PLOS ONE

RESEARCH ARTICLE

Intervention on default contagion under partial information in a financial network

Yang Xu *

Industrial Engineering and Operations Research, University of California at Berkeley, Berkeley, CA, United States of America

* yangxu@berkeley.edu

Abstract

We study the optimal interventions of a regulator (a central bank or government) on the illiquidity default contagion process in a large, heterogeneous, unsecured interbank lending market. The regulator has only partial information on the interbank connections and aims to minimize the fraction of final defaults with minimal interventions. We derive the analytical results of the asymptotic optimal intervention policy and the asymptotic magnitude of default contagion in terms of the network characteristics. We extend the results of Amini, Cont and Minca's work to incorporate interventions and adopt the dynamics of Amini, Minca and Sulem's model to build heterogeneous networks with degree sequences and initial equity levels drawn from arbitrary distributions. Our results generate insights that the optimal intervention policy is "monotonic" in terms of the intervention cost, the closeness to invulnerability and connectivity. The regulator should prioritize interventions on banks that are systematically important or close to invulnerability. Moreover, the regulator should keep intervening on a bank once having intervened on it. Our simulation results show a good agreement with the theoretical results.

OPEN ACCESS

Citation: Xu Y (2019) Intervention on default contagion under partial information in a financial network. PLoS ONE 14(1): e0209819. https://doi.org/10.1371/journal.pone.0209819

Editor: Fenghua Wen, Central South University, CHINA

Received: November 15, 2017

Accepted: December 8, 2018

Published: January 15, 2019

Copyright: © 2019 Yang Xu. This is an open access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: All relevant data are within the paper and its Supporting Information files.

Funding: The author received no specific funding for this work.

Competing interests: The author has declared that no competing interests exist.

Introduction

Introduction

The systemic risk in a financial network has been drawing more and more interest of regulators and researchers, especially after the Asian financial crisis in the late 1990's and the more recent economic recession during 2007-2009. Financial institutions (hereafter, banks) are connected to form an interbank network to allow liquidity reallocation between the banks in that banks with liquidity surpluses can lend to banks with liquidity deficits. However, the interbank network may also introduce aggregate liquidity shortage and default contagion. Liquidity shock such as a run will cause some banks to default, leading to losses of their creditor banks through interbank connections, which may in turn result in losses of their creditors. In the financial crisis of 2007-2009 the interbank market dysfunctioned because the market participants perceived heightened counterparty risk and liquidity risk [1], and the severe reduction in transaction volume was a major contributing factor to the collapse of many banks. When the interbank market is stressed or freezes, the central bank as the lender of last resort has to

provide extensive short term liquidity support. For example, the Federal Reserve in the US established many facilities, including the traditional discount window, the Term Auction Facility (TAF), Primary Dealer Credit Facility (PDCF), and Term Securities Lending Facility (TSLF). The central bank or government also recapitalize the banks and provide risk capital in the form of a bailout. Naturally, we ask the following questions: Should the regulator (the central bank or government) intervene if the bankruptcy of one or several banks has occurred or is imminent? What is the optimal intervention policy of the regulator based on the measurable features of the network and the banks, such as the degrees of connectivity and the levels of capitalization? How much improvement can the optimal strategy achieve regarding the fraction of market protected from defaults?

To answer these questions we study the uncollateralized interbank funding market where the majority of interbank loans are overnight. The connections are constantly changing so the regulator may not know exactly all the connections over time. After some liquidity incident, e.g. runs on a few banks, some banks default, which initiates the default contagion process. During the process, the regulator has to intervene on the defaulted banks to prevent the contagion from spreading. Due to the system panics, no banks want to lend new loans to other banks but meanwhile they are still obliged to pay back their current loans which are due in the time frame of the model. So it is reasonable to fix the in and out degrees of banks and assume all the connections do not change any more after the inception of default contagion. Theoretically, the regulator would be able to find out the connections between the banks by communicating with the banks; however, the contagion process happens so fast that the regulator may not have ample time to find out the connections between the banks. In other words, the regulator will have to intervene while all the connections between the banks are unknown. So the regulator has only partial information because in the beginning the regulator only knows the initially defaulted banks and the magnitude of connections of each bank but the connections between the banks are unknown. Every time a default occurs the regulator learns the connections of the defaulted bank (i.e. the banks that are affected by the default), represented by the connections being revealed after the default.

We set up a probability space under which the financial network is generated by a uniform matching of the in and out degrees (a configuration model). A directed link in the network represents one unit of loan. In the following we may use "bank" and "node" interchangeably. Before an external shock to the system, each node has a positive equity level (the difference of total assets and liabilities), which indicates the number of defaulted loans due to the default of its debtors a node can withstand before it defaults. In other words, it is the "distance to default". After an external shock, some nodes in the system default initially and we set their equity levels to zero. We adopt the dynamics of the model in [2, 3]. When a node defaults, it defaults on all of its loans. We assume a zero recovery rate of the loan, i.e. the creditor receives zero value from the loan, which is the most realistic assumption for short term default as suggested in [4]. We assume there is a time span between a node's default and the time its creditor records the loan as a loss (by writing down the loan from its balance sheet). We model this time span by independently and identically distributed exponential random variables. After the affected node records the defaulted loan, it may request the regulator for interventions. If the regulator decides to intervene by replacing the defaulted loan or by infusing one unit of equity, the equity level of the affected node will stay the same, otherwise its equity level will decrease by one. Once the equity level reaches zero, the node defaults. We assume that the once a bank has defaulted it cannot become liquid again within the time horizon of the model because it is very unlikely for a bank that has declared default to gain enough capital in the short term considered in the model.

We emphasize that one essential feature of the model is partial information. Because it takes time and effort for the regulator to find out the exact connections in the interbank market and the default contagion process may progress very rapidly, the regulator may have to intervene even before it is able to figure out the connections in order to take early actions and save the maximum amount of banks from bankruptcy. During the contagion process, the regulator only knows the default set (the set of defaulted nodes) with some out-links revealed, but meanwhile other out-links remain hidden until the affected nodes record the defaulted loans. More generally, unlike the complete information assumption in other theoretical models, the partial information setting aligns with the reality better, as pointed out by [5]: "*Interbank exposure data are never publicly available, and in many countries nonexistent even for central regulators*".

Our methodology is illustrated in Fig 1. The regulator's goal is to minimize the number of final defaulted nodes with the minimum number of interventions. Thus we obtain a stochastic control problem $\min_{\mu_n} \mathrm{obj}_n(Gr_n, \mu_n)$ where the objective function depends on the graph Gr_n and the intervention sequence μ_n, shown in (1). We aim to solve it for the optimal intervention sequence μ_n^* and thus obtain the optimal objective function value $\mathrm{obj}_n(Gr_n, \mu_n^*)$, shown in (2). However, solving the problem with the usual dynamic programming approach will incur intractability problem because of the fast expansion of the state space as pointed out in [3], especially for a heterogeneous network. We take an alternative approach based on the fact that under some regularity conditions, the objective function converges as $n \to \infty$. We solve the asymptotic optimal control problem $\min_\mu \mathrm{obj}(Gr, \mu)$ in (3) where obj, Gr and μ are the limit forms of obj_n, Gr_n and μ_n, respectively, and obtain the optimal intervention μ^* and the objective function $\mathrm{obj}(Gr, \mu^*)$ in (4) which allow us to construct the optimal intervention sequence μ_n^* for a finite n through μ^* and approximate $\mathrm{obj}_n(Gr_n, \mu_n^*)$ with $\mathrm{obj}(Gr, \mu^*)$. Our results of the numerical experiments validate the approximation for networks with sizes close to the real financial networks.

Relations to previous literature

Our work is closely related to the current literature on the role of the central bank as the lender of last resort, including providing liquidity by a loan, recapitalizing the banks and bank bailout. These studies differ in their perspectives and focuses in their models. The influencing Diamond & Dybvig model [6] about the market panics and bank runs has two Nash equilibria: depositors withdraw only for their real expenditure needs or a bank run. [7] extend the Diamond & Dybvig model to a financial network of four banks to study the default contagion. The interbank network is formed to allocate liquidity among the banks to satisfy regional liquidity demands. In a complete market where the banks exchange deposits or in a disconnected market no contagion occurs while in an incomplete market where the banks do not exchange deposits with all other banks, high connectivity may entail contagion. The role of the central bank is thus to complete the market. [8] introduce a similar model but assume the depositors are uncertain about where they have to consume. In their model the central bank acts as a "crisis manager": when a bank is to be liquidated, the central bank has to organize the bypass of the defaulting bank in the payment network and provide liquidity to the banks that depend on the defaulting bank. [9] consider that the fire sales of the banking assets occur when a large number of banks default and investors outside the banking sector who are inefficient users may end up purchasing the liquidated assets. To avoid the allocation inefficiency, the regulator may bail out the banks directly or provides liquidity to surviving banks to purchase defaulting banks. [10] argue that the government should bail out banks in distress because it can provide liquidity more efficiently than private investors. [11] consider three forms of

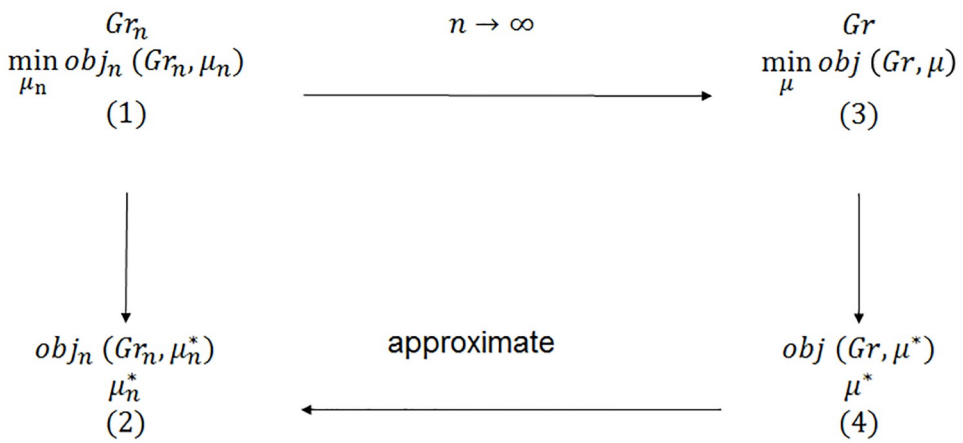

Fig 1. The methodology: Approximation of the finite network with the infinite network.

https://doi.org/10.1371/journal.pone.0209819.g001

regulator interventions: buying equity, purchasing assets and providing debt guarantee to alleviate debt overhang in a financial market but the regulator has limited information and resources. All the works discuss the optimal interventions based on equilibrium analysis. In contrary, our model emphasizes the aspect of the interbank market that it is a complex network and focus on how the regulator should make intervention decisions under the network dynamics. Moreover, some papers study the related problem about the banks in the interbank network bailing each other out, such as [12] and [13].

In addition to the theoretical studies, empirical studies abound. [14] analyze the data on the interbank transactions derived from the main euro area payment system and find that the European Central Bank took the role of the overnight unsecured interbank market in liquidity provision to the banks during the global financial crisis in 2008-2010. [15] analyze the daily transaction data and find that in 2008 counterparty risk plays a more important role than liquidity hoarding in reducing liquidity and increasing the cost of finance in the federal funds market in the US. By analyzing supervisory data of Germany, [16] find that regulatory interventions decrease liquidity creation while capital support does not affect it. [17] discuss the relations between liquidity regulations and the lender of last resort practice and argue that they are complementary rather than conflicting tools.

Our work is also related to the strand of literature on systemic risk and default contagion in the financial networks without considering regulator interventions. These works focus on understanding the dependence of the default contagion on various features of the financial network and the banks within it, including the degrees of connectivity, the equity levels and so on. Similarly, there are mainly two types of literature: empirical and theoretical. The empirical studies conduct statistical analyses on the interbank markets using data on interbank lending as far as they are available and provide an overview of the structural characteristics of the interbank network in different countries ([4, 18, 19]). The theoretical studies model the financial network with network models but differ in their assumptions about the network structure and approaches: some focus on "stylized" networks whose structures are hypothetical ([7, 20, 21, 22]) while others rely on simulations ([4, 23]). Among them, [24] and [25] propose random network models that allow more realistic and heterogeneous structures. [26] survey theoretical works on contagion and systemic risk in financial networks and categorize them according to different topics including network connectivity, bank heterogeneity, uncertainty in financial markets, and portfolio composition of the banks.

In regards to section Introduction, this paper is closely related to [24] and [2, 3]. [24] study the magnitude of default contagion in a heterogeneous network with given degree sequence and arbitrary distribution of weights and derive the analytical expressions of the asymptotic fractions of defaults in terms of the network characteristics. Our work incorporates interventions into a model proven to be equivalent to theirs. Thus if there are no interventions, the asymptotic fraction of final defaults will be the same as in [24]. [2] consider a stylized core-periphery financial network as an intermediary to provide liquidity to fund projects in outside economy but it may also incur contagion when the banks hoard liquidity. The regulator intervenes by providing loans to defaulting banks. [3] consider a similar core-periphery model where the regulator intervenes by injecting equity. We adopt the dynamics of their model that constructs the default set under interventions through a configuration model because the configuration model can be adapted to the contagion process [27]. But we differ from [2, 3] in two important ways. First, our model is a more general heterogeneous random network with degree sequences and initial equity levels drawn from arbitrary distributions. More importantly, [2, 3] focus on the benefits and costs of the connectivity in the presence of the regulator and draw conclusions mainly from numerical studies, while we focus on the optimal intervention policy and its relations to the network characteristics. Mathematically we have successfully addressed two major mathematical difficulties arising from considering interventions on a general complex network: Interventions introduce discontinuity into the asymptotic process thus the main supporting theorem used in [24] is no longer applicable directly; moreover, the high dimensional optimal control problem we obtain later is well known among control theorists difficult to solve, especially analytically. We give analytical formulations for the asymptotic optimal intervention policy as well as the asymptotic number of interventions and final defaulted banks. The asymptotic results provide a good approximation to real financial networks, which are heterogeneous and have several hundred to a thousand of banks thanks to the fast convergence behavior of our results.

Contributions

The main contribution is that we have proposed a new approach to determine the optimal intervention strategy on contagions in a large and heterogeneous financial network-to be specific, we derive the asymptotic optimal intervention strategy as the size of the network tends to infinity and then show that it is a good approximation of the optimal intervention strategy for a real financial network. This new approach has the advantage of avoiding the stylized model in the equilibrium analysis or the intractability of the dynamic programming approach in previous literature and enabling us to obtain analytical results. In light of this, we derive rigorous asymptotic results of the optimal strategy for the regulator and the fraction of final defaulted banks under the dynamics of the default contagion process in a heterogeneous network with a degree sequence and initial equity levels that can be drawn from arbitrary distributions. The analytical expressions are presented in terms of measurable features of the network. For convergence of the results, we assume the network is sparse as in [24] which is supported by the empirical studies of real financial networks ([28, 29]).

The key insights of our findings of the optimal strategies are summarized in the following. We should only consider intervening on a bank when it records the loss of a defaulted loan and is very close to default. The optimal intervention policy depends strongly on the intervention cost. The smaller the intervention cost, the more interventions are implemented. Moreover, the optimal intervention policy is "monotonic" with respect to the measurable features of the network. We should not intervene on the banks with out degrees in a certain range regardless of their other features; for those banks worth interventions, the larger the sum of initial

equity and accumulative interventions received, the earlier we should begin intervention on them; the time to start intervention on a node is also "monotonic" in its in and out degrees. Interestingly, once we begin intervening on a node, we keep intervening on it every time it records the loss of a defaulted loan. By comparing the fractions of final defaults under no interventions and the optimal intervention policy, we are able to quantify the improvement made by interventions in terms of the network features. This gives guidance for the maximum impact the regulator can have to offset the effects of default contagion.

The paper is organized as follows. We set up the model and introduce the stochastic control problem (SCP) in section Model description and dynamics. In section The asymptotic control problem we formulate the asymptotic control problem Eq (9) that gives the limit for the objective function of the SCP as the size of the network goes to infinity and present the necessary conditions for the optimal intervention policy, which lead to the main theorems. In section Numerical experiments we show the results of the numerical experiments to validate the approximation of Eq (4) by Eq (9). We present in Appendix A: Proofs all the proofs and in Appendix B: Wormald's theorem and Appendix C: Extended pontryagin maximum principle the two theorems used in the proofs as well as a list of notations in Appendix D: Preliminary list of notations.

Model description and dynamics

Basic setup

We consider default contagion in an unsecured interbank lending market under short term illiquidity risk based on the model of [2]. Due to the system panics, no banks want to lend new loans to other banks but meanwhile they are still obliged to pay back their current loans which are due in the time frame of the model, so we fix the in and out degrees of banks in the network, denoted as $(d^-(v), d^+(v))_{v\in[n]}$, where $[n] = \{1, \ldots, n\}$ the set of nodes. Let $m = \sum_{v\in[n]} d^-(v) = \sum_{v\in[n]} d^+(v)$.

Then we model the financial network with prescribed degree sequence $(d^-(v), d^+(v))_{v\in[n]}$ as an unweighted directed network $([n], \mathcal{E}_n)$, where $\mathcal{E}_n$ denotes the set of links. A directed link $(v, w) \in \mathcal{E}_n$ represents v borrows a unit of loan from w, i.e. v is obliged to repay w one unit of loan. We allow multiple loans to exist between two nodes and also self links representing the internal loans between different departments of the bank.

Now we set up a probability space $(G_{n,m}, \mathbb{P})$ where $G_{n,m}$ is the set of networks on n nodes with at most m directed links. So the random financial network with m directed links lives in this probability space and under $\mathbb{P}$ the law of the random link set $\mathcal{E}_n$ is determined as follows. We start with n unconnected nodes and assign node v $d^-(v)$ in-half-links and $d^+(v)$ out-half-links. An in-half-link represents an offer of a loan and an out-half-link a demand for a loan. Then the m in-half-links and m out-half-links are matched uniformly so that the borrowers and lenders are determined. The resulting random network is called the *configuration model*.

The uniform matching of the in and out-half-links allows us to construct the random network sequentially: at every step we can choose any out half-link by any rule and choose the in-half-link uniformly over all unconnected in-half-links to form a directed link. This is because conditional on any subset of connected links, the unconnected links also follow the uniform distribution. Moreover, the conditional law of unconnected links only depends on the number of connected links, not the matching history. Additionally we can restrict the matching to choosing only the out-half-links from the defaulted nodes so that we can model the development of the default set with their revealed out-links.

Remark 1. As a result of the uniform connection of in and out-half-links, a node gets selected with the probability proportional to the number of its unconnected in-half-links. The

rationale is that when the regulator searches for the lender of a defaulting loan, a bank that lends out more loans to other banks is more likely to be the lender and be affected by the defaulting loan.

Then we endow a node $v \in [n]$ with its initial equity level $e(v) \in \mathbb{N}_0 := \{0, 1, 2, \ldots\}$ which represents the number of defaulted loans v can tolerate until v defaults, so it is the "distance to default". Next after the system receives some external shock, some nodes default and the system begins to evolve. Define time 0 as the time right after the shock. Let $(\mathcal{G}_k)_{0 \le k \le m}$ be the filtration for the probability space $(G_{n,m}, \mathbb{P})$ which models the arrival of new information, i.e. the revealed link at each step. Because this implies that the the remaining equity of the selected node will decrease by one, $(\mathcal{G}_k)_{0 \le k \le m}$ also models the default contagion at the same time. Note in the following the network with the set of revealed links evolves in the space $G_{n,m}$ as the result of the contagion process.

Initial condition. Define the set of initially defaulted nodes $\mathcal{D}_0 := \{v \in [n] : e(v) \le 0\}$ and the σ-algebra representing the information available initially $\mathcal{G}_0 := \sigma((d^-(v), d^+(v), e(v))_{v \in [n]})$. Given the degree pairs for all nodes $(d^-(v), d^+(v))_{v \in [n]}$, envision for a node $v \in [n]$, there are $d^-(v)$ in-half-links each representing a loan another node is obliged to pay v and $d^+(v)$ out-half-links each representing a loan v is obliged to pay another node. Let c_k^v be the sum of initial equity and accumulative number of interventions on node v and l_k^v be the number of revealed in-links of node v at step k, so $c_0^v = e(v)$, $l_0^v = 0$. An example of the initial condition of a four node network is illustrated in Fig 2 with $\mathcal{D}_0 = \{1, 2\}$.

Dynamics. We adopt the dynamics from [2]. At the kth step for $k \in [1, m]$, if the out-links of nodes in $\mathcal{D}_{k-1}$ have not all been revealed, then the new link is revealed following the rule: an out-half-link of any node in $\mathcal{D}_{k-1}$ is picked by any rule and then connected uniformly to another unconnected in-half-link. Let (V_k, W_k) be a pair of random variables denoting the link from node V_k to node W_k being *revealed* at step k. We call W_k is *selected* at step k. Assume $(V_k, W_k) = (v, w)$, then the uniformity in connecting the half-links leads to the probability of w being selected conditional on $\mathcal{G}_{k-1}$ as

$$\begin{aligned} &\mathbb{P}(W_k = w | \mathcal{G}_{k-1}) \\ &= \frac{\text{number of } w\text{'s unrevealed in} - \text{links at } k-1}{\text{total number of unrevealed in} - \text{links at } k-1} \\ &= \frac{d^-(w) - l_{k-1}^w}{m - (k-1)}. \end{aligned} \tag{1}$$

So a node is selected with the probability proportional to the number of its unrevealed (unconnected) in-half-links. After a directed link (v, w) is revealed, then proceed with the following steps:

- Update $\mathcal{G}_k = \sigma(\mathcal{G}_{k-1} \cup \{(v, w)\})$.
- Update the number of revealed out-links: $l_k^w = l_{k-1}^w + 1$ and $l_k^\eta = l_{k-1}^\eta$ for $\eta \ne w$.
- Determine the intervention $\mu_k \in \{0, 1\}$ $\mathcal{G}_k$ measurable at step k for the selected node w. Note that $c_k^\eta \le l_k^\eta$ indicates that the node η has defaulted by step k. Because we do not intervene on defaulted node, $\mu_k = 0$ if $c_{k-1}^w \le l_{k-1}^w$.
- Update $c_k^w = c_{k-1}^w + \mu_k^w$, otherwise $c_k^\eta = c_{k-1}^\eta$ for $\eta \ne w$.
- Update the default set $\mathcal{D}_k$. If $c_k^w \le l_k^w$ and $w \notin \mathcal{D}_{k-1}$, then $\mathcal{D}_k = \mathcal{D}_{k-1} \cup \{w\}$, otherwise $\mathcal{D}_k = \mathcal{D}_{k-1}$.

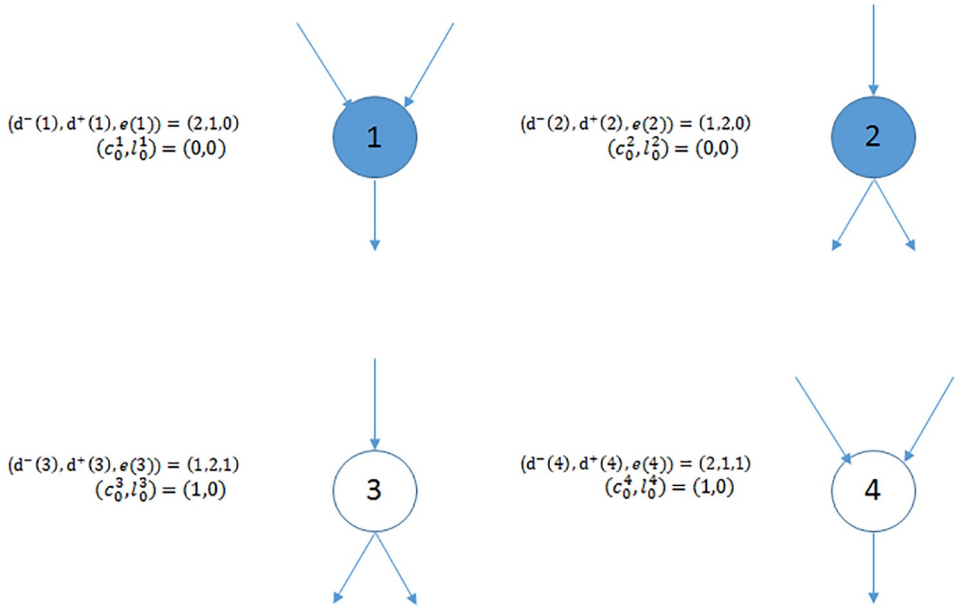

Fig 2. Financial network before default contagion occurs. $(d^-(v), d^+(v))$ are the degrees and $e(v)$ is the initial equity of node v. The nodes in the initial default set $\mathcal{D}_0 = \{1, 2\}$ are marked in blue.

https://doi.org/10.1371/journal.pone.0209819.g002

If all out-links from $\mathcal{D}_k$ have been revealed, the process ends and let the process end time be $T_n = k$, otherwise repeat the process. Define D_{T_n} as the number of defaulted nodes by the process end time T_n. In Fig 3 we show the first step of the dynamics for the network in Fig 2. The link (1, 4) is revealed (connected) and node 4 is selected with probability $\frac{2}{6}$. Because the node 4 is liquid, the regulator needs to decide whether to intervene. Node 4 remains liquid if it receives one intervention, or it will default and be included in the default set.

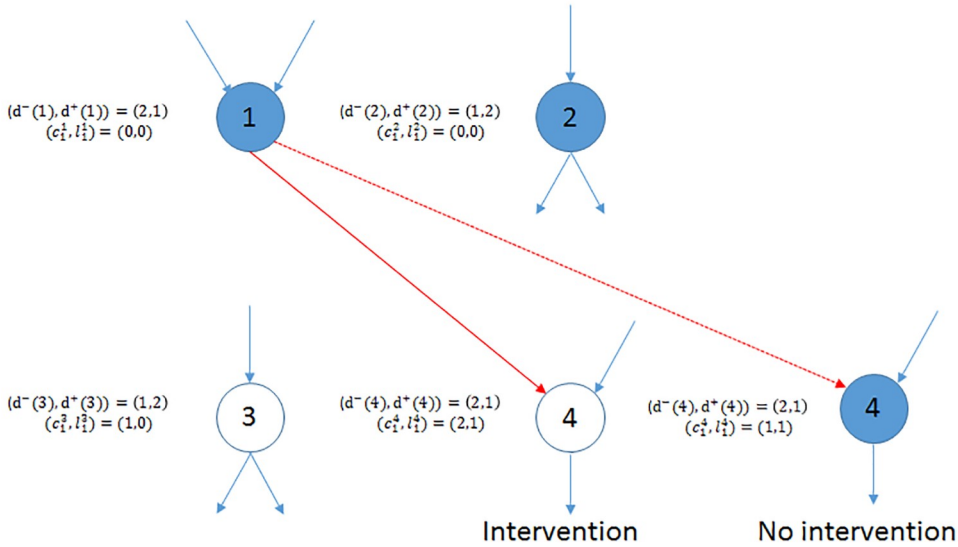

Fig 3. Dynamics at step one. Node 4 is selected with probability $\frac{2}{6}$ and a link between 1 and 4 is revealed. If one intervention is applied, node 4 remains liquid, otherwise its equity level will decrease by one and node 4 defaults.

https://doi.org/10.1371/journal.pone.0209819.g003

Remark 2. The notion of partial information is reflected by the fact that there are always unrevealed links before the process ends and the regulator cannot make decisions depending on the knowledge of it. The regulator can only be certain that if the remaining equity is nonpositive ($c_k^\eta \leq l_k^\eta$) then the node η has defaulted at step k. On the other hand, every out-half-link from the default set represents a defaulted loan which is possible to impact every node in the network at a later time. So all the current liquid vulnerable nodes are subject to default at a later time and the regulator should take this into account when making intervention decisions at each step.

Let R_k be the accumulative number of interventions by step k and $D_k = |\mathcal{D}_k|$ be the number of defaults at step k. The regulator aims to minimize the number of defaulted nodes by T_n with the minimum amount of interventions, so we define the objective function as a linear combination of the (scaled) number of interventions and defaults by the end of the process T_n as

$$J_n = \mathbb{E}(K\frac{R_{T_n}}{n} + \frac{D_{T_n}}{n}|\mathcal{G}_0), \tag{2}$$

where $K > 0$ is the relative "cost" of an intervention. Further by the definition of $c_{T_n}^v$ and noting that a node defaults at last if $c_{T_n}^v \leq l_{T_n}^v$, i.e. the number of defaulted loans exceeds the total of the initial equity level and the number of interventions received by T_n, we can express R_{T_n} and D_{T_n} as

$$\begin{aligned} R_{T_n} &= \sum_{v\in[n]}(c_{T_n}^v - e(v)), \\ D_{T_n} &= \sum_{v\in[n]}\mathbb{1}_{(c_{T_n}^v \leq l_{T_n}^v)}. \end{aligned} \tag{3}$$

Now we define the stochastic optimal control problem as

$$\min_{\mu\in\mathbb{U}} J_n, \tag{4}$$

where $\mu = (\mu_k)_{1\leq k\leq m}$, $\mu_k \in \{0, 1\}$ and $\mathbb{U}$ contains all $(\mathcal{G}_k)_{0\leq k\leq m}$ adapted process μ.

The asymptotic control problem

Assumptions and definitions

We assume that a bank cannot become liquid again once it has defaulted, thus we cannot save defaulted banks. This assumption is reasonable in the setting of default contagion in a stressed network and a short time window. Nor do we intervene on invulnerable nodes, because they never default but intervening on them will only prevent us from saving the banks that are very close to default especially when the interventions are costly.

In the model description we only intervene on the node that is selected at each step. Now we show that even if the regulator intervenes on multiple nodes and applies more than one unit of credit every time, it will not be better.

Proposition 1. For the stochastic control problem Eq (4), we only need to consider intervening on a node that, when selected, has only one unit of equity remaining.

We see proposition 1 implies that it is never optimal to intervene on a node if it is not selected or has more than one unit of equity remaining when selected. Let (i, j, c, l) be the state of a node, meaning it has the in and out degree (i, j), sum of the initial equity and the number of interventions c and l revealed in-links. Note that by definition $l \leq i$. We characterize nodes with states because nodes with the same state have the same probability of being selected at each step and are statistically the same in influencing other nodes. Note in particular:

1. $c = 0$ denotes that the node has defaulted initially.
2. $c - l$ denotes the remaining equity or "distance to default", i.e. the number of times of being selected before a node defaults without interventions. Thus $c \leq l$ means the node has defaulted.
3. Because $l \leq i$ by definition, $i < c$ implies that a node is invulnerable, i.e. even all loans lent out to the counterparties are written down from the balance sheet, the node still has positive remaining equity. On the contrary, $0 < c \leq i$ denotes the node has the possibility to default, i.e. vulnerable.
4. In the beginning of the contagion process, all nodes have $l = 0$, i.e. are in states of the form $(i, j, c, 0)$.

Then we define the state of the system at each step. Note that the number of nodes that have defaulted initially ($c = 0$) or invulnerable ($i < c$) in the beginning will not change throughout the process, so we only need to keep track of the nodes that are initially vulnerable ($0 < c \leq i$) and currently liquid and if needed, we can always calculate the number of defaulted nodes at any time in the process. Further note that the possible states throughout the process for nodes that are vulnerable in the beginning and liquid at a later step are

$$\Gamma := \{(i, j, c, l) : 0 \leq i, 0 \leq j, 0 \leq l < c \leq i \text{ or } c = i + 1, l = i\}. \tag{5}$$

Note particularly the state $(i, j, i + 1, i)$ is the result that a node in state $(i, j, i, i - 1)$ is selected and receives one intervention and thus becomes invulnerable.

Definition 1. (State variable S_k) Let $S_k^{i,j,c,l}$ denote the number of nodes that are vulnerable initially and are in state (i, j, c, l) at step k, for $k = 0, \ldots, m$ and $S_k := (S_k^{i,j,c,l})_{(i,j,c,l)\in\Gamma}$ be the state of the system. Note in the following we may use α to represent $(i, j, c, l)\in\Gamma$ and write S_k^{α} instead of $S_k^{i,j,c,l}$ to simplify the notation.

Recall $m = m(n)$ is the number of the total in (or out) degree of the network, which is also the maximum steps of the process. Throughout this paper we follow the convention that the superscript (usually a multi-index) denotes the state and the subscript denotes the time (discrete or continuous), e.g. $s_\tau^{i,j,c,l}$, $u_\tau^{i,j,c,c-1}$, $s_t^{i,j,c,l}$, $w_t^{i,j,c,l}$ and $u_t^{i,j,c,c-1}$ in the following. Then we define the empirical probability of in, out degrees and initial equity levels.

Definition 2. (Empirical probability) Define the empirical probability of the triplet (in degree, out degree, initial equity level) as

$$P_n(i, j, c) = \frac{1}{n}|\{v \in [n]|d^-(v) = i, d^+(v) = j, e(v) = c\}|. \tag{6}$$

Note that $\sum_{c\geq 0} P_n(i, j, c) = \frac{1}{n}|\{v \in [n]|d^-(v) = i, d^+(v) = j\}|$ represents the empirical probability of the in and out degree pair (i, j).

Previously we use W_k to denote the selected node at step k. Now with a little abuse of notation, let W_k denote the state of the selected node at step k, $k = 1, \ldots, m$, so $W_k \in \Gamma^+ := \{(i, j, c, l): 0 \leq i, 0 \leq j, 0 \leq c, 0 \leq l \leq i\}$. We consider a Markovian control policy $G_n = (g_1^{(n)}(S_0, W_1), \ldots, g_m^{(n)}(S_{m-1}, W_m))$ where $g_{k+1}^{(n)} : \mathbb{N}_0^{|\Gamma|} \times \Gamma^+ \to \{0, 1\}$ specifies the intervention at step $k + 1$ on the selected node which has state W_{k+1} given the state S_k and the superscript (n) shows the dependence on n.

Letting $P_n = (P_n(i, j, c))_{i,j,0\le c\le i}$, we rewrite the terms J_n as $J_{G_n}(P_n)$, R_{T_n} as $R_{T_n}(G_n, P_n)$ and D_{T_n} as $D_{T_n}(G_n, P_n)$ in Eq 4 based on G_n and P_n, so

$$\begin{aligned} R_{T_n}(G_n, P_n) &= \sum_{k=1}^{T_n} g_k^{(n)}(S_{k-1}, W_k), \\ D_{T_n}(G_n, P_n) &= n\sum_{i,j} P_n(i, j, 0) + n\sum_{i,j,1\le c\le i} P_n(i, j, c) - \sum_{(i,j,c,l)\in\Gamma} S_{T_n}^{i,j,c,l} \\ &= n\sum_{i,j,0\le c\le i} P_n(i, j, c) - \sum_{(i,j,c,l)\in\Gamma} S_{T_n}^{i,j,c,l}. \end{aligned} \tag{7}$$

Note that the first equality for $D_{T_n}(G_n, P_n)$ holds because the nodes that default at the end of the process consist of two parts: the nodes that have defaulted initially i.e. $n\sum_{i,j} P_n(i, j, 0)$ and those nodes that are vulnerable initially and default during the process i.e. $n\sum_{i,j,1\le c\le i} P_n(i, j, c) - \sum_{(i,j,c,l)\in\Gamma} S_{T_n}^{i,j,c,l}$.

Assumption 1. Consider a sequence $([n], \mathcal{E}_n)$ of random networks, indexed by the size of the network n. For each $n \in \mathbb{N}$, $(d^-(v))_{v\in[n]}$, $(d^+(v))_{v\in[n]}$ are sequences of nonnegative integers with $\sum_{v\in[n]} d^-(v) = \sum_{v\in[n]} d^+(v)$ and such that for some probability distribution p on $\mathbb{N}_0^3$ independent of n with $\lambda := \sum_{i,j,c} ip(i, j, c) = \sum_{i,j,c} jp(i, j, c) < \infty$, the following holds

1. $P_n(i, j, c) \to p(i, j, c)\ \forall\ i, j, c \ge 0$ as $n \to \infty$.
2. $\sum_{v\in[n]}[(d^-(v))^2 + (d^+(v))^2] = O(n)$.

Note that the second assumption implies by uniform integrability that $\frac{m(n)}{n} \to \lambda$ as $n \to \infty$ and recall that $m(n) := \sum_{v\in[n]} d^-(v) = \sum_{v\in[n]} d^+(v)$. Since $k \le m(n)$, for large n it holds that $\frac{k}{n} \le \frac{m(n)}{n} \le \lambda + 1$. Assumption 1 essentially implies the network is sparse which is justified in many empirical study literature on the structure of financial networks [24].

Remark 3. We previously defined that the vector P_n only includes $P_n(i, j, c)$ in the range $0 \le c \le i$, i.e. the fractions of initially defaulted and vulnerable nodes. Accordingly define $p := (p(i, j, c))_{i,j,0\le c\le i}$, i.e. the vector p only includes $p(i, j, c)$ in the range $0 \le c \le i$.

Next we present our assumptions on the control functions $g_k^{(n)}$.

Assumption 2. Define $\Phi := \{(i, j, c, c-1): 0 \le i, 0 \le j, 1 \le c \le i\}$. Let $G_n = (g_1^{(n)}, \ldots, g_m^{(n)})$ be the a control policy (a sequence of control functions) for the contagion process on a network of size n where n is large enough such that $\frac{m(n)}{n} \le \lambda + 1$. Assume that

$$g_{k+1}^{(n)}(s, w) = \begin{cases} u_{\frac{k}{n}}^{i,j,c,c-1} & \text{if } w = (i, j, c, c-1) \in \Phi \\ 0 & \text{otherwise,} \end{cases} \tag{8}$$

for $0 \le k \le m-1$. $u_\tau^{i,j,c,c-1} = u^{i,j,c,c-1}(\tau)$ where $u^{i,\,j,\,c,\,c-1}$: $[0, \lambda + 1] \to \{0, 1\}$ is a piecewise constant function on $[0, \lambda + 1]$, i.e. there is a partition of the interval into a finite set of intervals such that $u^{i,\,j,\,c,\,c-1}$ is constant 0 or 1 on each interval. Let $u = (u^\beta)_{\beta\in\Phi}$ and Π contain all piecewise constant vector function u on $[0, \lambda + 1]$.

Note that Φ includes possible states having the distance to default equal to one and $\Phi \subset \Gamma$. Further $g_{k+1}^{(n)}(s, w) = 0$ for $w \notin \Phi$ follows from proposition 1. In the following we may use β to represent $(i, j, c, c-1) \in \Phi$ and write u_τ^β instead of $u_\tau^{i,j,c,c-1}$ to simplify the notation.

Remark 4. By this assumption the function u is independent of the state but only a function of time. This implies that the control function $g_{k+1}^{(n)}(s, w)$ depends on the scaled time $\frac{k}{n}$ and the state of the currently selected node w but not on the state s. We will show it suffices to consider

such control policy G_n later after proposition 3 because given a function u, we can predict a deterministic process to which the scaled stochastic contagion process converges in probability at any time as the size of the network $n \to \infty$. Moreover, this type of control policies is the one that can be solved in the optimal control problem Eq (39) we will introduce later.

In summary, assumption 1 assumes the convergence of the empirical probabilities of the in and out degrees and the initial equity. On the other hand, proposition 3 indicates that the control functions depend on the scaled time and the state of the currently selected node. These two assumptions allow us to define the following asymptotic control problem by ensuring that the limits in the objective function are well defined.

Definition 3. For a sequence of networks with P_n and G_n satisfying assumption 1 and assumption 2, respectively, define the asymptotic control problem as

$$\begin{aligned} & \min_{u\in\Pi} \lim_{n\to\infty} J_{G_n}(P_n) \\ = \quad & \min_{u\in\Pi} \lim_{n\to\infty} K\mathbb{E}\frac{R_{T_n}(G_n,P_n)}{n} + \mathbb{E}\frac{D_{T_n}(G_n,P_n)}{n}. \end{aligned} \tag{9}$$

In the following we will show the limits in Eq (9) are well defined by applying Wormald's theorem [30].

Dynamics of the default contagion process with interventions

Recall that R_k is the accumulative number of interventions up to step k, so

$$\begin{aligned} R_0 &= 0, \\ R_k &= \sum_{\ell=1}^{k} g_\ell(S_{\ell-1}, W_\ell) \\ &= \sum_{\ell=1}^{k}\sum_{\beta\in\Phi} \mathbb{1}_{(W_\ell=\beta)} u^{\beta}_{\frac{\ell-1}{n}}. \end{aligned} \tag{10}$$

We shall show that $(S_k, R_k)_{k=0,\ldots,m}$ is a controlled Markov chain given a control policy G_n. In Fig 4 we illustrate for the same (i, j) pair the state space as well as their transition relations between the states.

To describe the transition probabilities, assume the state of the selected node at step $k + 1$ is $W_{k+1} = (i, j, c, l)$, for $k = 0, \ldots, m - 1$, there are three possibilities:

The selected node has defaulted, i.e. $c \leq l$ or the node is invulnerable, i.e. $c > i$, then $S_{k+1} = S_k$, $R_{k+1} = R_k$.

1. The selected node is vulnerable but has the "distance to default" greater than one, i.e. $c - l \geq 2$, then the node w is selected with probability $\frac{(i-l)S_k^{i,j,c,l}}{m-k}$ and

$$\begin{aligned} S_{k+1}^{i,j,c,l} &= S_k^{i,j,c,l} - 1, \\ S_{k+1}^{i,j,c,l+1} &= S_k^{i,j,c,l+1} + 1, \\ R_{k+1} &= R_k, \end{aligned} \tag{11}$$

while other entries of S_{k+1} are the same as S_k.

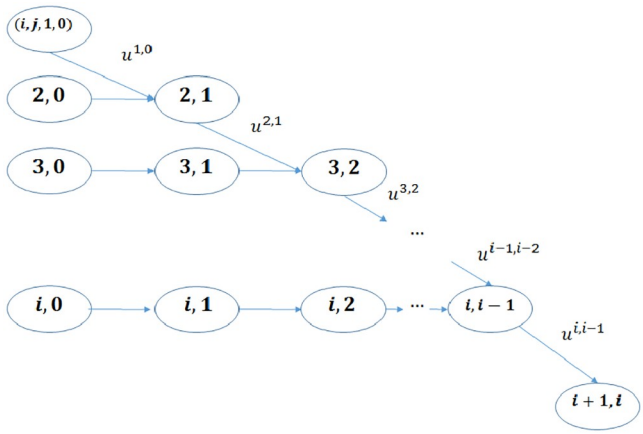

Fig 4. The state space for the same (i, j) pair, $0 \leq i$, $0 \leq j$ and their transition relations.

https://doi.org/10.1371/journal.pone.0209819.g004

2. The selected node has the "distance to default" of one, i.e. $c - l = 1$, then the node is selected with probability $\frac{(i-c+1)S_k^{i,j,c,c-1}}{m-k}$ and by assumption 2,

$$\begin{aligned} S_{k+1}^{i,j,c,c-1} &= S_k^{i,j,c,c-1} - 1, \\ S_{k+1}^{i,j,c+1,c} &= S_k^{i,j,c+1,c} + g_{k+1}^{(n)}(S_k, (i, j, c, c-1)) \\ &= S_k^{i,j,c+1,c} + u_{\frac{k}{n}}^{i,j,c,c-1}, \\ R_{k+1} &= R_k + u_{\frac{k}{n}}^{i,j,c,c-1}, \end{aligned} \tag{12}$$

while other entries of S_{k+1} are the same as S_k.

Let $(\mathcal{F}_k)_{k=0,\ldots,m}$ be the natural filtration of S_k, $\Delta S_k^\alpha = S_{k+1}^\alpha - S_k^\alpha$, $\alpha \in \Gamma$ and $\Delta R_k = R_{k+1} - R_k$, it follows that

$$\begin{aligned} \mathbb{E}[\Delta S_k^{i,j,c,0} | \mathcal{F}_k] &= -\frac{iS_k^{i,j,c,0}}{m-k} \qquad \text{for } 1 \leq c \leq i, \\ \mathbb{E}[\Delta S_k^{i,j,c,l} | \mathcal{F}_k] &= \frac{(i-l+1)S_k^{i,j,c,l-1}}{m-k} - \frac{(i-l)S_k^{i,j,c,l}}{m-k} \\ &\quad \text{for } 3 \leq c \leq i, 1 \leq l \leq c-2, \\ \mathbb{E}[\Delta S_k^{i,j,c,c-1} | \mathcal{F}_k] &= \frac{(i-c+2)S_k^{i,j,c-1,c-2}}{m-k} u_{\frac{k}{n}}^{i,j,c-1,c-2} \\ &\quad + \frac{(i-c+2)S_k^{i,j,c,c-2}}{m-k} - \frac{(i-c+1)S_k^{i,j,c,c-1}}{m-k} \\ &\quad \text{for } 2 \leq c \leq i, \\ \mathbb{E}[\Delta S_k^{i,j,i+1,i} | \mathcal{F}_k] &= \frac{S_k^{i,j,i,i-1}}{m-k} u_{\frac{k}{n}}^{i,j,i,i-1}, \\ \mathbb{E}[\Delta R_k | \mathcal{F}_k] &= \sum_{(i,j,c,c-1) \in \Phi} \frac{(i-c+1)S_k^{i,j,c,c-1}}{m-k} u_{\frac{k}{n}}^{i,j,c,c-1}. \end{aligned} \tag{13}$$

Convergence of the default contagion process with interventions

Based on the dynamics of the contagion process under interventions described in the previous section, we will show that the state variable S_k, the accumulative interventions R_k, the number of defaults D_k and the number of unrevealed out-links from the default set D_k^- (defined later) after being scaled by n all converge to a deterministic process which depends on the solution of the system of ODEs we will present now. Then we are able to show that the stochastic control problem Eq (4) converges to the asymptotic control problem Eq (9).

Definition 4. (ODEs of s_τ) Given a set of piecewise constant function $u = (u^\beta)_{\beta\in\Phi}$ on $[0, \lambda]$, i.e. $u \in \Pi$, define the system of ordinary differential equations (ODEs) of $s_\tau = (s_\tau^\alpha)_{\alpha\in\Gamma}$ as

$$\begin{aligned}
\frac{ds_\tau^{i,j,c,0}}{d\tau} &= -\frac{is_\tau^{i,j,c,0}}{\lambda-\tau} \qquad \text{for } 1 \le c \le i,\\
\frac{ds_\tau^{i,j,c,l}}{d\tau} &= \frac{(i-l+1)s_\tau^{i,j,c,l-1}}{\lambda-\tau} - \frac{(i-l)s_\tau^{i,j,c,l}}{\lambda-\tau}\\
&\quad \text{for } 3 \le c \le i, 1 \le l \le c-2,\\
\frac{ds_\tau^{i,j,c,c-1}}{d\tau} &= \frac{(i-c+2)s_\tau^{i,j,c-1,c-2}}{\lambda-\tau}u_\tau^{i,j,c-1,c-2} + \frac{(i-c+2)s_\tau^{i,j,c,c-2}}{\lambda-\tau}\\
&\quad -\frac{(i-c+1)s_\tau^{i,j,c,c-1}}{\lambda-\tau} \text{ for } 2 \le c \le i,\\
\frac{ds_\tau^{i,j,i+1,i}}{d\tau} &= \frac{s_\tau^{i,j,i,i-1}}{\lambda-\tau}u_\tau^{i,j,i,i-1}.
\end{aligned} \tag{14}$$

The ODEs can be expressed in the form $\frac{ds_\tau}{d\tau} = h(\tau, s_\tau; u_\tau)$ where $h = (h^\alpha)_{\alpha\in\Gamma}$.

For what is needed below we analyze the solutions of the ODEs in definition 4 for a subinterval of $[0, \lambda]$ on which u_τ is a constant vector function.

Proposition 2. Let $s_\tau = (s_\tau^\alpha)_{\alpha\in\Gamma}$ satisfy the system of ordinary differential equations in definition 4 with the initial conditions $s_{\tau_1} = s_1 := (s_1^\alpha)_{\alpha\in\Gamma}$ and assume u_τ is a constant vector function $u_\tau = b := (b^\beta)_{\beta\in\Phi}$ in the interval $[\tau_1, \tau_2) \subseteq [0, \lambda)$ where $b^\beta \in \{0, 1\}$ is a constant, then the solution s_τ on $[\tau_1, \tau_2)$ is

$$\begin{aligned}
s_\tau^{i,j,c,l} &= \left(\frac{\lambda-\tau}{\lambda-\tau_1}\right)^{i-l}\sum_{r=0}^{l} s_1^{i,j,c,r}\binom{i-r}{l-r}\left(1-\frac{\lambda-\tau}{\lambda-\tau_1}\right)^{l-r}\\
&\quad \text{for } 2 \le c \le i, 0 \le l \le c-2,
\end{aligned} \tag{15}$$

$$\begin{aligned}
s_\tau^{i,j,c,c-1} &= \left(\frac{\lambda-\tau}{\lambda-\tau_1}\right)^{i-c+1}\sum_{r=0}^{c-1}\sum_{q=r+1}^{c}\prod_{k=q}^{c-1} b^{i,j,k,k-1}s_1^{i,j,q,r}\binom{i-r}{c-1-r}\left(1-\frac{\lambda-\tau}{\lambda-\tau_1}\right)^{c-1-r}\\
&\quad \text{for } 1 \le c \le i,
\end{aligned} \tag{16}$$

$$s_\tau^{i,j,i+1,i} = s_1^{i,j,i+1,i} + \sum_{r=0}^{i-1}\sum_{q=r+1}^{i}\prod_{k=q}^{i} b^{i,j,k,k-1}s_1^{i,j,q,r}\left(1-\frac{\lambda-\tau}{\lambda-\tau_1}\right)^{i-r}, \tag{17}$$

where $\prod_{k=c}^{c-1} b^{i,j,k,k-1} := 1$. As a direct result, if we take the initial condition $s_{\tau_1}^{i,j,c,l} = p(i,j,c)\mathbb{1}_{(l=0)}$

for $(i, j, c, l) \in \Gamma$ at $\tau_1 = 0$, it follows that

$$\begin{aligned} s_\tau^{i,j,c,l} &= p(i,j,c)\binom{i}{l}(1-\frac{\tau}{\lambda})^{i-l}(\frac{\tau}{\lambda})^l \\ &\text{for } 2 \le c \le i,\ 0 \le l \le c-2. \end{aligned} \tag{18}$$

Remark 5. We discuss some properties of s_τ. Observe that the ODEs are "separable" in that $s_\tau^{i,j,c,l}$ only depends on the entries of s_τ and p with the same (i, j). Fix an (i, j) pair, define $\Gamma_{i,j} := \{(c, l): 0 \le l < c \le i \text{ or } c = i + 1, l = i\}$. If u_τ is a constant vector of 1's on $[\tau_1, \tau_2) \subseteq [0, \lambda)$, then we can show after some algebra that for any $\tau \in [\tau_1, \tau_2)$,

$$\sum_{(c,l)\in\Gamma_{i,j}} s_\tau^{i,j,c,l} = \sum_{(c,l)\in\Gamma_{i,j}} s_{\tau_1}^{i,j,c,l}. \tag{19}$$

If there exists some c_0 such that $1 \le c_0 \le i$, $u_\tau^{i,j,c_0,c_0-1} = 0$, then $\sum_{(c,l)\in\Gamma_{i,j}} s_\tau^{i,j,c,l} < \sum_{(c,l)\in\Gamma_{i,j}} s_{\tau_1}^{i,j,c,l}$. Since the initial condition is $s_0^{i,j,c,l} = p(i,j,c)\mathbb{1}_{(l=0)}$ for $1 \le c \le i$, it follows that

$$\sum_{(c,l)\in\Gamma_{i,j}} s_\tau^{i,j,c,l} \le \sum_{1\le c\le i} p(i,j,c). \tag{20}$$

In the following part our goal is to approximate $\frac{R_k}{n}$ and $\frac{D_k}{n}$ as $n \to \infty$ given a function u. However, the number of variables depends on n, so we need to bound the terms associated with large in or out degrees. Fix $\epsilon > 0$ and by assumption 1 we have that

$$\lambda = \sum_{i,j,c} ip(i,j,c) = \sum_{i,j,c} jp(i,j,c) < \infty, \tag{21}$$

then there exists an integer M^ϵ such that

$$\sum_{i\ge M^\epsilon}\sum_{j,c} ip(i,j,c) + \sum_{j\ge M^\epsilon}\sum_{i,c} jp(i,j,c) < \epsilon, \tag{22}$$

so letting $i \vee j = \max\{i, j\}$, we have

$$\begin{aligned} &\sum_{i\vee j\ge M^\epsilon,c} jp(i,j,c) \\ = &\sum_{i\ge M^\epsilon}\sum_{j<M^\epsilon}\sum_c jp(i,j,c) + \sum_{i\ge M^\epsilon}\sum_{j\ge M^\epsilon}\sum_c jp(i,j,c) + \sum_{i<M^\epsilon}\sum_{j\ge M^\epsilon}\sum_c jp(i,j,c) \\ \le &\sum_{i\ge M^\epsilon}\sum_{j<M^\epsilon}\sum_c ip(i,j,c) + \sum_{i\ge M^\epsilon}\sum_{j\ge M^\epsilon}\sum_c jp(i,j,c) + \sum_{i<M^\epsilon}\sum_{j\ge M^\epsilon}\sum_c jp(i,j,c) \\ < &\epsilon. \end{aligned} \tag{23}$$

We can prove similarly that there exists an integer L^ϵ such that $\sum_{i\vee j\ge L^\epsilon,c} ip(i,j,c) < \epsilon$, but without loss of generality we write M^ϵ instead of L^ϵ in what follows. Moreover, by assumption 1, as $n \to \infty$,

$$\sum_{i,j,c} iP_n(i,j,c) = \sum_{i,j,c} jP_n(i,j,c) \to \lambda < \infty, \tag{24}$$

so for n large enough, we can show that

$$\begin{aligned} \sum_{i \vee j \geq M^{\epsilon}, c} jP_n(i,j,c) &< \epsilon, \\ \sum_{i \vee j \geq M^{\epsilon}, c} iP_n(i,j,c) &< \epsilon. \end{aligned} \tag{25}$$

So we define the integer M^{ϵ} formally.

Definition 5. Given any $\epsilon > 0$, define M^{ϵ} as the integer such that

$$\begin{aligned} \sum_{i \vee j \geq M^{\epsilon}, c} ip(i,j,c) &< \epsilon, \\ \sum_{i \vee j \geq M^{\epsilon}, c} jp(i,j,c) &< \epsilon. \end{aligned} \tag{26}$$

Accordingly, define

$$\begin{aligned} \Gamma^{\epsilon} &:= \{(i,j,c,l) : i \vee j < M^{\epsilon}, 0 \leq l < c \leq i \text{ or } c = i+1, l = i\}, \\ \Phi^{\epsilon} &:= \{(i,j,c,c-1) : i \vee j < M^{\epsilon}, 0 \leq c \leq i\}, \\ \hat{\lambda} &:= \lambda - \epsilon, \end{aligned} \tag{27}$$

where $a \vee b = \max\{a, b\}$.

Next we show that the scaled state variable S_k and R_k converges in probability to the solution of the ODEs in definition 4 given the function u. The difficulty in proving proposition 3 arises from the fact that the right sides of the ODEs for s_τ in definition 4 are discontinuous due to interventions so the auxiliary Wormald's theorem in Appendix B: Wormald's theorem is not applicable and needs to be adapted.

Proposition 3. Consider a sequence of networks with initial conditions $(P_n)_{n \geq 1}$ satisfying assumption 1 and let $(G_n)_{n \geq 1}$ be the sequence of control policies for the contagion process on the sequence of networks and $(G_n)_{n \geq 1}$ satisfy assumption 2 with the function $u = (u^{\beta})_{\beta \in \Phi^{\epsilon}}$, then

$$\begin{aligned} \sup_{0 \leq k \leq n\hat{\lambda}} \frac{S_k^{\alpha}}{n} - s^{\alpha}_{\frac{k}{n}} &= O\left(n^{-\frac{1}{4}}\right), \\ \sup_{0 \leq k \leq n\hat{\lambda}} \frac{\tilde{R}_k}{n} - \tilde{r}_{\frac{k}{n}} &= O\left(n^{-\frac{1}{4}}\right), \end{aligned} \tag{28}$$

with probability $1 - O\left(n^{\frac{1}{4}} \exp\left(-n^{\frac{1}{4}}\right)\right)$ and $\alpha \in \Gamma^{\epsilon}$, where $s_\tau = (s^{\alpha}_{\tau})_{\alpha \in \Gamma^{\epsilon}}$ is the solution for the ODEs in definition 4 with the initial conditions $s_0^{i,j,c,l} = p(i,j,c)\mathbb{1}_{(l=0)}$ and

$$\begin{aligned} \tilde{R}_0 &= 0, \\ \tilde{R}_k &= \sum_{\ell=1}^{k} \sum_{\beta \in \Phi^{\epsilon}} \mathbb{1}_{(W_\ell = \beta)} u^{\beta}_{\frac{\ell-1}{n}}, \end{aligned} \tag{29}$$

and

$$\tilde{r}_\tau = \int_0^\tau \sum_{(i,j,c,c-1) \in \Phi^{\epsilon}} \frac{(i-c+1)s_t^{i,j,c,c-1}}{\lambda - t} u_t^{i,j,c,c-1} dt. \tag{30}$$

From proposition 3 we see that given $(P_n)_{n\geq 1}$ and $(G_n)_{n\geq 1}$ satisfying assumption 1 and assumption 2, respectively, the scaled stochastic process $\frac{S_k}{n}$ converges to the deterministic process $s_{\frac{k}{n}}$ for any k in $[0, n\hat{\lambda}]$. This justifies assumption 2 on the control policy because given a control policy G_n depending on the function u, we can predict with high probability the scaled stochastic contagion process at any time k.

Next we discuss the convergence of the scaled number of defaults $\frac{D_k}{n}$ and the process end time $\frac{T_n}{n}$. Note in definition 6, definition 7 and proposition 4 it is not required that $i \vee j < M^{\epsilon}$

Definition 6. Define D_k^- the number of unrevealed out links from the default set at step k.

Recall that D_k denotes the number of defaulted nodes at step k which consist of two parts: the nodes that have defaulted initially $n\sum_{i,j} P_n(i,j,0)$ and those that are vulnerable initially and default by step k, i.e. $n\sum_{i,j,1\leq c\leq i} P_n(i,j,c) - \sum_{(i,j,c,l)\in\Gamma} S_k^{i,j,c,l}$, thus

$$\begin{aligned} D_k = & \; n\sum_{i,j} P_n(i,j,0) + n\sum_{i,j,1\leq c\leq i} P_n(i,j,c) - \sum_{(i,j,c,l)\in\Gamma} S_k^{i,j,c,l} \\ = & \; n\sum_{i,j,0\leq c\leq i} P_n(i,j,c) - \sum_{(i,j,c,l)\in\Gamma} S_k^{i,j,c,l}. \end{aligned} \tag{31}$$

Similarly, among all defaulted nodes at step k the nodes with out degree j consist of two parts: the nodes that have defaulted initially $n\sum_i P_n(i,j,0)$ and those nodes that are vulnerable initially and default by step k, $n\sum_{i,1\leq c\leq i} P_n(i,j,c) - \sum_{i,0\leq l<c\leq i \text{ or } c=i+1,l=i} S_k^{i,j,c,l}$, thus

$$\begin{aligned} D_k^- = & \; \sum_j j(n\sum_i P_n(i,j,0) + n\sum_{i,1\leq c\leq i} P_n(i,j,c) - \sum_{i,0\leq l<c\leq i \text{ or } c=i+1,l=i} S_k^{i,j,c,l}) - k \\ = & \; n\sum_{i,j,0\leq c\leq i} jP_n(i,j,c) - \sum_{(i,j,c,l)\in\Gamma} jS_k^{i,j,c,l} - k. \end{aligned} \tag{32}$$

Correspondingly we make the following definitions to approximate $\frac{D_k}{n}$ and $\frac{D_k^-}{n}$ as $n \to \infty$.

Definition 7. Define

$$\begin{aligned} d_\tau = & \; \sum_{i,j,0\leq c\leq i} p(i,j,c) - \sum_{(i,j,c,l)\in\Gamma} s_\tau^{i,j,c,l}, \\ d_\tau^- = & \; \sum_{i,j,0\leq c\leq i} jp(i,j,c) - \sum_{(i,j,c,l)\in\Gamma} js_\tau^{i,j,c,l} - \tau. \end{aligned} \tag{33}$$

Proposition 4. Based on definition 6 and definition 7, it follows that

$$\begin{aligned} \sup_{0\leq k\leq n\hat{\lambda}} \left| \frac{D_k^-}{n} - d_{\frac{k}{n}}^- \right| & \leq o_p(1) + 2\epsilon, \\ \sup_{0\leq k\leq n\hat{\lambda}} \left| \frac{D_k}{n} - d_{\frac{k}{n}} \right| & \leq o_p(1) + 2\epsilon. \end{aligned} \tag{34}$$

To summarize the results we have so far, we have shown in proposition 3 and proposition 4 that the state variable S_k, the accumulative interventions R_k, the number of defaults D_k and the number of unrevealed out-links from the default set D_k^- after being scaled by n all converge to a deterministic process which depends on the solution of ODEs in definition 4. This convergence applies to any k before $n\hat{\lambda}$. By definition 6, $T_n = \min\{0 \leq k \leq m : D_k^- = 0\}$. Additionally define $\tau_f = \inf\{0 \leq \tau \leq \lambda : d_\tau^- = 0\}$. Next we show that when $\frac{T_n}{n}$ converges in probability to τ_f, then $\frac{R_{T_n}}{n}$ and $\frac{D_{T_n}}{n}$ also converge in probability to the corresponding deterministic variables,

r_{τ_f} and d_{τ_f}, which in light of the boundedness of $\frac{R_{T_n}}{n}$ and $\frac{D_{T_n}}{n}$ further implies convergence in expectations, thus the limits in Eq (9) are well defined.

Proposition 5. Consider a sequence of networks with initial conditions $(P_n)_{n\geq 1}$ satisfying assumption 1 and let $(G_n)_{n\geq 1}$ be the sequence of control policies for the contagion processes on the sequence of networks and $(G_n)_{n\geq 1}$ satisfy assumption 2 with the function u. If $\tau_f = \lambda$, or $\tau_f < \lambda$ and $\frac{d}{d\tau} d^-_{\tau_f} < 0$, it follows that as $n \to \infty$,

$$\begin{aligned} \frac{R_{T_n}(G_n, P_n)}{n} &\xrightarrow{p} r_{\tau_f}(u, p), \\ \frac{D_{T_n}(G_n, P_n)}{n} &\xrightarrow{p} d_{\tau_f}(u, p). \end{aligned} \tag{35}$$

where

$$r_{\tau_f} = \int_0^{\tau_f} \sum_{(i,j,c,c-1)\in\Phi} \frac{(i-c+1)s_t^{i,j,c,c-1}}{\lambda - t} u_t^{i,j,c,c-1} dt. \tag{36}$$

Further it follows that as $n \to \infty$,

$$\begin{aligned} \mathbb{E}\frac{R_{T_n}(G_n, P_n)}{n} &\to r_{\tau_f}(u, p), \\ \mathbb{E}\frac{D_{T_n}(G_n, P_n)}{n} &\to d_{\tau_f}(u, p). \end{aligned} \tag{37}$$

Under the conditions in proposition 5, the asymptotic control problem Eq (9) becomes

$$\min_{u\in\Pi} K \cdot r_{\tau_f}(u, p) + d_{\tau_f}(u, p). \tag{38}$$

In the following let $u_\tau = (u_\tau^\beta)_{\beta\in\Phi}$ and $u = (u_\tau)_{\tau\in[0,\lambda]}$.

Substituting the expressions of $r_{\tau_f}(u, p)$ and $d_{\tau_f}(u, p)$ in Eqs (36) and (33) respectively into Eq (38) and putting together the system of ordinary differential equations of $s_\tau = (s_\tau^\alpha)_{\alpha\in\Gamma}$, i.e. $\frac{d}{d\tau} s_\tau = h(\tau, s_\tau; u_\tau)$ in definition 4 as well as the condition that determines τ_f, i.e. $d^-_{\tau_f} = 0$, we attain the following deterministic optimal control problem.

$$\begin{aligned} \min_{u,\tau_f} \quad & K \cdot r_{\tau_f}(u, p) \quad + d_{\tau_f}(u, p) \\ \text{st} \quad & \frac{d}{d\tau} s_\tau = h(\tau, s_\tau; u_\tau) \\ & s_0^{i,j,c,l} = p(i, j, c)\mathbb{1}_{(l=0)} \\ & d^-_{\tau_f} = 0 \\ & u_\tau^\beta \in \{0, 1\}\ \forall \beta \in \Phi \\ & \tau_f \in [0, \lambda), \end{aligned} \tag{39}$$

where $\frac{d}{d\tau}s_\tau = h(\tau, s_\tau; u_\tau)$ is defined as in definition 4 and

$$\begin{aligned} r_{\tau_f}(u,p) &= \int_0^{\tau_f} \sum_{i,j,1\le c\le i} \frac{(i-c+1)s_\tau^{i,j,c,c-1}}{\lambda-\tau} u_\tau^{i,j,c,c-1} d\tau, \\ d_{\tau_f}(u,p) &= \sum_{i,j,0\le c\le i} p(i,j,c) - \sum_{(i,j,c,l)\in\Gamma} s_{\tau_f}^{i,j,c,l}, \\ d_{\tau_f}^- &= \sum_{i,j,0\le c\le i} jp(i,j,c) - \sum_{(i,j,c,l)\in\Gamma} js_{\tau_f}^{i,j,c,l} - \tau_f. \end{aligned} \tag{40}$$

Some difficulties arise because Eq (39) is an infinite dimensional optimal control problem. In light of assumption 1, it suffices to solve a finite dimensional problem to approximate the objective function of the infinite dimensional problem. First we define the finite dimensional optimal control problem.

Definition 8. (FOCP) For $\epsilon > 0$, recall M^ϵ as in definition 5. Define the finite dimensional optimal control problem (FOCP) as Eq (39) with the indexes (i, j) restricted to $i \vee j < M^\epsilon$.

Remark 6. The restriction of (i, j) to $i \vee j < M^\epsilon$ indicates that we use only $p(i, j, c)$, $i \vee j < M^\epsilon$, $0 \le c \le i$ in the calculation. It is equivalent to setting $p(i, j, c) = 0$ for $i \vee j \ge M^\epsilon$, $0 \le c \le i$ while keeping $p(i, j, c)$ for $i \vee j < M^\epsilon$, $0 \le c \le i$ unchanged, which implies asymptotically nodes with $i \vee j \ge M^\epsilon$ are all invulnerable. By the solution of the ODE in proposition 2, it implies that $s_\tau^\alpha = 0$, for $\alpha \in \Gamma\backslash\Gamma^\epsilon$. Note we use tilde sign with the variables to indicate the indexes (i, j) are in the range $i \vee j < M^\epsilon$, for example, $\tilde{r}_{\tau_f}$, $\tilde{d}_{\tau_f}$ and $\tilde{d}_{\tau_f}^-$.

We have the following lemma regarding the objective functions of the infinite and finite dimensional optimal control problems.

Lemma 1. Let $\zeta(u, \tau_f, p) := K\, r_{\tau_f}(u, p) + d_{\tau_f}(u, p)$ be the objective function for the infinite dimensional Eq (39) and $\tilde{\zeta}(u, \tau_f, p) := K\, \tilde{r}_{\tau_f}(u, p) + \tilde{d}_{\tau_f}(u, p)$ for (FOCP). Let (u^*, τ_f^*) and $(\tilde{u}, \tilde{\tau}_f)$ be the optimal solutions for the infinite dimensional Eq (39) and (FOCP), respectively, then for the same p we have that

$$\left|\tilde{\zeta}(\tilde{u}, \tilde{\tau}_f, p) - \zeta(u^*, \tau_f^*, p)\right| < (K+1)\epsilon. \tag{41}$$

By lemma 1 we only need to solve the finite dimensional optimal control problem (FOCP) in definition 8. Because ϵ can be arbitrarily small, we can approximate the objective function of the infinite dimensional problem to any precision. Given p for (FOCP), the Pontryagin's maximum principle provides the necessary conditions for the optimal control $\tilde{u}$ and end time $\tilde{\tau}_f$. We can obtain the optimal asymptotic number of interventions $\tilde{r}_{\tilde{\tau}_f}$ and fraction of final defaults $\tilde{d}_{\tilde{\tau}_f}$, which lead to the main results of our work. In the next section we focus on solving (FOCP) and suppress the tilde sign for the variables for notational convenience.

Necessary conditions for the optimal control problem

In the following we solve the finite dimensional optimal control problem (FOCP) in definition 8. Throughout this section we understand that the degrees are in the bounded range $i \vee j < M^\epsilon$ unless specified otherwise. We also suppress the tilde sign for notational convenience. Let $t = t(\tau) := -\ln(\lambda - \tau)$, $t_0 := t(0) = -\ln\lambda$ and $t_f := -\ln(\lambda - \tau_f)$. Note $t(\tau)$ is a strictly increasing function of τ and so is the inverse function $\tau = \tau(t)$. We remark that we assume in the following that $\tau < \lambda$ which implies $t_f < \infty$, but later we can see that the solutions of s_τ, u_τ and w_τ do allow $\tau = \lambda$. Then we can reformulate the optimal control problem Eq (39) into an autonomous one,

i.e. the differential equations of the system dynamics do not depend on time explicitly. Let $u_t = (u_t^\beta)_{\beta\in\Phi^\epsilon}$ and $u = (u_t)_{t\geq t_0}$ (note previously $u = (u_\tau)_{\tau\in[0,\lambda]}$. Additionally, we allow an arbitrary starting time t_0 here).

$$\min_{u,t_f} \quad K\cdot r_{t_f}(u,p) \quad + d_{t_f}(u,p) \tag{42}$$

$$\begin{aligned} \text{st} \qquad \frac{d}{dt}s_t &= h(s_t;u_t) \\ s_{t_0}^{i,j,c,l} &= p(i,j,c)\mathbb{1}_{(l=0)} \\ d_{t_f}^- &= 0 \\ u_t^\beta &\in \{0,1\},\ \forall\beta\in\Phi^\epsilon \\ t_f &\in [0,\infty), \end{aligned} \tag{43}$$

where $\frac{d}{dt}s_t = h(s_t;u_t)$ denotes the system of differential equations

$$\begin{aligned} \frac{ds_t^{i,j,c,0}}{dt} &= -is_t^{i,j,c,0} \text{ for } 1\leq c\leq i, \\ \frac{ds_t^{i,j,c,l}}{dt} &= (i-l+1)s_t^{i,j,c,l-1} - (i-l)s_t^{i,j,c,l} \\ &\quad \text{for } 3\leq c\leq i,\ 1\leq l\leq c-2, \\ \frac{ds_t^{i,j,c,c-1}}{dt} &= (i-c+2)s_t^{i,j,c-1,c-2}u_t^{i,j,c-1,c-2} + (i-c+2)s_t^{i,j,c,c-2} \\ &\quad -(i-c+1)s_t^{i,j,c,c-1} \\ &\quad \text{for } 2\leq c\leq i, \\ \frac{ds_t^{i,j,i+1,i}}{dt} &= s_t^{i,j,i,i-1}u_t^{i,j,i,i-1}, \end{aligned} \tag{44}$$

and

$$r_{t_f}(u,p) \quad = \int_{t_0}^{t_f} \sum_{i,j,1\leq c\leq i} (i-c+1)s_t^{i,j,c,c-1}u_t^{i,j,c,c-1}dt, \tag{45}$$

$$d_{t_f}(u,p) \quad = \sum_{i,j,0\leq c\leq i} p(i,j,c) - \sum_{(i,j,c,l)\in\Gamma^\epsilon} s_{t_f}^{i,j,c,l}, \tag{46}$$

$$\begin{aligned} d_{t_f}^- &= \sum_{i,j,0\leq c\leq i} jp(i,j,c) - \sum_{(i,j,c,l)\in\Gamma^\epsilon} js_{t_f}^{i,j,c,l} - \tau_f \\ &= \sum_{i,j,0\leq c\leq i} jp(i,j,c) - \sum_{(i,j,c,l)\in\Gamma^\epsilon} js_{t_f}^{i,j,c,l} - \lambda(1-e^{t_0-t_f}). \end{aligned} \tag{47}$$

Note that Eq (47) follows from $\frac{1}{\lambda} = e^{t_0}$ and thus $\tau_f = \lambda - e^{-t_f} = \lambda\left(1-\frac{1}{\lambda}e^{-t_f}\right) = \lambda(1-e^{t_0-t_f})$.

To determine the necessary conditions for the optimal terminal time t_f^* and optimal control u_t^* in Eq (42), we need the Extended Pontryagin Maximum Principle in Appendix C: Extended pontryagin maximum principle. Then we put together the objective function Eq (43) and the necessary conditions to construct the optimization problem Eq (86) we will introduce later.

Applying the Extended Pontryagin Maximum Principle to the optimal control problem Eq (42) yields the following necessary conditions of optimality. Note in order to simplify notations, we suppress the apostrophe "*" in the following. In other words, we use s_t, u_t, w_t, v, t_f instead of $s_t^*, u_t^*, w_t^*, v^*, t_f^*$ to denote the optimal values.

Proposition 6. (Necessary conditions of optimality) Let $(s_t, u_t)_{t\in[t_0,\, tf]}$ be the optimal state and control trajectory of Eq (42) where t_f is the optimal terminal time. Define the Hamiltonian function

$$\begin{aligned}\mathcal{H}(s_t, u_t, w_t) = & \sum_{i,j,1\le c\le i} w_t^{i,j,c,0}(-is_t^{i,j,c,0}) \\ & + \sum_{i,j,2\le c\le i,1\le l\le c-1} w_t^{i,j,c,l}\left[(i-l+1)s_t^{i,j,c,l-1} - (i-l)s_t^{i,j,c,l}\right] \\ & + \sum_{i,j,2\le c\le i+1} (K + w_t^{i,j,c,c-1})(i-c+2)s_t^{i,j,c-1,c-2}u_t^{i,j,c-1,c-2},\end{aligned} \tag{48}$$

then there exists a piecewise continuously differentiable vector function $w_t = (w_t^\alpha)_{\alpha\in\Gamma^\in} \in \hat{C}^1[t_0,\infty)^{|\Gamma^\in|}$ and a scalar $v \in \mathbb{R}$ such that the following conditions hold:

1. The optimal control u_t satisfies that $\forall t \in [t_0, t_f]$, $1 \le c \le i$,if $s_t^{i,j,c,c-1} > 0$,

$$u_t^{i,j,c,c-1} = \begin{cases} 0 & \text{if } w_t^{i,j,c+1,c} > -K \\ 1 & \text{if } w_t^{i,j,c+1,c} < -K \\ 0 \text{ or } 1 & \text{if } w_t^{i,j,c+1,c} = -K, \end{cases} \tag{49}$$

if $s_t^{i,j,c,c-1} = 0$,

$$u_t^{i,j,c,c-1} = 0 \text{ or } 1. \tag{50}$$

2. For $2 \le c \le i$, $0 \le l \le c-2$,

$$\frac{d}{dt} w_t^{i,j,c,l} = (i-l)(w_t^{i,j,c,l} - w_t^{i,j,c,l+1}), \tag{51}$$

and for $1 \le c \le i$,

$$\frac{d}{dt} w_t^{i,j,c,c-1} = (i-c+1)(w_t^{i,j,c,c-1} - (K + w_t^{i,j,c+1,c})u_t^{i,j,c,c-1}), \tag{52}$$

and

$$\frac{d}{dt} w_t^{i,j,i+1,i} = 0. \tag{53}$$

We denote the set of ordinary differential equations for w_t as $\frac{d}{dt} w_t = h'(w_t; u_t)$.

3. $\mathcal{H}(s_t, u_t, w_t)$ is a constant for $t \in [t_0, t_f]$.

4. Transversal conditions

$$w_{t_f}^{i,j,c,l} = vj - 1 \qquad \forall (i,j,c,l) \in \Gamma^{\epsilon}, \tag{54}$$

$$\mathcal{H}(s_{t_f}, u_{t_f}, w_{t_f}) = -ve^{-t_f}, \tag{55}$$

$$\begin{aligned} d_{t_f}^{-} &= \sum_{i,j,0\le c\le i} jp(i,j,c) - \sum_{(i,j,c,l)\in\Gamma^{\epsilon}} js_{t_f}^{i,j,c,l} - \lambda(1 - e^{t_0 - t_f}) \\ &= 0. \end{aligned} \tag{56}$$

Remark 7. (Singular control policy) Observe that if the coefficient of $u_t^{i,j,c,c-1}$ in the Hamiltonian function $\mathcal{H}(s_t, u_t, w_t)$ Eq (48), i.e. $(i - c + 1)(K + w_t^{i,j,c+1,c})s_t^{i,j,c,c-1}$ vanishes, $u_t^{i,j,c,c-1} = 0$ or 1 both satisfy conditions of the Extended Pontryagin Maximum Principle in Appendix C: Extended pontryagin maximum principle i.e. minimizing $\mathcal{H}(s_t, u_t, w_t)$. In other words, the Extended Pontryagin Maximum Principle cannot determine the optimal control $u_t^{i,j,c,c-1}$ in this case. Moreover, since $i - c + 1 > 0$, so if $(K + w_t^{i,j,c+1,c})s_t^{i,j,c,c-1} = 0$ can be sustained over some interval $(\theta_1, \theta_2) \subset [t_0, t_f]$, then $u_t^{i,j,c,c-1}$ can be 0 or 1 at any time on (θ_1, θ_2) and switch arbitrarily often between 0 and 1. In the terminology of optimal control theory, the control u_t is "singular" on (θ_1, θ_2) and the corresponding portion of the state trajectory s_t on (θ_1, θ_2) is called a singular arc. Further note that $(K + w_t^{i,j,c+1,c})s_t^{i,j,c,c-1} = 0$, $t \in (\theta_1, \theta_2)$ implies two cases: $s_t^{i,j,c,c-1} = 0$ or $s_t^{i,j,c,c-1} > 0$, $w_t^{i,j,c+1,c} = -K$, $t \in (\theta_1, \theta_2)$. We can show that in the first case any feasible control function $u_t^{i,j,c,c-1}$ will not affect other entries of s_t and the second case only occurs when $c = i$ and $(\theta_1, \theta_2) = (t_0, t_f)$.

Solutions of the necessary conditions

Throughout this section we understand that the degrees are in the bounded range $i \vee j < M^{\epsilon}$ unless specified otherwise. A well known fact in the control theorists community is that solving for the optimal (s_t, u_t, w_t) from the necessary conditions presented in proposition 6 is difficult, especially analytically. In what follows, we solve for the optimal (s_t, u_t, w_t) in three steps.

First, solve the the two-point boundary value problem (TPBVP) consisting of the differential equations for s_t in Eq (42) and w_t in condition (2) of proposition 6 where for s_t the boundary values are given at $t = t_0$ and for w_t at $t = t_f$ as follows.

$$\begin{aligned} \frac{d}{dt}s_t &= h(s_t; u_t), \\ s_{t_0}^{i,j,c,l} &= p(i,j,c)\mathbb{1}_{(l=0)} \qquad \forall (i,j,c,l) \in \Gamma^{\epsilon}, \\ \frac{d}{dt}w_t &= h'(w_t; u_t), \\ w_{t_f}^{i,j,c,l} &= vj - 1 \qquad \forall (i,j,c,l) \in \Gamma^{\epsilon}, \end{aligned} \tag{57}$$

and additionally the optimal control policy u_t satisfies Eq (50) in condition 1 of proposition 6. We solve for (s_t, u_t, w_t) in terms of the auxiliary variables (v, t_f, t_s) and in the following we call these expressions in terms of (v, t_f, t_s) as the solutions of (s_t, u_t, w_t).

Second, because the optimal (s_t, u_t, w_t) satisfies the two equations in the necessary conditions of proposition 6, i.e. the Hamiltonian function Eq (55) at $t = t_f$ and the terminal condition

Eq (56), while minimizing the objective function Eq (43), we have the following optimization problem for (s_t, u_t, w_t):

$$\min_{s_t,u_t,w_t} \quad K \cdot it_{t_f}(u,p) + d_{t_f}(u,p) \tag{58}$$

$$\text{st} \quad \mathcal{H}(s_{t_f}, u_{t_f}, w_{t_f}) = -ve^{-t_f} \tag{59}$$

$$d^-_{t_f} = 0, \tag{60}$$

$$\tag{61}$$

where

$$\begin{aligned} r_{t_f}(u,p) &= \int_{t_0}^{t_f} \sum_{i,j,1\le c\le i} (i-c+1) s_t^{i,j,c,c-1} u_t^{i,j,c,c-1} dt, \\ d_{t_f}(u,p) &= \sum_{i,j,0\le c\le i} p(i,j,c) - \sum_{(i,j,c,l)\in\Gamma^\epsilon} s_{t_f}^{i,j,c,l}, \\ d^-_{t_f} &= \sum_{i,j,0\le c\le i} jp(i,j,c) - \sum_{(i,j,c,l)\in\Gamma^\epsilon} js_{t_f}^{i,j,c,l} - \lambda(1-e^{t_0-t_f}). \end{aligned} \tag{62}$$

After substituting (s_t, u_t, w_t) expressed in terms of (v, t_f, t_s) into the optimization problem (58), we are able to solve the optimal (v, t_f, t_s) based on which we can calculate the optimal (s_t, u_t, w_t).

Third, we change the variables from (v, t_f, t_s) to (v, y, z) for further simplification and attain the optimization problem Eq (86) later.

Now we carry out the first step. The main difficulty of solving problem Eq (57) comes from the fact that w_t and s_t depend on u_t which depends on w_t and s_t recursively through Eq (50). To disentangle the recursive dependence, the idea is to analyze the properties of s_t based on which we can either derive the properties of w_t or the explicit solutions of w_t depending on signs of $vj - 1 + K$. Then by Eq (50) we attain the optimal control policy u_t which leads to the solution of s_t.

It turns out that we only need w_{t_f} as well as u_t and s_t in (58). From problem Eq (57) we know that $w^\beta_{t_f} = vj - 1$ for $\beta \in \Gamma^\epsilon$. For u_t and s_t, we give out their solutions in the following directly due to the limited space of the paper. The solutions of u_t and s_t can be verified by substituting into problem Eq (57).

Lemma 2. The optimal control policy u_t in terms of the variables (v, t_f, t_s) is given as below. For $1 \le c \le i$ except $c = i$ and $vj - 1 = -K$, $\forall t \in [t_0, t_f]$,

$$u_t^{i,j,c,c-1} = \mathbb{1}_{(t \ge t^{i,j,c})}, \tag{63}$$

where

$$t^{i,j,c} = \begin{cases} t_f & \text{if } vj-1 \ge -K \\ t_f + \ln\left(1 + \dfrac{K+vj-1}{(i-c)K}\right) & \text{if } vj-1 < -K \text{ and } 1 \le c < i + \dfrac{K+vj-1}{K(1-e^{t_0-t_f})} \\ t_0 & \text{otherwise,} \end{cases} \tag{64}$$

If $vj - 1 = -K$, $\forall t \in [t_0, t_f]$,

$$u_t^{i,j,i,i-1} = \mathbb{1}_{(t \geq t_s)} \text{ for some } t_s \in [t_0, t_f]. \tag{65}$$

The following is the solution for s_t.

Lemma 3. Letting $p(i, j, i + 1) = 0$, under the optimal control policy in lemma 2, we have the following solutions of s_t for the two-point boundary value problem (TPBVP).

1. For $2 \leq c \leq i$, $0 \leq l \leq c - 2$ and $c = 1$, $l = 0$,

$$s_t^{i,j,c,l} = p(i,j,c)\binom{i}{l}(e^{t_0-t})^{i-l}(1-e^{t_0-t})^l. \tag{66}$$

2. If $vj - 1 < -K$, consider $t \in [t^{i,j,h}, t^{i,j,h-1})$, for some $1 \leq h \leq i$ where

$$t^{i,j,h} = \begin{cases} t_f & \text{if } h = 0 \\ t_f + \ln\left(1 + \dfrac{K + vj - 1}{(i-h)K}\right) & \text{if } 1 \leq h < i + \dfrac{K + vj - 1}{K(1 - e^{t_0 - t_f})} \\ t_0 & \text{otherwise.} \end{cases} \tag{67}$$

If $1 \leq c < h$,

$$s_t^{i,j,c,c-1} = p(i,j,c)\binom{i}{c-1}(e^{t_0-t})^{i-c+1}(1-e^{t_0-t})^{c-1}. \tag{68}$$

If $h \leq c \leq i + 1$,

$$\begin{aligned} s_t^{i,j,c,c-1} &= \binom{i}{c-1}(e^{t_0-t})^{i-c+1}\sum_{m=h}^{c} p(i,j,m) \\ &\sum_{n=0}^{m-1}\binom{c-1}{n}\left(1-e^{t_0-t^{i,j,m}}\right)^n(e^{t_0-t^{i,j,m}} - e^{t_0-t})^{c-1-n}. \end{aligned} \tag{69}$$

3. If $vj - 1 > -K$, for $1 \leq c \leq i + 1$, $t \in [t_0, t_f]$,

$$s_t^{i,j,c,c-1} = p(i,j,c)\binom{i}{c-1}(e^{t_0-t})^{i-c+1}(1-e^{t_0-t})^{c-1}. \tag{70}$$

4. If $vj - 1 = -K$, for $1 \leq c \leq i$, $t \in [t_0, t_f]$,

$$s_t^{i,j,c,c-1} = p(i,j,c)\binom{i}{c-1}(e^{t_0-t})^{i-c+1}(1-e^{t_0-t})^{c-1}, \tag{71}$$

and

$$s_t^{i,j,i+1,i} = p(i,j,i)\mathbb{1}_{\{t_s \leq t\}}[(1-e^{t_0-t})^i - (1-e^{t_0-t_s})^i], \tag{72}$$

where $t_s \in [t_0, t_f]$.

Since Eqs (56) and (55) require the state variable value particularly at $t = t_f$, we can apply lemma 3 at $t = t_f$ to attain s_{t_f}. Next we proceed to the second step, i.e. we substitute (s_t, u_t, w_{t_f}) in terms of (v, t_f, t_s) into the optimization program (58) which leads to the following results.

Proposition 7. Based on the solutions of optimal s_t in lemma 3 (particularly s_{t_f}), u_t in lemma 2 and $w^{\alpha}_{t_f} = vj - 1, \forall \alpha \in \Gamma^{c}$, letting

$$t^{i,j,c} = \begin{cases} t_f & \text{if } K + vj - 1 \geq 0 \text{ or } c = 0 \\ t_f + \ln(1 + \frac{K + vj - 1}{(i-c)K}) & \text{if } K + vj - 1 < 0 \text{ and } 1 \leq c < i + \frac{K + vj - 1}{K(1 - e^{t_0 - t_f})}, \\ t_0 & \text{otherwise,} \end{cases} \tag{73}$$

the Hamiltonian equation Eq (60) at $t = t_f$ is equivalent to

$$\begin{aligned} &\sum_j \max\{-K, vj - 1\} \sum_i i \sum_{c=1}^{i} p(i,j,c) \sum_{m=c}^{i} \binom{i-1}{m-1} (e^{t_0 - t_f})^{i-m+1} \\ &\sum_{n=0}^{c-1} \binom{m-1}{n} \left(1 - e^{t_0 - t^{i,j,c}}\right)^{n} (e^{t_0 - t^{i,j,c}} - e^{t_0 - t_f})^{m-1-n} \\ &= v\lambda e^{t_0 - t_f}. \end{aligned} \tag{74}$$

The terminal condition Eq (61) is equivalent to

$$\begin{aligned} &\sum_i \sum_j j \{ \sum_{c=0}^{i} p(i,j,c) \sum_{n=c}^{i} \binom{i}{n} (1 - e^{t_0 - t^{i,j,c}})^{n} (e^{t_0 - t^{i,j,c}})^{i-n} \\ &-\mathbb{1}_{(vj-1=-K)} p(i,j,i) [(1 - e^{t_0 - t_f})^{i} - (1 - e^{t_0 - t_s})^{i}] \} \\ &= \lambda (1 - e^{t_0 - t_f}). \end{aligned} \tag{75}$$

And the objective function Eq (59) becomes

$$\begin{aligned} &K \cdot r_{t_f}(u,p) + d_{t_f}(u,p) \\ &= K \sum_i \sum_j \{ \sum_{c=1}^{i} p(i,j,c) \\ &\sum_{m=c}^{i} \sum_{n=0}^{c-1} (m - c + 1) \binom{i}{m} \binom{m}{n} (e^{t_0 - t_f})^{i-m} (1 - e^{t_0 - t^{i,j,c}})^{n} (e^{t_0 - t^{i,j,c}} - e^{t_0 - t_f})^{m-n} \\ &+\mathbb{1}_{(vj-1=-K)} p(i,j,i) [(1 - e^{t_0 - t_f})^{i} - (1 - e^{t_0 - t_s})^{i}] \} \\ &+ \sum_j \sum_i \{ \sum_{c=0}^{i} p(i,j,c) \sum_{n=c}^{i} \binom{i}{n} (1 - e^{t_0 - t^{i,j,c}})^{n} (e^{t_0 - t^{i,j,c}})^{i-n} \\ &-\mathbb{1}_{(vj-1=-K)} p(i,j,i) [(1 - e^{t_0 - t_f})^{i} - (1 - e^{t_0 - t_s})^{i}] \}. \end{aligned} \tag{76}$$

For the third step, we further simplify the expressions in proposition 7. Define

$$\begin{aligned} y &:= 1 - e^{t_0 - t_f} = \frac{\tau_f}{\lambda}, \\ z &:= 1 - e^{t_0 - t_s}, \\ x^{i,j,c,c-1} &:= 1 - e^{t_0 - t^{i,j,c}} \\ &= \begin{cases} y & \text{if } K + vj - 1 \geq 0 \text{ or } c = 0 \\ 1 - (1-y)\dfrac{(i-c)K}{(i-c+1)K + vj - 1} & \text{if } K + vj - 1 < 0 \text{ and } 1 \leq c < i + \dfrac{K + vj - 1}{Ky} \\ 0 & \text{otherwise,} \end{cases} \end{aligned} \tag{77}$$

where the first equality follows from $t_0 = -\ln \lambda$ and $t_f = -\ln(\lambda - \tau_f)$. Because $t_0 \leq t_s \leq t_f$ and the function $1 - e^{t_0 - t}$ is increasing in t, $0 \leq z \leq y \leq 1$. As a result, we can substitute the new variables (y, v, z) into the objective function Eq (59), the Hamiltonian equation Eq (60) and the terminal condition Eq (61). Moreover, we add the definition of $x^{i,j,c,c-1}$ and $0 \leq z \leq y \leq 1$. Then we obtain a new optimization problem defined as Eq (86) in section Main results. After solving Eq (86) for (y^*, v^*, z^*), we are able to calculate u_t^* and s_t^* (or u_τ^* and s_τ^* after changing the time index) in order to present theorem 2 and theorem 3.

Main results

Contagion process with no interventions. We first present the theorem when no interventions are provided in the contagion process. For $\epsilon > 0$, recall M^ϵ is defined as in definition 5 and note that all indexes (i, j) are in the range $i \vee j < M^\epsilon$ in what follows.

Definition 9. (I function) Define a function I: $[0, 1] \to [0, 1]$ as

$$I(y) := \frac{1}{\lambda} \sum_{i \vee j < M^\epsilon} j \sum_{c=0}^{i} p(i, j, c) \mathbb{P}(\mathrm{Bin}(i, y) \geq c) \tag{78}$$

where $\mathrm{Bin}(i, y)$ denotes a binomial random variable with i trials and the probability of occurrence y, so $\mathbb{P}(\mathrm{Bin}(i, y) \geq c) = \sum_{m=c}^{i} \binom{i}{m} y^m (1-y)^{i-m}$. $I(y)$ is constructed to represent the asymptotic scaled total out degree from the default set at scaled time y and attains its form Eq (78) from the solution of a set of differential equations.

Since $I(0) = \frac{1}{\lambda} \sum_{i \vee j < M^\epsilon} j p(i, j, 0) \geq 0$, and from the definition of λ,

$$I(1) = \frac{1}{\lambda} \sum_{i \vee j < M^\epsilon} j \sum_{c=0}^{i} p(i, j, c) \leq 1, \tag{79}$$

and $I(y)$ is continuous and increasing, it has at least one fixed point in $[0, 1]$. Further define

$$J(y) := \sum_{i \vee j < M^\epsilon} \sum_{c=0}^{i} p(i, j, c) \mathbb{P}(\mathrm{Bin}(i, y) \geq c). \tag{80}$$

Theorem 1. (Extends from theorem 3.8 of [24]) Consider a sequence of networks with initial conditions $(P_n)_{n \geq 1}$ satisfying assumption 1 where $p = (p(i, j, c))_{i,j,0 \leq c \leq i}$ such that $p(i, j, c) = 0$ for $i \vee j \geq M^\epsilon$, $0 \leq c \leq i$ and no interventions are implemented, let $y^* \in [0, 1]$ be the smallest fixed point of I.

1. If $y^* = 1$, then asymptotically almost all nodes default during the contagion process, i.e.

$$\frac{D_{T_n}}{n} \xrightarrow{p} 1. \tag{81}$$

2. If $y^* < 1$ and it is a stable fixed point, i.e. $I'(y^*) < 1$, then asymptotically the fraction of final defaulted nodes

$$\frac{D_{T_n}}{n} \xrightarrow{p} J(y^*). \tag{82}$$

Particularly, if $I(0) = 0$ and $I'(0) < 1$, then

$$\frac{D_{T_n}}{n} \xrightarrow{p} \sum_{i \vee j < M^\epsilon} p(i, j, 0). \tag{83}$$

Remark 8. Theorem 1 states that the stopping time of the default contagion process is fully governed by the smallest fixed point of the function $I(y)$ and the asymptotic fraction of defaulted nodes at the end of the process can be 1, 0 or a fractional number, representing almost all nodes default, almost no nodes default or a partial number of nodes default, respectively.

Contagion process with interventions. We present the theorems for the contagion process with interventions as the result of solving the finite dimensional optimal control problem (FOCP) in definition 8. For $\epsilon > 0$, recall M^ϵ is defined as in definition 5. By lemma 1, the optimal objective function value of (FOCP) is within $(K + 1)\epsilon$ distance to the optimal objective function value of the infinite dimensional Eq (39). That is why we solved (FOCP) and present the results below with (i, j) in the range $i \vee j < M^\epsilon$ in the following. From remark 6, for a given vector $p = (p(i, j, c))_{0 \le i, 0 \le j, 0 \le c \le i}$, the restriction of (i, j) to $i \vee j < M^\epsilon$ indicates that we use only $p(i, j, c)$, $i \vee j < M^\epsilon$, $0 \le c \le i$ in the calculation. It is equivalent to setting $p(i, j, c) = 0$, $i \vee j \ge M^\epsilon$, $0 \le c \le i$ while keeping $p(i, j, c)$, $i \vee j < M^\epsilon$, $0 \le c \le i$ unchanged, which implies asymptotically nodes with $i \vee j \ge M^\epsilon$ are all invulnerable (Note all $p(i, j, c)$ need not to sum up to one).

First we define the optimization problem Eq (86) based on which we present theorem 2 and theorem 3.

Definition 10. ($\tilde{I}$ and $\tilde{J}$ function) Let $x = (x^\beta)_{\beta \in \Phi^\epsilon}$ where $x^\beta = x^\beta(y, v)$ and $p = (p(i, j, c))_{i, j, 0 \le c \le i}$. We define the functions $\tilde{I}(y, v, z)$ and $\tilde{J}(y, v, z)$ as

$$\begin{aligned} \tilde{I}(y, v, z) \quad &= \frac{1}{\lambda} \sum_{i \vee j < M^\epsilon} j \\ &\left[\sum_{c=0}^{i} p(i, j, c) \mathbb{P}(\mathrm{Bin}(i, x^{i,j,c,c-1}) \ge c) - \mathbb{1}_{(vj-1=-K)} p(i, j, i)(y^i - z^i) \right], \end{aligned} \tag{84}$$

$$\begin{aligned} \tilde{J}(y, v, z) \quad &= \sum_{i \vee j < M^\epsilon} \\ &\left[\sum_{c=0}^{i} p(i, j, c) \mathbb{P}(\mathrm{Bin}(i, x^{i,j,c,c-1}) \ge c) - \mathbb{1}_{(vj-1=-K)} p(i, j, i)(y^i - z^i) \right]. \end{aligned} \tag{85}$$

Note we may write $\tilde{I}(y; v, z)$ and $\tilde{J}(y; v, z)$ to indicate that we treat y as the variable and v, z as the fixed parameters. To present the main results, we define the following optimization problem.

Definition 11. Define the following optimization problem.

$$\min_{y,v,z} \quad K \cdot \tilde{r}(y, v, z) + \tilde{J}(y, v, z) \tag{86}$$

$$\text{st} \quad (1-y)\tilde{H}(y, v) = \lambda v(1-y) \tag{87}$$

$$\tilde{I}(y, v, z) = y \tag{88}$$

$$x^{i,j,c,c-1} = \begin{cases} y & \text{if } K + vj - 1 \geq 0 \text{ or } c = 0 \\ 1 - (1-y)\frac{(i-c)K}{(i-c+1)K+vj-1} & \text{if } K + vj - 1 < 0 \\ & \text{and } 1 \leq c < i + \frac{K+vj-1}{Ky} \\ 0 & \text{otherwise} \end{cases} \tag{89}$$

$$\forall (i, j, c, c-1) \in \Phi^{\epsilon} \tag{90}$$

$$\begin{aligned} 0 \leq z \leq y \leq 1 \\ y, v, z \in \mathbb{R}, \end{aligned} \tag{91}$$

where

$$\begin{aligned} \tilde{r}(y, v, z) &= \sum_{i \vee j < M^{\epsilon}} \{ \sum_{c=1}^{i} p(i, j, c) \\ &\quad \sum_{m=c}^{i} \sum_{n=0}^{c-1} (m - c + 1) \mathbb{P}(\text{Multin}(i, x^{i,j,c,c-1}, y - x^{i,j,c,c-1}, 1 - y) = (n, m - n, i - m)) \\ &\quad + \mathbb{1}_{(vj-1=-K)} p(i, j, i)(y^i - z^i) \}, \\ \tilde{H}(y, v) &= \sum_{i \vee j < M^{\epsilon}} \max\{-K, vj - 1\} i \sum_{c=1}^{i} p(i, j, c) \\ &\quad [\mathbb{P}(\text{Bin}(i-1, y) \geq c - 1) - \mathbb{P}(\text{Bin}(i-1, x^{i,j,c,c-1}) \geq c)], \end{aligned} \tag{92}$$

where $\text{Bin}(i, y)$ denotes a binomial random variable in i trials with the probability of occurrence y, so $\mathbb{P}(\text{Bin}(i, y) \geq c) = \sum_{m=c}^{i} \binom{i}{m} y^m (1-y)^{i-m}$ and $\text{Multin}(i, x, y, 1 - x - y) = (a, b, i - a - b)$ denotes a multinomial distribution in i trials with the probabilities of occurrence of each of three types being x, y and $1 - x - y$, and turns out to have a, b and $i - a - b$ occurrences of each type, respectively, so $\mathbb{P}(\text{Multin}(i, x, y, 1 - x - y) = (a, b, i - a - b)) = \binom{i}{a,b,i-a-b} x^a y^b (1 - x - y)^{i-a-b}$.

Remark 9. A feasible solution (y, v, z) has its own meanings for the optimal control problem Eq (39) on the deterministic process $(s_\tau)_{\tau \in [0,\lambda]}$: $y = \frac{\tau_f}{\lambda}$ is the scaled end time of the process; v is an intervention indicator in that we should intervene on nodes with out degree j satisfying $vj - 1 \leq -K$ and v also determines the starting time of the intervention; z specifies the starting time of the intervention for nodes in state $(i, j, i, i - 1)$ when $vj - 1 = -K$. Moreover, the

auxiliary variables $(x^{i,j,c,c-1})_{(i,j,c,c-1)\in\Phi^\epsilon}$ specifies the starting time of interventions for nodes with different (i, j) values.

Then we are ready to present the next main theorem about the optimal control policy.

Theorem 2. For the asymptotic control problem Eq (9), then

1. Consider a sequence of networks with initial conditions $(P_n)_{n\geq 1}$ satisfying assumption 1 where $p = (p(i, j, c))_{i,j,0\leq c\leq i}$ such that $p(i, j, c) = 0$ for $i \vee j \geq M^\epsilon$, $0 \leq c \leq i$.
2. Let $(G_n)_{n\geq 1}$ be the sequence of control policies for the contagion process on the sequence of networks and $(G_n)_{n\geq 1}$ satisfy assumption 2.
3. Let (y^*, v^*, z^*) be the optimal solution for the optimization problem Eq (86).

If $y^* = 1$, or $y^* \in [0, 1)$ and it is a stable fixed point of $\tilde{I}(y; v^*, z^*)$, i.e. $\frac{d}{dy}\tilde{I}(y^*; v^*, z^*) < 1$, the optimal control policy $G_n^* = (g_1^{(n)*}, \ldots, g_m^{(n)*})$ satisfies that for $0 \leq k \leq m-1$,

$$g_{k+1}^{(n)*}(s, w) = \begin{cases} \mathbb{1}_{(k\geq n\lambda(x^*)^{i,j,c,c-1})} & \text{if } w = (i,j,c,c-1) \in \Phi^\epsilon \text{ except} \\ & c = i \text{ and } v^*j - 1 = -K \\ \mathbb{1}_{(k\geq n\lambda z^*)} & \text{if } w = (i,j,i,i-1) \text{ and } v^*j - 1 = -K \\ 0 & \text{otherwise,} \end{cases} \tag{93}$$

where

$$(x^*)^{i,j,c,c-1} = \begin{cases} y^* & \text{if } K + v^*j - 1 \geq 0 \text{ or } c = 0 \\ 1 - (1-y^*)\frac{(i-c)K}{(i-c+1)K+v^*j-1} & \text{if } K + v^*j - 1 < 0 \text{ and} \\ & 1 \leq c < i + \frac{K+v^*j-1}{Ky^*} \\ 0 & \text{otherwise,} \end{cases} \tag{94}$$

for $(i, j, c, c-1)\in\Phi^\epsilon$.

The next theorem states conclusions for the asymptotic fraction of final defaulted nodes under the optimal policy satisfying theorem 2.

Theorem 3. For the asymptotic control problem Eq 9, then

1. Consider a sequence of networks with initial conditions $(P_n)_{n\geq 1}$ satisfying assumption 1 where $p = (p(i, j, c))_{i,j,0\leq c\leq i}$ such that $p(i, j, c) = 0$ for $i \vee j \geq M^\epsilon$, $0 \leq c \leq i$.
2. Let $(G_n)_{n\geq 1}$ be the sequence of control policies for the contagion process on the sequence of networks and $(G_n)_{n\geq 1}$ satisfy assumption 2.
3. Let (y^*, v^*, z^*) be the optimal solution for the optimization problem Eq (86).

Then under the optimal control policy G_n^* as in theorem 2, we have the following conclusions for the asymptotic fraction of final defaulted nodes.

1. If $y^* = 1$, then asymptotically almost all nodes default during the default contagion process, i.e.

$$\frac{D_{T_n}}{n} \xrightarrow{p} 1. \tag{95}$$

2. If $y^* \in [0, 1)$ and it is a stable fixed point of $\tilde{I}(y; v^*, z^*)$, i.e. $\frac{d}{dy}\tilde{I}(y^*; v^*, z^*) < 1$, then asymptotically the fraction of final defaulted nodes

$$\frac{D_{T_n}}{n} \xrightarrow{p} \tilde{J}(y^*, v^*, z^*), \tag{96}$$

where $x^{i,j,c,c-1}$ in $\tilde{J}$ is

$$x^{i,j,c,c-1} = \begin{cases} y^* & \text{if } K + v^* j - 1 \geq 0 \text{ or } c = 0 \\ 1 - (1 - y^*)\frac{(i-c)K}{(i-c+1)K + v^* j - 1} & \text{if } K + v^* j - 1 < 0 \\ & \text{and } 1 \leq c < i + \frac{K + v^* j - 1}{K y^*} \\ 0 & \text{otherwise,} \end{cases} \tag{97}$$

for $(i, j, c, c-1) \in \Phi^\epsilon$. Particularly, if $y^* = 0$ and $\frac{d}{dy}\tilde{I}(0; v^*, z^*) < 1$, then

$$\frac{D_{T_n}}{n} \xrightarrow{p} \sum_{i \vee j < M^\epsilon} p(i, j, 0), \tag{98}$$

i.e. the final defaulted nodes only consist of the initially defaulted nodes.

In theorem 3 the first case indicates that the network is highly vulnerable and interventions are costly, then the regulator rather lets the whole network default without implementing any interventions, while in the second case interventions are less expensive or the contagion effect is not as high, it is better for the regulator to implement interventions to counteract the contagion process.

Discussion and summary. The key to solve Eq (86) depends on solving the two equations Eqs (88) and (89). First we claim that the optimal v^* for Eq (86) must be nonpositive.

Lemma 4. For Eq (86), the optimal $v^* \leq 0$.

Here we give an algorithm to solve Eq (86) numerically.

Algorithm 1. Solving Eq (86) numerically.

1. Assume $v = 0$,
   a. if $K = 1$, then solve Eqs (88) and (89) by e.g. Newton's method, for y and z;
   b. if $K \neq 1$, then Eq (86) does not depend on z, so solve for y and let $z \leq y$ arbitrary.
2. Assume $v < 0$,
   a. if $K = 1$, then Eq (86) does not depend on z, so solve for y and v such that $v < 0$ and let $z \leq y$ arbitrary;
   b. if $K \neq 1$,
   c. let $y = z$ and solve Eqs (88) and (89) for y and v such that $0 \leq y \leq 1$ and $v < 0$;
   d. if additionally $K > 1$, let $v = \frac{1-K}{j}$ for $j > 0$ and solve Eqs (88) and (89) for y and z such that $0 \leq z \leq y \leq 1$ for each $j \in \{1, \ldots, M^\epsilon\}$.
3. Choose among all the solutions above (if any) the one that minimizes the objective function Eq (87).

Recall that a node is in state (i, j, c, l) if it has in and out degree pair (i, j), the sum of initial equity and accumulative interventions c (called total buffer) and l revealed in-links. Similar to

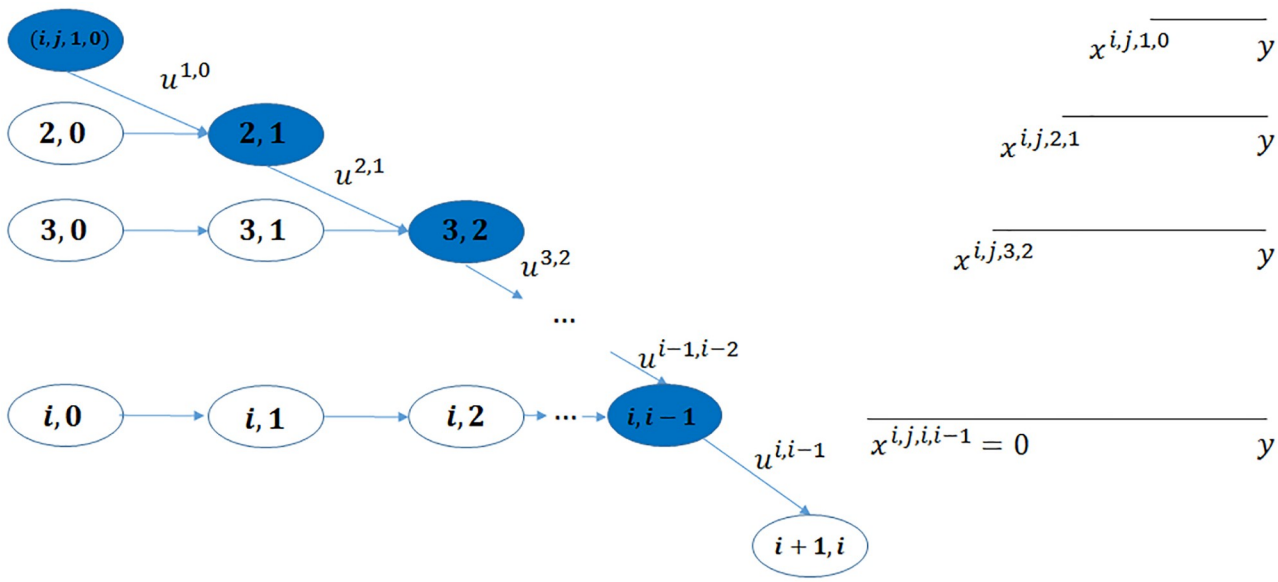

Fig 5. Optimal intervention policy summary where the states indicating one unit of equity remaining are colored in blue.

https://doi.org/10.1371/journal.pone.0209819.g005

[24], we call the in-links to a node that has only one unit of equity remaining ("distance to default" equal to one) as "contagious" links. So a node in state $(i, j, 1, 0)$ has i contagious links and a node in state $(i, j, 2, 1)$ has $i - 1$ contagious links and so on, as shown in Fig 5 the states associated with contagious links are colored in blue. The insights from the optimal interventions policy are summarized as follows.

1. It is never optimal to intervene on a node if it is not selected or has at least two units of remaining equity when selected. Thus the optimal control policy described in theorem 2 only specifies the optimal intervention decision on a node that, when selected, has one unit of equity remaining, i.e. $l = c - 1$. In other words, the use of interventions is to counteract the effects of contagious links.

2. The optimal control policy depends strongly on K, the relative cost of interventions. At a higher K value, interventions are costly and the regulator rather lets the contagion to evolve without any interventions. At a lower K value, the regulator implements more and more interventions, even a "complete" intervention strategy, that is, intervening on every selected node having the "distance to default" of one since the beginning of the process.

3. The optimal control policy is "monotonic" concerning the number of out degree of a node. There exists a cutoff value of the out degree such that it is only optimal to intervene on a node with out degree larger than this cutoff value and not otherwise, regardless of its in degree, total buffer and revealed in-links. For nodes with out degree equal to the cutoff value, we have the singular control case that only those in state $(i, j, i, i - 1)$ needs interventions and the starting time of interventions is determined by the variable z from the optimization problem Eq (86).

4. For nodes that are candidates to receive interventions, the starting time of interventions (depends on the variable $x^{i,j,c,c-1}$) is "monotonic" in terms of the total buffer. The higher the total buffer is, the earlier we should begin to intervene as illustrated in Fig 5 that $x^{i,j,c,c-1}$ is decreasing in c. Moreover, the starting time is also "monotonic" in terms of the in and out degree. For the same out degree, the smaller the in degree is, the earlier the intervention

should begin. If $v < 0$, the larger the out degree is, the earlier the intervention should begin. The economic meaning is that we should focus on systematically important nodes as well as nodes that are close to invulnerability and thus easier to save. In other words, we achieve our objective by protecting the nodes that would incur large impact on the network after defaulting and by bringing easy-to-save nodes into invulnerable states.

5. Once we have begun intervening on a node we should keep implementing interventions on it every time it is selected. In other words, we do not allow nodes that have received interventions to default. This is an interesting result. In the partial information setting, [3] consider the interventions on banks that record more write-downs later and default as the "wasted government money" and mention the tradeoff associated with intervention: potentially wasted money versus less capital depletion. But our finding is that there is no wasted money under the optimal policy and the tradeoff is the high intervention cost versus less magnitude of defaults.

Indeed, following the optimal policy, we are able to achieve earlier termination time of the contagion process and smaller fraction of final defaulted nodes. We can quantify the improvement by comparing $\tilde{I}$ and $\tilde{J}$ in theorem 3 with I and J defined in theorem 1, respectively. Note in the following we suppress the apostrophe "$*$" and the indexes (i, j) are in the range $i \vee j < M^{\epsilon}$.

1. $\tilde{I}(y, v, z)$ plays the same role as $I(y)$ in theorem 1, which represents the asymptotic scaled total out degree from the default set at scaled time y. Since

$$\begin{aligned} & I(y) - \tilde{I}(y, v, z) \\ = \; & \frac{1}{\lambda} \sum_{i \vee j < M^{\epsilon}} j \{ \sum_{c=0}^{i} p(i, j, c) [\mathbb{P}(\mathrm{Bin}(i, y) \geq c) - \mathbb{P}(\mathrm{Bin}(i, x^{i,j,c,c-1}) \geq c)] \\ & + \mathbb{1}_{(vj-1=-K)} p(i, j, i)(y^i - z^i) \}, \end{aligned} \tag{99}$$

and note that

$$x^{i,j,c,c-1} \leq y \text{ for } (i, j, c, c-1) \in \Phi^{\epsilon}, \tag{100}$$

thus

$$\mathbb{P}(\mathrm{Bin}(i, y) \geq c) - \mathbb{P}(\mathrm{Bin}(i, x^{i,j,c,c-1}) \geq c) \geq 0. \tag{101}$$

Then for the same initial conditions $p = (p(i, j, c))_{i \vee j < M^{\epsilon}, 0 \leq c \leq i}$, the smallest fixed point of $\tilde{I}(y; v^*, z^*)$ is no greater than that of $I(y)$, which implies that the default contagion process under optimal interventions terminates no later than under no interventions.

2. Similarly $\tilde{J}(y, v, z)$ plays the same role as $J(y)$ in theorem 1, which represents the asymptotic fraction of final defaulted nodes under the optimal control policy. The difference

$$\begin{aligned} & J(y) - \tilde{J}(y, v, z) \\ = \; & \sum_{i \vee j < M^{\epsilon}} \{ \sum_{c=0}^{i} p(i, j, c) [\mathbb{P}(\mathrm{Bin}(i, y) \geq c) - \mathbb{P}(\mathrm{Bin}(i, x^{i,j,c,c-1}) \geq c)] \\ & + \mathbb{1}_{(vj-1=-K)} p(i, j, i)(y^i - z^i) \} \geq 0 \end{aligned} \tag{102}$$

quantifies the fraction of nodes that are prevented from default because of the optimal control policy.

Numerical experiments

Introduction

While the theoretical framework described before allows heterogeneous networks with degree sequences and initial equity levels drawn from arbitrary distributions, we present a relatively simplified case in numerical experiments for illustration purpose, in which the degree sequences and initial equity levels satisfy specific distributions. Consider a sequence of networks with the number of nodes n growing to infinity, whose in and out degrees are between 1 and 10, and each node's in degree equal to its own out degree, i.e. $d^-(v) = d^+(v)$, $v \in [n]$, respectively, so we call either the in or out degree as the degree of the node. This allows us to combine two indexes i and j into one index i, so the state of a node becomes (i, c, l) and the empirical probability P_n and the limiting probability p of the degree and initial equity become $P_n(i, c)$ and $p(i, c)$ respectively. Additionally we assume the initial equity levels between 1 and 10. In sum, we consider the degree and initial equity level pair in the set $\Gamma' := \{(i, c) \in \mathbb{N}_0^2 : 1 \leq i \leq 10, 0 \leq c \leq 10\}$.

Next we decide on the limiting probability p. Note that Γ' contains three initial types of nodes: defaulted (with $c = 0$), vulnerable (with $c \leq i$) and invulnerable (with $c > i$). In this numerical experiment, we set the total fraction of initial defaults as ξ and assume the fraction of initial defaults is the same across all degrees, i.e. $p(i, 0) = \frac{\xi}{10}$ for $i \in [1, 10]$. For the initially liquid nodes, the joint probability of the degree and initial equity conditional on being liquid is constructed through a binormal copula with correlation ρ and two marginal probabilities. The marginal probabilities of the degree and initial equity are assumed to follow the Zipf's law, i.e.

$$\begin{aligned} \mathbb{P}(\deg = i) &= \frac{i^{-(1+a_1)}}{\sum_{l=1}^{10} l^{-(1+a_1)}}, \\ \mathbb{P}(\text{initial equity} = c) &= \frac{c^{-(1+a_2)}}{\sum_{l=1}^{10} l^{-(1+a_2)}}, \end{aligned} \tag{103}$$

where $a_1, a_2 > 0$. The Zipf's law is a form of the power law with Pareto tails, which is observed for the distribution of the degrees and equity levels of the financial networks in many empirical studies, see e.g. [19, 29].

In a network of size n with the joint probability $P_n(i, c)$ of the degree and initial equity, a contagion process under interventions occurs as described in section Dynamics. Recall that we only need to consider intervening on a node that, when selected, has only one unit of equity left, i.e. a node with "distance to default" equal to one. Here we consider two types of intervention policies, the optimal policy and the alternative policy: intervening on nodes with degree between 8 and 10 and "distance to default" equal to one from the beginning of the process. The alternative policy implies interventions on nodes with high degrees and close to default, representing the usual policy employed by the central bank or government in a real financial crisis.

Our objective is to verify the convergence in probability of $\frac{R_{T_n}}{n}$ and $\frac{D_{T_n}}{n}$ as well as the convergence of the scaled termination time $\frac{T_n}{m}$ as stated in proposition 5. Moreover, we shall study the convergence rate of the standard deviation and IQR (interquartile range) to examine if the asymptotic variables provide good approximations under realistic n values.

Under the optimal policy in the form given in theorem 2, the limits for $\frac{R_{T_n}}{n}$, $\frac{D_{T_n}}{n}$ and $\frac{T_n}{m}$ as $n \to \infty$ are $\tilde{r}(y^*, v^*, z^*)$, $\tilde{J}(y^*, v^*, z^*)$ and y^*, respectively in Eq (86) where (y^*, v^*, z^*) is the optimal solution. On the other hand, the alternative policy is that for $0 \le k \le m-1$,

$$g_{k+1}^{(n),\text{alt}}(s, w) = \begin{cases} \mathbb{1}_{(k \ge n\lambda x_{\text{alt}}^{i,c,c-1})} & \text{if } w = (i, c, c-1) \in \Phi' \\ 0 & \text{otherwise,} \end{cases} \tag{104}$$

where for $(i, c, c-1) \in \Phi'$,

$$x_{\text{alt}}^{i,c,c-1} = \begin{cases} 0 & \text{if } i \in \{8, 9, 10\} \\ y & \text{otherwise,} \end{cases} \tag{105}$$

and y is the solution of $\frac{1}{\lambda}\sum_{i=0}^{10} i \sum_{c=0}^{i} p(i, c)\mathbb{P}(\text{Bin}(i, x_{\text{alt}}^{i,c,c-1}) \ge c) = y$. Then the limits for $\frac{R_{T_n}}{n}$, $\frac{D_{T_n}}{n}$ and $\frac{T_n}{m}$ as $n \to \infty$ can be calculated as:

$$\begin{aligned} \frac{R_{T_n}}{n} &\xrightarrow{p} \sum_{i=0}^{10}\sum_{c=1}^{i} p(i, c) \\ &\quad \sum_{m=c}^{i}\sum_{n=0}^{c-1}(m - c + 1)\mathbb{P}(\text{Multin}(i, x_{\text{alt}}^{i,c,c-1}, y - x_{\text{alt}}^{i,c,c-1}, 1 - y) \\ &\quad = (n, m - n, i - m)), \\ \frac{D_{T_n}}{n} &\xrightarrow{p} \sum_{i=0}^{10}\sum_{c=0}^{i} p(i, c)\mathbb{P}(\text{Bin}(i, x_{\text{alt}}^{i,c,c-1}) \ge c), \\ \frac{T_n}{m} &\xrightarrow{p} y, \end{aligned} \tag{106}$$

where $\mathbb{P}(\text{Bin}(i, y) \ge c) = \sum_{m=c}^{i}\binom{i}{m} y^m (1-y)^{i-m}$ and $\mathbb{P}(\text{Multin}(i, x, y, 1 - x - y) = (a, b, i - a - b)) = \binom{i}{a,b,i-a-b} x^a y^b (1 - x - y)^{i-a-b}$.

Simulation

The set up. We have the following setup.

1. A sequence of six networks with increasing number of nodes $n \in \{5^4, 6^4, \ldots, 10^4\}$ and there are 100 runs for each network under either intervention policy.
2. To determine the asymptotic fraction $p(\cdot, \cdot)$ of the degree and initial equity pair (i, c) where $(i, c) \in \Gamma'$, we set the following parameters.
   a. The fraction of initial defaults $\xi = 0.5$, indicating half of the nodes have defaulted. As stated before, we assume in this numerical experiment that the fraction of initial defaults is the same across all degrees, thus $p(i, 0) = \frac{\xi}{10}$ for $i \in [1, 10]$.
   b. The probability of the degree and initial equity for liquid nodes $p(i, e)$, $i \in \{1, \ldots, 10\}$, $e \in \{1, \ldots, 10\}$ is determined by a binormal coupula with the exponents of the marginal probabilities of the degree and initial equity $(a_1, a_2) = (0.8, 0.7)$ and the correlation coefficient $\rho = 0.9$. Note that a smaller a_1 indicates larger fraction of nodes with higher degrees, thus higher connectivity and a smaller a_2 indicates larger fraction of nodes with higher initial equities, and ρ implies how likely that higher degree nodes have higher initial equities.

Table 1. Theoretical limits of $\frac{R}{n}$, $\frac{D}{n}$ and $\frac{T}{m}$ under the optimal and alternative policies.

Policies	$\frac{R}{n}$	$\frac{D}{n}$	$K\frac{R}{n}+\frac{D}{n}$	$\frac{T}{m}$
Optimal	0.306	0.503	0.657	0.728
Alternative	0.019	0.821	0.830	0.866

https://doi.org/10.1371/journal.pone.0209819.t001

3. After determining the asymptotic fraction $p(\cdot, \cdot)$, we construct a sequence of empirical fractions $P_n(\cdot, \cdot)$ for each network that converge to $p(\cdot, \cdot)$ by

$$P_n(i, c) = \frac{[np(i, c)]}{n} \qquad (i, c) \in \Gamma', \tag{107}$$

where $[\cdot]$ is the round function. In other words, the number of nodes with degree i and initial equity c are $[np(i, c)]$ for a network of n nodes.

4. We consider two intervention policies described as before.

5. The relative cost for the interventions $K = 0.5$.

In the following we suppress T_n in the subscripts and write $\frac{R}{n}$, $\frac{D}{n}$ and $\frac{T}{m}$ in stead of $\frac{R_{T_n}}{n}$, $\frac{D_{T_n}}{n}$ and $\frac{T_n}{m}$ respectively.

Theoretical results. The theoretical limits of $\frac{R}{n}$, $\frac{D}{n}$ and $\frac{T}{m}$ under the optimal and alternative policies are summarized in Table 1.

We first verify that the objective function $K\frac{R}{n}+\frac{D}{n}$ in Eq (86) is indeed less under the optimal policy. Moreover, compared with the alternative policy, the optimal policy intervens more but results in smaller fraction of final defaulted nodes and ends the contagion process earlier.

Simulation results. We show the plots for $\frac{R}{n}$, $\frac{D}{n}$ and $\frac{T}{m}$ under the optimal and alternative policies as shown in Figs 6–11.

1. Under either policy and for each variable, there are four plots in each figure. The first two plots are two boxplots. The above boxplot visualizes five summary statistics (min, mean −standard deviation, mean, mean+ standard deviation, max) while the bottom boxplot uses another set of summary statistics (1st quartile−1.5IQR, 1st quartile, median, 3rd quartile, 3rd quartile+ 1.5IQR) and the data outside the range are treated as outliers, where IQR stands for interquartile range, i.e. the difference between the third and the first quartiles.

2. The blue dashed horizontal line in each plot indicates the theoretical limits of $\frac{R}{n}$, $\frac{D}{n}$ and $\frac{T}{m}$ based on $p(\cdot, \cdot)$ and the red solid line in each box indicates the theoretical limits of those values calculated with $P_n(\cdot, \cdot)$ for each n. We calculate the theoretical values in both ways because for small n, $P_n(\cdot, \cdot)$ determined by Eq (107) has a relatively large rounding error and thus deviates a bit from $p(\cdot, \cdot)$ but calculating using $P_n(\cdot, \cdot)$ instead of $p(\cdot, \cdot)$ can effectively remove the deviations in the inputs to the model. However, given $p(\cdot, \cdot)$, $P_n(\cdot, \cdot)$ is different for different n values, thus the theoretical values of a variable calculated with $P_n(\cdot, \cdot)$ are also different for different n's.

3. The black dots in the boxplots indicates the results of 100 runs and they are jittered by a random amount left and right to avoid overplotting. From the black dots we can see the distributions of the results. Note that the black dots in the above and bottom boxplots show

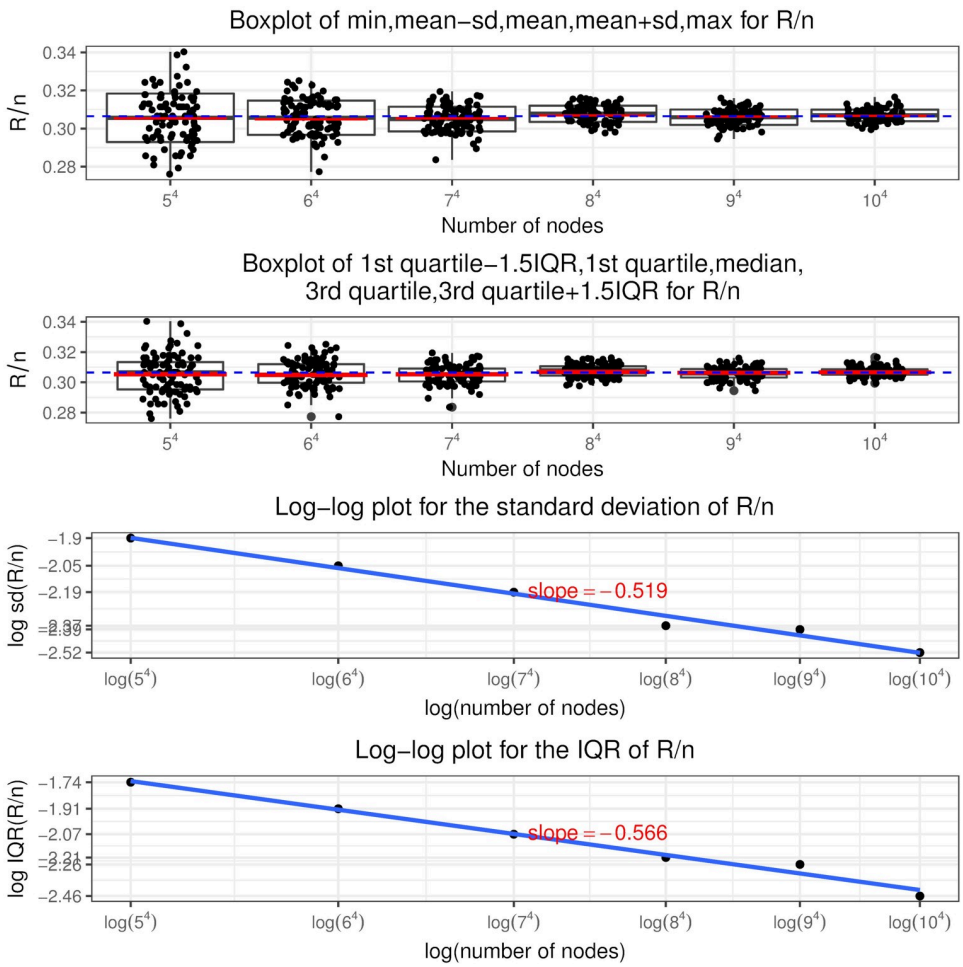

Fig 6. The boxplot and log-log plot of standard deviation and IQR for R/n under optimal policy.

https://doi.org/10.1371/journal.pone.0209819.g006

the same results for each n. They look different because they are jittered left and right by a different random amount.

4. The last two plots in every figure shows the log-log plot of the standard deviation and IQR of $\frac{R}{n}$, $\frac{D}{n}$ and $\frac{T}{m}$ against n and a fitted straight line with the slope.

From the simulation results, we make the following conclusions.

1. From the boxplots of $\frac{R}{n}$, $\frac{D}{n}$ and $\frac{T}{m}$ under both intervention policies, we observe that the mean or median converge to the calculated theoretical value with shrinking standard deviation or IQR. Because the theoretical value is a constant given the joint probability of degree and initial equity $p(\cdot, \cdot)$, convergence of mean to the theoretical value with variance converging to zero is equivalent to convergence in probability, this observation provides evidence for the convergences in probability of $\frac{R}{n}$, $\frac{D}{n}$ and $\frac{T}{m}$ to their theoretical values, thus proving proposition 5 and theorem 3.

2. Be comparing the blue dashed line and the red solid line we see that the mean or median is closer to the red solid line, i.e. the theoretical value calculated with $P_n(\cdot, \cdot)$ instead of $p(\cdot, \cdot)$.

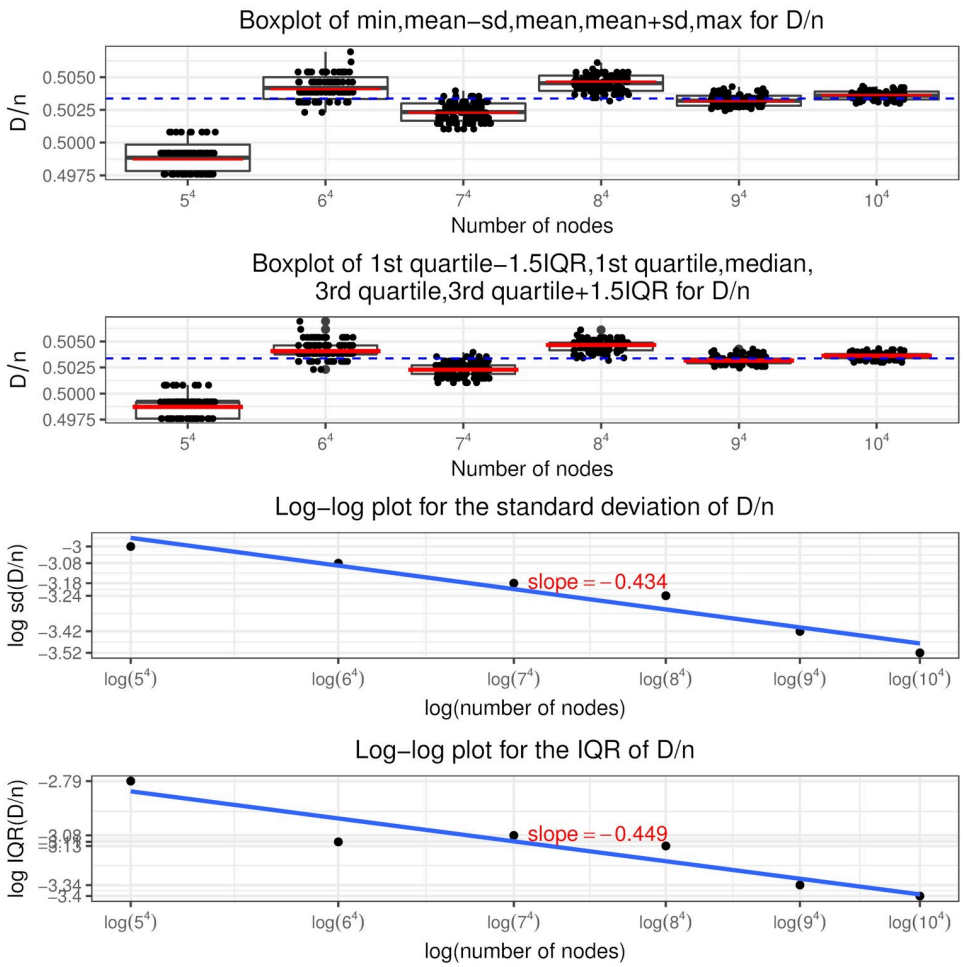

Fig 7. The boxplot and log-log plot of standard deviation and IQR for D/n under optimal policy.

https://doi.org/10.1371/journal.pone.0209819.g007

This reflects the rounding error caused by Eq (107) in the inputs into the calculation. By using the more accurate fraction we observe that the closeness of the mean or median to the theoretical value does not vary in n although the results of different runs are more and more concentrated around the mean or median as n grows.

3. The log-log plots of the standard deviation and IQR of each variable with the fitted straight lines further show that both of them decrease with power law tails, i.e. in the form of $z = Cx^{-a}$ where C is a constant and $a > 0$ is the exponent. The absolute value of the slope of the straight line serves as the exponent. It is interesting to observe that the exponents for the standard deviation and IQR of different variables are in the range $0.4 \sim 0.5$ under both intervention policies. This implies that the dispersions of all variables converge to zero at roughly the same rate under both policies.

Conclusion. From the simulation part we can make the following conclusions.

1. The convergences of $\frac{R}{n}$, $\frac{D}{n}$ and $\frac{T}{m}$ to their theoretical values (stated in proposition 5) are supported by the simulation results. It is worth noting that the closeness of the mean or median

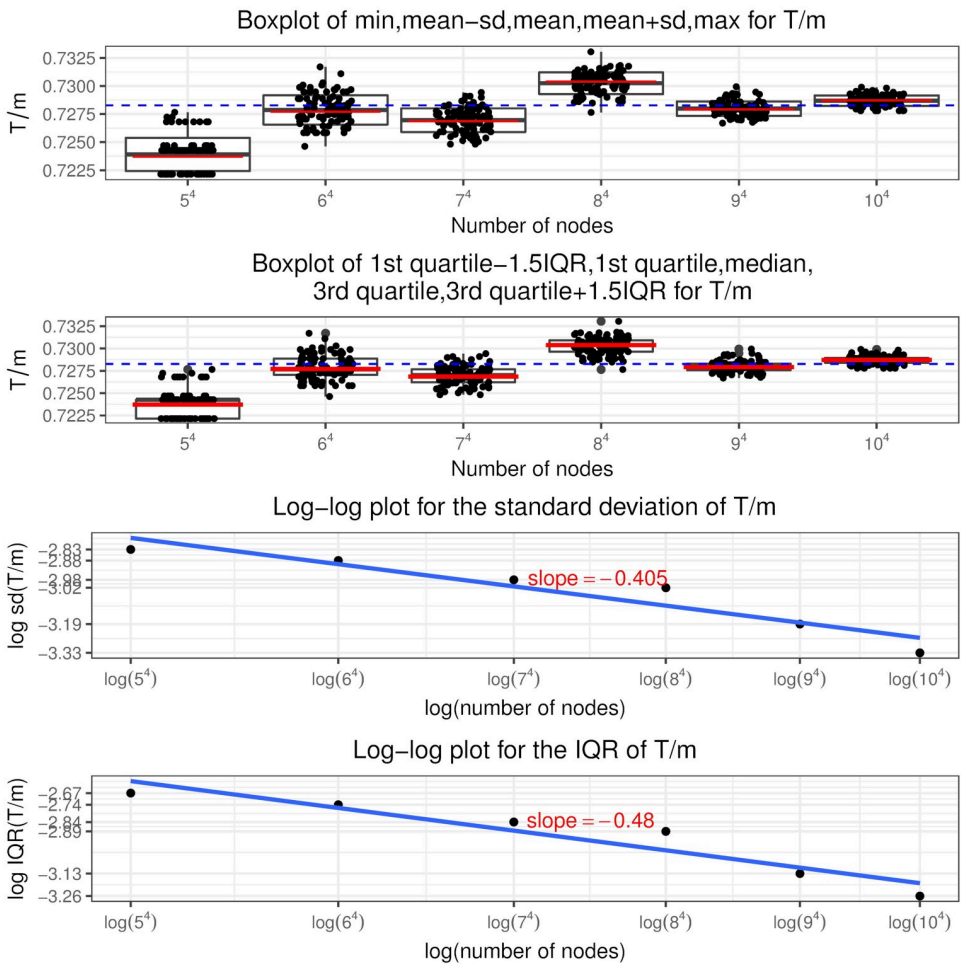

Fig 8. The boxplot and log-log plot of standard deviation and IQR for T/m under optimal policy.

https://doi.org/10.1371/journal.pone.0209819.g008

to the theoretical value does not vary for different n after the rounding error in the initial fractions are removed, but the dispersion of the variable shrinks as n grows.

2. The dispersion of each variable decreases following a power law. The exponents are close to each other under both intervention policies and for all variables, indicating a uniform convergence rate of the dispersions of all the variables under both policies.

Appendix A: Proofs

Proof of proposition 1

We give a proof in words similar to the proof of proposition 3.4 in [2] for a different objective function of optimizing the value of the financial system at the end of the process under some budget constraint. We observe that the objective function J_n depends on the set of defaulted nodes only through its cardinality. Any node will affect the states of other nodes only after it defaults because the set of unrevealed out links of the default set determining the contagion process grows only after a node defaults. And it is possible for a default to occur only when a node has one unit of equity (distance to default equal to one) at the time of being selected.

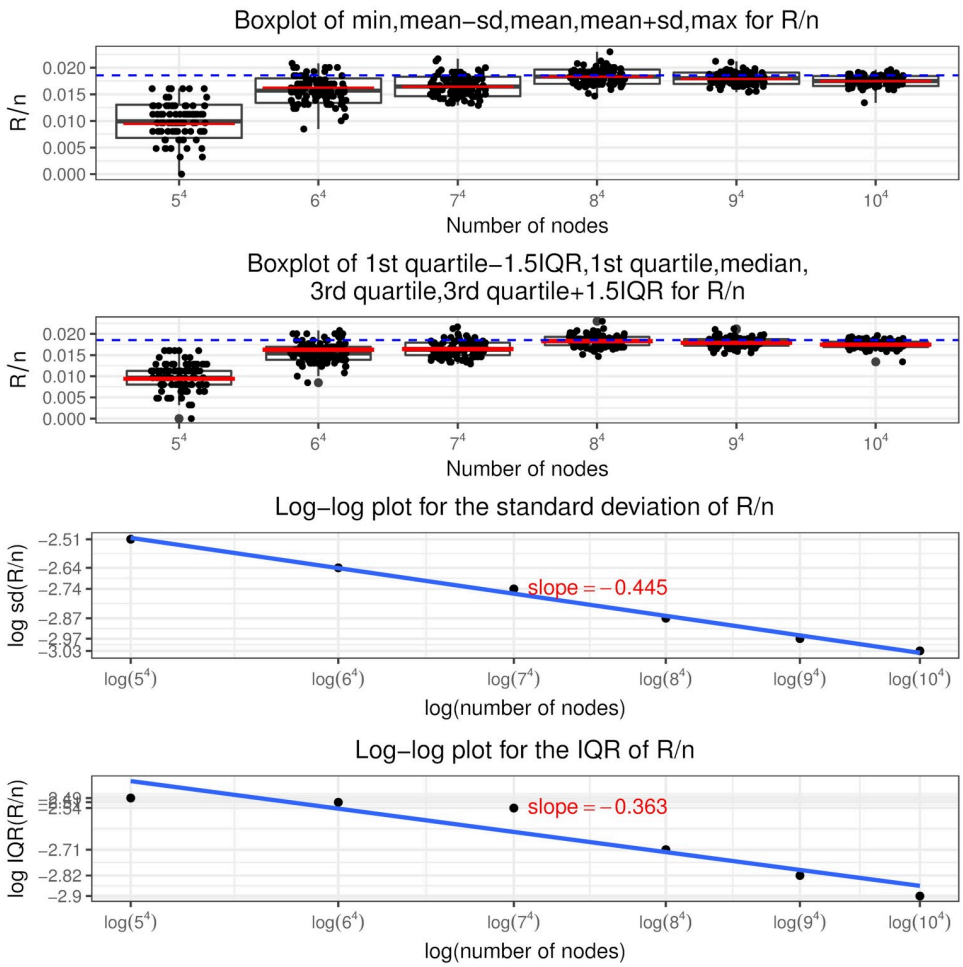

Fig 9. The boxplot and log-log plot of standard deviation and IQR for *R/n* under alternative policy.

https://doi.org/10.1371/journal.pone.0209819.g009

Before that time, the equity only decreases by one every time it is selected. Moreover, there is always a chance to intervene on a node before it defaults. However, if we intervene on a node that is not selected at the current step or has more than one units of remaining equity when selected, it is possible that the node may not be selected in the following steps before the process ends in which case we implemented redundant interventions without reducing the number of defaults.

Proof of proposition 2

Proof. Assume $u_\tau = b := (b^\beta)_{\beta\in\Phi}$ a constant vector for $\tau \in [\tau_1, \tau_2) \subseteq [0, \lambda)$, $0 \le \tau_1 < \tau_2 < \infty$ and $b^\beta \in \{0, 1\}$. Note that the ODEs are "separable" in that $s_\tau^{i,j,c,l}$ only depends on the entries of s_τ with the same (i, j), so we can only focus on the system of ODEs with the same (i, j). For the same (i, j), define $\Gamma_{i,j} := \{(c, l):0 \le l < c \le i \text{ or } c = i + 1, l = i\}$ and suppress (i, j) in the superscripts in definition 4, then we obtain the system of ODEs for $\tau \in [\tau_1, \tau_2)$ with the initial condition $s_{\tau_1} = s_1 := (s_1^{c,l})_{(c,l)\in\Gamma_{i,j}}$.

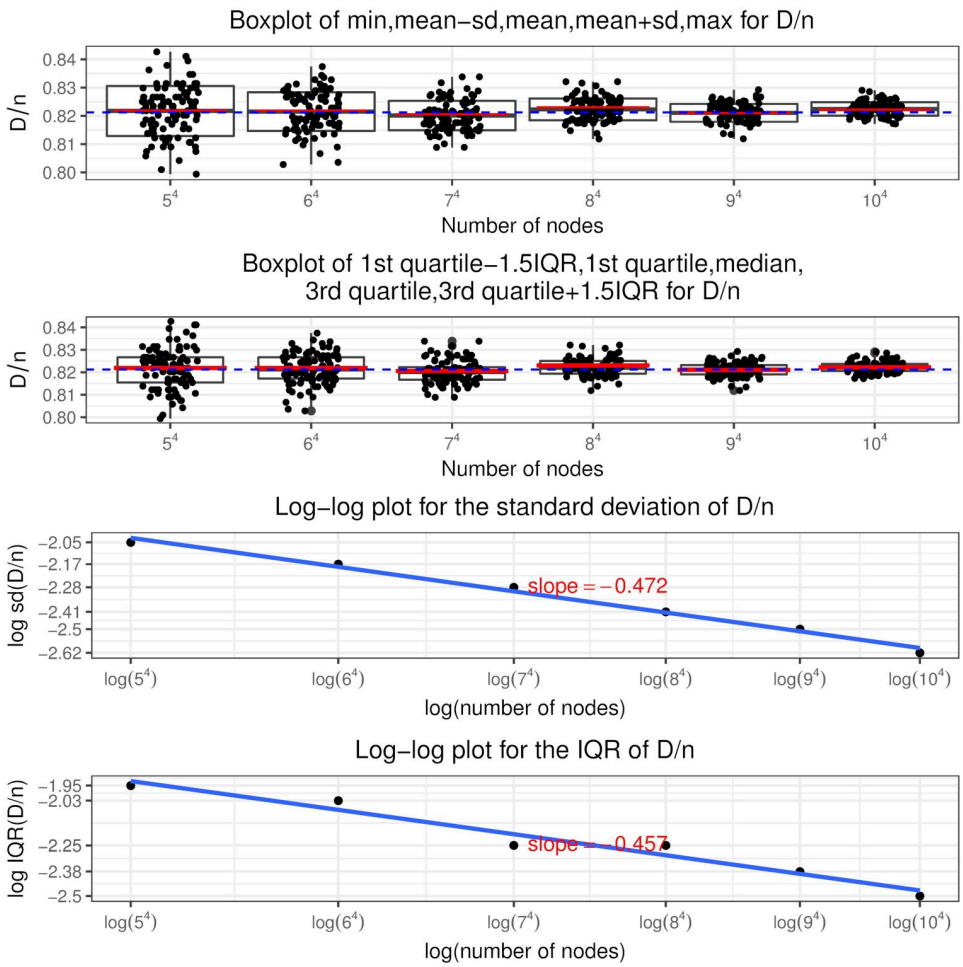

Fig 10. The boxplot and log-log plot of standard deviation and IQR for *D/n* under alternative policy.

https://doi.org/10.1371/journal.pone.0209819.g010

Letting $t = -\ln(\lambda - \tau)$, $t_1 = -\ln(\lambda - \tau_1)$ and $t_2 = -\ln(\lambda - \tau_2)$, we have an autonomous system of ODEs for $t \in [t_1, t_2)$ that

$$\begin{aligned}
\frac{ds_t^{c,0}}{dt} &= -is_t^{c,0} \text{ for } 1 \le c \le i, \\
\frac{ds_t^{c,l}}{dt} &= (i-l+1)s_t^{c,l-1} - (i-l)s_t^{c,l} \\
&\quad \text{for } 3 \le c \le i,\ 1 \le l \le c-2, \\
\frac{ds_t^{c,c-1}}{dt} &= (i-c+2)s_t^{c-1,c-2}b^{c-1,c-2} \\
&\quad +(i-c+2)s_t^{c,c-2} - (i-c+1)s_t^{c,c-1} \\
&\quad \text{for } 2 \le c \le i, \\
\frac{ds_t^{i+1,i}}{dt} &= s_t^{i,i-1}b^{i,i-1},
\end{aligned} \tag{108}$$

with the initial condition $s_{t_1} = s_1$.

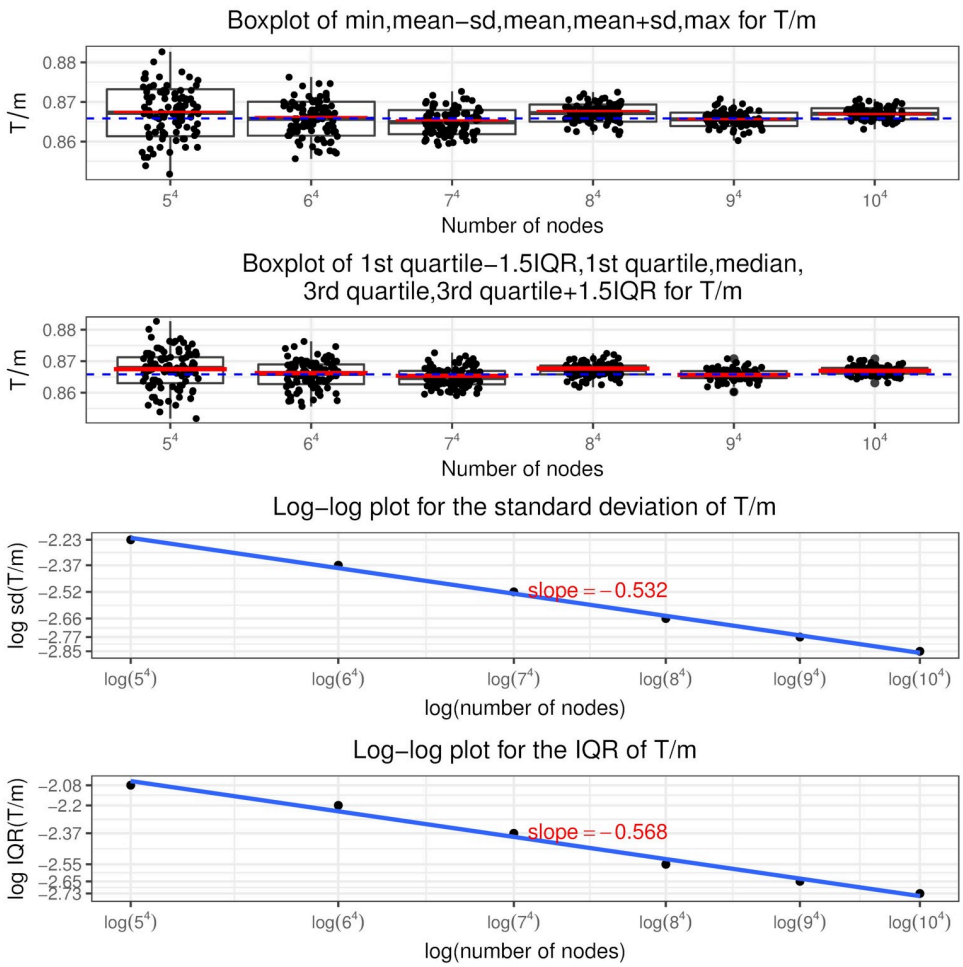

Fig 11. The boxplot and log-log plot of standard deviation and IQR for *T/m* under alternative policy.

https://doi.org/10.1371/journal.pone.0209819.g011

By induction, we can prove the solution s_t on $[t_1, t_2)$ is

$$\begin{aligned} s_t^{c,l} &= e^{(i-l)(t_1-t)} \sum_{r=0}^{l} s_1^{c,r} \binom{i-r}{l-r} (1-e^{t_1-t})^{l-r} \\ &\quad \text{for } 2 \le c \le i,\ 0 \le l \le c-2, \\ s_t^{c,c-1} &= e^{(i-c+1)(t_1-t)} \sum_{r=0}^{c-1} \sum_{q=r+1}^{c} \prod_{k=q}^{c-1} b^{k,k-1} s_1^{q,r} \binom{i-r}{c-1-r} (1-e^{t_1-t})^{c-1-r} \\ &\quad \text{for } 1 \le c \le i, \\ s_t^{i+1,i} &= s_1^{i+1,i} + \sum_{r=0}^{i-1} \sum_{q=r+1}^{i} \prod_{k=q}^{i} b^{k,k-1} s_1^{q,r} (1-e^{t_1-t})^{i-r}, \end{aligned} \tag{109}$$

where $\prod_{k=c}^{c-1} b^{k,k-1} := 1$.

By changing the variable t to τ by $t = -\ln(\lambda - \tau)$, Eqs (15), (16) and (17) follow. Let the initial condition be $s_{\tau_1}^{i,j,c,l} = p(i,j,c)\mathbb{1}_{(l=0)}$ for $(i, j, c, l) \in \Gamma$ at $\tau_1 = 0$, then Eq (18) follows from Eq (15).

Proof of proposition 3

Proof. For the following proof we need to adapt the Wormald's theorem in Appendix B: Wormald's theorem. For notational convenience we suppress the tilde sign for $\tilde{R}$, $\tilde{r}$. Since $\frac{m}{n} \to \lambda$ as $n \to \infty$, for the given ϵ and $\hat{\lambda} = \lambda - \in$, we can find $n_0 \in \mathbb{N}$, such that $0 < \hat{\lambda} < \frac{m}{n} < \lambda + 0.1$ for $n \geq n_0$. Let $z = (z^\alpha)_{\alpha \in \Gamma^\epsilon}$ and

$$U = \left\{(\tau, z, r) \in \mathbb{R}^{|\Gamma^\epsilon|+2} : \ -\epsilon < \tau < \hat{\lambda}, \ -\epsilon < z^\alpha < 1.1\,, -\epsilon < r < \lambda + 0.1\right\}, \tag{110}$$

then U contains the closure of

$$\{(0, z, 0) : \mathbb{P}(S_0^\alpha = z^\alpha n, \ \forall \alpha \in \Gamma^\epsilon, R_0 = 0) \neq 0 \text{ for some } n\}. \tag{111}$$

Define the stopping time $T_U = \min\{1 \leq k \leq m : \left(\frac{k}{n}, \frac{S_k}{n}, \frac{R_k}{n}\right) \notin U\}$.

By definition 1 and definition of R_k, $0 \leq S_k^\alpha \leq n$, $\alpha \in \Gamma^\epsilon$ and $0 \leq R_k \leq (\lambda + 0.1)n$ hold $\forall k \geq 0$ and $n \geq n_0$. Recall that $S_k = (S_k^\alpha)_{\alpha \in \Gamma^\epsilon}$ and $\frac{S_k}{n} = \left(\frac{S_k^\alpha}{n}\right)_{\alpha \in \Gamma^\epsilon}$. The following conditions hold:

1. For $0 \leq k < T_U$ and $\alpha \in \Gamma^\epsilon$,

$$\begin{aligned} |S_{k+1}^\alpha - S_k^\alpha| &\leq 1, \\ |R_{k+1} - R_k| &\leq 1, \end{aligned} \tag{112}$$

   i.e. $\rho_1 = 1$.

2. There exists $\rho_2 = O(n^{-1})$ such that for $0 \leq k < T_U$ and $\alpha \in \Gamma^\epsilon$,

$$\begin{aligned} \left|\mathbb{E}\left(S_{k+1}^\alpha - S_k^\alpha | \mathcal{F}_k\right) - h^\alpha\left(\frac{k}{n}, \frac{S_k}{n}\right)\right| &\leq \rho_2, \\ \left|\mathbb{E}\left(R_{k+1} - R_k | \mathcal{F}_k\right) - h_0\left(\frac{k}{n}, \frac{S_k}{n}\right)\right| &\leq \rho_2, \end{aligned} \tag{113}$$

   where $h = (h^\alpha)_{\alpha \in \Gamma^\epsilon}$ and h_0 are

$$\begin{aligned} h^{i,j,c,l}(t,z) &= \begin{cases} -\dfrac{iz^{i,j,c,0}}{\lambda - t} & \text{if } 1 \leq c \leq i, l = 0 \\ \dfrac{(i-l+1)z^{i,j,c,l-1}}{\lambda - t} - \dfrac{(i-l)z^{i,j,c,l}}{\lambda - t} & \text{if } 3 \leq c \leq i, 1 \leq l \leq c-2 \\ \dfrac{(i-c+2)z^{i,j,c-1,c-2}}{\lambda - t} u_t^{i,j,c-1,c-2} & \\ + \dfrac{(i-c+2)z^{i,j,c,c-2}}{\lambda - t} - \dfrac{(i-c+1)z^{i,j,c,c-1}}{\lambda - t} & \text{if } 2 \leq c \leq i \\ \dfrac{z^{i,j,i,i-1}}{\lambda - t} u_t^{i,j,i,i-1} & \text{if } c = i, l = i-1, \end{cases} \\ h_0(t,z) &= \sum_{(i,j,c,c-1) \in \Phi^\epsilon} \frac{(i-c+1)z^{i,j,c,c-1}}{\lambda - t} u_t^{i,j,c,c-1}. \end{aligned} \tag{114}$$

In particular, Eq (113) follows from

$$\begin{aligned}
&\Bigg|\sum_{(i,j,c,c-1)\in\Phi^{\epsilon}} \frac{(i-c+1)S_k^{i,j,c,c-1}}{m-k} u_{\frac{k}{n}}^{i,j,c,c-1} \\
&\quad - \sum_{(i,j,c,c-1)\in\Phi^{\epsilon}} \frac{(i-c+1)\frac{S_k^{i,j,c,c-1}}{n}}{\lambda-\frac{k}{n}} u_{\frac{k}{n}}^{i,j,c,c-1}\Bigg| \\
\leq &\sum_{(i,j,c,c-1)\in\Phi^{\epsilon}} \left|\frac{(i-c+1)\frac{S_k^{i,j,c,c-1}}{n}}{\frac{m}{n}-\frac{k}{n}} u_{\frac{k}{n}}^{i,j,c,c-1} - \frac{(i-c+1)\frac{S_k^{i,j,c,c-1}}{n}}{\lambda-\frac{k}{n}} u_{\frac{k}{n}}^{i,j,c,c-1}\right| \\
= &\ O(n^{-1}).
\end{aligned} \tag{115}$$

However, for $\beta \in \Phi^{\epsilon}$, h^{β} and h_0 are not Lipschitz continuous because u_{τ}^{β} can have step changes on $[0, \lambda)$. So we need to adapt the proof. Note that u_{τ}^{β} is piecewise constant $\{0, 1\}$ valued function thus $h^{\beta}(\tau, s)$ and $h_0(\tau, s)$ are Lipschitz continuous in each interval where $\tilde{u} = (u^{\beta})_{\beta\in\Phi^{\epsilon}}$ is a constant vector and then we can apply the Wormald's theorem in Appendix B: Wormald's theorem. In the following define $s_{\tau}(\tau', x)$ as the solution of the ODEs,

$$\frac{d}{d\tau} s_{\tau} = h(\tau, s_{\tau}), \tag{116}$$

with initial condition at τ', $s_{\tau'} = x := (x^{\alpha})_{\alpha\in\Gamma^{\epsilon}}$.

In what follows define the points where any component of $\tilde{u}_{\tau}$ has a step change. $\tau_l := \inf\{\tau > \tau_{l-1} : u_{\tau}^{\beta} \text{ has a step change for some } \beta \in \Phi^{\epsilon}\} \wedge \hat{\lambda}$ for $l \geq 1$ and $\tau_0 := 0$. Also let $k_l = \lceil n\tau_l \rceil$, where $\lceil\cdot\rceil$ is the ceiling function. As a result, $k_l - 1 < n\tau_l \leq k_l$. Recall the initial condition $s_0 = (s_0^{\alpha})_{\alpha\in\Gamma^{\epsilon}}$ with $s_0^{i,j,c,l} = p(i,j,c)\mathbb{1}_{(l=0)}$. Because every u^{β} for $\beta \in \Phi^{\epsilon}$ has only a finite number of step changes on $[0, \lambda)$ and Φ^{ϵ} is a finite set, there are in total a finite number of step changes for all the component functions of $\tilde{u}$ on $[0, \lambda)$.

Then by the Wormald's theorem, let $\rho = n^{-\frac{1}{4}}$, it follows that

$$\sup_{0\leq k\leq k_1-1} \frac{S_k^{\alpha}}{n} - s_{\frac{k}{n}}^{\alpha}(0, s_0) = O(n^{-\frac{1}{4}}) \tag{117}$$

with probability $1 - O\left(n^{\frac{1}{4}} \exp\left(-n^{\frac{1}{4}}\right)\right)$, $\forall\alpha \in \Gamma^{\epsilon}$. Note that we will write "with probability $1 - O\left(n^{\frac{1}{4}} \exp\left(-n^{\frac{1}{4}}\right)\right)$" as whp hereinafter.

In particular, we have that

$$\frac{S_{k_1-1}^{\alpha}}{n} - s_{\frac{k_1-1}{n}}^{\alpha}(0, s_0) = O(n^{-\frac{1}{4}}) \text{ whp.} \tag{118}$$

Additionally by the Wormald's theorem again we have that

$$\sup_{k_1\leq k\leq k_2-1} \frac{S_k^{\alpha}}{n} - s_{\frac{k}{n}}^{\alpha}\left(\frac{k_1}{n}, \frac{S_{k_1}}{n}\right) = O(n^{-\frac{1}{4}}) \text{ whp.} \tag{119}$$

Note that

$$\left|\frac{S_{k_1}^{\alpha}}{n} - \frac{S_{k_1-1}^{\alpha}}{n}\right| \leq \frac{1}{n} \qquad \forall\alpha \in \Gamma^{\epsilon}, \tag{120}$$

and by the Lipschitz continuity of $s^{\alpha}_{\tau}(0, s_0)$ on $(0, \tau_1^-)$,

$$s^{\alpha}_{\frac{k_1-1}{n}}(0, s_0) - s^{\alpha}_{\tau_1}(0, s_0) = O(n^{-1}). \tag{121}$$

So by Eqs (118), (120) and (121), we have

$$\begin{aligned} \left|\frac{S^{\alpha}_{k_1}}{n} - s^{\alpha}_{\tau_1}(0, s_0)\right| \leq & \left|\frac{S^{\alpha}_{k_1}}{n} - \frac{S^{\alpha}_{k_1-1}}{n}\right| + \left|\frac{S^{\alpha}_{k_1-1}}{n} - s^{\alpha}_{\frac{k_1-1}{n}}(0, s_0)\right| \\ & + \left|s^{\alpha}_{\frac{k_1-1}{n}}(0, s_0) - s^{\alpha}_{\tau_1}(0, s_0)\right| \\ = & \; n^{-1} + O\left(n^{-\frac{1}{4}}\right) + O(n^{-1}) \;\; \text{whp.} \end{aligned} \tag{122}$$

Thus we have that

$$\left\|\frac{S_{k_1}}{n} - s_{\tau_1}(0, s_0)\right\| = O\left(n^{-\frac{1}{4}}\right) + O(n^{-1}) \;\; \text{whp.} \tag{123}$$

where $\|\eta\|$ is the norm for $\eta \in \mathbb{R}^{|\Gamma^{\epsilon}|}$. We do not specify the norm because all norms in $\mathbb{R}^l$ are equivalent, $l \in \mathbb{N}$. From proposition 2 we see that the partial derivatives of $s^{\alpha}_{\tau}(\tau', x)$ with respect to the initial time τ' and every entry of x are continuous in τ' and every entry of x respectively, and are bounded for any τ in a subinterval of $[0, \hat{\lambda})$ on which $\tilde{u}$ is a constant vector function, i.e.

$$\left\|\frac{\partial s^{\alpha}_{\tau}(\tau', x)}{\partial(\tau', x)}\right\| \leq M_1 < \infty \tag{124}$$

where M_1 is a constant. Recall that $|\frac{k_1}{n} - \tau_1| < n^{-1}$, so by Eqs (123) and (124), it follows from the fundamentals of calculus (e.g. theorem 9.19 and 9.21 in [31]) that

$$\begin{aligned} & s^{\alpha}_{\tau}(\frac{k_1}{n}, \frac{S_{k_1}}{n}) - s^{\alpha}_{\tau}(\tau_1, s_{\tau_1}(0, s_0)) \\ = & \; s^{\alpha}_{\tau}(\frac{k_1}{n}, \frac{S_{k_1}}{n}) - s^{\alpha}_{\tau}(0, s_0) \\ = & \; O\left(n^{-\frac{1}{4}}\right) + O(n^{-1}) \;\; \text{whp,} \end{aligned} \tag{125}$$

for $\tau \in (\tau_1, \tau_2)$. So it follows from Eq (119) that $\forall \alpha \in \Gamma^{\epsilon}$,

$$\sup_{k_1 \leq k \leq k_2 - 1} \frac{S^{\alpha}_{k}}{n} - s^{\alpha}_{\frac{k}{n}}(0, s_0) = O\left(n^{-\frac{1}{4}}\right) \;\; \text{whp.} \tag{126}$$

Similarly for R_k, define $r_{\tau}(\tau', x, y)$ as the solution of

$$\frac{d}{d\tau} r_{\tau} = h_0(\tau, s_{\tau}), \tag{127}$$

with the initial condition at τ', $(s_{\tau'}, r_{\tau'}) = (x, y)$. Applying the Wormald's theorem for R_k and r_{τ}

gives that,

$$\begin{aligned} \sup_{0\le k\le k_1-1} \frac{R_k}{n} - r_{\frac{k}{n}}(0,s_0,0) &= O\left(n^{-\frac{1}{4}}\right) \text{ whp,} \\ \sup_{k_1\le k\le k_2-1} \frac{R_k}{n} - r_{\frac{k}{n}}(\frac{k_1}{n},\frac{S_{k_1}}{n},\frac{R_{k_1}}{n}) &= O\left(n^{-\frac{1}{4}}\right) \text{ whp.} \end{aligned} \tag{128}$$

In particular,

$$\frac{R_{k_1-1}}{n} - r_{\frac{k_1-1}{n}}(0,s_0,0) = O\left(n^{-\frac{1}{4}}\right) \text{ whp,} \tag{129}$$

Further note that

$$\left|\frac{R_{k_1}}{n} - \frac{R_{k_1-1}}{n}\right| \le \frac{1}{n} \qquad \forall \alpha \in \Gamma^{\epsilon}, \tag{130}$$

and by the Lipschitz continuity of $r_\tau(0, s_0, 0)$ on $(0, \tau_1^-)$,

$$r_{\frac{k_1-1}{n}}(0,s_0,0) - r_{\tau_1}(0,s_0,0) = O(n^{-1}). \tag{131}$$

So by Eqs (129), (130) and (131) we have

$$\begin{aligned} \left|\frac{R_{k_1}}{n} - r_{\tau_1}(0,s_0,0)\right| \le & \left|\frac{R_{k_1}}{n} - \frac{R_{k_1-1}}{n}\right| + \left|\frac{R_{k_1-1}}{n} - r_{\frac{k_1-1}{n}}(0,s_0,0)\right| \\ & + \left|r_{\frac{k_1-1}{n}}(0,s_0,0) - r_{\tau_1}(0,s_0,0)\right| \\ = & \; n^{-1} + O\left(n^{-\frac{1}{4}}\right) + O(n^{-1}) \text{ whp.} \end{aligned} \tag{132}$$

Here we apply the fact we shall prove later that the partial derivatives of $r_\tau(\tau', x, y)$ with respect to the initial time τ' and every entry of x and y are continuous in τ', every entry of x and y respectively, and are bounded for any τ in a subinterval of $[0, \hat{\lambda})$ on which $\tilde{u}$ is a constant vector function, i.e.

$$\left\|\frac{\partial r_\tau(\tau',x,y)}{\partial(\tau',x,y)}\right\| \le M_2 < \infty \tag{133}$$

for some constant M_2. Recall that $|\frac{k_1}{n} - \tau_1| < n^{-1}$, so by Eqs (123), (132) and (133), it follows from the fundamentals of calculus that

$$\begin{aligned} & r_\tau(\frac{k_1}{n},\frac{S_{k_1}}{n},\frac{R_{k_1}}{n}) - r_\tau(\tau_1, s_{\tau_1}(0,s_0), r_{\tau_1}(0,s_0,0)) \\ = & \; r_\tau(\frac{k_1}{n},\frac{S_{k_1}}{n},\frac{R_{k_1}}{n}) - r_\tau(0,s_0,0) \\ = & \; O\left(n^{-\frac{1}{4}}\right) + O(n^{-1}) \text{ whp,} \end{aligned} \tag{134}$$

for $\tau \in (\tau_1, \tau_2)$. So it follows from Eq (128) that

$$\sup_{k_1\le k\le k_2-1} \frac{R_k}{n} - r_{\frac{k}{n}}(0,s_0,0) = O\left(n^{-\frac{1}{4}}\right) \text{ whp.} \tag{135}$$

We can repeat the above procedure every time any u_τ^β has a step change, $\beta \in \Phi^\epsilon$ and there are only a finite number of step changes in $[0, \lambda)$. Because $s_\tau^\alpha \leq 1$ and $r_\tau \leq \lambda$, $d_\infty((s_\tau, r_\tau), \partial U) \geq 0.1 \geq Cn^{-\frac{1}{4}}$, for a sufficiently large constant C. Thus the supremum of τ that (s_τ, r_τ) can be extended to the boundary of U is $\hat{\lambda}$, i.e. in Eq (177) of the Wormald's theorem in Appendix B: Wormald's theorem,

$$\begin{aligned} \sigma &= \sup\left\{\tau \geq 0 : d_\infty((s_\tau, r_\tau), \partial U) \geq Cn^{-\frac{1}{4}}\right\} \\ &= \hat{\lambda}. \end{aligned} \tag{136}$$

So it follows that

$$\begin{aligned} \sup_{0 \leq k \leq n\hat{\lambda}} \frac{S_k^\alpha}{n} - s_{\frac{k}{n}}^\alpha(0, s_0) &= O\left(n^{-\frac{1}{4}}\right) \text{ whp}, \\ \sup_{0 \leq k \leq n\hat{\lambda}} \frac{R_k}{n} - r_{\frac{k}{n}}(0, s_0, 0) &= O\left(n^{-\frac{1}{4}}\right) \text{ whp}. \end{aligned} \tag{137}$$

At last we prove the claim that the partial derivatives of $r_\tau(\tau', x, y)$ with respect to the initial time τ' and every entry of x and y are all continuous and bounded as in Eq (133) for any τ in a subinterval of $[0, \hat{\lambda})$ on which $\tilde{u}$ is a constant vector function $b = (b^\beta)_{\beta \in \Phi^\epsilon}$. Note first that r_τ with initial condition $\bar{s} = (s_{\tau'}, r_{\tau'})$ at $\tau = \tau'$ in a subinterval of $[0, \hat{\lambda})$ on which $\tilde{u} = b$ satisfies that

$$r_\tau = r_{\tau'} + \int_{\tau'}^{\tau} \sum_{(i,j,c,c-1) \in \Phi^\epsilon} \frac{(i - c + 1) b^{i,j,c,c-1}}{\lambda - y} s_y^{i,j,c,c-1}(\tau', s_{\tau'}) dy. \tag{138}$$

We shall prove the boundedness by showing the boundedness of $\| \frac{\partial r_\tau}{\partial \bar{s}} \|$ and $| \frac{\partial r_\tau}{\partial \tau'} |$, seperately. First we take the derivatives of r_τ with respect to the initial condition $\bar{s}$ and obtain

$$\frac{\partial r_\tau}{\partial \bar{s}} = e^{\text{last}} + \int_{\tau'}^{\tau} \sum_{(i,j,c,c-1) \in \Phi^\epsilon} \frac{(i - c + 1) b^{i,j,c,c-1}}{\lambda - y} \frac{\partial s_y^{i,j,c,c-1}(\tau', s_{\tau'})}{\partial \bar{s}} dy, \tag{139}$$

where e^{last} is a vector of zeros except an entry of one at the last. The continuity of every entry in $\frac{\partial r_\tau}{\partial \bar{s}}$ is obvious. For boundedness,

$$\left\| \frac{\partial r_\tau}{\partial \bar{s}} \right\| \leq 1 + \int_{\tau'}^{\tau} \sum_{(i,j,c,c-1) \in \Phi^\epsilon} \frac{(i - c + 1) b^{i,j,c,c-1}}{\lambda - y} \left\| \frac{\partial s_y^{i,j,c,c-1}(\tau', s_{\tau'})}{\partial \bar{s}} \right\| dy. \tag{140}$$

By Eq (124), $\left\| \frac{\partial s_y^{i,j,c,c-1}}{\partial \bar{s}} \right\| < M_1$, so $\| \frac{\partial r_\tau}{\partial \bar{s}} \|$ is bounded. Next we take the derivative of r_τ with respect to the initial time τ' by the Leibniz integral rule and obtain that

$$\begin{aligned} \frac{\partial r_\tau}{\partial \tau'} = & - \sum_{(i,j,c,c-1) \in \Phi^\epsilon} \frac{(i - c + 1) b^{i,j,c,c-1}}{\lambda - \tau'} s_{\tau'}^{i,j,c,c-1}(\tau', s_{\tau'}) \\ & + \int_{\tau'}^{\tau} \sum_{(i,j,c,c-1) \in \Phi^\epsilon} \frac{(i - c + 1) b^{i,j,c,c-1}}{\lambda - y} \frac{\partial s_y^{i,j,c,c-1}(\tau', s_{\tau'})}{\partial \tau'} dy, \end{aligned} \tag{141}$$

where $s_{\tau'}^{i,j,c,c-1}(\tau', s_{\tau'}) = s_{\tau'}^{i,j,c,c-1}$. The continuity of $\frac{\partial r_\tau}{\partial \tau'}$ follows. By Eq (124), $\left| \frac{\partial s_y^{i,j,c,c-1}(\tau', s_{\tau'})}{\partial \tau'} \right|$ is bounded, so $| \frac{\partial r_\tau}{\partial \tau'} |$ is bounded. We have proved Eq (133).

Proof of proposition 4

Proof. For some $[\tau_1, \tau_2) \subseteq [0, \lambda)$ on which u_τ is a constant vector function, by remark 5 we have for every fixed (i, j) pair and $\Gamma_{i,j} = \{(c, l):0 \leq l < c \leq i \text{ or } c = i + 1, l = i\}$ that

$$\sum_{(c,l)\in\Gamma_{i,j}} s_\tau^{i,j,c,l} \leq \sum_{1\leq c\leq i} p(i, j, c), \tag{142}$$

and thus it follows from Eq (23) that

$$\begin{aligned} 0 \leq & \sum_{i\vee j\geq M^\epsilon}\sum_{0\leq c\leq i} jp(i, j, c) - \sum_{(i,j,c,l)\in\Gamma\setminus\Gamma^\epsilon} js_\tau^{i,j,c,l} \\ \leq & \sum_{i\vee j\geq M^\epsilon}\sum_{0\leq c\leq i} jp(i, j, c) < \epsilon. \end{aligned} \tag{143}$$

Similarly because by the definition of $S_k^{i,j,c,l}$ for $1 \leq k \leq m$, for fixed (i, j) pair, $(i, j, c, l) \in \Gamma$,

$$0 \leq \sum_{c,l} \frac{S_k^{i,j,c,l}}{n} \leq \sum_{1\leq c\leq i} P_n(i, j, c), \tag{144}$$

thus it follows from Eq (25) that

$$\begin{aligned} 0 \leq & \sum_{i\vee j\geq M^\epsilon}\sum_{0\leq c\leq i} jP_n(i, j, c) - \sum_{(i,j,c,l)\in\Gamma\setminus\Gamma^\epsilon} j\frac{S_k^{i,j,c,l}}{n} \\ \leq & \sum_{i\vee j\geq M^\epsilon}\sum_{0\leq c\leq i} jP_n(i, j, c) < \epsilon. \end{aligned} \tag{145}$$

For any k where $0 \leq k \leq \hat{\lambda}$, by proposition 3, it follows that

$$\begin{aligned} \left|\frac{D_k^-}{n} - d_{\frac{k}{n}}^-\right| < & \left|\sum_{i\vee j<M^\epsilon, 0\leq c\leq i} jP_n(i, j, c) - \sum_{(i,j,c,l)\in\Gamma^\epsilon} j\frac{S_k^{i,j,c,l}}{n}\right. \\ & \left. -\left(\sum_{i\vee j<M^\epsilon, 0\leq c\leq i} jp(i, j, c) - \sum_{(i,j,c,l)\in\Gamma^\epsilon} js_{\frac{k}{n}}^{i,j,c,l}\right)\right| + 2\epsilon \\ = & \left|\sum_{i\vee j<M^\epsilon, 0\leq c\leq i} j(P_n(i, j, c) - p(i, j, c))\right. \\ & \left. - \sum_{(i,j,c,l)\in\Gamma^\epsilon} j\left(\frac{S_k^{i,j,c,l}}{n} - s_{\frac{k}{n}}^{i,j,c,l}\right)\right| + 2\epsilon \\ \leq & \sum_{i\vee j<M^\epsilon, 0\leq c\leq i} j|P_n(i, j, c) - p(i, j, c)| \\ & + \sum_{(i,j,c,l)\in\Gamma^\epsilon} j\left|\frac{S_k^{i,j,c,l}}{n} - s_{\frac{k}{n}}^{i,j,c,l}\right| + 2\epsilon \\ \leq & M^\epsilon|\Gamma^\epsilon|(o(1) + o_p(1)) + 2\epsilon \\ = & o_p(1) + 2\epsilon, \end{aligned} \tag{146}$$

and similarly,

$$\begin{aligned}
\left|\frac{D_k}{n} - d_{\frac{k}{n}}\right| < & \left|\sum_{i\vee j<M^\epsilon, 0\le c\le i} P_n(i,j,c) - \sum_{(i,j,c,l)\in\Gamma^\epsilon} \frac{S_k^{i,j,c,l}}{n}\right. \\
& \left. -\left(\sum_{i\vee j<M^\epsilon, 0\le c\le i} p(i,j,c) - \sum_{(i,j,c,l)\in\Gamma^\epsilon} s_{\frac{k}{n}}^{i,j,c,l}\right)\right| + 2\epsilon \\
= & \left|\sum_{i\vee j<M^\epsilon, 0\le c\le i} (P_n(i,j,c) - p(i,j,c))\right. \\
& \left. - \sum_{(i,j,c,l)\in\Gamma^\epsilon}\left(\frac{S_k^{i,j,c,l}}{n} - s_{\frac{k}{n}}^{i,j,c,l}\right)\right| + 2\epsilon \\
\le & \sum_{i\vee j<M^\epsilon, 0\le c\le i} |P_n(i,j,c) - p(i,j,c)| \\
& + \sum_{(i,j,c,l)\in\Gamma^\epsilon}\left|\frac{S_k^{i,j,c,l}}{n} - s_{\frac{k}{n}}^{i,j,c,l}\right| + 2\epsilon \\
\le & |\Gamma^\epsilon|(o(1) + o_p(1)) + 2\epsilon \\
= & o_p(1) + 2\epsilon.
\end{aligned} \tag{147}$$

Proof of proposition 5

Proof. By Eq (25) for n large enough and $1 \le k \le m$, we have

$$\begin{aligned}
& \frac{1}{n}\sum_{\ell=1}^{k}\sum_{i\vee j\ge M^\epsilon}\sum_{1\le c\le i} \mathbb{1}_{(W_\ell=(i,j,c,c-1))} u_{\frac{\ell-1}{n}}^{i,j,c,c-1} \\
\le & \frac{1}{n}\sum_{i\vee j\ge M^\epsilon}\sum_{1\le c\le i} i\, n P_n(i,j,c) \\
\le & \sum_{i\vee j\ge M^\epsilon, c} i P_n(i,j,c) < \epsilon.
\end{aligned} \tag{148}$$

The first inequality holds because the number of times nodes with states in the range $i \vee j \ge M^\epsilon$, $1 \le c \le i$ are selected during the process is bounded above by their total in degree. Similarly by Eq (23), for $\tau \le \hat{\lambda}$,

$$\begin{aligned}
& \int_0^\tau \sum_{i\vee j\ge M^\epsilon, 1\le c\le i} \frac{(i-c+1)s_t^{i,j,c,c-1}}{\lambda - t} u_t^{i,j,c,c-1} dt \\
\le & \int_0^\tau \sum_{i\vee j\ge M^\epsilon, 1\le c\le i} \frac{ip(i,j,c)}{\lambda - t} dt \\
\le & \epsilon \int_0^\tau \frac{1}{\lambda - t} dt \\
= & \epsilon \ln \frac{\lambda}{\lambda - \tau} \\
\le & \epsilon \ln \frac{\lambda}{\epsilon} = O(\epsilon).
\end{aligned} \tag{149}$$

For any k where $0 \le k \le n\hat{\lambda}$, by proposition 3 it follows that

$$\begin{aligned}\left|\frac{R_k}{n} - r_{\frac{k}{n}}\right| \le & \ |\frac{\tilde{R}_k}{n} + \frac{1}{n}\sum_{\ell=1}^{k}\sum_{i\vee j\ge M^\epsilon}\sum_{1\le c\le i}\mathbb{1}_{(W_\ell=(i,j,c,c-1))}u^{i,j,c,c-1}_{\frac{\ell-1}{n}} \\ & -(\tilde{r}_{\frac{k}{n}} + \int_0^{\frac{k}{n}}\sum_{i\vee j\ge M^\epsilon,1\le c\le i}\frac{(i-c+1)s_t^{i,j,c,c-1}}{\lambda - t}u_t^{i,j,c,c-1}dt)| \\ \le & \ \left|\frac{\tilde{R}_k}{n} - \tilde{r}_{\frac{k}{n}}\right| + O(\epsilon) \\ \le & \ o_p(1) + O(\epsilon),\end{aligned} \tag{150}$$

thus we have that

$$\sup_{0\le k\le n\hat{\lambda}}|\frac{R_k}{n} - r_{\frac{k}{n}}| = o_p(1) + O(\epsilon). \tag{151}$$

If $\tau_f = \lambda$, it implies that $d_\tau^- > 0$ for $\tau \in (0, \hat{\lambda})$, then it follows from proposition 4 that $\frac{T_n}{n} = \hat{\lambda} + O(\epsilon) + o_p(1)$. Then because at each step there is at most one more node defaulting, $\frac{D_{T_n}}{n} = \frac{D_{\lfloor n\hat{\lambda}\rfloor}}{n} + O(\epsilon) + o_p(1)$ and from proposition 4 again, $\frac{D_{\lfloor n\hat{\lambda}\rfloor}}{n} = d_{\hat{\lambda}} + O(\epsilon) + o_p(1)$. $\lfloor\cdot\rfloor$ denotes the floor function. Further by the continuity of d_τ on $[0, \lambda]$, $\frac{D_{T_n}}{n} = d_\lambda + O(\epsilon) + o_p(1)$. Similarly, by Eq (151) and the continuity of r_τ on $[0, \lambda]$, we have that $\frac{R_{T_n}}{n} = r_\lambda + O(\epsilon) + o_p(1)$.

If $\tau_f < \lambda$ and $\frac{d}{d\tau}d^-_{\tau_f} < 0$, by definition 4, s_τ is continuous and thus by Eq (33) d_τ^- is also continuous. So there exists some $\tau' > 0$ such that $d_\tau^- < 0$ for $\tau \in (\tau_f, \tau_f + \tau')$ by the continuity of d_τ^-. Since ϵ is arbitrary, let ϵ be small enough such that $\inf_{\tau\in(\tau_f,\tau_f+\tau')}d_\tau^- < -2\epsilon$ and $\hat{\tau}$ be the first time d_τ^- reaches the minimum. Because $d_{\hat{\tau}}^- < -2\epsilon$, then by proposition 4 $\frac{D^-_{\lfloor n\hat{\tau}\rfloor}}{n} < 0$ with high probability, so it holds that $\frac{T_n}{n} = \tau_f + O(\epsilon) + o_p(1)$. Again by the continuity of d_τ and r_τ on $[0, \lambda]$, proposition 4 and Eq (151), $\frac{D_{T_n}}{n} = d_{\tau_f} + O(\epsilon) + o_p(1)$ and $\frac{R_{T_n}}{n} = r_{\tau_f} + O(\epsilon) + o_p(1)$.

In both cases we conclude that Eq (35) holds by tending $\epsilon \to 0$.

To prove Eq (37), since $R_{T_n} \le m \le (\lambda + 0.1)n$ for large n and $D_{T_n} \le n$, $\frac{R_{T_n}(G_n,P_n)}{n}$ and $\frac{D_{T_n}(G_n,P_n)}{n}$ are bounded and thus uniformly integrable. For a sequence of uniformly integrable random variables, convergence in probability implies convergence in expectation. Therefore Eq (37) holds.

Proof of lemma 1

Proof. Solve (FOCP) for the optimal $(\tilde{u}, \tilde{\tau}_f)$. Note that $\tilde{u}^\beta = 0$ for $\beta \in \Phi\backslash\Phi^\epsilon$. If there exists some $p(i, j, c)>0$, $i \vee j \ge M^\epsilon$, $0 \le c \le i$, then by remark 5, at $\tilde{\tau}_f$ by summing over (i, j) pairs satisfying $i \vee j \ge M^\epsilon$ we can show that

$$\sum_{i\vee j\ge M^\epsilon,0\le c\le i}p(i,j,c) - \sum_{(i,j,c,l)\in\Gamma\backslash\Gamma^\epsilon}s^{i,j,c,l}_{\tilde{\tau}_f} \ge 0, \tag{152}$$

and by the definition of $\tilde{\tau}_f$ that

$$\begin{aligned} d^-_{\tilde{\tau}_f} = & \sum_{i\vee j\geq M^\epsilon, 0\leq c\leq i} jp(i,j,c) - \sum_{(i,j,c,l)\in\Gamma\setminus\Gamma^\epsilon} js^{i,j,c,l}_{\tilde{\tau}_f} \\ & + \sum_{i\vee j< M^\epsilon, 0\leq c\leq i} jp(i,j,c) - \sum_{(i,j,c,l)\in\Gamma^\epsilon} js^{i,j,c,l}_{\tilde{\tau}_f} - \tilde{\tau}_f \\ = & \sum_{i\vee j\geq M^\epsilon, 0\leq c\leq i} jp(i,j,c) - \sum_{(i,j,c,l)\in\Gamma\setminus\Gamma^\epsilon} js^{i,j,c,l}_{\tilde{\tau}_f} \geq 0. \end{aligned} \tag{153}$$

Now we construct a function u as $u_\tau = \tilde{u}_\tau$, $\tau \leq \tilde{\tau}_f$ and $u^\beta_\tau = 1$, $\tilde{\tau}_f < \tau$, $\beta \in \Phi$. Note that under u, there are always interventions after $\tilde{\tau}_f$, thus by remark 5, for a fixed (i, j) pair with the set $\Gamma_{i,j} = \{(c, l):0 \leq l < c \leq i \text{ or } c = i + 1, l = i\}$, $\sum_{(c,l)\in\Gamma_{i,j}} s^{i,j,c,l}_\tau$ will not change, i.e. $\sum_{(c,l)\in\Gamma_{i,j}} s^{i,j,c,l}_\tau = \sum_{(c,l)\in\Gamma_{i,j}} s^{i,j,c,l}_{\tilde{\tau}_f}$ for $\tau > \tilde{\tau}_f$. Let $\tau_f$be the solution of $d^-_{\tau_f} = 0$ under u, then it follows that

$$\begin{aligned} d_{\tau_f} - \tilde{d}_{\tilde{\tau}_f} = & \sum_{i,j,0\leq c\leq i} p(i,j,c) - \sum_{(i,j,c,l)\in\Gamma} s^{i,j,c,l}_{\tau_f} \\ & -(\sum_{i\vee j< M^\epsilon, 0\leq c\leq i} p(i,j,c) - \sum_{(i,j,c,l)\in\Gamma^\epsilon} s^{i,j,c,l}_{\tilde{\tau}_f}) \\ = & \sum_{i\vee j\geq M^\epsilon, 0\leq c\leq i} p(i,j,c) - (\sum_{(i,j,c,l)\in\Gamma} s^{i,j,c,l}_{\tilde{\tau}_f} - \sum_{(i,j,c,l)\in\Gamma^\epsilon} s^{i,j,c,l}_{\tilde{\tau}_f}) \\ = & \sum_{i\vee j\geq M^\epsilon, 0\leq c\leq i} p(i,j,c) - \sum_{(i,j,c,l)\in\Gamma\setminus\Gamma^\epsilon} s^{i,j,c,l}_{\tilde{\tau}_f} \\ \leq & \sum_{i\vee j\geq M^\epsilon, 0\leq c\leq i} p(i,j,c) < \epsilon, \end{aligned} \tag{154}$$

and similarly

$$\begin{aligned} r_{\tau_f} - \tilde{r}_{\tilde{\tau}_f} = & \sum_{i\vee j\geq M^\epsilon, 0\leq c\leq i} jp(i,j,c) - \sum_{(i,j,c,l)\in\Gamma\setminus\Gamma^\epsilon} js^{i,j,c,l}_{\tilde{\tau}_f} \\ \leq & \sum_{i\vee j\geq M^\epsilon, 0\leq c\leq i} jp(i,j,c) < \epsilon. \end{aligned} \tag{155}$$

By the definition of ζ and $\tilde{\zeta}$,

$$\zeta(u, \tau_f, p) \leq \tilde{\zeta}(\tilde{u}, \tilde{\tau}_f, p) + (K + 1)\epsilon. \tag{156}$$

Recall (u^*, τ^*_f) is the optimal solution for the infinite dimensional Eq (39). By remark 6, (FOCP) assumes that the high degree nodes are invulnerable and because the $(\tilde{u}, \tilde{\tau}_f)$ is the optimal solution for (FOCP), it provides the lower bound for the optimal objective function of the infinite dimensional Eq (39), i.e.

$$\tilde{\zeta}(\tilde{u}, \tilde{\tau}_f, p) \leq \zeta(u^*, \tau^*_f, p). \tag{157}$$

Let the objective function be $\zeta(u, \tau_f, p)$ under u, then by the optimality of (u^*, τ^*_f), we have that

$$\zeta(u^*, \tau^*_f, p) \leq \zeta(u, \tau_f, p). \tag{158}$$

In sum, we have that

$$\tilde{\zeta}(\tilde{u}, \tilde{\tau}_f, p) \leq \zeta(u^*, \tau_f^*, p) \leq \zeta(u, \tau_f, p) \leq \tilde{\zeta}(\tilde{u}, \tilde{\tau}_f, p) + (K+1)\epsilon. \tag{159}$$

Thus the conclusion follows.

Proof of proposition 6

Proof. We apply the Extended Pontryagin Maximum Principle (EPMP) in Appendix C: Extended pontryagin maximum principle. First we present the correspondence of a notation A in EPMP and B in our application in the form $A \to B$.

$$\begin{aligned}
t &\to t,\\
t_0 &\to t_0,\\
t_f &\to t_f,\\
(x_t^i)_{i\in\{1,\ldots,n_x\}} &\to (s_t^\alpha)_{\alpha\in\Gamma^\epsilon},\\
(u_t^i)_{i\in\{1,\ldots,n_u\}} &\to (u_t^\beta)_{\beta\in\Phi^\epsilon},\\
U &\to \{0,1\},\\
\mathring{\lambda} &\to \mathring{w},\\
(\lambda_t^i)_{i\in\{1,\ldots,n_x\}} &\to (w_t^\alpha)_{\alpha\in\Gamma^\epsilon},\\
\ell(t, x_t, u_t) &\to K\sum_{i,j,1\leq c\leq i}(i-c+1)s_t^{i,j,c,c-1}u_t^{i,j,c,c-1},\\
\phi(t_f, x_{t_f}) &\to d_{t_f},\\
\psi_k(t_f, x_{t_f}) = 0, k = 1,\ldots,n_\psi &\to d_{t_f}^- = 0.
\end{aligned} \tag{160}$$

Let $(\mathring{w}, w_t)$ be the adjoint variables then $\mathring{w} = 1$, since otherwise the necessary conditions of optimality becomes independent of the cost functional in Eq (42). The Hamiltonian function Eq (48) is a direct result of Eq (183). Note that $n_\psi = 1$ and

$$\psi(t, s) = \sum_{i,j,0\leq c\leq i} jp(i,j,c) - \sum_{(i,j,c,l)\in\Gamma^\epsilon} js^{i,j,c,l} - \lambda(1 - e^{t_0 - t}). \tag{161}$$

Taking partial derivative yields $\frac{\partial}{\partial s}\psi(t_f, s_{t_f}) = (j, j, \ldots, j)$ which has rank 1.

Since the Hamiltonian function is affine in the control variable u_t, by condition (1) of EPMP, we attain that, for $1 \leq c \leq i$,

$$u_t^{i,j,c,c-1} = \begin{cases} 0 & \text{if } (K + w_t^{i,j,c+1,c})s_t^{i,j,c,c-1} > 0\\ 1 & \text{if } (K + w_t^{i,j,c+1,c})s_t^{i,j,c,c-1} < 0\\ 0 \text{ or } 1 & \text{if } (K + w_t^{i,j,c+1,c})s_t^{i,j,c,c-1} = 0. \end{cases} \tag{162}$$

By distinguishing the two cases $s_t^{i,j,c,c-1} > 0$ and $s_t^{i,j,c,c-1} = 0$, we have the equivalent form in Eq (50).

Taking partial derivative of $\mathcal{H}$ with regard to s yields the differential equations of w_t in condition (2). Note that $\mathcal{H}$ is autonomous, then according to condition (3) of EPMP, $\mathcal{H}(s_t, u_t, w_t)$ is a constant for $t \in [t_0, t_f]$, which is condition (3).

Then define

$$\begin{aligned}\Psi(t,s) := \quad & \sum_{i,j,0\le c\le i} p(i,j,c) - \sum_{(i,j,c,l)\in\Gamma^{\epsilon}} s^{i,j,c,l} \\ & +\nu\Big(\sum_{i,j,0\le c\le i} jp(i,j,c) - \sum_{(i,j,c,l)\in\Gamma^{\epsilon}} js_{t_f}^{i,j,c,l} - \lambda(1-e^{t_0-t_f})\Big)\end{aligned} \tag{163}$$

and taking partial derivatives with respect to s and t respectively by condition (4) of EPMP together with the terminal condition Eq (47) leads to condition (4).

Proof of theorem 1

Proof. For the contagion process without intervention, we relate our model to the auxiliary model used in the proof of theorem 3.8 in [24].

Recall that in Dynamics we are given a set of nodes $[n]$ and the degree sequence $(d^-(v), d^+(v))_{v\in[n]}$ as well as the initial equity levels $(e(v))_{v\in[n]}$ and the network is constructed sequentially by matching any out half-link from the default set to a uniformly chosen unconnected in half-link at every step. For each node v we assign each in half-link a number ranging in $\{1,\ldots, d^-(v)\}$. Let $\sum^v$ be the set of all permutations of the in half-links of node $v \in [n]$, then a permutation $\pi \in \sum^v$ specifies the order in which the in half-links are connected.

Because every in half-link of v represents one unit of loan, v will default after $e(v)$ of its in half-links have been connected (or $e(v)$ of its in links have been revealed) for every permutation $\pi \in \Sigma^v$. So the default threshold $\theta(v, \pi)$ for node v if the order in which the in half-links are connected is specified by π is $\theta(v, \pi) = e(v)$, $\forall\pi \in \Sigma^v$. Further our assumption 1 is equivalent to the assumption 4.1 and 4.2 in [24]. Moreover, under no intervention, the random graph generated in Dynamics conforms to the model defined in definition 5.4 in [24] with in and out degree sequences $(d^-(v), d^+(v))_{v\in[n]}$ and default thresholds $(e(v))_{v\in[n]}$ So by theorem 3.8 in [24] we achieve the conclusions of theorem 1.

Proof of theorem 2

Proof. To simplify the notations we suppress the apostrophe "*". In lemma 2 we have presented the optimal control policy $(u_t)_{t\in[t_0,tf]}$ in terms of $t,t_0,t_f,t_s,t^{i,j,c}$. Recall that in Eq (77) we have the following relations,

$$\begin{aligned} y &= 1-e^{t_0-t_f}, \\ z &= 1-e^{t_0-t_s}, \\ x^{i,j,c,c-1} &= 1-e^{t_0-t^{i,j,c}} \\ &= \begin{cases} y & \text{if } K+vj-1\ge 0 \text{ or } c=0 \\ 1-(1-y)\frac{(i-c)K}{(i-c+1)K+vj-1} & \text{if } K+vj-1<0 \\ & \text{and } 1\le c<i+\frac{K+vj-1}{Ky} \\ 0 & \text{otherwise,} \end{cases}\end{aligned} \tag{164}$$

as well as $t = -\ln(\lambda - \tau)$, $t_0 = -\ln\lambda$, so we can change the variable from t to τ. Particularly we

apply mapping $f(t) = 1 - e^{t_0 - t}$ which is strictly increasing in t, then we have

$$\begin{aligned} f(t) &= \frac{\tau}{\lambda}, \\ f(t_0) &= 0, \\ f(t^{i,j,c}) &= x^{i,j,c,c-1}, \\ f(t_s) &= z, \\ f(t_f) &= y. \end{aligned} \tag{165}$$

We replace each variable t, t_0, t_f, t_s, $t^{i,j,c}$ in lemma 2 with its corresponding variable in Eq (165) resulting in the expressions for $u_\tau^{i,j,c,c-1}$. At last by assumption 2 on the relations between the control policy $G_n = (g_1^{(n)}, \ldots, g_m^{(n)})$ and the function u, we have the conclusion in theorem 2.

Proof of theorem 3

Proof. In proposition 7 we have obtained the expressions for $d_{t_f}^-$ and d_{t_f} with $i \vee j < M^\epsilon$ in terms of (v, t_f, t_s), after change of variables to (v, y, z) with $y = \frac{\tau_f}{\lambda}$ we have the following expressions for $d_{\tau_f}^-$ and d_{τ_f} with their relations to $\tilde{I}(y; v, z)$ and $\tilde{J}(y; v, z)$ in definition 10.

$$\begin{aligned} d_{\tau_f}^- &= \sum_{i \vee j < M^\epsilon} j \Bigg[\sum_{c=0}^{i} p(i,j,c) \mathbb{P}(\mathrm{Bin}(i, x^{i,j,c,c-1}) \geq c) \\ &\quad - \mathbb{1}_{(vj-1=-K)} p(i,j,i) \left(\left(\frac{\tau_f}{\lambda} \right)^i - z^i \right) \Bigg] - \tau_f \\ &= \lambda \left(\tilde{I}\left(\frac{\tau_f}{\lambda}; v, z \right) - \frac{\tau_f}{\lambda} \right), \\ d_{\tau_f} &= \sum_{i \vee j < M^\epsilon} \Bigg[\sum_{c=0}^{i} p(i,j,c) \mathbb{P}(\mathrm{Bin}(i, x^{i,j,c,c-1}) \geq c) \\ &\quad - \mathbb{1}_{(vj-1=-K)} p(i,j,i) \left(\left(\frac{\tau_f}{\lambda} \right)^i - z^i \right) \Bigg] \\ &= \tilde{J}\left(\frac{\tau_f}{\lambda}; v, z \right). \end{aligned} \tag{166}$$

Suppose (y^*, v^*, z^*) is an optimal solution for the optimization problem Eq (86) and note that y^* is the smallest fixed point of $\tilde{I}(y; v^*, z^*)$ and $y^* = \frac{\tau_f^*}{\lambda}$.

1. If $y^* = 1$, then $\tau_f^* = \lambda$. By the definition of $d_{\tau_f}^-$, it can only occur when $\sum_{i \vee j < M^\epsilon} j \sum_{c=0}^{i} p(i,j,c) = \lambda$ and $z^* = \frac{\tau_f^*}{\lambda} = 1$, thus we have $d_{\tau_f^*} = d_\lambda = 1$, then by proposition 5,

$$\frac{D_{T_n}}{n} \xrightarrow{p} 1, \tag{167}$$

which proves (1) of theorem 3.

2. If $y^* < 1$ and $\tilde{I}'(y^*; v^*, z^*) < 1$, then $\tau_f^* < \lambda$ and $\frac{d}{d\tau} d^-_{\tau_f^*} = \tilde{I}'\left(\frac{\tau_f^*}{\lambda}; v^*, z^*\right) - 1 < 0$. Again it follows from proposition 5,

$$\frac{D_{T_n}}{n} \xrightarrow{p} d_{\tau_f^*} = \tilde{J}(y^*; v^*, z^*). \tag{168}$$

which proves (2) of theorem 3. This concludes the proof of theorem 3.

It is important to note that the two cases in theorem 3 corresponds to $\tau_f^* = \lambda$, and $\tau_f^* < \lambda$, $\frac{d}{d\tau} d^-_{\tau_f^*} < 0$, respectively. By proposition 5 they guarantees that the limits of $\mathbb{E}\frac{R_{T_n}(G_n,P_n)}{n}$ and $\mathbb{E}\frac{D_{T_n}(G_n,P_n)}{n}$ in Eq 9 as $n \to \infty$ are well defined, which are r_{τ_f} and d_{τ_f}, respectively.

Proof of lemma 4

Proof. In the following we suppress "$*$". Note first that if $v > 0$, $x^{i,j,c,c-1}$ is increasing in j. This implies that $x^{i,j_1,c,c-1} < x^{i,j_2,c,c-1}$ for the two states in Φ^ϵ, $(i, j_1, c, c-1)$ and $(i, j_2, c, c-1)$ where $j_1 < j_2$. By theorem 2, this further implies that at some step k such that $n\lambda x^{i,j_1,c,c-1} \le k \le n\lambda x^{i,j_2,c,c-1}$, we should intervene on a node in state $(i, j_1, c, c-1)$ when it is selected at k but not on a node in state $(i, j_2, c, c-1)$. But this control policy is not optimal because both nodes are the same except the out degree and the node in state $(i, j_2, c, c-1)$ is systematically more important.

Appendix B: Wormald's theorem

The following is from [30]. Let $a \ge 2$ be a fixed integer and $((Y_t^l)_{1\le l\le a})_{t\ge 0}$ denote a sequence of real valued random variables indexed by n with its natural filtration $(\mathcal{F}_t)_{t\ge 0}$. Assume that there is a constant $C_0 > 0$ such that $|Y_t^l| \le C_0 n$ for $\forall n, t \ge 0$ and $1 \le l \le a$. Let $f_l : \mathbb{R}^{a+1} \to \mathbb{R}$ be functions and $U \subseteq \mathbb{R}^{a+1}$ be some bounded connected open set containing the closure of

$$\{(0, z^1, \ldots, z^a) : \mathbb{P}(Y_0^l = z^l n,\ 1 \le l \le a) \ne 0 \text{ for some } n\}. \tag{169}$$

Define the stopping time $T_U = \inf\left\{t \ge 1 : \left(\frac{t}{n}, \frac{Y_t^1}{n}, \ldots, \frac{Y_t^a}{n}\right) \notin U\right\}$. Assume the following three conditions are satisfied:

1. (Boundedness) For some function $\rho_1 = \rho_1(n) \ge 1$ and $\forall t < T_U$ and $1 \le l \le a$,

$$|Y_{t+1}^l - Y_t^l| \le \rho_1. \tag{170}$$

2. (Trend) For some function $\rho_2 = \rho_2(n) = o(1)$ and $\forall t < T_U$ and $1 \le l \le a$,

$$\left|\mathbb{E}\left(Y_{t+1}^l - Y_t^l | \mathcal{F}_t\right) - f_l\left(\frac{t}{n}, \frac{Y_t^1}{n}, \ldots, \frac{Y_t^a}{n}\right)\right| \le \rho_2. \tag{171}$$

3. (Lipschitz continuity) The functions $(f_l)_{1\le l\le a}$ are continuous and satisfies a Lipschitz condition on

$$U \cap \{(t, z^1, \ldots, z^a) : t \ge 0\} \tag{172}$$

with the same Lipschitz constant for each l.

Then the following holds:

1. For $(0, \hat{z}^1, \ldots, \hat{z}^a) \in U$ the system of differential equations

$$\frac{dz^l}{ds} = f_l(s, z^1, \ldots, z^a), \qquad 1 \le l \le a \tag{173}$$

has a unique solution in U for $z^l : \mathbb{R} \to \mathbb{R}$ passing through

$$z_0^l = \hat{z}^l, \qquad 1 \le l \le a \tag{174}$$

and which extends to points arbitrarily close to the boundary of U.

2. Let $\rho > \rho_2$ and $\rho = o(1)$. For a sufficiently large constant C, with probability $1 - O\left(\frac{\rho_1}{\rho} \exp\left(-\frac{n\rho^3}{\rho_1^3}\right)\right)$, it holds that

$$\sup_{0 \le t \le n\sigma} \left(\frac{Y_t^l}{n} - z_{\frac{t}{n}}^l \right) = O(\rho) \tag{175}$$

where z_s^l is the solution in (1) with

$$z_0^l = \frac{Y_0^l}{n} \tag{176}$$

and

$$\sigma = \sigma(n) = \sup\{s \ge 0 : \ d_\infty(((z_s^l)_{1 \le l \le a}), \partial U) \ge C\rho\}, \tag{177}$$

where $d_\infty(u, v) = \max_{1 \le i \le j} |u_i - v_i|$ for $u = (u_1, \ldots, u_j) \in \mathbb{R}^j$ and $v = (v_1, \ldots, v_j) \in \mathbb{R}^j$.

Appendix C: Extended pontryagin maximum principle

The following is from [32]. Consider the optimal control problem to minimize the cost functional including a terminal term

$$\mathcal{J}(u, t_f) := \int_{t_0}^{t_f} \ell(t, x_t, u_t) dt + \phi(t_f, x_{t_f}), \tag{178}$$

with fixed initial time t_0 and free terminal time t_f, subject to the dynamical system

$$\dot{x}_t = f(t, x_t, u_t); x_{t_0} = x_0, \tag{179}$$

where the vector function $x \in \hat{C}^1[t_0, T]^{n_x}$ represents the state variables characterizing the behavior of the system at any time instant t, and some general terminal constraints

$$\psi_k(t_f, x_{t_f}) = 0, k = 1, \ldots, n_\psi. \tag{180}$$

The admissible controls shall be taken in the class of piecewise continuous functions

$$u \in \mathcal{U}[t_0, T] := \{u \in \hat{C}[t_0, T]^{n_u} : u_t \in U \text{ for } t_0 \le t \le t_f\}, \tag{181}$$

with $t_f \in [t_0, T]$, where $T > t_0$ and the nonempty, possibly closed and nonconvex set U denotes the control region.

Suppose ℓ and f are continuous and have continuous first partial derivatives with respect to (t, x, u) on $[t_0, T] \times \mathbb{R}^{n_x} \times \mathbb{R}^{n_u}$, and also ϕ and $\psi := (\psi_k)_{k=1,\ldots,n_\psi}$ are continuous and have continuous first partial derivatives with respect to (t, x) on $[t_0, T] \times \mathbb{R}^{n_x}$. Suppose that the terminal

constraints Eq (180) satisfy the constraint qualification

$$\operatorname{rank}\left(\frac{\partial\psi}{\partial x}(t_f^*, x_{t_f^*}^*)\right) = n_\psi \tag{182}$$

where $\frac{\partial\psi}{\partial x}(t_f^*, x_{t_f^*}^*)$ denotes the Jacobian matrix of the partial derivatives of components of ψ with respect to x evaluated at $(t_f^*, x_{t_f^*}^*)$. Define the Hamiltonian function

$$\mathcal{H}(t, x, u, \mathring{\lambda}, \lambda) = \mathring{\lambda}\ell(t, x, u) + \lambda^T f(t, x, u). \tag{183}$$

Let $(u^*, t_f^*) \in \hat{C}[t_0, T]^{n_u} \times [t_0, T)$ denote a minimizer for the problem, and $x^* \in \hat{C}^1[t_0, T]$ the optimal state, then there exists a n_x dimensional piecewise continuously differentiable vector function λ_t^* and $\mathring{\lambda}^* \in \{0, 1\}$ ($(\mathring{\lambda}^*, \lambda_t^*)$ are called adjoint variables) and a Lagrange multiplier vector $\nu^* \in \mathbb{R}^{n_\psi}$ such that $(\mathring{\lambda}^*, \lambda_t^*) \neq 0$ for every $t \in [t_0, t_f^*]$ and the following conditions hold:

1. The function $\mathcal{H}(t, x_t^*, w, \mathring{\lambda}^*, \lambda_t^*)$ attains its minimum on U at $w = u_t^*$ for every $t \in [t_0, t_f^*]$, i.e.

$$\mathcal{H}(t, x_t^*, w, \mathring{\lambda}^*, \lambda_t^*) \geq \mathcal{H}(t, x_t^*, u_t^*, \mathring{\lambda}^*, \lambda_t^*), \forall w \in U. \tag{184}$$

2. $(x_t^*, u_t^*, \mathring{\lambda}^*, \lambda_t^*)$ verifies the equations

$$\begin{aligned} \frac{d}{dt}x_t^* &= f(t, x_t^*, u_t^*), \\ \frac{d}{dt}\lambda_t^* &= -\frac{\partial}{\partial x}\mathcal{H}(t, x_t^*, u_t^*, \mathring{\lambda}^*, \lambda_t^*) \end{aligned} \tag{185}$$

   at each instant t of continuity of u^* and $\mathring{\lambda}^* \in \{0, 1\}$.

3. $\mathcal{H}(t, x_t^*, u_t^*, \mathring{\lambda}^*, \lambda_t^*) = \mathcal{H}(t_f^*, x_{t_f^*}^*, u_{t_f^*}^*, \mathring{\lambda}^*, \lambda_{t_f^*}^*) - \int_t^{t_f^*} \frac{\partial}{\partial t}\mathcal{H}(\tau, x_\tau^*, u_\tau^*, \mathring{\lambda}, \lambda_\tau^*)d\tau$. Therefore, if $\frac{\partial}{\partial t}\mathcal{H} = 0$, i.e. $\mathcal{H}$ is autonomous, then $\mathcal{H}$ is a constant over time.

4. (Transversal condition) Define $\Psi(t, x) := \mathring{\lambda}^*\phi(t, x) + \nu^{*T}\psi(t, x)$, then

$$\begin{aligned} \lambda_{t_f^*}^* &= \frac{\partial}{\partial x}\Psi(t_f^*, x_{t_f^*}^*), \\ \mathcal{H}(t_f^*, x_{t_f^*}^*, u_{t_f^*}^*, \mathring{\lambda}_{t_f^*}^*, \lambda_{t_f^*}^*) &= -\frac{\partial}{\partial t}\Psi(t_f^*, x_{t_f^*}^*) \end{aligned} \tag{186}$$

   together with the terminal condition Eq (180) at $t = t_f^*$, i.e. $\psi_k(t_f^*, x_{t_f^*}^*) = 0$ for $k = 1, \ldots, n_\psi$.

5. The optimal control u^* may or may not be continuous; in the latter case we have a corner point. In particular, the conditions that must hold at any corner point $\theta \in [t_0, t_f^*]$ are

$$\begin{aligned} x_{\theta^-}^* &= x_{\theta^+}^*, \\ \lambda_{\theta^-}^* &= \lambda_{\theta^+}^*, \\ \mathcal{H}(\theta^-, x_\theta^*, u_{\theta^-}^*, \mathring{\lambda}^*, \lambda_\theta^*) &= \mathcal{H}(\theta^+, x_\theta^*, u_{\theta^+}^*, \mathring{\lambda}^*, \lambda_\theta^*). \end{aligned} \tag{187}$$

Proof. See theorem 3.33 and theorem 3.34 in [32].

Appendix D: Preliminary list of notations

$\mathcal{D}_k$: the default set at step k;

$D_k = |\mathcal{D}_k|$;

D_{T_n}: the number of defaulted nodes by the end of the process T_n;

D_k^-: the number of unrevealed out links from the default set at step k;

$\mathcal{E}_n$: the set of links in a random network;

Gr_n: a graph on n nodes;

$G_{n,m}$: the set of networks on n nodes with at most m directed links;

$(\mathcal{G}_k)_{0\le k\le m}$: the filtration on the default contagion process;

$G_n = (g_1^{(n)}, \ldots, g_m^{(n)})$: a control policy for a network of size n;

$I(y)$,$J(y)$, $\tilde{I}(y, v, z)$, $\tilde{J}(y, v, z)$: special functions defined for theorem 1, theorem 2 and theorem 3;

K: the "cost" of an intervention relative to a defaulted node;

M^ϵ: an integer defined based on ϵ;

$\mathbb{N}_0 := \{0, 1, 2, \ldots\}$;

$\mathbb{N} := \{1, 2, \ldots\}$, the set of positive integers;

$\mathbb{P}$: probability measure on the set $G_{n,m}$;

$P_n(i, j, c)$: the empirical probability of the degrees and initial equity levels and $P_n = (P_n(i, j, c))_{i,j,0\le c\le i}$;

Q_k: the set of hidden out links from the default set at step k;

R_k: the accumulative number of interventions by step k;

R_{T_n}: the accumulative number of interventions by the end of the process T_n;

$S_k^{i,j,c,l}$: the number of nodes that are vulnerable initially and in state (i, j, c, l) at step k and $S_k := (S_k^\alpha)_{\alpha\in\Gamma}$;

T_n: the contagion process end time;

$\mathbb{U}$: the set of all $(\mathcal{G}_k)_{0\le k\le m}$ adapted process μ;

(V_k, W_k): a pair of random variables denoting the link from node V_k to node W_k revealed at step k; By abuse of notation, W_k denotes the state of the node selected at k when there is no confusion;

$b := (b^\beta)_{\beta\in\Phi}$: a vector of $\{0, 1\}$ constants;

c_k^v: the sum of initial equity and accumulative interventions of node v at step k;

$d^-(v)$,$d^+(v)$: the in and out degree of a node v;

d_τ^-: the asymptotic number of unrevealed out links of the default set at time τ;

d_τ: the asymptotic fraction of defaults at time τ;

$e(v)$: the initial equity of a node;

$g_k^{(n)}$: the control function at step k for a network of size n;

h: the set of ODEs of s_τ and s_t;

h': the set of ODEs of w_t;

l_k^v: the number of revealed in links of node v at step k;

$m = m(n)$: the total in or out degree of a network of size n, maybe index variable as well;

n: the number of nodes, may be index variable as well;

$[n] = \{1, \ldots, n\}$;

$u_\tau^{i,j,c,c-1}$: a piecewise constant function; $u_\tau = (u_\tau^\beta)_{\beta\in\Phi}$;

r_τ: the asymptotic scaled number of interventions by time τ;

v: the Lagrange multiplier;

(v_k, w_k): the link from node v_k to node w_k revealed at step k;
$w_t^{i,j,c,l}$: the adjoint variable;
ϵ: a positive real;
λ: the asymptotic mean of in (out) degree;
$\hat{\lambda} := \lambda - \in$;
μ_k: intervention at step k and $\mu = (\mu_k)_{1 \le k \le m}$;
μ_n: in Introduction denotes an intervention sequence for a network of n nodes;
τ_f: the end time in of the asymptotic process;
$\|\|$: any norm of $\mathbb{R}^l$ for some $l \in \mathbb{N}$;
$\lceil \cdot \rceil$: the ceiling function;
$[\cdot]$: the round function;
$\tilde{}$: the corresponding variable resulted by the constraint $i \vee j < M^\epsilon$, e.g. $\tilde{R}_k$, $\tilde{u}$, $\tilde{\tau}_f$, $\tilde{r}_\tau$, $\tilde{d}_\tau$;
Bin(i, y): a binomial random variable in i trials with the probability of occurrence y;
Multin$(i, x, y, 1 - x - y) = (a, b, i - a - b)$: a multinomial distribution in i trials with the probabilities of occurrence of each of three types being x, y and $1 - x - y$, and turns out to have a, b and $i - a - b$ occurrences of each type.

Important sets:

$$\begin{aligned}
\Phi &:= \{(i,j,c,c-1) : 0 \le i, 0 \le j, 1 \le c \le i\}, \\
\Phi^\epsilon &:= \{(i,j,c,c-1) : i \vee j < M^\epsilon, 1 \le c \le i\}. \\
\Gamma^+ &:= \{(i,j,c,l) : 0 \le i, 0 \le j, 0 \le c, 0 \le l \le i\}, \\
\Gamma &:= \{(i,j,c,l) : 0 \le i, 0 \le j, 0 \le l < c \le i \text{ or } c = i+1, l = i\}, \\
\Gamma^\epsilon &:= \{(i,j,c,l) : i \vee j < M^\epsilon, 0 \le l < c \le i \text{ or } c = i+1, l = i\}, \\
\Gamma_{i,j} &:= \{(c,l) : 0 \le l < c \le i \text{ or } c = i+1, l = i\}.
\end{aligned} \tag{188}$$

Supporting information

S1 Dataset. Dataset generated from the numerical experiments and based on which Figs 6–11 are presented. The spreadsheet contains two tables which includes respectively, under the optimal and alternative intervention policies the scaled number of interventions $\frac{R_{T_n}}{n}$, the scaled number of defaults $\frac{D_{T_n}}{n}$, the scaled process end time $\frac{T_n}{m}$ and the objective function value for the number of nodes $n = 5^4, 6^4, \ldots, 10^4$ and there are 100 runs for each n. The dataset is publicly available via https://figshare.com/articles/Simulation_result/7477562 with DOI 10.6084/m9.figshare.7477562.
(XLSX)

Acknowledgments

The author thanks Lisa R Goldberg, Ilan Adler, Mariana Olvera-Cravioto and Alexander D Shkolnik for their valuable comments and suggestions.

Author Contributions

Conceptualization: Yang Xu.

Formal analysis: Yang Xu.

Methodology: Yang Xu.

Resources: Yang Xu.

Validation: Yang Xu.

Visualization: Yang Xu.

Writing – original draft: Yang Xu.

References

1. Vollmer U. In: Backhaus J, editor. Lender of Last Resort. New York, NY: Springer; 2014. p. 1–6.
2. Amini H, Minca A, Sulem A. Control of interbank contagion under partial information. Journal of Financial Mathematics. 2015; 6(1):1195–1219. https://doi.org/10.1137/140981538
3. Amini H, Minca A, Sulem A. Optimal equity infusions in interbank networks. Journal of Financial Stability. 2017; 31:1–17. https://doi.org/10.1016/j.jfs.2017.05.008
4. Cont R, Moussa A, Santos EBe. Network Structure and Systemic Risk in Banking Systems. SSRN eLibrary. 2010; p. 327–368.
5. Hurd TR. Contagion! The Spread of Systemic Risk in Financial Networks; 2015.
6. Diamond DW, Dybvig PH. Bank Runs, Deposit Insurance, and Liquidity. Journal of Political Economy. 1983; 91(3):401–419. https://doi.org/10.1086/261155
7. Allen F, Gale D. Financial contagion. Journal of Political Economy. 2000; 1(108):1–33. https://doi.org/10.1086/262109
8. Freixas X, Parigi B, Rochet JC. Systemic Risk, Interbank Relations and Liquidity Provision by the Central Bank. Journal of Money, Credit and Banking. 2000; 3(3):611–638. https://doi.org/10.2307/2601198
9. Acharya VV, Yorulmazer T. Cash-in-the-market pricing and optimal resolution of bank failures. Review of Financial Studies. 2008; 21(6):2705–2742. https://doi.org/10.1093/rfs/hhm078
10. Gorton G, Huang L. Liquidity, efficiency, and bank bailouts. American Economic Review. 2004; 94 (3):455–483. https://doi.org/10.1257/0002828041464650
11. Philippon T, Schnabl P. Efficient Recapitalization. Journal of Finance. 2013; 68(1):1–42. https://doi.org/10.1111/j.1540-6261.2012.01793.x
12. Rogers LCG, Veraart LAM. Failure and Rescue in an Interbank Network. Management Science. 2013; 59(4):882–898. https://doi.org/10.1287/mnsc.1120.1569
13. Leitner Y. Financial networks: Contagion, commitment, and private sector bailouts. Journal of Finance. 2005; 60:2925–2953. https://doi.org/10.1111/j.1540-6261.2005.00821.x
14. Garcia-de Andoain C, Heider F, Hoerova M, Manganelli S. Lending-of-last-resort is as lending-of-last-resort does: Central bank liquidity provision and interbank market functioning in the euro area. Journal of Financial Intermediation. 2016; 28:32–47. https://doi.org/10.1016/j.jfi.2016.01.003
15. Afonso G, Kovner A, Schoar A. Stressed, not frozen: The federal funds market in the financial crisis. Journal of Finance. 2011; 66(4):1109–1139. https://doi.org/10.1111/j.1540-6261.2011.01670.x
16. Berger AN, Bouwman CHS, Kick T, Schaeck K. Bank liquidity creation following regulatory interventions and capital support. Journal of Financial Intermediation. 2016; 26:115–141. https://doi.org/10.1016/j.jfi.2016.01.001
17. Carlson MA, Duygan-Bump B, Nelson WR. Why Do We Need Both Liquidity Regulations and a Lender of Last Resort? A Perspective from Federal Reserve Lending During the 2007-09 U.S. Financial Crisis; 2015. Available from: http://www.ssrn.com/abstract=2573767.
18. Furfine CH. Interbank Exposures: Quantifying the Risk of Contagion. Journal of Money, Credit and Banking. 2003; 35(1):pp. 111–128. https://doi.org/10.1353/mcb.2003.0004
19. Boss M, Elsinger H. An empirical analysis of the network structure of the Austrian interbank market. Financial stability Report. 2004;(7):77–87.
20. May RM, Arinaminpathy N. Systemic risk: the dynamics of model banking systems. Journal of the Royal Society, Interface. 2010; 7(46):823–838. https://doi.org/10.1098/rsif.2009.0359 PMID: 19864264
21. Haldane AG, May RM. Systemic risk in banking ecosystems. Nature. 2011; 469:351–355 https://doi.org/10.1038/nature09659 PMID: 21248842
22. Gai P, Haldane A, Kapadia S. Complexity, concentration and contagion. Journal of Monetary Economics. 2011; 58:453–470 https://doi.org/10.1016/j.jmoneco.2011.05.005
23. Nier E, Yang J, Yorulmazer T, Alentorn A. Network models and financial stability. Journal of Economic Dynamics and Control. 2007; 31(6):2033–2060 https://doi.org/10.1016/j.jedc.2007.01.014

24. Amini H, Cont R, Minca A. Resilience to contagion in financial networks. Mathematical Finance. 2013; 00(0):1–37.
25. Eisenberg L, Noe TH. Systemic risk in financial systems. Management Science. 2001; 47(2):236–249 https://doi.org/10.1287/mnsc.47.2.236.9835
26. Chinazzi M, Fagiolo G. Systemic Risk, Contagion, and Financial Networks: A Survey. SSRN Electronic Journal. 2013 https://doi.org/10.2139/ssrn.2243504
27. van der Hofstad R. Random Graphs and Complex Networks; 2014.
28. Furfine CH. The Microstructure of the Federal Funds Market. Financial Markets, Institutions and Instruments. 1999; 8(5):24–44. https://doi.org/10.1111/1468-0416.00031
29. Bech ML, Atalay E. The topology of the federal funds market. Physica A: Statistical Mechanics and its Applications. 2010; 389(22):5223–5246. https://doi.org/10.1016/j.physa.2010.05.058
30. Wormald NC. The Differential Equation Method for Random Network Processes and Greedy Algorithms. In: Lectures on Approximation and Randomized Algorithms; 1999. p. 73–155.
31. Rudin W. Principles of Mathematical Analysis. 3rd ed. McGraw-Hill Education; 1976. Available from: http://onlinelibrary.wiley.com/doi/10.1002/cbdv.200490137/abstract.
32. Chachuat BC. Nonlinear and Dynamic Optimization: From Theory to Practice—IC-32: Spring Term 2009. EPFL; 2009.