Integrating Sentiment-Driven Dynamics in Financial Contagion: Insights from Interbank and Social Media Networks

Wagner Eduardo Schuster*

Extended Abstract

Recent financial crises have highlighted the necessity of understanding the intricate dynamics within financial networks to effectively assess systemic risk. While existing research predominantly focuses on interbank contagion effects, this study takes a novel approach by investigating the additional potential impact of sentiment-driven bank runs on systemic risk. Drawing upon Diamond and Dybvig's framework, I examine the potential effects of self-fulfilling prophecies within financial networks, utilizing sentiment analysis derived from Twitter data to capture non-fundamental factors and integrate them into traditional financial contagion simulations.

The study utilizes a comprehensive dataset encompassing the balance sheets of Europe's largest 50 banks, alongside unstructured big data collected from Twitter. The analysis employs three main methodologies: interbank network estimation, sentiment analysis, and financial contagion simulations. To estimate interbank connections, I used the Minimum Density approach which optimizes linkages to reflect observed lending and borrowing patterns among financial institutions. The sentiment analysis assigns sentiment scores to individual tweets related to financial institutions, which are then used to construct a sentiment network. Employing Graphical Gaussian Models, this network estimates interconnections between banks based on their partial correlation, with a Lasso penalty ensuring a sparse solution. Furthermore, I propose a methodology to incorporate the effect of bank runs into interbank financial contagion simulations, providing insights into the additional effects on both the number of affected banks and the total assets lost.

The resulting interbank network reveals a core-periphery structure, underscoring the significance of certain banks, such as Crédit Agricole (AGRI), which exhibit high centrality despite not ranking as the largest in terms of total assets, owing to their interbank exposures. In contrast, the sentiment network reveals distinct connections among banks compared to interbank connections, capturing non-fundamental factors related to social media users' sentiments.

Financial contagion simulations highlight the significance of incorporating sentiment-driven bank runs into systemic risk assessments. While BNP Paribas (BNP) emerges as the most significant contributor to contagion in the absence of bank runs, sentiment connections reshape the contagion landscape, implicating banks like Lloyds Banking (LLOY), Barclays (BARC), and Coöperatieve Rabobank (RAB) due to their sentiment connections with other banks. Notably, banks with previously lower centrality now exhibit increased importance in the network, a crucial finding for regulators.

The study emphasizes the need for enhanced monitoring of both interbank and sentiment networks to identify early warning signals of systemic risk. Additionally, the dynamic nature of sentiment analysis enables near real-time computation of metrics, facilitating the rapid reconstruction of the networks and examination of potential contagion effects. By incorporating sentiment-driven contagion into the analysis, this study offers a more comprehensive framework for assessing systemic risk and understanding the interplay between fundamental and non-fundamental factors in financial networks.

*Erasmus Mundus Joint Master Degree QEM - Models and Methods of Quantitative Economics from Université Paris 1 Panthéon-Sorbonne, Università Ca' Foscari Venezia, and Universitat Autònoma de Barcelona

These findings contribute to the literature on financial contagion and have significant implications for regulators and policymakers, highlighting the importance of proactive monitoring and intervention strategies in mitigating systemic risk.

KEYWORDS: Systemic Risk; Financial Networks; Sentiment Analysis; Financial Contagion.

JEL codes: G01; G21; D85.

1 Introduction

Systemic risk is a critical concern in today's world, as externalities created by failures or distress in one bank can generate spillover effects, potentially leading to broader systemic crises. The interconnectedness and complexity of financial institutions and markets make identifying and mitigating systemic risk challenging.

Numerous studies have examined the structure and dynamics of financial networks to understand how shocks and contagion propagate through the system, potentially leading to systemic risk (Allen & Gale, 2000; Upper & Worms, 2004; Upper, 2010; Mistrulli, 2011; Battiston et al., 2012; Anand et al., 2015; Cerchiello et al., 2017). However, most of this literature focuses on contagion effects in the interbank market.

My approach is different: I plan to examine additional sentiment-driven effects, such as bank runs as in the self-fulfilling prophecy outlined by Diamond and Dybvig (1983). My research will explore the potential implications for systemic risk when incorporating sentiment analysis into network models, rather than focusing solely on the deterioration of banks' fundamentals.

Existing approaches often rely on structured data, which may overlook critical insights from unstructured sources. Advances in Artificial Intelligence, specifically in natural language processing (NLP) and large language models (LLMs), now enable the processing and extraction of meaningful information from large volumes of unstructured text. This research seeks to bridge that gap by integrating these technologies into the network analysis of financial systems.

Specifically, my research will address the following questions:

- How do sentiment-driven effects, such as bank runs, impact the propagation of financial contagion in the interbank market?
- Can sentiment analysis of unstructured big data improve the measurement and understanding of the dynamics between fundamental and non-fundamental factors in bank runs?
- How does the inclusion of unstructured big data aid in identifying systemically important banks and early warning indicators for financial contagion?

By exploring the impacts of sentiment-driven effects on Europe's largest banks, my research aims to provide valuable insights for regulatory agencies concerning macroprudential and targeted supervision. It offers a comprehensive framework for assessing bank interconnectedness and conducting stress test simulations to measure the potential impacts of events such as bank runs. Additionally, it aims to identify systemically important banks and provide early warnings to signal financial vulnerabilities, enabling timely policy interventions.

2 Literature review: Financial Networks

Allen and Gale's (2000) seminal work has been instrumental in understanding financial contagion arising from credit interlinkages among banks. They demonstrate that the propagation of an unforeseen liquidity shock and its systemic implications are heavily influenced by the pattern of interconnectedness among banks.

In this sense, network analysis is a valuable tool for understanding and analyzing systemic risk. Jackson and Pernoud (2021) highlight that measuring the contribution of banks to systemic risk involves identifying central, significant, or systemic nodes using financial centrality concepts. However, traditional centrality measures often do not apply to Financial Networks.

To measure a bank's systemic importance, one effective approach is to utilize financial contagion simulations (Battiston et al., 2012; Mistrulli, 2011). However, a common issue in such analyses is the lack of specific interbank connection data, necessitating the estimation of these networks for accurate contagion simulations. This study will follow Anand et al. (2015) in estimating interbank connections using the Minimum Density (MD) method. In this approach, banks avoid spreading their borrowing and lending across the entire system due to the prohibitive costs associated with information processing, risk management, and creditworthiness checks, which are manageable only for the largest banks.

Additionally, I aim to incorporate sentiment-driven effects as an amplification channel within conventional interbank financial contagion simulations. Using both direct measures (interbank exposures and balance sheet information) and indirect measures (non-fundamental aspects), this approach will provide a holistic understanding of systemic risk. For this, sentiment analysis from social media will be employed.

Social media has emerged as a significant factor in bank runs, rapidly spreading information and opinions that can fuel panic and lead to deposit withdrawals. Shiller (2019) emphasizes the role of shared narratives and group sentiments in driving economic events. Data from Twitter (Cerchiello et al., 2017; Cookson et al., 2023), the Bank of England, Reuters, broker reports (Nyman et al., 2021), and Eastmoney Net (Fan et al., 2021) has been used to underscore the importance of sentiment analysis in understanding bank interconnections and risk contagion.

Traditional approaches often rely on limited structured data. This study will leverage advances in NLP and LLMs to process unstructured text and extract meaningful information, thereby bridging the gap in network analysis of financial systems.

3 Methodology and data

The objective of this section is to provide an overview of the methodology and data set utilized in this paper. Firstly, the dataset employed in the study is presented, followed by an outline of the methodologies employed. Lastly, an illustrative example is presented to showcase the practical application and effectiveness of the proposed approach.

I used mostly the software R version 4.2.2 with R-Studio 2023.03.1.

3.1 Data

In this study, I collected data on the 50 largest banks in Europe based on their assets. However, three banks were excluded from the analysis: two from Russia and one from Turkey. The decision to exclude these banks was primarily based on the limitation of the sentiment analysis algorithm used, which does not support these languages.

The dataset is divided into two main groups: the first group comprises balance sheet data for the banks, while the second group consists of tweets gathered from Twitter for sentiment analysis.

Regarding the first group, I analyzed individually all the financial reports of each bank in the last period of 2022 (December/22). From these financial reports I collected the following data:

- Ticker (in case of banks listed)
- Headquarter (HQ)
- Total Assets and Total Liabilities
- Total Equity

- Tier 1 Equity (CET1)
- Interbank Assets and Interbank Liabilities

This data can be seen in **Table 1**:

Table 1: Balance Sheet Europe's largest 50 banks

Rank	Institution Name	Ticker	HQ	Total Assets	Total Liabilities	Equity	IB Assets	IB Liabilities	CET1
1	HSBC Holdings	HSBA	UK	2805.44	2620.05	185.38	6.90	63.10	112.81
2	BNP Paribas	BNP	FR	2666.38	2539.82	126.56	32.62	124.72	91.83
3	Crédit Agricole	AGRI	FR	2167.62	2094.14	73.48	567.64	284.17	40.62
4	Banco Santander	SAN	ES	1734.66	1637.07	97.59	63.69	80.34	73.39
5	Barclays	BARC	UK	1717.29	1638.72	78.58	11.36	22.67	53.18
6	Groupe BPCE	BPCE	FR	1531.13	1448.58	82.56	97.69	139.12	69.67
7	Société Générale	GLE	FR	1486.82	1414.04	72.78	66.90	132.99	48.64
8	Deutsche Bank Aktiengesellschaft	DBK	DE	1336.79	1264.46	72.33	14.39	7.20	48.10
9	UBS Group AG	UBSG	CH	1044.39	990.28	54.11	13.99	10.97	42.99
10	Lloyds Banking	LLOY	UK	995.90	941.98	53.91	12.06	8.24	36.15
11	Intesa Sanpaolo	ISP	IT	975.68	921.60	54.09	32.88	137.48	40.77
12	ING Groep N.V.	INGA	NL	967.82	917.40	50.41	35.10	56.63	41.97
13	Crédit Mutuel	MUT	FR	885.09	828.34	56.75	57.97	63.22	50.89
14	UniCredit	UCG	IT	857.77	796.12	61.65	57.80	131.34	51.44
15	NatWest	NWG	UK	816.90	775.50	41.41	8.10	23.19	28.35
16	Standard Chartered	STAN	UK	770.12	723.14	46.98	36.11	26.30	32.30
17	La Banque Postale	POST	FR	745.64	705.28	40.37	67.10	26.44	13.61
18	Banco Bilbao Vizcaya Argentaria	BBVA	ES	713.14	662.53	50.61	25.23	28.92	42.49
19	Coöperatieve Rabobank	RAB	NL	628.51	582.16	46.36	11.12	31.26	38.37
20	DZ BANK AG	DZ	DE	627.04	558.92	68.12	123.44	186.79	18.76
21	Nordea Bank Abp	NDA	FI	594.84	563.44	31.40	4.57	32.87	20.28
22	CaixaBank	CABK	ES	592.23	557.97	34.26	12.19	12.77	27.49
24	Danske Bank	DANSKE	DK	505.52	483.98	21.54	8.17	18.64	15.47
25	Commerzbank AG	CBK	DE	477.44	388.73	88.71	15.14	41.40	23.90
26	ABN AMRO Bank N.V.	ABN	NL	379.58	356.77	22.81	2.98	17.51	19.51
27	KBC Group NV	KBC	BE	355.87	335.06	20.81	29.42	35.67	16.82
28	Landesbank Baden-Württemberg	LAND	DE	324.17	278.62	45.55	81.28	84.08	13.53
29	Erste Group Bank	EBS	AT	323.86	298.56	25.30	18.44	28.82	20.44
30	Skandinaviska Enskilda Banken AB	SEB	SE	319.47	300.97	18.50	6.98	4.59	14.74
31	Svenska Handelsbanken	SHB	SE	312.31	294.59	17.73	0.85	7.39	14.34
33	Nationwide Building Society	NBS	UK	308.46	289.28	19.18	3.24	28.43	15.58
34	DNB Bank ASA	DNB	NO	295.09	271.47	23.62	1.87	16.17	17.70
35	Raiffeisen Gruppe	RAIF	CH	282.61	261.84	20.77	2.21	14.09	20.72
36	Bayerische Landesbank	BAY	DE	259.30	230.88	28.42	61.44	60.96	11.36
37	Swedbank AB	SWED	SE	258.16	242.24	15.92	5.12	6.59	13.03
38	Banco de Sabadell	SAB	ES	251.38	238.16	13.22	4.70	11.37	9.99
39	Nykredit	NYK	DK	214.96	201.95	13.01	1.56	3.74	11.83
40	Raiffeisen Bank International AG	RBI	AT	207.06	188.29	18.76	15.60	33.61	15.64

Rank	Institution Name	Ticker	HQ	Total Assets	Total Liabilities	Equity	IB Assets	IB Liabilities	CET1
41	Zürcher Kantonalbank	KANT	CH	201.20	187.81	13.39	2.96	39.33	12.88
42	Banco BPM	BAMI	IT	189.69	176.29	13.39	5.49	32.64	8.62
43	Belfius Bank	BEL	BE	179.47	167.85	11.62	4.14	1.87	10.72
44	OP Financial Group	OP	FI	175.52	161.18	14.34	0.80	12.30	12.57
45	BPER Banca	BPE	IT	152.30	144.42	7.88	9.48	22.00	6.61
46	Bank of Ireland Group	BIRG	IE	151.32	139.39	11.93	3.18	3.68	7.54
47	AIB Group plc	A5G	IE	129.75	117.49	12.26	1.50	0.23	9.00
49	Banca Monte dei Paschi di Siena	BMPS	IT	120.17	112.14	8.03	3.26	21.38	7.60
50	Norddeutsche Landesbank Girozentrale	NORD	DE	109.33	103.03	6.30	13.11	28.66	5.68

NOTES: 1) all values are in Billions of Euros. 2) Sberbank of Russia, VTB Bank from Russia, and Türkiye Cumhuriyeti Ziraat Bankasi A.S. from Turkey are removed due to the different alphabet used in their mother language which is not supported by the sentimental analysys algorithm used. These banks would be the 23rd, 32th, and 48th in the rank, respectively. 3) For the banks not listed in Stock Market, the column "Ticker" present an acronym. 4) Deutsche Bank Aktiengesellschaft does not specify the amount of interbank assets and liabilities, just the balance. Therefore, I assumed this value as liabilities and twice it as assets (to match the balance, since it was positive value in the assets). 5) Danske Bank A/S does not inform his CET1, just total equity. Therefore, I used the average ratio CET1/Equity (71.81%) of the other banks to estimate it.

Regarding the descriptive statistics of this data, there are banks representing 14 different countries, with the majority of them being British (6), French (6), or German (6). In the other hand, Norway only has one bank among the largest 50.

It is important to emphasize that the total amount of assets, which sums up for all 47 banks, is 33,145.22 billion Euros. The largest bank, HSBA (HSBC Holdings), has 2,805.44 billion Euros, accounting for 8.5% of the total. On the other hand, the smallest bank, NORD (Norddeutsch Landesbank Girozentrale), has 109.33 billion Euros, representing 0.32% of the total.

Another significant point to highlight is the interbank liabilities, which are used later to establish the interconnections among banks. The ranking of the largest banks based on interbank liabilities would be quite different from their ranking based on assets. The top three banks in terms of interbank liabilities would be AGRI (Crédit Agricole), DZ (DZ Bank), and BPCE (Groupe BPCE). Their rankings based on assets were 3rd, 20th, and 6th, respectively.

With respect to the Twitter data, I constructed a query in order to collect the relevant tweets for each one of the banks. I retrieved the tweets that commented the name of bank (or how it is know for in Twitter) or yet the Ticker for those banks listed.

Together with that, I filter for tweets containing a word belonging to a financial taxonomy that was developed by Cerchiello et al. (2017). This financial taxonomy was developed by the authors based on their knowledge of which balance sheet information may affect financial risk.

To ensure comprehensive coverage, this financial taxonomy was translaed to the local language of each bank's headquarters. Therefore, I conducted searches for tweets in both English and the local language of each bank's headquarters. This allowed for a broader range of tweets to be included.

Finally, whenever possible, I prioritized tweets with engagement, such as likes, retweets, or comments, as they indicate a higher level of interaction with the content.

In total I collected over 50 thousand tweets in the period between 01/01/2022 to 31/12/2022.

For more details on the query and how this data was collected, it is possible to check the code in the Appendix (Part I - Search Query used to retrieve data from Twitter).

In the sequence, I detail the methodologies used in this work.

3.2 Direct Financial Network: interbank connections trough Balance Sheet

Typically, we have information on the assets and liabilities of each bank in the interbank market, but not on the specific interconnections between banks. In such cases, estimating the interbank network is necessary when performing contagion simulations or evaluating other risk metrics.

To estimate the interbank connections among the financial institutions, I followed the methodology proposed by Anand et al. (2015). The authors state that establishing network linkages represent a costs for banks. Contrary to the Maximum Entropy approach (ME) proposed by Upper and Worms (2004), in this methodology banks do not spread their borrowing and lending across the entire system, since the costs in terms of information processing, risk management and creditworthiness checks would be prohibitive for all but the largest banks. For this reason, the authors propose a Minimum Density (MD) approach to create interbank networks.

Minimizing the total number of linkages necessary for allocating interbank positions results to be consistent with total lending and borrowing observed for each bank. Formally, let c represent the fixed cost of establishing a link, A_i the assets of bank i and L_i the liabilities. Then the MD approach can be formulated as a constrained optimization problem for the matrix Z which contains the interbank assets and liabilities of the n banks. As follows:

$$\begin{aligned} \min_{z} = c \sum_{i=1}^{N} \sum_{j=1}^{N} 1_{[Z_{ij}>0]} \qquad s.t. \\ \sum_{j=1}^{N} Z_{ij} = A_i \qquad \forall i = 1, 2, ..., N \\ \sum_{i=1}^{N} Z_{ij} = L_j \qquad \forall j = 1, 2, ..., N \\ Z_{ij} \geq 0 \qquad \forall i, j \end{aligned} \tag{1}$$

where the indicator function 1 equals one only if bank i lends to bank j. The constraints can be softened by assigning penalties for deviations from the marginals (i.e., A_i and L_j) as follows:

$$\begin{aligned} AD_i = \left(A_i - \sum_{j} Z_{ij} \right), \\ LD_i = \left(L_i - \sum_{j} Z_{ji} \right) \end{aligned} \tag{2}$$

where LD_i measures bank i's current deficit; i.e. how much its bilateral borrowing falls short of the total amount it needs to raise, L_i, which is also the amount to be matched by the solution being constructed, Z. Substituting into objective function one gets the value function for the optimization program:

$$V(Z) = -c \sum_{i=1}^{N} \sum_{j=1}^{N} 1_{[Z_{ij>0}]} - \sum_{i=1}^{N} [\alpha_i AD_i^2 + \gamma_i LD_i^2] \tag{3}$$

The code used to to implement this method can be seen in the Appendix (Part III - R code for constructing the Financial Networks)

3.3 Financial Contagion

For simulating the interbank contagion, I will follow the works of Battiston et al. (2012a) and Mistrulli (2011). In the simulation of interbank financial contagion in Mistrulli (2011) all banks raising funds in the interbank market are allowed to fail one at a time.

The next step involves calculating the losses incurred by banks that have provided loans to the failed bank. If these losses exceed the lenders' tier-1 capital, which includes capital and reserves, the lenders would also experience a default. The simulation then proceeds by checking if the banks that failed in the initial iteration lead to the failure of other banks. In each subsequent iteration, the banks that failed in the previous iteration are removed from the set of banks that could be affected by contagion. The simulation continues until at least one bank defaults.

Formally, let B represent the set of banks and X_{ij} the funds that bank $j \in B$ borrows from bank $i \in B$, where $X_{ij} \geq 0 \quad \forall (i,j) \in BxB$ and $X_{ii} = 0 \quad \forall i \in B$. Additionally, let $c_i > 0$ represent the initial Tier 1 capital endowment of bank i, and $\alpha \in [0,1]$ the rate of loss (i.e. the incidence of losses due to contagion in the interbank exposure). Finally, let $z \in B$ denote the first bank that defaults because of some idiosyncratic shock, and define $D_z^n \subseteq B$ and $S_z^n \subseteq B$ as the set of banks that default and survive, respectively, at the nth step of the contagion path initiated by bank z, as follows:

$$\begin{aligned} D_z^n &= \{k \in B : c_{k,z}^n \leq 0 | c_{k,z}^{n-1} > 0\} \\ S_z^n &= \{k \in B : c_{k,z}^n > 0\} \qquad \forall n \geq 1 \end{aligned} \tag{4}$$

Where $c_{k,z}^n$, which is the capital of bank k at the nth step of the contagion initiated by bank z, is equal to:

$$c_{k,z}^n = c_{k,z}^{n-1} - \alpha \sum_{j \in D_z^{n-1}} X_{kj}, \qquad \forall n \geq 1 \textit{ and } \forall k \neq z \tag{5}$$

The contagion path is represented by the following:

$$\begin{array}{ll} \textit{Failed banks} & \textit{Surviving banks} \\ D_Z^0 = \{Z\} & S_Z^0 = \{k \in B \backslash \{Z\}\} \\ D_Z^1 = \{k \in S_Z^0 : c_{k,z}^1 = (c_k - \alpha X_{kz}) \leq 0\} & S_Z^1 = \{k \in S_Z^0 : c_{k,z}^1 = (c_k - \alpha X_{kz}) > 0\} \\ D_Z^2 = \{k \in S_Z^1 : c_{k,z}^2 = (c_{k,z}^1 - \alpha \sum_{j \in D_z^1} X_{kj}) \leq 0\} & S_Z^2 = \{k \in S_Z^1 : c_{k,z}^2 = (c_{k,z}^1 - \alpha \sum_{j \in D_z^1} X_{kj}) > 0\} \\ \ldots & \ldots \\ D_Z^n = \{k \in S_Z^{n-1} : c_{k,z}^n = (c_{k,z}^{n-1} - \alpha \sum_{j \in D_z^{n-1}} X_{kj}) \leq 0\} & S_Z^n = \{k \in S_Z^{n-1} : c_{k,z}^n = (c_{k,z}^{n-1} - \alpha \sum_{j \in D_z^{n-1}} X_{kj}) > 0\} \\ \ldots & \ldots \\ D_Z^N = \{\emptyset\} & S_Z^N = \{k \in S_Z^{N-1} : c_{k,z}^N = (c_{k,z}^{N-1} - \alpha \sum_{j \in D_z^{N-1}} X_{kj}) > 0\} \end{array} \tag{6}$$

and the process stops after N iterations when no additional default occurs.

Based on the literature in Financial Contagion, I have made some important assumptions. Firstly, in line with Battiston et al. (2012a) I have adopted their rationale which assumes that agents are unable to recover during the duration of the cascade. Consequently, I will assume that the rate of loss (α) is equal to one. This choice is motivated by the fact that I am conducting a stress test and aiming to measure a worst-case scenario situation.

Also based on Battiston (2012a), another assumption I will make is that agents lack certainty regarding the robustness level of their counterparties, represented by the Tier 1 capital. Consequently, they are unable to anticipate whether a significant default cascade will occur or not.

Finally, one crucial assumption emphasized in numerous papers that conducted financial contagion simulations, such as Battiston et al. (2012a), Upper and Worms (2004), and Mistrulli (2015), pertains to the structure of banks once the simulations commence. Accordingly, I assume that the occurrence of bank failure triggering contagion is unexpected, which implies that the pattern of interbank lending remains constant throughout the entire contagion process. While this assumption may initially appear strict, it acknowledges the fact that domino effects may take place over a very short period of time, where banks have minimal opportunities for adjustments or strategic maneuvers.

Lastly, as highlighted by Mistrulli (2011), the contagion mechanism used in these works focuses on a specific channel for contagion while disregarding other potential sources that may interact with the propagation of contagion within the interbank market. This precise aspect forms the crux of my research. By incorporating the possibility of bank runs as an amplification mechanism in financial contagion, I aim to make a contribution to the existing literature.

The code used to to implement this method can be seen in the Appendix (Part IV - R code for simulating Financial Contagion).

3.4 Sentiment Network: unstructured big data from Twitter

In order to construct the Sentiment Network, I initially performed sentiment analysis on each of the retrieved tweets. Subsequently, I utilized the sentiment scores assigned to the banks to build the Sentiment Network.

3.4.1 Measuring the sentiment

Once the tweets were collected as stated in the "Data" section, each obtained tweet has been classified into a sentiment class measure. The higher the measure, the more positive the sentiment (or value) that the tweet assigns to the bank under analysis.

The sentiment classification has been carried out employing the NRC Emotion Lexicon. The NRC (National Research Council) lexicon consists of approximately 14,000 words or terms and is a popular resource used in sentiment analysis tasks. It is a sentiment lexicon that consists of a list of words and their associated sentiment scores. Each word in the lexicon is tagged with one or more emotions or sentiment categories, such as joy, sadness, anger, fear, trust, anticipation, surprise, and disgust.

Additionally, for this work I have used the extend multi-language version of the vocabulary as I need to measure the sentiment in the local language of each bank.

The procedure for measuring the sentiment of each tweet involves several steps. Firstly, I utilized Google's Compact Language Detector to determine the language used in each tweet. Although Twitter has its own language detection algorithm, I observed a significant number of mismatches, which were notably reduced by employing Google's algorithm.

After language identification, the next step involves using the openNLP sentence tokenizer for tokenization. Tokenization plays a crucial role in sentiment analysis as it breaks down a sentence into individual tokens. This process enables a more detailed analysis and understanding of sentiment by taking into account factors such as negation, sentiment modifiers, context, and noise removal.

For instance, in the sentence "I do not like this product," tokenization ensures the accurate interpretation of the negative sentiment by recognizing the presence of "not." Similarly, in the sentence "The movie was extremely captivating," tokenization identifies the sentiment modifier "extremely," which intensifies the positive sentiment expressed. Furthermore, tokenization helps eliminate noise such as punctuation marks, allowing the focus to be on meaningful words.

In the sequence, I implemented a loop to analyze each bank and its corresponding Twitter data. For each tweet, the language used was identified and then provided as input to the NRC Emotion Lexicon algorithm. This step ensured that the sentiment analysis was conducted in the appropriate language, taking into account the specific linguistic nuances and context.

Finally, I computed the average sentiment for each bank by aggregating the sentiment measurements on a daily basis. This approach allowed me to obtain a consolidated sentiment value for each bank, which will be used in estimating the sentiment network.

The details of the procedure can be seen in the Appendix (Part II - R code for treating the unstructured big data retrieved from Twitter)

3.4.2 Construct the Sentimental Network to connect the banks

Accordingly to Fan et al. (2021), Cerchiello et al. (2017) and Cerchiello and Giudici (2016), Graphical Gaussian Models (GGM) are well suited to estimate interconnections between a large set of financial institutions.

In Cerchiello et al. (2017) the authors use GGM to estimate the relationships between banks. To do so they begin by estimating the relationship among banks through their partial correlation following the work of Lauritzen (1996). Partial correlations can be estimated assuming that the observations follow a GGM, in which the covariance matrix (Σ) is constrained by the conditional independences described by a graph.

More formally, let $X = (X_1, ..., X_N) \in R^N$ be a N-dimensional random vector distributed according to a multivariate normal distribution $\mathcal{N}(\mu, \sigma)$ and, assuming stationarity, $\mu = 0$. In addition, the authors assume throughout that the covariance matrix Σ is not singular.

Let $G = (V, E)$ be an undirected graph, with vertex set $V = 1, ..., N$, and edge set $E = VxV$, a binary matrix, with elements e_{ij}, that describe whether pairs of vertices are (symmetrically) linked between each other ($e_{ij} = 1$), or not ($e_{ij} = 0$).

If the vertices V of this graph are put in correspondence with the random variables $X_1, ..., X_N$, the edge set E induces conditional independence on X through the Markov properties as stated by Lauritzen (1996). In particular, the pairwise Markov property determined by G states that, for all $1 \leq i < j \leq N$:

$$e_{ij} = 0 \Leftrightarrow X_i \perp X_j | X_{V \setminus \{i,j\}} \tag{7}$$

that is, the absence of an edge between vertices i and j is equivalent to independence between the random variables X_i and X_j, conditionally on all other variables $x_{v \setminus \{i,j\}}$.

Whittaker (1990) proved that this independence corresponds to a partial correlation of zero i.e.:

$$X_i \perp X_j | X_{V \setminus \{i,j\}} \Leftrightarrow \rho_{ijV} = 0 \tag{8}$$

where

$$\rho_{ijV} = \frac{-\theta^{ij}}{\sqrt{\theta^{ii}\theta^{jj}}} \tag{9}$$

denotes the ijth partial correlation, i.e., the correlation between X_i and X_j conditionally on the remaining variables $X_{v \setminus \{i,j\}}$.

Therefore, ρ_{ij} expresses the direct influence of a financial institution on another, which can be used to measure the relationships between institutions.

If the value of θ_{ij} is zero (not zero), an edge does not exist (does exist) between vertex i and vertex j, and X_i and X_j are (are not) conditional independent.

Accordingly to Fan et al. (2021) constructing a visual connectedness network of financial institutions can be enhanced by considering partial correlations. When measuring the correlation of network nodes, it is important to differentiate between direct and indirect correlations. Simple correlation coefficients fail to capture this distinction, while partial correlation coefficients excel at quantifying direct correlations. By excluding the influence of other variables, the partial correlation coefficient specifically analyzes the correlation between two variables, providing a more accurate measure of direct correlation.

Therefore the GGM can be defined as the family of all N-variate normal distributions that satisfies the constraints induced by the graph $G = (V, E)$ on the partial correlation, as follows:

$$e_{ij} = 0 \Leftrightarrow \rho_{ijV} = 0, \qquad \forall \quad 1 \leq i < j \leq N \tag{10}$$

Moreover, let Θ be the inverse of the variance-covariance matrix Σ, i.e., Σ^{-1}. Θ is also referred to as a precision matrix and θ_{ij} are its elements.

According to the above analysis, the precision matrix contains the structural information of the graph model, so the structural learning of the Graphical Gaussian model can be transformed into the problem of calculate the precision matrix.

Let S be the covariance matrix of samples, that is, $S = \Sigma_{n=1}^{N} X^{(n)} X^{(n)T}/N$. The maximum likelihood estimation problem of the precision matrix can be expressed as:

$$\max_{\Theta>0} log|\Theta| - tr(S\Theta) \tag{11}$$

where $\Theta > 0$ denotes that all the elements of the matrix Θ are positive; $tr(\cdot)$ indicate trace of the matrix; and $|\Theta|$ denotes determinant of Θ.

However, the resulting model is too complex to be interpreted in reality. This occurs because the maximum likelihood estimation method cannot generate sparse solutions. In order to solve this problem, following Fan et al. (2021) and Banulescu & Dumitrescu (2015) a Lasso approach can be used by applying L_1-norm penalty to:

$$\hat{\Theta} = arg \max_{\Theta>0} log|\Theta| - tr(S\Theta) - \lambda||\Theta||_1 \tag{12}$$

where $\lambda \geq 0$; Θ is a positive definite matrix; $||\Theta||_1 = \sum_{j=1} \sum_{i=1} |\Theta_{ij}|$ is L_1-norm penalty for Θ. Since the L_1-norm penalty is used, the solution of Eq. (12), i.e., $\hat{\Theta}$, will be sparse. In another words, the proportion of zero elements in the precision matrix $\hat{\Theta}$ is larger. These zero elements represent the conditional independence between X_i and X_j and the absence of an edge between vertices i and j.

The remaining issue pertains to determining the optimal value for λ since different values can yield varying partial correlations and, consequently, different connections in the Sentiment Network. While some authors, such as Cerchiello et al. (2017), argue that the choice of λ is arbitrary, the literature does provide some methods for obtaining the best value for λ.

Cross-validation is a widely used technique for model selection and hyperparameter tuning. In the case of Lasso, cross-validation can help determine the optimal λ value by assessing the model's performance on different subsets of the data.

The process involves dividing the available data into multiple subsets or folds. The model is trained on a combination of these folds and evaluated on the remaining fold. This process is repeated several times, with different fold combinations, and the performance metric (such as mean squared error or accuracy) is averaged across all iterations. By testing various λ values and selecting the one that results in the best performance metric, the cross-validation approach helps identify the optimal λ for the Lasso method, which is the approach I used in this work.

The details for this procedures can be seen in the Appendix (Part III - R code for constructing the Financial Networks)

In summary, two or more banks will be connected when users exhibit similar sentiment towards these banks. In other words, the network connections are formed based on the similarity of sentiment expressed by users regarding the banks.

The main assumption underlying this approach is that if bank runs occur, they are more likely to spread among banks that have established connections in the sentiment expressed by social media users. This assumption suggests that the spread of a bank run is not random but rather influenced by the interconnected sentiment dynamics among banks.

3.5 Financial Contagion with Bank Runs effects

To incorporate the effect of bank runs in the interbank financial contagion approach, I propose a methodology that can be summarized in the following steps:

- for each bank, analyze the other banks that are sentiment connected to it
- check if the banks sentiment connected were not already contagioned by the interbank simulation. In case negative, allow these banks also to fail due to the bank run effect
- check the interbank connections from these new banks that are assumed to fail
- check if there are no duplicate cases (banks that have already been considered fail due to the interbank contagion)
- compute the number of additional banks that would fail in this approach
- compute the additional total assets from these banks
- compute the difference from this simulation from the simulation of only interbank contagion

Therefore, following these steps on simulations for each bank, I can compute the additional effects of considering bank runs in the simulation, both in number of banks and in total assets lost.

The details of this procedure can be seen in the Appendix (Part IV - R code for simulating Financial Contagion)

3.6 An illustration

In this section I provide a simple example to helps illustrate the model and the methodologies proposed. Consider **Figure 1**, where five stylized banks are represented. In this picture it is showed banks A, B, C, D, and E and theirs respective assets, capital (tier 1 capital) and interbank liabilities. This values are (100,10,30), (150,20,20), (120,12,0), (80,8,6), and (120,12,20) respectively.

To simplify the example, let's assume that bank A's interbank liabilities are evenly distributed among banks B, C, and D. It's important to note that this simplified scenario does not precisely reflect the behavior of the MD method. Instead, it is more similar to the ME method. However, for the purpose of this illustration, treating it in a similar manner would yield comparable results and make the analysis easier.

To consider the simulation of financial contagion in this approach, I first simulate that Bank A fails. In this case, I will consider that this bank will not pay the whole liabilities to the banks B, C, and D. What we have in this case is that Banks B and C, for having more financial robustness (they have capital of 20 and 12 respectively) could absorb this default of bank A and will not fail. However, bank D which has only 8 of capital would fail.

After that, the approach is to go for a second round where I need to check the impacts of the interbank liabilities of bank D, which has fail in the first round. Again for simplicity I will consider that bank D also split its interbank liabilities equally among banks E and C. Bank E, as it happened to banks B and C in the first round, has enough capital to absorb this loss and will not fail. However, now when we look to bank C, this bank would also fail considering that it is necessary to keep track of the defaults in the previous round. Therefore, bank C would be short of 13 (10 from default of Bank A in the first round plus 3 in the default of Bank D in the second round) and would also fail.

The simulation continues for another round with the analyse of the interbank liabilities of bank C now. However, since this bank has 0 of interbank liabilities the simulation would stop here.

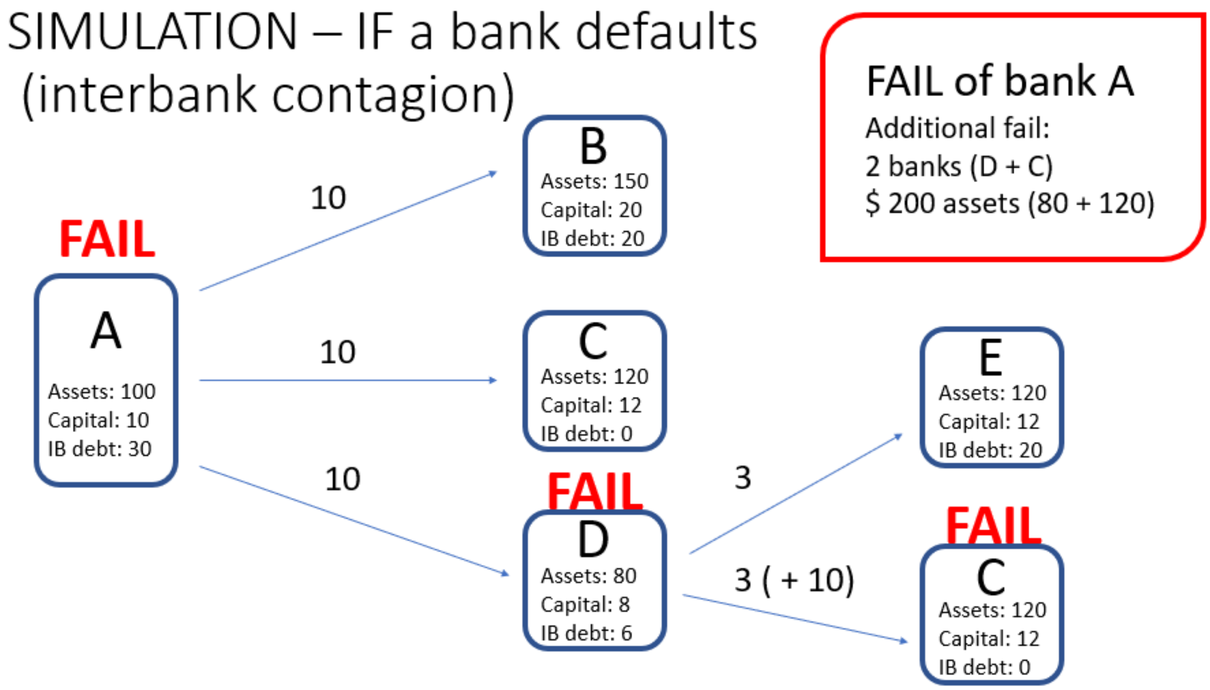

Figure 1: Simulation interbank contagion

In this case, the default of Bank A would cause the additional default of other 2 banks (D and C) and the additional amount of assets lost in this simulation would be 200 (80 from bank D and 120 from bank C).

This examples follows closely the works of Acenad et al. (2015) and Battiston et al. (2012a).

In **Figure 2** it is possible observe the representation of the network connecting the illustrated banks.

Now in my expansion, I will add the effects of a bank run in the simulation.

Consider **Figure 3**. Initially, it presents the same example as depicted in **Figure 1**, illustrating the interbank market contagion. However, in this scenario, I introduce the additional effects of bank run contagion.

In this case, Bank A is "sentiment connected" to Bank E. Therefore, when simulating the default of Bank A, I will also consider that, in addition to the interbank contagion effects explained earlier, Bank E would also fail due to a bank run contagion originating from Bank A. This is because in a simulation where Bank A experiences a bank run event and fails, it is likely that Bank E would also be affected by a bank run due to their sentiment connection.

Finally, in this case, I have also to analyze for the effects of interbank connections of bank E. Considering that, Bank G would also fail once it would not receive 10 from Bank E and its capital is only 8.

Therefore, now it is possible to analyze the effects of the default simulation of Bank A and compare the results in both cases, i.e., considering just the contagion in the interbank market and considering also the additional effect of bank runs.

As before we had that the simulation of the default in Bank A would cause the additional defaults of other 2 banks and 200 in assets. Now, considering the effects of a bank run, the effect of a simulated default in

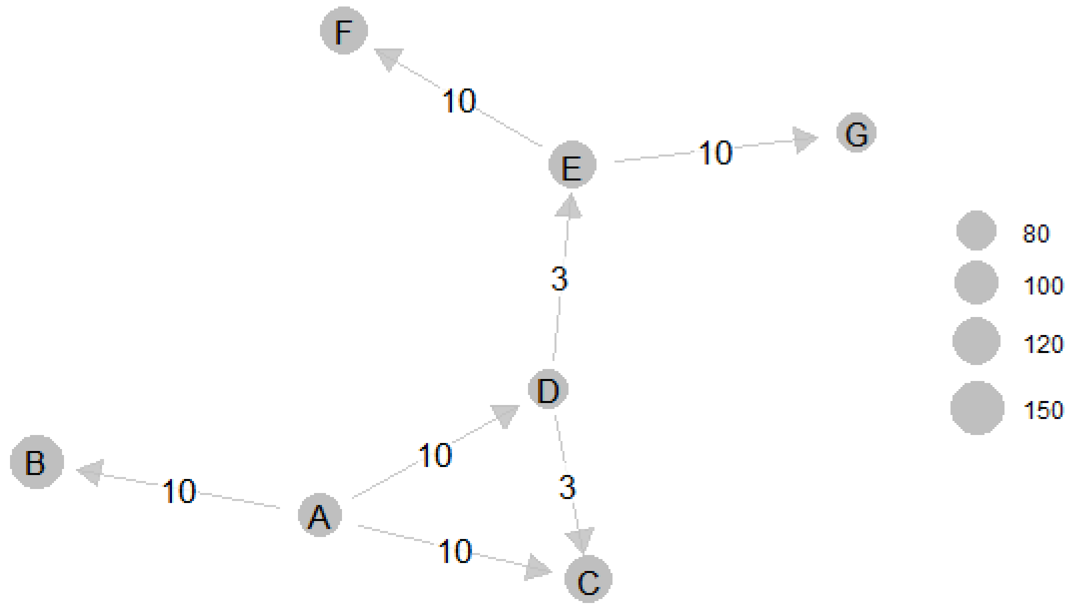

Figure 2: Interbank network with the illustration data

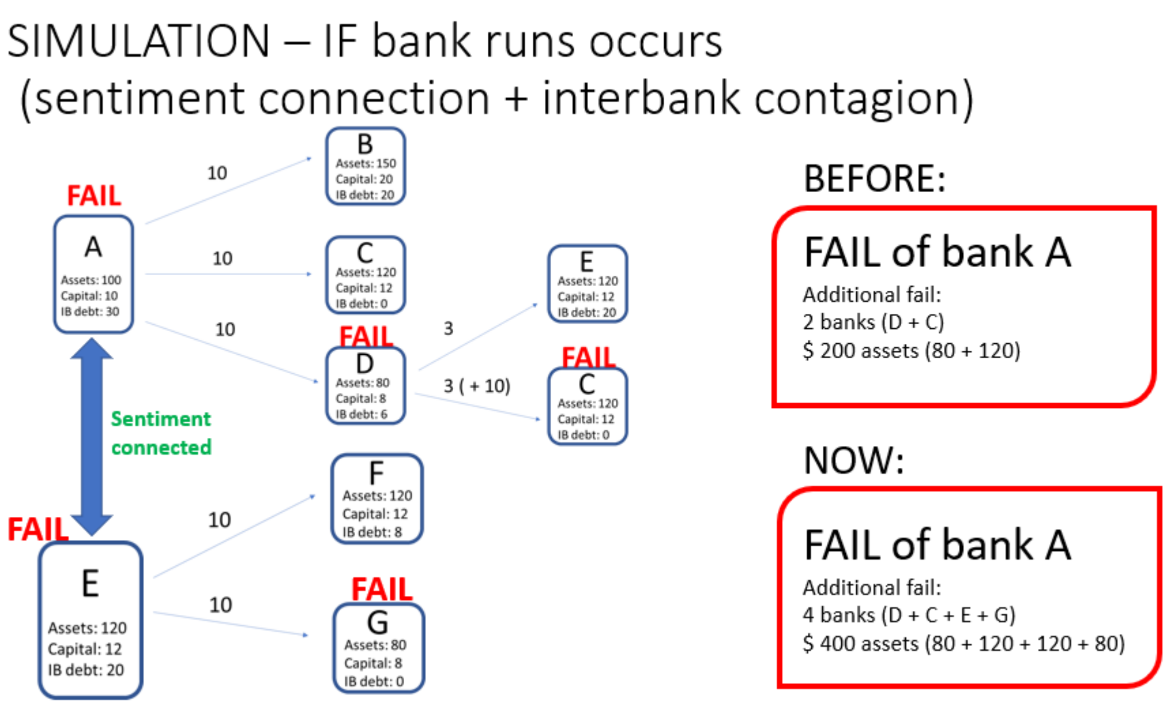

Figure 3: Simulation interbank contagion with bank run effects

Bank A would cause the additional default of 4 banks (D, C, E, and F) and the additional loss of 400 in assets.

With this methodology it is possible to quantify the additional impacts that a bank run could cause in the system as well as access which banks would spread more default in the system (more central and systemically important).

Finally, it is crucial to emphasize that Sentiment Analysis is a highly dynamic process. As a matter of fact, these metrics could be computed in virtually real-time basis, providing a significant advantage over certain conventional indicators.

Let's consider an example: suppose that today Bank Z experiences a bank run (for any given reason). Using this methodology, it would be possible to reconstruct the sentiment network connections within a few days, or even hours, and analyze the potential contagion effects. Moreover, it enables us to identify other banks that are more likely to be affected by the bank run through sentiment analysis.

This scenario becomes highly plausible when we have concrete information confirming that Bank Z indeed encountered a bank run. In such cases, if users who extensively comment on the bank run of Bank Z also display a similar pattern of behavior towards another bank, let's say Bank X, it is reasonable to assume that Bank X is susceptible to experiencing a bank run alongside Bank Z.

4 Results

In this section I present my results. I start by showing the results of the estimated Financial Network created for the simulation on the interbank connections within the financial institutions' data. After that, I present the Sentiment Network constructed using the unstructured big data retrieved from Twitter.

In the sequence, I show the results of the simulations and how the contagious effect affect the institutions. Finally, I compare the effects considering only interbank connections with those of my method that includes bank run effects.

Therefore, this section provide two main approaches to analyzing the financial contagion and possible bank runs effect: visually and quantitatively.

4.1 Financial Networks

In this section, I will present the networks related to interbank connections and the sentiment network, starting with the interbank connections. As explained in the methodology, it is often not possible to directly observe the actual network; instead, we only have access to the marginals (assets and liabilities). Consequently, it becomes necessary to estimate the adjacency matrix before conducting contagion simulations or calculating other network metrics. To address this, I employed the minimum density estimation method proposed by Anand et al. (2015).

This method considers the sparsity of interbank networks and aims to estimate a network with the fewest possible links that still satisfy a predetermined set of constraints. By doing so, it provides a more accurate representation of the real interbank network, thereby enhancing the precision of contagion simulations and other risk assessments.

Figure 4 illustrates the estimated interbank network.

In this figure, nodes are weighted based on the total assets of each bank, and the colors indicate the countries where the banks are headquartered.

The resulting network demonstrates sparsity and disassortativity. Sparsity refers to a network with relatively few connections compared to the total possible connections, resulting in few edges between nodes. Disassortativity, or disassortative mixing, describes a network where nodes tend to connect to nodes with different attributes or characteristics. Highly connected nodes are more likely to be connected to less connected nodes.

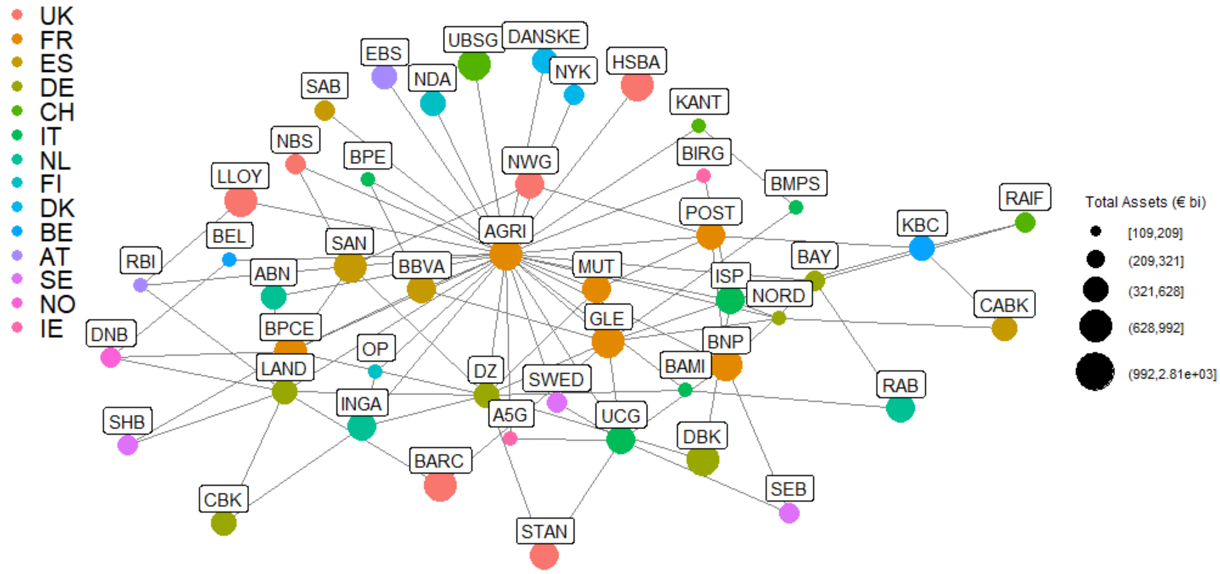

Figure 4: Financial Networks using Balance Sheet data

These characteristics are typical of a core-periphery network, which is commonly observed in interbank financial networks. In a core-periphery network, a densely connected core of nodes interacts strongly among themselves, while the periphery consists of nodes with fewer connections.

In fact, observing the network, it becomes apparent that AGRI appears to be the most central bank, which aligns with its high interbank asset value. Despite not being the largest in terms of total assets, AGRI's significant interbank assets contribute to its centrality within the network. On the other hand, banks such as NIK, BEL, SEB, and BIRG, with low interbank liabilities, occupy the "periphery" of the network. These banks have only a few connections, reflecting their relatively small debts, which are primarily directed towards larger banks like AGRI and BNP. These observations align with the assumptions of the ME methodology.

Moving forward, in the sequence I present the Sentiment Network constructed using the unstructured big data retrieved from Twitter.

This Network can be seen in **Figure 5**.

As in the previous figure, here nodes are weighted based on the total assets of each bank, and the colors indicate the countries where the banks are headquartered.

First of all, it is possible to noticed that the connections among banks in this case differ significantly from the interbank connections in the previous case. This was expected, as stated by Cerchiello et al. (2017), considering that the two networks aim to capture distinct information. The interbank network focuses on the connections between banks, while this network is constructed based on the sentiment expressed by social media users towards those banks.

This distinction is beneficial because if both networks captured similar issues, the simulation considering both would duplicate these connections, leading to an overestimate of contagion. Furthermore, it would not align with the overall objective of this work.

It is important to note that is is expected from the sentiment network to also capture non-fundamental factors related to social media users' sentiment, which are not included in the first network. Additionally, the comments in tweets cannot contain information about interbank connections since these connections are often unknown, even to central banks and regulators, as stated by Anand et al. (2015).

With that in mind, let's analyze the connections in **Figure 5**. We can observe certain connections that are to be expected, such as those between French banks POST and AGRI, and GLE and BNP. Similarly, the

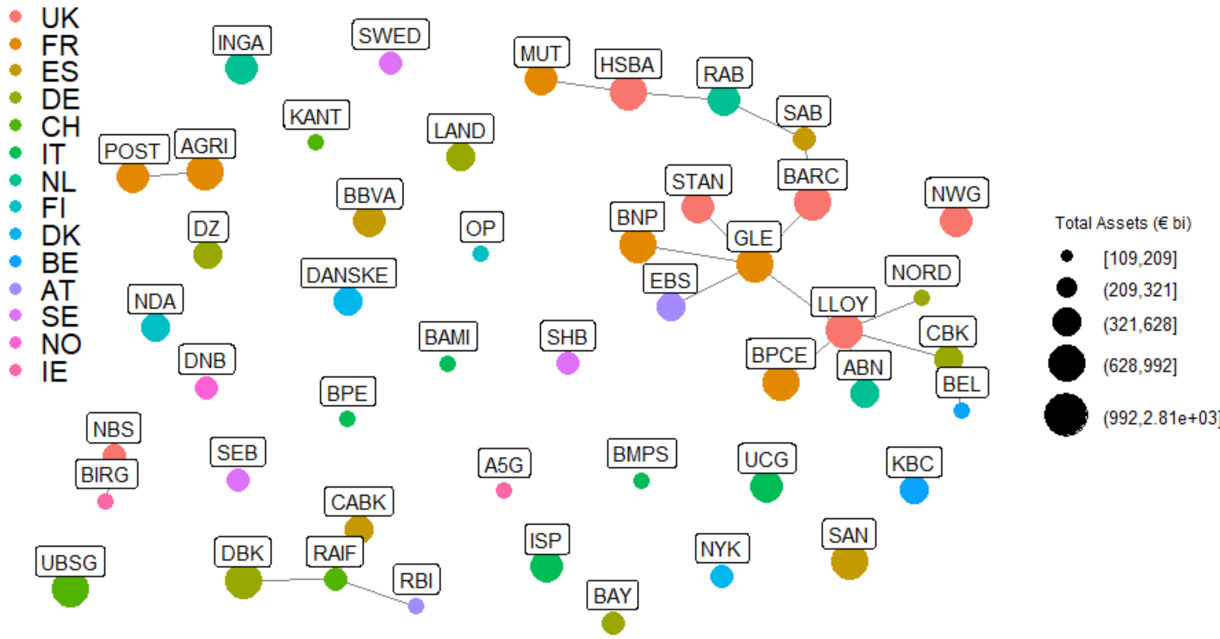

Figure 5: Sentiment Network using unstructured big data from Twitter

British BNS is connected to the Irish BIRG, and the German DBK is connected to the Swiss RAIF and the Austrian RBI.

However, it is important to note that connections are not necessarily limited to banks from the same country or language. These connections reflect the sentiment towards the institutions, regardless of language. Therefore, the presence of other connections is completely normal and somewhat expected.

The crucial aspect to emphasize is the dynamic nature of Sentiment Analysis. This process enables the computation of metrics in near real-time, as emphasized in the illustration chapter, enabling the identification of new connections as sentiments towards banks evolve.

By applying this methodology, it becomes feasible to reconstruct sentiment network connections within a short timeframe, ranging from a few days to mere hours. Such rapid analysis facilitates the examination of potential contagion effects. This situation becomes highly plausible when concrete information confirms the occurrence of a bank run at one bank. In such cases, if users extensively commenting on the bank run of this particular bank exhibit a similar behavioral pattern towards another bank, such as Bank X, it is reasonable to assume that Bank X is also susceptible to experiencing a bank run simultaneously.

4.2 Financial Contagion

After examining the visual aspects of the methodology's results, this section delves into the quantitative findings.

To begin with, the outcomes of the financial contagion simulations considering solely interbank connections are presented. **Table 2** displays the top 10 banks ranked by the potential contagion effect, measured by the total assets that would be impacted in the event of their failure.

Table 2: Rank of top 10 IB Financial Contagion (by assets default)

Bank	Institution Name	Rank	Add Banks Default	N	Add Assets Default	% Total
BNP	BNP Paribas	2	AGRI, BPCE, GLE, ISP, MUT, UCG, STAN, POST, DZ, KBC, LAND, EBS, BAY	13	11310.13	34.12
UCG	UniCredit	14	AGRI, BPCE, GLE, ISP, MUT, STAN, POST, DZ, KBC, LAND, EBS, BAY	12	10452.36	31.54
GLE	Société Générale	7	AGRI, BPCE, ISP, MUT, STAN, POST, DZ, KBC, LAND, EBS, BAY	11	8965.54	27.05
ISP	Intesa Sanpaolo	11	AGRI, BPCE, MUT, STAN, POST, DZ, KBC, LAND, EBS, BAY	10	7989.86	24.11
HSBA	HSBC Holdings	1	AGRI, BPCE, MUT, STAN, POST, DZ, LAND, EBS	8	7374.69	22.25
INGA	ING Groep N.V.	12	AGRI, BPCE, MUT, STAN, POST, DZ, LAND, EBS	8	7374.69	22.25
MUT	Crédit Mutuel	13	AGRI, BPCE, STAN, POST, DZ, LAND, EBS	7	6489.60	19.58
BPCE	Groupe BPCE	6	AGRI, MUT, STAN, POST, DZ, LAND, EBS	7	5843.55	17.63
AGRI	Crédit Agricole	3	BPCE, MUT, STAN, POST, DZ, LAND, EBS	7	5207.07	15.71
SAN	Banco Santander	4	STAN, DZ, LAND, RBI	4	1928.39	5.82

NOTES: 1) assets are in Billions of Euros. 2) For the banks not listed in Stock Market, the column "Bank" present an acronym, otherwise it is the Ticker.

The table highlights that BNP would have the most significant impact on the financial system if it were to fail. Its failure would lead to the failure of 13 other banks, resulting in a combined impact of 11,310 billion Euros. This amount represents over 34% of the total assets of all banks in the system.

Following BNP, the banks UCG, GLE, ISP, HSBA, and INGA are listed, and their failures would account for impacts ranging from around 30% to 20% of the system's total assets.

Interestingly, the spread of systemic contagion does not necessarily require banks to be extremely large in terms of total assets. A case in point is bank UCG, which ranks only 14th in terms of total assets. This effect may occur because banks can have significant interbank connections, leading to a domino effect. Furthermore, these banks may have substantial interbank liabilities despite their size. In fact, bank UCG would rank 6th if consider interbank liabilities.

Now turning to the results that incorporate the effects of bank runs in the simulations, I will analyze **Table 3**, which presents the top 10 banks also ranked by their total assets.

Table 3: Rank of top 10 Financial Contagion with Bank Run Effects
(by add assets default)

Bank	Institution Name	Rank	Sentiment Connected	Add Banks Default	Add Assets Default	% Total
GLE	Société Générale	7	BNP, BARC, LLOY, STAN, EBS	15	15202.88	45.87
LLOY	Lloyds Banking	10	BPCE, GLE, CBK, ABN, NORD	16	12386.52	37.37
BNP	BNP Paribas	2	GLE	13	11310.13	34.12
BARC	Barclays	5	GLE, SAB	13	10703.74	32.29

Bank	Institution Name	Rank	Sentiment Connected	Add Banks Default	Add Assets Default	% Total
RAB	Coöperatieve Rabobank	19	HSBA, SAB	11	10690.81	32.25
UCG	UniCredit	14		12	10452.36	31.54
EBS	Erste Group Bank	29	GLE	11	10128.50	30.56
SAB	Banco de Sabadell	38	BARC, RAB	11	9979.79	30.11
STAN	Standard Chartered	16	GLE	11	9682.24	29.21
RAIF	Raiffeisen Gruppe	35	DBK, CABK, RBI	12	9620.09	29.02

NOTES: 1) assets are in Billions of Euros. 2) For the banks not listed in Stock Market, the column "Bank" present an acronym, otherwise it is the Ticker.

In this scenario, it is possible to observe a significant change in the top-ranked banks that would cause the most substantial systemic contagion in the event of their failure.

While GLE and BNP maintain their positions in the top ranks, banks such as LLOY, BARC, and RAB now emerge as contributors to significant systemic contagion.

Additionally, the total amount lost in this scenario is larger, with GLE potentially accounting for up to 45% of total assets. The other banks, up to RAB, would fall within the range of up to 32% impact on total assets.

The change in rankings is primarily attributed to the fact that these banks are sentiment connected to other banks that would experience significant contagion in case of distress. As a result, the contagion effect is considerably amplified.

To gain a clearer understanding of this comparison, let's refer to **Table 4**, which provides a comprehensive overview of all the banks in both scenarios: counting only interbank connections and accounting for the amplification mechanism through bank runs.

Table 4: Comparison Financial Contagion with IB and with Bank Run Effects + IB (by total assets add)

Bank	Defaults IB	Assets IB	Defaults IB+BR	Assets IB+BR	Add Banks	Add Assets
LLOY	0	0.00	16	12386.52	16	12386.52
RAB	1	259.30	11	10690.81	10	10431.51
EBS	0	0.00	11	10128.50	11	10128.50
SAB	0	0.00	11	9979.79	11	9979.79
STAN	0	0.00	11	9682.24	11	9682.24
RAIF	0	0.00	12	9620.09	12	9620.09
BARC	3	1721.34	13	10703.74	10	8982.40
POST	0	0.00	7	6629.04	7	6629.04
GLE	11	8965.54	15	15202.88	4	6237.34
MUT	7	6489.60	8	9295.04	1	2805.44
CBK	0	0.00	2	1175.36	2	1175.36
ABN	0	0.00	1	995.90	1	995.90
NORD	0	0.00	1	995.90	1	995.90
BPCE	7	5843.55	8	6839.45	1	995.90
HSBA	8	7374.69	10	8262.50	2	887.81
BEL	0	0.00	1	477.44	1	477.44
BIRG	0	0.00	1	308.46	1	308.46
DBK	0	0.00	1	282.61	1	282.61
CABK	1	109.33	2	391.94	1	282.61
RBI	3	1721.34	4	2003.95	1	282.61

Bank	Defaults IB	Assets IB	Defaults IB+BR	Assets IB+BR	Add Banks	Add Assets
NBS	0	0.00	1	151.32	1	151.32
BNP	13	11310.13	13	11310.13	0	0.00
AGRI	7	5207.07	7	5207.07	0	0.00
SAN	4	1928.39	4	1928.39	0	0.00
UBSG	0	0.00	0	0.00	0	0.00
ISP	10	7989.86	10	7989.86	0	0.00
INGA	8	7374.69	8	7374.69	0	0.00
UCG	12	10452.36	12	10452.36	0	0.00
NWG	1	745.64	1	745.64	0	0.00
BBVA	1	152.30	1	152.30	0	0.00
DZ	1	770.12	1	770.12	0	0.00
NDA	0	0.00	0	0.00	0	0.00
DANSKE	0	0.00	0	0.00	0	0.00
KBC	1	745.64	1	745.64	0	0.00
LAND	2	1397.16	2	1397.16	0	0.00
SEB	0	0.00	0	0.00	0	0.00
SHB	0	0.00	0	0.00	0	0.00
DNB	0	0.00	0	0.00	0	0.00
BAY	0	0.00	0	0.00	0	0.00
SWED	0	0.00	0	0.00	0	0.00
NYK	0	0.00	0	0.00	0	0.00
KANT	0	0.00	0	0.00	0	0.00
BAMI	0	0.00	0	0.00	0	0.00
OP	0	0.00	0	0.00	0	0.00
BPE	0	0.00	0	0.00	0	0.00
A5G	0	0.00	0	0.00	0	0.00
BMPS	0	0.00	0	0.00	0	0.00

NOTES: 1) assets are in Billions of Euros. 2) For the banks not listed in Stock Market, the column "Bank" present an acronym, otherwise it is the Ticker.

This table is ranked based on the total assets that are added to the contagion when bank runs are included in the simulations, compared to considering only interbank contagion. In other words, it represents the difference between the contagion effects when just interbank contagion are considering and when both bank runs and interbank connections are taken into account.

In this table, it is evident that bank LLOY contributes the most to the system's defaults after considering the amplification caused by bank runs. The significance of bank LLOY's contribution stems from the fact that, initially, it would not cause any defaults based solely on interbank contagion. The same holds true for the other seven banks at the top of the table (excluding banks RAB and BARC, which would cause 1 and 3 defaults through interbank connections, respectively).

The total amount of assets that would be lost, considering only interbank connections, sums up to 80,558 billion euros. Comparing when bank runs are taken into consideration, this number increases to 174,276 billion euros, representing an 116.34% increase (over twice as much). Nevertheless, it's important to note that this measure purely holds for a hypothetical comparison among the simulation, as the simulation should be considered only individually for each bank.

Although there may be convergence in results for certain banks, such as bank GLE, it is important to emphasize that the inclusion of the proposed amplification mechanism leads to significant changes in the results. This observation highlights the criticality of integrating different measures, as it allows for a more comprehensive understanding. By combining multiple measures of systemic importance, one can obtain a more accurate and holistic perspective of the potential risks that financial contagion poses to the stability of the financial system.

In light of the results obtained, it is evident that banks that were not highly central in the network before now exhibit increased centrality. This finding is of particular importance and should attract the attention of regulators.

Comparing these results with existing literature, it is crucial to emphasize that the outcomes of this study lean towards a pessimistic perspective. However, it is essential to view these results as a stress test in a worst-case scenario approach, offering insights into the potential vulnerabilities of the financial system.

The outcomes suggest a high degree of vulnerability to contagion effects, especially when subjected to bank runs amplification channel. Such scenarios can initiate a chain reaction of defaults and amplify losses, leading to significant disruptions in the financial system and potentially affecting the broader economy.

It is important to note that these results should not be interpreted too literally or used as the sole determinant of a financial institution's systemic importance. The additional stress indicator, represented by the bank runs amplification, can serve as another measure of systemic importance. Therefore, it should be considered alongside other factors such as the institution's size, interconnectedness, and criticality to the overall functioning of the financial system.

Finally, the distinction between fundamental and non-fundamental networks is noteworthy. This observation emphasizes the significance of including networks that might capture non-fundamental aspects in the calculation of financial contagion and centrality for banks, as it provides valuable information that would otherwise be overlooked. Considering both types of networks can contribute to a more comprehensive understanding of the dynamics and potential risks within the financial system.

5 Conclusion

The results obtained from the analysis provide valuable insights into the potential effects of bank runs and sentiment-driven contagion in the financial system. The combination of interbank connections and sentiment networks allows for a more comprehensive understanding of systemic risk and its propagation through the system.

The estimated interbank network demonstrates a core-periphery structure, where a densely connected core of banks interacts strongly among themselves, while the periphery consists of banks with fewer connections. This finding aligns with previous studies on interbank networks and reflects the hierarchical nature of the financial system. The centrality of certain banks, such as AGRI, highlights their importance in the network due to their high interbank assets. On the other hand, banks with low interbank liabilities occupy the periphery of the network, indicating their relatively smaller debts directed towards larger banks.

The sentiment network constructed from Twitter data provides a different perspective on the connections between banks. It captures non-fundamental factors, reflecting users' sentiments towards banks regardless of their country or language. This network is dynamic in nature and allows for near real-time analysis, enabling the identification of new connections as sentiments evolve. The sentiment network provides additional information that complements the interbank network and contributes to a more holistic understanding of contagion dynamics.

The simulation results highlight the potential impact of bank failures on the financial system. In the case of interbank contagion without considering bank runs, BNP emerges as the bank with the most significant potential contagion effect, followed by UCG, GLE, ISP, and HSBA. However, when bank runs are included in the simulations, the ranking of banks experiencing significant contagion changes. Banks like LLOY, BARC, and RAB emerge as contributors to systemic contagion due to their sentiment connections with other banks.

The comparison between interbank contagion and contagion with bank runs amplification shows a substantial increase in the potential losses and defaults in the system when bank runs are considered. Banks that initially would not cause any defaults based solely on interbank contagion can become significant contributors to systemic risk when sentiment-driven bank runs are taken into account. The inclusion of the amplification mechanism through bank runs provides a more comprehensive perspective on systemic importance and potential vulnerabilities.

These results emphasize the importance of integrating multiple measures of systemic risk to gain a more accurate understanding. By considering both interbank connections and sentiment-driven bank runs, regulators can identify systemically important banks that may not be apparent solely based on traditional measures. The findings suggest a high degree of vulnerability to contagion effects and highlight the need for proactive monitoring and risk management strategies to maintain financial stability.

Nevertheless, it is important to interpret these results with caution and not view them as definitive determinants of a bank's systemic importance. The simulations conducted in this study represent a stress test scenario and provide insights into potential vulnerabilities rather than precise predictions. Further research and analysis are necessary to validate and refine the findings.

The study also has some limitations. The sentiment analysis based on Twitter data may not capture the entire spectrum of market sentiment, and the sentiment expressed on social media platforms may not always align with market realities. The estimation of the interbank network relies on the availability and quality of data, which may have limitations. Additionally, the simulations conducted in this study represent a simplified model of financial contagion and do not account for all possible factors and complexities of real-world scenarios. Finally, there are additional limitations to consider in this study. Firstly, the sentiment analysis algorithm used relies on a general vocabulary rather than a specialized finance vocabulary. This limitation arises from the lack of available algorithms that encompass multi-language capabilities with a specific finance focus, which was a necessary requirement for this research. Secondly, similar to many other papers in the field of financial contagion, the actual interbank connections remain unknown, necessitating the estimation of interbank relationships as undertaken in this study.

In conclusion, the research presented in this paper contributes to the existing literature on systemic risk management and financial contagion. By incorporating bank runs and sentiment-driven contagion into the analysis, the study offers a more comprehensive measurement framework for assessing systemic risk and understanding the dynamics between fundamental and non-fundamental factors. The findings highlight the potential vulnerabilities of the financial system and provide insights for regulators to enhance risk management and supervision practices. Further research, as outlined above, can continue to refine and expand our understanding of systemic risk, enhance the measurement and prediction of financial contagion, and explore the applicability of the suggested methodology in real-world cases of distress for result comparison.

References

Acharya, V. V., Pedersen, L. H., Philippon, T., & Richardson, M. (2017). Measuring systemic risk. *The review of financial studies*, 30(1), 2-47. DOI: https://doi.org/10.1093/rfs/hhw088

Allen, F., & Gale, D. (2000). Financial contagion. *Journal of political Economy*, 108(1), 1-33. DOI: https://doi.org/10.1086/262109

Anand, K., Craig, B. and G. von Peter (2015). Filling in the blanks: network structure and interbank contagion. *Quantitative Finance* 15:4, 625-636. DOI: https://doi.org/10.1080/14697688.2014.968195

Banulescu, G. D., & Dumitrescu, E. I. (2015). Which are the SIFIs? A Component Expected Shortfall approach to systemic risk. *Journal of Banking & Finance*, 50, 575-588. DOI: https://doi.org/10.1016/j.jbankfin.2014.01.037

Battiston, S., Gatti, D. D., Gallegati, M., Greenwald, B., & Stiglitz, J. E. (2012a). Default cascades: When does risk diversification increase stability?. *Journal of Financial Stability*, 8(3), 138-149. DOI: https://doi.org/10.1016/j.jfs.2012.01.002

Battiston, S., Puliga, M., Kaushik, R., Tasca, P., & Caldarelli, G. (2012b). DebtRank: Too Central to Fail? Financial Networks, the FED and Systemic Risk. *Scientific Reports*, 2, 541.

Billio, M., Getmansky, M., Lo, A. W., & Pelizzon, L. (2012). Econometric measures of connectedness and systemic risk in the finance and insurance sectors. *Journal of financial economics*, 104(3), 535-559. DOI: https://doi.org/10.1016/j.jfineco.2011.12.010

Bricco, M. J., & Xu, M. T. (2019). *Interconnectedness and contagion analysis: A practical framework.* International Monetary Fund. DOI: https://doi.org/10.5089/9781513516226.001

Calomiris, C. W., & Mason, J. (1997). *Contagion and bank failures during the Great Depression: The June 1932.* Chicago banking panic. http://www.jstor.org/stable/2951329

Cerchiello, P., & Giudici, P. (2016). Big data analysis for financial risk management. *Journal of Big Data*, 3(1), 1-12.

Cerchiello, P., Giudici, P., & Nicola, G. (2017). Twitter data models for bank risk contagion.*Neurocomputing*, 264, 50-56. DOI: https://doi.org/10.1016/j.neucom.2016.10.101

Cookson, J. A., Fox, C., Gil-Bazo, J., Imbet, J. F., & Schiller, C. (2023). Social media as a bank run catalyst. Available at SSRN: https://ssrn.com/abstract=4422754 or http://dx.doi.org/10.2139/ssrn.4422754

Diamond, D. W., & Dybvig, P. H. (1983). Bank runs, deposit insurance, and liquidity. *Journal of political economy*, 91(3), 401-419. https://www.jstor.org/stable/1837095

Elliott, M., Golub, B., & Jackson, M. O. (2014). Financial networks and contagion. American Economic Review, 104(10), 3115-3153. DOI: http://dx.doi.org/10.2139/ssrn.2175056

Fan, X., Wang, Y., & Wang, D. (2021). Network connectedness and China's systemic financial risk contagion——An analysis based on big data. *Pacific-Basin Finance Journa*l, 68, 101322. DOI: https://doi.org/10.1016/j.pacfin.2020.101322

Iyer, R., & Puri, M. (2012). Understanding bank runs: The importance of depositor-bank relationships and networks. *American Economic Review*, 102(4), 1414-1445.

Jackson, M. O., & Pernoud, A. (2021). Systemic risk in financial networks: A survey. Annual Review of Economics, 13, 171-202. DOI: http://dx.doi.org/10.2139/ssrn.3651864

Lauritzen, S. L. (1996). *Graphical models* (Vol. 17). Clarendon Press.

Mistrulli, P. E. (2011). Assessing financial contagion in the interbank market: Maximum entropy versus observed interbank lending patterns. *Journal of Banking & Finance*, 35(5), 1114-1127. DOI: https://doi.org/10.1016/j.jbankfin.2010.09.018

Nopp, C., & Hanbury, A. (2015, September). Detecting risks in the banking system by sentiment analysis. In Proceedings of the 2015 conference on empirical methods in natural language processing (pp. 591-600).

Nyman, R., Kapadia, S., & Tuckett, D. (2021). News and narratives in financial systems: exploiting big data for systemic risk assessment. *Journal of Economic Dynamics and Control*, 127, 104119. DOI: https://doi.org/10.1016/j.jedc.2021.104119

Rönnqvist, S., & Sarlin, P. (2015). Bank networks from text: interrelations, centrality and determinants. Quantitative Finance, 15(10), 1619-1635. DOI: https://doi.org/10.1080/14697688.2015.1071076

Shiller, R. J. (2019). Narrative economics: How stories go viral and drive major economic events. Princeton University Press.

Upper, C. (2011). Simulation methods to assess the danger of contagion in interbank markets. *Journal of financial stability*, 7(3), 111-125. DOI: https://doi.org/10.1016/j.jfs.2010.12.001

Upper, C. and A. Worms (2004). Estimating bilateral exposures in the German interbank market: Is there a danger of contagion? European Economic Review 48, 827-849. DOI: https://doi.org/10.1016/j.euroecorev.2003.12.009

Whittaker, J. (1990). Graphical Models in Applied Multivariate Statistics Wiley New York 1. Zbl0732, 62056.

Appendix

Part I - Search Query used to retrieve data from Twitter

```
# Search Query
# Description: query used to retrieve data from Twitter to each bank
# initial filters: financial vocabulary from Cerchiello et al. (2017); date in 2022;
# just in English; and tweets with engagement (retweets and/or likes)
# Author: Wagner Eduardo Schuster (2023/05)

# Examples:
## 1st try: with keywords, just in English, and with engagement
(bank OR @name_in_twitter OR $ticker) AND (commissions OR "labour costs" OR deposits OR
interbank OR management OR "interest margin" OR subsidiaries OR capital OR "loan losses"
OR loans) since:2022-01-01 until:2022-12-31 lang:en filter:has_engagement

## 2nd try: keywords also in the language of the bank
## Example 13: ING Groep N.V. - ENGLISH/DUTCH
(@ing_news OR "ing bank" OR "ing groep" OR $INGA.AS OR $NYSE:ING) AND
(commissions OR commissies OR "labour costs" OR "arbeidskosten" OR deposits OR
bankstortingen OR interbank OR interbancair OR management OR  "interest margin" OR
renteverschil OR subsidiaries OR dochterondernemingen OR capital OR kapitaal OR
"loan losses" OR "verliezen op leningen" OR loans OR leningen)
since:2022-01-01 until:2022-12-31 filter:has_engagement

## 3rd try: no keywords but more engagement
## Example: 31: Skandinaviska Enskilda Banken AB (publ)
(@sebgroup OR "seb group" OR "skandinaviska enskilda banken" OR $SEB.A)
since:2022-01-01 until:2022-12-31 filter:has_engagement (min_retweets:10 OR min_faves:10)
```

Part II - R code for treating the unstructured big data retrieved from Twitter

```r
# Twitter data
# Description: load data from twitter, check duplicates, detect language,
# perform sentimental analysis, take average by date and merge into a final data frame
# Author: Wagner E. Schuster (2023/05)

# set-up
rm(list=ls());gc()
setwd("~/QEM/BARCELONA/Thesis")

# list of banks (50 largest - 2 Russians and 1 Turkish)
banks_names <- list.files("Data/Twitter")
banks <- list()

# load Twitter data
## package readxl
library(readxl)
library(data.table)
```

```r
# loop
for (i in 1:length(banks_names)){
  ## load files
  tweet <- read_xlsx(paste0("Data/Twitter/",banks_names[i]))
  ## take just necessary columns (id, tweet text, and time)
  tweet <- tweet[,c(1,3,5)]
  ## remove possible duplicates by single column
  tweet <- tweet[!duplicated(tweet$`Tweet ID`), ]
  ## format date
  tweet$`Tweet Posted Time` <- as.Date(tweet$`Tweet Posted Time`)
  ## rename
  names(tweet) <- c("ID", "date", "tweet" )

  ## convert to data table (faster)
  tweet <- as.data.table(tweet)

  ## add to banks list
  banks[[i]] <- tweet
}

# include names for each bank (without de code and extension)
banks_names_clean <- sub('\\.xlsx$', '', substr(banks_names,4,100))
names(banks) <- banks_names_clean


# sentiment analysis
library(syuzhet)

# languages from NRC (from syuzhet)
languages <- c("english", "french", "italian", "spanish", "german", "dutch",
               "danish", "finnish", "swedish", "catalan", "portuguese")

# no norwegian, plus catalan and portuguese
# uses tokenization: implements the openNLP sentence tokenizer

# detect language of the tweet
# (the algorithm from twitter does not seem to work well)
# I will use detect language (Google's Compact Language Detector)
#install.packages("cld2")
library(cld2)

# loop to identify language in each tweet for each bank
for (i in 1:length(banks_names)){
  # identify language
  banks[[i]]$lang <- detect_language(text = banks[[i]]$tweet
                                     , plain_text = FALSE, lang_code = FALSE)
  # transform lower case
  banks[[i]]$lang <- tolower(banks[[i]]$lang)
  # keep only post on languages supported by NRC
  banks[[i]] <- banks[[i]][banks[[i]]$lang %in% languages,]
}
```

```r
# loop to perform sentiment analysis accordingly to the correct language
# data table to be faster
# (which still takes over 1 hour - without df it takes 1 hours for each bank)

# need double loop to look first each bank than each row accordingly to the lang
for (i in 1:length(banks_names)){
  for (row in 1:nrow(banks[[i]])){
    # compute sentiment accordingly to the language of each tweet
    banks[[i]][row, sent := get_sentiment(banks[[i]]$tweet[row], method="nrc",
                                    language = banks[[i]]$lang[row])]
 }
}


# construct final table
## defining start date
start_date <- as.Date("2022/01/01")
## defining end date
end_date <- as.Date("2022/12/31")
## generating range of dates
date <- seq(start_date, end_date,"days")
## final data frame
SENTIMENT <- as.data.table(date)

## loop to compute average sentiment by day and merge with final data frame
for (i in 1:length(banks_names)){
  # average by group
  agg <- aggregate(banks[[i]]$sent, list(banks[[i]]$date), FUN=mean)
  # rename 1st column to date
  names(agg)[1] <- "date"
  # merge
  SENTIMENT <- merge(SENTIMENT,agg,by="date",all.x=TRUE)
  # replace NA for the ZERO
  SENTIMENT$x[is.na(SENTIMENT$x)] <- mean(SENTIMENT$x, na.rm = TRUE)
  # rename for the correct name of the bank
  names(SENTIMENT)[i+1] <- banks_names_clean[i]
}

# save file
library(xlsx)
write.xlsx(SENTIMENT, "Data/Twitter_sentiment.xlsx")
```

Part III - R code for constructing the Financial Networks

```r
# Financial Networks
# Description: construct the Financial Networks using:
# 1) twitter sentiment; and 2) balance sheet
# Author: Wagner E. Schuster (2023/05)

#### setup ####
rm(list=ls());gc()
setwd("~/QEM/BARCELONA/Thesis")
```

```r
# load data
library(xlsx)
## Twitter sentiment
twitter_sent <- xlsx::read.xlsx("Data/Twitter_sentiment.xlsx",1)
### skip 1st column (noise)
twitter_sent <- twitter_sent[,-1]

## Balance Sheet
balance_sheet <- xlsx::read.xlsx("Data/BS_interbank.xlsx",1)
### take just important columns
balance_sheet <- balance_sheet[1:47,1:15]
### rename
names(balance_sheet)[c(2,3,5,6,10,11)] <- c("Institution Name",
                                            "Ticker",
                                            "Total Assets",
                                            "Total Liabilities",
                                            "IB Assets",
                                            "IB Liabilities")

#### twitter data ####

# Gaussian Graphical Model
## using package GLASSOO for compute partial autocorrelation
library(GLASSOO)
# using pachake CVglasso to define best lambda using information criteria
library(CVglasso)

## choose best lamda using information criteria of the cross validation test
## (maybe create a loop to compute many times the estimations)
### loglik
cv_result_ll <- CVglasso(as.matrix(twitter_sent[,-1])
                         , nlam=1000, crit.cv="loglik")
### AIC
cv_result_aic <- CVglasso(as.matrix(twitter_sent[,-1])
                          , nlam=1000, crit.cv="AIC")
### BIC
cv_result_bic <- CVglasso(as.matrix(twitter_sent[,-1])
                          , nlam=1000, crit.cv="BIC")
### decide best
mean(c(cv_result_ll$Tuning[2], cv_result_aic$Tuning[2], cv_result_bic$Tuning[2]))

# graphs of cross-validations
# layout(matrix(1:4,nrow = 2, byrow = TRUE))
# plot(cv_result_ll, main="loglik")
# plot(cv_result_aic, main="AIC")
# plot(cv_result_bic, main="BIC")

## best option: 0.15

# compute partial correlations using glasso penalty
glasso_twitter <- GLASSO(X = twitter_sent[,-1], lam=0.15)

# save precision matrix (omega)
```

```r
pcorr_twitter <- glasso_twitter$Omega

# compute partial correlations from the inverse of precision matrix (Theta)
library(corpcor)
pcorr_twitter <- round(cor2pcor(solve(pcorr_twitter)),6)

# rename (I will give a short for those without sticker)
colnames(pcorr_twitter) <- balance_sheet$Ticker

# networks
library(igraph)

# Make an Igraph object from this matrix:
network_twitter <- graph_from_adjacency_matrix(pcorr_twitter
                            , weighted=T, mode="undirected", diag=F)
# plot
library("qgraph")
library(ggplot2)
library(GGally)
library(intergraph)

# Create a color palette based on the number of unique countries
num_countries <- length(unique(balance_sheet$HQ))
color_palette <- scales::hue_pal()(num_countries)

# Map colors to countries
country_colors <- setNames(color_palette, unique(balance_sheet$HQ))
col <- country_colors[balance_sheet$HQ]


## weighted by assets (billions of Euros)
ggnet2(network_twitter, size = balance_sheet$`Total Assets`, label = TRUE
       ,label.size = 4, size.cut = 5, color = col)

# label outside
ggnet2(network_twitter, size = balance_sheet$`Total Assets`, label = FALSE
       ,label.size = 4, size.cut = 5, color = col) +
  theme(legend.text = element_text(size = 10)) +
  geom_label(aes(label = balance_sheet$Ticker),nudge_y = 0.05) +
  guides(size = guide_legend(
    title = expression(paste("Total Assets (", "\u20AC", " bi)")),
    override.aes = list(size = c(3, 5, 7, 9, 11))
  ))

## save
saveRDS(network_twitter, "Data/Networks/network_twitter.Rds")


#### balance sheet data ####
# Illustration
ExampleBS <- xlsx::read.xlsx("Data/Illustration.xlsx",1)
ExampleMD <- xlsx::read.xlsx("Data/Illustration.xlsx",2)
ExampleMD <- as.matrix(ExampleMD)
```

```r
## rownames and colnames for the matrix
rownames(ExampleMD) <- colnames(ExampleMD) <- ExampleBS$Ticker

## graph (direct, nodes weighted by tot assets and edges label using debts)
ggnet2(network(t(ExampleMD), directed = TRUE, names.eval = "weights"), label = TRUE,
       arrow.gap = 0.05, arrow.size = 10, size = ExampleBS$Total.Assets
       ,edge.color = "gray",  # Customize edge color
       edge.alpha = 0.8
       # labels sizes by the debt (need to take in row order the non zeros of MD)
       ,edge.label = t(ExampleMD)[t(ExampleMD) != 0])

# real data
## Loads the package
library(NetworkRiskMeasures)
## Minimum Density Estimation
### guarantee tot ib assets = tot ib liabilities (no ib out of the 50)
balance_sheet$`IB Assets` <- (sum(balance_sheet$`IB Liabilities`)
                             * (balance_sheet$`IB Assets`/sum(balance_sheet$`IB Assets`)))
### MD computation
set.seed(18) # seed for reproducibility
MD <- matrix_estimation(rowsums = balance_sheet$`IB Assets`,
                        colsums = balance_sheet$`IB Liabilities`, method = "md")

## rownames and colnames for the matrix
rownames(MD) <- colnames(MD) <- balance_sheet$Ticker

## graph
ggnet2(MD, size = balance_sheet$`Total Assets`, label = TRUE
       , label.size = 4, size.cut = 5, color = col)

ggnet2(t(MD), size = balance_sheet$`Total Assets`, label = TRUE
       , label.size = 4, size.cut = 5, color = col)


## graph label outside
ggnet2(MD, size = balance_sheet$`Total Assets`, label = FALSE
       , label.size = 4, size.cut = 5, color = col) +
  theme(legend.text = element_text(size = 8)) +
  geom_label(aes(label = balance_sheet$Ticker),nudge_y = 0.05) +
  guides(size = guide_legend(
    title = expression(paste("Total Assets (", "\u20AC", " bi)")),
    override.aes = list(size = c(3, 5, 7, 9, 11))
  ))


# save
write.xlsx(MD, "Data/MD.xlsx")
saveRDS(gmd, "Data/Networks/network_BS.Rds")
```

Part IV - R code for simulating Financial Contagion

```r
# Financial Contagion Simulations
# Description: simulate financial contagion in the interbank market
# and with the additional effect of bank runs
# Author: Wagner E. Schuster (2023/06)

# setup
rm(list=ls());gc()
setwd("~/QEM/BARCELONA/Thesis")

## load data
library(xlsx)
### Balance Sheet
balance_sheet <- xlsx::read.xlsx("Data/BS_interbank.xlsx",1)
#### take just important columns
balance_sheet <- balance_sheet[1:47,1:15]
#### rename
names(balance_sheet)[c(2,3,5,6,10,11)] <- c("Institution Name",
                                            "Ticker",
                                            "Total Assets",
                                            "Total Liabilities",
                                            "IB Assets",
                                            "IB Liabilities")

### interbank connections
MD <- xlsx::read.xlsx("Data/MD.xlsx",1)
#### take out 1st column (noise)
MD <- MD[,-1]

#### SIMULATIONS interbank contagion ####
# table to collect results
SIMULATIONS <- as.data.frame(balance_sheet$Ticker)
SIMULATIONS$add_banks_default <- NA
SIMULATIONS$banks_default <- NA
SIMULATIONS$add_assets_default <- NA

# loop to compute the simulations ib
for (i in 1:length(balance_sheet$Ticker)){
  # first shock (bank i fail)
  capital_sim <- as.matrix(balance_sheet$CET1 - MD[,i])

  # is there any default?
  if (any(capital_sim < 0)){

    # 1st round simulation
    # capital_sim minus (-) sum of columns MD of banks in default
    # by using "MD[,capital_sim<0]" I take the respective columns
    # in MD where row capital < 0 (default)
    capital_sim_round1 <- as.matrix(capital_sim
        # need to put a column with zeros in case it is just one default
                                    - rowSums(cbind(0,MD[,capital_sim<0])))
```

```r
    # initiate simulations next rounds
    capital_sim_rounds <- capital_sim_round1

    # guarantee capital bank i won't be negative (otherwise it would be counted again)
    # (it is already discounted at the very begging in capital_sim)
    capital_sim_rounds[i] <- 0

    PREV <- capital_sim
    NEXT <- capital_sim_rounds

    # is there any additional default? number of defaults bigger than before
    while (table(NEXT < 0)["TRUE"] > table(PREV < 0)["TRUE"]){
      PREV <- capital_sim_rounds

      # by doing the computation always in respect to initial cap_sim,
      # I avoid duplicate defaults
      capital_sim_rounds <- as.matrix(#capital_sim_round1
                                capital_sim
                # sum IB liab of banks that default in the last round
                # need to put a column with zeros in case it is just one default
                                - rowSums(cbind(0,MD[,capital_sim_rounds < 0])))

      # guarantee capital bank i won't be negative (otherwise it would be counted again)
      # (it is already discounted at the very begging in capital_sim)
      capital_sim_rounds[i] <- 0

      NEXT <- capital_sim_rounds
    }

    # remove the bank simulated in the count of additional (in case it is counted)
    capital_sim_rounds[i] <- 0
  }

  if (any(capital_sim < 0)){
    # write in the table the results
    SIMULATIONS$add_banks_default[i] <- length(balance_sheet$Ticker[capital_sim_rounds < 0])
    SIMULATIONS$banks_default[i] <- paste(balance_sheet$Ticker[capital_sim_rounds < 0], collapse=", ")
    SIMULATIONS$add_assets_default[i] <- sum(balance_sheet$`Total Assets`[capital_sim_rounds < 0])

  }else{
    # if there is no default in 1st shock
    SIMULATIONS$add_banks_default[i] <- 0
    SIMULATIONS$banks_default[i] <- "none"
    SIMULATIONS$add_assets_default[i] <- 0
  }
}

# save results
write.xlsx(SIMULATIONS, "Data/SIMULATIONS_IB.xlsx")


#### SIMULATIONS interbank contagion + Bank Runs ####
```

```r
# 1st: check banks sentiment connected
# recover sentiment connections from the network
network_twitter <- readRDS("Data/Networks/network_twitter.Rds")

# table to collect results
CONNECTIONS <- as.data.frame(balance_sheet$Ticker)
CONNECTIONS$num_connections <- NA
CONNECTIONS$banks_connected <- NA

# loop to extract the connections
for (i in 1:length(balance_sheet$Ticker)){
  # extract connections
  connections <- names(unlist(network_twitter[[i]]))
  # take string after point
  names <- sub(".*\\.(.*)$", "\\1", connections)
  # write in the table
  CONNECTIONS$num_connections[i] <- length(connections)
  CONNECTIONS$banks_connected[i] <- paste(names, collapse=", ")
}

# 2nd: compute simulations IB + BR
## table to collect results
SIMULATIONS_BR <- as.data.frame(balance_sheet$Ticker)
SIMULATIONS_BR$tot_banks_default <- NA
SIMULATIONS_BR$banks_default_br <- NA
SIMULATIONS_BR$tot_assets_default <- NA

## loop to compute the simulations ib + br
for (i in 1:length(balance_sheet$Ticker)){

  # Get the name of the banks sentiment connected
  current_banks <- strsplit(CONNECTIONS$banks_connected[i], ", ")[[1]]

  # Find the corresponding row(s) in the dataset where the bank names match
  matching_rows <- balance_sheet[balance_sheet$Ticker %in% current_banks, ]
  # transform numeric
  matching_rows <- as.numeric(rownames(matching_rows))

  # first shock (bank i fail) - IB
  capital_sim <- as.matrix(balance_sheet$CET1 - MD[,i])

  # now take out effect from BR
  capital_sim <- capital_sim -
    rowSums(cbind(0,MD[,c(matching_rows)]))

  # from now it's similar code as for IB contagion...

  # is there any default? (more than the defaults in 1st round BR)
  if (any(capital_sim < 0)){
    # 1st round simulation
    # capital_sim minus (-) sum of columns MD of banks in default
    # by using "MD[,capital_sim<0]" I take the respective columns in MD where
    # row capital < 0 (default)
```

```r
  capital_sim_round1 <- as.matrix(capital_sim
                  # need to put a column with zeros in case it is just one default
                                  - rowSums(cbind(0,MD[,capital_sim<0])))

  # initiate simulations next rounds
  capital_sim_rounds <- capital_sim_round1

  # guarantee capital bank i won't be negative (otherwise it would be counted again)
  # (it is already discounted at the very begging in capital_sim)
  capital_sim_rounds[i] <- 0

  # same for banks 1st BR
  capital_sim_rounds[matching_rows] <- 0

  PREV <- capital_sim
  NEXT <- capital_sim_rounds

  # is there any additional default? number of defaults bigger than before
  while (table(NEXT < 0)["TRUE"] > table(PREV < 0)["TRUE"]){
    PREV <- capital_sim_rounds

    # by doing the computation always in respect to initial cap_sim I avoid duplicate defaults
    capital_sim_rounds <- as.matrix(#capital_sim_round1
      capital_sim
      # sum IB liab of banks that default in the last round
      # need to put a column with zeros in case it is just one default
      - rowSums(cbind(0,MD[,capital_sim_rounds < 0])))

    # guarantee capital bank i won't be negative (otherwise it would be counted again)
    # (it is already discounted at the very begging in capital_sim)
    capital_sim_rounds[i] <- 0

    # guarantee also capital banks in 1st BR won't be negative for the same reason
    capital_sim_rounds[matching_rows] <- 0

    NEXT <- capital_sim_rounds
  }

  # force default banks in BR (make capital negative)
  #capital_sim[as.numeric(rownames(matching_rows))] <- -1
  capital_sim_rounds[c(matching_rows)] <- -1

  # remove the bank simulated in the count of additional (in case it is counted)
  capital_sim_rounds[i] <- 0
  # retrieve names and assets from banks that fail
}
if (any(capital_sim < 0)){
  # write in the table the results
  SIMULATIONS_BR$tot_banks_default[i] <- length(balance_sheet$Ticker[capital_sim_rounds < 0])
  SIMULATIONS_BR$banks_default_br[i] <- paste(balance_sheet$Ticker[capital_sim_rounds < 0],
                                              collapse=", ")
  SIMULATIONS_BR$tot_assets_default[i] <- sum(balance_sheet$`Total Assets`[capital_sim_rounds < 0])
```

```r
  }else if (length(matching_rows)>0){
    # need to check if there were sentiment connections
    SIMULATIONS_BR$tot_banks_default[i] <- length(matching_rows>0)
    SIMULATIONS_BR$banks_default_br[i] <- CONNECTIONS$banks_connected[i]
    SIMULATIONS_BR$tot_assets_default[i] <- sum(balance_sheet$`Total Assets`[matching_rows])

  }else{
    # if there is no default in 1st shock
    SIMULATIONS_BR$tot_banks_default[i] <- 0
    SIMULATIONS_BR$banks_default_br[i] <- "none"
    SIMULATIONS_BR$tot_assets_default[i] <- 0
  }
}


# 3rd: merge IB and IB + BR and compute comparisions
SIMULATIONS_BR <- merge(CONNECTIONS, SIMULATIONS_BR,  by="balance_sheet$Ticker", sort = FALSE)
## bring variable rank to order and total assets for later
SIMULATIONS_BR <- merge(SIMULATIONS_BR, balance_sheet[,c(1,3,5,2)],
              by.x="balance_sheet$Ticker",
              by.y="Ticker"
              ,sort = FALSE)

# save results
write.xlsx(SIMULATIONS_BR, "Data/SIMULATIONS_BR.xlsx")

# 4th: merge with simulations with only IB and compute comparisons
FINAL <- merge(SIMULATIONS, SIMULATIONS_BR, by="balance_sheet$Ticker", sort = FALSE)

# take difference
# keep just important variables
# when read it adds 1 column
#COMPARISON2 <- COMPARISON[,c(1,3,4,5,18,17,19)]
#FINAL2 <- FINAL[,c(1,2,3,4,16,15,17)]
FINAL2 <- FINAL[,c(1,2,3,4,5,6,7,8,9,10,11)]

# rename
names(FINAL2) <- c("Bank", "N IB", "Banks IB", "Assets IB", "N Connections", "Banks Connected",
                        "N IB + BR", "Banks IB + BR", "Assets IB + BR", "Rank", "Assets")
# compute additional effect bank runs
#COMPARISON2$add_banks <- COMPARISON2$`N IB + BR` - COMPARISON2$`N IB`
#COMPARISON2$add_assets <- COMPARISON2$`Assets IB + BR` - COMPARISON2$`Assets IB`
FINAL2$add_banks <- FINAL2$`N IB + BR` - FINAL2$`N IB`
FINAL2$add_assets <- FINAL2$`Assets IB + BR` - FINAL2$`Assets IB`

# save results
write.xlsx(FINAL2, "Data/FINAL.xlsx")
```