Interbank market formation through reinforcement learning and risk aversion

Anqi Liu[a], Cheuk Yin Jeffrey Mo[a], Mark Paddrik[b], Steve Y. Yang[a,*]

[a] *Stevens Institute of Technology, Financial Engineering Division,*
1 Castle Point on Hudson, Hoboken, NJ 03070

[b] *Office of Financial Research, U.S. Department of the Treasury,*
717 14th Street NW, Washington, DC 20220

Abstract

In this study, we propose a multi-agent model to simulate bank lending and borrowing behaviors which can generate realistic interbank network dynamics. Using data from 2001 to 2014 that covers around 6600 banks, we model individual bank decisions using the temporal difference reinforcement learning algorithm based on banks' lending preferences and environment, and we then generate the interbank market dynamics from the empirical data. This dynamic model allows us to construct interbank networks as they change with bank risk preferences, and thus facilitates the analysis of the banking systems stability. The model successfully replicates the key characteristics of interbank lending and borrowing relationships that have been documented in the recent literature. A key finding of this model is that individual bank's risk aversion choice leads to unique interbank market structures that suggest the macro risk preference of the market.

Keywords: Interbank lending market, Contagion risk, Multi-agent system, Financial crisis

JEL: D85, G17, G21, L14

[*] Corresponding author

Email addresses: aliu@stevens.eud (Anqi Liu), cmo1@stevens.edu (Cheuk Yin Jeffrey Mo), Mark.Paddrik@ofr.treasury.gov (Mark Paddrik), syang14@stevens.edu (Steve Y. Yang)

1. Introduction

Prior to the 2008-2009 financial crisis, central banks and market regulators were primarily concerned with banks' performance through a microprudential lens, examining individual banks' asset and liability portfolios to control risk. While this perspective has been commonly used to identify problematic banks, it does not consider the system-wide implications of other troubled banks. As banks are highly interconnected, a marcoprudential approach, which incorporates interbank relationships, would help identify potential latent fragility of the network but also macro risk preferences and tolerances.

During the crisis, it was evident that the interbank market behaviors suggested a heightened concern for counterparty risk that reduced liquidity and increased the cost of financing for weaker banks (Afonso et al., 2011). Banks overall were less likely to lend liquid assets to each other. Large banks, which play a central role in this market, increased their liquidity buffers (Berrospide, 2012), forcing medium and small banks to look for new sources of liquidity. However, due to the latent nature of these interbank activities, most of the current literature focus on either highly stylized network models or empirical examinations of limited observations in some countries (Boss et al., 2004; Iori and Gabbi, 2008; Roukny et al., 2014) to gain insight on counterparty risk and liquidity risk contagion.

In reality, banks are autonomous decision makers, who respond or adapt to market changes to achieve individual objectives, which are dynamic in nature. This characteristic is largely not present in a pure optimization theoretical framework, which has led to a series of attempts to study endogenous interbank network formation models (Georg (2013); Ladley (2013); Lux (2015)).We develop a multi-agent system where agents are capable of learning using a temporal difference reinforcement learning algorithm based on individual bank performance goals and the changing macro environment surrounding those choices.

Our study falls into the endogenous interbank network formation literature and bridges a few gaps in the existing reconstruction literature. First, instead of using a global optimization approaches to rebuild networks, we form a interbank network dynamically through individual banks preference. Second, in order to achieve a banks specific performance objectives, we allow banks to be adaptive to environment by changing lending and borrowing policies through reinforcement learning.

We collect data from U.S. Federal Deposit Insurance Corporation (FDIC) and build a model to represent U.S. interbank lending system, aiming to investigate how different institutions interact, how risk preferences influence lending/borrowing decisions, and finally how these endogenous interactions lead to different agent converging policies. This model captures the dynamic nature of interbank lending networks, including overnight debts (federal funds), short-term and long-term debts markets. We simulate the banks' lending and borrowing behaviors according to statistical patterns of individual banks and general behavior patterns from the empirical findings. This framework reconstructs interbank exposures of autonomous banks by having each learn how best to achieve their risk preferences through policy selection. These result are then validated against empirical findings.

The primary contribution of this study is to develop an dynamic interbank market model with learning agents to reconstruct the interbank network based on bank performance data and behavioral patterns. From network-based model perspective, we design two experiments to answer the following questions.

1. Can a multi-agent system with learning agents reconstruct the dynamics of the interbank networks?

2. Does agent's risk preference influence change make the system less prone-to contagion?

To answer these questions, we first calibrate the multi-agent model with the U.S. bank balance sheet data, and then examine the characteristics of networks generated from the learning agents. Two experiments are conducted in this study to investigate the network property changes of U.S. interbank market with different bank risk profiles in lending to other banks upon receiving a lending request. First, we look at two different interbank networks of banks, one with a loosening lending policy, where banks are most likely to lend to one another, and one with a tightening lending policy, where banks tend to set higher lending standards, respectively. We then compute and compare network properties of the modeled interbank networks under different risk preference settings.

Results show that as the network stabilizes at a certain level of risk preferences, the network degree begins to decline as bank agents continue to tighten their lending policies. We find the following: 1) the interbank lending market functions better as it avoids shocks to the liquidity of agents; 2) the shock propagation is lessened because failing bank networks are isolated. Overall, our model brings data to modeling financial system contagion problems, and bridge the gap in the existing literature of interbank network contagion literature where most of the existing stress testing methods implicitly assume that banks are highly stylized with no suboptimal behaviors.

The rest of the paper is organized as follows. In section 2, we provide background literature. We then present the interbank lending model framework in section 3, and in section 4 we introduce reinforcement learning agents into the interbank framework to build a multi-agent system. And then we discuss the data used in the study in section 5. In section 6, we describe validation of the model and discuss the convergence of the learning process. Finally, we conduct two experiments and discuss findings accordingly in section 7. We then conclude the paper in section 8.

2. Background and related literature

2.1. Interbank network topology

Empirical studies about interbank network structure, especially for overnight market, have been applied to explain system-wide risk exposures. Boss et al. (2004), Iori and Gabbi (2008), and Roukny et al. (2014) discover similar network characteristics of banking system in Austrian, Italy and German. They suggested that interbank market presents sparsity, power law degree distribution, small clustering, and small world properties. In a similar but more comprehensive study, Cont et al. (2013) investigated Brazilian interbank bilateral exposures based on balance sheet data and documented that total network fails to satisfy the small world conditions in terms of arbitrary small clustering coefficient.

Other studies have explored bilateral exposures from the perspective of bank behavioral preferences. By categorizing banks into two groups, small and large, Cocco et al. (2009) concluded that small banks rely on large banks to borrow funds, while large banks tend to hold consistent relationships with familiar counterparties because of lower interest payments.

2.2. Systemic risk and interbank network extrapolation

Pioneering works by Allen and Gale (2000); Eisenberg and Noe (2001) highlighted the role of interconnection on systemic risk. The 2008-2009 financial crisis revealed the fact that regulators and market participants had very limited information to examine financial network linkage and identify risk channels. To address this problem, interbank market data and network models were applied to examine the cause of the 2008 financial crisis (Iori et al. (2006); Elliott et al. (2014)).

However, though there exists a growing consensus that examining the banking system through of interbank exposures is important much of the interbank lending data is not easily assessable in many countries. Existing research has focused trying to derive this data through network formation algorithms, which can be summarized into two camps.

The first utilizes linear algebra concepts to approximate the bilateral exposures by setting balance sheet lending and borrowing as marginals and then fills in the adjacency matrix. The predominate method is Maximum Entropy which follows a simple risk-sharing rule and splits loans to banks as even as possible (Upper and Worms (2004)). However, the real interbank system is sparse, as having many relationships for small banks is quite costly (Craig and Von Peter (2014)). As a result, a series of algorithms have been designed to optimize different network features (Basel Committee Supervision on Banking (2015)). Minimum Density (Anand et al. (2015)), has been one of the leading condensers due to its accuracy in bilateral exposure estimators despite it underestimates the total number of links.

The second group of researchers focus on the strategic network formation approach which setup rules on how the network reach equilibrium by maximizing agents' utility (Jackson and Wolinsky (1996); Acemoglu et al. (2015)). Gofman (2016) further improved these methods by incorporating agent trading decisions when build endogenous network and analyze systemic risk.

2.3. Multi-agent systems in interbank networks

As an alternative method to the theoretic approaches, a multi-agent system consists of autonomous agents with independent behavior rules, connections between agents, and the exotic environment (Macal and North (2010)). Compared with statistical and mathematical models, multi-agent systems are better at replicating real social phenomena, adaptive agents' behaviors, and information diffusion among agents (Gilbert and Terna (2000); Macy and Willer (2002)). Providing a platform for endogenously learning network formation, and have been applied on network topologies and contagion risk among banks (Georg

(2013); Ladley (2013); Lux (2015)). In addition, further extension has replicated multi-layered network structures hinging on multiple types of interbank loans (Kok and Montagna (2013)).

2.4. Multi-agent learning systems

In a multi-agent system, there are often two prominent types of agents: Zero-intelligence agents and learning agents. Zero-intelligence agents are described as those who "do not seek or maximize profits, and does not observe, remember, or learn" (Gobe and Sunder (1993)). Zero-intelligence agents have been widely adapted to simulate market trading environment in which traders are represented by these agents (Othman (2008)).

On the other hand, learning agents can improve their performance through learning from past experiences. Reinforcement learning methods are often used in developing learning agents. The method allows agents to select an action for each state that tend to maximize the expected future value of the reinforcement signal (Sutton and Barto (1988)). Various learning algorithms are used in reinforcement learning and the most common approaches are temporal difference learning and Q learning. Temporal difference learning is a value prediction method that updates its estimate of its previous estimation (bootstrap) after receiving each reinforcement signal for every action taken. Q learning is one of the most important algorithms used to optimize the expected future reward for each state-action pair. The optimal policy can be learned implicitly using Q learning.

Ramanauskas (2008) investigates how reinforcement learning can be applied in an artificial stock market modeling and discusses the benefits of implementing reinforcement learning methods compared to the rational expectation equilibrium method. He finds that reinforcement learning allows the investor agents to act in a more realistic and out-of-equilibrium situation. More closely connected to our model is the work done by Lux (2015). In this paper, Lux builts a dynamic multi-agent learning system that banks within the system will experience regular deposit shocks and they have to act accordingly to remain solvent. They rely on the fundamental reinforcement learning concept to select their counterparties by updating the trust factor between banks.

3. Multi-agent interbank lending framework

This section presents a multi-agent model to simulate the U.S. interbank lending market. We design an iterative dynamic framework in which banks settle debts, process payments, and update financial reports in each iteration. In this model, one iteration is equivalent to a quarter, or 3 months. The model's initialization and primary activities are modeled in every iteration is described as following.

First the model is initializes the system with a lending network with 6600 banks linked by interbank debts. There are two types of banks and three types of debts (see Table 1). This initial network is calculated by balance sheet data in 2006 using *maximum entropy* approach. However, knowing that this algorithm solves the bilateral exposure problem by generating too many links comparing with empirical findings of the real U.S. overnight interbank market, we then allow the model to reorganize itself through the multi-agent learning process.

Table 1: Interbank lending network setting

Nodes	Large banks	The largest four domestic banks: Bank of America, Citibank, J.P. Morgan Chase Banks, and Wells Fargo Bank
	Small banks	Other banks
Links	Overnight debts	Federal funds, usually expire overnight
	Short-term debts	Federal securities, usually expire within 1 month
	Long-term debts	Loans expire less than 1 year

In every iteration, banks first process payments to existing debts. According to expiration of the three types of interbank lending, overnight debts are paid fully, short-term debt payments and long-term debt payments are draw from

$U(99\%, 100\%)$ and $U(25\%, 100\%)$ respectively. We adopt Eisenburg-Noe iterative clearing vector algorithm to clear payments (Eisenberg and Noe, 2001). In this process, banks may face liquidity or solvency issue that prevent them to fulfill their obligations (see equation (1) and (2)). In this case, the lenders realizes written-down for debts with defaulting banks according to the Eisenburg-Noe algorithms *pro rata* loss distribution method.

Secondly banks search to settle debts by actively searching for counterparties. Generally, each bank sends borrowing requests to potential lenders, and the lenders response with approval or rejection. Detailed decision making policies are associated with agent reinforcement learning process which will be introduced in Section 4. At the end of each iteration, banks process balance sheet updates to realize collected payments and latest borrowing and lending. Moreover, we incorporate retained earnings based on an empirical distribution $Beta(17.36, -0.1, 0.3)$.

$$E_i(t) < A_i(t) - L_i(t) \tag{1}$$

$$C_i(t) < ON_i^p(t) + ST_i^p(t) + LT_i^p(t) \tag{2}$$

where $ON_i^p(t)$, $ST_i^p(t)$, and $LT_i^p(t)$ are bank i' payments of overnight borrowing, short-term borrowing, and long-term borrowing on period t.

4. Bank lending-borrowing with reinforcement learning

In this section, we present a learning method to simulate bank's behavior in the U.S. interbank lending and borrowing market. This method utilizes reinforcement learning to help guide bank decisions to lend and borrow based on the public and private information available to them. The remaining of this section covers the reinforcement learning framework, and how bank's lending and borrowing needs are matched to form the interbank network.

4.1. Reinforcement learning framework

Reinforcement learning is a computational technique to allow agents to learn and determine the ideal behavior based on past experience. More specifically, agents learn by receiving reinforcement signals by interacting with the environment. The objective of a reinforcement learning agent is to maximize its accumulative reward in the future. A simplified representation of the reinforcement learning framework is presented in Figure 1.

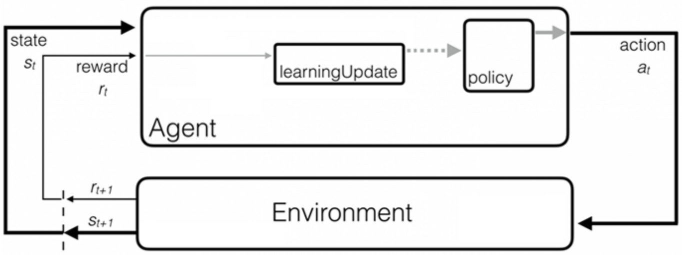


Figure 1: Agent environment interaction in reinforcement learning

There are three elements that drive the interaction between an agent and the environment: state, action, and reward. Upon evaluating the current state, the agent selects an action that maximizes their expected cumulative reward in the long run. The process of selecting an action based on a given state is typically specified in agent's policy. The environment will return a reward signal that agent can use to learn and improve their future evaluation of states. In our model, agents adopt the temporal difference approach to improve their policy in the current state based on the environment. This approach does not involve policy optimization, but instead agents have a target policy toward which they select their counterparties for new debts.

4.2. Banking system states

A state refers to all the current information, both public and private, received by an agent to determine what to do next. The decision to select an action will be based on the agent's policy and their evaluation of the current state. The view on a certain state may be altered upon receiving a reinforcement signal that results from performing the action. In other words, agent may change their perspective of how good an action is for a given state when they receive updated signal from the environment.

In the model, banks keep track of two scores of all other banks in order to pick the counterparties of new debts: a size score, S^{size}, and a relationship score, S^{relation}. The size score is calibrated through the comparison of a banks size to the average size of existing counterparties (This is consistent with the preference of banks to build relationships with large banks).

$$\begin{aligned} S_{i,j}^{\text{size}}(t) &= \log A_j(t) - \frac{\sum_{k,k\neq i} \log A_k(t-1)\mathbb{I}_{i,k}(t-1)}{\sum_{k,k\neq i} \mathbb{I}_{i,k}(t-1)} \\ \mathbb{I}_{i,k}(t) &= \begin{cases} 1, \text{ If } i \text{ and } k \text{ are connected on period } t \\ 0, \text{ Otherwise} \end{cases} \end{aligned} \tag{3}$$

where $S_{i,j}^{\text{size}}(t)$ is the size score of bank j evaluated by bank i on the period t, A_j is the total assets of bank j, $\mathbb{I}_{i,k}(t)$ is a binary variable for keeping track previous debt obligations.

The relationship score captures the preference of continuing business with existing counterparties. Therefore, the score will be updated upon receiving the reinforcement signal from the lending and borrowing actions a bank make in each iteration. In the model, temporal different (TD) method is adopted to update the relationship score, and the updating process is explained and illustrated by a simplified stylized network in Section 4.4. In addition, banks also hold private information about their own target ratios and evaluate the status of their balance sheet to help guide the lending and borrowing policy.

4.3. Actions - bank's lending and borrowing policy

In reinforcement learning, the objective of a policy is to map the current state to an action that optimizes the expected sum of discounted utilities. Agents can either adopt a fixed policy to select an action (passive learning) or learn to select the optimal action based on previous reward signals received (active learning). In this model, banks are modeled as passive learning agents that rely on a fixed policy to select their lending and borrowing counterparties. However, banks may learn to select better counterparites by learning and updating the utilities of states.

In the model, banks first check their current balance sheet ratios and the target ratios to establish their demand of lending and borrowing in overnight, short-term, or long-term markets. For example, if bank i's current overnight lending to asset ratio is lower than its target, it will look for a borrower in the overnight market. But if the bank's current overnight borrowing to liability ratio is lower than its target, it will look for a lender in the overnight market. Once the bank reaches all of its targets it will not want to lend or borrow in all the markets. A score-driven process is then applied to select their counterparites for initializing new debts each period.

Targeting borrowing expectation, first each bank would send borrowing request to large banks based on S^{relation} because of higher opportunity to obtain funds. As long as the targets are not met entirely, it sends requests to highest scoring bank. If the targets are still not completed, it goes to the highest S^{size} bank and ask for new debts. This process would continue until the requests are satisfied perfectly or no more fund is available in the market.

Upon receiving a borrowing request, banks go through two key questions: 1) should they provide new debts to borrowers? 2) how much to lend to each requester? Possessing lending preferences, each bank with space in its ON^l, LT^l, or ST^l follows a similar scoring system as described in equation (4). Accordingly, banks go through each requests and make decisions until lending targets are completed or no more request left to fill. However, banks may refuse to accept requests from other banks in the market even though they have capacity, and they utilize an S-shaped function, $p(S_{i,j}^{\text{total}})$, assess the chance that lending bank i would settle new debts to borrowing bank j, and pick partners.

$$S_{i,j}^{\text{total}} = \omega S_{i,j}^{\text{relation}} + (1-\omega) S_{i,j}^{\text{size}} \tag{4}$$

where $s(i,j)$ is the score that borrower i assigns to lender j. It is the weighted average of relationship score and size score of bank j. We set equal weights to these scores so that $\omega = 0.5$.

$$p(S_{i,j}^{\text{total}}) = \frac{1}{1+\exp\left(\alpha + \beta \times S_{i,j}^{\text{total}}\right)} \tag{5}$$

In this function, $p(S_{i,j}^{\text{total}})$ presents the probability that borrower i lends to lender j, and α and β respectively control intercept and slope. α is a real number, and a larger α implies a lower probability of lending to a bank scoring 0. Considering different preferences of large banks and small banks, we set values from uniform distribution $U(-1,1)$ for large banks and values from uniform distribution $U(3,5)$ for small banks. Under this setting, more lending fund flows from large banks to small banks. β is a negative real number, and a larger β means that the probability, $p(S_{i,j}^{\text{total}})$, moves slower from 0 to 1. It also implies a tighter lending policy such that fewer borrowers get debts. We set β to $U(-1.1,-0.9)$ as default values.

Following a uniform distribution, lending banks decide the fraction they want to lend out from their lending pool. The new debt amount is set as the lower value between the one determined by the lending bank and requested by the borrowing bank. This new debt established is observed by both banks as a reward that their bank balance ratios move closer to their respective target ratio. In summary, the flow of the bank's policy can be represented and illustrated by Figure 2.

4.4. Temporal difference learning update

The TD algorithm combines the characteristics of both the Monte Carlo methods and dynamic programming. Similar to Monte Carlo method, TD algorithm is a model-free approach that is able to evaluate the value of a given state by

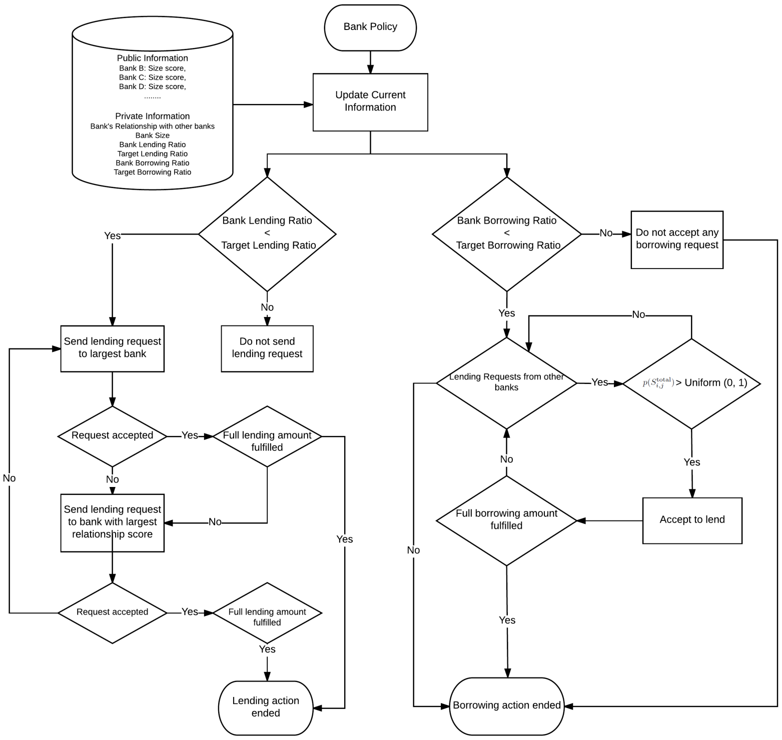

Bank Policy
Public Information
Bank B: Size score,
Bank C: Size score,
Bank D: Size score,
....
Private Information
Bank's Relationship with other banks
Bank Size
Bank Lending Ratio
Target Lending Ratio
Bank Borrowing Ratio
Target Borrowing Ratio
Update Current Information
Bank Lending Ratio < Target Lending Ratio
Yes
No
Bank Borrowing Ratio < Target Borrowing Ratio
No
Do not accept any borrowing request
Yes
Send lending request to largest bank
Do not send lending request
No
Lending Requests from other banks
Yes
$p(S_{i,j}^{\text{total}})$ > Uniform (0, 1)
Request accepted
Yes
Full lending amount fulfilled
No
No
Yes
Send lending request to bank with largest relationship score
No
Yes
No
Full borrowing amount fulfilled
Accept to lend
No
Request accepted
Yes
Full lending amount fulfilled
Yes
Yes
Lending action ended
Borrowing action ended

Figure 2: Bank lending and borrowing policy flowchart

learning directly from past experiences. In addition, it also incorporates the idea of bootstrapping from dynamic programming that the estimation of their update is based on a previous estimate. In this model, we use $TD(0)$ method to allow banks to learn from past lending and borrowing actions. $TD(0)$ is formulated as follows:

$$V(s_t) = V(s_t) + \alpha[r_{t+1} + \gamma V(s_{t+1}) - V(s_t)] \tag{6}$$

where $V(s_t)$ is the estimate of expected sum of discounted rewards at time t, r_{t+1} denotes the observed reward, α is the learning rate, and γ is a discount factor for $V(s_{t+1})$.

In the model, at the beginning of each iteration (a quarter), each bank determines its lending and borrowing actions according to the bank policy described in previous section. Lending and borrowing actions are decided by the size score of each bank and relationship score between two banks. Since size score is determined by the total assets presented in the balance sheet, banks do not update the size score using the temporal difference method. However, since the nature of relationship score is to capture a bank's tendency to keep existing relationships, banks update their relationship score with other banks by using the $TD(0)$ algorithm. The updating process is described below.

When bank i and bank j established a new relationship, relationship score between the two banks, $S_{i,j}^{\text{relation}}$ will be updated using equation (6). At the beginning of each quarter, TD learning is applied to every bank to update their relationship score. In equation (6), V_t denotes the bank's evaluation of its relationship score with all other banks. V_{t+1} is the estimation of the sum of debts established in the future and it can be formulated as equation (7). In this model, banks assume that all debts received from that counterparty equal the last reward observed in the future. Therefore, equation (7) can be expended and simplified to equation (8).

$$V(s_{t+1}) = \sum_{k=0}^{\infty} \gamma^k \times R_{t+k+2} \tag{7}$$

$$\begin{aligned} V(s_{t+1}) &= \sum_{k=0}^{\infty} \gamma^k \times D_{t+k+2} \\ &= (\gamma \times D_{t+2}) + (\gamma \times D_{t+3}) + (\gamma \times D_{t+4}) + ... \\ &= D_{t+1}(1 + \gamma^1 + \gamma^2 + ...) \\ &= \frac{D_{t+1}}{(1-\gamma)} \end{aligned} \tag{8}$$

where D_t is the new debts between two banks in period t.

As a result, the $TD(0)$ updating process of relationship score can be expressed as:

$$S_{i,j}^{\text{relation}}(t) = S_{i,j}^{\text{relation}}(t) + \alpha[D_{t+1} + \gamma \frac{D_{t+1}}{(1-\gamma)} - S_{i,j}^{\text{relation}}(t)] \tag{9}$$

We make further assumption that the learning rate α is equal to $1-\gamma$, resulting in:

$$\begin{aligned} S_{i,j}^{\text{relation}}(t) &= S_{i,j}^{\text{relation}}(t) + \alpha[D_{t+1} + \frac{(1-\alpha)}{(\alpha)}D_{t+1} - S_{i,j}^{\text{relation}}(t)] \\ &= S_{i,j}^{\text{relation}}(t) + D_{t+1} - \alpha S_{i,j}^{\text{relation}}(t) \\ &= (1-\alpha)S_{i,j}^{\text{relation}}(t) + D_{t+1} \end{aligned} \tag{10}$$

4.5. Illustration of TD updating using a stylized network

To further illustrate the updating process, consider a stylized network where there are only 4 banks engaging in lending and borrowing actions. Assume the actions of banks result in a lending/borrowing situation shown in Figure 3, and we only consider the bank As perspective. As banks begin to lend and borrow in attempts to reach the target ratio, they do not have any prior relationship with each other. Therefore, relationship score is initialized to be zero and the bank's decision will based solely on the size score at the beginning of the simulation.

From the perspective of bank A, it accepts to lend 30 units to bank B after going through its policy and the amount of new debts established (30) will be added to the prior relationship score. In the second quarter, bank A accepts lending requests from both bank B and bank D. Given that bank A has prior relationship with bank B, the updated relationship score between the two banks will be $S_{A,D}^{relation} = (1-\alpha)S_{A,D}^{relation} + D_{t+1}$. The new debts established from the second quarter will simply be added to the relationship score between bank A and bank D. Following a similar updating process, the only action bank A took in the third quarter was borrowing 20 units from bank C. Therefore the resulting relationship scores from the perspective of bank A are: $S_{A,B}^{relation} = (1-\alpha)S_{A,B}^{relation}$, $S_{A,C}^{relation} = 20$, and $S_{A,D}^{relation} = (1-\alpha)S_{A,D}^{relation}$.

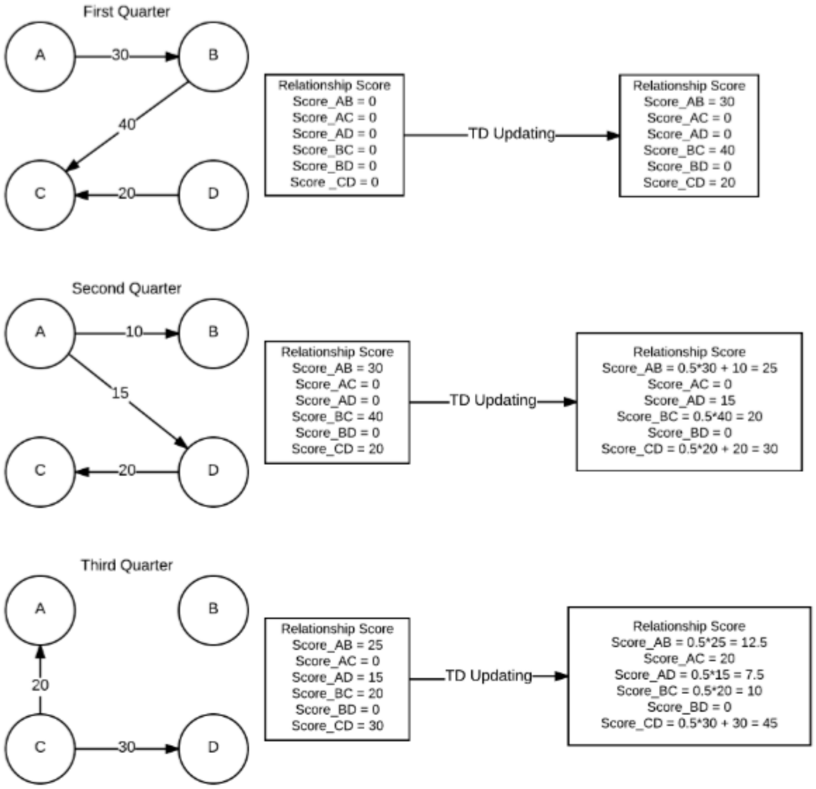


Figure 3: Stylized interbank network TD update illustration

5. Data

Financial reports are one of the key information sources that disclose banks' financial fundamentals and business conditions. The Federal Reserve, FDIC, and OCC require all U.S. regulated banks to submit quarterly reports known as *Federal Financial Institutions Examination Council Reports of Condition and Income*. Those banks include national banks, state member banks and insured state nonmember banks. Similar to regulators that rely on balance sheets to monitor banks' liquidity status and banking system structures, we use balance sheets data from March, 2001 to December, 2014, covering around $10,000$ banks (see Table 2).

Table 2: Description of the bank balance sheet

Assets, A		**Liabilities,** L	
Overnight lending: federal funds, ON^l	Interbank lending	Overnight borrowing: federal funds, ON^b	Interbank borrowing
Short-term lending: federal securities, ST^l		Short-term borrowing: federal securities, ST^b	
Long-term lending: loans due from banks, LT^l		Long-term borrowing: loans due to banks, LT^b	
Cash and balance due, C		Other liabilities, OL	
Other assets, OA		**Equity,** E	

Notes: This description of a banks balance sheet focuses on major bank lending and borrowing channels, i.e. overnight, short-term, and long-term markets. The rest of the balance sheet is condensed into cash or other asset or liabilities. The notations introduced here for aspect of the balance sheet will be used throughout this paper.

In the model, we assume that banks target at a series of balance sheet ratios (see Table 3), associated with banks' decisions of lending and borrowing. These targets guarantee that banks maintain their lending-borrowing preferences for overnight, short-term, and long-term debts. Additionally, we use the equity multiplier,the ratio of its total assets to its equity, to control a bank's expected leverage.

Table 3: Bank balance sheet ratio

Equity Multiplier	$\frac{E_i}{A_i}$
Overnight Lending, Borrowing Ratio	$\frac{ON_i^l}{A_i}$, $\frac{ON_i^b}{L_i}$
Short-term Lending, Borrowing Ratio	$\frac{ST_i^l}{A_i}$, $\frac{ST_i^b}{L_i}$
Long-term Lending, Borrowing Ratio	$\frac{LT_i^l}{A_i}$, $\frac{LT_i^b}{L_i}$

6. Model validation

We run a number of validation exercises to ensure that model can produce inter-bank markets resembling of the real interbank markets. The model is first initialized based on 2001 financial data. The distribution of the four selected ratios is validated according to simulation data of 20 quarters and empirical data from 2001 to 2006. A comparison of the distributions of the four observed versus simulated ratios shows that from a balance sheet perspective, the simulation closely resembles real bank lending and borrowing behaviors. And then we compare the network topology features to those described in the current literature.

6.1. Network properties validation

We validate overnight lending market with the U.S. federal fund market. Soramäki et al. (2006) collected interbank transactions in 2006 from Fedwire and analyzed the empirical network topology. To compare the network structures, we conducted 100 experiments with the same number of agents (see Table 4).

Table 4: Interbank network property comparison

	Num. of Nodes	Average Degree	Clustering Coefficient	Power Law
U.S. Federal Market	6600	15	0.53	2.15
Model (100 runs)	6600	14.78	0.36	2.39

6.2. Convergence of relationship score learning

Banks use temporal difference learning to update their relationship score as they receive reinforcement signals from borrowing and lending actions. In this experiment, we study the effectiveness of applying $TD(0)$ method in a complex interbank network environment. Sutton and Barto (1988) provides a thorough proof that $TD(0)$ method coverages to a optimal evaluation for linear problems. However, effectiveness of applying $TD(0)$ method in complex real-world problems remain unclear. Tesauro (1992) investigates the effectiveness of TD method on complex practical issues such as the game of backgammon and self-play. In this model, we validate the use of temporal difference learning by investigating the convergence of relationship score (the TD updating target) over time.

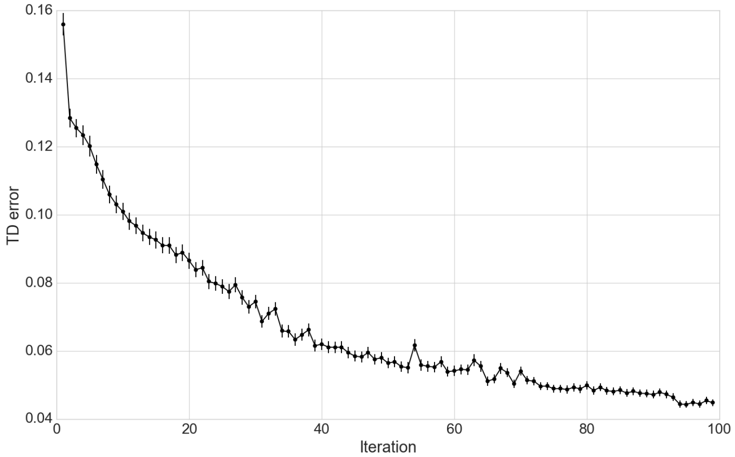


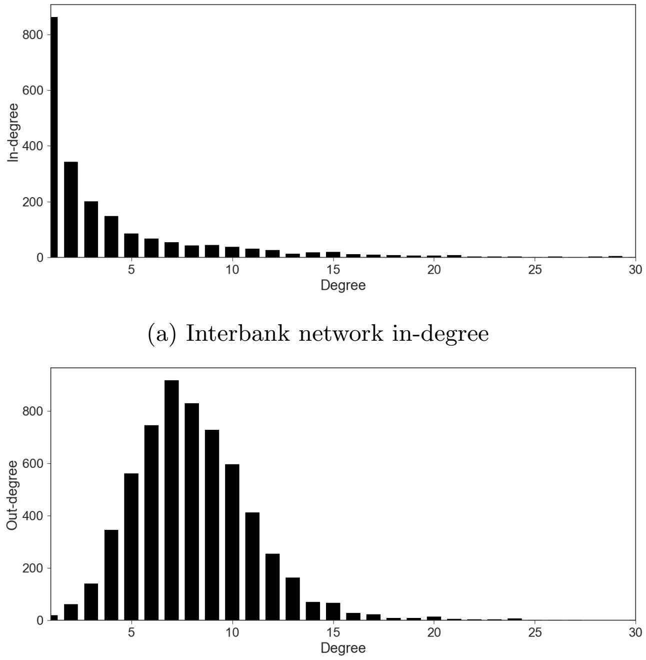
Figure 4: Learning convergence of TD target

The experiment runs for 100 iterations of bank lending and borrowing interactions. At each iteration, the mean squared error of relationship score is recorded and computed for the entire bank population. From figure 4, we can see that the mean squared error of relationship score converges as the system runs for 100 iterations.

7. Experiments and discussions

7.1. Interbank network topologies

Interbank network degree is asymmetric. Both in-degree and out-degree are heavily skewed right, indicating that the majority of bank agents has few established relationships (see Figure 5). Moreover, we observe power-law decaying patter in the in-degree is distribution (see Figure 5a). Obviously, banks prefer to control a lower number of lenders. Over 10% of banks only borrow from one lender. However, it is not common to minimize the number of borrowers due to consideration of risk diversification.



(a) Interbank network in-degree

(b) Interbank network out-degree

Figure 5: Interbank network degree distribution

The multi-agent learning model simulates the banking system dynamics and provides a realistic tool to investigate problematic issues in the interbank lending system. In this experiment, we present the interbank network structure after 50 iterations of the model and discuss the preliminary results that show the differences between a loose lending policy and a risk-adverse lending policy.

We present two network structures that identify the lending and borrowing relationships between banks. Figure 6 shows the interbank network when banks adopt a loose lending policy and Figure 7 shows the network when banks become more risk adverse in accepting lending requests. In both network plots, we labeled the four large bank agents as "1,2,3,4" because empirical studies have suggested that they tend to establish more lending and borrowing relationships

Table 5: Interbank network topology

Overnight	Average degree	Clustering coefficient	Power law	Average path
Loose Policy	15.12	0.35	2.39	2.34
Risk-Adverse Policy	11.51	0.19	2.42	2.66
Short-term	Average degree	Clustering coefficient	Power law	Average path
Loose Policy	1.04	0.43	2.44	2.30
Risk-Adverse Policy	1.04	0.53	2.29	2.21
Long-term	Average degree	Clustering coefficient	Power law	Average path
Loose Policy	2.42	0.40	2.14	2.44
Risk-Adverse Policy	2.42	0.57	2.15	2.28

than other banks (Cocco et al. (2009)). From these results we discover that large banks tend to form their own clusters when they employ a loose lending policy. On the other hand, this phenomenon becomes less obvious when they adopt a more risk-adverse policy.

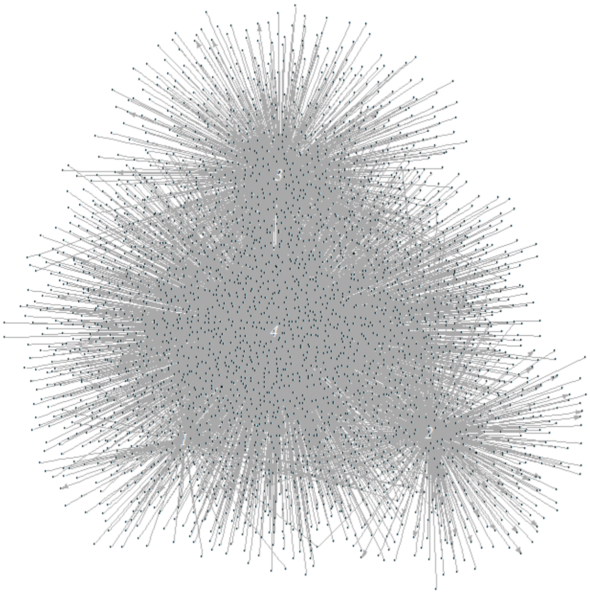


Figure 6: Interbank network formation with a loose lending policy at $\alpha = -1$

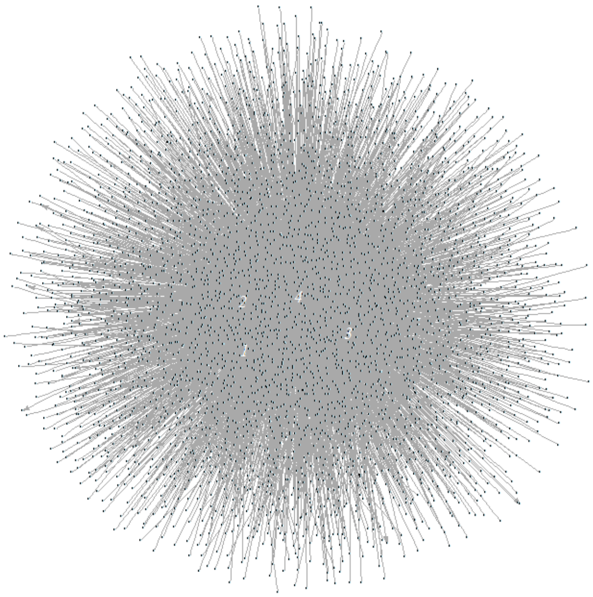
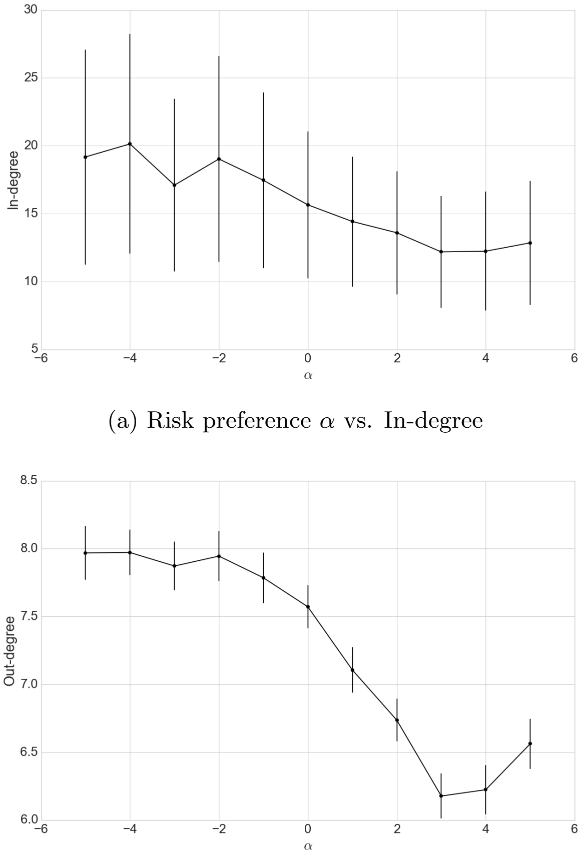


Figure 7: Interbank network formation with a risk-adverse policy at $\alpha = 5$

7.2. Network adaptation and risk preferences

The bank lending policy (equation (5)) shows that α is the risk tolerance parameter. A small α represents higher risk tolerance. Put it differently, when α decreases, banks loosen lending policy and have higher chance to lend out. In this experiment, we study the network property adaptation by altering α from -5 to 5. We explore changes in average degree, clustering coefficient, and average shortest path.

As the model forms a directed network in which edges are pointing from lender to borrower. We can measure how the calibration of α impacts the average number of in-degree (lending) and out-degree (borrowing) relationships. We observe similar patterns for the two measures. The network degree is stable when $\alpha < -1$, however degree decreases with increasing α (see Figure 8). This result confirms that a tightened lending policy reduces the amount of debt, or counterparities, in interbank network. Moreover, the 95% confidence interval of in-degree is much larger than that of out-degree (see Figure 8). These results are consistent with our finding of asymmetric distributions for lending and borrowing. With an increasing α, it becomes harder for banks to find lenders such that the in-degree is decreasing with a tighter confidence interval (see Figure 8a).



(a) Risk preference α vs. In-degree

(b) Risk preference α vs. Out-degree

Figure 8: Risk preference α experiments

The average shortest path, as shown in Figure 9, agrees with the hypothesis that interconnectedness of banks declining as α increases. We observe longer distance among banks when α is positive, indicating that the whole system is less connected.

Inspecting Figure 10, we find the clustering coefficient of the network becomes

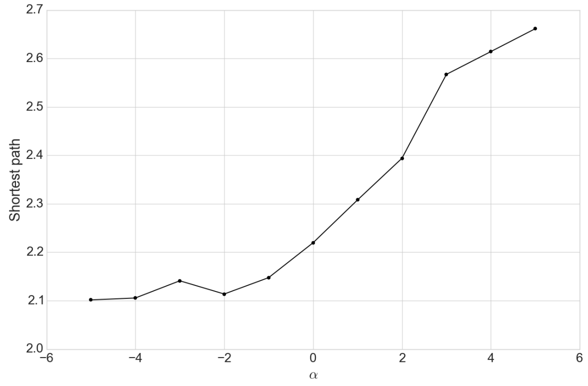


Figure 9: Risk preference α vs. shortest path

less clustered as the risk adverseness of banks increases. This phenomenon shows that when banks become more averse to lending to other banks, the network becomes less clustered thus more spare than complete. This is validated by also a decreased degree and increasing average shortest path, which also indicate a loosely connected network.

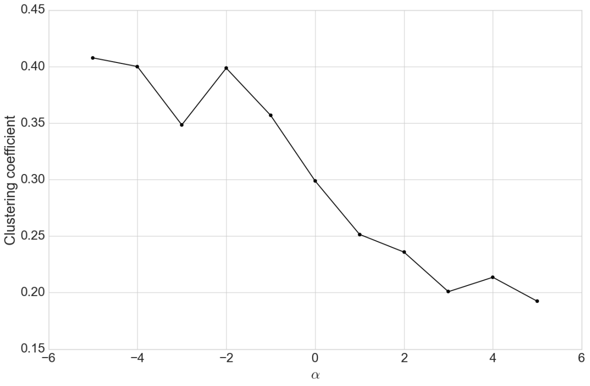


Figure 10: Risk preference α vs. clustering coefficient

8. Conclusion

This study proposes a multi-agent learning model to reconstruct interbank lending networks and examine their dynamics. Alongside with balance sheet ratios, we incorporate a reinforcement learning framework to help guide bank decisions to lend and borrow based on a given set of information. Banks learn and update their choice parameters based on past interaction within the interbank network. Unlike the traditional global optimization methods, our model is able to capture the individual preferences of banks while allowing them to vary their policies to the environmental conditions presented. Thus, providing a analytical framework to consider the implications of macroprudential events/policies on individual preference and overall expectation for how the interbank market mircostructure will evolve.

Two experiments are conducted in this study to investigate the network property adaptation of bank's risk preference in lending capital to other banks upon receiving a lending request. First, we look at two different interbank networks of banks adopting a loose lending policy and a tightened lending policy respectively. We compare the network properties of the modeled interbank lending networks, and show that given a certain level of risk preference, the network degree begins to decline as bank continue to tighten their lending policies.

Secondly, we examine the interconnectedness of bank agents by computing the clustering coefficients and average shortest path amount banks. This suggests that as banks tighten their lending policies, the network becomes more sparse and thus more concentrated on which banks lend to. As a result, the interbank network is less at risk to concerns of contagion; though banks are less likely to find new counterparties, they are generally more capable of sustaining stresses, thus demonstrating how individual choices presents beneficial overall market structure condition due to the adaptive learning capabilities of the banks. This finding seems to support the general observation during the financial crisis.

Lastly, the methodology proposed in this study offers a new perspective on how one can utilize reinforcement learning to model bank agents in an interbank lending setting with real data. Combined with the use of financial data, such as that used in this study, we provide guidance on how central banks and regulators can consider building more functional models for examining the interbank market. Allowing lessons to be draw on how policy decisions can lead to new interbank lending market dynamics, can impact bank behavior and choice, and can influence shifts in the markets preferred mircostructure. Future research might include exploring how banks dynamically select the optimal policy based on the observed states, and also consider how monetary policy influences competition in interbank lending.

Reference

Acemoglu, D., Ozdaglar, A. E., and Tahbaz-Salehi, A. (2015). Systemic risk in endogenous financial networks. *Available at SSRN 2553900.*

Afonso, G., Kovner, A., and Schoar, A. (2011). Stressed, not frozen: The federal funds market in the financial crisis. *The Journal of Finance*, 66(4):1109–1139.

Allen, F. and Gale, D. (2000). financial contagion. *The Journal of Political Economy*, 108(1):1–33.

Anand, K., Craig, B., and Von Peter, G. (2015). Filling in the blanks: Network structure and interbank contagion. *Quantitative Finance*, 15(4):625–636.

Basel Committee Supervision on Banking (2015). *Making supervisory stress tests more macroprudential : Considering liquidity and solvency interactions and systemic risk.* Number November.

Berrospide, J. M. (2012). Bank liquidity hoarding and the financial crisis: an empirical evaluation.

Boss, M., Elsinger, H., Summer, M., and Thurner 4, S. (2004). Network topology of the interbank market. *Quantitative Finance*, 4(6):677–684.

Cocco, J. F., Gomes, F. J., and Martins, N. C. (2009). Lending relationships in the interbank market. *Journal of Financial Intermediation*, 18(1):24–48.

Cont, R., Moussa, A., and Santos, E. B. (2013). Network Structure and Systemic Risk in Banking Systems. In *Handbook on Systemic Risk*, pages 327–368.

Craig, B. and Von Peter, G. (2014). Interbank tiering and money center banks. *Journal of Financial Intermediation*, 23(3):322–347.

Eisenberg, L. and Noe, T. (2001). Systemic risk in financial systems. *Management Science*, 47(2):236–249.

Elliott, M., Golub, B., and Jackson, M. O. (2014). Financial networks and contagion. *American Economic Review*, 104(10):3115–3153.

Georg, C. P. (2013). The effect of the interbank network structure on contagion and common shocks. *Journal of Banking and Finance*, 37(7):2216–2228.

Gilbert, N. and Terna, P. (2000). How to build and use agent-based models in social science. *Mind & Society*, 1(1):57–72.

Gobe, D. and Sunder, S. (1993). Allocative Efficiency of Markets with Zero-Intelligence Traders: Market as a Partial Substitute for Individual Rationality. *The Journal of Political Economy*, 101(1):119–137.

Gofman, M. (2016). Efficiency and stability of a financial architecture with too-interconnected-to-fail institutions. *Journal of Financial Economics*.

Iori, G. and Gabbi, G. (2008). A Network Analysis of the Italian Overnight Money Market. *Journal of Economic Dynamics and Control*, 32(1):259–278.

Iori, G., Jafarey, S., and Padilla, F. G. (2006). Systemic risk on the interbank market. *Journal of Economic Behavior & Organization*, 61(4):525–542.

Jackson, M. O. and Wolinsky, A. (1996). A strategic model of social and economic networks. *Journal of Economic Theory*, 71(1):44–74.

Kok, C. and Montagna, M. (2013). Multi-layered Interbank Model for Assessing Systemic Risk.

Ladley, D. (2013). Contagion and risk-sharing on the inter-bank market. *Journal of Economic Dynamics and Control*, 37(7):1384–1400.

Lux, T. (2015). Emergence of a core-periphery structure in a simple dynamic model of the interbank market. *Journal of Economic Dynamics and Control*, 52(September 2013):A11–A23.

Macal, C. M. and North, M. J. (2010). Tutorial on agent-based modelling and simulation. *Journal of Simulation*, 4(3):151–162.

Macy, M. W. and Willer, R. (2002). From Factors to Factors: Computational Sociology and Agent-Based Modeling. *Annual Review of Sociology*, 28(1):143–166.

Othman, A. (2008). Zero-Intelligence Agents in Prediction Markets. *Proceedings of Seventh International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2008)*, pages 879–886.

Ramanauskas, T. (2008). Agent-Based Financial Modelling As an Alternative To the Standard Representative-Agent Paradigm. *Monetary Studies (Bank of Lithuania)*, 12(2):5–21.

Roukny, T., Georg, C.-P., and Battiston, S. (2014). A network analysis of the evolution of the german interbank market. *Deutsche Bundesbank. Discussion Paper*, (22/2014).

Soramäki, K., Bech, M. L., Arnold, J., and Robert, J. (2006). Glass, and walter e. beyeler," the topology of interbank payment flows," federal reserve bank of new york. Technical report, Staff Report.

Sutton, R. S. and Barto, A. G. (1988). Reinforcement Learning: An Introduction.

Tesauro, G. (1992). Practical Issues in Temporal Difference Learning . 277:257–277.

Upper, C. and Worms, A. (2004). Estimating bilateral exposures in the German interbank market: Is there a danger of contagion? *European Economic Review*, 48(4):827–849.