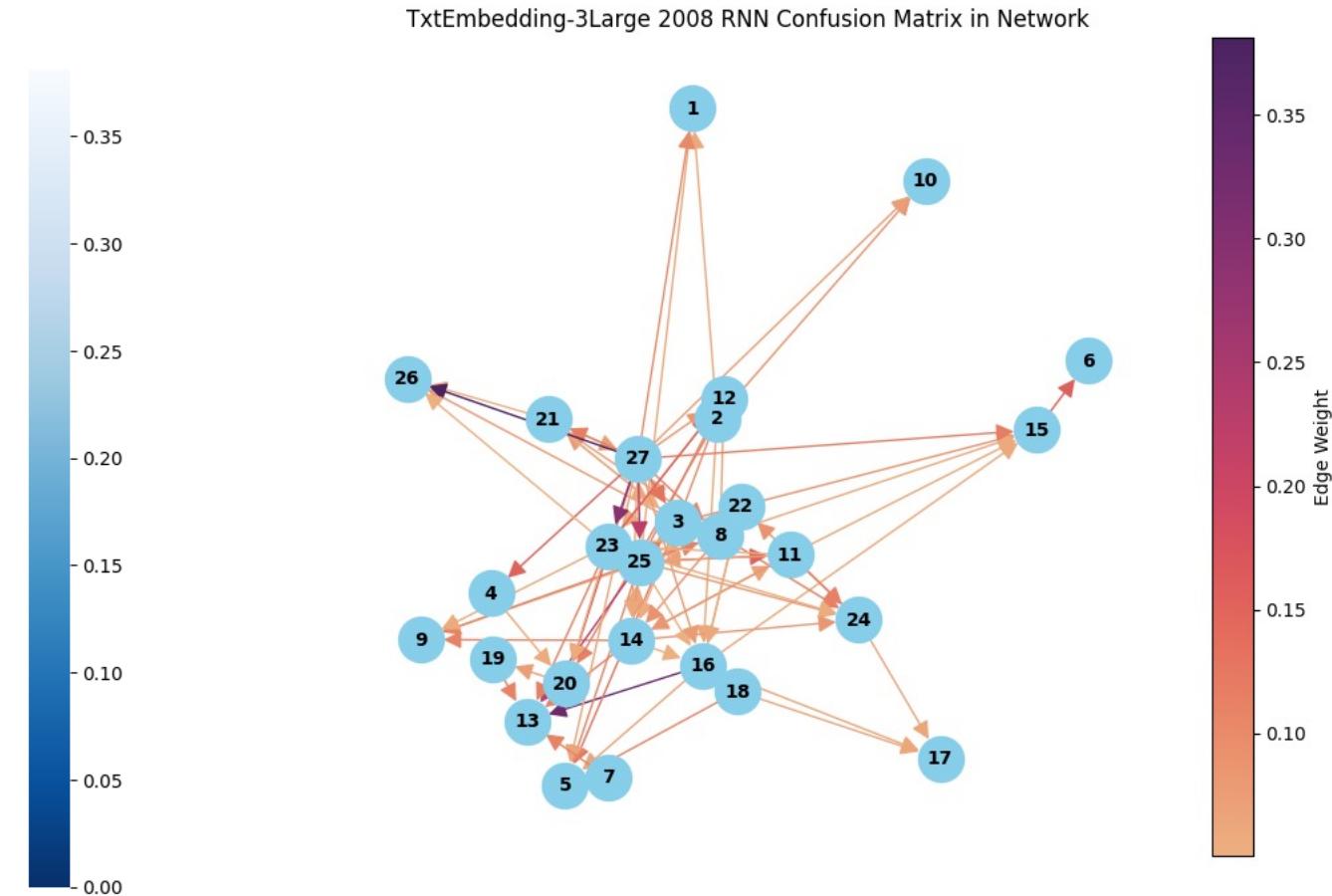
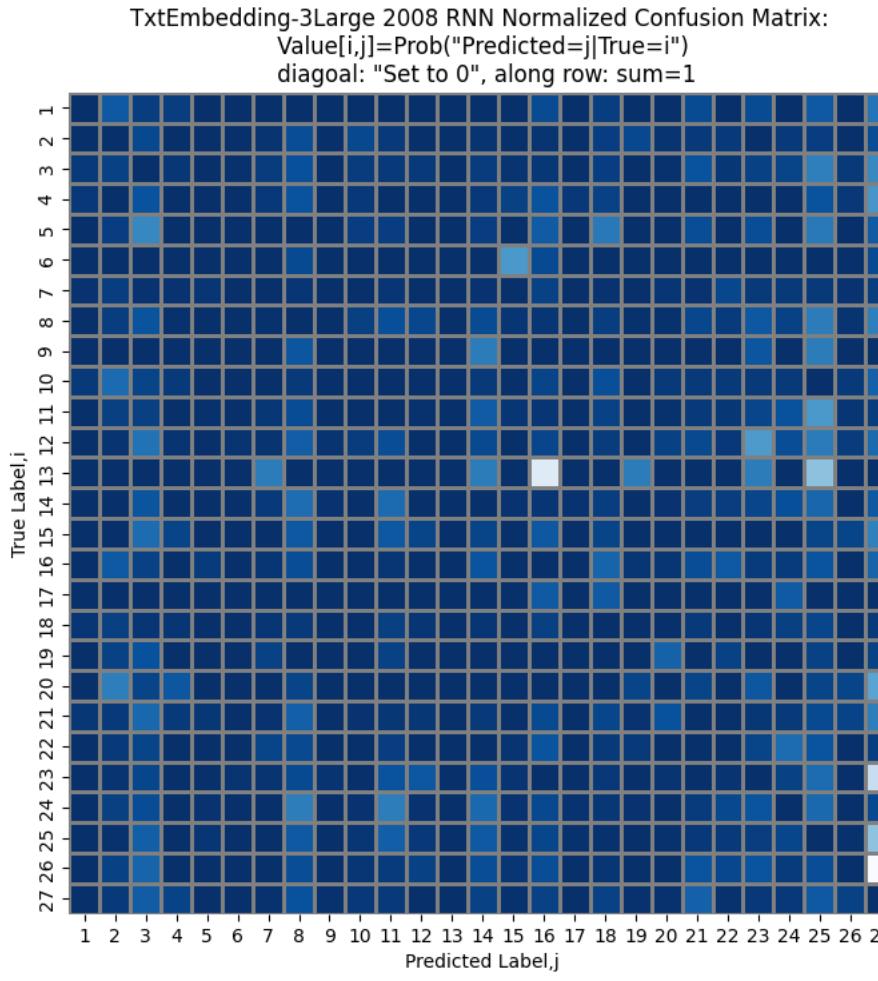


1. Generating embedding from the news via OpenAI API with fixed dimensions of embedding vector.
2. t-SNE reduced the daily news vector(1536) dimension into X-Y axis, and averaged daily for given year
3. 2007:
 - ❖ Bear, Lehman, Merrill are close, and GS is not too far
 - ❖ Upper left corner contains several retail banks
4. 2008-2009:
 - ❖ Firms move closer, the x-axis and y-axis are narrower
 - ❖ 2009: Citi Group, Morgan Stanley, and American Express are not too different from the bankrupt firms (Lehman, Bear, and Merrill Lynch)

Financial KG with Embedding data using ML approach

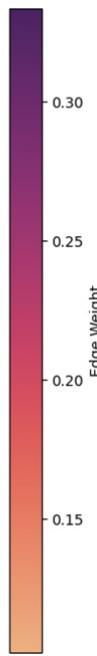
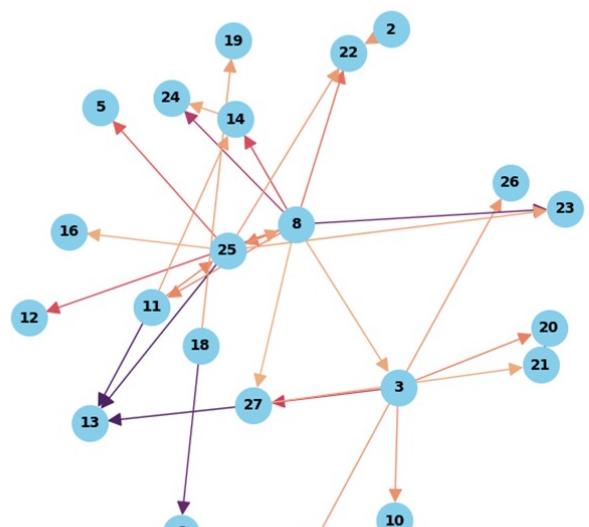
- Using text embedding and Recurrent Neural Network(RNN) to draw KG.
- RNN is used to predict the label of 27 firms(1-27) based on the embedding vector of the news from OpenAI model.
- Confusion matrix calculated from the testing dataset (20%) is used as the edge;
- Confusion matrix indicate the similarity across 27 firms, just like **correlation**:
 - ❖ For example, based the embedded news, what is the probability of confusing/misclassifying Morgan Stanley (#14) vs JPM (#12): $P(\text{Predicted}=14 \mid \text{True}=12)=20\%$
 - ❖ Confusion matrix is normalized to ensure probability sum up to 1; the diagonal is zeroed out.
 - ❖ Benchmarking performed between 2nd (ADA) and 3rd (Text-Large) generation of OpenAI model shows quite similar results.

Financial KG with Embedding data (for illustration purpose)

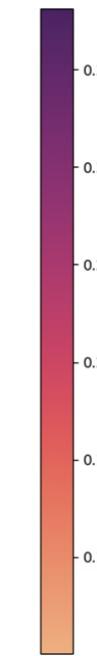
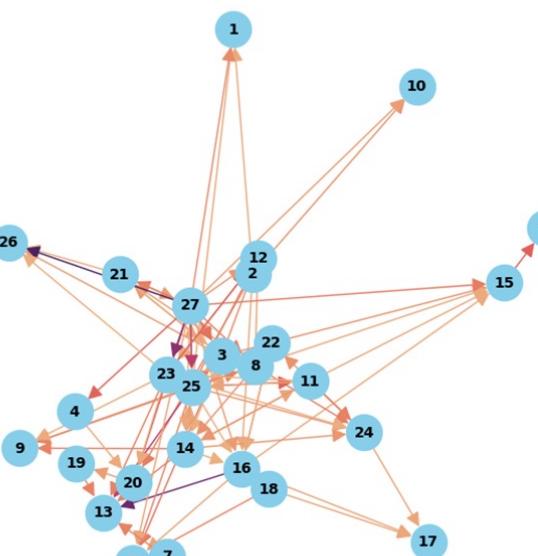


More results for 2007-2009

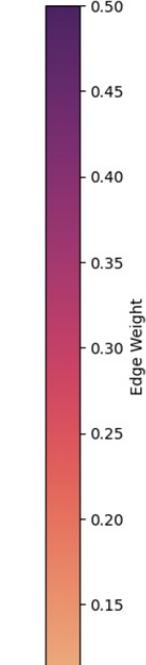
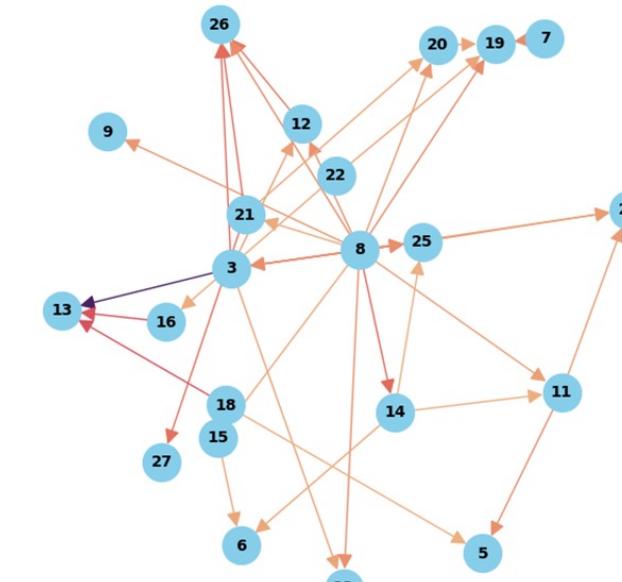
TxtEmbedding-3Large 2007 RNN Confusion Matrix in Network



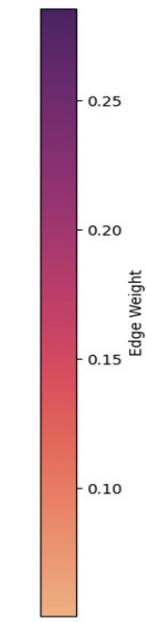
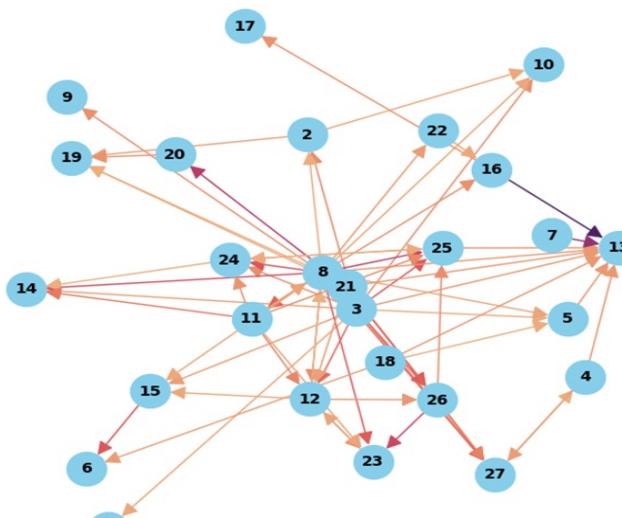
TxtEmbedding-3Large 2008 RNN Confusion Matrix in Network



TxtEmbedding-3Large 2009 RNN Confusion Matrix in Network



TxtEmbedding-3Large RNN Confusion Matrix in Network



Name	Number
Aflac	1
American Express	2
Bank of America	3
BB & T (Truist)	4
Franklin Resources	5
Bank of NY Mellon	6
Berkshire Hathaway	7
Citi Group	8
CME Group	9
Capital One	10
Goldman Sachs	11

JP Morgan	12
MetLife	13
Morgan Stanley	14
PNC Bank	15
Prudential	16
Simon Property	17
State Street	18
Travelers Group	19
US BankCorp	20
Wells Fargo	21
AIG	22
Bear Sterns	23
Lehman Brothers	24
Merrill Lynch	25
Washington Mutual	26
Wachovia	27

Systemic Risk Questionnaires based on the graph

- Integrate the empirical results with the following qualitative questions:

- ❖ Is network denser in 2008 (i.e. higher systemic risk in 2008)?
- ❖ Can model predict defaults of Lehman and Bear Sterns?
- ❖ Can model predict the trouble of AIG?
- ❖ Who are in the center of the network?
- ❖ Who are closest to Bear Sterns and Lehman Brothers?

- **Conclusion:** Graph model with NLP, Volatility and Liquidity metrics cannot predict systemic risk of 2008.

Example: Results of Empirical observation

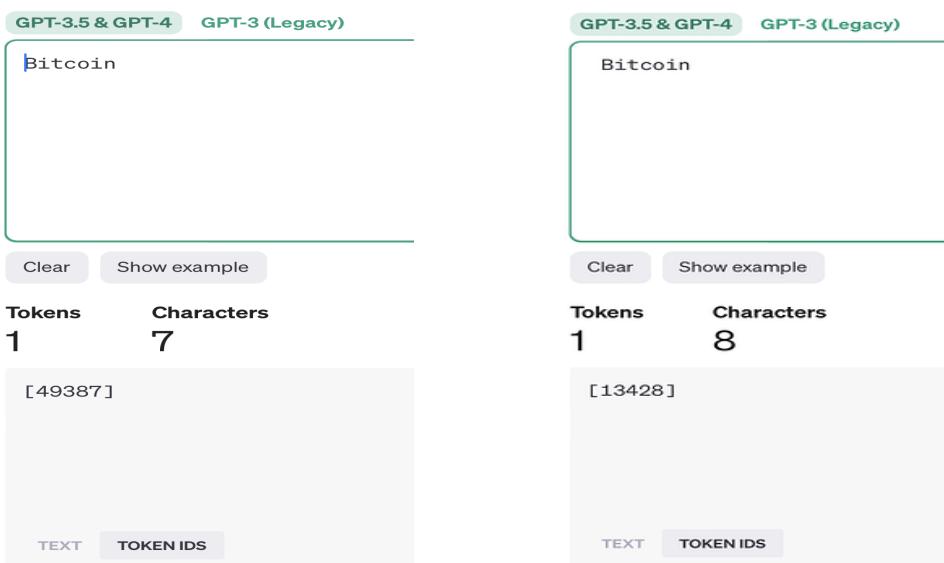
		Is network denser in 2008?	Predict trouble of AIG?	Center of network in 2007	Center of network in 2008	Center of network in 2009	Who are closest to Lehman and Bear Sterns
1	t-SNE	yes	no	n.a.	n.a.	n.a.	GS, MS, C, MER
2	frequency count	yes	no	BAC, GS, C	BSC	BAC, C	GS, MER, MS
3	OpenAI-ada-002	yes	no	BAC, GS	BAC, GS, C	BAC, GS, C	GS, C
4	OpenAI-3large	yes	no	C, MER	BAC, C, AIG, BSC, MER	C	C
5	centrality score	n.a.	n.a.	C > GS > (BAC, MER, MS) > BSC	BSC > C > MS	C > BAC > MS > WFC > GS	n.a.
6	volatility	yes	no	GS, BSC	everyone	BRK, WFC, JPM	too hard to tell
7	liquidity index	yes	no	GS	BRK, WB, MS	BRK	everyone

Other applications in Finance

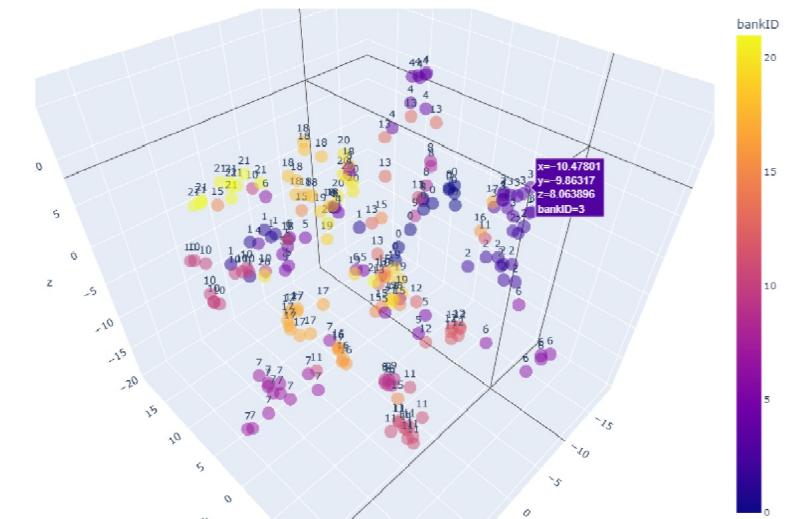
- Use OpenAI embedded vector for asset pricing: predict stock return ([Appendix D](#))
- Mutual Fund selection and similarity analysis ([Fund2Vec](#))
- Market Sentiment Analysis
- Bank network analysis

Observations with the Open AI embedding Model

- Embeddings are **not** based on an analysis of input text, but on the OpenAI **pre-trained model** from the internet sources
- Embedding vectors are highly correlated with many common mean value, [post-processing](#) is recommended
- Textual data must be **normalized** before applying the Open AI embedding model (e.g., “Bitcoin” vs “ Bitcoin” with different token)



3-D T-SNE



Appendix A: Partial Correlation/Conditional Correlation

Partial Correlation example:

- Apple and MSFT, all have strong correlation with SP500, we wants to know if we remove the impact of SP500, what is the effect of conditional/residual correlation?
- *Return of Apple stock: X, Return of MSFT stock: Y, Return of SP500: M*
- $\rho(X, Y|M) = \frac{\rho(X,Y) - \rho(X,M)\rho(Y,M)}{\sqrt{[1-\rho^2(X,M)][1-\rho^2(Y,M)]}}$
- Overall, lower than the Pearson correlation
- For higher dimension, the precision matrix can be solved via Lasso method (*Partial Correlation Graphical LASSO [2104.10099] Partial Correlation Graphical LASSO (arxiv.org)*)

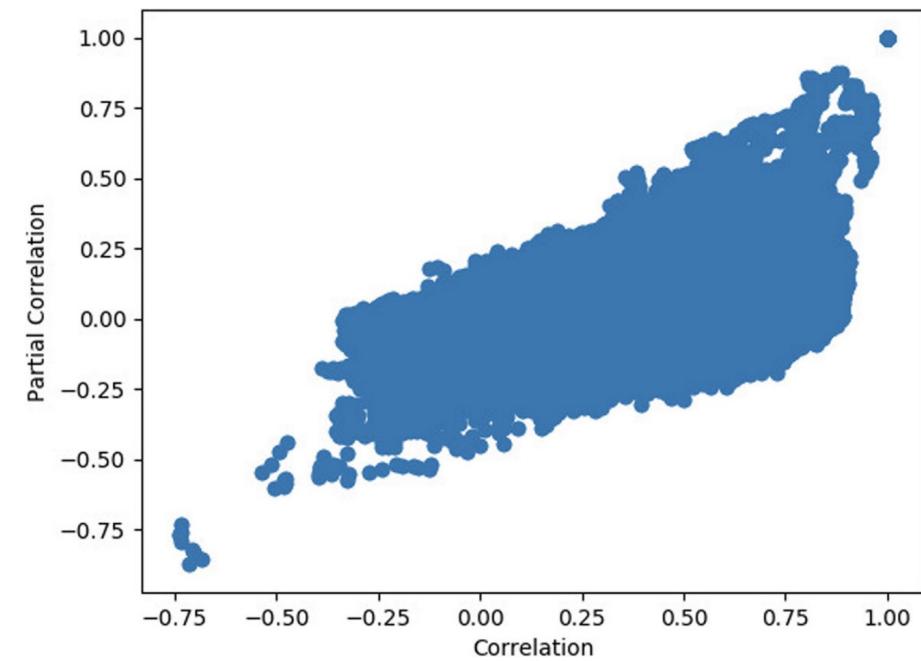


Fig. 4 Scatter plot of the correlation coefficient for an edge against that of the partial correlation coefficient for each of the 140 networks in the dataset. The partial correlation coefficients are in general smaller than their corresponding correlation coefficients which is to be expected if the indirect correlations are reduced, however there is still a relationship between the two

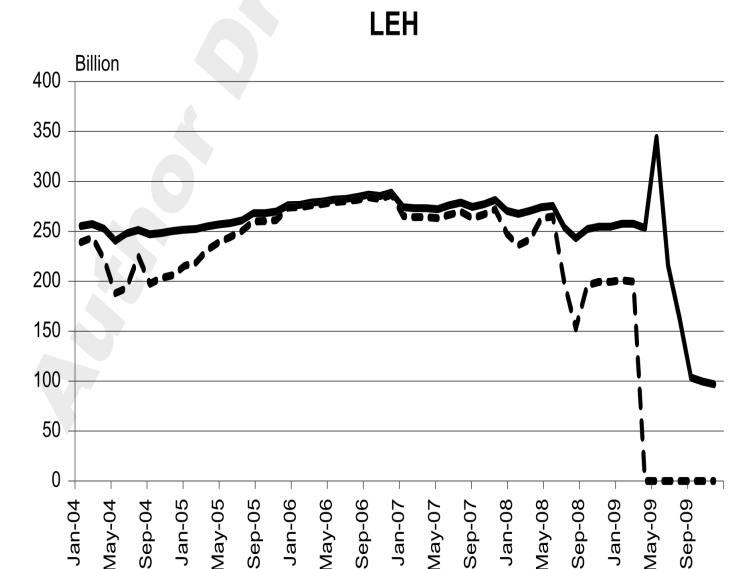
Partial Corr v Corr(*Partial correlation in financial networks, Applied Network Science 2020*)

Appendix B(1): Liquidity Discount in a nutshell*

- Liquidity metrics that inspired by the 2008 Financial Crisis when asset price started to **deviated from economic fundamentals**(e.g., Bear Stearns' stock price drop, MBS difficult to sell)
- Financial institutions can economically default (e.g., Countrywide Financial) vs. liquidity default (e.g., Lehman) and could cross the default boundary
- Case study demonstrate that Lehman's Liquidity issue started back to 2007**

EXHIBIT 5

Liquid and Illiquid Asset Values of Lehman



Liquidity discount ratio: explains the asset sell-off when Asset fundamentals are not changing too much:

- ❖ Liquidity discount ratio is modeled based the ratio between Asset value under economically default and Liquidity default.
- ❖ Asset value is modeled as a put option follow the Merton model and compound option model (Geske 1977), incorporating the leverage factor, CDS Spread etc.
- ❖ **An index** was created for Liquidity Discount ratio including over 200 financial firms from 2006-2014 (**Data source** used in the Systemic risk study ***)

Reference:

**A liquidity Discount, Journal of Fixed Income 2012*

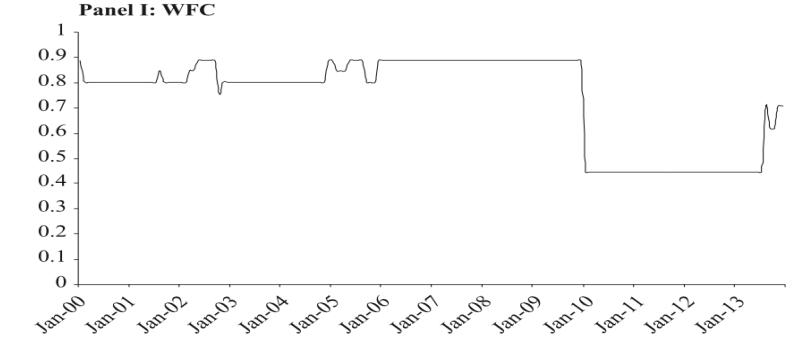
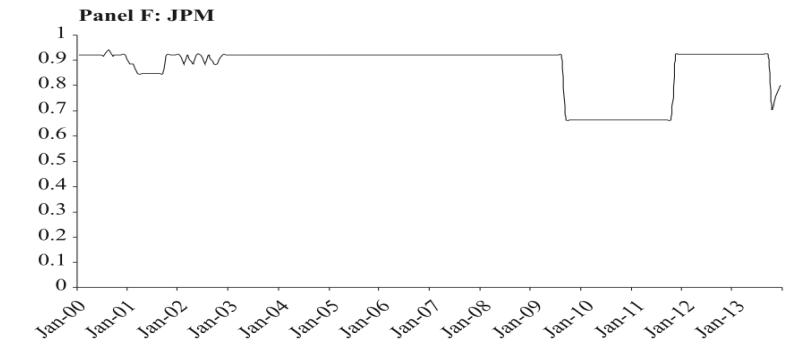
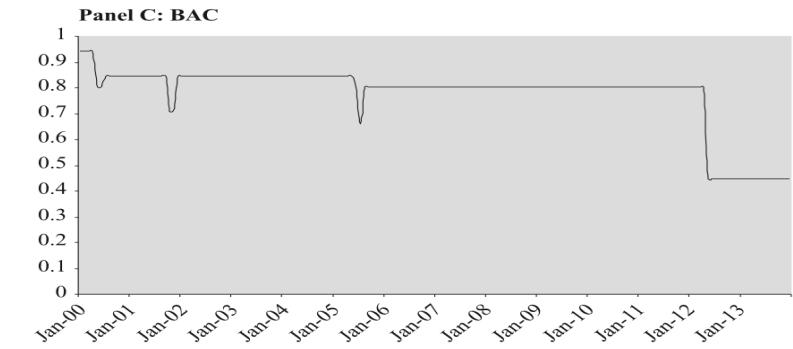
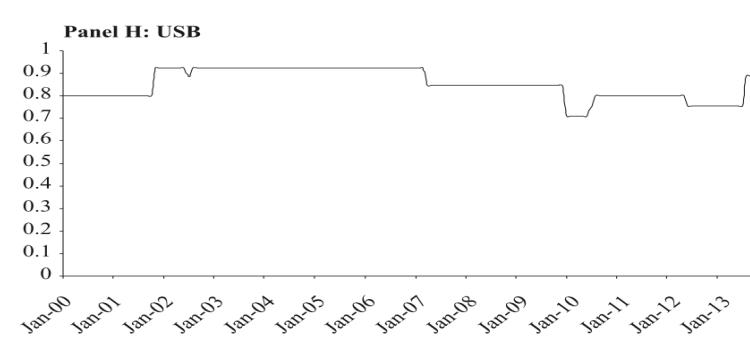
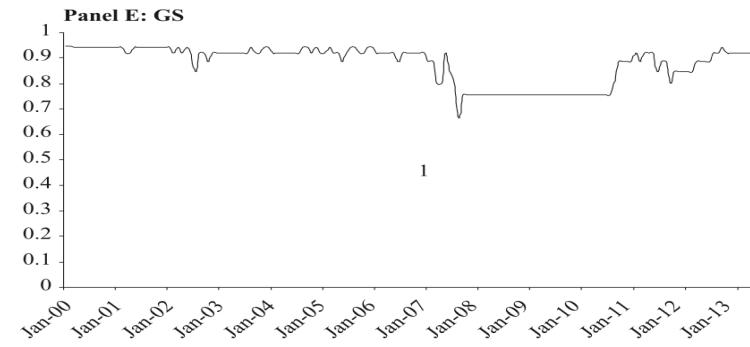
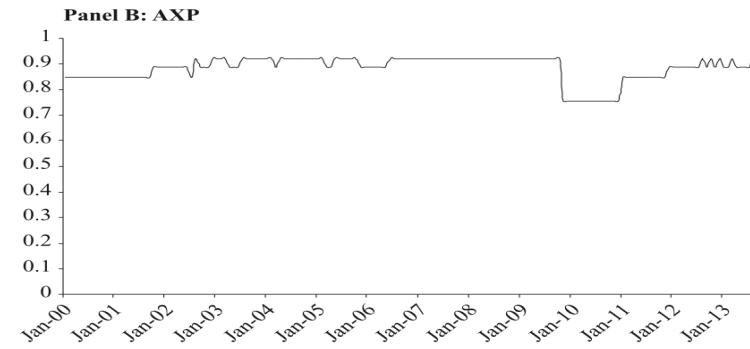
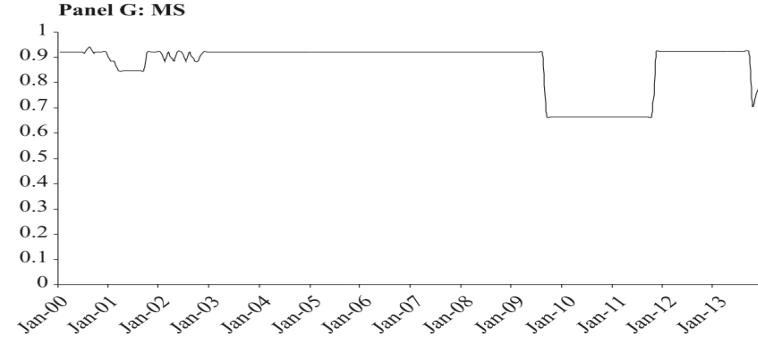
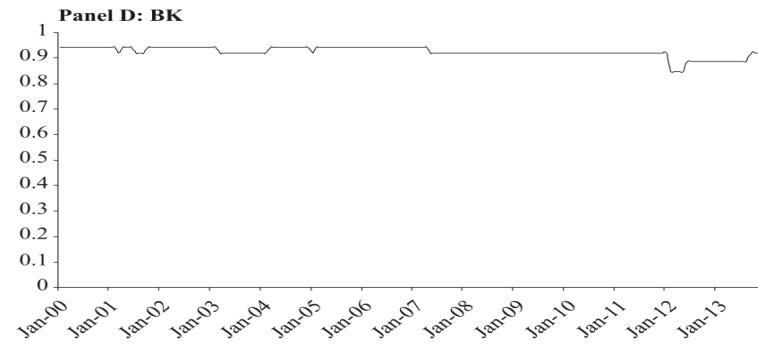
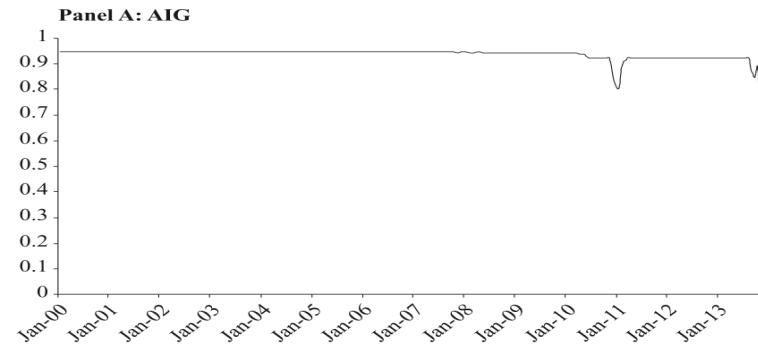
***Valuing Financial Assets with Liquidity Discount: An Implication for Basel III Journal of Fixed Income 2013*

****A Liquidity Index, Journal of Fixed Income 2016*

Appendix B(2)

EXHIBIT 10

Liquidity Discount in Asset Values (percentage)

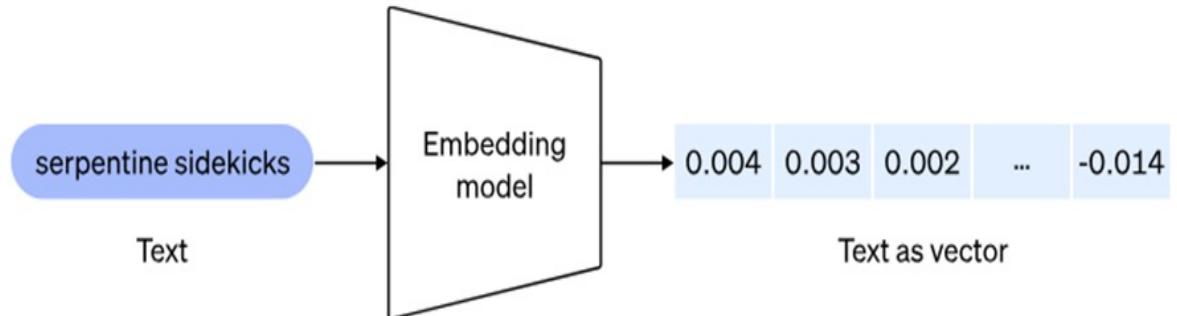


Reference: A Closed-Form Solution to the Liquidity Discount Problem: With an Application to the Liquidity Crisis Chen, Ren-Raw; Li, Bo. The Journal of Fixed Income; (Fall 2015)

Appendix C-1

- OpenAI API: Tokenized model is charged per token.
- 1 token = 4.63 Chars, 10^9 tokens = \$100
- Ada-002: transform textual data into a fixed numerical array 1536x1, token limit of 8191
- text-embedding-3-large: third generation model with 3072-dimension fixed length vector for each news
- New Features: can be shrink to 252 dimensions without losing detail of the information

```
1 # Define prompt
2 prompt_template= """
3 Write a summary within 100 words of the following:
4 "{text}"
5 SUMMARY:
6 """
7 prompt= PromptTemplate.from_template(prompt_template)
8
9 # Define LLM chain
10 llm= ChatOpenAI(
11     temperature= 0,
12     model_name= "gpt-3.5-turbo")
13
14 llm_chain= LLMChain(
15     llm= llm,
16     prompt= prompt)
17
18 theSummary= llm_chain.run(theText)
19 print(theSummary)
```



	GPT-4 Turbo	GPT-4 32K
Provider	OpenAI	OpenAI
Input Context Window	128K tokens	32.8K tokens
Maximum Output Tokens	4,096 tokens	Not specified.
Release Date	November 6th, 2023 8 months ago	March 14th, 2023 over 1 year ago

Pricing

GPT-4 Turbo is roughly 6.0x cheaper compared to GPT-4 32K for input tokens and roughly 4.0x cheaper for output tokens.

	GPT-4 Turbo	GPT-4 32K
Input	\$10.00 per million tokens	\$60.00 per million tokens
Output	\$30.00 per million tokens	\$120.00 per million tokens

Source: [OpenAI](#)

Example: API Usage ,hyperparameter like **temperature** can be tuned

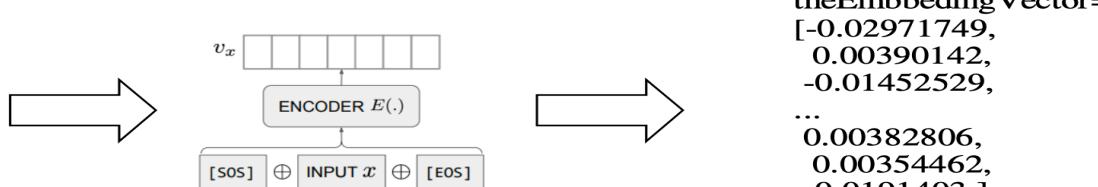
Appendix C-2

text
↓

		India Insurance News		December 1, 2016 Thursday 6:30 AM EST Fitch Affirms Aflac's Ratings; Outlook Negative				
		LENGTH: 1126 words Dec. 1 -- Fitch Ratings has affirmed the 'A+' Insurance Financial Strength (IFS) ratings of Aflac Inc.'s (Aflac) insurance subsidiaries. Fitch has also affirmed Aflac's Issuer Default Rating (IDR) at 'A' and senior debt ratings at 'A-'. The Rating Outlook remains Stable for Aflac's holding company ratings and Negative for its insurance subsidiaries. A full ratings list follows at the end of this release. KEY RATING DRIVERS The affirmation of Aflac's ratings reflect the company's extremely strong competitive position in the supplemental accident and health insurance markets in Japan and the U.S., extremely strong earnings profile and very strong capitalization. The ratings also incorporate the impact of slow economic growth in Aflac's key Japanese market and its sizeable exposure to Japanese sovereign risk including significant concentration in Japanese government bonds (JGB). Aflac's IFS rating is capped at one notch above Japan's sovereign rating of 'A', due to Aflac Japan's exposure to Japan's economy. Aflac Japan comprises more than two-thirds of						
A	B	C	D	E	F	G	H	I
1	bankID	Name	Content	nChars	nWords	nSents	nTokens	openAI_emb
2	0	AFLAC	India Insurance News	7280	1257	35	1716 [-0.02971749, 0.00390142, -0.01452529, ... 0.00382806, 0.00354462, -0.0191403]	1536 dim
3	1	AFLAC	India Insurance News	6725	1185	34	1648 [-0.02405601, -0.00344566, -0.01790806, ... 0.00398813, -0.00153196, -0.02785996]	
4	2	AFLAC	India Insurance News	14410	2657	103	3235 [-0.0389362, -0.00374613, -0.01878639, ... 0.00291858, 0.00856072, -0.01637418]	

Figure 4: OpenAI’s text-embedding-ada-002

```
theInputText= “
India Insurance News
December 1, 2016
Thursday 6:30 AM EST
Fitch Affirms Aflac's
Ratings; Outlook Negative
LENGTH: 1126 words
Dec. 1 -- Fitch Ratings has
affirmed the 'A+' Insurance
Financial Strength (IFS)
ratings of Aflac Inc.'s
(Aflac) insurance
subsidiaries. Fitch has also
affirmed Aflac's Issuer
Default Rating (IDR) at 'A'
and senior debt ratings at
'A-'. ....”
```



Appendix D

Edit Database Cell

Mode: Text

The Washington Post November 13, 2007 Tuesday Suburban Edition WALL STREET SECTION: FINANCIAL; Pg.D02 LENGTH: 784 words WALL STREET Banks Join Nasdaq

1 Trading Portal Wall Street's biggest investment banks, including Goldman Sachs and Morgan Stanley, plan to drop their competing platforms for trading unregistered securities and join the Nasdaq Stock Market's system. Twelve firms will cooperate to create the Portal Alliance for so-called 144A securities, Nasdaq said in a statement. Nasdaq's Portal system currently lists 700 equity securities and has gained almost 100 subscribers since it was launched in August. CONSUMER SAFETY Study Sheds Light on Toy Recalls Toy recalls in the United States increased between 1992 and 2006, regardless of where the toys were produced, according to a new study by a researcher at the University of Manitoba's I.H. Asper School of Business. And recalls of toys made in China have not increased along with the rise in toy imports from China, the study found. "Blaming China or any other country is unlikely to enhance toy safety," Hari Bapuji, the author, wrote. "The more important question ... is: Why are the toys becoming increasingly unsafe?" LABOR 4 Ford Plants Approve Contract United Auto Workers members at Ford Motor plants in Kentucky, Missouri and Ohio voted to approve a four-year contract, moving the agreement closer to ratification. Ford has committed to produce new vehicles at various plants. The accord also calls for new workers to be paid less than current employees and shifts responsibility for retiree health care to a union-run fund. MERGERS & ACQUISITIONS Winemakers Agree to Combine Constellation Brands, the world's largest winemaker, agreed to buy Fortune Brands' wine unit for \$885 million, which would boost its U.S. vineyard holdings by 13 percent. The purchase would include five wineries and 1,500 acres of vineyards in California, the companies said. Constellation would gain the Clos du Bois, Geyser Peak and Gary Farrell labels. IBM to Buy Cognos IBM agreed to buy Cognos for \$5 billion in cash, its largest acquisition. The deal for Cognos, which already had a business partnership with IBM, is expected to be completed in the first quarter. With Cognos, IBM would better compete with Oracle and SAP in providing software that tracks corporate performance. EXECUTIVES Adobe Names New CEO Adobe Systems, the biggest maker of graphic-design software, promoted Shantanu Narayen to chief executive. Bruce Chizen will remain chief executive until Dec. 1 and serve out the rest of his term on Adobe's board, the company said. Adobe also announced that sales may be at the high end of its forecast for the quarter. INSURANCE Board Coverage Costs Fall The cost of insurance for corporate boards dropped at a slower rate in the third quarter than earlier this year, falling by 16 percent in the three months ended Sept. 30 from a year earlier. Insurance broker Aon Corp., which polled firms, also said 19 class-action suits have been filed by shareholders this year against corporate boards, surpassing last year's total. INVESTING Icahn Funds Post Quarterly Loss Hedge funds managed by Carl Icahn, left, posted their first quarterly decline since they opened three years ago because of volatile markets and losses on stocks including WCI Communities and Lear. The funds, run by Icahn Capital Management, together fell 1.5 percent in the third quarter. The firm reversed more than \$27 million of incentive fees as a result of the drop, according to a filing by Icahn Enterprises, parent of the fund-management business. TELECOM EchoStar Shares Drop 16 Percent Stock in EchoStar Communications, the second-largest U.S. satellite television service, dropped the most in 7 1/2 years after third-quarter subscriber growth fell short of some analysts' estimates. EchoStar shares fell \$7.68, or 16 percent, to \$40.83 in Nasdaq Stock Market trading, the steepest decline since April 2000. The deteriorating housing and credit markets and increased competition hurt demand and will continue to slow customer growth, EchoStar said in a regulatory filing. TELECOM Google Contest for Cell Software Google is offering \$10 million in prizes for people who build the best software to enhance the company's upcoming cellphone operating system. The operating system will be based on computer code that can be openly distributed among programmers, which Google hopes will encourage developers to create new software and improvements that could spawn new uses for smart phones. EARNINGS Tyson Foods said fourth-quarter profit was \$32 million compared with a year-ago loss of \$56 million. Revenue rose 6.4 percent to \$6.88 billion. Higher prices for poultry and beef

Type of data currently in cell: Text / Numeric

Apply

Appendix D.1

Textual Data Processed with SpaCy to extract tags

The Washington Post **ORG** November 13, 2007 **DATE** Tuesday **DATE** Suburban Edition WALL STREET SECTION: FINANCIAL; Pg. D02 LENGTH: 784 **CARDINAL** words WALL STREET Banks Join Nasdaq Trading Portal Wall Street's biggest investment banks, including Goldman Sachs **ORG** and Morgan Stanley **ORG**, plan to drop their competing platforms for trading unregistered securities and join the Nasdaq Stock Market's **ORG** system. Twelve **CARDINAL** firms will cooperate to create the Portal Alliance **ORG** for so-called 144A **CARDINAL** securities, Nasdaq **ORG** said in a statement. Nasdaq **ORG**'s Portal system currently lists 700 **CARDINAL** equity securities and has gained almost 100 **CARDINAL** subscribers since it was launched in August **DATE**. CONSUMER SAFETY Study Sheds Light on Toy Recalls Toy recalls in the United States **GPE** increased between 1992 and 2006 **DATE**, regardless of where the toys were produced, according to a new study by a researcher at the University of Manitoba's **ORG** I.H. Asper School of Business **ORG**. And recalls of toys made in China **GPE** have not increased along with the rise in toy imports from China **GPE**, the study found. Blaming China **WORK_OF_ART** or any other country is unlikely to enhance toy safety," Hari Bapuji **PERSON**, the author, wrote. "The more important question . . . is: Why are the toys becoming increasingly unsafe?" LABOR 4 **CARDINAL** Ford **ORG** Plants Approve Contract United Auto Workers **ORG** members at Ford Motor **ORG** plants in Kentucky **GPE**, Missouri **GPE** and Ohio **GPE** voted to approve a four-year **DATE** contract, moving the agreement closer to ratification. Ford **ORG** has committed to produce new vehicles at various plants. The accord also calls for new workers to be paid less than current employees and shifts responsibility for retiree health care to a union-run fund.

MERGERS & ACQUISITIONS Winemakers Agree to Combine Constellation Brands **ORG**, the world's largest winemaker, agreed to buy Fortune Brands' **ORG** wine unit for \$885 million **MONEY**, which would boost its U.S. **GPE** vineyard holdings by 13 percent **PERCENT**. The purchase would include five **CARDINAL** wineries and 1,500 acres **QUANTITY** of vineyards **LOC** in California **GPE**, the companies said. Constellation would gain the Clos du Bois **ORG**, Geyser Peak **PERSON** and Gary Farrell **PERSON** labels. IBM **ORG** to Buy Cognos **PERSON** IBM **ORG** agreed to buy Cognos **PRODUCT** for \$5 billion **MONEY** in cash, its largest acquisition. The deal for Cognos **PRODUCT**, which already had a business partnership with IBM **ORG**, is expected to be completed in the first quarter **DATE**. With Cognos **ORG**, IBM **ORG** would better compete with Oracle **ORG** and SAP **ORG** in providing software that tracks corporate performance. EXECUTIVES Adobe Names New **ORG** CEO Adobe Systems **ORG**, the biggest maker of graphic-design software, promoted Shantanu Narayen **ORG** to chief executive. Bruce Chizen **PERSON** will remain chief executive until Dec. 1 **DATE** and serve out the rest of his term on Adobe **ORG**'s board, the company said. Adobe **PERSON** also announced that sales may be at the high end of its forecast for the quarter **DATE**. INSURANCE Board Coverage Costs Fall **ORG** The cost of insurance for corporate boards dropped at a slower rate in the third quarter than earlier this year **DATE**, falling by 16 percent **PERCENT** in the three months ended Sept. 30 **DATE** from a year earlier **DATE**. Insurance broker Aon Corp. **ORG**, which polled firms, also said 19 **CARDINAL** class-action suits have been filed by shareholders this year **DATE** against corporate boards, surpassing last year's **DATE** total. INVESTING Icahn Funds Post Quarterly Loss Hedge funds managed by Carl Icahn **PERSON**, left, posted their first **ORDINAL** quarterly **DATE** decline since they opened three years ago **DATE** because of volatile markets and losses on stocks including WCI Communities **ORG** and Lear. The funds, run by Icahn Capital Management **ORG**, together fell 1.5 percent **PERCENT** in the third quarter **DATE**. The firm reversed more than \$27 million **MONEY** of incentive fees as a result of the drop, according to a filing by Icahn Enterprises **ORG**, parent of the fund-management business. TELECOM **GPE** EchoStar **PRODUCT** Shares Drop 16 Percent **PERCENT** Stock in EchoStar Communications, the second **ORDINAL**-largest U.S. **GPE** satellite television service, dropped the most in 7 1/2 years **DATE** after third-quarter **DATE** subscriber growth fell short of some analysts' estimates. EchoStar **PRODUCT** shares fell \$ 7.68 **MONEY**, or 16 percent **PERCENT**, to \$ 40.83 **MONEY**.

Appendix E

- Find the stock return for 27 firms
- Use +1/-1 as the stock return indicate profit/loss
- Using Embedded data as the feature to predict the classification of +1/-1
- Relevant research: [Expected Returns and Large Language Models](#)

Algo/Accuracy	OpenAI's text-embedding-ada-002 1563Dim	text-embedding-3-large 3072Dim	text-embedding-3-large-252Dim Reduced
MLP Classifier	0.60	0.56	0.54
Support Vector Classifier	0.55	0.55	0.54
K Neighbors Classifier	0.57	0.54	0.55
Logistic Regression	0.54	0.52	0.53
Random Forest Method	0.61	0.55	0.55