Missing Financial Network Data

Allison Oldham Luedtke*

Jinfeng Chen

Henry Fisher

Yanet Tadele

October 1, 2023

Abstract

Little data exists describing the links of the US financial network. Using a computational model of interbank lending, we show that this lack of data can lead to erroneous model predictions. We find that missing a single loan in the network can lead to large differences in the predicted number of unpaid loans and total dollars repaid. This missing data could mean implementing policies that are designed to improve macroeconomic stability but that could actually lead to substantial destabilization. These results are robust across multiple network sampling regimes.

Keywords: interbank lending; financial networks; incomplete data

JEL Classifications: C63, E58, G21, L14

*St. Olaf College, Northfield, Minnesota. We are very grateful to Andrew Skemp and Demetri Lord for their excellent research assistance on this project.

1. Introduction

The collapse of Silicon Valley Bank and Signature Bank sparked a renewed interest in preventing cascading bank failures like what amplified the impact of the 2008 financial crisis. Memories of double-digit unemployment, plummeting home values, and declining retirement accounts spurred concern about another major downturn. When Lehman Brothers failed in the fall of 2008, many banks struggled or failed in the following weeks. This resulted in a 700 billion dollar bailout of the financial sector from the US government to avoid a more catastrophic collapse of the banking industry. Economic researchers and policymakers have recognized the importance of controlling financial contagion in order to prevent the worst symptoms of the Great Recession. Avoiding an economic downturn of similar magnitude is a vital task for policymakers. Understanding the lending relationships between banks is paramount for averting future financial crises. This paper illustrates the impact of missing financial data on our ability to respond to such crises.

Theoretical work has been done to understand how the interbank lending network as a whole affects financial stability (Jackson 2010; Schweitzer et al. 2009; Hasman 2013). Researchers have characterized the simultaneously robust-yet-fragile nature of networks in the face of negative shocks (Acemoglu, Ozdaglar, and Tahbaz-Salehi 2015; Chinazzi and Fagiolo 2013). If the interbank lending network is too interconnected, it serves to propagate the shock to many banks throughout the network. If it is not interconnected enough, banks must rely on only a few banks for repayment and are particularly vulnerable to shocks. However, little empirical work exists on this question because there is no data set that describes the specific loan relationships that exist in the US financial system.[1] Soramäki et al. 2007 describe the structure of the interbank lending network, finding it resembles a scale-free network with a few tightly connected core banks.

This paper defines the effect one loan can have on the overall interbank lending network. If a loan in the network is absent from the data, will projections about bank default and loan repayment be incorrect? And if so, by how much? We simulate interbank lending networks to measure the difference in repayment characteristics between networks that differ by exactly one link. We measure the difference in the total interbank loan shortfall across the whole network of banks, the number of banks unable to pay their loans in full, and the number of loans that are not paid in full.

In our simulations, banks lend to each other following the degree distribution described by

[1] This is not true for all countries. Imai and Takarabe (2011) use data describing banks in Japan to investigate whether banking integration contributes to the propagation of financial contagion (Imai and Takarabe 2011).

Soramäki et al. 2007. The banks are represented as nodes in the network, while the edges represent nodes. We then calculate the equilibrium repayments for each loan in the network. To do this, we adapt the network model of interbank lending described in Acemoglu, Ozdaglar, and Tahbaz-Salehi 2015. In the model, each bank has outside investments in addition to their loans. The returns from these investments, the banks' senior obligations, and their other assets determine how much of their loans each bank repays. If a bank fails to repay its loans, the banks that loaned it money will be able to repay fewer of their loans. It is through this effect that financial contagion spreads throughout the network.

We need final results here

The ability of economists and policymakers to provide and implement accurate schemes for the stable financial system is dependent upon their abilities to predict and analyze the entire market. Currently, there is a lack of accessible interbank lending data, which precludes researchers and policymakers from achieving a full picture of the financial networks. We show that any missing data can skew understanding of how these lending networks will act. Missing just one link leads to changes in total repayments, or network-wide shifts of millions of dollars. Our results indicate that more accurate repayment predictions require more complete data.

2. Network Model

The focus of this paper is to model the dynamics of loan repayments within the American interbank lending network and the implications for economic stability. Within this network, loan repayments facilitate the flow of funds between banks. We use a model that incorporates cash flows, senior obligations, and repayment equilibrium behavior. Interruptions to repayments can set off a chain reaction among interconnected banks. Through the introduction of missing data scenarios, such as adding or omitting loan links, this paper demonstrates how seemingly minor data errors can have significant repercussions on repayment cascades. Such cascades can lead to a surge in unpaid loans and even bank defaults.

The behavior of each individual bank is modeled in the simulation. Each bank is obligated to repay its debts to other banks to the extent it can. We adopt a model from Acemoglu, Ozdaglar, and Tahbaz-Salehi 2015. We denote the face value of a loan from bank i to bank j y_{ij}. Bank j repays to bank i some number r_{ij} between 0 and y_{ij}. The value of r_{ij} depends on the funds that bank j has at its disposal. Bank j's available resources depend not only on its additional cash and the investments it makes but also on the ability of its debtors to repay their loans. The model therefore describes an equilibrium repayment network for all of the banks in the system.

If there is no loan between two banks in the network, $y_{ij} = r_{ij} = 0$. The amount banks repay to each other is dependent upon their current assets and obligations. Current obligations may prevent them from repaying interbank loans on time and in full. These consist of payments owed to firms, individuals, and other private and public entities. We denote these "senior obligations" payments that banks must repay before the repayment to other banks, as v. Included in this are taxes, wages, and rent. All senior obligations must be repaid in full before the payment of interbank lending can begin.

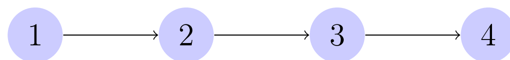
Each bank has an amount of cash on hand c_j, and accumulates a random return on individual investment, z_j. These both contribute to the total amount of assets available to the bank, or the cash flow, h_j. A bank's ability to pay back interbank loans in full is determined by whether or not h_j is sufficient to cover both senior and junior obligations. If h_j is insufficient to cover the repayment of senior obligations, bank j will be unable to repay its loan in full. Each bank's ability to repay other banks in the network is determined by the repayment quantities and consistencies of other banks. The repayment equilibrium model is described by:

$$r_{ij} = \frac{y_{ij}}{y_j} \max\{\min\{y_j, c + A_j + \sum_{j \neq s} r_{js} - v\}, 0\} \tag{1}$$

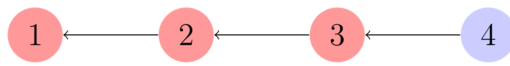
We quantify banks' decision-making processes. In our network, banks use the repayment equation to decide at what degree is feasible to repay the loan. If $h_j - v > y_j$, then the bank can repay all of its loans y_j in full and therefore repay each loan y_{ij} in full. If $0 < h_j - v < y_j$, then bank j will repay each bank a fraction of the loan $\left(\frac{y_{ij}}{y_j}\right)(h_j - v)$. Finally, if $h_j - v < 0$, the bank defaults on all of its loans since it does not have enough cash on hand to cover its senior obligations. Banks will always find a small amount to pay back between y_{ij} and 0. This value can never be negative, because, in reality, banks can never repay a negative sum. If we do this for all the loans borrowed from bank j, we will know how bank j repays all its debts. The cash flow equation tells us how much capital a bank has on hand. The cash on hand of every bank is equal to its cash on hand, plus its investments or easily liquifiable cash equivalents, plus all the repayments from other banks. If a bank cannot repay its loans, it can use that money to pay its senior obligations so that it will not fall into default.

Since each bank relies on the payments of other banks within the network, the repayment model allows us to quantify the consequences of loan delinquency for the network as a whole. Suppose bank 1 loaned to bank 2 and bank 2 to bank 3 and so on as shown in Figure 1. These loans denoted as y_{12} and y_{23} respectively. The amount that bank 2 repays bank 1, r_{12}, depends upon the repayments of bank 3 to bank 2 and bank 4 to bank 3, r_{23} and r_{34}. Suppose bank 3 is delinquent on its loan due to a low return on its outside investments A_2, meaning the loan from the bank is not paid back in full. Therefore, bank 2 does not have enough cash to cover enough senior obligations to repay their loan in full from bank 1, making bank 2 delinquent on their loan, as seen in Figure 2.

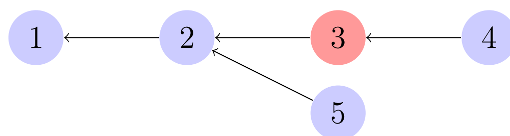
Figure 1



(a) Network of original loans (y_{ij})



(b) Network of original repayments (r_{ij})



(c) Network of new repayments with an added link(r_{ij})

Notes:

However, consider if we missed a link in the network of loans from bank 2 to bank 5, y_{25}. Let's assume that Bank 5 had a good return on their outside investments and can repay their loan in full. This will allow bank 2 to repay its loan to bank 1 in full, preventing the cascading bank failures that could lead to a financial system crash. This example is magnified when we incorporate more banks into the network

3. Simulations

To simulate missing data in a network, we add a link between two randomly chosen banks. Hence, our results come from network generation simulations that match observable network characteristics such as shape and degree distribution. Specifically, we simulate the repayment amounts of banks within the network when faced with an additional, previously unaccounted-for link.

To identify the effects of the missing loan in a network, we simulate financial networks in the following ways: We first generate a random network of loans that aligns with the actual degree distribution. Then I find the repayment equilibrium. We then add a loan in a random spot with a random cost associated and finally compare repayment equilibria among networks with and without and without an added link. We contextualize those results within aggregate economic outcomes. If we assume this new network with the additional link is the actual network, and the previous network represents the network that is readily accessible, the disparity in repayment equilibrium can be associated with the missing link in the network.

Our simulations measure five characteristics of the repayment network. Those include the total change in shortfall, loans with shortfall, banks with shortfall, banks that fail to repay some of their loans, and banks that fail to repay anything. These five chosen indicators measure the effects of the presence of an added link within a financial network.

The total change in shortfall measures the amount of money not repaid across the entire network. It is calculated by taking the sum of the difference between each respective element of the Y_{ij} and R_{ij} matrices ($\sum(y_{ij} - r_{ij})$). The total loans with a shortfall is a count of all the loans in each network that are not fully repaid. The change in loans with shortfall measures the effect of the additional link on the amount of loans that are fully repaid. The number of banks with a shortfall is measured by counting the nonzero values that result from subtracting the row sums of R_{ij} from the column sums of Y_{ij}. This measures the number of banks that fail to repay all their loans. The number of banks that fail to repay some of their loans is measured by computing the difference between the column sums of the R_{ij} matrix and the column sums of the Y_{ij} matrix. The resulting number of nonzero elements equals the number of banks that failed to repay at least some proportion of their loans. If the resulting difference for a sum of column j equals the sum of column j in the Y_{ij} matrix, then that bank has failed to repay anything.

We generate a random directed network to model the financial network using the power-law cluster graph model. This approach combines properties of both small-world networks and scale-free networks, making it a suitable choice for stimulating complex systems like financial networks. With small-world properties, the model ensures efficient and rapid information dissemination, vital for quickly transmitting financial transactions and responding to market shocks. On the other hand, by incorporating scale-free properties, the model captures highly connected nodes, reflecting critical entities that influence the system's stability. Financial networks' inherent complexity is accurately represented through this combination, allowing us to study emergent behaviors, cascading effects, and systemic risks arising from intricate interactions. The level of clustering further enhances the model's fidelity, as it accurately captures interconnected institutions or sectors exhibiting correlated behavior or risk profiles. The generation process introduced randomness through the randomization of edges, which adds complexity to the resulting network. Using this we generate a random directed network.

In order for an additional loan to be added between two banks, there must not already be a loan between the two banks in the same direction within the original network.

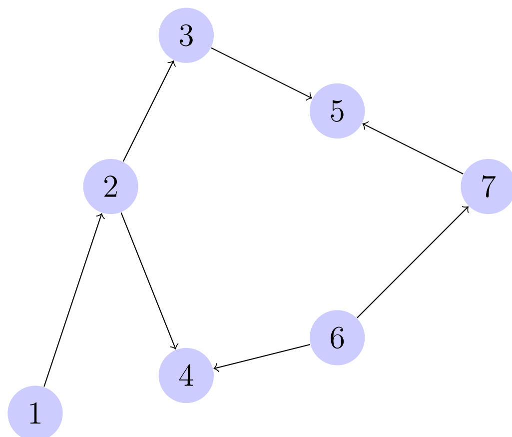
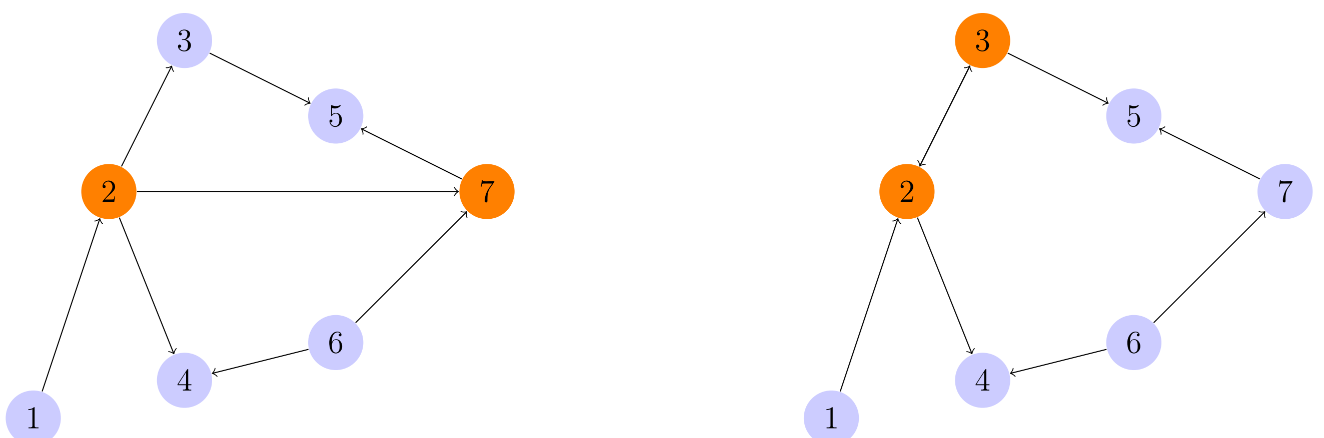


Figure 2: Original network without an added loan



(a) Loan added between two banks without any pre-existing loan between them

(b) Loan added between two banks with a pre-existing loan in the opposite direction

Figure 3: Networks with an added Loan

To make the simulation relevant to the actual banking network, the parameters in the model are matched as closely as possible to the measured characteristics of the financial network. Each bank is assumed to have both borrowed from other banks and lent to other banks. Additionally, We assume every loan in the network is of equal value. This ensures that any result from the additional link is attributable to the extra link itself, not how much the loan is worth. Each loan is set to be worth 100 million dollars. This is both within the middle 50 percent of measured interbank loan values, as well as the most common loan amount. [*cite NY Fed paper*].

Each bank is assumed to have average liquidity and senior obligations. I keep constant two parameters for every bank in the network: cash on hand (c) and senior obligations (v). Cash on hand is calculated to be the total liquidity in the US banking system divided by the total number of FDIC-insured financial institutions in the United States. The measured liquidity in the US was 75 billion dollars in total [*cite FRED*]. The total number of FDIC-insured institutions is 4706 [*cite FDIC*]. Thus, we have 4706 nodes in our network, and our c variable $=$ 16 million dollars. Employee wages are used as a proxy for total senior obligations. The average bank employee makes 75 thousand dollars *cite*, and the banking sector has 2,215,318 employees, which means the average monthly payroll obligation is about three million dollars.

I set the outside project result, A_j, to be randomly assigned to each bank. By doing so, the simulation reflects that in every time period, some banks do well and some banks do poorly. In the simulation, the values for A_j are generated using a normal distribution centered around ten million dollars with a standard distribution of five million dollars.

Our simulations consist of networks with 100 banks. This is because the system of equations described by figure 1, grows exponentially with the number of banks. It is computationally impractical to simulate networks with the full 4706 banks. We've found that 100 nodes are the maximum number of banks that can run in a reasonable amount of time while still yielding meaningful results. We ran a simulation ranging from twenty to five-thousand banks which printed our four parameters to determine their volatility in relation to a given number of banks. Each parameter shows varying levels of accuracy at n $=$ 100, however, in most cases, 100 banks output values in the same order of magnitude as the intended value. With readily accessible resources, one hundred banks are significantly more feasible than a complete network, while not undermining the integrity of tests' legitimacy.

4. Results

The results of our simulation indicate that a difference of one link can lead to substantial changes in important aggregate outcomes. We consider two outcomes before and after the addition of a random link: the number of loans that are not paid in full and the total dollar value shortfall in repaid loans. We use paired t-tests to test the hypothesis that these two measures are the same before and after the addition of the link. We reject this hypothesis (at the 1% level) in both cases.

Figure 4: Difference in Unpaid Loan Dollars

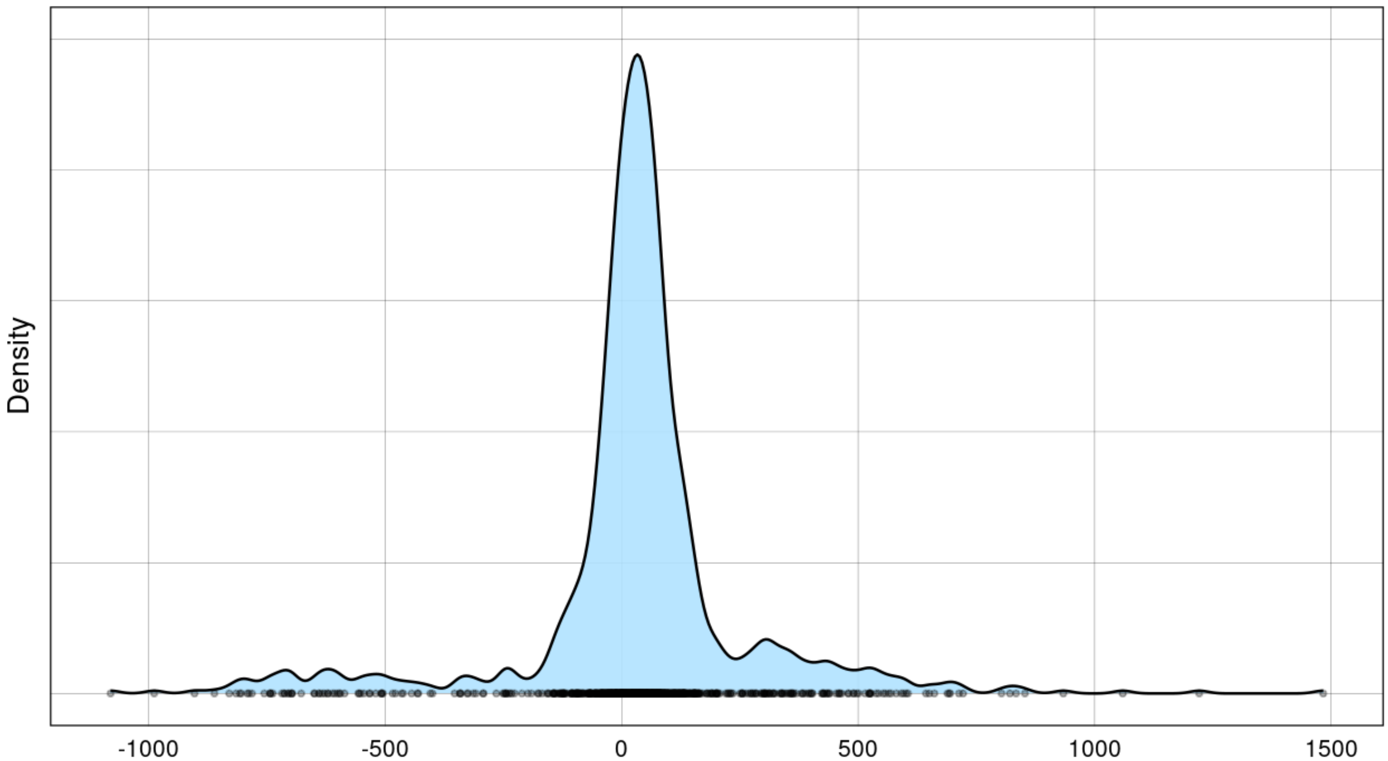


Figure 2 depicts the distribution of the difference between the total unpaid dollars across the entire network before the addition of the link and after the addition of the link. The peak of the distribution is to the right of 0. The average difference across our experiments was \$32.54 dollars. This was statistically significantly different from zero with a p-value of 0.00.

Figure 5: Difference in Delinquent Loans

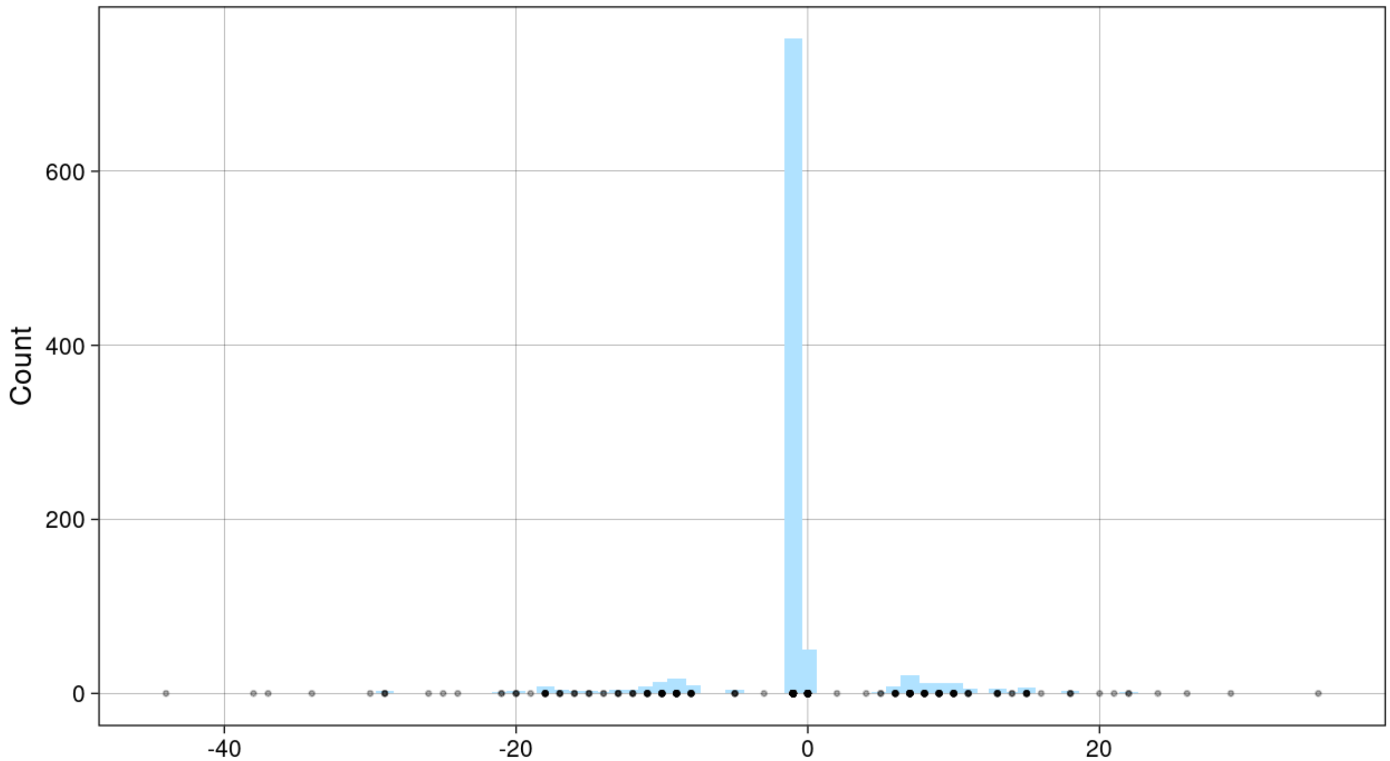


Figure 5 depicts the distribution of the difference between the number of loans that are not paid in full across the entire network before the addition of the link and after the addition of the link. The peak of the distribution is to the left of 0. The average difference across our experiments was 1.05 loans. This was statistically significantly different from zero with a p-value of 0.00.

From these results, we can see that if we miss one link in the network, it is most likely that we will miss a loan that is fully repaid and tens of millions of dollars. Though we only miss one link, the difference in the outcomes will not be small. If we want to correctly predict a crisis, it is important and necessary not to miss any links.

5. Conclusion

This paper demonstrates the effect missing a single loan can have in the network of interbank lending. These small network differences can have large consequences. The addition of one loan can shift repayment equilibrium outcomes by hundreds of millions or even billions of dollars in either direction. This paper illustrates the importance of having perfect data on the entire network of interbank loans. If a difference of one loan can precipitate such a large swing in projected outcomes, having perfect data is all the more important. Correct policy decisions depend on our ability to understand the financial network. Imperfections in our understanding of the network could cause us to choose imperfect policy and wrongly project outcomes.

References

Acemoglu, Daron, Asuman Ozdaglar, and Alireza Tahbaz-Salehi (Feb. 2015). "Systemic Risk and Stability in Financial Networks". In: *American Economic Review* 105.2, pp. 564–608. ISSN: 0002-8282. DOI: 10.1257/aer.20130456. URL: https://www.aeaweb.org/articles?id=10.1257/aer.20130456 (visited on 01/21/2023) (Go back).

Chinazzi, Matteo and Giorgio Fagiolo (2013). "Systemic Risk, Contagion, and Financial Networks: A Survey". In: *SSRN Electronic Journal*. ISSN: 1556-5068. DOI: 10.2139/ssrn.2243504. URL: http://www.ssrn.com/abstract=2243504 (visited on 06/13/2019) (Go back).

Hasman, Augusto (2013). "A Critical Review of Contagion Risk in Banking". In: *Journal of Economic Surveys* 27.5. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1467-6419.2012.00739.x, pp. 978–995. ISSN: 1467-6419. DOI: 10.1111/j.1467-6419.2012.00739.x. URL: https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1467-6419.2012.00739.x (visited on 01/21/2023) (Go back).

Imai, Masami and Seitaro Takarabe (Jan. 2011). "Bank Integration and Transmission of Financial Shocks: Evidence from Japan". In: *American Economic Journal: Macroeconomics* 3.1, pp. 155–183. ISSN: 1945-7707. DOI: 10.1257/mac.3.1.155. URL: https://www.aeaweb.org/articles?id=10.1257/mac.3.1.155&ArticleSearch%5Bwithin%5D%5Barticletitle%5D=1&ArticleSearch%5Bwithin%5D%5Barticleabstract%5D=1&ArticleSearch%5Bwithin%5D%5Bauthorlast%5D=1&ArticleSearch%5Bq%5D=network&JelClass%5Bvalue%5D=0&journal=6&from=j (visited on 03/23/2023) (Go back).

Jackson, Matthew O. (2010). *Social and economic networks*. OCLC: 254984264. Princeton, NJ: Princeton Univ. Press. 504 pp. ISBN: 978-0-691-14820-5 978-0-691-13440-6 (Go back).

Schweitzer, Frank, Giorgio Fagiolo, Didier Sornette, Fernando Vega-Redondo, Alessandro Vespignani, and Douglas R. White (July 24, 2009). "Economic Networks: The New Challenges". In: *Science* 325.5939, pp. 422–425. ISSN: 0036-8075, 1095-9203. DOI: 10.1126/science.1173644. URL: https://science.sciencemag.org/content/325/5939/422 (visited on 06/15/2019) (Go back).

Soramäki, Kimmo, Morten L. Bech, Jeffrey Arnold, Robert J. Glass, and Walter E. Beyeler (June 1, 2007). "The topology of interbank payment flows". In: *Physica A: Statistical Mechanics and its Applications* 379.1, pp. 317–333. ISSN: 0378-4371. DOI: 10.1016/j.physa.2006.11.093. URL: http://www.sciencedirect.com/science/article/pii/S0378437106013124 (visited on 06/15/2019) (Go back).